



Skolkovo Institute of Science and Technology

Skolkovo Institute of Science and Technology

# ON THE PERFORMANCE OF QUANTUM APPROXIMATE OPTIMIZATION

*Doctoral Thesis*

by

VISHWANATHAN AKSHAY

DOCTORAL PROGRAM IN COMPUTATIONAL AND DATA  
SCIENCE AND ENGINEERING

Supervisor

Professor, Jacob Daniel Biamonte,  
Doctor of Philosophy, Doctor of Physical and Mathematical Sciences

Moscow – 2023

© Vishwanathan Akshay 2023

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgement is made, and has not been submitted for any other degree.

Candidate: Vishwanathan Akshay  
Supervisor: Prof. Jacob Daniel Biamonte

## Abstract

Contemporary approaches in quantum algorithms development center around the so-called variational model for quantum computation, wherein a classical co-processor in tandem with a noisy intermediate-scale quantum (or NISQ) device, minimizes an objective function to accomplish some computational task. Specifically, one prepares a quantum state by applying a sequence of parameterized quantum gates onto some initial, easy to prepare starting state. This parameterized state, called ansatz, is then tuned iteratively via classical outer-loop optimization routines to minimize the expected value of some classical cost function. The cost function ensures that the relevant computation is realized when the iterative procedure attains the minimum. In this thesis, we study Quantum Approximate Optimization Algorithm (QAOA), a variational approach towards combinatorial optimization. By considering a noiseless setting, we investigate fundamental effects which aid or limit the algorithm's use case. Firstly, we study the performance of QAOA applied to random instances of constraint satisfiability. Here we discover an underparameterization effect, which we call reachability deficits. Specifically, we observe the quality of approximations returned by QAOA to inversely correlate with problem instances constrains-to-variable ratio (called problem density). This effect limits shallow depth circuits to only address low density sub-class of instances and forces the need for employing deeper circuits. Secondly, we consider training QAOA circuits and show that the optimal circuit parameters feature a concentration effect, called parameter concentrations. We analytically demonstrate this effect for variational state preparation and recover an inverse polynomial dependence on the problem size. Our results imply that an initial guess for circuit parameters could be obtained by considering optimal parameters found at small problem sizes, thus allowing one to leverage this effect to reduce the classical cost associated with circuit training. Finally, we attempt to address the open problem of estimating the circuit depth needed for QAOA to succeed. We conjecture a logistic saturation behaviour for the circuit depth as a function of problem density to recover  $\epsilon$ -tolerant approximations. We test our prediction against simulated data and find the model to be capable of describing the data within a  $3\sigma$  confidence. Based on this empirical approach, we also recover a linear trend for circuit-depth scaling with respect to problem sizes, a result that presents optimistic prospects for quantum advantage within the framework of QAOA.

# Publications

The thesis is based on the following publications:

1. V. Akshay, H. Philathong, M.E.S. Morales, and J. Biamonte. Reachability deficits in quantum approximate optimization. *Physical Review Letters*, 124(9):090504, 2020
2. V. Akshay, H. Philathong, I. Zacharov, and J. Biamonte. Reachability deficits in quantum approximate optimization of graph problems. *Quantum*, 5:532, 2021
3. V. Akshay, D. Rabinovich, E. Campos, and J. Biamonte. Parameter concentrations in quantum approximate optimization. *Physical Review A*, 104(1):L010401, 2021
4. V. Akshay, H. Philathong, E. Campos, D. Rabinovich, I. Zacharov, X. Zhang, and J. Biamonte. Circuit depth scaling for quantum approximate optimization. *Physical Review A*, 106:042438, 2022

# Contents

<b>List of symbols and abbreviations</b>	<b>8</b>
<b>List of figures</b>	<b>9</b>
<b>Introduction</b>	<b>12</b>
<b>1 Concepts in quantum computation</b>	<b>17</b>
1.1 Definitions and notations . . . . .	17
1.2 Quantum model of computation . . . . .	20
<b>2 Ising formulation of classical optimization problems</b>	<b>22</b>
2.1 Background . . . . .	22
2.2 Decision and optimization . . . . .	23
2.3 Vector space embedding . . . . .	24
2.4 Combinatorial optimization problems . . . . .	24
2.4.1 Boolean satisfiability . . . . .	25
2.4.2 Maximum cut . . . . .	27
2.4.3 Graph optimization . . . . .	28
2.5 Computational phase transitions . . . . .	29
2.5.1 Review of some computational phase transitions . . . . .	29
2.6 Ising model . . . . .	30
2.7 Physics inspired algorithms for minimizing Ising Hamiltonians . . . . .	31
2.7.1 Simulated annealing . . . . .	31
2.7.2 Adiabatic quantum computation . . . . .	32
2.8 Discussion . . . . .	33

<b>3</b>	<b>Quantum approximate optimization</b>	<b>35</b>
3.1	Background . . . . .	35
3.2	Variational approach to Hamiltonian minimization . . . . .	37
3.3	Quantum approximate optimization algorithm . . . . .	38
3.4	Performance metrics . . . . .	40
3.5	Recovering adiabatic evolution . . . . .	43
3.6	Variational Grover search . . . . .	45
3.7	Discussion . . . . .	49
 <b>4</b>	 <b>Underparameterization effects in QAOA</b>	 <b>50</b>
4.1	Background . . . . .	50
4.2	Reachability deficits . . . . .	51
4.3	Performance of QAOA applied to MAX-SAT . . . . .	52
4.3.1	Numerical details . . . . .	53
4.3.2	Empirical observations . . . . .	54
4.4	Reachability deficits in triangle free regular graphs . . . . .	57
4.5	Reachability deficits in graph optimization . . . . .	63
4.5.1	Numerical details . . . . .	64
4.5.2	Empirical observations . . . . .	66
4.6	Discussion . . . . .	71
 <b>5</b>	 <b>Concentration effects in QAOA</b>	 <b>72</b>
5.1	Background . . . . .	72
5.2	Concentration effects in QAOA . . . . .	73
5.3	Instance concentration in SK model . . . . .	74
5.3.1	Explicit form for moment generating function . . . . .	75
5.3.2	Evaluation of moments . . . . .	78
5.4	Parameter concentrations in variational state preparation . . . . .	80
5.4.1	Parameter concentration at $p=1$ . . . . .	80
5.4.2	Parameter concentration for $p=2$ . . . . .	82
5.4.3	Numerical results for $p \geq 3$ . . . . .	83
5.5	Discussion . . . . .	85

<b>6 Circuit depth scaling in QAOA</b>	<b>87</b>
6.1 Background . . . . .	87
6.2 Logistic saturation . . . . .	88
6.3 Numerical study on the predictive model . . . . .	89
6.3.1 Empirical observations . . . . .	91
6.4 Discussion . . . . .	91
<b>Conclusion</b>	<b>93</b>
<b>Bibliography</b>	<b>98</b>
<b>A QAOA software implementation</b>	<b>107</b>
A.1 Ansatz implementation . . . . .	107
A.2 Layerwise heuristic optimization strategy . . . . .	108
A.3 Software benchmarks . . . . .	109

# List of symbols and abbreviations

$\mathbb{N}, \mathbb{N}_0$	Field of natural numbers and natural numbers inclusive of zero
$\mathbb{C}$	Field of complex numbers
$\mathbb{R}$	Field of real numbers
$\mathcal{H}$	Finite dimensional Hilbert space of $n$ qubits
$\dim(\mathcal{H})$	Dimension of the Hilbert space
$A^\dagger$	Hermitian conjugate of a matrix
$\mathbb{1}$	Identity matrix
$ \psi\rangle$	A unit norm pure quantum state
$\text{herm}_{\mathbb{C}}[d]$	The space of Hermitian operators of dimension $d$
$X_j, Y_j, Z_j$	Pauli matrices acting on the $j^{\text{th}}$ qubit
$\mathcal{O}(g(x))$	Asymptotic upper-bound
$p$	Circuit depth of QAOA
$\beta, \gamma$	Vectors of variational parameters of a QAOA circuit
$\alpha$	Problem density–ratio of a problem instances constrains to variables
P	The class of problems that can be solved in polynomial time by a deterministic Turing machine
NP	The class of problems that can be verified in polynomial time by a deterministic Turing machine given a candidate solution
CNF	Conjunctive Normal Form
VQA	Variational Quantum Algorithms
QAOA	Quantum Approximate Optimization Algorithm
NISQ	Noisy Intermediate Scale Quantum

# List of figures

4-1	$f(\alpha)$ as defined in Def. 3.10 vs clause density for 2-SAT for differing QAOA depths. Data points show the average value obtained over 300 randomly generated instances for $n = 10$ with error bars indicating the standard deviation. Plots also show improved performance for higher depths. Figure reprocessed from [1]. . . . .	56
4-2	$f(\alpha)$ as defined in Def. 3.10 vs clause density for 3-SAT for differing QAOA depths. Data points show the average value obtained over 300 randomly generated instances for $n = 10$ with error bars indicating the standard deviation. Plots also show improved performance for higher depths. Figure reprocessed from [1]. . . . .	56
4-3	Average approximation ratios versus qubit count $n$ , for the three families of graphs: (i) planar grid graphs (blue), (ii) 3-regular graphs (orange), and (iii) SK model or complete graphs (green). Each data point represents the average performance of depth $p = 3$ QAOA over statistics of 300 randomly generated instances with error bars indicating the standard deviation. Figure reprocessed from [2]. . . . .	67
4-4	Success probability versus qubit count $n$ for the three families of graphs: (i) planar grid graphs (blue), (ii) 3-regular graphs (orange), and (iii) SK model or complete graphs (green). Each data point represents the average performance of depth $p = 3$ QAOA over statistics of 300 randomly generated instances with error bars indicating the standard deviation. Figure reprocessed from [2]. . . . .	67

4-5	Error in approximations versus graph density for depth $p = 3, 6, 9$ QAOA. Each data point represents the average performance over statistics of 300 uniform random graph instances, $G_{n,m}$ with nodes $n = 10$ and error bars indicating the standard deviation. Figure reprocessed from [2]. . . . .	68
4-6	Success probability versus graph density for depth $p = 3, 6, 9$ QAOA. Each data point represents the average performance over statistics of 300 uniform random graph instances, $G_{n,m}$ with nodes $n = 10$ and error bars indicating the standard deviation. Figure reprocessed from [2]. . . . .	68
4-7	Comparing (top) error in approximation and (bottom) success probability of the three graph families. For even numbers of qubits, the left vertical pair represents the planar grid to random graph comparison, and the right vertical pair represents the 3-regular to random graph comparison. For odd numbers of qubits, the grid to random graph comparison is shown (since 3-regular graphs can be generated only with an even number of nodes). Each data point represents the average performance of depth $p = 3$ QAOA over statistics of 300 randomly generated instances. By comparison, we do not observe any topology related biases in QAOA performance that can be of statistical relevance. Figure reprocessed from [2]. . . . .	69
4-8	Performance landscape for depth $p = 3$ QAOA on uniform random graph instances $G_{n,m}$ , across varying graph densities, and for graphs with 10 up to 20 nodes. Contours represent polynomial fits of average error in approximations over statistics of 300 instances generated for each density and node count. We observe a rapid fall-off region in approximation quality experienced beyond intermediate-density and low instance sizes. Figure reprocessed from [2]. . . . .	70

---

5-1	Comparison of numerical results (dots) and our fitting for optimal parameters (lines) $\beta_5$ (bottom;orange), $\gamma_5$ (top;blue) with respect to the number of qubits $n$ . The dots represent the parameters obtained numerically via maximization of the overlap function. Figure reprocessed from [3]. . . . .	84
5-2	Parameter concentrations visualized for $p = 5$ depth QAOA and varying number of qubits $n$ . Two symmetric branches of optimal parameters are well distinguished. Parameter concentration is seen within each of the branches. Figure reprocessed from [3]. . . . .	85
6-1	Average critical depth $p^*$ versus clause density. Data points represent the average $p^*$ calculated according to Eq. (6.2) across problem density for MAX-2-SAT with tolerance $\epsilon = 0.3$ . Colors correspond to different problem sizes $n = \{5, 10, 15\}$ with errorbars representing $3\sigma$ for the estimated mean. Solid curves represent the least squares fit of the data to the predictive model as described in Eq. (6.3). . . . .	92
6-2	Fitting parameters versus number of qubits for QAOA on MAX-2-SAT with tolerance $\epsilon = 0.3$ . Note that the recovered critical density $\alpha_c \approx 1$ , and growth rate $\kappa$ , show very little variability with respect to number of qubits. In contrast, the the saturation value $p_{max}$ illustrates a linear trend in the considered range of 5 to 15 qubits. . . . .	92
A-1	Run-time (in seconds) for generating the full state-vector as a function of QAOA depth, $p$ and problem size, $n$ . Dots represent average times computed for a statistic of 100 cases at each depth and problem size.	110

# Introduction

## General topical characterization

As information is necessarily represented in physical media, the laws of physics govern how information is processed. On the microscopic scale, quantum mechanics ultimately dictate these laws, which differ considerably from classical physics describing macroscopic systems. Such quantum systems are known to exhibit non-intuitive behaviors such as superposition and entanglement, which lack a local deterministic description [5, 6]. Therefore, one might wonder how a model of computation based on quantum mechanics, might behave. The fundamental building block of a quantum computer are its information processing units, called qubits, and the relevant operations that can be performed, called quantum gates. Mathematically, a qubit is represented as a vector in a two-dimensional complex vector space, and a quantum system described by  $n$  such individual qubits collectively represents a state space of dimension  $2^n$ . Due to this curse of dimensionality, naively simulating quantum systems on a classical computer becomes intractable [7, 8]. In contrast, a quantum computer would circumvent this problem by naturally emulating quantum systems. Such ideas were further formulated and developed in the work of David Deutsch [9]—what we now call the Church–Turing–Deutsch principle, which asserts that a universal (quantum) computing device can simulate any physical process. Therefore, one would expect a quantum model of computation to be inherently more powerful [10]. Although not formally proven, ample empirical evidence to support this claim. Moreover, beyond simulating quantum physics, many computational problems exist that are believed to be beyond the scope of any classical computer [11]. These problems arise in the form of e.g., combinatorics, groups or properties of graphs,

etc. Early works of: Deutsch-Jozsa [12], Simon [13], Shor [14], and Grover [15] provably demonstrate prospects for quantum advantage over best known classical algorithms by elegantly designing quantum algorithms. However, realizing such algorithms in this era of Noisy-Intermediate-Scale-Quantum (or NISQ) devices [16] remains challenging due to the need for error correction [17, 18] and the limited number of quantum operations that can be performed.

Motivated by these practical challenges, modern quantum programming revolves around a new model for quantum computation called the variational model [19, 20]. Such models, partly inspired by machine learning [21–24], exploit a hybrid quantum-classical approach. Unlike traditional algorithms, such as in Shor’s [14], variational quantum algorithms (or VQAs) consist of an iterative procedure where parameterized quantum circuits (sequences of quantum gates) are trained by an outer loop classical optimization step in order to minimize an objective function. A typical example of a VQA is the Quantum Approximate Optimization Algorithm (or QAOA), designed to approximate solutions to combinatorial optimization problems [25–35]. Although VQAs have recently attracted wide-spread adoption due to resilience over systematic limitations of near-term devices—variability in pulse timing and limited coherence times [36–39], the prospects of employing such algorithms, even in the noiseless setting, remain largely open. In this thesis, we explore this notion restricted to QAOA.

## Thesis goals

The main goal of this thesis is to study, in the noiseless setting, the algorithmic performance of QAOA and to investigate fundamental effects which aid or limit the algorithm’s use case for minimizing classical Ising Hamiltonians. To realize this objective, we:

1. Study the algorithmic performance of QAOA on minimizing Ising Hamiltonians.
2. Interrelate and study different performance metrics that are widely used in the literature.

3. Study the limiting performance of fixed depth QAOA.
4. Study a folklore concentration effect for QAOA circuit parameters.
5. Study a predictive model for estimating circuit depth needed for QAOA to succeed.

## Statements defended

The thesis defends the following main claims:

1. The ratio of problem instances constraints to variables (called *problem density*) induces underparameterization in fixed depth QAOA. This effect is termed as reachability deficits. [1].
  - QAOA on random MAX-3-SAT and random MAX-2-SAT instances exhibits reachability deficits [1].
  - Onset of deficits for QAOA on random MAX-3-SAT and random MAX-2-SAT instances numerically coincide with the computational phase transition point at problem density 1 [1].
  - QAOA on graph minimization problems – corresponding to minimizing 2-local Ising model with uniform random couplings drawn from  $\{-1, +1\}$ , exhibit reachability deficits [2].
2. The optimal parameters for the fixed depth QAOA admit a concentration effect [3]. Specifically, concentrations imply that, given a set of optimal parameters  $\beta_n^*$  and  $\gamma_n^*$ , for  $n$  qubits, a set of optimal parameters for  $n + 1$  qubits,  $\beta_{n+1}^*$  and  $\gamma_{n+1}^*$  can be found such that:

$$|\beta_{n+1}^* - \beta_n^*|^2 + |\gamma_{n+1}^* - \gamma_n^*|^2 \sim \text{poly}^{-1}(n).$$

- Parameters concentrate for  $p = 1, 2$  QAOA on variational state preparation with concentration scaling [3]:

$$|\beta_{n+1}^* - \beta_n^*|^2 + |\gamma_{n+1}^* - \gamma_n^*|^2 \sim n^{-4},$$

- Parameters concentrate for  $p \geq 3$  depth QAOA on variational state preparation for up to  $p = 5$  and 17 qubits with observed scaling being the same as for  $p = 1, 2$  [3].
3. The circuit depth required for QAOA to recover  $\epsilon$  tolerant performance on random MAX-2-SAT instances on up to 15 qubits can empirically be described by a logistic function on problem density [4],

$$p^*(\alpha) \approx \frac{p_{max}}{1 + e^{-\kappa(\alpha - \alpha_c)}}.$$

## Scientific novelty

1. We discovered an effect, now called reachability deficits, which places a fundamental limitation on the performance of fixed depth QAOA, while prior art such as in [26, 40] focused on showing performance advantage over classical algorithms by considering specific instances which allow for analytical performance guarantees.
2. We identify low depth performance benefits of QAOA to only address instances with low problem densities and observe such benefits to wane for densities beyond the computational phase transition point in combinatorial optimization problems [41]. Such a categorization was previously unknown.
3. We develop a new type of concentration effect for optimal parameters and prove its existence for the problem of variational state preparation.
4. We develop an empirical model based on the limiting performance of QAOA which predicts the required circuit depth needed for QAOA to recover near optimal solutions to problem instances of MAX-2-SAT.

## Presentation and validation of the results

The main results compiled in this thesis are based on published articles in peer reviewed journals. Some of the results have also been presented as posters at the

following conferences:

1. International Conference on Quantum Technologies (July 12-16, 2021, Moscow)
2. International Conference on Quantum Technologies (July 15-19, 2019, Moscow)

Validity of the results are supported by numerical experiments and/or mathematical proofs wherever applicable.

