

Missing Puzzle Pieces in the Performance Landscape of the Quantum Approximate Optimization Algorithm

Elisabeth Wybo and Martin Leib

IQM Quantum Computers, Georg-Brauchle-Ring 23-25, 80992 München, Germany

We consider the maximum cut and maximum independent set problems on random regular graphs in the infinite-size limit, and calculate the energy densities achieved by QAOA for high degrees up to $d = 100$. Such an analysis is possible because the reverse causal cones of the operators in the Hamiltonian are with high probability associated with tree subgraphs, for which efficient classical contraction schemes can be developed. We combine the QAOA analysis with state-of-the-art upper bounds on optimality for both problems. This yields novel and better bounds on the approximation ratios achieved by QAOA for large problem sizes. We show that the approximation ratios achieved by QAOA improve as the graph degree increases for the maximum cut problem. However, QAOA exhibits the opposite behavior for the maximum independent set problem, i.e. the achieved approximation ratios decrease when the degree of the problem is increased. This phenomenon is explainable by the overlap gap property for large d , which restricts local algorithms (like QAOA) from reaching near-optimal solutions with high probability. In addition, we use the QAOA parameters determined on the tree subgraphs for small graph instances, and in that way outperform classical algorithms like Goemans-Williamson for the maximum cut problem and minimal greedy for the maximum independent set problem. In this way we circumvent the parameter optimization problem and are able to compute the expected approximation ratios.

1 Introduction

The advent of quantum computing promises advantage over classical computing in areas as diverse as quantum simulation [1–3], machine learning [4–7] and optimization [8, 9]. To unequivocally profit from this however, fault tolerant operation of quantum computers is needed. In the current era of Noisy Intermediate Scale Quantum (NISQ) devices [10], it is not possible to run the deep quantum algorithms equipped with proofs of quantum advantage. Therefore, a new research topic of quantum heuristics with substantially reduced number of required gate operations is currently flourishing. The reduced number of gates for these heuristics comes typically at the cost of unclear prospects for quantum advantage caused by a lack of techniques for their complexity analysis. Compared to other quantum algorithms for the NISQ era, there is however already a lot known about performance and time-to-solution for the Quantum Approximate Optimization Algorithm (QAOA) [11–17] which the current work is adding a missing puzzle piece to.

Two basic insights have enabled analytical investigations of the average performance of the QAOA: First, the fact that the expected performance of QAOA is the sum over terms which are only affected by a subset of gates, called the Reverse Causal Cone (RCC), which is limited by the number of layers in the QAOA circuit and not the actual problem size. Second, the focus on either all-to-all or random regular problem graphs. Random d -regular graphs have the convenient property that the p -local environments of the nodes and edges, are asymptotically (with growing number of graph nodes N) trees. This is because the probability of having small loops vanishes [18].

Elisabeth Wybo: elisabeth.wybo@meetiqm.com

Martin Leib: martin.leib@meetiqm.com

Consequently the topology of the qubits involved in a single RCC of QAOA is also, with high probability a tree. Therefore, it follows that the asymptotic performance of QAOA with depth p , is fully determined by its performance on those tree structures [11, 19], which allow to use efficient classical contraction schemes. These locality arguments have already been explored in the original QAOA paper for 3-regular graphs [11]. Additionally, it was also possible to calculate optimal, instance- and problem size-independent, parameters thereby creating an optimization loop free variation of QAOA termed ‘tree QAOA’ [20–24].

With the help of the above described techniques, it was possible to show that QAOA outperforms on average the classical worst-case performance guarantee of the Goemans-Williamson (GW) algorithm for the Maximum Cut (MaxCut) problem on high-girth regular graphs, if p is large enough [25, 26]. The GW algorithm [27] is the optimal efficient classical approximate algorithm for this problem with a performance guarantee, assuming that the unique games conjecture is true [28]. It should be noted, however, that some classical algorithms, while lacking a worst-case performance guarantee, perform better on average than the GW guarantee [29–32]. In particular, in Ref. [32] it was shown that an approximate message passing algorithm finds solutions that are arbitrarily close to the optimal solution in the large degree limit. Also in Ref. [33] it was shown that there exists a classical polynomial time algorithm that can beat QAOA in many cases. For the MaxCut problem, the asymptotic QAOA angles have been determined explicitly for the smallest p [11, 16, 22, 34], either using numerical contraction [21, 26], or by evaluating recursive formulas that describe the contraction sequence analytically [19, 35, 36]. The numerical approach of Refs. [21, 26] is limited to fairly small degrees. On the other hand, in the approach of Ref. [19] the degree does not enter in the complexity of the evaluation of the variational energy. The tree angles have been evaluated up to $p = 11$ for MaxCut on 3-regular graphs [26]. This also implies that the asymptotic performance of QAOA on this problem is known explicitly up to depth $p = 11$. However, the performance scaling with p is not formally known. Summarizing, it can be said that fixed depth QAOA exhibits a constant, non-vanishing performance as function of the problem size N , which can also be extended to Sherrington-Kirkpatrick models on all-to-all connected problem graphs [35].

In contrast to this relative success of QAOA in solving the MaxCut problem, it has been explicitly shown that constant depth QAOA has strong sub-optimal performance for the Maximum Independent Set (MIS) problem [37, 38]. Indeed, the minimal requirement to find solutions close to the optimum is ‘to see the whole graph’, thus that $p \in O(\log(N))$, making QAOA an algorithm with increasing query complexity. This observation of widely differing algorithm performance for two NP-hard problem classes, MaxCut and MIS, is not restricted only to quantum algorithms or QAOA but also extends to the performance of classical algorithms.

A recent, very successful attempt at explaining this phenomenon employs the so-called Overlap Gap Property (OGP) that can be linked to algorithmic hardness. Problems that do have an OGP, like MIS in limit of large graph degree, possess an intricate property associated to the clustering of good solutions [39–42]. Solutions that are close to optimality for instances which possess OGP are either very similar or very different, implying that there are remote clusters of good solutions in the state space.

It can be shown that the existence of an OGP implies that there is an a-priori gap between what p -local algorithms (like QAOA) can possibly achieve, and optimality [43]. In contrast to this, for problems without OGP [44–46], we may be able to find a p -local algorithm that produces solutions $(1 - \epsilon)$ close to optimality with high probability. The statement ‘with high probability’ is important, and implies that there is no guarantee that such a good $(1 - \epsilon)$ approximation can be always reached, as otherwise complexity theoretic statements like the unique games conjecture [28], and therefore also $P \neq NP$, would collapse. For MaxCut on d -regular graphs, such an approximate $(1 - \epsilon)$ classical algorithm has been proposed in Ref. [32]. However, in this algorithm ϵ cannot be chosen arbitrarily small for finite d . The performance gap of QAOA in solving problems with and without OGP becomes bigger for random d -regular graphs as d increases. The regime of finite $d \geq 3$ has been treated with QAOA with fixed p for MaxCut with less stringent bounds on optimality [26]. Such an analysis suggests that the approximation ratios achieved by QAOA in solving MaxCut are decreasing with growing d , which is a qualitatively wrong result. In this work, we show they are not decreasing by using tighter bounds on optimality. For the MIS problem the same regime $d \geq 3$ has not been explicitly investigated yet with QAOA to the best of our

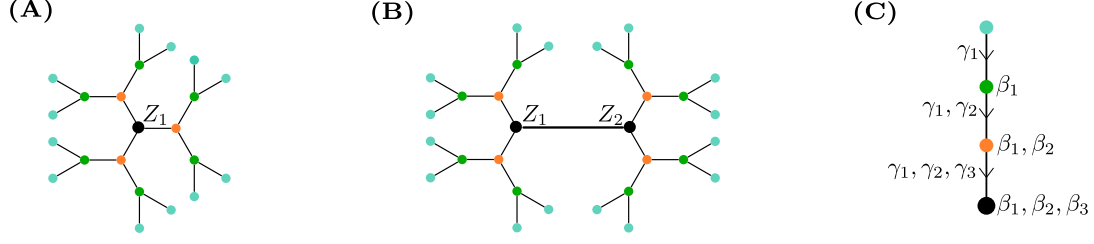


Figure 1: (A,B) The 1-tree and 2-tree subgraphs at level $p = 3$, for degree $d = 3$. (C) The path from outer leaf to root, together with the angle dependencies of each layer. The root variable is shown in black, the other variables are color coded according to their layer: $m = 1$ orange, $m = 2$ green, $m(= p) = 3$ blue.

knowledge.

In this work we therefore provide an explicit performance analysis of QAOA on these two problems. For this we extend the technique of the recursive formula of Ref. [19] to MIS, enabling performance calculations in the large but finite d regime. We start by showing how to treat the MaxCut problem and the MIS problem on equal footing using the language of Ising models with local fields. With a range of numerical improvements in evaluating the recursive formula, we are able to determine tree QAOA performance for Ising models with local fields $H = \sum_{ij \in E} Z_i Z_j + h \sum_{i \in V}$ for $N \rightarrow \infty$ on d -regular graphs with d up to 100.

In addition, we combine these results with state-of-art upper bounds on optimality. For MaxCut, upper bounds on the cut fraction in the asymptotic limit for d -regular graphs have been determined in Refs. [47, 48]. For MIS, upper bounds on the independence ratio in the asymptotic limit for d -regular graphs have been determined in Refs. [49, 50]. Taking into account these findings yields higher performance guarantees for QAOA in the asymptotic limit than previously assumed, and changes insights on the qualitative behaviour in the large d regime. This is mainly relevant for MaxCut: for instance in Ref. [26] it is observed that asymptotic tree QAOA outperforms the GW guarantee at $p \geq 11$. However, by combining the QAOA performance with state-of-the-art bounds on optimality, it can be realized that this actually already happens when $p \geq 4$ for $N \rightarrow \infty$. This observation is only true in the large N limit, but however gives hope that an effective quantum utility regime can be reached in the near term for this problem, with a local quantum algorithm.

Combining the better bounds with the extended method of the recursive formula, we are in the end able to show a manifestly different behavior of QAOA in solving the MaxCut and MIS problems for increasing graph degrees. The MIS problem has an OGP regime in the large d limit [51], but such regime is not proven for finite d . However, for this problem we observe a decreasing QAOA performance with increasing d . On the other hand for MaxCut, we do not observe such a decreasing performance. Despite these findings, there is still hope for MIS that with a quantum algorithm breaking locality, i.e. for QAOA with $p \in \Omega(\log(N))$, near-optimal performances can still be achieved in polynomial-time on a quantum computer. On the other hand, the performance of QAOA with $p \in O(\log(N))$ appears limited as classical algorithms are provably optimal up to the OGP bound [43]. However, there is also future hope that hybrid approaches can provide near-optimal performances in overall polynomial time [52–58].

Our paper is organized as follows: In Sec. 2, we introduce the QAOA algorithm and discuss its $N \rightarrow \infty$ limit for regular graphs which is called ‘tree QAOA’. In Sec. 3, we discuss the MaxCut and MIS problems and show how they can be unified as Ising models. Here, we also define the performance metrics that we will measure on the QAOA ansatz state in order to get the expected solution quality of the sampled solutions. In Sec. 4, we briefly review the concept of Overlap Gap Property and local algorithms, and provide a short overview about the different hardness regimes of the two problems, and the implications for QAOA. In Sec. 5, we present and discuss our results. The main result of our paper is shown in Fig. 5, where the asymptotic performances of QAOA for both problems are summarized for many different graph degrees. We conclude in Sec. 6, and provide the full tree calculations in Appendix A.

2 QAOA and tree QAOA

In this section, we review the QAOA algorithm, and summarize its asymptotic limit in the number of nodes N for d -regular graphs, which we will refer to as ‘tree QAOA’. QAOA is a variational algorithm that is designed to find low-energy states of a diagonal Hamiltonian H . Our work considers the Ising Hamiltonian on random d -regular graphs $G(V, E) \in \mathbb{G}(N, d)$, with V the vertex set $|V| = N$, and E the edge set $|E| = Nd/2$,

$$H = \frac{1}{\sqrt{d}} \left(\sum_{ij \in E} Z_i Z_j + h \sum_{i \in V} Z_i \right). \quad (1)$$

Here Z_i is the Pauli- Z operator associated with the i -th qubit, we use the convention that $Z_i |z\rangle = z_i |z\rangle$ with $z_i \in \{-1, 1\}$ and $z = (z_1, \dots, z_N)$. To find low-energy states, one optimizes a depth- p parameterized circuit ansatz that has been introduced in Ref. [11] and that consists of alternating layers of mixing and phase-separator unitaries

$$|\gamma, \beta\rangle = e^{-i\beta_p B} e^{-i\gamma_p H} \dots e^{-i\beta_1 B} e^{-i\gamma_1 H} |+\rangle, \quad (2)$$

where $B = \sum_i X_i$ is the mixing operator with X_i the Pauli- X operator associated to the i -th qubit. The initial state is an eigenstate of the mixing operator, and is defined as $|+\rangle = \frac{1}{\sqrt{2^N}} \sum_z |z\rangle$. Using the Ritz’ variational principle one can find optimal parameters $\beta = (\beta_1, \dots, \beta_p)$ and $\gamma = (\gamma_1, \dots, \gamma_p)$ in a classical outer optimization loop such that the variational energy of the classical cost function, in our case Eq. (1), is minimized

$$\min_{\gamma, \beta} \langle \gamma, \beta | H | \gamma, \beta \rangle. \quad (3)$$

This guarantees that the measurement samples z from the prepared quantum state $|\gamma, \beta\rangle$ on average have low energy $H(z) = \langle z | H | z \rangle$ in a noiseless scenario. Ideally, the optimal solution z_{opt} , for which $H(z_{opt}) = \min_z H(z)$, is sampled with non-vanishing probability. The QAOA ansatz resembles a trotterization of the quantum annealing protocol and therefore the adiabatic theorem guarantees that this is true for $p \rightarrow \infty$ with unit probability, but this is not practically attainable. It can however be shown that the variational method, where the trotter-step sizes are optimized, always performs better compared to continuous evolution if the total evolution time is finite [59]. Therefore, in this paper we want to investigate which energies (3) can be reached in the limit of large instances $N \rightarrow \infty$ for *fixed* p .

The energy expectation value (3) consists of a sum of local expectation values, because the summations in Eq. (1) run respectively over the edges and vertices of the problem graph. As the problem graph also determines the two-qubit gates in the QAOA ansatz, it follows that $\langle \gamma, \beta | Z_i | \gamma, \beta \rangle$ only depends on gates acting on qubits that are within distance p in the problem graph from vertex i . All other gates commute with the measurement operator Z_i and undergo a unitary cancellation. This is true in a similar fashion for $\langle \gamma, \beta | Z_i Z_j | \gamma, \beta \rangle$ with $ij \in E$, i.e. only gates acting on qubits that are within distance p from either i or j in the problem graph contribute. The qubits that are part of the causal cone of a local operator form a subgraph of the original problem graph. We will refer to the set of gates acting on such a subgraph, and determining the expectation value of a local operator, as the Reverse Causal Cone (RCC) of the operator. When uniformly sampling large random regular graphs, the probability of having small loops (e.g. triangles) in the graph is exponentially suppressed [18]. From this it follows that all the causal-cone subgraphs have the same topology of regular trees when $N \rightarrow \infty$. Therefore, the energy density measured on the QAOA ansatz state reduces to

$$\lim_{N \rightarrow \infty} \frac{\langle \gamma, \beta | H | \gamma, \beta \rangle}{N} = \frac{\sqrt{d}}{2} \langle \gamma, \beta | Z_1 Z_2 | \gamma, \beta \rangle^{2\text{-tree}} + \frac{h}{\sqrt{d}} \langle \gamma, \beta | Z_1 | \gamma, \beta \rangle^{1\text{-tree}}, \quad (4)$$

where H is given by Eq. (1). The expectation values $\langle \gamma, \beta | Z_1 Z_2 | \gamma, \beta \rangle^{2\text{-tree}}$ and $\langle \gamma, \beta | Z_1 | \gamma, \beta \rangle^{1\text{-tree}}$ are defined on a QAOA state which consists of a reduced amount of spin variables that are arranged in tree subgraphs. The edge correlator $\langle \gamma, \beta | Z_1 Z_2 | \gamma, \beta \rangle^{2\text{-tree}}$ describes the expectation value corresponding to a causal-cone subgraph that is a d -regular tree of depth p with two roots. Similarly the onsite expectation value $\langle \gamma, \beta | Z_1 | \gamma, \beta \rangle^{1\text{-tree}}$ corresponds to a single-root tree subgraph. For $d = 3$ and $p = 3$ these causal-cone subgraphs are shown in Fig. 1.

2.1 Summary of the recursive iterations

In Appendix A, we provide recursive formulas to compute the two terms in Eq. (4). These generalize the formulas derived in Ref. [19] by breaking the $\mathbb{Z}_2$ symmetry of MaxCut, and by including a block symmetry in the iterations. In this section, we will only summarize the results of A.1, and refer to the Appendix A for the full derivation and for more compact iterations in which the number of terms is reduced.

The two terms needed to evaluate Eq. (4) can be worked out as

$$\begin{aligned} \langle \gamma, \beta | Z_1 Z_2 | \gamma, \beta \rangle^{2\text{-tree}} \\ = \sum_{\mathbf{a}, \mathbf{b}} a_0 b_0 f(\mathbf{a}) f(\mathbf{b}) H_{d-1}^{(p)}(\mathbf{a}) H_{d-1}^{(p)}(\mathbf{b}) \exp\left(i \frac{1}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{ab}\right) \exp\left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma} \cdot (\mathbf{a} + \mathbf{b})\right), \end{aligned} \quad (5)$$

and

$$\langle \gamma, \beta | Z_1 | \gamma, \beta \rangle^{1\text{-tree}} = \sum_{\mathbf{a}} a_0 f(\mathbf{a}) H_d^{(p)}(\mathbf{a}) \exp\left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{a}\right). \quad (6)$$

Here the summation $\sum_{\mathbf{a}}$ ($\sum_{\mathbf{a}, \mathbf{b}}$) means summing over one basis set (two independent basis sets) and runs over all possible vectors

$$\mathbf{a} = (a_1, \dots, a_p, a_0, a_{-p}, \dots, a_{-1}), \quad a_i \in \{-1, 1\},$$

describing bitstrings of length $2p + 1$ (similar for $\mathbf{b}$). Following the notation of Ref. [19], we also arrange the QAOA angles corresponding to the phase separator in a vector of length $2p + 1$

$$\mathbf{\Gamma} = (\gamma_1, \dots, \gamma_p, 0, -\gamma_p, \dots, -\gamma_1).$$

The element-wise product between two bitstrings is denoted as $\mathbf{ab}$, and the inner product between two vectors as $\mathbf{\Gamma} \cdot \mathbf{a}$.

We define the following function which isolates the dependency of the QAOA mixing angles

$$\begin{aligned} f(\mathbf{a}) = \frac{1}{2} \langle a_1 | e^{i\beta_1 X} | a_2 \rangle \langle a_2 | e^{i\beta_2 X} | a_3 \rangle \dots \langle a_p | e^{i\beta_p X} | a_0 \rangle \\ \langle a_0 | e^{-i\beta_p X} | a_{-p} \rangle \dots \langle a_{-3} | e^{-i\beta_2 X} | a_{-2} \rangle \langle a_{-2} | e^{-i\beta_1 X} | a_{-1} \rangle. \end{aligned}$$

where we omit the dependency of β in the function argument to allow for compact notations. Factors of this form intuitively arise from the insertion of a resolution of the identity after the application of each QAOA layer which is indeed how these formulas are derived.

Finally, the recursive iteration that describes the contraction over the branches of the tree is given by

$$H_{d-1}^{(m)}(\mathbf{a}) = \left[\sum_{\mathbf{b}} f(\mathbf{b}) H_{d-1}^{(m-1)}(\mathbf{b}) \exp\left(i \frac{1}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{ab}\right) \exp\left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{b}\right) \right]^{d-1}, \quad (7)$$

with $m = 1, \dots, p$ and $H_{d-1}^{(0)}(\mathbf{a}) \equiv 1$. The level $m = 1$ describes the contraction over the outermost leaves, while $m = p$ corresponds to the contraction level directly adjacent to the root(s). Note however, that the subgraph of the local term has only a single root variable. Therefore, it has d equivalent branches (and not just $d - 1$), see also Fig. 1. For this reason the final iteration, when $m = p$, needs to be modified in this case to

$$H_d^{(p)}(\mathbf{a}) = \left[\sum_{\mathbf{b}} f(\mathbf{b}) H_{d-1}^{(p-1)}(\mathbf{b}) \exp\left(i \frac{1}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{ab}\right) \exp\left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{b}\right) \right]^d. \quad (8)$$

In Sec. 5, we will discuss our results that follow from minimizing Eq. (4). The time complexity of evaluating the energy density with the above described recursive procedure scales as $O((p + 1)2^{4p+2})$. In Appendix A we derive (large) prefactor speed ups.

3 MaxCut and Maximum independent set as Ising models

In this section, we introduce two paradigmatic combinatorial optimization problems for which we will study the performance of QAOA in this work: Maximum Cut (MaxCut) and Maximum Independent Set (MIS). MaxCut refers to the task of finding a bipartition of the set of vertices of a given graph into two sets such that the number of edges of the graph connecting nodes between the two sets is maximal. The MIS problem asks for the largest set of vertices for a given graph such that no two vertices within this set are adjacent in the graph. The cost function for both MaxCut and MIS can be described as an Ising Hamiltonian of the form of Eq. (1) which enables us to treat them on a similar footing. We will consider both problems on d -regular graphs with $d \geq 3$. Note that for $d < 3$, the MaxCut and MIS problems are equivalent and trivial.

3.1 MaxCut

When $h = 0$ the model (1) corresponds to the standard Ising anti-ferromagnet, whose ground state encodes the solution of the MaxCut problem. Indeed, minimizing the energy when $h = 0$, corresponds to maximizing the expected number of cut edges (or anti-ferromagnetic interactions)

$$C = \frac{1}{2} \sum_{ij \in E} (1 - Z_i Z_j). \quad (9)$$

Suppose we have a QAOA ansatz state $|\gamma, \beta\rangle$, obtained by minimizing the energy given by Eq. (1), and we would like to assess the average quality of a candidate solution $\mathbf{z}$ sampled from this quantum state. For the MaxCut problem, finding a performance metric is straightforward, as every basis state $\mathbf{z}$ can be seen as a candidate solution, for which we can calculate the expected ‘cut fraction’ $c \in \{0, 1/|E|, \dots, 1\}$ with $|E| = Nd/2$,

$$c(\mathbf{z}) = \frac{1}{2|E|} \sum_{ij \in E} (1 - \langle \mathbf{z} | Z_i Z_j | \mathbf{z} \rangle).$$

In the limit $N \rightarrow \infty$, the expected cut fraction for the depth- p QAOA ansatz state minimizing the energy (1) is

$$c_p = \lim_{N \rightarrow \infty} \frac{1}{2|E|} \sum_{ij \in E} (1 - \langle \gamma, \beta | Z_i Z_j | \gamma, \beta \rangle) = \frac{1 - \langle \gamma, \beta | Z_1 Z_2 | \gamma, \beta \rangle^{2\text{-tree}}}{2}. \quad (10)$$

It was shown in Ref. [48] that there exist rigorous upper bounds c_{ub} on the optimal cut fraction c_{opt} ($c_{ub} > c_{opt}$) for d -regular graphs (and for sparse Erdős-Rényi graphs) in the limit $N \rightarrow \infty$. These bounds confirm the earlier conjecture presented in Ref. [60] and can be obtained explicitly from techniques in statistical physics that make use of the interpolation method. In practice, they can be obtained by solving a small variational problem. We have calculated the upper bounds in this way and listed them in Table 1 for different graph degrees.

Other upper bounds can be derived from complementary approaches, relating the improvement that is made on top of the average random sampling outcome to the Parisi constant [47]. However, we have found that the first approach yields a tighter (i.e. lower) upper bound on the optimal fraction, at least for small graph degrees $d < 10$.

With the help of the tighter upper bound on the optimal cut fraction c_{ub} , we can compute a tighter (i.e. higher) lower bound α_{MC} on the approximation ratio c_p/c_{opt} , and thus on the performance, of QAOA for MaxCut

$$c_p < \alpha_{MC} = \frac{c_p}{c_{ub}} < \frac{c_p}{c_{opt}} \leq 1.$$

In large parts of the QAOA literature, including the original QAOA paper, the expected cut fraction c_p of the QAOA state is considered as a lower bound on the performance. While this is naturally correct (a weaker lower bound is still a lower bound), it may suggest the misleading insight that the performance of QAOA is worse for MaxCut on graphs with higher degree [22]. This is indeed

degree d	c_{ub} (cut fraction MaxCut)	r_{ub} (independence ratio MIS)
3	0.92410	0.45400
4	0.86824	0.41635
5	0.83504	0.38443
6	0.80500	0.35799
7	0.78509	0.33567
8	0.76585	0.31652
9	0.75233	0.29987
10	0.73877	0.28521
20	0.67023	0.19732
50	0.60820	0.11079
100	0.57665	0.06787

Table 1: The upper bounds on the optimal cut fraction c_{opt} and independence ratio r_{opt} for d -regular graphs in the asymptotic limit $N \rightarrow \infty$. For the MaxCut problem, these upper bounds have been obtained by solving the variational problem derived in Ref. [48]. For the MIS problem, the upper bounds have been taken directly from Ref. [61], see also Refs. [49, 50].

wrong as in the extreme limit of (large) complete graphs, maximal cuts have size $|E|/2$, and such cuts are given on average by random sampling (which is equivalent to $p = 0$ QAOA, or to QAOA with all angles 0). Hence, the MaxCut problem is easy in this limit, and the performance of QAOA, as measured by the above approximation ratio α_{MC} , should therefore improve.

3.2 Maximum independent set

The MIS problem is about finding the largest subset of vertices of a given graph such that none of them is sharing an edge. The MIS problem is solved by finding the ground state of

$$H_{\lambda}^{MIS} = \lambda \sum_{ij \in E} N_i N_j - \sum_{i \in V} N_i \quad (11)$$

with $\lambda > 1$, and the ‘number operator’ $N_i = \frac{Z_i + 1}{2}$. Therefore $\langle z_i | N_i | z_i \rangle = (z_i + 1)/2 \in \{0, 1\}$. We assume that if $N_i |z_i\rangle = |z_i\rangle$, the i -th vertex is part of the chosen set. The second term in Eq. (11) is attempting to maximize the number of vertices in the set, i.e. the number of 1’s in $|z\rangle$, while the first term ensures their independence by adding an energy penalty for having two adjacent vertices in the set. For regular graphs, and in terms of Pauli-Z operators, this Hamiltonian becomes

$$H_{\lambda,d}^{MIS} = \frac{\lambda}{4} \sum_{ij \in E} Z_i Z_j + \frac{\lambda d - 2}{4} \sum_{i \in V} Z_i + \frac{\lambda d N}{8} - \frac{N}{2}.$$

After rescaling $H_{\lambda,d}^{MIS} \rightarrow 4/\lambda H_{\lambda,d}^{MIS}$, we see that this reduces to Eq. (1) with $h = d - 2/\lambda$ up to an irrelevant constant. Hence, the ground state of Eq. (1) corresponds to the solution of the maximum independent set if $h \in]d - 2, d]$. However, the low-lying excited states might not correspond to independent sets. Note that for h ’s outside this interval, even the ground state might not correspond to an independent set. For both reasons, a simple pruning protocol can be devised: if we have a candidate solution $\mathbf{z} \in \{-1, 1\}^{\otimes N}$ with energy $H_{\lambda=1,d}^{MIS}(\mathbf{z}) = \langle \mathbf{z} | H_{\lambda=1,d}^{MIS} | \mathbf{z} \rangle < 0$, we can prune this bitstring to an independent set of at least size $-H_{\lambda=1,d}^{MIS}(\mathbf{z})$. Indeed, assume that $z_i = z_j = 1$ and that ij is an edge in G . If we remove this edge by flipping one of the variables randomly, the resulting state has at most the same energy $H_{\lambda=1,d}^{MIS}(\mathbf{z})$. States with positive energy $H_{\lambda=1,d}^{MIS}(\mathbf{z}) > 0$ have less vertices chosen than edges existing that connect them. This implies that the above described pruning procedure might end up in a trivial independent set containing only one vertex. For states with positive energy, we can thus make the worst-case (trivial) assumption that they correspond to a single-node solution by keeping one of the 1’s and flipping all others. Note, however, that the single node solution is ‘optimal’ for complete graphs. Indeed, for complete graphs we have that the lower bound in the spectrum is $\min_{\mathbf{z}} H_{\lambda=1,d}^{MIS}(\mathbf{z}) = -1$.

Now, we would like to ask the same question as before for MaxCut: suppose we have a QAOA ansatz state $|\gamma, \beta\rangle$, obtained by minimizing the energy given by Eq. (1), what is the average quality of a candidate solution $\mathbf{z}$ sampled from this quantum state? For the MIS problem, not every basis state $\mathbf{z}$ can be seen as a candidate solution, as not every basis state corresponds to an independent set. However, as discussed above, basis states with $H_{\lambda=1,d}^{MIS}(\mathbf{z}) < 0$ are easily prunable, to independent sets of at least size $-H_{\lambda=1,d}^{MIS}(\mathbf{z})$. The independence ratio is given by the number of variables in the independent set divided by the total number of variables. As performance metric for MIS, will thus consider

$$r(\mathbf{z}) = -\frac{H_{\lambda=1,d}^{MIS}(\mathbf{z})}{N} = -\frac{1}{4N} \sum_{ij \in E} \langle \mathbf{z} | Z_i Z_j | \mathbf{z} \rangle + \frac{2-d}{4N} \sum_{i \in V} \langle \mathbf{z} | Z_i | \mathbf{z} \rangle + \frac{4-d}{8},$$

which is only meaningful when positive. In the asymptotic limit the expected independence ratio of bitstrings sampled from a QAOA state with negative energy $\langle H_{\lambda=1,d}^{MIS} \rangle$ is given by

$$\begin{aligned} r_p &= \lim_{N \rightarrow \infty} \left[-\frac{1}{4N} \sum_{ij \in E} \langle \gamma, \beta | Z_i Z_j | \gamma, \beta \rangle + \frac{2-d}{4N} \sum_{i \in V} \langle \gamma, \beta | Z_i | \gamma, \beta \rangle + \frac{4-d}{8} \right] \\ &= -\frac{d}{8} \langle \gamma, \beta | Z_i Z_j | \gamma, \beta \rangle^{2\text{-tree}} + \frac{2-d}{4} \langle \gamma, \beta | Z_i | \gamma, \beta \rangle^{1\text{-tree}} + \frac{4-d}{8}. \end{aligned}$$

Similarly, as for the cut fraction for the MaxCut problem, the optimal (largest) independence ratio for MIS on d -regular graphs r_{opt} can be upper bounded for $N \rightarrow \infty$ [49, 50]. So we can derive a $r_{ub} > r_{opt}$, which can then be used to form a lower bound α_{MIS} on the approximation ratio r_p/r_{opt} achieved by QAOA on this problem in limit of large N

$$r_p < \alpha_{MIS} = \frac{r_p}{r_{ub}} < \frac{r_p}{r_{opt}} \leq 1.$$

The upper bounds r_{ub} are also listed in Table 1 for different degrees. These have been taken from Ref. [61].

With the above introduced Ising model and performance metrics α_{MC} and α_{MIS} , we will be able to investigate the performance of tree QAOA in solving the MaxCut and MIS problems as a function of the local field. From this section, it may seem intuitive that MIS problem is hard, as the magnitude of the local terms in Eq. (1) exactly balances the magnitude of the interactions. (Recall that the MIS regime is realized when $h \in [d-2, d]$.) Therefore, the MIS problem is realized exactly in the ‘critical’ h regime of Eq. (1).

It has indeed been shown that the MIS problem exhibits a hardness property known as the ‘overlap gap property’, and that local algorithms cannot find high quality solutions. We will review these topics in the next section.

4 Local algorithms and overlap gap property

Independent sets of more than half the size of the maximum independent set of random regular graphs above a certain degree are either very similar i.e. contain almost the same set of vertices, or very different i.e. contain almost disjoint sets of vertices [51]. In other words, there is a gap in the overlap of large independent sets of graphs, which is why this phenomenon is called the Overlap Gap Property (OGP) of the MIS problem. Other NP-hard optimization problems like MaxCut are not believed to exhibit this phenomenon [44]. There is analytical evidence that this fundamental difference between the combinatorial optimization problems has a profound impact on the performance of local algorithms, classical and quantum alike [46]. Therefore, in this section we will briefly review p -local algorithms, and their performance limitations caused by OGP for the MIS problem.

An intuitive definition of generic p -local algorithms (be it quantum or classical) is given in Ref. [46]. We review this definition tailored to our purposes. We start by defining the p -local environment of a vertex $v_i \in V$ in the problem graph $G(V, E)$ as the subgraph $B_G^p(v_i) \subseteq G(V, E)$

induced by vertex set $\{v_j \in V \mid d_G(v_i, v_j) \leq p\}$. Here the graph distance between two vertices $d_G(v_i, v_j)$ is the number of edges in the shortest path between v_i and v_j .

We consider a probabilistic algorithm $\mathcal{A}$ that generates samples from a distribution $\mathcal{A}_G(z_1, \dots, z_N)$ of binary variables $z_1, \dots, z_N$ that are associated to each vertex of a specific graph G . We say that $\mathcal{A}$ is a p -local algorithm when it satisfies two criteria:

- (i) When $d(v_i, v_j) > 2p$, $\mathcal{A}_G(z_i) = \sum_{z_1, \dots, z_N \setminus z_i} \mathcal{A}_G(z_1, \dots, z_N)$ is statistically independent from $\mathcal{A}_G(z_j) = \sum_{z_1, \dots, z_N \setminus z_j} \mathcal{A}_G(z_1, \dots, z_N)$, meaning that the marginals are statistically independent if the vertices associated to the variables do not have overlapping environments.
- (ii) The marginals are equal $\mathcal{A}_G(z_i) = \mathcal{A}_G(z_j)$ if their p -local environments are isomorphic, $B_G^p(v_i) \simeq B_G^p(v_j)$.

In this sense QAOA is a local algorithm because (i) Variables with non-overlapping RCCs are not mutually entangled and also not classically correlated. Hence, they do not share any mutual information and are therefore statistically independent. (ii) Observables depend solely on the topology of their RCCs for a given set of angles.