# Chapter 1

## Concepts in quantum computation

This chapter reviews the basic concepts in quantum computation and introduces the standard notation used throughout the thesis.

### 1.1 Definitions and notations

The fundamental building block in quantum computation is the notion of qubits, objects which represent the state or configuration of a quantum mechanical system. Mathematically, a qubit is represented as a two-dimensional vector in the complex euclidean space,  $\mathbb{C}^2$ . A standard choice of basis, called the computational basis, in  $\mathbb{C}^2$  are the orthonormal vectors,  $\{|0\rangle, |1\rangle\}$ , following convention of the Dirac notation.

**Remark 1.1.** *An arbitrary state  $|\psi\rangle \in \mathbb{C}^2$  admits a unique decomposition in the computational basis as follows:*

$$|\psi\rangle = a|0\rangle + b|1\rangle, \tag{1.1}$$

with  $a, b \in \mathbb{C}$ .

Similarly to the single-qubit case, the states of an  $n$  qubit system are represented by vectors in  $[\mathbb{C}^2]^{\otimes n}$ . The computational basis vectors are then simply obtained via the tensor product of individual basis vectors,  $\{|0\rangle, |1\rangle\}^{\otimes n}$ . The space of  $n$  qubits forms a finite-dimensional Hilbert space  $\mathcal{H}$ , when equipped with an inner-product.

**Definition 1.1** (Inner-Product). *The standard inner product is defined as:*

$$\begin{aligned} \langle \cdot | \cdot \rangle : [\mathbb{C}^2]^{\otimes n} \times [\mathbb{C}^2]^{\otimes n} &\rightarrow \mathbb{C}, \\ \langle \phi | \psi \rangle &= \sum_{k=0}^{2^n-1} \phi_k^* \psi_k. \end{aligned} \tag{1.2}$$

Here,  $\phi_k, \psi_k \in \mathbb{C}$  are the coefficients in some basis decomposition of  $|\phi\rangle$  and  $|\psi\rangle$ .

**Remark 1.2.** *The 2-norm of a vector  $|\psi\rangle \in \mathcal{H}$  denoted by  $\|\psi\|_2$  satisfies:*

$$\|\psi\|_2 = \sqrt{\langle \psi | \psi \rangle}. \tag{1.3}$$

Operations on qubits are prescribed by linear maps from a Hilbert space  $\mathcal{H}$ , onto itself, denoted by  $\mathcal{L}(\mathcal{H})$ .

**Remark 1.3** (Matrix representation of linear maps). *For the finite-dimensional Hilbert space  $\mathcal{H}$ , of dimension  $d$ , any linear map  $A \in \mathcal{L}(\mathcal{H})$  has a matrix representation:*

$$A = \sum_{j,k=1}^d a_{jk} |j\rangle\langle k|, \tag{1.4}$$

for an arbitrary choice of orthonormal basis  $\{|k\rangle\}_{k=1}^d \in \mathcal{H}$  and  $a_{jk}$  being the matrix elements of  $A$ .

**Definition 1.2.** *Projectors, Unitaries, and Hermitian operators are defined as:*

1. **Projectors**

$P \in \mathcal{L}(\mathcal{H})$  is a projector when  $P^2 = P$ .

2. **Unitaries**

$U \in \mathcal{L}(\mathcal{H})$  is unitary when  $U^\dagger U = U U^\dagger = \mathbb{1}$ .

3. **Hermitian operators**

$H \in \mathcal{L}(\mathcal{H})$  is Hermitian when  $H^\dagger = H$ .

Notice that the space of linear maps  $\mathcal{L}(\mathcal{H})$  itself forms a Hilbert space when equipped with an inner product.

**Definition 1.3** (Hilbert-Schmidt inner product). *Similar to the standard inner product, the Hilbert-Schmidt inner product for  $A, B \in \mathcal{L}(\mathcal{H})$  is given by:*

$$\langle A|B \rangle_{HS} = \frac{1}{\dim(\mathcal{H})} \text{Tr}(A^\dagger B). \quad (1.5)$$

From the definition of the Hilbert-Schmidt inner product, it can be seen that the set of Pauli matrices along with the identity operator forms an orthogonal basis for linear operators on a single qubit.

**Definition 1.4** (Pauli Matrices). *The three  $2 \times 2$  Pauli matrices  $X, Y, Z$  are:*

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

**Remark 1.4** (Eigenvalues and Eigenvectors of Pauli matrices). *Pauli matrices have the eigenvalues  $\lambda = \pm 1$  and the eigenvectors corresponding to these eigenvalues are:*

1.  $X : |\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$  corresponding to  $\lambda = \pm 1$ .
2.  $Y : |y_\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm i|1\rangle)$  corresponding to  $\lambda = \pm 1$ .
3.  $Z : |0\rangle$  and  $|1\rangle$  corresponding to  $\lambda = 1$  and  $\lambda = -1$  respectively.

For linear operators on  $n$  qubits, an orthogonal basis set can be constructed by the tensor product of Pauli matrices and the identity operator,  $\{\mathbf{1}, X, Y, Z\}^{\otimes n}$ . We call such basis elements Pauli strings.

**Definition 1.5** (Cardinality and Locality of linear operators). *Given  $\mathcal{H}$ , the Hilbert space of  $n$  qubits and  $A \in \mathcal{L}(\mathcal{H})$ , a basis decomposition of  $A$ , in terms of Pauli strings, can be obtained as:*

$$A = \sum_{\alpha} c_{\alpha} \bigotimes_{k=1}^n \sigma_k^{\alpha_k}, \quad (1.6)$$

where  $\alpha \in \{0, 1, 2, 3\}^{\times n}$  and  $\sigma_k^{\alpha_k}$  represent the corresponding Pauli matrix according to the labeling  $\sigma_k^0 = \mathbb{1}_k, \sigma_k^1 = X_k, \sigma_k^2 = Y_k$  and  $\sigma_k^3 = Z_k$ . The cardinality,  $|A|$ , is then defined as the number of non-zero coefficients  $c_{\alpha} \in \mathbb{C}$  in Eq. (1.6) and the locality given by the maximum number of non-identity terms taken over the Pauli strings,

$$\bigotimes_{k=1}^n \sigma_k^{\alpha_k}.$$

An implication of the basis decomposition in Eq. (1.6) is that the space of Hermitian operators, denoted by  $\text{herm}_{\mathbb{C}}[2^n]$  can be thought of as the real span over Pauli strings.

**Remark 1.5.** *The space of Hermitian operators in  $\mathcal{L}(\mathcal{H})$  can be defined as:*

$$\text{herm}_{\mathbb{C}}[2^n] = \text{span}_{\mathbb{R}} \left\{ \bigotimes_{k=1}^n \sigma_k^{\alpha_k} \mid \alpha_k \in \{0, 1, 2, 3\} \right\}. \quad (1.7)$$

Another important property to note is that every Hermitian operator,  $H \in \text{herm}_{\mathbb{C}}[2^n]$  gives rise to a unitary operator or sometimes called propagators via exponentiation.

**Definition 1.6** (Matrix exponentiation). *Given a matrix  $A \in \text{Mat}_{\mathbb{C}}[2^n]$ , the matrix exponential of  $A$ , denoted by  $\exp A$  is defined as a series sum:*

$$\exp A = \sum_{k=0}^{\infty} \frac{A^k}{k!}. \quad (1.8)$$

**Definition 1.7** (Propagators). *Given  $H \in \text{herm}_{\mathbb{C}}[2^n]$ , the propagator of  $H$  is defined as:*

$$U(\theta) = \exp(-i\theta H), \quad (1.9)$$

where  $\theta \in [0, 2\pi)$ .

## 1.2 Quantum model of computation

Similar to classical computers, where computation is performed by transforming input states of  $n$  bits into outputs, quantum computers perform series of unitary operations called quantum gates to transform input states of  $n$  qubits. The entire sequence of unitary operations along with the input state is what is referred to as a quantum circuit.

Measurement in quantum mechanics prescribes that the quantum states upon measurement collapse into eigenstates of some measurement observable, which are Hermitian operators in  $\mathcal{L}(\mathcal{H})$ . Given a measurement observable  $A \in \text{herm}_{\mathbb{C}}[2^n]$  with

eigenstates  $\{|\lambda_k\rangle\}_{k=1}^{2^n}$  and the corresponding eigenvalues  $\lambda_k$ , the measurement of a state  $|\psi\rangle = \sum_{k=1}^{2^n} c_k |\lambda_k\rangle$  returns a measurement value  $\lambda_k$  with probability  $|c_k|^2$  up to some normalization  $\sum_{k=1}^{2^n} |c_k|^2$ . Therefore, it is convenient to introduce a unit norm restriction on the Hilbert space  $\mathcal{H}$ , to represent the states of qubits. Unless otherwise stated, in the thesis we shall make this restriction implicitly and define the expected value of Hermitian operators with respect to a state  $|\psi\rangle \in \mathcal{H}$  as:

**Definition 1.8** (Expectation value). *Given  $A \in \text{herm}_{\mathbb{C}}[2^n]$ , the expected value given a unit norm state  $|\psi\rangle \in \mathcal{H}$  is defined as:*

$$\langle \psi | A | \psi \rangle = \sum_{k=1}^{2^n} \lambda_k |c_k|^2, \quad (1.10)$$

where  $\lambda_k$  are the eigenvalues of  $A$  and  $c_k$  are the coefficients of  $|\psi\rangle$  in the eigenbasis of  $A$ .

# Chapter 2

## Ising formulation of classical optimization problems

In this chapter, we introduce classical combinatorial optimization problems and show how finding solutions to problem instances are equivalent to minimizing Ising Hamiltonians. This equivalence allows for a bridge between statistical physics and computation, which is also briefly explored in the chapter.

### 2.1 Background

Computational complexity aims to classify problems according to their worst-case resource usage [42]. Traditionally, in theoretical computer science, two important classical complexity classes are  $P$  and  $NP$ . These classes separate decision problems:  $P$  consists of the set of problems that can be solved in polynomial time on a classical computer, whereas  $NP$  consists of problems which admit polynomial time verification of candidate solutions. One of the most famous problems in computer science is whether  $P \stackrel{?}{=} NP$  [11, 43]. Not only is this question relevant to computer scientists, answering this question has immense practical significance, as many real-world problems inescapably belong to  $NP$ , problems such as in logistics, scheduling, and drug manufacturing, to name a few [44–49].

Decades of research has been spent on attempting polynomial time algorithms to solve the most difficult problems in  $NP$ , aptly named as  $NP$ -Complete problems.

Such problems feature a sense of completeness as all problems in NP polynomially reduced to an instance of NP-Complete problem [50]. Finding a polynomial time algorithm for any one NP-Complete problem proves  $P = NP$ , yet no such algorithm has ever been reported, leading to the wide spread belief that  $P \neq NP$ . Furthermore, many problems in physics are proven to be NP-Complete [41, 51]. This connection to physics motivated several physics inspired algorithms such as simulated annealing [52] and random walk algorithms [53] in hopes of solving NP-Complete problems. Taking a step further, modern approaches, including quantum, rely on building physical computing devices that bootstrap physics to naturally solve NP problems [54–58]. Although such models were initially proposed as a way to outperform classical computation, NP-Complete problems still remain out of reach and therefore provide a good test bed to ascertain the computational power of these models.

## 2.2 Decision and optimization

We will start with the notion of NP-Decision problems [41]. Such problem instances can be represented by pseudo-Boolean functions,

$$f : \mathbb{B}^{\times n} \longrightarrow \mathbb{N}_0. \quad (2.1)$$

Since these are NP-problems, we will demand that for all inputs  $y \in \mathbb{B}^{\times n}$ ,  $f(y)$  can be evaluated in  $\text{poly}(n)$ .

**Definition 2.1** (NP-Decision problem). *The task of a decision problem is to determine the promise i.e., deciding whether a given instance is:*

$$\text{YES instance} \implies \exists y \in \mathbb{B}^{\times n} \mid f(y) = 0,$$

$$\text{NO instance} \implies \forall y \in \mathbb{B}^{\times n}, f(y) \geq 1.$$

The optimization version of NP-Decision problem can be defined in a similar manner.

**Definition 2.2** (Optimization version of NP-Decision problems). *The task of an optimization problem is to determine inputs:*

$$y^* \in \arg \min_{y \in \mathbb{B}^{\times n}} f(y).$$

*Equivalently, optimization problems are sometimes stated as finding the minimum:*

$$f(y^*) = \min_{y \in \mathbb{B}^{\times n}} f(y).$$

## 2.3 Vector space embedding

Consider the following maps between logical bits and qubits:

$$\begin{aligned} \mathbb{B} &\longrightarrow \mathbb{C}^2 \\ \times &\longrightarrow \otimes. \end{aligned} \tag{2.2}$$

Under this map, the pseudo-Boolean function in Eq. (2.1) induces an operator  $H_f \in \mathcal{L}(\mathcal{H})$  as:

$$H_f = \sum_{y \in \mathbb{B}^{\times n}} f(y) |y\rangle\langle y|. \tag{2.3}$$

In this thesis, we shall call such operators  $H_f$ , as problem Hamiltonians. Notice that under this vector space embedding [47, 59], optimization problems (minimization or maximization with respect to appropriate sign change) as defined in Def. 2.2 interrelates Hamiltonian minimization as:

$$\min_{y \in \mathbb{B}^{\times n}} f(y) \equiv \min_{|\psi\rangle \in \mathcal{H}} \langle \psi | H_f | \psi \rangle. \tag{2.4}$$

## 2.4 Combinatorial optimization problems

We will briefly review some of the NP-Complete problems and their corresponding NP-Hard optimization versions [41]. The thesis will explore these problems in the context of QAOA [26], and therefore, we provide the explicit constructions of the

considered combinatorial optimization problem instances into problem Hamiltonians.

### 2.4.1 Boolean satisfiability

Boolean satisfiability, or SAT, is the problem of determining satisfiability of Boolean formulae. More specifically, one aims to decide whether a given formula can be satisfied by assigning truth values to the variables. Limited to  $k$ -variables or literals per clause, and expressed in the conjunctive normal form (CNF),  $k$ -SAT was proven to be NP-Complete for  $k \geq 3$  [60].

**Example 2.1.** Consider the following example of a 3-SAT instance with 2 clauses and 5 variables:

$$f(x_1, x_2, x_3, x_4, x_5) = (x_1 \vee \neg x_2 \vee x_3) \wedge (x_1 \vee x_4 \vee \neg x_5). \quad (2.5)$$

This instance is satisfiable since variable assignment  $x_1 = 1, x_2 = 1, x_3 = 0, x_4 = 0, x_5 = 0$  output  $f(x_1, x_2, x_3, x_4, x_5) = 1$ .

**Definition 2.3** (Random  $k$ -SAT instances). Let  $X$  denote the set of  $n$  Boolean variables,  $X = \{x_j\}_{j=1}^n$ . A random clause  $C_l$  is then defined as the disjunction over  $k$ -tuples of random distinct literals chosen from  $X$ ,

$$C_l = \bigvee_{j=1}^k (y_j)^{l_j}, \quad (2.6)$$

where  $y_j \in X$  and  $l_j \in \{0, 1\}$  references the negation of the variable  $y_j$ .  $k$ -SAT instances are then constructed by the conjunction of  $m$  such randomly chosen distinct clauses,

$$f(x_1, x_2, \dots, x_n) = \bigwedge_{l=1}^m C_l. \quad (2.7)$$

**Remark 2.1** (Clause density). To study random instances of  $k$ -SAT, we introduce an order parameter for instance generation called clause density which is defined as:

$$\alpha = \frac{m}{n}, \quad (2.8)$$

where  $m$  is the number of clauses in the  $k$ -SAT instance and  $n$  is the number of variables.

The optimization version of Boolean satisfiability, called maximum satisfiability or MAX- $k$ -SAT, is the canonical NP-Hard optimization problem (for  $k \geq 2$ ) [61]. Here, one seeks to determine variable assignments that maximize the number of satisfiable clauses in a given instance.

**Definition 2.4** (MAX- $k$ -SAT). *Given a  $k$ -SAT instance with  $m$  clauses and  $n$  variables as in Eq. (2.7), let  $g(x)$  be the pseudo-Boolean function obtained by replacing the conjunction in Eq. (2.7) by the summation,*

$$g(x) = \sum_{l=1}^m C_l, \quad (2.9)$$

where  $x \in \mathbb{B}^{\times n}$ . MAX- $k$ -SAT aims to determine:

$$x^* \in \arg \max_{x \in \mathbb{B}^{\times n}} g(x). \quad (2.10)$$

In order to construct a problem Hamiltonian for MAX- $k$ -SAT, we consider the following maps:

$$\begin{aligned} x_j &\longrightarrow P_j^0, & \neg x_j &\longrightarrow P_j^1, \\ \wedge &\longrightarrow +, & \vee &\longrightarrow \otimes. \end{aligned} \quad (2.11)$$

Here,  $P_j^1 = (|1\rangle\langle 1|)_j$  and  $P_j^0 = (|0\rangle\langle 0|)_j$  are rank one projectors acting on the  $j^{\text{th}}$  qubit. This construction ensures that each clause is mapped onto projectors,  $P_{ij\dots k}^{\alpha\beta\dots\gamma} = P_i^\alpha \otimes P_j^\beta \otimes \dots \otimes P_k^\gamma$ , which penalize unsatisfiable assignments  $x \in \mathbb{B}^{\times n}$  with a penalty  $\langle x | P_{ij\dots k}^{\alpha\beta\dots\gamma} | x \rangle = 1$ .

The full Hamiltonian  $H_{\text{SAT}}$ , is then constructed by summing over all the clauses in the instance,

$$H_{\text{SAT}} = \sum_{l=1}^m \mathcal{C}_l \{P_{ij\dots k}^{\alpha\beta\dots\gamma}\}, \quad (2.12)$$

where  $\mathcal{C}_l$  assigns the value of  $\alpha, \beta, \dots, \gamma \in \{0, 1\}$ , corresponding to the negation of the literals indexed by  $i, j, \dots, k \in \{1, 2, \dots, n\}$  in the  $l^{\text{th}}$  clause. Note that the

minimization,  $\min_{|\psi\rangle \in \mathcal{H}} \langle \psi | H_{\text{SAT}} | \psi \rangle$  is equivalent to the maximization of Eq. (2.9).

**Remark 2.2** (2-Local SAT Hamiltonians). *The projectors appearing in Eq. (2.11) can be decomposed into the Pauli basis as:*

$$P_j^\alpha = \frac{1}{2} (\mathbb{1} + (-1)^\alpha Z_j). \quad (2.13)$$

*Considering 2-SAT instances and expanding Eq. (2.12) into the Pauli basis, one can show that  $H_{2\text{SAT}}$  admits a 2-local Hamiltonian of the form:*

$$H_{2\text{SAT}} \sim \sum_{j < k} J_{jk} Z_j Z_k + \sum_j h_j Z_j, \quad (2.14)$$

*with appropriate coefficients  $h_j$  and  $J_{jk}$ .*

## 2.4.2 Maximum cut

Given an undirected graph, maximum cut or MAX-CUT is the NP-Hard [62] optimization problem of partitioning nodes into two complementary subsets (say  $S$  and  $S'$ ) in such a way that the number of edges between the two subsets are maximized, commonly referred to as the MAX-CUT value of the graph.