We now review the definition of the OGP tailored to the MIS problem on regular graphs. Further details can be found in Refs. [37, 51]. We start by considering a multiplicative factor $\mu \in [0, 1]$, that quantifies the distance between a candidate solution and the optimum. For a given graph G , we can thus define the set of μ -good solutions (where we will assume that the bitstring $\mathbf{z}$ corresponds to an independent set of G)

$$\mathcal{S}(\mu, G) = \{\mathbf{z} : r(\mathbf{z}) \geq \mu r_{opt}\}.$$

Now let us consider two random regular graph instances G_0 and G_1 , defined on N vertices, each with their corresponding sets of μ -good solutions. Let us define for every $0 < \theta \leq \mu$, two sets of pairs $(\mathbf{z}_0, \mathbf{z}_1)$ with $\mathbf{z}_i \in \mathcal{S}(\mu, G_i)$, $i = 0, 1$, such that

$$\mathcal{S}^{similar}(\mu, \theta, G_0, G_1) = \{(\mathbf{z}_0, \mathbf{z}_1) : r(\mathbf{z}_0 \wedge \mathbf{z}_1) \geq \theta r_{opt}\},$$

and

$$\mathcal{S}^{different}(\mu, \theta, G_0, G_1) = \{(\mathbf{z}_0, \mathbf{z}_1) : r(\mathbf{z}_0 \wedge \mathbf{z}_1) \leq \theta r_{opt}\}.$$

These sets thus contain pairs of good solutions whose independence ratio of the intersection normalized to the optimal independence ratio, is either equal, or smaller or bigger than θ . Here the intersection of two independent sets means taking the logical AND product between their bitstrings, i.e. element-wise $z_i \wedge z_j = 1$ if and only if $z_i = 1$ and $z_j = 1$.

We now define interpolation graphs G_t between the two random regular graphs G_0 and G_1 . Here $t \in \{0, 1/|E|, \dots, 1\}$ determines the fraction of edges that are selected from G_1 and added to G_t . At the same time a fraction of $(1 - t)$ edges are selected from G_0 and added to G_t . To have similarity between G_t and $G_{t+\Delta}$ with $\Delta = |E|^{-1}$, the previously added (removed) edges from G_1 (G_0) should remain the same, such that G_t and $G_{t+\Delta}$ only differ by two edges in the typical case. For the sake of stating the OGP property of MIS, it does not matter that G_t is not regular at every step in the interpolation. The average degree of G_t is still d , which is the important fact.

Theorem 4.1. *See Ref. [38]. The MIS problem on d -regular graphs has OGP when d is large enough. This means that there exist a μ^* such that for every $\mu > \mu^*$ there exists $0 < \theta_1 < \theta_2 < \mu$, such that for N large enough*

$$\mathcal{S}^{similar}(\mu, \theta_1, G_{t_1}, G_{t_2}) \cap \mathcal{S}^{different}(\mu, \theta_2, G_{t_1}, G_{t_2}) = \emptyset, \quad \forall 0 \leq t_1, t_2 \leq 1$$

with high probability (i.e. a probability converging to 1 exponentially fast in N), and

$$\mathcal{S}^{similar}(\mu, \theta_1, G_0, G_1) = \emptyset$$

with high probability.

Hence, the OGP means that for bounded degree random graph instances G_{t_1} and G_{t_2} , pairs of μ -good solutions are either ‘similar’ (they have at least normalized intersection ratios of θ_2), or ‘different’ (they have at most normalized intersection ratios of θ_1) with high probability. Pairs of μ -good solutions that do not satisfy this property are exponentially rare. Additionally, when two graph instances do not share an a priori similarity, like G_0 and G_1 , it is already exponentially rare to have pairs of similar μ -good solutions with normalized intersection ratios larger than θ_1 . This means that somewhere in the interpolation between G_0 and G_1 , there must be a ‘jump’ allowing for pairs of similar solutions.

It is shown in Ref. [51] that the presence of OGP for MIS on regular graphs with large d obstructs local algorithms, like p -local QAOA [37, 38], from finding $(\mu > \frac{1}{2} + \frac{1}{2\sqrt{2}})$ -good solutions with non-vanishing probability. We briefly sketch why. Let us construct a sequence of coupled independent sets for the interpolation graphs G_t , which we call z_t . This sequence starts from an independent set z_0 obtained by taking a single sample from running the depth- p QAOA algorithm for G_0 , potentially followed by pruning. QAOA exhibits strong Hamming weight concentration [37]. Therefore, $r(z_0)$ will not differ much from its average r_p^0 with high probability. Because we know that the next graph G_Δ only differs by at most two edges from G_0 , we can construct an independent set z_Δ that is similar to z_0 . Indeed, as G_Δ only differs by at most two edges from G_0 , there exist QAOA samples for G_Δ that will only be different in at most $4 \max_i |B_G^p(v_i)|$ bits from z_0 . We take z_Δ to be one of those. Notice that for fixed p and large N , the subgraph sizes are vanishing compared to N , so that is why z_0 and z_Δ are considered similar. Like this we can construct the sequence of similar solutions z_t .

We now assume that depth- p QAOA can create large independent sets of G_t

$$r_p^t > \mu r_{opt}. \quad (12)$$

Notice that concentration over graph instances [18] implies that assumption (12) must be valid $\forall t$. However this assumption is in direct violation with the OGP of MIS, see Theorem 4.1, because it would then be possible with high probability to have similar large independent sets, for which $z_0 \wedge z_t$ is large, without having a sudden jump to different large independent sets, for which $z_0 \wedge z_{t+\Delta}$ is small, for a certain t . Therefore, we conclude that the expected independence ratio obtained from QAOA is suboptimal for large enough N and d

$$r_p \leq \mu r_{opt}.$$

It is shown in Ref. [62] that asymptotically $\mu^* \xrightarrow{>} 1/2$, for $d \rightarrow \infty$. Hence, there is a performance gap that can be as large as $1/2$ between the output of local algorithms and optimality. (Famously, this is also true for Erdős-Rényi graphs that are both sparse and dense [63, 64].) Indeed, it is shown in Ref. [65] that the optimal (i.e. maximum) independence ratio asymptotically converges to $2 \log(d)/d$, for $d \rightarrow \infty$. However, it is shown in Ref. [62] that the largest independence ratios that can be obtained by local algorithms are asymptotically $\log(d)/d$, when $d \rightarrow \infty$.

These findings refuted the hope [66] that local algorithms may be able to find maximum independent sets in random regular graphs. However, the statements are only proven for ‘sufficiently large’ d , which leaves open the question if OGP holds for the smallest d , and if it holds, what would be the value of μ^* . From Ref. [61], some lower bounds for the possible onset of OGP can however be deduced, see Table 2. Hence, the degrees considered in this paper seem still rather far from the asymptotic d regime. This also implies that even if there is OGP existing for the smallest d , the regime in which it exists is small. Indeed, for instance for $d = 3$, Table 2 means that there exists a local algorithm that can create independent sets with ratios of at least $0.9809 r_{opt}$. Hence if the MIS problem for $d = 3$ would have an OGP, μ^* must be bigger than this value. In other words, local algorithms fail the earliest in creating independent sets with ratios larger $0.9809 r_{opt}$ if the MIS for $d = 3$ problem would have an OGP.

In general, a proven presence of OGP for a problem class, yields a limitation on the performance of *all* local algorithms [42, 46]. On the other hand, it is a standing conjecture that the MaxCut (and Sherrington-Kirkpatrick) problems do not have OGP [44–46], and that there is thus no such performance gap set by a certain μ^* for local algorithms. Indeed, under the assumption that these problems have no OGP, approximate message-passing algorithms have been devised that with high probability do not see such a gap [32, 67]. In Ref. [32] in particular, such an algorithm for

degree d	Lower bound on μ^* for MIS
3	0.9809
4	0.9705
5	0.9346
6	0.9300
7	0.9255
8	0.9098
9	0.9057
10	0.9021
20	0.8808
50	0.8583
100	0.8427

Table 2: The lower bounds on the possible onset of OGP for the MIS problem for d -regular graphs in the asymptotic limit $N \rightarrow \infty$. These bounds are the approximation ratios achieved by the linear-time classical algorithm presented in Ref. [61], which is the state-of-the-art classical local polynomial-time algorithm solving this problem.

the MaxCut problem is constructed that finds solutions that are $(1 - \epsilon)$ close to optimality with high probability. It relies however on the condition that $d > O(1/\epsilon)$. Hence ϵ can not be made arbitrarily small yet.

The conjecture of ‘no OGP’ for MaxCut suggests that there are no bounds on the expected performance of tree QAOA in the asymptotic limit when p is increased. On the other hand, for MIS the regime with OGP can certainly not be reached with p -local QAOA when the RCCs do not span the whole system [37, 38]. The tree QAOA, which is the subject of this paper, however considers the asymptotic limit $N \rightarrow \infty$ and fixed p . In that case, it is also expected that even when increasing p indefinitely, the performance of tree QAOA is bounded away from optimality [68].

In the next section, we will evaluate the performance of tree QAOA explicitly for both problems.

5 Tree QAOA results

In this section, we first present the results from numerically evaluating the recursive tree QAOA formulas [see Eqs. (6),(5),(7) and (8)] in order to minimize the Ising energy density Eq. (4). These formulas were summarized in their simplest form in Sec. 2, and are derived (and made faster, by prefactors) in Appendix A. Second, we recycle the angles obtained from the tree QAOA to solve finite-size instances. In this way, we see that for degree $d = 3, 4$, we outperform the GW algorithm on average for MaxCut with those fixed angles [22, 26]. For MIS, we outperform a minimal greedy algorithm for $d = 3$.

5.1 Performance of tree QAOA in the asymptotic limit

As a first step, we will optimize the QAOA angles such that the energy density Eq. (4) is minimized as a function of the local field. Secondly, we will evaluate the performance metrics presented in Sec. 3 on the resulting ansatz state for both problems. Here, the operator in the phase separator of the QAOA ansatz is the same as the objective function that is minimized (i.e. Eq. (1)), but is not necessarily equivalent to the performance metric. Indeed for MaxCut they are only directly related when $h = 0$, and for MIS when $h = d - 2$. Later we will however restrict to these cases. We first investigate the (fixed) performance metrics for both problems as a function of the local field because in that way we can view the performance for different degrees and fields in a unified way. Like that, we can also investigate the QAOA performance of MaxCut in the h -regime corresponding to MIS, and conversely. Such analysis leads to interesting observations like the fact that the performance of QAOA in solving the MIS problem does not vary much with h when $h < d$ and both d and p are small.

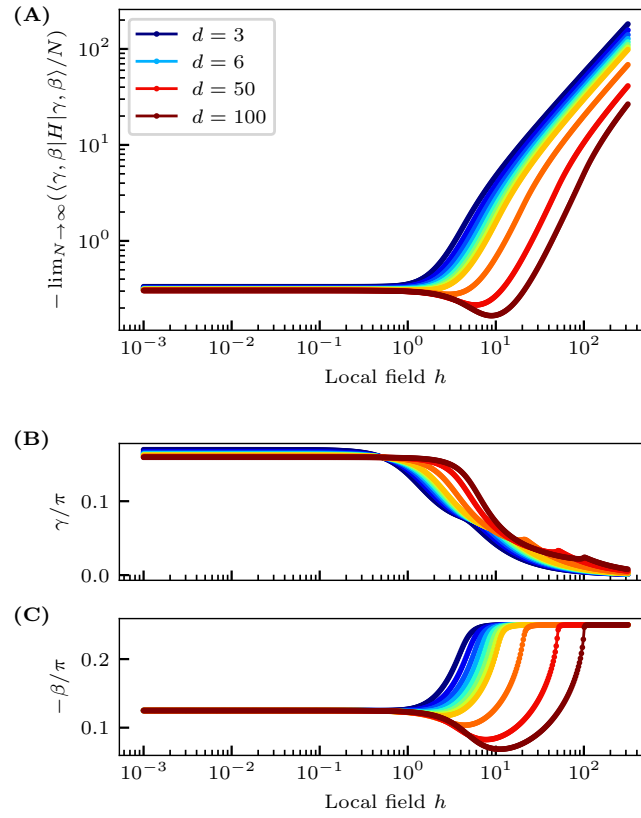


Figure 2: **(A)** The asymptotic energy densities achieved by $p = 1$ QAOA as a function of the local field [see Eq. (4)]. The colors indicate the different graph degrees listed in Table 1, only a selection is shown in the legend. **(B,C)** The respective QAOA angles to obtain these energy densities.

5.1.1 Changing the local field

In Fig. 2(A), we show the resulting energy density Eq. (4) as a function of the local field for $p = 1$. The corresponding QAOA angles to obtain these energy densities are shown in Fig. 2(B,C). The behavior of these is quite intuitive: for small $h < 1$ there is an anti-ferromagnetic regime (constant non-trivial γ), for $1 < h < d$ a transitional ‘critical’ regime (changing non-trivial angles), and for $d < h$ a trivial regime where all spins simply anti-align with the local field (realized exactly by the $\exp(i\pi/4 \sum_i X_i)$ rotation, and vanishing γ).

Next, we evaluate the performance of these optimized QAOA ansätze in solving the MaxCut and MIS problems. The approximation ratio we evaluate as performance metric for MIS (see Sec. 3) can be seen as an order parameter that detects ‘critical order’ in the QAOA state that is optimized for H as a function of h . Similarly, the MaxCut approximation ratio simply detects anti-ferromagnetic order.

In Fig. 3, we show the corresponding lower bounds on the QAOA approximation ratios for the two problems as a function of the local field in the limit $N \rightarrow \infty$. We recall that these approximation ratios take into account the state-of-the-art asymptotic upper bounds listed in Table 1. We also recall that a zero approximation ratio would mean cutting no edges for MaxCut, or finding a vanishing independence ratio with N for MIS. The behaviour of these approximation ratios is interesting. First, we can observe the manifestly different problem hardness when d grows larger in the performance of QAOA for both problems. On the one hand, MaxCut in the large d limit becomes ‘easier’. Indeed, when approaching complete graphs, cutting half of the edges (or random sampling), becomes optimal. However, on the other hand MIS is highly non trivial as it has a proven OGP when d grows larger [51]. This is reflected in a vanishing performance of QAOA when d is increased. Interestingly, for MIS on 3-regular graphs, the approximation ratio achieved by $p = 1$ QAOA is surprisingly constant when $h < 3$ [see Fig. 3(B)]. This implies that solutions for MIS on 3-regular graphs with independence ratios that are $\sim 60\%$ of the optimum can be equally well sampled from a QAOA circuit with angles optimized for the MaxCut problem. This constant behavior is specific to low degrees and relatively low approximation ratios, and is expected to vanish with increasing d (or p).

There is a sharp transition in the achieved approximation ratios for both problems when $h/d > 1$. This transition is due to the fact that the trivial regime is entered where all spins simply anti-align with h , leading to a diamagnetic state that represents an empty set and that cuts no edges. As QAOA captures this behavior more accurately with increasing p , this transition gets also sharper with increasing p . This can be seen from Fig. 4, where $d = 6$ is fixed, and p increased.

For the MIS case, the performance vanishes when h falls below some threshold value if d is large enough. From Figs. 3(B) and 4(B), it can be seen that this happens when $d = 6$ for $p = 1$. Increasing h towards $d - 2$ resolves this behavior, but notice that also increasing p helps to resolve this behavior. For $d = 6$, taking $p = 2$ seems already sufficient [see Fig. 4(B)].

As expected from the problem definition in Sec. 3, the MIS approximation ratio reaches a peak when h is around d , marking the critical regime of the Ising model Eq. (1) and corresponding to a positive independence constraint. Interestingly, however, for small p the maximal performance seems to be reached when $h \lesssim d - 2$ corresponding to an independence constraint $\lambda \lesssim 1$. This indicates that for QAOA with small p , it is not a requirement that h is chosen in such a way that the ground state of Eq. (1) corresponds to the maximum independent set. However, when p is increased and when therefore the energy distribution of QAOA shifts towards lower energies, the peak in optimal performance shifts towards a field that corresponds to an independence constraint $\lambda > 1$. This behavior can be seen in Fig. 4(B), where the boundaries of the region with independence constraint $\lambda > 1$ are indicated by the two dashed lines.

5.1.2 Fixing the local field

Now, we fix the local field to $h = 0$ for MaxCut and to $h = d - 2$ for MIS (corresponding to $\lambda = 1$), and investigate the performance of tree QAOA with increasing p for different graph degrees. For MaxCut the performance of tree QAOA has been investigated in the literature before [19, 21, 22, 26], it even has been considered for $p = 1$ in the original QAOA paper [11]. However, our current work extends these analyses by: (i) further improving the achieved approximation ratios by employing the upper bounds as discussed in Sec. 3, and (ii) extending the analysis to degrees

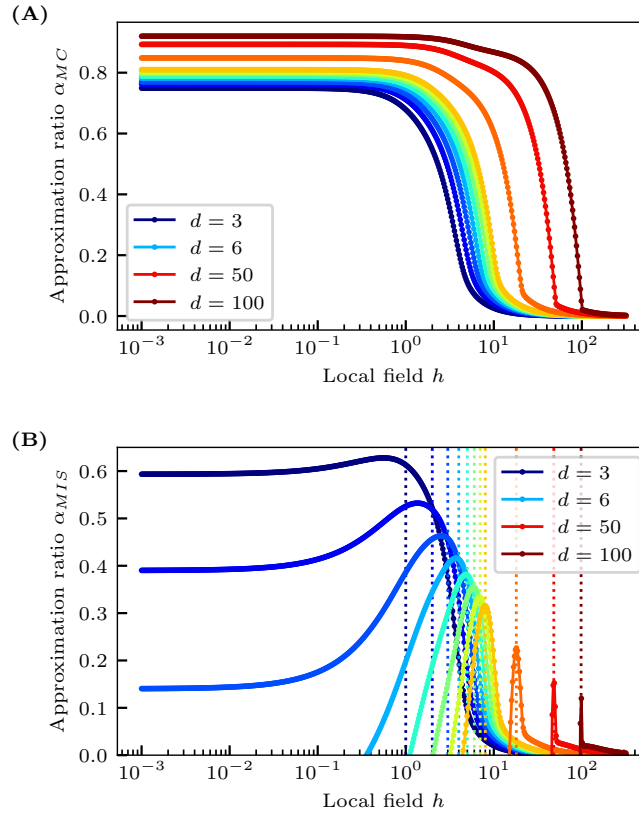


Figure 3: **The approximation ratios for (A) MaxCut and (B) MIS** obtained by $p = 1$ QAOA optimized as a function of the local field for $N \rightarrow \infty$. The colors indicate the different graph degrees listed in Table 1, only a selection is shown in the legend. The dotted vertical lines indicate local fields $h = d - 2$. The QAOA angles are chosen such that the energy density given by Eq. (4) is minimized [see Fig. 2]. The approximation ratios α_{MC} and α_{MIS} as discussed in Sec. 3, are evaluated on the resulting QAOA state.

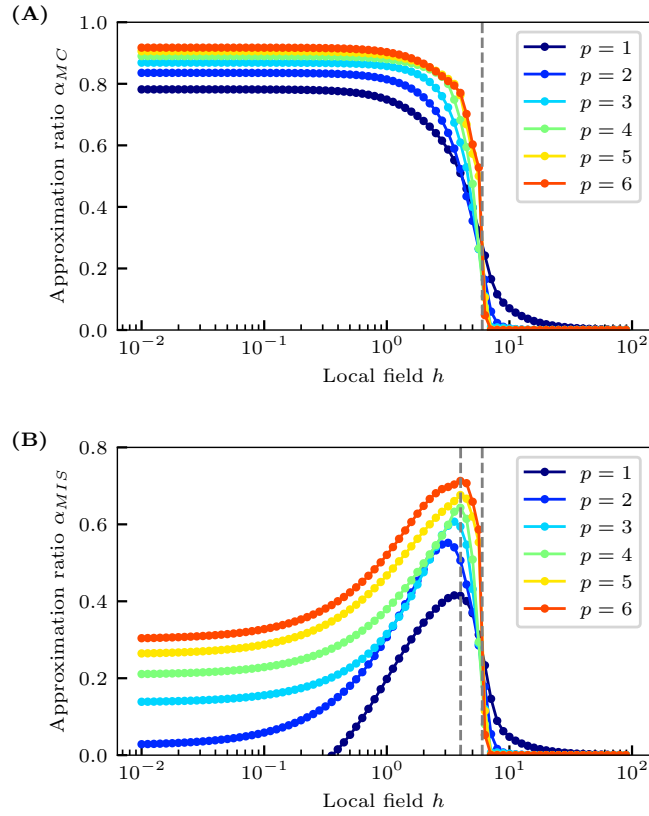


Figure 4: **The approximation ratios obtained by QAOA for (A) MaxCut, and (B) MIS on 6-regular graphs** for various p in the limit $N \rightarrow \infty$. The QAOA ansatz is optimized to minimize the energy given by Eq. (1). Therefore, there is a sharp transition in solution quality at $h = d$ for both problems. The ground state of Eq. (1) corresponds to the solution of MaxCut when $h \ll d$, and to the solution of MIS when $h \in [d-2, d]$. In these indicated regions the approximation ratio must therefore increase with p .

larger than $d = 3$. To make the first point explicit, it is stated in the original QAOA paper that on 3 regular graphs, QAOA is guaranteed to find approximation ratios of at least ~ 0.69 (also for $N \rightarrow \infty$). By incorporating the upper bound on the cut fraction of 3-regular graphs, we realize the approximation ratio is actually at least ~ 0.75 for $N \rightarrow \infty$, see Fig. 5(A). Here, we compare the improved approximation ratios obtained by QAOA, with the guaranteed approximation ratio of the Goemans-Williamson (GW) algorithm [27] and random sampling. The GW algorithm is guaranteed to find a solution to the MaxCut problem with an approximation ratio of at least $\alpha_{GW} = \min_{x \in [0,1]} 2 \arccos(x)/(\pi(1-x)) = 0.87\dots$ for any instance. Indeed, denoting the cut fraction of an average GW solution as c_{GW} , we have that

$$\alpha_{GW} c_{opt} \leq \alpha_{GW} c_{sdp} \leq c_{GW},$$

where c_{opt} is the cut fraction corresponding to the maximum cut, and c_{sdp} the (optimal) solution of the semi-definite program following the relaxation of the binary variables to unit vectors on the N -dimensional sphere. The first inequality directly follows from the fact this solution provides an upper bound on c_{opt} , i.e. $c_{opt} \leq c_{sdp}$, because it is a relaxation. The second inequality follows from the fact that the GW rounding produces a cut whose size is on average at least a fraction α_{GW} of the semi-definite relaxation optimum.

Thus, the knowledge of a tighter upper bound on c_{opt} does not alter the performance guarantee α_{GW} of the GW algorithm. So the approximation ratios introduced in Sec. 3 compare directly to α_{GW} , and already pass α_{GW} at $p = 4$ for the lowest degrees. However, we expect that the practically achieved GW approximation ratios will improve with increasing d . In particular, the GW results will never fall below random sampling $(2c_{ub})^{-1}$ shown by the red line in Fig. 5(A). On the other hand, for small d , we already pass α_{GW} at $p = 4$ which leaves open the possibility that shallower circuits than previously assumed could already provide a performance and runtime advantage over the GW algorithm.

We also note that in Fig. 5(A), the gap in approximation ratio between random sampling and QAOA shrinks upon increasing d . Based on this, it is also an interesting question of how QAOA performs relative to random sampling. For this, we can easily define a modified approximation ratio resembling the Q-score ratio defined in Ref. [69] (see Appendix A.4). Then, we can conclude that similarly to the local classical algorithm of Ref. [70], QAOA achieves constant approximation ratios relative to random sampling in the large d limit. This finding is consistent with the observation of a vanishing gap between random sampling and QAOA in Fig. 5(A). We also determine the constant explicitly in Appendix A.5 for $p = 1$ and conclude that

$$\alpha_{MC}^{\text{relative}}(d \rightarrow \infty, p = 1) = \frac{1}{2\sqrt{e}P^*} \approx 0.397\dots,$$

meaning that QAOA with $p = 1$ is around 40% better than random sampling in the large d limit.

Now, we analyze the performance of tree QAOA for solving MIS problem. We show the obtained approximation ratios in Fig. 5(B). We compare to the performance guarantee of the minimal greedy algorithm. The minimal greedy algorithm makes locally minimal random choices. It consists out of the following steps: (i) Randomly select a vertex from G that has the lowest degree. (ii) Add this vertex to the independent set, and delete all its neighbors from G . (iii) Repeat (i) and (ii) until the the graph G is empty (i.e. has no edges left). (iv) Add possible remaining vertices to the independent set. This algorithm is guaranteed to find independent sets with approximation ratios $3/(d+2)$ [72]. We also compare to the best known classical linear-time algorithm of Ref. [61], see dashed grey line. (These approximation ratios are the values that we considered as a lower bound for μ^* in Table 2.) As can be seen, there is an increasing gap with d between the approximation reached by QAOA and this algorithm. Therefore, it seems that the performance of tree QAOA for MIS is vanishing upon increasing d . This observation can be made explicit for $p = 1$ [see Appendix A.5]. We would expect this to be true as well when $p > 1$ and finite, because in the limit $N \rightarrow \infty$ we would expect that the locality restrictions apply in the same way for $p > 1$ as for $p = 1$. In practice, we are however mostly interested in solving the problems for small and fixed d . Then, the optimal parameters obtained from the tree QAOA could also be used for small problems, of problem size corresponding to the currently available QPU's. Illustrating that this approach achieves good performances is the topic of the next section.

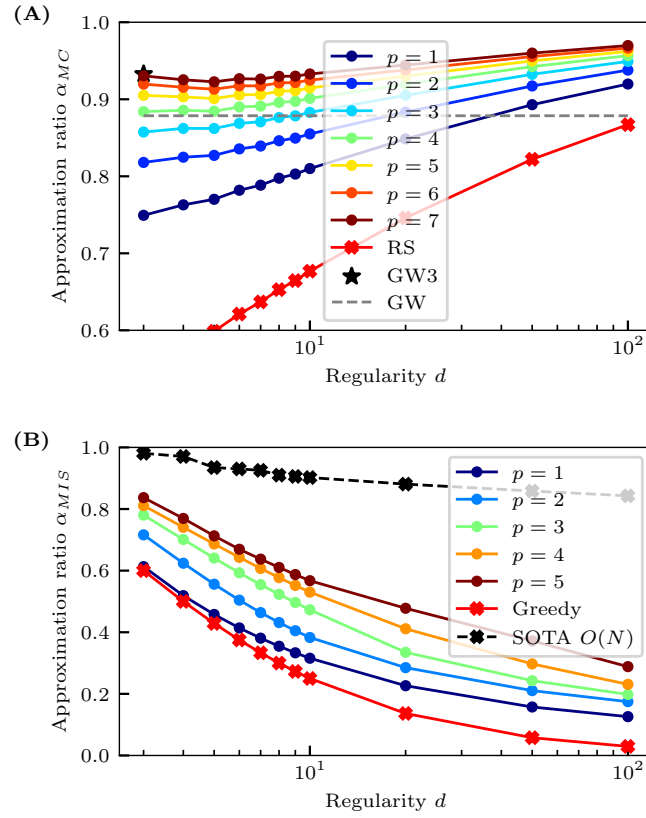


Figure 5: **The approximation ratios for (A) MaxCut and (B) MIS** obtained by depth- p QAOA for $N \rightarrow \infty$. For MaxCut, we compare with the GW guarantee and Random Sampling (RS). For 3-regular graphs there is a specialized version of the GW algorithm, discussed in Ref. [71], achieving a worst-case approximation ratio of 0.9326 which is indicated by the star. For MIS, we compare to the performance guarantee of a minimal greedy search [72], and the state-of-the-art linear-time prioritized search algorithm of Ref. [61].

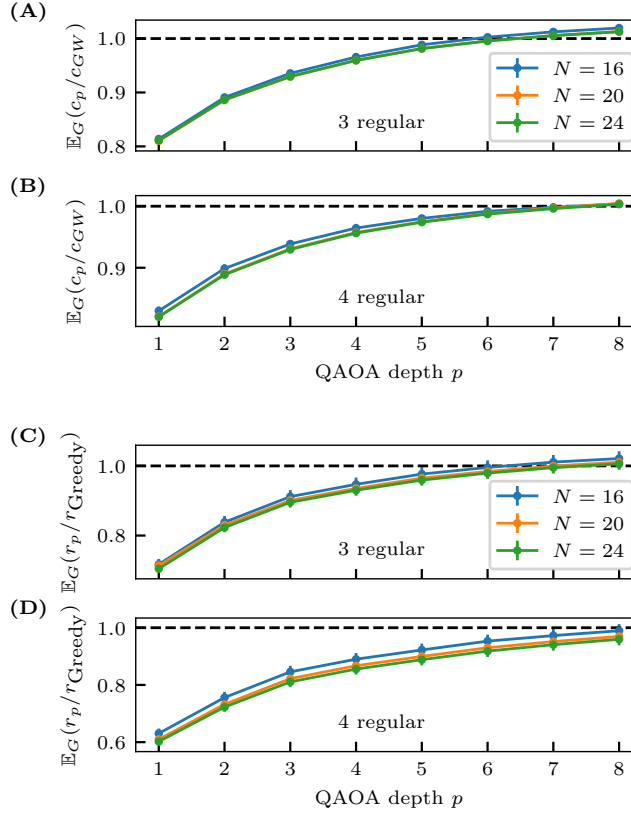


Figure 6: **Performance comparison between QAOA with predetermined tree angles and classical algorithms** on small instances of 3 and 4 regular graphs. The tree angles have been determined by the techniques outlined in Appendix A. **(A,B) MaxCut QAOA compared to GW.** The comparison is made on an instance-by-instance basis: we compare the expectation value of the number of cut edges measured on the QAOA state with the average number of cut edges based on 100 GW runs. **(C,D) MIS QAOA compared to minimal greedy.** We sampled 200 random graph instances, and the error bars show $3\bar{\sigma}$ with $\bar{\sigma}$ the standard error of the mean. The QAOA angles used for these simulations are included in Appendix B.

5.2 Tree QAOA angles applied to finite-size problems

The goal of this section is to compare how well QAOA with the tree angles obtained in the previous section performs for finite-size instances. In the previous section [see Fig. 5], we compared the expected QAOA outcomes with the performance guarantees of classical algorithms: GW for MaxCut and minimal greedy for MIS. However, it may be that in practice these classical algorithms perform significantly better than their lower performance bound. At the same time, the tree angles are only guaranteed to be optimal in the asymptotic limit. However, for MaxCut, it has been observed before that these angles also work well away from the asymptotic limit [21, 26].

To shine a clearer light on these issues, our goal is to make an explicit comparison for both MaxCut and MIS to the classical algorithms. The protocol we use is the following: (i) Use a set of tree angles determined as before (by making use of the formulas in Appendix A) to create a QAOA ansatz state for the sampled finite-size problem instances $G(N, d)$. Hence, we do not apply a procedure to fine tune the QAOA angles, but use the same set of fixed angles for all instances. (ii) Measure the performance metrics, i.e. the expectation values of Sec. 3, on this QAOA ansatz state. (iii) Run a classical probabilistic algorithm on the same instance a certain number of times, and take the average. (iv) Compare the outcomes of both algorithms.

For MaxCut, we set $h = 0$ in Eq. (1) and compare the fixed angle QAOA to the GW algorithm [27]. We measure the expected number of cut edges in the prepared QAOA state for every instance. For each instance, we then run the GW algorithm 100 times, and take the average. We observed that in practice, in our setup, the GW performs on average significantly better than its

worst-case guarantee α_{GW} . This also explains why we only observe an improvement for $p = 8$ for 3 and 4 regular graphs, see Fig. 6(A,B). This is in contrast to the asymptotic limit, where tree QAOA already outperforms the *lower bound* at $p = 4$ [see Fig. 5]. Although the approximation ratios shown in Fig. 5 clearly increase when growing d , we observe for higher p a non-monotonic regime when d is small. This possibly explains the slightly worse performance for $d = 4$ compared to $d = 3$ in Fig. 6(A,B).

For MIS, we set $h = d - 2$ in Eq. (1) and compare to the minimal greedy algorithm. Such greedy algorithms are widely used algorithms, but their performance is far from optimal for *large graphs*, even among local algorithms. Indeed, from Fig. 5(B), it is clear that we are far from outperforming the state-of-the-art (SOTA) classical local linear-time algorithm that solves the MIS problem on large regular graphs [61]. However, in the extreme *finite-size scenario*, we actually found that the minimal greedy algorithm outperforms the SOTA algorithm which justifies the choice of using it for this comparison here. In addition, it has a straightforward implementation and a performance guarantee. This comparison is shown in Fig. 6(C,D) for 3 and 4 regular graphs. We see that tree QAOA is outperforming the minimal greedy algorithm for $p = 8$ in the case $d = 3$, however this is not anymore the case for $d = 4$. First, these results confirm that, similarly to the GW algorithm, also the greedy algorithm performs significantly better on average than its worst-case guarantee. Second, these results for QAOA on small system sizes seem to also reflect the $N \rightarrow \infty$ finding that the MIS problem becomes more challenging for QAOA as d increases.

The tree angles we have used for the simulations in this section are included in Appendix B.

6 Conclusion

In this work, we have provided a comprehensive analysis of the performance landscape of tree QAOA for the MaxCut and MIS problems on random d -regular graph by unifying them as Ising models. Our approach extends the recursive formula technique introduced in Ref. [19] to handle Ising models with a local field, allowing us to evaluate the QAOA performance for both problems for large but finite graph degrees up to $d = 100$. We combined these results with state-of-the-art upper bounds on optimality which led to our main result: the MaxCut and MIS approximation ratios achieved by QAOA in the limit $N \rightarrow \infty$ shown in Fig. 5 for many different graph degrees d .

Our analysis demonstrates that QAOA outperforms the Goemans-Williamson (GW) guarantee at much shallower depths than previously assumed. Specifically, we show that asymptotic tree QAOA can surpass GW for MaxCut when $p \geq 4$ in the large graph limit ($N \rightarrow \infty$), rather than requiring a depth of $p \geq 11$ [26].