Let  $G = (V, E)$  denote a graph with a set of vertices  $V$  ( $|V| = n$ ) and edges  $E$  ( $|E| = m$ ). To construct the pseudo-Boolean function for this problem, we introduce binary variables  $\{x_k\}_{k=1}^n$  to represent which subset vertex  $v_k \in V$  belongs in:

$$x_k = \begin{cases} 0 & \text{if } x_k \in S, \\ 1 & \text{if } x_k \in S'. \end{cases} \quad (2.15)$$

Consider an edge  $\langle j, k \rangle \in E$ , we seek a pseudo-Boolean function such that equal assignments,  $x_j = x_k = 1$  or  $0$ , do not contribute to the MAX-CUT value.

**Remark 2.3** (Kernel embedding for equal assignments). *Given an edge  $\langle j, k \rangle \in E$ , the following pseudo-Boolean function embeds equal assignments,  $x_j = x_k = 1$  or  $0$ , into its Kernel,*

$$g(x_j, x_k) = x_j + x_k - 2x_j x_k. \quad (2.16)$$

**Definition 2.5** (MAX-CUT). *Given  $G = (V, E)$  with set of vertices  $V$  ( $|V| = n$ ) and edges  $E$  ( $|E| = m$ ), a candidate cut value of  $G$  for an assignment  $x \in \mathbb{B}^{\times n}$  is given by:*

$$g(x) = \sum_{\langle j,k \rangle \in E} (x_j + x_k - 2x_j x_k). \quad (2.17)$$

*MAX-CUT aims to determine:*

$$x^* \in \arg \max_{x \in \mathbb{B}^{\times n}} g(x). \quad (2.18)$$

In order to construct the problem Hamiltonian corresponding to Eq. (2.17), we map binary variables  $x_j \rightarrow \frac{1}{2}(\mathbb{1} + Z_j)$ , where  $Z_j$  is the Pauli-Z matrix acting on the  $j^{\text{th}}$  qubit. The cut values as defined in Eq. (2.17) can then be recovered as the expected value of the following 2-local Hamiltonian,

$$H = \frac{1}{2} \sum_{\langle j,k \rangle \in E} (\mathbb{1} - Z_j Z_k). \quad (2.19)$$

Note that maximizing the cut value in Eq. (2.17) is equivalent to minimizing  $\sum_{\langle j,k \rangle \in E} Z_j Z_k$  terms in Eq. (2.19).

### 2.4.3 Graph optimization

As a generalization to MAX-CUT, one may introduce constraints by imposing edge weights prescribed by some matrix  $W \in \mathbb{R}^{n \times n}$  with elements  $w_{jk} = w_{kj} \neq 0 \forall \langle j, k \rangle \in E$ . This optimization problem, sometimes called weighted MAX-CUT or in general graph optimization involves maximizing the pseudo-Boolean function,

$$g(x) = \sum_{\langle j,k \rangle \in E} w_{jk} (x_j + x_k - 2x_j x_k), \quad (2.20)$$

which is equivalent to minimizing a corresponding 2-local Hamiltonian,

$$H = \sum_{\langle j,k \rangle \in E} w_{jk} Z_j Z_k. \quad (2.21)$$

**Remark 2.4** (Erdos-Renyi random graphs). *A random graph instance corresponds*

to sampling from a uniform distribution over all possible graphs containing  $n$ -vertices and  $m$ -edges. Similar to the clause density in the case of MAX- $k$ -SAT, an order parameter for instance generation (called graph density) can be defined as:

$$\alpha = \frac{m}{n}. \quad (2.22)$$

## 2.5 Computational phase transitions

In physics, a phase transition corresponds to an abrupt, discontinuous change in the macroscopic properties of a physical system. Here, we will explore a similar feature in the context of combinatorial optimization problems called computational phase transitions [41].

Computational complexity of solving NP-problems prescribe an exponential behavior for the asymptotic time complexity. This, however, is stated in terms of an upper bound or a worst-case scenario considering all possible instances at problem size  $n$ . For typical cases, one considers asymptotic behavior for time complexity over random ensembles of instances generated according to some order parameter. Computational phase transitions are then defined as abrupt changes in the asymptotic behavior of time complexity with respect to the appropriate choice of order parameter.

**Definition 2.6** (Easy and Hard instances). *Regions of the ensemble space (typically characterized by an order parameter) of instances which require polynomially increasing time complexity are categorized as Easy instances. Regions elsewhere, instances are categorized as Hard.*

### 2.5.1 Review of some computational phase transitions

1. 3-SAT is NP-Complete and features a computational phase transition with an Easy-Hard-Easy pattern [63]. The best known lower and upper bounds for the critical clause density are:  $3.52 \leq \alpha_c \leq 4.53$ .
2. MAX-2, 3-SAT is NP-Hard and features a computational phase transition with

an Easy-Hard pattern at critical clause density  $\alpha_c \approx 1$  [64].

## 2.6 Ising model

As was shown previously, we treat optimization of combinatorial problem instances as Hamiltonian minimization and explored certain examples where instances map onto 2-local Hamiltonians. Although these Hamiltonians are artificially constructed, their form is similar to that of Hamiltonians of a physical system called the Ising Model [65].

Ising model, initially proposed to explain the magnetic properties of materials, consists of discrete variables or Ising spins arranged in a  $d$ - dimensional lattice. The spins take binary values,  $\sigma_k \in \{-1, +1\}$ , representing the magnetic dipole moment of a lattice site  $k$ , and are allowed to interact with neighboring spins via short range dipole-dipole couplings. In the presence of an external applied magnetic field, the model describes the energy of a configuration of spins  $\boldsymbol{\sigma}$ , via an Ising Hamiltonian,  $H(\boldsymbol{\sigma})$ .

**Definition 2.7** (Ising Hamiltonian). *Given  $\Lambda$ , a  $d$ -dimensional lattice of Ising spins  $\sigma_k, k \in \Lambda$ , let  $\langle j, k \rangle$  represent pairs of neighboring spins and  $h_j$  be the external local field applied to a site  $j$ . Then, for a configuration of spins  $\boldsymbol{\sigma} = (\sigma_k)_{k \in \Lambda}$ , the energy is given by:*

$$H(\boldsymbol{\sigma}) = \sum_{\langle j, k \rangle} J_{j, k} \sigma_j \sigma_k + \sum_j h_j \sigma_j. \quad (2.23)$$

**Remark 2.5** (Interaction types). *Based on the sign of the couplings  $J_{jk}$  in Eq. (2.23), interactions can be of two types:*

1) **Ferromagnetic:** Here  $J_{jk} < 0$  and therefore parallel configurations of spins lower the energy of the system.

2) **Anti-Ferromagnetic:** Here  $J_{jk} > 0$  and therefore antiparallel configurations of spins lower the energy of the system.

Each spin configuration  $\boldsymbol{\sigma}$ , constitutes a microstate for the physical system.

Thermodynamics dictates, in equilibrium with a thermal bath at some fixed temperature  $T$ , microstates of the system are described by the Gibbs's distribution:

$$\rho(\boldsymbol{\sigma}) = \frac{1}{\mathcal{Z}} \exp\left(-\frac{1}{T}H(\boldsymbol{\sigma})\right), \quad (2.24)$$

where the partition function  $\mathcal{Z} = \sum_{\boldsymbol{\sigma}} (\exp(-\frac{1}{T}H(\boldsymbol{\sigma})))$ .

Macroscopic properties such as energy density and magnetization per degree of freedom, are then given by ensemble averages over microstates at some fixed temperature  $T$ .

**Remark 2.6** (Extremas). *Based on the order parameter  $T$ , we encounter two extreme cases:*

- 1) *Temperature  $T = \infty$ : The probability  $\rho(\boldsymbol{\sigma})$  becomes independent of the energy  $H(\boldsymbol{\sigma})$ , each microstate  $\boldsymbol{\sigma}$  is equiprobable and the system is said to be in a disordered phase.*
- 2) *Temperature  $T = 0$ : The probability  $\rho(\boldsymbol{\sigma})$  collapses into the global minima of  $H(\boldsymbol{\sigma})$ , i. e., the ground states. The system is frozen into these states.*

## 2.7 Physics inspired algorithms for minimizing Ising Hamiltonians

As one might infer from the physics of the Ising model, by tuning the inverse temperature in Eq. (2.24), one can drive the system into its low-energy configuration. Since solutions to combinatorial optimization problem instances map onto ground states of some Ising type Hamiltonian, this idea, from a computational standpoint, has been explored extensively to generate physics-inspired algorithms, albeit heuristic, to approximate solutions to combinatorial problem instances. We review some of these algorithms here as they motivate the main topic of this thesis, Quantum Approximate Optimization Algorithm (explored in the forthcoming chapters).

### 2.7.1 Simulated annealing

Given a system specified by some Hamiltonian  $H(\sigma)$ , simulated annealing aims to prepare low-energy configuration of  $H(\sigma)$  [52]. The algorithm works by employing

a stochastic dynamics governed by the master equation:

$$\frac{d}{dt}\rho(\sigma, t) = \sum_{\sigma'} M(\sigma|\sigma', t)\rho(\sigma', t). \quad (2.25)$$

Here,  $M$  is the matrix of transition probabilities  $p(\sigma|\sigma', t)$ , for a configuration  $\sigma'$  at time  $t$  to transition into  $\sigma$  and  $\rho(\sigma, t = 0)$  is the initial distribution that is arbitrarily set. In order for the system to equilibrate into the Gibbs distribution  $\rho_{eq}(\sigma, t) = \frac{1}{Z} \exp(-\beta(t)H(\sigma))$  at some inverse temperature  $\beta(t)$ , the dynamics prescribed by  $M$  is artificially designed by forcing conservation of probabilities,

$$\sum_{\sigma} p(\sigma|\sigma', t) = 1, \quad (2.26)$$

and a detailed balance condition,

$$M(\sigma|\sigma', t)\rho_{eq}(\sigma', t) = M(\sigma'|\sigma, t)\rho_{eq}(\sigma, t). \quad (2.27)$$

Given that the system evolves for a sufficient period of time, the probability distribution converges to the equilibrium distribution. By setting  $\beta(t) = \beta \gg 1$ , simulated annealing recovers low-energy configurations for  $H(\sigma)$ .

## 2.7.2 Adiabatic quantum computation

Here we shall explore another method to recover low-energy configurations of some problem Hamiltonian. To establish how adiabatic quantum computation works [66, 67], we first recall the time evolution of a quantum mechanical system.

Given  $H(t)$ , the Hamiltonian describing the system at a time  $t$ , the time evolution of the physical state of the system is governed by the Schrödinger equation,

$$i \frac{d}{dt} |\psi(t)\rangle = H(t) |\psi(t)\rangle. \quad (2.28)$$

The quantum adiabatic theorem states that a system initially in its ground state tends to remain in its low energy configuration if the time dependent Hamiltonian varies slowly [68]. The theorem also provides a bound on the slowdown that is

required. Adiabatic quantum computation takes advantage of this principle.

Consider  $H(s)$  to be a time dependent Hamiltonian for  $0 \leq s \leq 1$  and let  $|\phi(s)\rangle$  denote its ground state. The goal is to evolve an initial state  $|\psi(0)\rangle = |\phi(0)\rangle$  by varying  $H(s)$ ,  $s : 0 \rightarrow 1$  slowly. This is done by introducing a delay factor  $\tau(s)$  into the Schrodinger equation such that,

$$i \frac{d}{ds} |\psi(s)\rangle = \tau(s) H(s) |\psi(s)\rangle. \quad (2.29)$$

If

$$\tau(s) \gg \frac{\left\| \frac{d}{ds} H(s) \right\|_2}{g(s)^2}, \quad (2.30)$$

where  $g(s)$  is the gap between the two smallest eigenvalues of  $H(s)$ , then the adiabatic theorem states that  $|\psi(1)\rangle \approx |\phi(1)\rangle$ . In adiabatic quantum computation,  $H(s)$  is suitable parameterized such that  $H(0)$  is a Hamiltonian with an easy-to-prepare known ground state and  $H(1)$  is the optimization problem instance. An example of one such parameterization is the following:

$$H(s) = (1 - s) H_{init} + s H_{final}. \quad (2.31)$$

In most cases,  $g(s)$  is hard to determine for all  $s$ , therefore a constant delay factor  $\tau(s) = \tau_c \in \mathcal{O}\left(\frac{\Delta_{max}}{g_{min}^2}\right)$ , is used and  $\Delta_{max} = \max_s \left\| \frac{d}{ds} H(s) \right\|_2$  and  $g_{min} = \min_s g(s)$ .

## 2.8 Discussion

In this chapter, we explored several combinatorial optimization problems and establish how such problems can be formulated as finding the ground state energy of the Ising model albeit with non-physical couplings. In its original formulation, the simplistic model was introduced to explain the magnetic behavior of certain materials known as ferromagnets [65]. Such materials exhibit a phase transition: going from a disordered phase into a ferromagnetic configuration, when cooled beyond some critical temperature. It is well known that one of the hallmarks of success in statistical mechanics is its ability to explain and predict this critical behavior.

However, for more exotic materials—systems that exhibit frustrations and disorder—a much richer and more complex behaviors are reported [69–71]. Such materials, called spin glasses, currently push the limits of statistical mechanics theories and hint towards the existence of a new phase of matter. Recent advancements such as Replica-Symmetry-Breaking allude towards such exotic glassy phases [72], however, a full description still proves to be challenging.

Spin glasses are also of particular interest in theoretical computer science. This comes from the fact that finding the lowest energy configurations of spin glasses is proven to be complete for the complexity class NP [62]. Since all problems in NP reduce to an instance of any one NP-Complete problem [50], the physics of driving spin glasses to their low energy configurations provides immense insights into developing efficient algorithms for solving NP problems. Indeed, a wide variety of physics inspired algorithms have been designed as a means for minimizing spin glass problems [54–58, 73]. Even in quantum computation, this bridge between statistical mechanics and computation is currently being explored in an attempt to establish the Church–Turing–Deutsch principle [9]. It is expected that when more powerful quantum computers are developed, a clear feature of supremacy will arise from their ability to solve NP problems.

# Chapter 3

## Quantum approximate optimization

In this chapter, we introduce the Quantum Approximate Optimization Algorithm (or QAOA) [26] and explore its use case as an algorithm to approximate ground states of some problem Hamiltonian. The chapter begins by introducing a variational approach to Hamiltonian minimization and describes QAOA in this setting. To study algorithmic performance, different metrics are widely used in the literature. The chapter includes a brief review of the important ones and provides a theorem for interrelating different performance metrics. Since the algorithm was inspired by adiabatic quantum computation and, in principle, can approximate any adiabatic evolution in the infinite depth limit [66], a brief overview of the relation between these two approaches is elaborated next [67]. Finally, the chapter ends by recovering the well-known Grover’s unstructured database search [15] in the variational framework.

### 3.1 Background

To fundamentally understand what can and cannot be achieved in a prescribed model of computation, one has to invoke the notion of universality [9]. Simply stated, universality refers to the ability to perform or realize arbitrary operations. In the case of classical computers, a model of computation is said to be universal if it can be shown, within the constraints of the model, that any arbitrary input bitstring can be transformed into any given output bitstring [74, 75]. Universality for

quantum computers can analogously be defined. The standard model for universal quantum computation is based on the gate model or circuit model, where operations are performed as a sequence of unitaries or gates selected from a universal set [10, 76, 77]. Though universal, practical realizations of this model feature limited capabilities owing to systematic constraints such as noise and low coherence times. Such devices, called Noisy-Intermediate-Scale-Quantum devices (NISQ), therefore possess very limited capabilities [16–18]. Another approach towards universality is in the setting of adiabatic quantum computation [78]. Though discussed briefly in the previous chapter, the general framework can be extended into what is known as the ground state model of computation, wherein the ground state of some suitable Hamiltonian corresponds to the output of a quantum circuit. Therefore, universality can be argued by considering Hamiltonian minimization via adiabatic evolution as a means to prepare the output of any given quantum circuit [79]. Furthermore, this approach naturally provides a path towards solving a wide class of NP-Hard combinatorial optimization problems [67]. However, the need for designing physical systems capable of realizing such artificial Hamiltonians poses a significant developmental challenge.

Considering the best of both approaches, a new model of quantum computing originates, known as the variational model [19, 20, 80]. This model exploits a hybrid setup of a quantum processor coupled to an outer-loop classical optimization routine. More specifically, a NISQ device is programmed to prepare a family of parameterized quantum states called ansatz [26, 81, 82]. The ansatz are then measured in the Pauli basis to compute the expected value of some Hamiltonian to be minimized. If the class of Hamiltonians is restricted to polynomially bounded cardinality, then a classical computer can recover expected values efficiently from the measurement statistics. These values are then passed onto an iterative classical optimization routine which variationally adjusts the ansatz to obtain an approximate minimum for the expected value. Such an approach allows for alleviating some of the systematic limitations of the NISQ devices and overcomes the need for designing artificial systems described by programmable Hamiltonians. Moreover, variational models have recently been shown to also admit universality and therefore have promising

prospects in the NISQ era of quantum computing [27, 28, 79].

## 3.2 Variational approach to Hamiltonian minimization

Consider a problem Hamiltonian  $H$ . We wish to approximate the ground state of  $H$ . This can be done variationally on a quantum computer by constructing parameterized quantum states called ansatz [26, 81, 82]. Typically, one achieves this by applying a sequence of parameterized unitaries or gates onto some initial state.

**Definition 3.1** (Ansatz). *Given a sequence of parameterized unitaries  $\{U_k(\theta_k)\}_{k=1}^p$ ,  $\theta_k \in \mathbb{R}$  that can be realized, and for some fixed circuit depth  $p \in \mathbb{N}$ , an ansatz state is defined as the parameterized quantum state:*

$$|\psi(\boldsymbol{\theta})\rangle = \prod_{k=1}^p U_k(\theta_k) |0\rangle^{\otimes n}. \quad (3.1)$$

By varying over the parameter space,  $\boldsymbol{\theta} \in \mathbb{R}^{\times p}$ , one can generate a variational state space as:

$$\Omega = \bigcup_{\boldsymbol{\theta}} \left\{ |\psi(\boldsymbol{\theta})\rangle \right\}. \quad (3.2)$$

Note that in general  $\Omega \subseteq \mathcal{H}$ , where  $\mathcal{H}$  is the Hilbert space of  $n$ -qubits. The variational approach involves searching for the ground state of  $H$  in the variational state space  $\Omega$ , by minimizing the expected value of the Hamiltonian:

$$\min_{\psi \in \Omega} \langle \psi | H | \psi \rangle \geq \min_{\phi \in \mathcal{H}} \langle \phi | H | \phi \rangle. \quad (3.3)$$

Such a minimization can be performed via a hybrid classical-quantum feedback loop. Specifically, one starts with a guess for variational parameters  $\boldsymbol{\theta}$ , and prepares the state according to Eq. (3.1). Measurement of this state allows for computing the objective function in Eq. (3.3) which is then passed onto a classical outer-loop

optimizer. The routine terminates with the optimizer returning a set of optimal parameters  $\boldsymbol{\theta}^*$ , which then allows for preparing a candidate ground state  $|\psi(\boldsymbol{\theta}^*)\rangle$ . The challenge in the variational approach is to construct variational state spaces that are expressive enough to approximate the true ground state of  $H$ .

### 3.3 Quantum approximate optimization algorithm

QAOA [26] is an instance of the class of variational quantum algorithms. Here, we describe how the algorithm works.

Given an optimization problem instance and its corresponding problem Hamiltonian  $H$ , the usual QAOA procedure is as follows:

1. Generate ansatz states,  $|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle$  on a quantum computer, where  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_p)$  and  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$  are tunable real parameters over some fixed range. The state is prepared by applying a sequence of  $2p$ -parameterized unitary gates acting on the reference state  $|+\rangle^{\otimes n}$ , which is the symmetric superposition of all  $2^n$  computational basis states.

$$|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = \prod_{k=1}^p U(\gamma_k, \beta_k) |+\rangle^{\otimes n}, \quad (3.4)$$

where

$$U(\gamma_k, \beta_k) = \exp(-i\beta_k H_x) \cdot \exp(-i\gamma_k H). \quad (3.5)$$

The Hamiltonian  $H_x = \sum_{k=1}^n X_k$ , is called the mixer Hamiltonian and performs a mixing of phases propagated by  $H$ ; typically a diagonal matrix in the computational basis. A convention we follow throughout the thesis is to call  $p$ , the QAOA depth.

2. Measurement of the state in Eq. (3.4) is performed to compute the expected value,  $\langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | H | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle$ . This is an approximation which can be calculated efficiently on a classical computer.
3. Steps 1 and 2 are repeated, and a classical optimization algorithm is used to assign a set of optimal parameters,  $\boldsymbol{\gamma}^*$  and  $\boldsymbol{\beta}^*$  that minimize  $\langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | H | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle$ .

$|\psi(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)\rangle$  then recovers a ground state approximation for  $H$ .

**Remark 3.1** (Improving approximations). *Standard fixed ranges for the parameters are  $\gamma_k \in [0, 2\pi)$ , and  $\beta_k \in [0, \pi]$  for all  $k = 1, 2, \dots, p$ . It can be seen that QAOA approximations improve with the QAOA depth  $p$  as:*

$$\begin{aligned} \langle + |^{\otimes n} H | + \rangle^{\otimes n} &= \frac{1}{2^n} \text{Tr}(H) \leq \langle \psi | H | \psi \rangle_{p=1} \leq \langle \psi | H | \psi \rangle_{p=2} \cdots \\ &\cdots \leq \langle \psi | H | \psi \rangle_{p=l} \leq \langle \psi | H | \psi \rangle_{p=l+1}, \end{aligned} \quad (3.6)$$

where we make use of the notation:

$$\langle \psi | H | \psi \rangle_{p=l} \equiv \min_{\boldsymbol{\gamma}, \boldsymbol{\beta}} \langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | H | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle, \quad (3.7)$$

and  $\boldsymbol{\gamma} \in [0, 2\pi)^{\times l}$  and  $\boldsymbol{\beta} \in [0, \pi)^{\times l}$ .

Note that though approximations improve with QAOA depth, there are two main challenges: 1) it comes at the cost of classically optimizing over an exponentially increasing parameter space and 2) deeper circuits are limited to noise in practical implementations. Therefore, in most cases, the QAOA depths are fixed and studied in a low depth setting.

**Remark 3.2** (Variations–Quantum Alternating Operator Ansatz [83]). *Standard implementation of QAOA relies on constructing ansatz via an alternating sequence as shown in Eq. (3.5), with  $H_x = \sum_k X_k$  and some problem Hamiltonian say  $H_z$ . Note that for phase mixing, the commutator  $[H_x, H_z] \neq 0$ . In this regard, QAOA ansatz can be generalized into:*

$$|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = \prod_{k=1}^p U(\gamma_k, \beta_k) |\phi\rangle, \quad (3.8)$$

where

$$U(\gamma_k, \beta_k) = \exp(-i\beta_k H_1) \cdot \exp(-i\gamma_k H_2). \quad (3.9)$$

In this construction, the commutator  $[H_1, H_2] \neq 0$  and  $|\phi\rangle$  is chosen to be the ground state of  $H_1$ . Such a construction is still valid for QAOA and is termed Quantum

*Alternating Operator Ansatz.*

### 3.4 Performance metrics

Here, we briefly review some of the metrics used to study the algorithmic performance of QAOA.

**Definition 3.2** (Error in Approximations). *Given a problem Hamiltonian  $H$ , whose ground states we wish to approximate using  $p$ -depth QAOA, a metric to assess the algorithmic performance is given by error in approximation:*

$$f = \min_{\psi \in \Omega} \langle \psi | H | \psi \rangle - \min_{\phi \in \mathcal{H}} \langle \phi | H | \phi \rangle. \quad (3.10)$$

Here,  $\Omega$  represents the variational state space available for a  $p$ -depth QAOA ansatz,

$$\Omega = \bigcup_{\gamma, \beta} \left\{ |\psi(\gamma, \beta)\rangle \right\}. \quad (3.11)$$

QAOA exactly recovers the ground states whenever  $f = 0$ .

**Definition 3.3** (Residual Energy). *Given a problem Hamiltonian  $H$ , whose ground states we wish to approximate using  $p$ -depth QAOA, let  $H_{min}$  and  $H_{max}$  represent the ground state energy and maximal excited state energy of  $H$ . The residual energy is then defined as:*

$$E_{res} = \frac{\langle \psi(\gamma, \beta) | H | \psi(\gamma, \beta) \rangle - H_{min}}{H_{max} - H_{min}}. \quad (3.12)$$

QAOA exactly recovers the ground states whenever  $E_{res} = 0$ .

**Definition 3.4** (Approximation Ratio). *Given a problem Hamiltonian  $H > 0$ , whose ground states we wish to approximate using  $p$ -depth QAOA, let  $H_{min}$  represent the ground state energy of  $H$ . The approximation ratio is then defined as:*

$$r = \frac{H_{min}}{\langle \psi(\gamma, \beta) | H | \psi(\gamma, \beta) \rangle}. \quad (3.13)$$

QAOA exactly recovers ground states whenever  $r = 1$

**Definition 3.5** (Success Probability or Ground State Overlap). *Given a problem Hamiltonian  $H$ , whose ground states we wish to approximate using  $p$ -depth QAOA, let  $\{|gs_k\rangle\}_{k=1}^d$  be the set of possibly  $d$ -degenerate ground states of  $H$ . Then the ground state space overlap is given by,*

$$\eta = \sum_{k=1}^d |\langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | gs_k \rangle|^2. \quad (3.14)$$

QAOA exactly recovers the ground states whenever  $\eta = 1$ .

Note that the metrics, error in approximations and the residual energy are trivially related by the unit normalization given by  $H_{\max} - H_{\min}$ . Moreover, this normalization cannot be calculated a priori without a brute-force search for  $H_{\max}$  and  $H_{\min}$ .

Another important caveat to note is that although Defs. 3.2, 3.3 and 3.4 are indeed valid performance metrics to study QAOA, they do not directly indicate proximity (in terms of a distance metric) to the true ground state space of  $H$ . Ground state overlap circumvents this issue; however, one cannot calculate Eq. (3.14) in an actual experiment as it requires knowledge of the set of possibly  $d$ -degenerate ground states of  $H$ .

Taking these into account, we formulate a relation between the ground state overlap and the error in approximations.

**Theorem 3.1** (Stability of approximations [79]). *Let  $H \geq 0$  be an optimization problem instance with 1)  $d$ -degenerate ground state energy,  $\lambda_0$  2) minimal spectral gap,  $\Delta$  and 3) maximal excited state energy,  $\lambda_{\max}$ . Then, for any ground state approximation,  $|\psi\rangle$  satisfying  $\langle \psi | H | \psi \rangle \leq \Delta$ , the error in approximations and ground state overlap satisfy:*

$$1 - \frac{\langle \psi | H | \psi \rangle - \lambda_0}{\Delta} \leq \eta(\psi) \leq 1 - \frac{\langle \psi | H | \psi \rangle - \lambda_0}{\lambda_{\max} - \lambda_0}. \quad (3.15)$$

*Proof.* The Hamiltonian in its eigenbasis can be expressed as,

$$H = \lambda_0 \sum_{j=1}^d |\phi_j\rangle\langle\phi_j| + \sum_{k>d} \lambda_k |\phi_k\rangle\langle\phi_k|. \quad (3.16)$$

Here, we reorder the eigenbasis so that  $\lambda_0 < \lambda_1 \leq \dots \leq \lambda_k \dots \leq \lambda_{max}$ . Consider the expectation value of the Hamiltonian on an arbitrary state  $|\psi\rangle$ ,

$$\langle \psi | H | \psi \rangle = \lambda_0 \sum_{j=1}^d |\langle \phi_j | \psi \rangle|^2 + \sum_{k>d} \lambda_k |\langle \phi_k | \psi \rangle|^2. \quad (3.17)$$

Assuming the gap  $\Delta = \lambda_1 - \lambda_0$ , is known,

$$\begin{aligned} \langle \psi | H | \psi \rangle &= \lambda_0 \sum_{j=1}^d |\langle \phi_j | \psi \rangle|^2 + \sum_{k>d} \lambda_k |\langle \phi_k | \psi \rangle|^2 \\ &\geq \lambda_0 \left( \sum_{j=1}^d |\langle \phi_j | \psi \rangle|^2 \right) + (\lambda_0 + \Delta) \left( \sum_{k>d} |\langle \phi_k | \psi \rangle|^2 \right) \\ &= \lambda_0 \left( \sum_{j=1}^d |\langle \phi_j | \psi \rangle|^2 \right) + (\lambda_0 + \Delta) \left( 1 - \sum_{j=1}^d |\langle \phi_j | \psi \rangle|^2 \right) \\ &= \lambda_0 + \Delta \left( 1 - \sum_{j=1}^d |\langle \phi_j | \psi \rangle|^2 \right). \end{aligned} \quad (3.18)$$

Rearranging the terms then gives the required lower bound as:

$$\sum_{j=1}^d |\langle \phi_j | \psi \rangle|^2 \geq 1 - \frac{\langle \psi | H | \psi \rangle - \lambda_0}{\Delta}. \quad (3.19)$$

For the upper bound, we substitute for each excited state energies  $\lambda_k$  in Eq. (3.17) by  $\lambda_{max}$ ,

$$\begin{aligned} \langle \psi | H | \psi \rangle &\leq \lambda_0 \sum_{j=1}^d |\langle \phi_j | \psi \rangle|^2 + \lambda_{max} \sum_{k>d} |\langle \phi_k | \psi \rangle|^2 \\ &= \lambda_0 \left( \sum_{j=1}^d |\langle \phi_j | \psi \rangle|^2 \right) + \lambda_{max} \left( 1 - \sum_{j=1}^d |\langle \phi_j | \psi \rangle|^2 \right). \end{aligned} \quad (3.20)$$

Rearranging the terms now gives the required upper bound as:

$$\sum_{j=1}^d |\langle \phi_j | \psi \rangle|^2 \leq 1 - \frac{\langle \psi | H | \psi \rangle - \lambda_0}{\lambda_{max} - \lambda_0}. \quad (3.21)$$

■

### 3.5 Recovering adiabatic evolution

Although QAOA is introduced in the thesis as a variational approach to Hamiltonian minimization, the algorithm was originally inspired by adiabatic quantum computation [26, 67]. It can be shown that in the limit of QAOA depth  $p \rightarrow \infty$ , QAOA has the potential to recover any adiabatic evolution. This is particularly important as adiabatic evolution guarantees exact recovery of ground states, albeit with long adiabatic evolution times. Here, we describe how a discretization of adiabatic evolution results in a form that resembles the QAOA ansatz. We also derive how the errors scale in this approximation, thus providing an upper bound on the circuit depth of QAOA to approximate a given adiabatic evolution [67].

Consider a problem Hamiltonian  $H_z$ , whose ground states we wish to prepare. In adiabatic quantum computing, one considers the time evolution of the ground state of some easy-to-prepare initial Hamiltonian  $H_x$ . To describe this evolution we recall the Schrödinger equation with a time-dependent Hamiltonian,

$$H(t) = \left(1 - \frac{t}{T}\right) H_x + \frac{t}{T} H_z, \quad (3.22)$$

with  $t \in [0, T]$  and  $T \in \mathcal{O}\left(\frac{\Delta_{max}}{g_{min}^2}\right)$  to satisfy adiabatic theorem. Evolution of the ground state of  $H_x$ , say  $|\phi(0)\rangle$  under  $H(t)$  is then given by:

$$i \frac{d}{dt} |\phi(t)\rangle = H(t) |\phi(t)\rangle. \quad (3.23)$$

Solution of the time dependent Schrödinger equation in Eq. (3.23) is given by the time evolution operator,

$$|\phi(T)\rangle = U(T) |\phi(0)\rangle, \quad (3.24)$$

where

$$U(T) = \mathcal{T} \left\{ \exp \left( -i \int_0^T H(t) dt \right) \right\}. \quad (3.25)$$

Here,  $\mathcal{T}$  is the time ordering operator. By the adiabatic theorem,  $|\phi(T)\rangle$  is guaranteed to be a close approximation to the ground state of  $H_z$ . Therefore, the goal is

to obtain an explicit form for  $U(T)$  as a QAOA sequence. To do this, we consider a series of approximations. First, we remove time ordering by discretizing  $t : 0 \rightarrow T$  into a total of  $m$  steps with  $\Delta t = \frac{T}{m}$ . Eq. (3.25) then can be approximated as:

$$\begin{aligned} U'(T) &= \prod_{k=1}^m \exp \{-iH(k\Delta t) \Delta t\}, \\ &= \prod_{k=1}^m \exp \left\{ -iH'_k \left( \frac{T}{m} \right) \right\}. \end{aligned} \quad (3.26)$$

Here,

$$H'_k = \left( 1 - \frac{k}{m} \right) H_x + \frac{k}{m} H_z. \quad (3.27)$$

To estimate the errors in approximating  $U(T)$  via  $U'(T)$ , we use the following lemma from [67]:

**Lemma 3.1.** *Let  $H(t)$  and  $H'(t)$  be two time-dependent Hamiltonians for  $0 \leq t \leq T$ , and let  $U(T)$  and  $U'(T)$  be the respective unitaries that they induce. If the difference between the Hamiltonians is limited by  $\|H(t) - H'(t)\|_2 \leq \delta \forall t$ , then the distance between the induced transformations is bounded by  $\|U(T) - U'(T)\|_2 \leq \sqrt{2T\delta}$ .*

Since in our case,  $\|H(t) - H'_{j(t)}\|_2$ , where  $j(t) = \left\lceil \frac{tm}{T} \right\rceil$  is equal to  $\frac{1}{m} \|H_z - H_x\|_2 \in \mathcal{O} \left( \frac{\text{poly } n}{m} \right)$ , the errors caused in the discretization step is bounded by  $\mathcal{O} \left( \sqrt{\frac{T \cdot \text{poly } n}{m}} \right)$ .

Now we will consider on top of the discretization, a second approximation for  $U'(T)$ ,  $U''(T)$  by using the Baker-Campbell-Hausdorff [84] formula which induces errors as:

$$\| \exp(A + B) - \exp A \cdot \exp B \|_2 \in \mathcal{O} (\|AB\|_2). \quad (3.28)$$

Applying this formula to Eq. (3.26) and simplifying we get,

$$U''(T) = \prod_{k=1}^m \exp \left\{ -i \frac{T}{m} \left( 1 - \frac{k}{m} \right) H_x \right\} \cdot \exp \left\{ -i \frac{T}{m} \left( \frac{k}{m} \right) H_z \right\}. \quad (3.29)$$

The errors caused in using the Baker-Campbell-Hausdorff formula in this specific

case are bounded by the following,

$$\|U'(T) - U''(T)\|_2 \in \mathcal{O}\left(\frac{T^2}{m} \text{poly } n\right). \quad (3.30)$$

Since this error is larger than what we get from discretization, the total error in approximating  $U(T)$  is given by:

$$\|U(T) - U''(T)\|_2 \in \mathcal{O}\left(\frac{T^2}{m} \text{poly } n\right). \quad (3.31)$$

Therefore, setting  $m \sim \mathcal{O}(T^2 \text{poly } n)$ , one can bound the errors and recover the adiabatic evolution given by  $U(T)$ . Note that immediately from Eq. (3.29), one can see that the explicit form resembles that of QAOA ansatz as in Eq. (3.9) with QAOA depth  $p = m$ ,  $\beta_k = \frac{T}{m} \left(1 - \frac{k}{m}\right)$  and  $\gamma_k = \frac{T}{m} \left(\frac{k}{m}\right)$ .

## 3.6 Variational Grover search

Here we recover Grover's search algorithm in the variational approach. Grover's search relates to searching an index from an unstructured database of size  $N = 2^n$  [15].

**Definition 3.6** (Unstructured search). *Consider an unstructured database  $S$  indexed by  $\{0, 1\}^{\times n}$ . Let  $f : \{0, 1\}^{\times n} \rightarrow \{0, 1\}$  be a Boolean function (aka. black box) such that:*

$$f(j) = \begin{cases} 1 & \text{iff } j = \omega \\ 0 & \text{otherwise.} \end{cases} \quad (3.32)$$

Find  $\omega \in \{0, 1\}^{\times n}$ .

Consider this problem in the variational approach. To begin with, we shall construct a problem Hamiltonian for the search problem as defined in Def. 3.6.

**Remark 3.3.** *Let  $H_z = \mathbb{1} - |\omega\rangle\langle\omega|$  for some  $\omega \in \{0, 1\}^{\times n}$ . The expected value of  $H_z$  is minimized by  $|\omega\rangle$ . For any other state  $|\psi\rangle \neq \omega$ ,  $\langle\psi|H_z|\psi\rangle = 1 - |\langle\psi|\omega\rangle|^2 > 0$ .*

To construct the ansatz state, we choose  $H_2 = (|+\rangle\langle+|)^{\otimes n}$ ,  $H_1 = |\omega\rangle\langle\omega|$  and the initial state  $|\phi\rangle = |+\rangle^{\otimes n}$ . Therefore for a  $p$ -depth quantum circuit, we create the variational ansatz as:

$$|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = \prod_{k=1}^p \exp\{-i\beta_k H_2\} \cdot \exp\{-i\gamma_k H_1\} |+\rangle^{\otimes n}, \quad (3.33)$$

and then perform  $\min_{\boldsymbol{\gamma}, \boldsymbol{\beta}} \langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | H_z | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle$ .