For the MIS problem, QAOA at any depth $p \geq 1$ outperforms the lower bound of the greedy algorithm. However, QAOA is far from outperforming the best classical local polynomial-time algorithm for this problem [61]. For higher degree, the performance of QAOA with fixed p is bounded away from optimality for the MIS problem due to the presence of OGP [37, 38]. The increased hardness of the MIS problem with increasing d is clearly manifested by the decrease of approximation ratios achieved by tree QAOA [see Fig. 5]. This behavior is opposite for MaxCut which is believed not to exhibit an OGP. In this case the QAOA approximation ratios are indeed not vanishing with increasing d .

In addition, we also showed that QAOA has a promising performance when the tree angles are taken as ‘fixed angles’ for small problems. For MaxCut, this setting is similar as in Ref. [22] and we showed that QAOA with fixed angles outperforms the GW algorithm at $p = 8$ for $d = 3, 4$ when comparing on average to explicit GW runs. Such an analysis seems to also partly hold for the MIS problem when comparing to minimal greedy at low degree. Indeed, in the case of $d = 3$, we have outperformed the minimal greedy algorithm with QAOA at depth $p = 8$ when comparing to explicit greedy runs for small instances. However, when increasing the problem size, the minimal greedy algorithm performs strongly suboptimal.

Our findings highlight the importance of problem structure in determining the algorithmic performance. Problems such as MaxCut, which lack an OGP regime can be effectively addressed by QAOA with a shallow number of layers. In contrast, problems with an OGP regime, like MIS, require more sophisticated quantum algorithms that break locality to achieve near-optimal solutions.

In order to boost performance by breaking locality, different forms of the mixing Hamiltonian could be considered [73], this also allows to ‘warm start’ from a good initial state obtained by a classical algorithm. We however also believe that hybrid quantum-classical approaches could be promising [52–58]. Such approaches combine quantum input, for instance $\langle Z_i \rangle$ and $\langle Z_i Z_j \rangle$ measurements on a quantum state, e.g. a QAOA state, with a classical algorithm that modifies the problem structure. In particular, when the quantum circuit is non-local and sees the whole graph such approaches may be very effective, although less reachable on the near term. It is our hope that the tree QAOA angles will serve useful for these approaches, especially for the MIS problem, and that expected ‘fixed-angle’ performances [31] will also be derived for such approaches. In that way, expensive and hard parameter optimization loops can be avoided [74–76].

Acknowledgements – We thank Thomas Cope and Jalil Khatibi Moqadam for useful discussions. We also thank Thomas Cope for careful reading and comments on the manuscript. This project is supported by the Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag through the project Quantum-enabling Services and Tools for Industrial Applications (QuaST). QuaST aims to facilitate the access to quantum-based solutions for optimization problems.

Code availability – We provide an implementation of the recursive formulas of Appendix A.3 and pre-calculated tree angles in our publicly available Python package `iqm-qaoa` <https://docs.meetiqm.com/iqm-qaoa/>.

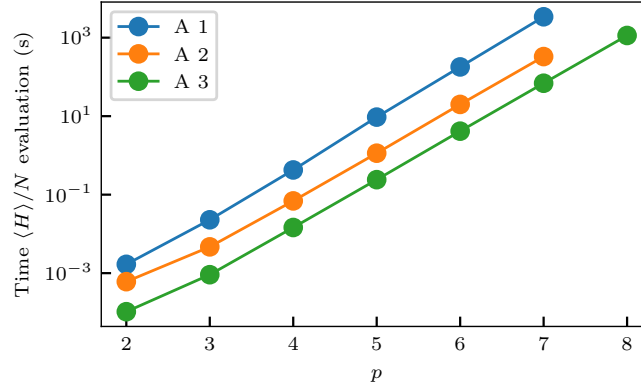


Figure 7: The computational time, on a single core of a standard laptop, to evaluate Eq. (4) at fixed QAOA angles using the formulas derived respectively in Sections A.1, A.2 and A.3. The speedups obtained are prefactors, the overall scaling remains $O(2^{4p})$.

A Calculation of the QAOA expectation values on trees

Our goal is to derive expressions that analytically describe the expectation values in Eq. (4) as recursive formulas. These ideas have been pioneered in Refs. [19, 35]. In particular Ref. [19] contains very elegant expressions for the QAOA correlators in the $d \rightarrow \infty$ limit. However the finite- d case received less attention in their work. The calculations presented in this work are in close correspondence to the ones presented in Ref. [19] with the following additions, each discussed in a different section: (i) We have a local field in our model (1), and thus have no $\mathbb{Z}_2$ symmetry, resulting in less symmetric expressions, see A.1. (ii) We grow the basis size at every recursion step, resulting in a faster recursion, see A.2. (iii) We associate a symmetry label t with every basis state, characterizing how ‘time-reversal’ symmetric it is, see A.3. The corresponding computational times of evaluating Eq. (4) according to the three different procedures are shown in Fig. 7. In Fig. 8, we sketch symbolically how these speedups are achieved.

A.1 A recursive iteration with local field

In this section, our goal is to derive similar expressions as in Ref. [19] that take into account a local field. For this, we work out

$$\langle \gamma, \beta | Z_1 Z_2 | \gamma, \beta \rangle^{2\text{-tree}} = \langle + | e^{i\gamma_1 H} e^{i\beta_1 B} \dots e^{i\gamma_p H} e^{i\beta_p B} Z_1 Z_2 e^{-i\beta_p B} e^{-i\gamma_p H} \dots e^{-i\beta_1 B} e^{-i\gamma_1 H} | + \rangle,$$

and

$$\langle \gamma, \beta | Z_1 | \gamma, \beta \rangle^{1\text{-tree}} = \langle + | e^{i\gamma_1 H} e^{i\beta_1 B} \dots e^{i\gamma_p H} e^{i\beta_p B} Z_1 e^{-i\beta_p B} e^{-i\gamma_p H} \dots e^{-i\beta_1 B} e^{-i\gamma_1 H} | + \rangle.$$

Here, H is given by Eq. (1) and is defined on the trees shown in Fig. 1. These trees are fully characterized by the degree d and the QAOA depth p . As a first step, we can insert $2p + 1$ resolutions of identity, labelled by superscripts $1, 2, \dots, p, 0, -p, \dots, -2, -1$. The basis vectors $\tilde{z}^{[m]}$ have a dimension that is equal to the number of variables in the tree, i.e.

$$\tilde{z}^{[m]} = (z_1^{[m]}, z_2^{[m]}, \dots, z_{N_{\text{tree}}}^{[m]}), \quad z_i^{[m]} \in \{-1, 1\} \text{ with } i = 1, \dots, N_{\text{tree}}$$

For the two different tree variants, the number of variables are

$$N_{2\text{-tree}} = 2 \frac{(d-1)^{p+1} - 1}{d-2}, \quad N_{1\text{-tree}} = 1 + d \frac{(d-1)^p - 1}{d-2}$$

With this, we have that

$$\begin{aligned} & \langle \gamma, \beta | Z_1 Z_2 | \gamma, \beta \rangle^{2\text{-tree}} \\ &= \frac{1}{2^{N_{2\text{-tree}}}} \sum_{\{\tilde{z}\}} \left[z_1^{[0]} z_2^{[0]} \left(\langle \tilde{z}^{[1]} | e^{i\beta_1 B} | \tilde{z}^{[2]} \rangle \langle \tilde{z}^{[2]} | e^{i\beta_2 B} | \tilde{z}^{[3]} \rangle \dots \langle \tilde{z}^{[p]} | e^{i\beta_p B} | \tilde{z}^{[0]} \rangle \right. \right. \\ & \quad \left. \langle \tilde{z}^{[0]} | e^{-i\beta_p B} | \tilde{z}^{[-p]} \rangle \dots \langle \tilde{z}^{[-3]} | e^{-i\beta_2 B} | \tilde{z}^{[-2]} \rangle \langle \tilde{z}^{[-2]} | e^{-i\beta_1 B} | \tilde{z}^{[-1]} \rangle \right) \\ & \quad \left. \exp \left(i\gamma_1 H(\tilde{z}^{[1]}) + \dots + i\gamma_p H(\tilde{z}^{[p]}) - i\gamma_1 H(\tilde{z}^{[-1]}) - \dots - i\gamma_p H(\tilde{z}^{[-p]}) \right) \right]. \end{aligned}$$

Here $\sum_{\{\tilde{z}\}}$ means summing over all $2p+1$ basis sets, that each have dimension $2^{N_{2\text{-tree}}}$, $H(\tilde{z}^{[m]})$ represents the energy of the bitstring. Using the same notation as in Ref. [19], we now define the following vectors of length $(2p+1)$

$$\mathbf{\Gamma} = (\gamma_1, \dots, \gamma_p, 0, -\gamma_p, \dots, -\gamma_1),$$

and

$$\mathbf{z}_k = (z_k^{[1]}, \dots, z_k^{[p]}, z_k^{[0]}, z_k^{[-p]}, \dots, z_k^{[-1]}), \quad z_k^{[m]} \in \{-1, 1\} \text{ with } k \in V_{\text{tree}}.$$

Then, we have that the mixing part of the above expression factorizes as

$$\begin{aligned} & \prod_{k \in V_{\text{tree}}} \left(\langle z_k^{[1]} | e^{i\beta_1 X_k} | z_k^{[2]} \rangle \langle z_k^{[2]} | e^{i\beta_2 X_k} | z_k^{[3]} \rangle \dots \langle z_k^{[p]} | e^{i\beta_p X_k} | z_k^{[0]} \rangle \right. \\ & \quad \left. \langle z_k^{[0]} | e^{-i\beta_p X_k} | z_k^{[-p]} \rangle \dots \langle z_k^{[-3]} | e^{-i\beta_2 X_k} | z_k^{[-2]} \rangle \langle z_k^{[-2]} | e^{-i\beta_1 X_k} | z_k^{[-1]} \rangle \right) \\ &= 2^{|V_{\text{tree}}|} \prod_{k \in V_{\text{tree}}} f(\mathbf{z}_k) \end{aligned}$$

where we defined the function $f(\mathbf{z}_k) \equiv f(\mathbf{z}_k; \beta_1, \dots, \beta_p)$

$$\begin{aligned} f(\mathbf{z}_k) &= \frac{1}{2} \langle z_k^{[1]} | e^{i\beta_1 X_k} | z_k^{[2]} \rangle \langle z_k^{[2]} | e^{i\beta_2 X_k} | z_k^{[3]} \rangle \dots \langle z_k^{[p]} | e^{i\beta_p X_k} | z_k^{[0]} \rangle \\ & \quad \langle z_k^{[0]} | e^{-i\beta_p X_k} | z_k^{[-p]} \rangle \dots \langle z_k^{[-3]} | e^{-i\beta_2 X_k} | z_k^{[-2]} \rangle \langle z_k^{[-2]} | e^{-i\beta_1 X_k} | z_k^{[-1]} \rangle. \end{aligned}$$

Here, the expectation values are

$$\langle z_k^{[m]} | e^{i\beta_i X_k} | z_k^{[n]} \rangle = \begin{cases} \cos(\beta_i), & \text{if } z_k^{[m]} = z_k^{[n]} \\ i \sin(\beta_i), & \text{otherwise,} \end{cases}$$

and normalization implies that $\sum_{\mathbf{z}_k} f(\mathbf{z}_k) = 1$. With these notations, we have that

$$\begin{aligned} & \langle \gamma, \beta | Z_1 Z_2 | \gamma, \beta \rangle^{2\text{-tree}} \\ &= \sum_{\{\mathbf{z}\}} \left[z_1^{[0]} z_2^{[0]} \left(\prod_{k \in V_{2\text{-tree}}} f(\mathbf{z}_k) \right) \exp \left(i \frac{1}{\sqrt{d}} \sum_{kl \in E_{2\text{-tree}}} \mathbf{\Gamma} \cdot \mathbf{z}_k \mathbf{z}_l \right) \exp \left(i \frac{h}{\sqrt{d}} \sum_{k \in V_{2\text{-tree}}} \mathbf{\Gamma} \cdot \mathbf{z}_k \right) \right]. \quad (13) \end{aligned}$$

Here $\sum_{\{\mathbf{z}\}}$ should now be read as summing over all $N_{2\text{-tree}}$ basis sets, that each have dimension 2^{2p+1} , $\mathbf{z}_k \mathbf{z}_l = (z_k^{[1]} z_l^{[1]}, z_k^{[2]} z_l^{[2]}, \dots)$ is the elementwise product of the two bitstrings, and $\mathbf{\Gamma} \cdot \mathbf{z}_k$ the standard inner product between vectors. Up to now we have not exploited the tree structure yet. This structure is however quite easy to exploit: there are only connections between generations and therefore it is natural to rearrange the summations such that we start summing from the outer children (or leaves) of the tree. Let us label an outer child by the position index u , and call the parent of that child $p(u)$. The dependency in Eq. (13) on this outer leaf can be separated as

$$\sum_{\mathbf{z}_u} f(\mathbf{z}_u) \exp \left(i \frac{1}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{z}_u \mathbf{z}_{p(u)} \right) \exp \left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{z}_u \right).$$

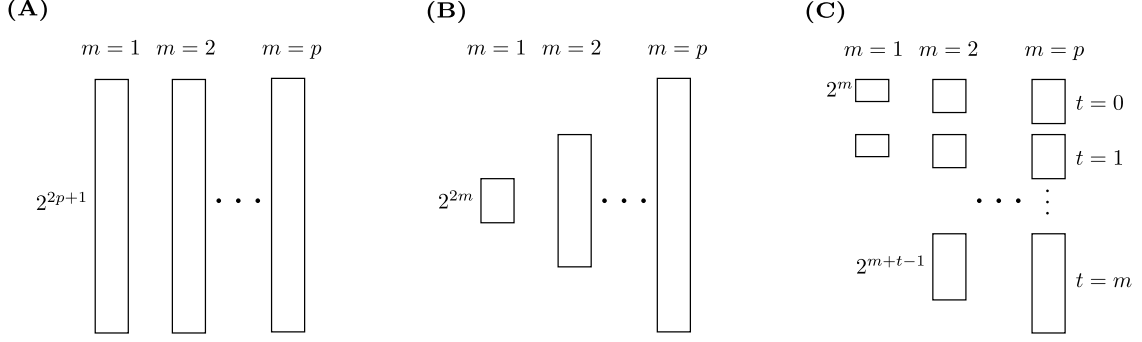


Figure 8: Structure of the basis at every iteration step $m \in \{1, 2, \dots, p\}$. (A) In the most naive way the basis has the same size 2^{2p+1} at each step. (B) For the first simplification, we grow the basis size iteratively. (C) For the second simplification, we split the basis in symmetry sectors, labelled by $t \in \{0, \dots, m\}$.

As there are exactly $d - 1$ (independent) incoming leaves in the parent node $p(u)$, we can define

$$H_{d-1}^{(1)}(\mathbf{z}_{p(u)}) = \left[\sum_{\mathbf{z}_u} f(\mathbf{z}_u) \exp\left(i \frac{1}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{z}_u \mathbf{z}_{p(u)}\right) \exp\left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{z}_u\right) \right]^{d-1}. \quad (14)$$

Then, we can move inwards and separate the dependencies of $\mathbf{z}_{p(u)}$, and so on. This will lead to the iterations

$$H_{d-1}^{(m)}(\mathbf{a}) = \left[\sum_{\mathbf{b}} f(\mathbf{b}) H_{d-1}^{(m-1)}(\mathbf{b}) \exp\left(i \frac{1}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{a} \mathbf{b}\right) \exp\left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{b}\right) \right]^{d-1}, \quad (15)$$

where we have lightened the notation and now simply assume that $\mathbf{a}$ is a bitstring associated to a parent, and $\mathbf{b}$ with a child. We also have that $H_{d-1}^{(0)} \equiv 1$. After p iterations, only the local fields and the edge between the two roots (labelled by 1 and 2) of the 2-tree remain, so finally we have that

$$\begin{aligned} & \langle \gamma, \beta | Z_1 Z_2 | \gamma, \beta \rangle^{2\text{-tree}} \\ &= \sum_{\mathbf{z}_1 \mathbf{z}_2} z_1^{[0]} z_2^{[0]} f(\mathbf{z}_1) f(\mathbf{z}_2) H_{d-1}^{(p)}(\mathbf{z}_1) H_{d-1}^{(p)}(\mathbf{z}_2) \exp\left(i \frac{1}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{z}_1 \mathbf{z}_2\right) \exp\left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma} \cdot (\mathbf{z}_1 + \mathbf{z}_2)\right). \end{aligned}$$

For the 1-tree, the procedure is exactly the same. However, the root (labelled by 1) has now d equivalent branches. Therefore, the final iteration needs to be modified to

$$H_d^{(p)}(\mathbf{a}) = \left[\sum_{\mathbf{b}} f(\mathbf{b}) H_{d-1}^{(p-1)}(\mathbf{b}) \exp\left(i \frac{1}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{a} \mathbf{b}\right) \exp\left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{b}\right) \right]^d.$$

while the earlier iterations remain unchanged. Therefore, the local expectation value becomes

$$\langle \gamma, \beta | Z_1 | \gamma, \beta \rangle^{1\text{-tree}} = \sum_{\mathbf{z}_1} z_1^{[0]} f(\mathbf{z}_1) H_d^{(p)}(\mathbf{z}_1) \exp\left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{z}_1\right).$$

The time complexity is the same for every step in the recursion in this case. Therefore the overall time complexity for the evaluation of the energy density [see Eq. (4)] is $O((p+1)2^{4p+2})$. As a difference with Ref. [19], note that $H_d^{(m)}(\mathbf{a})$ is generally not real anymore. This is a consequence of breaking the $\mathbb{Z}_2$ symmetry when $h \neq 0$.