Since both the mixer and the problem Hamiltonians  $H_2$  and  $H_1$  are projectors, we can simplify the unitaries in Eq. (3.33) as:

$$\begin{aligned} \exp\{-i\gamma_k H_1\} &= \exp\{-i\gamma_k |\omega\rangle\langle\omega|\} \\ &= \mathbb{1} + (e^{-i\gamma_k} - 1) |\omega\rangle\langle\omega|, \\ \exp\{-i\beta_k H_2\} &= \exp\{-i\beta_k (|+\rangle\langle+|)^{\otimes n}\} \\ &= \mathbb{1} + (e^{-i\beta_k} - 1) (|+\rangle\langle+|)^{\otimes n}. \end{aligned} \quad (3.34)$$

Now let us look at the transformation of the initial state  $|+\rangle^{\otimes n}$  by one layer of the quantum circuit.

$$|\psi(\gamma, \beta)\rangle = (\mathbb{1} + (e^{-i\beta} - 1) (|+\rangle\langle+|)^{\otimes n}) \cdot (\mathbb{1} + (e^{-i\gamma} - 1) |\omega\rangle\langle\omega|) |+\rangle^{\otimes n}, \quad (3.35)$$

which simplifies as:

$$|\psi(\gamma, \beta)\rangle = A |\omega\rangle + \frac{B}{\sqrt{N-1}} \sum_{y \neq \omega} |y\rangle, \quad (3.36)$$

for appropriate coefficients  $A, B$ . Since the input state  $|+\rangle^{\otimes n}$  also admits a form similar to Eq. (3.36), a two dimensional representation for this unitary transformation can be constructed as:

$$|\psi(\gamma, \beta)\rangle = M(\gamma, \beta) \begin{bmatrix} A_0 \\ B_0 \end{bmatrix}, \quad (3.37)$$

where,  $A_0$  and  $B_0$  are the coefficients of  $|+\rangle^{\otimes n}$  written in the form of Eq. (3.36) and

$$M(\gamma, \beta) = \begin{bmatrix} e^{-i\gamma} \left( 1 + \frac{(e^{-i\beta} - 1)}{N} \right), & (e^{-i\beta} - 1) \frac{\sqrt{N-1}}{N} \\ e^{-i\gamma} (e^{-i\beta} - 1) \frac{\sqrt{N-1}}{N}, & 1 + (e^{-i\beta} - 1) \frac{N-1}{N} \end{bmatrix}. \quad (3.38)$$

In general, one can write a  $p$ -depth ansatz state as:

$$|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = A_p |\omega\rangle + B_p \frac{1}{\sqrt{N-1}} \sum_{y \neq \omega} |y\rangle, \quad (3.39)$$

where the amplitudes of one step can be related to the amplitudes of the next step via the recursive application of  $M$ .

Consider a single layer of the ansatz as in Eq. (3.37), minimizing  $H_z$  corresponds to maximizing overlap with the  $|\omega\rangle$  as:

$$\min_{\gamma, \beta} \langle \psi(\gamma, \beta) | H_z | \psi(\gamma, \beta) \rangle = \max_{\gamma, \beta} |\langle \omega | \psi(\gamma, \beta) \rangle|^2. \quad (3.40)$$

The explicit form of this inner product is given by:

$$\langle \omega | \psi(\gamma, \beta) \rangle = e^{-i\gamma} \left( 1 + \frac{(e^{-i\beta} - 1)}{N} \right) A_0 + (e^{-i\beta} - 1) \frac{\sqrt{N-1}}{N} B_0. \quad (3.41)$$

Note that by taking the absolute value, and since  $A_0, B_0 \in \mathbb{R}_+$ , we obtain,

$$|\langle \omega | \psi(\gamma, \beta) \rangle| \leq \left| \left( 1 + \frac{(e^{-i\beta} - 1)}{N} \right) \right| \cdot A_0 + \left| (e^{-i\beta} - 1) \right| \cdot \frac{\sqrt{N-1}}{N} B_0, \quad (3.42)$$

which is independent of  $\gamma$ . Therefore, by taking maximization over  $\beta$  we obtain,

$$\begin{aligned} \max_{\beta} |\langle \omega | \psi(\gamma, \beta) \rangle| &\leq \max_{\beta} \left| \left( 1 + \frac{(e^{-i\beta} - 1)}{N} \right) \right| \cdot A_0 \\ &+ \max_{\beta} \left| (e^{-i\beta} - 1) \right| \cdot \frac{\sqrt{N-1}}{N} B_0. \end{aligned} \quad (3.43)$$

Since the second term in Eq. (3.43) is weighted with a factor of  $\sqrt{N-1}$  which is

dominant in terms of maximization,

$$\arg \max_{\beta} |\langle \omega | \psi(\gamma, \beta) \rangle| = \pi. \quad (3.44)$$

Substituting  $\beta = \pi$  in Eq. (3.41), we immediately recover,

$$\arg \max_{\gamma} |\langle \omega | \psi(\gamma, \beta) \rangle| = \pi. \quad (3.45)$$

In terms of optimal unitaries, we obtain  $U_1 = \exp \{-i\pi H_1\} = \mathbb{1} - 2|\omega\rangle\langle\omega|$  and  $U_2 = \exp \{-i\pi H_2\} = \mathbb{1} - 2 \cdot (|+\rangle\langle+|)^{\otimes n}$ . Note that these are exactly Grover's original search and diffusion operators [15], recovered in the variational approach.

To recover Grover's speedup, we set  $\gamma_k = \beta_k = \pi \forall k \in [1, p]$ . Then the matrix  $M$  becomes,

$$M = \begin{bmatrix} \frac{2-N}{N} & -\frac{2\sqrt{N-1}}{N} \\ \frac{2\sqrt{N-1}}{N} & \frac{2-N}{N} \end{bmatrix}, \quad (3.46)$$

This can be written as a rotation with  $\cos(\theta) = \frac{2-N}{N}$ , and  $\sin(\theta) = \frac{2\sqrt{N-1}}{N}$ . The initial state  $|+\rangle^{\otimes n}$  can also be written in terms of  $\theta$  as:

$$|+\rangle^{\otimes n} = \sin(\theta/2) |\omega\rangle + \cos(\theta/2) |Y\rangle, \quad (3.47)$$

where  $|Y\rangle$  is the unit norm state,  $|Y\rangle = \frac{1}{\sqrt{N-1}} \sum_{y \neq \omega} |y\rangle$ . After  $p$  iterations the state becomes:

$$|\psi_p\rangle = \sin((p+1/2)\theta) |\omega\rangle + \cos((p+1/2)\theta) |Y\rangle, \quad (3.48)$$

and the probability of measuring  $|\omega\rangle$  is

$$|\langle \omega | \psi_p \rangle|^2 = \sin^2((1/2 + p)\theta), \quad (3.49)$$

which is maximal for

$$\sin^2((1/2 + p)\theta) = 1 \quad (3.50)$$

$$(1/2 + p)\theta = \pi/2 \quad (3.51)$$

$$p = \frac{\pi}{2\theta} - \frac{1}{2} \quad (3.52)$$

$$p = \frac{\pi}{4 \arcsin\left(\frac{1}{\sqrt{N}}\right)} - \frac{1}{2}. \quad (3.53)$$

For  $N \gg 1$ ,  $\arcsin\left(\frac{1}{\sqrt{N}}\right) \approx \frac{1}{\sqrt{N}}$ , thus (3.53) becomes  $p \approx \frac{\pi}{4}\sqrt{N}$ , recovering Grover's scaling for circuit depth.

### 3.7 Discussion

In this chapter, we introduced a variational approach to Hamiltonian minimization and explored the connection between adiabatic quantum computation and QAOA. A point of discussion is that, we assumed the adiabatic time  $T$ , is known. In general,  $T$  depends on the minimum gap of the parameterized Hamiltonian  $H(t)$ , and one of the major drawbacks of adiabatic quantum computation is that this gap can vanish exponentially in the number of qubits [85]. Although, in principle, QAOA can recover adiabaticity with  $p \rightarrow \infty$ , the hope is that QAOA can make use of non-adiabatic pathways to minimize  $H_z$ .

Yet another area of rapid development is in ansatz design. Specifically, in QAOA, ansatz are constructed by an alternating sequence of two non-commuting propagators. Recent approaches such as in [30], the specific choice of the mixing Hamiltonian is motivated by leveraging additional information such as symmetries of the ground states and/or hard constraints that are imposed by the specific optimization problem. In such cases, the generated variational state space only consists of feasible states. Therefore, faster convergence may be expected if the QAOA mixers are designed specifically with respect to the optimization problem.

# Chapter 4

## Underparameterization effects in QAOA

In this chapter, we introduce and investigate the limiting performance for QAOA circuits when minimizing random instances of MAX-SAT as well as optimization problems on graphs. In both cases, our results indicate problem density-induced underparameterization in QAOA circuits, an effect which we call *reachability deficits* [1]. The chapter first defines reachability deficits before moving onto describing the numerical study done to assess performance of fixed-depth QAOA circuits applied to MAX-SAT instances where reachability deficits were first discovered. We next analytically show the existence of this effect for  $p = 1$  QAOA, considering instances of MAX-CUT on triangle-free  $d$  regular graphs. To further motivate our results, the chapter also includes numerical demonstrations of reachability deficits for random graph optimization problems [2].

### 4.1 Background

QAOA is an instance of the class of variational algorithms, specifically designed to solve instances of combinatorial optimization problems [26]. The algorithm has been shown to approximate solutions to problems such as MAX-CUT [40], perform Grover's search algorithm [24, 86], and execute optimal protocols for quantum control [87]. Recent milestones include: experimental demonstration of  $3p$ -QAOA using

twenty-three qubits [36], universality results [27, 28], as well as several results that aid and improve the original implementation of the algorithm [29–31]. Owing to this rapid development, QAOA has become a cornerstone of contemporary quantum algorithms development with growing range of applications. Despite these successes, the algorithm indeed has limitations.

Recent findings suggest that randomly parameterized quantum circuits suffer from *barren plateaus* resulting in an exponentially low probability of finding correct solutions [88, 89]. This finding applies to variational algorithms in general, and therefore, one might expect them in QAOA as well. Another potential weakness of the variational approach and, by extension, to QAOA, is the classical optimization step. Specific to QAOA, recent results show that classical algorithms in certain restrictive settings can outperform depth 1 QAOA [90], forcing the need for employing higher depth versions. Immediately, such a requirement comes at the cost of classically optimizing several parameters simultaneously, and therefore, quickly becomes cumbersome beyond low depth. Such optimizations alone may themselves be NP-Hard computationally [91]. Furthermore, due to this optimization step, analytical approaches to understand performance guarantees only describe QAOA in restrictive settings such as: limited to specific instances e.g. fully connected graphs and/or single depth [33–35]. Despite these challenges, provable advantages such as recovering a near-optimal query complexity in Grover’s search [86] and pathways towards quantum advantage [32] have been reported, yet an understanding of the ultimate limitations of the algorithm remains largely open.

## 4.2 Reachability deficits

In the variational approach, a parameterized quantum circuit (or variational circuit) of some depth  $p$ , can prepare states from a variational state space,

$$\Omega_p = \bigcup_{\theta} \{ |\psi(\theta)\rangle \}, \quad (4.1)$$

with  $|\psi(\boldsymbol{\theta})\rangle$  given by Eq. (3.1). Note that this variational state space for low depths  $p$ , may not cover the whole Hilbert space, and therefore minimization over  $\Omega$  does not guarantee convergence to the exact ground state energy.

**Definition 4.1** (Underparameterization). *Let  $\Omega_p$  represent the variational state space accessible for a parameterized quantum circuit of depth  $p$ . Given some problem Hamiltonian to be minimized  $H$ , the circuit is underparameterized, whenever:*

$$\min_{\psi \in \Omega} \langle \psi(\boldsymbol{\theta}) | H | \psi(\boldsymbol{\theta}) \rangle - \min_{\phi \in \mathcal{H}} \langle \phi | H | \phi \rangle > 0 \implies \Omega_p \subset \mathcal{H}. \quad (4.2)$$

Based on our empirical finding, we observe that for QAOA circuits of some fixed depth  $p$ , and for problem Hamiltonians of density  $\alpha$ , the accessible variational state space shows a nontrivial dependence on problem density  $\alpha$ , causing a density-induced underparameterization. This effect is termed *reachability deficit* [1].

**Definition 4.2** (Reachability deficits). *Given some problem Hamiltonian  $H(\alpha)$ , on  $n$  qubits and of density  $\alpha$ , let  $\Omega$  represent the variational state space accessible for a  $p$ -depth QAOA circuit:*

$$\Omega = \bigcup_{\boldsymbol{\gamma}, \boldsymbol{\beta}} \{ |\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle \}, \quad (4.3)$$

where  $|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle$  as in Eq. (3.4). *Reachability deficits are  $\alpha$  induced underparameterization:*

$$\min_{\psi \in \Omega} \langle \psi | H(\alpha) | \psi \rangle - \min_{\phi \in \mathcal{H}} \langle \phi | H(\alpha) | \phi \rangle > 0. \quad (4.4)$$

### 4.3 Performance of QAOA applied to MAX-SAT

To study the performance of QAOA on Boolean Satisfiability, we select 2 and 3-SAT instances as our testbed. The choice of studying these two classes separately is motivated by the fact that any  $k$ -SAT instance can be Karp reduced to a 3-SAT instance with polynomial overheads, while reduction to 2-SAT requires exponential overheads in auxiliary variables. The classical complexity class for the decision versions of these two problems are also different, while 2-SAT is in P, 3-SAT is NP-Complete. Although QAOA aims to approximate solutions to the optimization versions of these

problems (MAX-SAT), which is NP-Hard for both cases, MAX-2-SAT only requires 2-local terms in the construction of its equivalent problem Hamiltonian whereas MAX-3-SAT necessitates the use of 3-local terms. It is worthwhile to study whether the algorithmic behavior of QAOA is sensitive to the locality of the problem Hamiltonian. One would expect such a difference, as the variational state space (as in Eq. (3.2)) generated in each of these cases may be different.

Furthermore, both problems feature a computational phase transition at clause density  $\alpha_c \sim 1$ , with an Easy-Hard pattern, and it is well known that classical SAT solvers require increasing resources when solving instances beyond  $\alpha_c \sim 1$  [61]. Though witnessed in classical algorithms, the signature of such phase transitions in quantum algorithms remains an open area of interest.

### 4.3.1 Numerical details

Here, we briefly go through the numerical details of the study. For an elaborate description on the software code emulating QAOA and the optimization heuristics used, we ask the reader to refer to the appendix.

1. To study the performance of QAOA, we initially generate random instances of 2 and 3-SAT according to Eq. (2.7). The problem Hamiltonians are then constructed according to the embedding schemes as described in Section 2.4.1. These Hamiltonians have the form:

$$H_{2\text{SAT}} = \sum_{j < k} J_{jk} Z_j Z_k + \sum_j h_j Z_j, \quad (4.5)$$

and

$$H_{3\text{SAT}} = \sum_{j < k < l} I_{jkl} Z_j Z_k Z_l + \sum_{j < k} J_{jk} Z_j Z_k + \sum_j h_j Z_j. \quad (4.6)$$

2. In generating instances, we fix the number of variables to  $n = 10$  and increase the clause density  $\alpha$  by  $\frac{1}{10}$  for the range  $\alpha \in [0.1, 4.0]$ .

3. To study average performance, we use the error in approximation,  $f$ , as defined in Def. 3.10 and consider a statistic of 300 random instances for densities  $\alpha \in [0.1, 4.0]$  for 2-SAT as well as 3-SAT.
4. Corresponding to each instance, a problem Hamiltonian  $H_{\text{SAT}}$ , is constructed and a standard QAOA circuit with depths  $p \in \{4, 8, 12\}$  is initialized with random initial parameters  $\boldsymbol{\gamma}$  and  $\boldsymbol{\beta}$  as:

$$|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = \prod_{k=1}^p \exp\{-i\beta_k H_x\} \cdot \exp\{-i\gamma_k H_{\text{SAT}}\} |+\rangle^{\otimes n}, \quad (4.7)$$

where we use the standard driver and ranges,  $H_x = \sum_{j=1}^n X_j$ ,  $\gamma_k \in [0, 2\pi)$  and  $\beta_k \in [0, \pi]$ .

5. To find optimal angles,

$$(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*) \in \arg \min_{(\boldsymbol{\gamma}, \boldsymbol{\beta})} \langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | H_{\text{SAT}} | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle, \quad (4.8)$$

we use a heuristic optimization strategy motivated by layer-wise training [92] and choose the Limited Broyden–Fletcher–Goldfarb–Shanno method (L-BFGS-B) [93] to find optimal parameters during each step of the heuristic strategy.

6. Once the optimization terminates, the error in the approximation is calculated, and the entire process is repeated for the next instance in the statistic.

### 4.3.2 Empirical observations

Here, we compile a list of empirical observations based on our numerical results (see Fig. 4-1 and Fig. 4-2).

1. The average performance (quantified by the error in approximation) for fixed depth QAOA becomes increasingly worse for clause densities  $\alpha > \alpha_c$  for both MAX-2-SAT and MAX-3-SAT.
2. Though increasing depth versions achieve better performance, the trend of worsening approximations for clause densities  $\alpha > \alpha_c$  is still reflected.

3. For low density instances,  $\alpha < \alpha_c$ ,  $p = 4$  QAOA exactly recovers the ground states for both MAX-2-SAT and MAX-3-SAT
4. The exhibited point for  $\alpha_c$ , based on our numerical investigation, coincides with the algorithmic phase transition point,  $\alpha_c \sim 1$ , for both MAX-2-SAT and MAX-3-SAT.

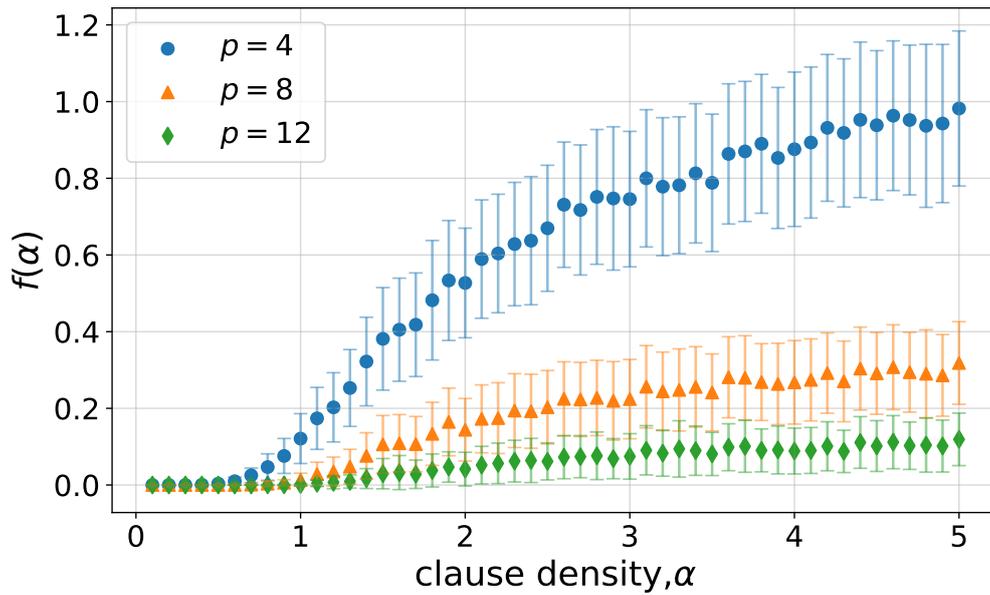


Figure 4-1:  $f(\alpha)$  as defined in Def. 3.10 vs clause density for 2-SAT for differing QAOA depths. Data points show the average value obtained over 300 randomly generated instances for  $n = 10$  with error bars indicating the standard deviation. Plots also show improved performance for higher depths. Figure reprocessed from [1].

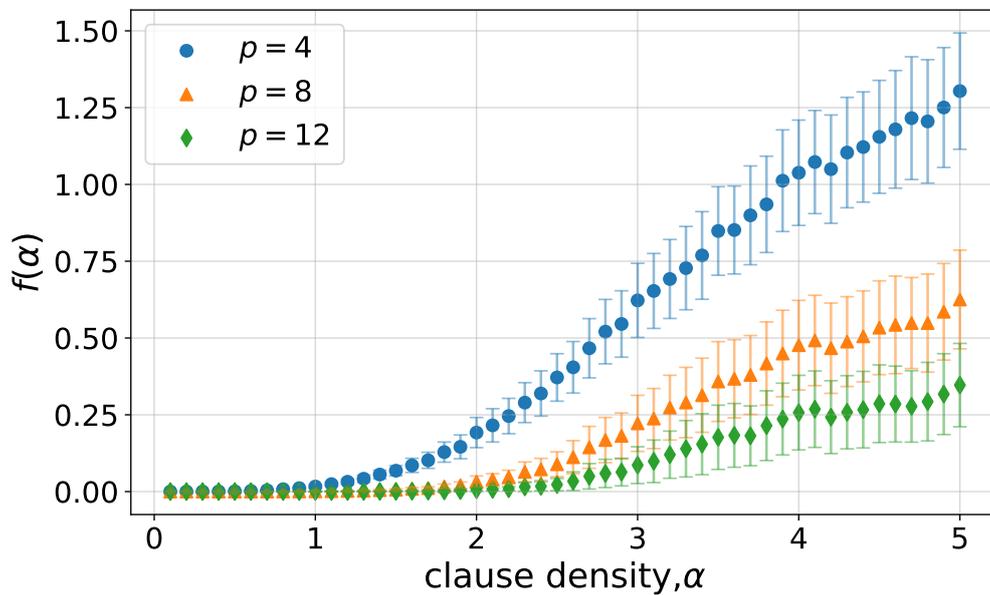


Figure 4-2:  $f(\alpha)$  as defined in Def. 3.10 vs clause density for 3-SAT for differing QAOA depths. Data points show the average value obtained over 300 randomly generated instances for  $n = 10$  with error bars indicating the standard deviation. Plots also show improved performance for higher depths. Figure reprocessed from [1].

## 4.4 Reachability deficits in triangle free regular graphs

In this section, we analytically demonstrate the existence of reachability deficits for the case of  $p = 1$  QAOA when applied to the problem of MAX-CUT on triangle-free  $d$  regular graphs. To show the existence, we first derive an explicit formula for the expected value of the MAX-CUT Hamiltonian [40], which we can analytically optimize to recover the optimal expected value. On the basis of this expression, we prove the existence of the effect.

Consider a graph  $G(V, E)$  with a set of vertices  $V$  ( $|V| = n$ ) and edges  $E$ . Maximizing the cut value of  $G$  corresponds to minimizing an Ising Hamiltonian of the form:

$$H = \sum_{\langle j,k \rangle=1}^{|E|} Z_j Z_k. \quad (4.9)$$

For  $p = 1$  QAOA, the generated ansatz takes the form:

$$\begin{aligned} |\psi(\gamma, \beta)\rangle &= \exp(-i\beta_k H_x) \cdot \exp(-i\gamma_k H) |+\rangle^{\otimes n}, \\ &= U_\gamma \cdot U_\beta |+\rangle^{\otimes n}, \end{aligned} \quad (4.10)$$

where  $H_x = \sum_{k=1}^n X_k$ . The expected value of the MAX-CUT Hamiltonian with respect to the ansatz state in Eq. (4.10), is then given by:

$$\begin{aligned} \langle \psi(\gamma, \beta) | H | \psi(\gamma, \beta) \rangle &= \sum_{\langle j,k \rangle=1}^{|E|} \langle \psi(\gamma, \beta) | Z_j Z_k | \psi(\gamma, \beta) \rangle \\ &= \sum_{\langle j,k \rangle=1}^{|E|} F_{j,k}(\gamma, \beta). \end{aligned} \quad (4.11)$$

Consider a single term in the expansion of Eq. (4.11) and dropping the  $\otimes n$  notation,

$$F_{j,k}(\gamma, \beta) = \langle + | U_\gamma^\dagger \cdot U_\beta^\dagger \cdot Z_j Z_k \cdot U_\beta \cdot U_\gamma | + \rangle. \quad (4.12)$$

Note that in the expression above, terms in  $U_\gamma$  and  $U_\beta$  that do not have support over qubit indices  $j$  or  $k$ , commute and cancel off. Therefore, when calculating the expected value over the edge  $\langle j, k \rangle$ , only nodes connected to qubits  $j$  or  $k$  or both

need to be accounted for. Thus, for  $p = 1$  QAOA, the expected value over an edge  $\langle j, k \rangle$  only requires knowledge of the subgraph structure generated by nodes one edge distance away from  $j$  and  $k$ . This reasoning can be generalized for arbitrary depth  $p$ , where contributing subgraphs are generated by including nodes that are  $p$ -distance away from  $j$  and  $k$ . Corresponding to different edges in  $G$ , different subgraph types would be generated, enumerating all such subgraph types then allows for an easy calculation of the expected value in Eq. (4.11) as:

$$\langle \psi(\gamma, \beta) | H | \psi(\gamma, \beta) \rangle = \sum_g w(g) F_g(\gamma, \beta), \quad (4.13)$$

where the sum is taken over subgraph types and  $w(g)$  corresponds to the frequency of each type  $g$ .

Here, our aim is to generate an explicit formula for the expectation value in Eq. (4.12) in terms of the parameters  $\gamma$  and  $\beta$ . For simplicity, we drop the indices on  $F$  in Eq. (4.12),

$$F(\gamma, \beta) = \langle + | U_\gamma^\dagger \cdot U_\beta^\dagger \cdot Z_j Z_k \cdot U_\beta \cdot U_\gamma | + \rangle. \quad (4.14)$$

To proceed, we make use of the following identities:

$$U_\beta^\dagger \cdot Z_j \cdot U_\beta = \cos(2\beta) \cdot Z_j + \sin(2\beta) \cdot Y_j. \quad (\text{i1})$$

$$U_\beta^\dagger \cdot Y_j \cdot U_\beta = \cos(2\beta) \cdot Y_j - \sin(2\beta) \cdot Z_j. \quad (\text{i2})$$

$$e^{i\gamma Z_j \cdot Z_m} \cdot Y_j \cdot e^{-i\gamma Z_j \cdot Z_m} = \cos(2\gamma) \cdot Y_j + \sin(2\gamma) \cdot X_j Z_m. \quad (\text{i3})$$

$$e^{i\gamma Z_j \cdot Z_m} \cdot X_j \cdot e^{-i\gamma Z_j \cdot Z_m} = \cos(2\gamma) \cdot X_j - \sin(2\gamma) \cdot Y_j Z_m. \quad (\text{i4})$$

Consider some Hamiltonian  $H_j$  of the form:

$$H_j = Z_j \otimes \sum_{m=1}^d Z_m, \quad (4.15)$$

by repeatedly applying Eqs. (i3), (i4), one can obtain an expansion in terms of Pauli strings:

$$e^{i\gamma H_j} \cdot Y_j \cdot e^{-i\gamma H_j} = \sum_{k=0}^d (-1)^{\left(\lceil \frac{k}{2} \rceil + 1\right)} (\cos(2\gamma))^{d-k} \cdot (\sin(2\gamma))^k \cdot X_j^k \cdot Y_j^{k+1} \{Z_m\}^k, \quad (\text{i5})$$

where we use the notation to denote  $k$ -local  $Z$ -Pauli strings,

$$\{Z_m\}^k = \sum_{i_1 < i_2 < \dots < i_k} Z_{i_1} \cdot Z_{i_2} \cdot \dots \cdot Z_{i_k}. \quad (4.16)$$

Similarly,

$$e^{i\gamma H_j} \cdot X_j \cdot e^{-i\gamma H_j} = \sum_{k=0}^d (-1)^{\lceil \frac{k}{2} \rceil} (\cos(2\gamma))^{d-k} \cdot (\sin(2\gamma))^k \cdot X_j^{k+1} \cdot Y_j^k \{Z_m\}^k. \quad (\text{i6})$$

From these identities one can see that under the transformations by the unitaries  $U_\gamma$  and  $U_\beta$ , the operator  $Z_j Z_k$  in Eq. (4.14) turns into a sum of Pauli strings over which the expectation value with respect to  $|+\rangle$  can be taken. Next, we do this systematically.

Consider the unitary transformation by  $U_\beta$ ,

$$\begin{aligned} U_\beta^\dagger \cdot Z_j Z_k \cdot U_\beta &= \cos^2(2\beta) \cdot Z_j Z_k + \sin^2(2\beta) \cdot Y_j Y_k \\ &\quad + \cos(2\beta) \sin(2\beta) [Z_j Y_k + Z_k Y_j]. \end{aligned} \quad (4.17)$$

Next, we apply the unitary transformation by  $U_\gamma$  to Eq. (4.17). Note that  $U_\gamma = \tilde{U}_\gamma e^{-i\gamma Z_j \cdot Z_k}$ , therefore, by using Eqs. (i3) and (i4) we have,

$$\begin{aligned}
 U_\gamma^\dagger \cdot U_\beta^\dagger \cdot Z_j Z_k \cdot U_\beta \cdot U_\gamma &= \cos^2(2\beta) \cdot Z_j Z_k + \sin^2(2\beta) \cdot \tilde{U}_\gamma^\dagger Y_j Y_k \tilde{U}_\gamma + \\
 &\quad \cos(2\beta) \sin(2\beta) \times \\
 &\quad \left[ \cos(2\gamma) \left( Z_j \otimes \tilde{U}_\gamma^\dagger Y_k \tilde{U}_\gamma \right) + \sin(2\gamma) \tilde{U}_\gamma^\dagger X_k \tilde{U}_\gamma \right. \\
 &\quad \left. + \cos(2\gamma) \left( Z_k \otimes \tilde{U}_\gamma^\dagger Y_j \tilde{U}_\gamma \right) + \sin(2\gamma) \tilde{U}_\gamma^\dagger X_j \tilde{U}_\gamma \right].
 \end{aligned} \tag{4.18}$$

Before simplifying this expression further, we take the expected value with respect to  $|+\rangle$ . Note that only Pauli strings containing pure  $X$  can contribute to the expected values since  $\langle +|Z|+\rangle = \langle +|Y|+\rangle = 0$  and  $\langle +|X|+\rangle = 1$ . Taking these into account, we have

$$\begin{aligned}
 F(\gamma, \beta) &= \sin^2(2\beta) \langle +| \tilde{U}_\gamma^\dagger Y_j Y_k \tilde{U}_\gamma |+\rangle + \\
 &\quad \cos(2\beta) \sin(2\beta) \sin(2\gamma) \times \\
 &\quad \left[ \langle +| \tilde{U}_\gamma^\dagger X_j \tilde{U}_\gamma |+\rangle + \langle +| \tilde{U}_\gamma^\dagger X_k \tilde{U}_\gamma |+\rangle \right].
 \end{aligned} \tag{4.19}$$

First, we evaluate  $\langle +| \tilde{U}_\gamma^\dagger X_j \tilde{U}_\gamma |+\rangle$ . Note that only terms that have support on qubit  $j$  can contribute, and the rest of the terms commute and cancel. Therefore, we can simplify  $\tilde{U}_\gamma = e^{-i\gamma H_j}$ , where  $H_j = Z_j \otimes \sum_{m=1}^{d_j-1} Z_m$ . Where  $d_j$  is the degree of the node  $j$  in the graph  $G$ . Now we can apply Eq. (i6) directly and see that the only term that can contribute is when  $l = 0$  in Eq. (i6). Therefore,

$$\langle +| \tilde{U}_\gamma^\dagger X_j \tilde{U}_\gamma |+\rangle = (\cos(2\gamma))^{d_j-1}, \tag{4.20}$$

and similarly,

$$\langle +| \tilde{U}_\gamma^\dagger X_k \tilde{U}_\gamma |+\rangle = (\cos(2\gamma))^{d_k-1}. \tag{4.21}$$

Now we evaluate  $\langle +| \tilde{U}_\gamma^\dagger Y_j Y_k \tilde{U}_\gamma |+\rangle$ . To do this we first factor  $\tilde{U}_\gamma = \hat{U} \cdot e^{-i\gamma H_{jk}}$ , where  $H_{jk}$  contains terms that have support on both qubits  $j$  and  $k$  and therefore

has the form,

$$\begin{aligned} H_{jk} &= Z_j \otimes \sum_{m=1}^t Z_m + Z_k \otimes \sum_{m=1}^t Z_m \\ &= (Z_j + Z_k) \otimes \hat{Z}_m, \end{aligned} \quad (4.22)$$

where  $t$  is the number of triangles formed by the edge  $\langle j, k \rangle$  in the corresponding  $p = 1$  subgraph. We make use of this factorization to simplify the calculation of  $\langle + | \tilde{U}_\gamma^\dagger Y_j Y_k \tilde{U}_\gamma | + \rangle$ . Notice that,

$$e^{i\gamma H_{jk}} Y_j Y_k e^{i\gamma H_{jk}} = \left( e^{i\gamma Z_j \hat{Z}_m} \cdot Y_j \cdot e^{-i\gamma Z_j \hat{Z}_m} \right) \otimes \left( e^{i\gamma Z_k \hat{Z}_m} \cdot Y_k \cdot e^{-i\gamma Z_k \hat{Z}_m} \right). \quad (4.23)$$

Therefore, we can use Eq. (i5) on both sides of the tensor product. By grouping Pauli strings free of  $Z$  terms we can get a simplified expression for Eq. (4.23) in the form:

$$\begin{aligned} e^{i\gamma H_{jk}} Y_j Y_k e^{i\gamma H_{jk}} &= \sum_{k=0}^t (\cos^2(2\gamma))^{t-k} \cdot (\sin^2(2\gamma))^k \times \\ &\quad (X_j X_k)^k \cdot (Y_j Y_k)^{k+1} + \hat{A}(Z_m), \end{aligned} \quad (4.24)$$

where  $\hat{A}(Z_m)$  is the operator composed of all other terms that contain free  $Z$  terms.

Now we can apply the rest of the unitary  $\hat{U}$ , to recover,

$$\begin{aligned} \langle + | \tilde{U}_\gamma^\dagger Y_j Y_k \tilde{U}_\gamma | + \rangle &= \sum_{k=0}^t \binom{t}{k} (\cos^2(2\gamma))^{t-k} \cdot (\sin^2(2\gamma))^k \times \\ &\quad \langle + | \hat{U}^\dagger (X_j X_k)^k \cdot (Y_j Y_k)^{k+1} \hat{U} | + \rangle + \\ &\quad \langle + | \hat{U}^\dagger \hat{A}(Z_m) \hat{U} | + \rangle. \end{aligned} \quad (4.25)$$

Note that  $\langle + | \hat{U}^\dagger \hat{A}(Z_m) \hat{U} | + \rangle = 0$  since the operator contains free  $Z$  terms and the only term that does not add  $Z$  terms when transformations under  $\hat{U}$  is when  $k = 0$

in Eqs. (i5) and (i6). Therefore, Eq. (4.25) further simplifies into,

$$\begin{aligned}
 \langle + | \tilde{U}_\gamma^\dagger Y_j Y_k \tilde{U}_\gamma | + \rangle &= (\cos(2\gamma))^{d_i+d_j-2(t+1)} \times \\
 &\sum_{k=0}^t \binom{t}{k} (\cos^2(2\gamma))^{t-k} \cdot (\sin^2(2\gamma))^k \times \\
 &(X_j X_k)^k \cdot (Y_j Y_k)^{k+1}, \\
 &= (\cos(2\gamma))^{d_i+d_j-2(t+1)} \sum_{k=\text{odd}}^t \binom{t}{k} (\cos^2(2\gamma))^{t-k} \cdot (\sin^2(2\gamma))^k \\
 &= \frac{1}{2} (\cos(2\gamma))^{d_i+d_j-2(t+1)} \cdot (1 - \cos^t(4\gamma)).
 \end{aligned} \tag{4.26}$$

Finally, substituting these into Eq. (4.14) we obtain the expression,

$$\begin{aligned}
 F(\gamma, \beta) &= \frac{1}{2} \left[ \sin^2(2\beta) \cdot (1 - \cos^t(4\gamma)) \cdot (\cos(2\gamma))^{d_i+d_j-2(t+1)} \right. \\
 &\quad \left. + \sin(4\beta) \cdot \sin(2\gamma) \left( (\cos(2\gamma))^{d_j-1} + (\cos(2\gamma))^{d_k-1} \right) \right].
 \end{aligned} \tag{4.27}$$

If we consider  $d$  regular, triangle-free graphs, then the full expression for the expected value of the MAX-CUT Hamiltonian can be calculated as:

$$\langle \psi(\gamma, \beta) | H | \psi(\gamma, \beta) \rangle = \frac{|E|}{2} (\cos(2\gamma))^{d-1} \cdot \sin(2\gamma) \cdot \sin(4\beta), \tag{4.28}$$

which has one set of optimal parameters as:

$$\arg \min_{(\gamma, \beta)} \langle \psi(\gamma, \beta) | H | \psi(\gamma, \beta) \rangle = \left( \frac{1}{2} \arctan \left( \frac{1}{\sqrt{d-1}} \right), \frac{\pi}{8} \right). \tag{4.29}$$

Substituting Eq. (4.29) in Eq. (4.28), we obtain the optimal value,

$$\langle \psi(\gamma^*, \beta^*) | H | \psi(\gamma^*, \beta^*) \rangle = \frac{|E|}{2} \frac{1}{\sqrt{d}} \left( \frac{d-1}{d} \right)^{\frac{d-1}{2}}. \tag{4.30}$$

Since in the limit  $d \rightarrow \infty$ ,  $\left(\frac{d-1}{d}\right)^{d-1} \rightarrow \frac{1}{e}$ ,

$$\langle \psi(\gamma^*, \beta^*) | H | \psi(\gamma^*, \beta^*) \rangle > \frac{|E|}{2} \frac{1}{\sqrt{d \cdot e}}. \quad (4.31)$$

For a triangle-free  $d$ -regular graph with  $|E|$  edges, the error in approximation becomes,

$$f(\alpha, n) = \alpha \cdot n \left(1 - \frac{1}{2\sqrt{d \cdot e}}\right) > 0, \quad (4.32)$$

thereby proving the existence of *reachability deficits* for  $p = 1$  QAOA circuits.