A.2 First speed-up: Expanding the basis size at every iteration

In the last section, we wrote down similar expressions as in Ref. [19] with a local field included. In this section, we will discuss a first speed-up of this iterative procedure. Up to now, we did not

take into account the explicit angle dependencies, that can be seen in Fig. 1(C). Taking this into account does, of course, not alter the final result, as unitary cancellations are implicit. However, taking this into account means that, for example, $H_{d-1}^{(1)}(\mathbf{a})$ should only depend on γ_1 , $H_{d-1}^{(2)}(\mathbf{a})$ only on $\gamma_1, \gamma_2, \beta_1$, and so on. This is however not the case in Eq. (15), but can be easily imposed yielding more compact iterations. We start by imposing the independence in Eq. (14), this gives

$$H_{d-1}^{(1)}(\mathbf{a}) \rightarrow H_{d-1}^{(1)}(\mathbf{a}^{(1)}) = \left[\cos\left(\frac{1}{\sqrt{d}} \mathbf{\Gamma}^{(1)} \cdot \mathbf{a}^{(1)}\right) \right]^{d-1} = \left[\cos\left(\frac{\gamma_1}{\sqrt{d}}(a_1 - a_{-1})\right) \right]^{d-1} \quad (16)$$

where we define the vectors of length $2m$

$$\mathbf{\Gamma}^{(m)} = (\gamma_1, \dots, \gamma_m, -\gamma_m, \dots, -\gamma_1),$$

and (to unify the notation with the QAOA angles, the subscripts now relate to the QAOA depth)

$$\mathbf{a}^{(m)} = (a_1, \dots, a_m, a_{-m}, \dots, a_{-1}), \quad a_i \in \{-1, 1\}. \quad (17)$$

Doing the same in the general case $H_{d-1}^{(m)}(\mathbf{a}) \rightarrow H_{d-1}^{(m)}(\mathbf{a}^{(m)})$ gives

$$\begin{aligned} H_{d-1}^{(m)}(\mathbf{a}^{(m)}) = & \quad (18) \\ & \left[\sum_{\mathbf{b}^{(m-1)}} g(\mathbf{b}^{(m-1)}) H_{d-1}^{(m-1)}(\mathbf{b}^{(m-1)}) \exp\left(i \frac{1}{\sqrt{d}} \mathbf{\Gamma}^{(m-1)} \cdot \mathbf{a}^{(m-1)} \mathbf{b}^{(m-1)}\right) \exp\left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma}^{(m-1)} \cdot \mathbf{b}^{(m-1)}\right) \right. \\ & \quad \frac{1}{2} \left(\langle b_{m-1} | e^{i\beta_{m-1}X} | 1 \rangle \langle 1 | e^{-i\beta_{m-1}X} | b_{-m+1} \rangle \exp\left(i \frac{\gamma_m}{\sqrt{d}}(a_m - a_{-m})\right) \right. \\ & \quad \left. \left. + \langle b_{m-1} | e^{i\beta_{m-1}X} | -1 \rangle \langle -1 | e^{-i\beta_{m-1}X} | b_{-m+1} \rangle \exp\left(-i \frac{\gamma_m}{\sqrt{d}}(a_m - a_{-m})\right) \right) \right]^{d-1}. \end{aligned}$$

In the last two lines, we have performed the summation over $b_m = b_{-m}$ explicitly. The function with reduced beta dependencies is now defined as

$$\begin{aligned} g(\mathbf{b}^{(m-1)}) &= \\ &= \langle b_1 | e^{i\beta_1 X} | b_2 \rangle \dots \langle b_{m-2} | e^{i\beta_{m-2} X} | b_{m-1} \rangle \langle b_{-(m-1)} | e^{-i\beta_{m-2} X} | b_{-(m-2)} \rangle \dots \langle b_{-2} | e^{-i\beta_1 X} | b_{-1} \rangle. \end{aligned}$$

This follows from the elimination of redundant variables in $f(\mathbf{b})$

$$\begin{aligned} \sum_{\substack{b_{m+1} \dots b_p b_0 \\ b_{-p} \dots b_{-(m-1)}}} f(\mathbf{b}) &= \frac{1}{2} \sum_{\substack{b_{m+1} \dots b_p b_0 \\ b_{-p} \dots b_{-(m-1)}}} \langle b_1 | e^{i\beta_1 X} | b_2 \rangle \langle b_2 | e^{i\beta_2 X} | b_3 \rangle \dots \langle b_p | e^{i\beta_p X} | b_0 \rangle \\ &\quad \langle b_0 | e^{-i\beta_p X} | b_{-p} \rangle \dots \langle b_{-3} | e^{-i\beta_2 X} | b_{-2} \rangle \langle b_{-2} | e^{-i\beta_1 X} | b_{-1} \rangle \\ &= \frac{1}{2} \langle b_1 | e^{i\beta_1 X} | b_2 \rangle \dots \langle b_{m-1} | e^{i\beta_{m-1} X} | b_m \rangle \langle b_m | b_{-m} \rangle \\ &\quad \langle b_{-m} | e^{-i\beta_{m-1} X} | b_{-(m-1)} \rangle \dots \langle b_{-2} | e^{-i\beta_1 X} | b_{-1} \rangle \\ &= \frac{1}{2} g(\mathbf{b}^{(m-1)}) \langle b_{m-1} | e^{i\beta_{m-1} X} | b_m \rangle \langle b_m | e^{-i\beta_{m-1} X} | b_{-(m-1)} \rangle. \end{aligned}$$

Then the final result can be modified to

$$\begin{aligned} \langle \gamma, \beta | Z_1 Z_2 | \gamma, \beta \rangle^{2\text{-tree}} &= \\ & \sum_{\mathbf{a}^{(p)} \mathbf{b}^{(p)}} g(\mathbf{a}^{(p-1)}) g(\mathbf{b}^{(p-1)}) H_{d-1}^{(p)}(\mathbf{a}^{(p)}) H_{d-1}^{(p)}(\mathbf{b}^{(p)}) \exp\left(i \frac{1}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{a}^{(p)} \mathbf{b}^{(p)}\right) \exp\left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma} \cdot (\mathbf{a}^{(p)} + \mathbf{b}^{(p)})\right) \\ & \quad \sum_{a_0 b_0} \frac{a_0 b_0}{4} \langle a_p | e^{i\beta_p X} | a_0 \rangle \langle a_0 | e^{-i\beta_p X} | a_{-p} \rangle \langle b_p | e^{i\beta_p X} | b_0 \rangle \langle b_0 | e^{-i\beta_p X} | b_{-p} \rangle. \end{aligned}$$

t	# bitstrings	bitstring
0	2^m	$\mathbf{a}_0^{(m)} = (a_1, \dots, a_{m-1}, a_m, a_m, a_{m-1}, \dots, a_1)$
1	2^m	$\mathbf{a}_1^{(m)} = (a_1, \dots, a_{m-1}, a_m, a_m, a_{m-1}, \dots, -a_1)$
2	2^{m+1}	$\mathbf{a}_2^{(m)} = (a_1, a_2, \dots, a_{m-1}, a_m, a_m, a_{m-1}, \dots, -a_2, a_{-1})$
$\vdots$	$\vdots$	$\vdots$
m	2^{2m-1}	$\mathbf{a}_m^{(m)} = (a_1, \dots, a_{m-1}, a_m, -a_m, a_{-(m-1)}, \dots, a_{-1})$

Table 3: Labelling of the basis $\{\mathbf{a}^{(m)}\}$, $|\{\mathbf{a}^{(m)}\}| = 2^{2m}$, according to t .

The summation over a_0 and b_0 can be worked out, reducing the scaling by a prefactor to $O(2^{4p})$. The onsite term simplified in this way becomes

$$\langle \gamma, \beta | Z_1 | \gamma, \beta \rangle^{1\text{-tree}} = \sum_{\mathbf{a}^{(p)}} g(\mathbf{a}^{(p-1)}) H_{d-1}^{(p)}(\mathbf{a}^{(p)}) \exp\left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma} \cdot \mathbf{a}^{(p)}\right) \sum_{a_0} \frac{a_0}{2} \langle a_p | e^{i\beta_p X} | a_0 \rangle \langle a_0 | e^{-i\beta_p X} | a_{-p} \rangle.$$

A.3 Second speed-up: Decomposing the basis into T-symmetric blocks

In this section, we discuss our last speed-up that is related to a splitting of the basis in blocks that carry a symmetry label. Some properties of these have been discussed in the Appendix of Ref. [19], with as goal providing simpler iterations in the $d \rightarrow \infty$ limit, but remain unexploited for the finite d iteration. To every bitstring of the form $\mathbf{a}^{(m)}$ [see Eq. (17)], we can associate a symmetry label $t \in \{0, 1, \dots, m\}$, $\mathbf{a}^{(m)} \rightarrow \mathbf{a}_t^{(m)}$ that characterizes how reflection symmetric the bitstring is with respect to the middle, i.e. when $t = 0$ the bitstring is completely symmetric, when $t = m$ it is not symmetric. This is summarized in Table 3. This label characterizes the fixed points in the iterative procedure. This is the most straightforward to see for $t = 0$, where we have that $H_d^{(m)}(\mathbf{a}_0^{(m)}) = 1$, $\forall m$. We can show this by induction. From Eq. (16) we see that this is indeed true for $m = 1$. Assuming that this is true at level $m - 1$, we have for level m [see Eq. (18)]

$$\begin{aligned} H_{d-1}^{(m)}(\mathbf{a}_0^{(m)}) &= \left[\sum_{\mathbf{b}^{(m-1)}} g(\mathbf{b}^{(m-1)}) \frac{1}{2} \langle b_{m-1} | b_{-(m-1)} \rangle \exp\left(i \frac{1}{\sqrt{d}} \mathbf{\Gamma}^{(m-1)} \cdot \mathbf{a}_0^{(m-1)} \mathbf{b}^{(m-1)}\right) \right. \\ &\quad \left. \exp\left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma}^{(m-1)} \cdot \mathbf{b}^{(m-1)}\right) \right]^{d-1} \\ &= \dots = \left[\sum_{b_1, b_{-1}} \frac{1}{2} \langle b_1 | b_{-1} \rangle \exp\left(i \frac{1}{\sqrt{d}} \gamma_1 a_1 (b_1 - b_{-1})\right) \exp\left(i \frac{h}{\sqrt{d}} \gamma_1 (b_1 - b_{-1})\right) \right]^{d-1} \\ &= 1. \end{aligned}$$

Similarly, if $t < m$ we have that $H_d^{(m)}(\mathbf{a}_t^{(m)}) = H_d^{(t)}(\mathbf{a}_t^{(t)})$. Here $\mathbf{a}^{(t)}$ is obtained from $\mathbf{a}^{(m)}$ by pruning away the symmetric bits, i.e.

$$\begin{aligned} \mathbf{a}_t^{(m)} &= (a_1, \dots, a_t, a_{t+1}, \dots, a_m, a_m, \dots, a_{t+1}, -a_t, \dots, a_{-1}) \\ &\rightarrow \mathbf{a}_t^{(t)} = (a_1, \dots, a_t, -a_t, \dots, a_{-1}). \end{aligned}$$

The proof is similar as before, indeed, from Eq. (18) we can immediately see that a sequence of simplifications is entered when $a_m = a_{-m}$. This means that at each iteration level m , we ‘only’ need to evaluate the 2^{2m-1} bitstrings that have $t = m$.

We can reduce this further by a factor of two by realizing that $H_{d-1}^{(m)}(R(\mathbf{a}_m^{(m)})) = [H_{d-1}^{(m)}(\mathbf{a}_m^{(m)})]^*$, where R is the reflection operator

$$R : (a_1, \dots, a_m, -a_m, \dots, a_{-1}) \rightarrow (a_{-1}, \dots, -a_t, a_t, \dots, a_1).$$

With these ingredients, we can work out the final contraction that takes into account the symmetry sectors

$$\begin{aligned} \langle \gamma, \beta | Z_1 Z_2 | \gamma, \beta \rangle^{2\text{-tree}} &= \sum_{t=1}^p \sum_{\mathbf{a}_t^{(t)} \mathbf{b}_t^{(t)}} G_t(a_t) G_t(b_t) g(\mathbf{a}_t^{(t-1)}) g(\mathbf{b}_t^{(t-1)}) H_{d-1}^{(t)}(\mathbf{a}_t^{(t)}) H_{d-1}^{(t)}(\mathbf{b}_t^{(t)}) \\ &\quad \exp \left(i \mathbf{\Gamma}^{(t)} \cdot \left(\frac{\mathbf{a}_t^{(t)} \mathbf{b}_t^{(t)} + h(\mathbf{a}_t^{(t)} + \mathbf{b}_t^{(t)})}{\sqrt{d}} \right) \right) \\ &\quad + 2 \sum_{t_a > t_b} \sum_{\mathbf{a}_{t_a}^{(t_a)} \mathbf{b}_{t_b}^{(t_b)}} G(a_{t_a}) \tilde{G}(b_{t_b}) g(\mathbf{a}_{t_a}^{(t_a-1)}) g(\mathbf{b}_{t_b}^{(t_b-1)}) H_{d-1}^{(t_a)}(\mathbf{a}_{t_a}^{(t_a)}) H_{d-1}^{(t_b)}(\mathbf{b}_{t_b}^{(t_b)}) \\ &\quad \exp \left(i \mathbf{\Gamma}^{(t_a)} \cdot \left(\frac{\mathbf{a}_{t_a}^{(t_a)} \mathbf{b}_{t_b}^{(t_b)} + h(\mathbf{a}_{t_a}^{(t_a)} + \mathbf{b}_{t_b}^{(t_b)})}{\sqrt{d}} \right) \right). \end{aligned} \quad (19)$$

Here, the function $G(a_t) \equiv G(a_t; \beta_t, \dots, \beta_p)$ result from the elimination of redundant symmetric variables

$$\begin{aligned} G(a_t) &= \sum_{a_{t+1}, \dots, a_p, a_0} \frac{a_0}{2} \langle a_t | e^{i\beta_t X} | a_{t+1} \rangle \langle a_{t+1} | e^{-i\beta_t X} | -a_t \rangle | \langle a_{t+1} | e^{i\beta_{t+1} X} | a_{t+2} \rangle |^2 \dots | \langle a_p | e^{i\beta_p X} | a_0 \rangle |^2 \\ &= -\frac{a_t}{2} i \sin(2\beta_t) \cos(2\beta_{t+1}) \dots \cos(2\beta_p). \end{aligned}$$

Notice that if $t = p$, we have that $G(a_p) = -\frac{a_p}{2} i \sin(2\beta_p)$. Similarly, we define

$$\begin{aligned} \tilde{G}(a_t) &= \sum_{a_{t+1}, \dots, a_p, a_0} \frac{a_0}{2} | \langle a_t | e^{i\beta_t X} | a_{t+1} \rangle |^2 | \langle a_{t+1} | e^{i\beta_{t+1} X} | a_{t+2} \rangle |^2 \dots | \langle a_p | e^{i\beta_p X} | a_0 \rangle |^2 \\ &= \frac{a_t}{2} \cos(2\beta_t) \cos(2\beta_{t+1}) \dots \cos(2\beta_p). \end{aligned}$$

For the onsite term, applying the same procedure yields

$$\langle \gamma, \beta | Z_1 | \gamma, \beta \rangle^{1\text{-tree}} = \sum_{t=1}^p \sum_{\mathbf{a}_t^{(t)}} G_t(a_t) g(\mathbf{a}_t^{(t-1)}) H_{d-1}^{(t)}(\mathbf{a}_t^{(t)}) \exp \left(i \frac{h}{\sqrt{d}} \mathbf{\Gamma}^{(t)} \cdot \mathbf{a}_t^{(t)} \right). \quad (20)$$

A.4 The performance of tree QAOA for MaxCut relative to random sampling

In Sec. 5, in particular in Fig. 5(A), we reported the QAOA approximation ratios for MaxCut as defined in Sec. 3. Here, we adopted the conventional form of the MaxCut cost function (9) and defined MaxCut approximation ratios that take into account upper bounds on optimality. For this reason, we obtained non-trivial approximation ratios for random sampling, or $p = 0$ QAOA, that are given by these upper bounds

$$\alpha_{MC} = \frac{c_0}{c_{ub}} = \frac{1}{2c_{ub}}$$

because random sampling cuts half of the edges and therefore the cut fraction is $c_0 = 1/2$. However, we could also be interested in the performance *relative* to random sampling. For this, we can define the relative approximation ratios obtained by level- p QAOA

$$\alpha_{MC}^{\text{relative}} = (c_p - 0.5)/(c_{ub} - 0.5). \quad (21)$$

Of course, this quantity can also be straightforwardly computed based on our tree calculation. The results are shown in Fig. 9 and it can be seen that, similarly to local classical algorithms [70], these relative approximation ratios approach a constant in the large d limit. For $p = 1$, we explicitly determine the value of this constant in the next section as $\frac{1}{2\sqrt{eP^*}}$ with P^* the Parisi constant A.5.