## 4.5 Reachability deficits in graph optimization

In this section, we numerically investigate the performance for QAOA on graph optimization problems. Since such optimization problem instances also embed into 2-local Ising Hamiltonians, one would expect the performance of a  $p$ -depth QAOA to be similar to what was observed for MAX-2-SAT. Instead of clause density as the order parameter of choice to study typical case performance, for graph instances we make use of the graph density i.e., the ratio of number of edges to node count in a given graph. Although there are several models to generate random instances, we use the Erdos-Renyi random graph model, where graphs with  $n$  nodes and  $m$  edges are sampled from a uniform distribution over all possible graphs with  $n$  nodes and  $m$  edges. In general, with appropriate edge weights and Ising spin encoding, minimization on graphs can represent solving a family of NP-Hard optimization problems such as MAX-CUT, Vertex Cover, and Maximum Independent Set, to name a few. In addition to studying QAOA performance on random graph instances, we also study QAOA on restrictive instances such as 3-regular graphs, planar grids, and non-planar fully connected graphs. The motivation for performing this restrictive study is that, unlike MAX-2-SAT instances which are hard to grasp in terms of their random problem structure, graph instances feature a certain level of ease in imposing structural constraints. These may include, bounded degree, topology (planar or non-planar graphs), and connectivity. Such constraints are practically important from a

hardware standpoint since current quantum processors have limited capabilities in realizing arbitrary instances [36].

### 4.5.1 Numerical details

Here, we briefly go through the numerical details of the study. For an elaborate description on the software code emulating QAOA and the optimization heuristics used, we ask the reader to refer to the appendix.

1. To study the performance of QAOA, we initially generate graph instances. Specifically, 4 families are considered:
  - (a)  $G_{n,m}$  for fixed  $n$ . Random instances are generated by initializing an empty graph on the vertex set of size  $|V| = n$ . Then,  $m$  edges are constructed in such a way that all possible  $\binom{n}{m}$  choices are equally likely with random weights drawn from  $\{-1, +1\}$ .
  - (b) 3-regular graphs. A 3-regular graph, also known as a cubic graph, is when each node in the graph has degree equal to 3. The density of the generated instance in this case is constant ( $density = 1.5$ ) since the number of edges required for constructing a 3-regular graph is given by  $3 \times \frac{n}{2}$ , for even  $n$ . Here, we also choose random weights drawn from  $\{-1, +1\}$
  - (c) 2-D grid graphs. These graphs are interesting mainly because of the inherent hardware connectivity of Google's Sycamore. Such graphs are generated by constructing a  $N \times N$  square grid and randomly deleting nodes on the boundaries to construct a random grid instance with  $n \leq N^2$  nodes. Unlike the case of regular graphs, 2-D grid graphs do not share a constant density across the generated instance. The maximum number of edges in a square/partially square grid is given by  $\lfloor 2n - 2\sqrt{n} \rfloor$ . Therefore, at most density of such instances is bounded above by 2.0 with random weights drawn from  $\{-1, +1\}$
  - (d) Fully connected. These graphs are highly non-planar and are constructed by considering all possible  $\binom{n}{2}$  edges with scaled random weights drawn

from  $\{-1, +1\}$  with a factor  $\frac{1}{\sqrt{n}}$ . These graphs have the maximum possible density of  $\frac{n-1}{2}$  and are commonly called instances of the Sherrington-Kirkpatrick model.

The problem Hamiltonians are then constructed according to the embedding schemes described in Section 2.4.3. These Hamiltonians have the form:

$$H = \sum_{\langle j,k \rangle \in E} w_{jk} Z_j Z_k. \quad (4.33)$$

2. We generate a statistic of 300 random instances for node sizes  $n = 10$  up to 20 for each of the graph families.
3. To study performance, we make use of the standard metrics described in Section 3.4.
4. Corresponding to each instance in the graph family, a problem Hamiltonian  $H$ , is constructed and a standard QAOA circuit with depths  $p = 3$  is initialized with random initial parameters  $\boldsymbol{\gamma}$  and  $\boldsymbol{\beta}$  as:

$$|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = \prod_{k=1}^p \exp\{-i\beta_k H_x\} \cdot \exp\{-i\gamma_k H\} |+\rangle^{\otimes n}, \quad (4.34)$$

where we use the standard driver and ranges,  $H_x = \sum_{j=1}^n X_j$ ,  $\gamma_k \in [0, 2\pi)$  and  $\beta_k \in [0, \pi)$ .

5. To find optimal angles,

$$(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*) \in \arg \min_{(\boldsymbol{\gamma}, \boldsymbol{\beta})} \langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | H | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle, \quad (4.35)$$

we use a heuristic optimization strategy motivated by layer-wise training and we choose the limited Broyden–Fletcher–Goldfarb–Shanno method (L-BFGS-B) to find optimal parameters during each step of the heuristic strategy.

6. Once the optimization terminates, performance metrics are calculated, and the whole process is repeated for the next instance in the statistic.

## 4.5.2 Empirical observations

Here we compile a list of empirical observations based on our numerical results (see Fig. 4-3 to Fig. 4-8 )

1. QAOA on random graph instances generated at varying densities exhibits reachability deficits.
2. The effect of reachability deficits is visually suppressed in the approximation ratio.
3.  $p = 3$  QAOA recovers exact solutions for random instances generated below the densities of 0.5.
4. Although increasing depths achieve better performance, approximations worsen for instances beyond density 0.5.
5. Absence of topology-dependent performance bias that can be attributed statistical relevance when comparing random graphs generated at approximately the same densities with the restrictive graph instances.

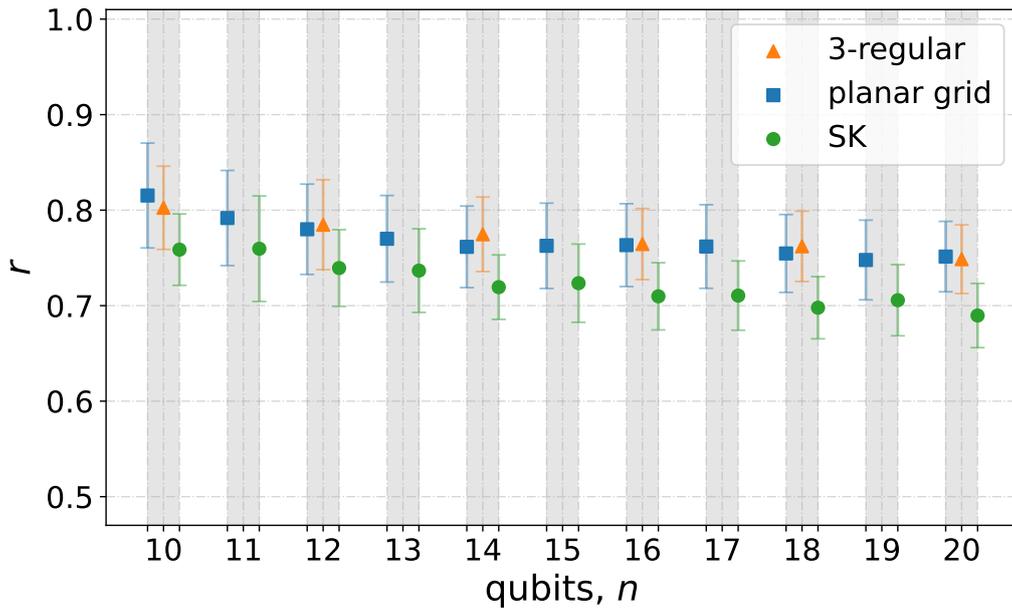


Figure 4-3: Average approximation ratios versus qubit count  $n$ , for the three families of graphs: (i) planar grid graphs (blue), (ii) 3-regular graphs (orange), and (iii) SK model or complete graphs (green). Each data point represents the average performance of depth  $p = 3$  QAOA over statistics of 300 randomly generated instances with error bars indicating the standard deviation. Figure reprocessed from [2].

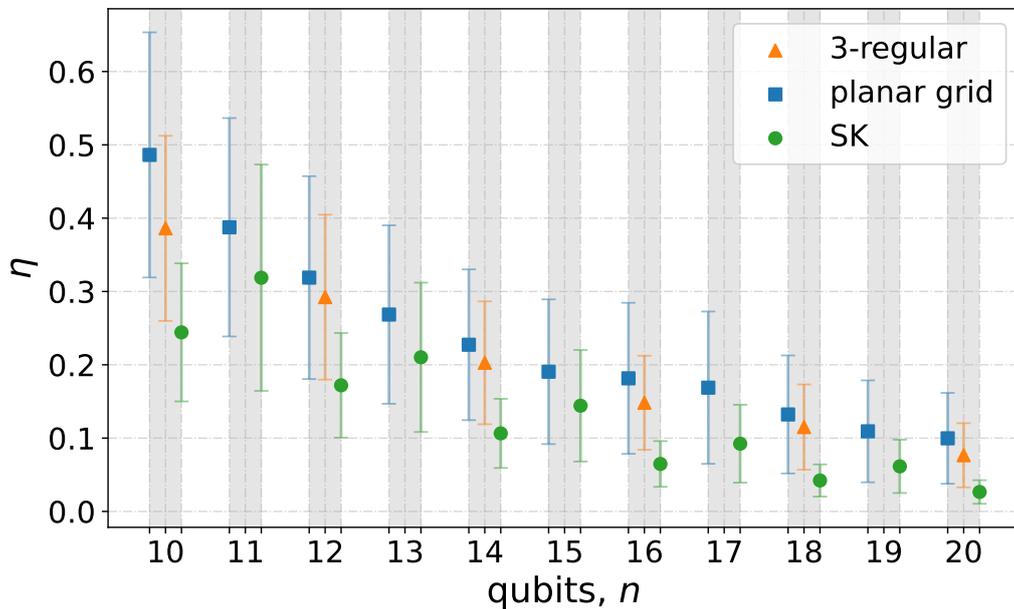


Figure 4-4: Success probability versus qubit count  $n$  for the three families of graphs: (i) planar grid graphs (blue), (ii) 3-regular graphs (orange), and (iii) SK model or complete graphs (green). Each data point represents the average performance of depth  $p = 3$  QAOA over statistics of 300 randomly generated instances with error bars indicating the standard deviation. Figure reprocessed from [2].

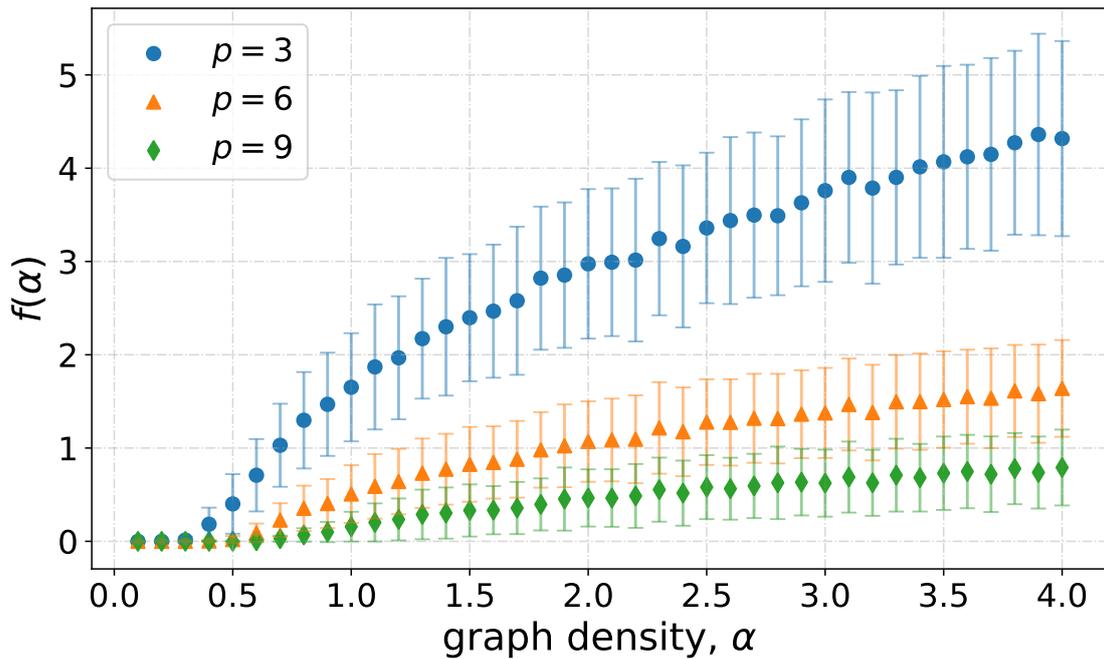


Figure 4-5: Error in approximations versus graph density for depth  $p = 3, 6, 9$  QAOA. Each data point represents the average performance over statistics of 300 uniform random graph instances,  $G_{n,m}$  with nodes  $n = 10$  and error bars indicating the standard deviation. Figure reprocessed from [2].

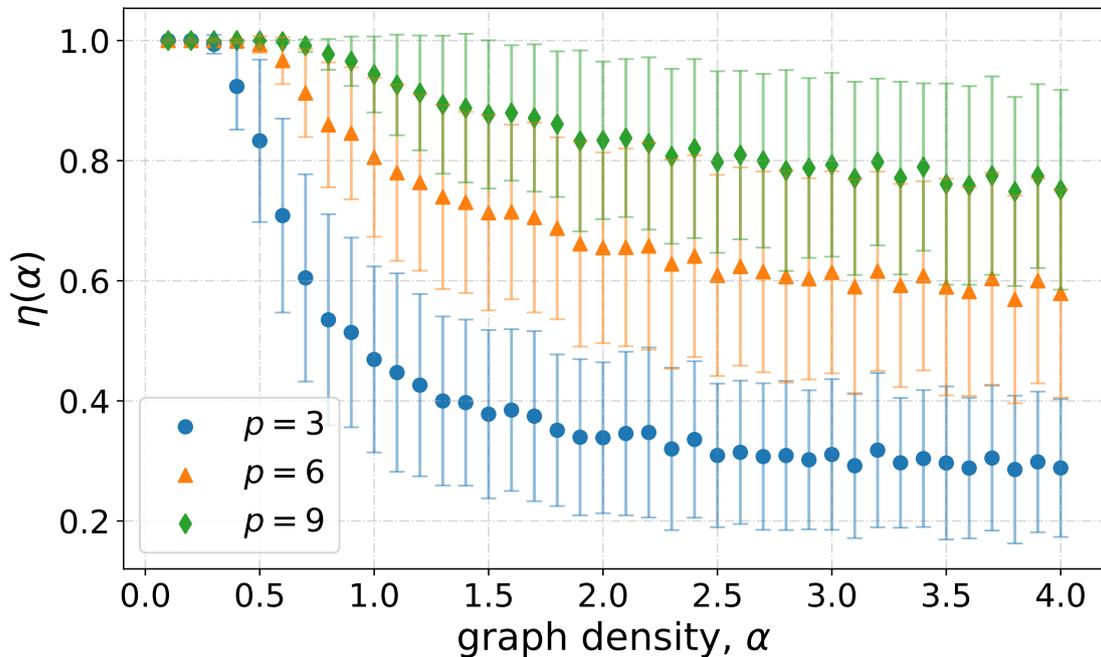


Figure 4-6: Success probability versus graph density for depth  $p = 3, 6, 9$  QAOA. Each data point represents the average performance over statistics of 300 uniform random graph instances,  $G_{n,m}$  with nodes  $n = 10$  and error bars indicating the standard deviation. Figure reprocessed from [2].

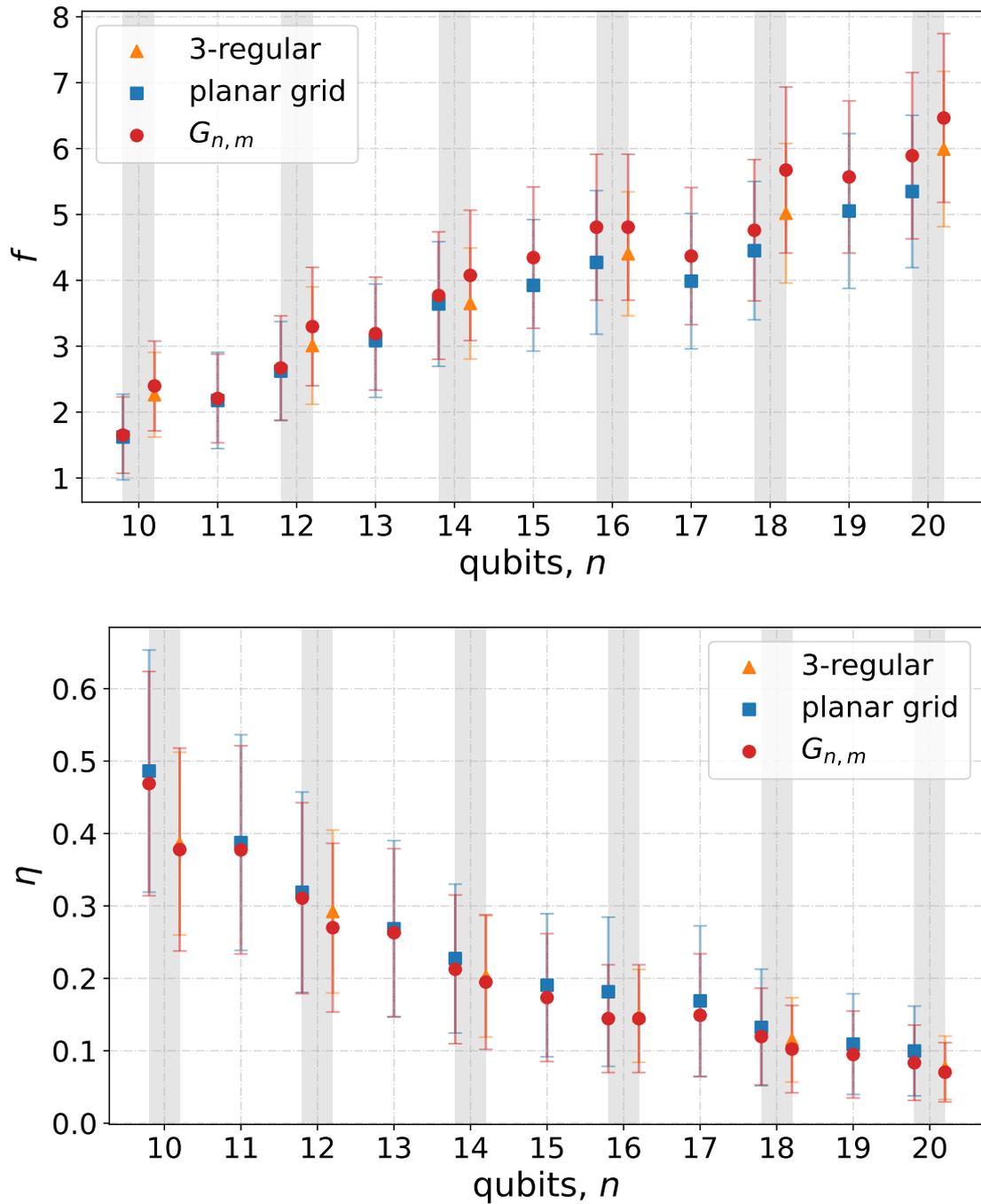


Figure 4-7: Comparing (top) error in approximation and (bottom) success probability of the three graph families. For even numbers of qubits, the left vertical pair represents the planar grid to random graph comparison, and the right vertical pair represents the 3-regular to random graph comparison. For odd numbers of qubits, the grid to random graph comparison is shown (since 3-regular graphs can be generated only with an even number of nodes). Each data point represents the average performance of depth  $p = 3$  QAOA over statistics of 300 randomly generated instances. By comparison, we do not observe any topology related biases in QAOA performance that can be of statistical relevance. Figure reprocessed from [2].