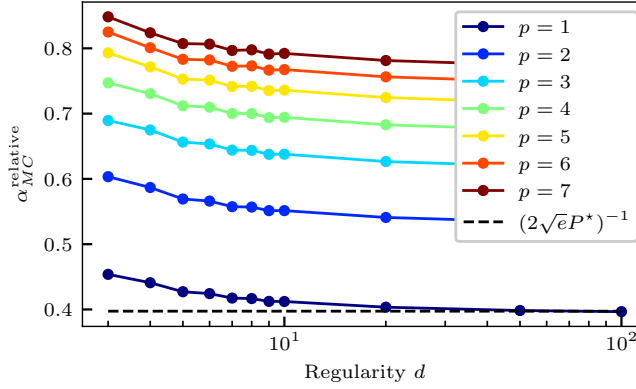


Figure 9: Relative approximation ratios achieved by QAOA that quantify the performance relative to random sampling.

A.5 The performance of tree QAOA for large d

In this section, we investigate the performance of tree QAOA in the large d limit with fixed $p = 1$. From the previous section, we can easily reduce the expectation values to $p = 1$, this gives

$$\begin{aligned} \langle \gamma, \beta | Z_1 Z_2 | \gamma, \beta \rangle^{2\text{-tree}} &= -\frac{1}{2} [\sin(2\beta)]^2 \left[\cos\left(\frac{4h\gamma}{\sqrt{d}}\right) - 1 \right] \left[\cos\left(\frac{2\gamma}{\sqrt{d}}\right) \right]^{2d-2} \\ &\quad + \sin(4\beta) \sin\left(\frac{2\gamma}{\sqrt{d}}\right) \cos\left(\frac{2h\gamma}{\sqrt{d}}\right) \left[\cos\left(\frac{2\gamma}{\sqrt{d}}\right) \right]^{d-1} \end{aligned}$$

where the two terms correspond to the two summations in Eq. (19), more precisely the cases where $t_a = t_b = 1$ and $t_a = 1, t_b = 0$. The local term becomes

$$\langle \gamma, \beta | Z_1 | \gamma, \beta \rangle^{1\text{-tree}} = \sin(2\beta) \sin\left(\frac{2h\gamma}{\sqrt{d}}\right) \left[\cos\left(\frac{2\gamma}{\sqrt{d}}\right) \right]^d,$$

which is just the single case $t_a = 1$ appearing in Eq. (20).

Tree QAOA with $p = 1$ aims to minimize Eq. (4), i.e. the following function

$$\begin{aligned} &\frac{h}{\sqrt{d}} \langle \gamma, \beta | Z_1 | \gamma, \beta \rangle^{1\text{-tree}} + \frac{\sqrt{d}}{2} \langle \gamma, \beta | Z_1 Z_2 | \gamma, \beta \rangle^{2\text{-tree}} \\ &= \frac{h}{\sqrt{d}} \sin(2\beta) \sin\left(\frac{2h\gamma}{\sqrt{d}}\right) \left[\cos\left(\frac{2\gamma}{\sqrt{d}}\right) \right]^d - \frac{\sqrt{d}}{4} [\sin(2\beta)]^2 \left[\cos\left(\frac{4h\gamma}{\sqrt{d}}\right) - 1 \right] \left[\cos\left(\frac{2\gamma}{\sqrt{d}}\right) \right]^{2d-2} \\ &\quad + \frac{\sqrt{d}}{2} \sin(4\beta) \sin\left(\frac{2\gamma}{\sqrt{d}}\right) \cos\left(\frac{2h\gamma}{\sqrt{d}}\right) \left[\cos\left(\frac{2\gamma}{\sqrt{d}}\right) \right]^{d-1}. \end{aligned}$$

In case of MaxCut, we have $h = 0$. Therefore trivially $\langle \gamma, \beta | Z_1 | \gamma, \beta \rangle^{1\text{-tree}} = 0$ and the correlator is given by

$$\langle \gamma, \beta | Z_1 Z_2 | \gamma, \beta \rangle^{2\text{-tree}} = \sin(4\beta) \sin\left(\frac{2\gamma}{\sqrt{d}}\right) \left[\cos\left(\frac{2\gamma}{\sqrt{d}}\right) \right]^{d-1}.$$

This expression is minimized by choosing $\beta^* = -\pi/8$ irrespective of d . The optimal gamma γ^* can be found by maximizing the remaining part which gives

$$\gamma^* = \frac{\sqrt{d}}{2} \cos^{-1} \left(\sqrt{\frac{d-1}{d}} \right), \quad (22)$$

for increasing d this becomes $\gamma^* \rightarrow \frac{\sqrt{d}}{2} \times \frac{1}{\sqrt{d}} = 1/2$. In this limit the expectation value becomes

$$\langle \gamma, \beta^* | Z_1 Z_2 | \gamma, \beta^* \rangle^{2\text{-tree}} \approx \frac{2\gamma}{\sqrt{d}} e^{-2\gamma^2} (1 + O(d^{-1})) \quad (23)$$

by a Taylor expansion. Therefore, the cut fraction obtained by tree QAOA is $1/2$ in the leading order which is the same as random sampling. However, as in Appendix A.4, we could also be interested in the performance relative to random sampling. For this we would like to compute the relative approximation ratio (21) in the large d limit restricted to $p = 1$ QAOA. From the above Eqs. (22) and (23), and by recalling the definition of c_p (10), we have that $c_1 - 0.5 = (2\sqrt{de})^{-1}$ for large d . For the denominator we can use the results of [47] that imply that $(c_{ub} - 0.5) \approx P^*/\sqrt{d}$ with $P^* \approx 0.7632\dots$ the Parisi constant. Therefore it follows that the relative approximation ratio converges to a constant in the large d limit that can be determined as

$$\alpha_{MC}^{\text{relative}}(d \rightarrow \infty, p = 1) = \frac{1}{2\sqrt{e}P^*} \approx 0.397\dots$$

This is consistent with our numerical results shown in the previous section A.4.

In case of MIS, $h \approx d$, the single-body term $\langle \gamma, \beta | Z_1 | \gamma, \beta \rangle^{1\text{-tree}}$ contributes. For increasing d , we have that $\beta^* \rightarrow -\pi/4$ and $\gamma^* \rightarrow \frac{\pi}{4\sqrt{d}}$. This means that in this limit $\langle \gamma^*, \beta^* | Z_1 Z_2 | \gamma^*, \beta^* \rangle^{2\text{-tree}} \rightarrow 1$ and $\langle \gamma^*, \beta^* | Z_1 | \gamma^*, \beta^* \rangle^{1\text{-tree}} \rightarrow -1$. Then, we have a vanishing independence ratio, and thus a vanishing MIS approximation ratio.

B Tree angles for MaxCut and MIS

We include the tree angles that were used in the finite-size simulations shown in Fig. 6 in the tables below. The angles below are obtained for the QAOA ansatz given in Eq. (2), for better table alignment we however show $\beta \rightarrow -\beta$. For MaxCut these angles are found by taking $h = 0$, and $h = 2$ for MIS.

QAOA depth p	Tree angles for MaxCut $d = 3$
1	$\gamma = (0.5330)$ $\beta = (0.3927)$
2	$\gamma = (0.4225, 0.7776)$ $\beta = (0.5549, 0.2924)$
3	$\gamma = (0.3653, 0.6914, 0.8114)$ $\beta = (0.6090, 0.4596, 0.2357)$
4	$\gamma = (0.3540, 0.6760, 0.8557, 1.0019)$ $\beta = (0.5996, 0.4343, 0.2968, 0.1590)$
5	$\gamma = (0.3111, 0.6115, 0.7119, 0.8697, 0.9993)$ $\beta = (0.6317, 0.5225, 0.3901, 0.2760, 0.1493)$
6	$\gamma = (0.2870, 0.5591, 0.6336, 0.7248, 0.8749, 0.9763)$ $\beta = (0.6359, 0.5344, 0.4633, 0.3600, 0.2585, 0.1388)$
7	$\gamma = (0.2682, 0.5342, 0.5960, 0.6496, 0.7426, 0.8845, 0.9750)$ $\beta = (0.6476, 0.5531, 0.4893, 0.4448, 0.3408, 0.2444, 0.1312)$
8	$\gamma = (0.2537, 0.5076, 0.5674, 0.6132, 0.6619, 0.7490, 0.8892, 0.9669)$ $\beta = (0.6492, 0.5555, 0.5013, 0.4690, 0.4202, 0.3195, 0.2310, 0.1229)$

QAOA depth p	Tree angles for MaxCut $d = 4$
1	$\gamma = (0.5236)$ $\beta = (0.3927)$
2	$\gamma = (0.4078, 0.7397)$ $\beta = (0.5341, 0.2830)$
3	$\gamma = (0.3545, 0.6514, 0.7543)$ $\beta = (0.5879, 0.4232, 0.2230)$
4	$\gamma = (0.3150, 0.5876, 0.6732, 0.7712)$ $\beta = (0.6050, 0.4778, 0.3613, 0.1875)$
5	$\gamma = (0.2909, 0.5468, 0.6033, 0.6872, 0.7844)$ $\beta = (0.6225, 0.5051, 0.4167, 0.3253, 0.1628)$
6	$\gamma = (0.2687, 0.5128, 0.5636, 0.6141, 0.6957, 0.7867)$ $\beta = (0.6293, 0.5232, 0.4528, 0.3883, 0.2981, 0.1459)$
7	$\gamma = (0.2537, 0.4890, 0.5317, 0.5757, 0.6214, 0.6976, 0.7885)$ $\beta = (0.6378, 0.5327, 0.4719, 0.4325, 0.3632, 0.2778, 0.1339)$
8	$\gamma = (0.2405, 0.4690, 0.5113, 0.5480, 0.5851, 0.6257, 0.7221, 0.8339)$ $\beta = (0.6405, 0.5385, 0.4817, 0.4526, 0.4101, 0.3452, 0.2605, 0.1198)$

QAOA depth p	Tree angles for MIS $d = 3$
1	$\gamma = (0.4299)$ $\beta = (0.3986)$
2	$\gamma = (0.3678, 0.7957)$ $\beta = (0.5175, 0.2642)$
3	$\gamma = (0.3260, 0.6720, 0.7582)$ $\beta = (0.5777, 0.3680, 0.2103)$
4	$\gamma = (0.2846, 0.6045, 0.7105, 0.7419)$ $\beta = (0.6046, 0.4686, 0.3396, 0.1846)$
5	$\gamma = (0.2722, 0.5640, 0.6536, 0.7082, 0.7908)$ $\beta = (0.6174, 0.4776, 0.4222, 0.3088, 0.1525)$
6	$\gamma = (0.2537, 0.5324, 0.6102, 0.6370, 0.7164, 0.7872)$ $\beta = (0.6350, 0.5082, 0.4523, 0.4136, 0.2866, 0.1533)$
7	$\gamma = (0.2316, 0.4843, 0.5687, 0.5937, 0.6238, 0.7266, 0.7577)$ $\beta = (0.6337, 0.5026, 0.4636, 0.4304, 0.3718, 0.2725, 0.1410)$
8	$\gamma = (0.2249, 0.4734, 0.5590, 0.5833, 0.6082, 0.6697, 0.7368, 0.7882)$ $\beta = (0.6428, 0.5181, 0.4779, 0.4519, 0.3974, 0.3589, 0.2566, 0.1203)$

QAOA depth p	Tree angles for MIS $d = 4$
1	$\gamma = (0.3376)$ $\beta = (0.4240)$
2	$\gamma = (0.3123, 0.8352)$ $\beta = (0.5169, 0.2407)$
3	$\gamma = (0.2756, 0.6606, 0.7660)$ $\beta = (0.5628, 0.3024, 0.1942)$
4	$\gamma = (0.2405, 0.5462, 0.6808, 0.6781)$ $\beta = (0.5745, 0.3541, 0.2618, 0.1563)$
5	$\gamma = (0.2245, 0.5178, 0.6307, 0.6445, 0.6808)$ $\beta = (0.6138, 0.4152, 0.3412, 0.2583, 0.1410)$
6	$\gamma = (0.2043, 0.4545, 0.5887, 0.6121, 0.6307, 0.7054)$ $\beta = (0.6162, 0.4200, 0.3845, 0.3231, 0.2229, 0.1152)$
7	$\gamma = (0.1981, 0.4406, 0.5595, 0.5777, 0.5887, 0.6775, 0.7116)$ $\beta = (0.6350, 0.4475, 0.4108, 0.3828, 0.3036, 0.1984, 0.1188)$
8	$\gamma = (0.1765, 0.3840, 0.5239, 0.5596, 0.5595, 0.6197, 0.6817, 0.6833)$ $\beta = (0.6278, 0.4452, 0.4267, 0.3991, 0.3570, 0.2689, 0.1946, 0.1085)$

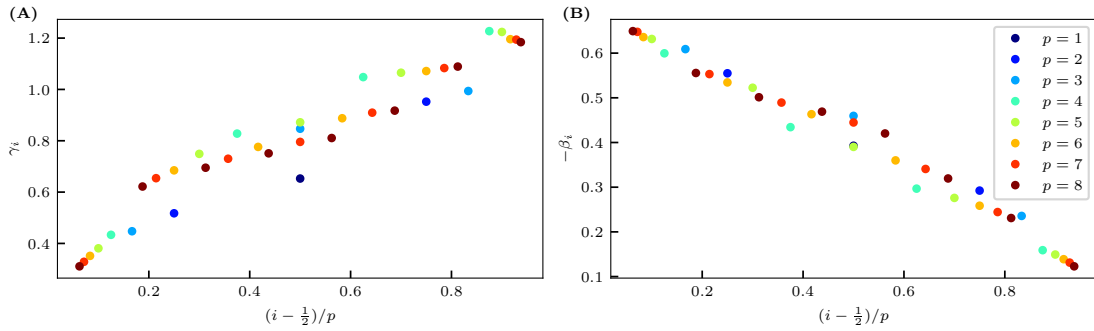


Figure 10: Tree angles for MaxCut $d = 3$.

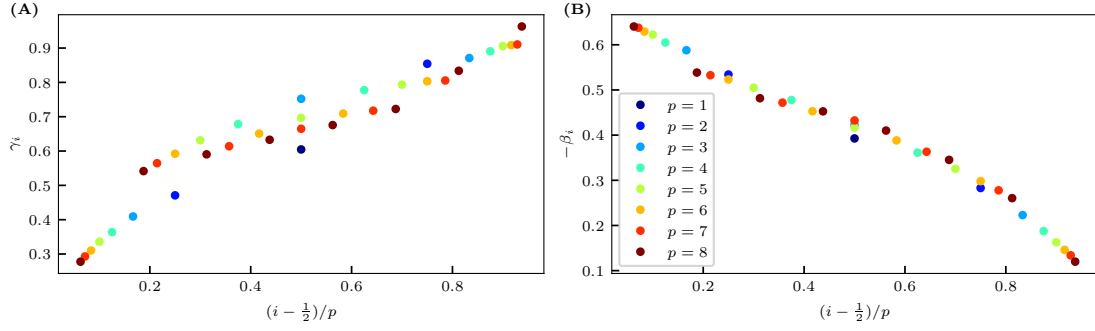


Figure 11: Tree angles for MaxCut $d = 4$.

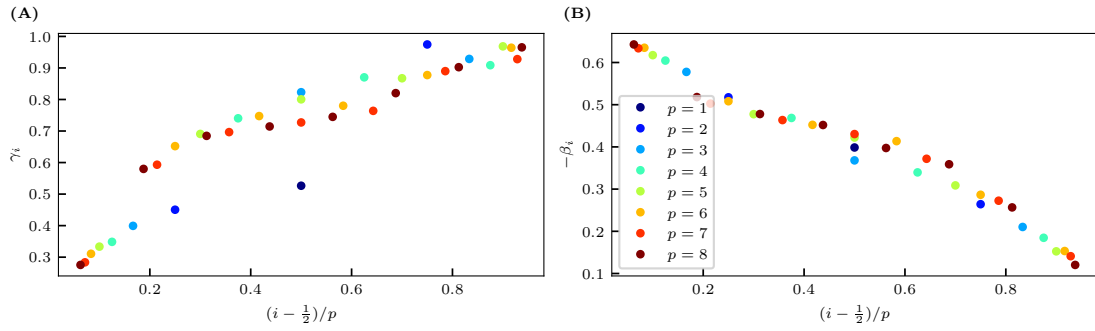


Figure 12: Tree angles for MIS $d = 3$.

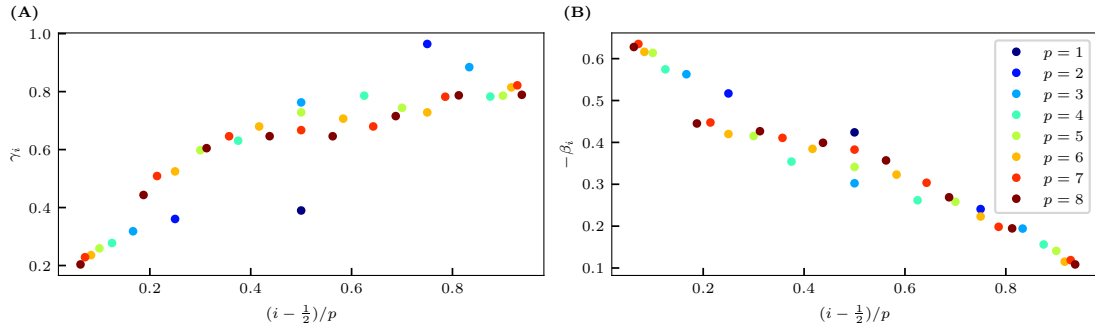


Figure 13: Tree angles for MIS $d = 4$.

References

- [1] Richard P. Feynman. “Simulating physics with computers”. *International Journal of Theoretical Physics* **21**, 467–488 (1982).
- [2] J. Ignacio Cirac and Peter Zoller. “Goals and opportunities in quantum simulation”. *Nature Physics* **8**, 264–266 (2012).
- [3] I. M. Georgescu, S. Ashhab, and Franco Nori. “Quantum simulation”. *Rev. Mod. Phys.* **86**, 153–185 (2014).
- [4] Aram W. Harrow, Avinandan Hassidim, and Seth Lloyd. “Quantum algorithm for linear systems of equations”. *Phys. Rev. Lett.* **103**, 150502 (2009).
- [5] Yunchao Liu, Srinivasan Arunachalam, and Kristan Temme. “A rigorous and robust quantum speed-up in supervised machine learning”. *Nature Physics* **17**, 1013–1017 (2021).
- [6] Hsin-Yuan Huang, Richard Kueng, and John Preskill. “Information-theoretic bounds on quantum advantage in machine learning”. *Phys. Rev. Lett.* **126**, 190505 (2021).
- [7] Ryan Sweke, Jean-Pierre Seifert, Dominik Hangleiter, and Jens Eisert. “On the Quantum versus Classical Learnability of Discrete Distributions”. *Quantum* **5**, 417 (2021).
- [8] Lov K. Grover. “A Fast quantum mechanical algorithm for database search” (1996). [arXiv:quant-ph/9605043](https://arxiv.org/abs/quant-ph/9605043).
- [9] Amira Abbas, Andris Ambainis, Brandon Augustino, Andreas Bäertschi, Harry Buhrman, Carleton Coffrin, Giorgio Cortiana, Vedran Dunjko, Daniel J. Egger, Bruce G. Elmegreen, Nicola Franco, Filippo Fratini, Bryce Fuller, Julien Gacon, Constantin Gocciulea, Sander Gribling, Swati Gupta, Stuart Hadfield, Raoul Heese, Gerhard Kircher, Thomas Kleinert, Thorsten Koch, Georgios Korpas, Steve Lenk, Jakub Marecek, Vanio Markov, Guglielmo Mazzola, Stefano Mensa, Naeimeh Mohseni, Giacomo Nannicini, Corey O’Meara, Elena Peña Tapia, Sebastian Pokutta, Manuel Proissl, Patrick Rebentrost, Emre Sahin, Benjamin C. B. Symons, Sabine Törnøw, Victor Valls, Stefan Woerner, Mira L. Wolf-Bauwens, Jon Yard, Sheir Yarkoni, Dirk Zechiel, Sergiy Zhuk, and Christa Zoufal. “Quantum optimization: Potential, challenges, and the path forward” (2023). [arXiv:2312.02279](https://arxiv.org/abs/2312.02279).
- [10] John Preskill. “Quantum computing in the nisq era and beyond”. *Quantum* **2**, 79 (2018) **2**, 79 (2018). [arXiv:1801.00862](https://arxiv.org/abs/1801.00862).
- [11] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. “A quantum approximate optimization algorithm” (2014). [arXiv:1411.4028](https://arxiv.org/abs/1411.4028).
- [12] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. “A quantum approximate optimization algorithm applied to a bounded occurrence constraint problem” (2014). [arXiv:1412.6062](https://arxiv.org/abs/1412.6062).
- [13] Edward Farhi and Aram W Harrow. “Quantum supremacy through the quantum approximate optimization algorithm” (2016). [arXiv:1602.07674](https://arxiv.org/abs/1602.07674).
- [14] Stuart Hadfield, Zhihui 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* **12.2** (2019): 34. (Special issue "Quantum Optimization Theory, Algorithms, and Applications") **12**, 34 (2017). [arXiv:1709.03489](https://arxiv.org/abs/1709.03489).
- [15] Ehsan Zahedinejad and Arman Zaribafyan. “Combinatorial optimization on gate model quantum computers: A survey” (2017). [arXiv:1708.05294](https://arxiv.org/abs/1708.05294).
- [16] Zhihui Wang, Stuart Hadfield, Zhang Jiang, and Eleanor G. Rieffel. “Quantum approximate optimization algorithm for maxcut: A fermionic view”. *Phys. Rev. A* **97**, 022304 (2018).
- [17] Kostas Blekos, Dean Brand, Andrea Ceschini, Chiao-Hui Chou, Rui-Hao Li, Komal Pandya, and Alessandro Summer. “A review on quantum approximate optimization algorithm and its variants”. *Physics Reports* **1068**, 1–66 (2023). [arXiv:2306.09198](https://arxiv.org/abs/2306.09198).
- [18] Eran Makover and Jeffrey McGowan. “Regular trees in random regular graphs” (2006).
- [19] Joao Basso, Edward Farhi, Kunal Marwaha, Benjamin Villalonga, and Leo Zhou. “The quantum approximate optimization algorithm at high depth for maxcut on large-girth regular graphs and the sherrington-kirkpatrick model”. In *Proceedings of the 17th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC ’22)*, 7:1–7:21, (2022) (2021). [arXiv:2110.14206](https://arxiv.org/abs/2110.14206).
- [20] Fernando G. S. L. Brandao, Michael Broughton, Edward Farhi, Sam Gutmann, and Hartmut Neven. “For fixed control parameters the quantum approximate optimization algorithm’s objective function value concentrates for typical instances” (2018). [arXiv:1812.04170](https://arxiv.org/abs/1812.04170).

- [21] Michael Streif and Martin Leib. “Training the quantum approximate optimization algorithm without access to a quantum processing unit”. *Quantum Science and Technology* **5**, 034008 (2020).
- [22] Jonathan Wurtz and Peter J. Love. “Maxcut qaoa performance guarantees for $p > 1$ ”. *Phys. Rev. A* **103**, 042612 (2021) **103**, 042612 (2020). [arXiv:2010.11209](#).
- [23] Alexey Galda, Xiaoyuan Liu, Danylo Lykov, Yuri Alexeev, and Ilya Safro. “Transferability of optimal qaoa parameters between random graphs” (2021). [arXiv:2106.07531](#).
- [24] Alexey Galda, Eesh Gupta, Jose Falla, Xiaoyuan Liu, Danylo Lykov, Yuri Alexeev, and Ilya Safro. “Similarity-based parameter transferability in the quantum approximate optimization algorithm” (2023). [arXiv:2307.05420](#).
- [25] Gavin E. Crooks. “Performance of the quantum approximate optimization algorithm on the maximum cut problem” (2018). [arXiv:1811.08419](#).
- [26] Jonathan Wurtz and Danylo Lykov. “Fixed-angle conjectures for the quantum approximate optimization algorithm on regular maxcut graphs”. *Phys. Rev. A* **104**, 052419 (2021).
- [27] Michel X. Goemans and David P. Williamson. “Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming”. *Journal of the ACM* **42**, 1115–1145 (1995).
- [28] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O’Donnell. “Optimal inapproximability results for max-cut and other 2-variable csps?”. *SIAM Journal on Computing* **37**, 319–357 (2007). [arXiv:https://doi.org/10.1137/S0097539705447372](#).
- [29] M. B. Hastings. “Classical and quantum bounded depth approximation algorithms” (2019). [arXiv:1905.07047](#).
- [30] G. G. Guerreschi and A. Y. Matsuura. “QAOA for max-cut requires hundreds of qubits for quantum speed-up”. *Scientific Reports* **9** (2019).
- [31] Danylo Lykov, Jonathan Wurtz, Cody Poole, Mark Saffman, Tom Noel, and Yuri Alexeev. “Sampling frequency thresholds for quantum advantage of quantum approximate optimization algorithm” (2022). [arXiv:2206.03579](#).
- [32] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. “Local algorithms for maximum cut and minimum bisection on locally treelike regular graphs of large degree” (2021). [arXiv:2111.06813](#).
- [33] Boaz Barak and Kunal Marwaha. “Classical algorithms and quantum limitations for maximum cut on high-girth graphs”. *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*; Article No. 14 (2021). [arXiv:2106.05900](#).
- [34] Kunal Marwaha. “Local classical max-cut algorithm outperforms $p = 2$ qaoa on high-girth regular graphs”. *Quantum* **5**, 437 (2021) **5**, 437 (2021). [arXiv:2101.05513](#).
- [35] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Leo Zhou. “The quantum approximate optimization algorithm and the sherrington-kirkpatrick model at infinite size”. *Quantum* **6**, 759 (2022).
- [36] Sami Boulebnane and Ashley Montanaro. “Predicting parameters for the quantum approximate optimization algorithm for max-cut from the infinite-size limit” (2021). [arXiv:2110.10685](#).
- [37] Edward Farhi, David Gamarnik, and Sam Gutmann. “The quantum approximate optimization algorithm needs to see the whole graph: A typical case” (2020). [arXiv:2004.09002](#).
- [38] Edward Farhi, David Gamarnik, and Sam Gutmann. “The quantum approximate optimization algorithm needs to see the whole graph: Worst case examples” (2020). [arXiv:2005.08747](#).
- [39] M. Mézard, T. Mora, and R. Zecchina. “Clustering of solutions in the random satisfiability problem”. *Phys. Rev. Lett.* **94**, 197205 (2005).
- [40] Dimitris Achlioptas, Amin Coja-Oghlan, and Federico Ricci-Tersenghi. “On the solution-space geometry of random constraint satisfaction problems”. *Random Structures & Algorithms* **38**, 251–268 (2011). [arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.20323](#).
- [41] David Gamarnik and Aukosh Jagannath. “The overlap gap property and approximate message passing algorithms for p -spin models” (2019). [arXiv:1911.06943](#).
- [42] David Gamarnik. “The overlap gap property: a geometric barrier to optimizing over random structures”. *Proceedings of the National Academy of Sciences* **118** (2021). [arXiv:2109.14409](#).
- [43] Antares Chen, Neng Huang, and Kunal Marwaha. “Local algorithms and the failure of log-depth quantum advantage on sparse random csps” (2023). [arXiv:2310.01563](#).

- [44] Wei-Kuo Chen, David Gamarnik, Dmitry Panchenko, and Mustazee Rahman. “Suboptimality of local algorithms for a class of max-cut problems”. *Annals of Probability* 2019, Vol. 47, No. 3, 1587-161847 (2017). [arXiv:1707.05386](#).
- [45] Antonio Auffinger, Wei-Kuo Chen, and Qiang Zeng. “The sk model is infinite step replica symmetry breaking at zero temperature”. *CPAM 2020* (2017). [arXiv:1703.06872](#).
- [46] Chi-Ning Chou, Peter J. Love, Juspreet Singh Sandhu, and Jonathan Shi. “Limitations of local quantum algorithms on random max-k-xor and beyond” (2021). [arXiv:2108.06049](#).
- [47] Amir Dembo, Andrea Montanari, and Subhabrata Sen. “Extremal cuts of sparse random graphs”. *Annals of Probability*, 2017, Vol 45, No. 2, 1190- 121745 (2015). [arXiv:1503.03923](#).
- [48] Amin Coja-Oghlan, Philipp Loick, Balázs F. Mezei, and Gregory B. Sorkin. “The ising anti-ferromagnet and max cut on random regular graphs” (2020). [arXiv:2009.10483](#).
- [49] B. D. McKay. “Independent sets in regular graphs of high girth”. *Ars Combinatoria*, 23A 179-185. (1987). url: <https://users.cecs.anu.edu.au/~bdm/papers/IndSetsRegGraphs.pdf>.
- [50] József Balogh, Alexandr Kostochka, and Xujun Liu. “Cubic graphs with small independence ratio” (2017). [arXiv:1708.03996](#).
- [51] David Gamarnik and Madhu Sudan. “Limits of local algorithms over sparse random graphs” (2013). [arXiv:1304.1831](#).
- [52] Sergey Bravyi, Alexander Kliesch, Robert Koenig, and Eugene Tang. “Obstacles to state preparation and variational optimization from symmetry protection”. *Phys. Rev. Lett.* 125, 260505 (2020) 125, 260505 (2019). [arXiv:1910.08980](#).
- [53] Eunok Bae and Soojoon Lee. “Recursive qaoa outperforms the original qaoa for the max-cut problem on complete graphs”. *Quantum Information Processing* 23 (2022). [arXiv:2211.15832](#).
- [54] Yash J. Patel, Sofiene Jerbi, Thomas Bäck, and Vedran Dunjko. “Reinforcement learning assisted recursive qaoa”. *EPJ Quantum Technol.* 11, 6 (2024)11 (2022). [arXiv:2207.06294](#).
- [55] Friedrich Wagner, Jonas Nüßlein, and Frauke Liers. “Enhancing quantum algorithms for quadratic unconstrained binary optimization via integer programming” (2023). [arXiv:2302.05493](#).
- [56] Maxime Dupont, Bram Evert, Mark J. Hodson, Bhuvanesh Sundar, Stephen Jeffrey, Yuki Yamaguchi, Dennis Feng, Filip B. Maciejewski, Stuart Hadfield, M. Sohaib Alam, Zhihui Wang, Shon Grabbe, P. Aaron Lott, Eleanor G. Rieffel, Davide Venturelli, and Matthew J. Reagor. “Quantum-enhanced greedy combinatorial optimization solver”. *Science Advances* 9, 45 (2023)9 (2023). [arXiv:2303.05509](#).
- [57] Maxime Dupont and Bhuvanesh Sundar. “Quantum relax-and-round algorithm for combinatorial optimization” (2023). [arXiv:2307.05821](#).
- [58] Jernej Rudi Finžgar, Aron Kerschbaumer, Martin J. A. Schuetz, Christian B. Mendl, and Helmut G. Katzgraber. “Quantum-informed recursive optimization algorithms” (2023). [arXiv:2308.13607](#).
- [59] Zhi-Cheng Yang, Armin Rahmani, Alireza Shabani, Hartmut Neven, and Claudio Chamon. “Optimizing Variational Quantum Algorithms Using Pontryagin’s Minimum Principle”. *Physical Review X* 7, 021027 (2017).
- [60] Lenka Zdeborová and Stefan Boettcher. “Conjecture on the maximum cut and bisection width in random regular graphs”. *J. Stat. Mech.* (2010) P02020 2010, P02020 (2009). [arXiv:0912.4861](#).
- [61] Raffaele Marino and Scott Kirkpatrick. “Large independent sets on random d -regular graphs with fixed degree d ” (2020). [arXiv:2003.12293](#).
- [62] Mustazee Rahman and Balint Virag. “Local algorithms for independent sets are half-optimal”. *Ann. Probab.* 45 (2017), no. 3, 1543-157745 (2014). [arXiv:1402.0485](#).
- [63] Amin Coja-Oghlan and Charilaos Efthymiou. “On independent sets in random graphs”. *Random Structures and Algorithms* 47 (2015) 436 - 486 47, 436-486 (2010). [arXiv:1007.1378](#).
- [64] Alexander S. Wein. “Optimal low-degree hardness of maximum independent set” (2020). [arXiv:2010.06563](#).
- [65] A.M Frieze and T Łuczak. “On the independence and chromatic numbers of random regular graphs”. *Journal of Combinatorial Theory, Series B* 54, 123-132 (1992).
- [66] Hamed Hatami, László Lovász, and Balázs Szegedy. “Limits of local-global convergent graph sequences” (2012). [arXiv:1205.4356](#).

- [67] Andrea Montanari. “Optimization of the sherrington-kirkpatrick hamiltonian” (2018). [arXiv:1812.10897](#).
- [68] Mark Xin Hong Goh. “The overlap gap property limits limit swapping in qaoa” (2024). [arXiv:2404.06087](#).
- [69] Simon Martiel, Thomas Ayrat, and Cyril Allouche. “Benchmarking quantum coprocessors in an application-centric, hardware-agnostic, and scalable way”. *IEEE Transactions on Quantum Engineering* **2**, 1–11 (2021).
- [70] Juho Hirvonen, Joel Rybicki, Stefan Schmid, and Jukka Suomela. “Large cuts with local algorithms on triangle-free graphs” (2014) [arXiv:1402.2543](#).
- [71] Eran Halperin, Dror Livnat, and Uri Zwick. “Max cut in cubic graphs”. *Journal of Algorithms* **53**, 169–185 (2004).
- [72] M. M. Hallórssen and J. Radhakrishnan. “Greed is good: Approximating independent sets in sparse and bounded-degree graphs”. *Algorithmica* **18**, 145–163 (1997).
- [73] Reuben Tate, Jai Moondra, Bryan Gard, Greg Mohler, and Swati Gupta. “Warm-Started QAOA with Custom Mixers Provably Converges and Computationally Beats Goemans-Williamson’s Max-Cut at Low Circuit Depths”. *Quantum* **7**, 1121 (2023).
- [74] Gian Giacomo Guerreschi and Mikhail Smelyanskiy. “Practical optimization for hybrid quantum-classical algorithms” (2017). [arXiv:1701.01450](#).
- [75] 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”. *Phys. Rev. X* **10**, 021067 (2020) **10**, 021067 (2018). [arXiv:1812.01041](#).
- [76] Lennart Bittel and Martin Kliesch. “Training variational quantum algorithms is np-hard”. *Phys. Rev. Lett.* **127**, 120502 (2021) **127**, 120502 (2021). [arXiv:2101.07267](#).