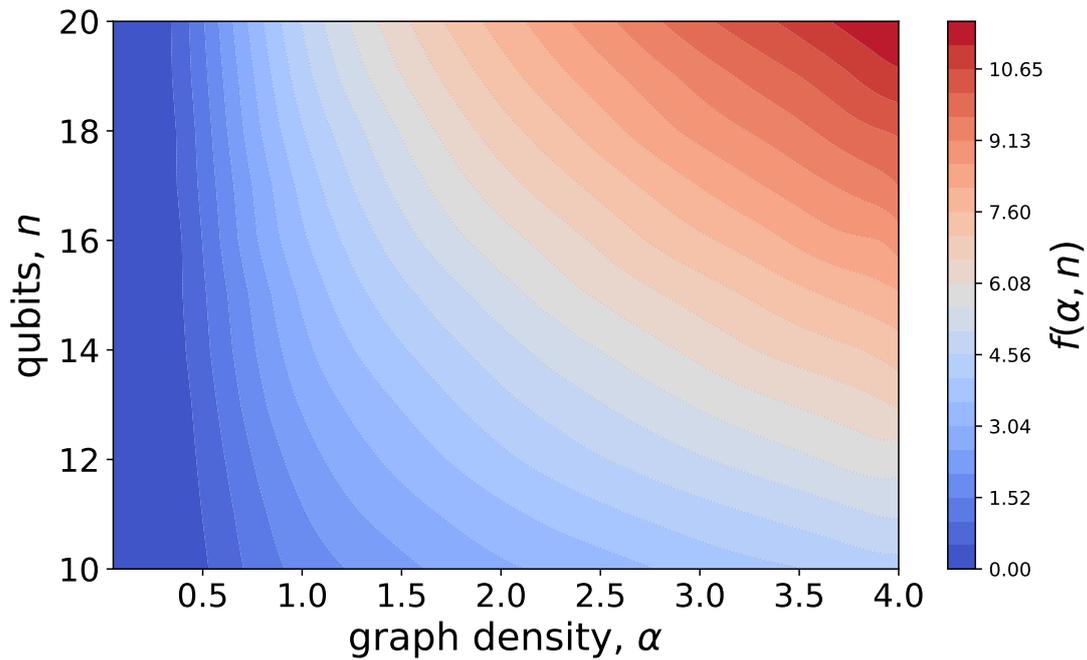


Figure 4-8: Performance landscape for depth  $p = 3$  QAOA on uniform random graph instances  $G_{n,m}$ , across varying graph densities, and for graphs with 10 up to 20 nodes. Contours represent polynomial fits of average error in approximations over statistics of 300 instances generated for each density and node count. We observe a rapid fall-off region in approximation quality experienced beyond intermediate-density and low instance sizes. Figure reprocessed from [2].

## 4.6 Discussion

QAOA has been applied many times throughout the literature, and many findings report its surprising success. However, findings to date appear to be implicitly constrained to instances of low problem density (the ratio of instances constraints to variables). Hence, the considered instances are not representative of statistically likely examples and are only representative of the low-density subset. It is precisely this low-density subset that does not exhibit reachability deficits.

Considering random instances of the satisfiability problem, we observe: instances with relatively low clause density require low-depth QAOA circuits, whereas for high-density instances, larger depth is required in order to approximate the minimum. The exhibited point of this crossover further correlates with the algorithmic phase-transition point for the considered combinatorial optimization problem. This feature is reminiscent of the behavior of increasing computational resources required for classical algorithms attempting solutions for problem instances beyond the algorithmic phase transition [94]. Such transitions seem to be a priori different; in physics, one is the onset of non-trivial macroscopic collective behavior in a system composed of a large number of elements that follow simple microscopic laws [95], and in computer science, one is the computational difficulty existing in a solution search process. However, computational difficulty is reflected both in solving problem instances and in simulating physical processes at such critical points [96, 97]. Furthermore, such algorithmic transitions are widely believed to be a common feature for NP problems [41], leading us to believe that our results extend beyond satisfiability. Indeed, by considering optimization problems on graphs where the problem density is given by the ratio of graph edges to graph nodes (called graph density), we empirically observed similar findings and were able to analytically demonstrate the existence of reachability deficits, albeit for a statistically insignificant case. Therefore, we are motivated to generalize the effect of reachability deficits as a fundamental limitation for QAOA.

# Chapter 5

## Concentration effects in QAOA

In this chapter, we consider effects related to training QAOA circuits i.e., in finding optimal parameters at fixed circuit depth. The chapter first introduces and formalizes what are known as *concentration effects* in QAOA [33, 35, 98, 99]. We then review the result of *instance concentrations* [33] for  $p = 1$  QAOA circuits in minimizing the Sherrington-Kirkpatrick model before describing our analytical results on *parameter concentrations* [3] for  $p = 1, 2$  QAOA on the problem of variational state preparation. The chapter will then include some numerical results on *parameter concentrations* before closing with a discussion on the prospects of reducing the classical computational cost by leveraging concentrations.

### 5.1 Background

Variational algorithms consist of an outer-loop classical optimization step that assigns parameters to a quantum circuit in order to minimize an objective Hamiltonian. This classical step involves finding the minima of a high-dimensional landscape possibly plagued with local minimas [100]. As such, gradient-free and gradient-based methods are utilized to achieve convergence to the global minimum. Specifically, in the case of QAOA, a  $p$ -depth ansatz is prescribed by a total of  $2p$  variational parameters. Therefore, by increasing depth, one encounters an exponentially large parameter space when searching for a candidate optimum. Furthermore, due to the phenomenon of *barren plateaus* [88, 89], gradient-based optimizers may end up fail-

ing at the task. Owing to these challenges, training parameterized quantum circuits in itself is an active area of research. Several strategies have been proposed to alleviate some of the challenges in training parameterized circuits [25, 29, 31, 101, 102]. Although they work for specific cases, these heuristic strategies may be suboptimal [103].

A different approach to reducing the complexity associated with the classical optimization step in QAOA is by leveraging *concentrations* [33, 35, 98, 99]. Concentrations arise in the literature as folklore: Though mentioned in numerical and even analytical studies [98, 99, 101], their precise definition, scaling behavior, and analytic prediction are lacking. State-of-the-art analytical approaches were based on the fully connected Sherrington-Kirkpatrick model. For general depth  $p$ , it was shown that QAOA becomes instance independent in the infinite system size limit ( $n \rightarrow \infty$ ) [33]. Although this result applies to concentrations with respect to instances, the scaling behavior of optimal parameters was not addressed. In addition to instance concentrations, several numerical studies report *distributions* over optimal parameters even when QAOA is considered on random instances [29, 99, 104]. Furthermore, such distributions have been empirically shown to behave non-trivially with respect to  $n$ , and therefore, add to the folklore of concentrations.

## 5.2 Concentration effects in QAOA

**Definition 5.1** (Instance Concentration). *Let  $\Gamma$  represent a problem class, and let  $H_g$  represent the corresponding problem Hamiltonians for random instances  $g \in \Gamma$  with fixed problem size  $|g| = n$ . For QAOA circuits of fixed depth  $p \in \mathbb{N}$ , instances concentrate whenever:*

$$\lim_{n \rightarrow \infty} \mathbb{E}_g [\langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | H_g^2 | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle] - \mathbb{E}_g^2 [\langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | H_g | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle] = 0. \quad (5.1)$$

**Definition 5.2** (Parameter Concentration). *Let  $\Gamma$  represent a problem class, and let  $H_g$  represent the corresponding problem Hamiltonians for random instances  $g \in \Gamma$ . For QAOA circuits of fixed depth  $p \in \mathbb{N}$ , given  $(\boldsymbol{\gamma}_n^*, \boldsymbol{\beta}_n^*) \in \arg \min_{\boldsymbol{\gamma}, \boldsymbol{\beta}} \langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | H_g | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle$ .*

Parameters concentrate whenever:

$$\begin{aligned} \forall (\beta_n^*, \gamma_n^*) \exists (\beta_{n+1}^*, \gamma_{n+1}^*) : \\ |\beta_{n+1}^* - \beta_n^*|^2 + |\gamma_{n+1}^* - \gamma_n^*|^2 \sim \text{poly}^{-1}(n). \end{aligned} \quad (5.2)$$

### 5.3 Instance concentration in SK model

In this section, we review the calculation done in [33] which analytically demonstrates the effect of instance concentration for the case of  $p = 1$  QAOA on the Sherrington-Kirkpatrick model.

SK model [105] is a mean-field approach to spin glasses. Here, we consider a system of  $n$ -spins with all-to-all connectivity. The associated Hamiltonian for such a system is described by:

$$H_{SK}(\boldsymbol{\sigma}) = \frac{1}{\sqrt{n}} \sum_{j < k} J_{jk} \sigma_j \sigma_k, \quad (5.3)$$

where  $\boldsymbol{\sigma} \in \{-1, 1\}^{\times n}$  represents a spin configuration and the couplings  $J_{jk}$ , are drawn independently from a normal distribution with mean 0 and variance 1. An interesting problem in this regard is to calculate the minimum energy density in the thermodynamic limit of  $n \rightarrow \infty$ . This quantity has been shown to exist [72] and, when calculated numerically, gives:

$$\lim_{n \rightarrow \infty} \min_{\boldsymbol{\sigma}} \frac{H_{SK}(\boldsymbol{\sigma})}{n} = -0.7631. \quad (5.4)$$

Though this celebrated result allows one to know the exact minimum in the thermodynamic limit, the associated spin configuration that attains this optima is hard to find using known methods of optimization. A few notable results include 1) Zero temperature simulated annealing does not promise recovery of the exact low-energy configurations [106] and 2) semidefinite programming methods recover spin configurations with energy density  $\frac{H_{SK}(\boldsymbol{\sigma})}{n} \approx -\frac{2}{\pi}$  [107]. Knowing these results, one can ask how well does QAOA perform in this setting? As was shown in [33],  $p = 11$  depth QAOA outperforms standard semidefinite programming. Here, instead, we

review their result specific to an effect in QAOA, called *instance concentrations*.

To begin with, we consider the typical performance of  $p = 1$  depth QAOA on instances of the SK-model as in Eq. (5.3).

**Definition 5.3** (Typical Performance of SK instances). *Given a  $p$ -depth QAOA circuit as in Eq. (3.4), let  $H$  represent an instance of the SK-model with random couplings  $\mathbf{J}$  drawn independently from the standard normal distribution (as in Eq. (5.3)). The typical performance is then defined as expected value, with respect to  $p$ -depth QAOA ansatz, averaged over instance:*

$$\mathbb{E}_{\mathbf{J}} \left[ \langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | \frac{H}{n} | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle \right]. \quad (5.5)$$

In order to obtain a closed form expression for Eq. (5.5) at  $p = 1$ , we consider the moment generating function,

$$F(t) = \mathbb{E}_{\mathbf{J}} \left[ \langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | e^{it \frac{H}{n}} | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle \right]. \quad (5.6)$$

Note that,

$$-i \frac{d}{dt} F(t) \Big|_{t=0} = \mathbb{E}_{\mathbf{J}} \left[ \langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | \frac{H}{n} | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle \right], \quad (5.7)$$

and similarly,

$$- \frac{d^2}{dt^2} F(t) \Big|_{t=0} = \mathbb{E}_{\mathbf{J}} \left[ \langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | \left( \frac{H}{n} \right)^2 | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle \right]. \quad (5.8)$$

Explicit expressions for Eqs. (5.7),(5.8) is what we need to prove instance concentrations for  $p = 1$  QAOA.

### 5.3.1 Explicit form for moment generating function

Consider  $F(t)$  for  $p = 1$  in Eq. (5.6) and working in the spin basis,  $\mathbf{z} \in \{-1, +1\}^{\times n}$ ,

$$\begin{aligned}
 F(t) &= \sum_{\mathbf{z}^m} \mathbb{E}_{\mathbf{J}} \left[ e^{it\frac{H}{n}(\mathbf{z}^m)} \langle + | e^{i\gamma\frac{H}{n}} \cdot e^{i\beta H_x} | \mathbf{z}^m \rangle \langle \mathbf{z}^m | e^{-i\beta H_x} \cdot e^{-i\gamma\frac{H}{n}} | + \rangle \right] \\
 &= \frac{1}{2^n} \sum_{\mathbf{z}^m, \mathbf{z}^1, \mathbf{z}^2} \mathbb{E}_{\mathbf{J}} \left[ f(\mathbf{z}^m, \mathbf{z}^1, \mathbf{z}^2) \langle \mathbf{z}^1 | e^{i\beta H_x} | \mathbf{z}^m \rangle \langle \mathbf{z}^m | e^{-i\beta H_x} | \mathbf{z}^2 \rangle \right].
 \end{aligned} \tag{5.9}$$

Here  $\mathbf{z}^m, \mathbf{z}^1, \mathbf{z}^2 \in \{-1, 1\}^{\times n}$  and we make use of the standard QAOA circuit with  $H_x \sum_{k=1}^n X_k$ . Next, we simplify the exponent,  $f(\mathbf{z}^m, \mathbf{z}^1, \mathbf{z}^2)$ ,

$$f(\mathbf{z}^m, \mathbf{z}^1, \mathbf{z}^2) = \exp \left\{ i \left( \frac{t}{n} H(\mathbf{z}^m) + \gamma [H(\mathbf{z}^1) - H(\mathbf{z}^2)] \right) \right\}. \tag{5.10}$$

Substituting the explicit form for the energies from Eq. (5.3), we obtain,

$$f(\mathbf{z}^m, \mathbf{z}^1, \mathbf{z}^2) = \exp \left\{ \frac{i}{\sqrt{n}} \sum_{j < k} J_{jk} \left( z_j^m z_k^m \cdot \frac{t}{n} + \gamma [z_j^1 z_k^1 - z_j^2 z_k^2] \right) \right\}. \tag{5.11}$$

Note that since we work in the spin basis and sum over all possible strings  $\mathbf{z}^m, \mathbf{z}^1, \mathbf{z}^2$ , we can redefine  $\mathbf{z}^1 \rightarrow \mathbf{z}^1 \cdot \mathbf{z}^m$  and  $\mathbf{z}^2 \rightarrow \mathbf{z}^2 \cdot \mathbf{z}^m$ , then we obtain,

$$\begin{aligned}
 f(\mathbf{z}^m, \mathbf{z}^1, \mathbf{z}^2) &= \exp \left\{ \frac{i}{\sqrt{n}} \sum_{j < k} J_{jk} z_j^m z_k^m \left( \frac{t}{n} + \gamma [z_j^1 z_k^1 - z_j^2 z_k^2] \right) \right\} \\
 &= \exp \left\{ \frac{i}{\sqrt{n}} \sum_{j < k} \tilde{J}_{jk} \left( \frac{t}{n} + \gamma [z_j^1 z_k^1 - z_j^2 z_k^2] \right) \right\} \\
 &= \prod_{j < k} \exp \left\{ \frac{i}{\sqrt{n}} \tilde{J}_{jk} \left( \frac{t}{n} + \gamma [z_j^1 z_k^1 - z_j^2 z_k^2] \right) \right\}
 \end{aligned} \tag{5.12}$$

Here,  $\tilde{J}_{jk} = \pm J_{jk}$ . Since  $\mathbb{E}_{\mathbf{J}}[\dots] = \mathbb{E}_{\tilde{\mathbf{J}}}[\dots]$ , we drop the tilde notation and make the substitution in Eq. (5.9). Simplifying this expression, we obtain,

$$\begin{aligned}
 F(t) &= \sum_{\mathbf{z}^1, \mathbf{z}^2} \mathbb{E}_{\mathbf{J}} [f(\mathbf{z}^1, \mathbf{z}^2)] \langle \mathbf{z}^1 | e^{i\beta H_x} | \mathbf{1} \rangle \langle \mathbf{1} | e^{i\beta H_x} | \mathbf{z}^2 \rangle \\
 &= \sum_{\mathbf{z}^1, \mathbf{z}^2} \prod_{j < k} \mathbb{E}_{\mathbf{J}} \left[ \exp \left\{ \frac{i}{\sqrt{n}} J_{jk} \left( \frac{t}{n} + \gamma [z_j^1 z_k^1 - z_j^2 z_k^2] \right) \right\} \right] \times \\
 &\quad g_{\beta}^* (\mathbf{z}^1) \cdot g_{\beta} (\mathbf{z}^2), \tag{5.13}
 \end{aligned}$$

or in simplified notation,

$$\sum_{\mathbf{z}^1, \mathbf{z}^2} \prod_{j < k} \mathbb{E}_{\mathbf{J}} \left[ \exp \left\{ \frac{i}{\sqrt{n}} J_{jk} \left( \frac{t}{n} + \Phi_{jk} (\mathbf{z}^1, \mathbf{z}^2) \right) \right\} \right] g_{\beta}^* (\mathbf{z}^1) \cdot g_{\beta} (\mathbf{z}^2), \tag{5.14}$$

where,

$$g_{\beta}^* (\mathbf{z}^1) = (\cos (\beta))^{\#\mathbf{1} \text{ in } \mathbf{z}^1} \cdot (i \sin (\beta))^{\#\mathbf{-1} \text{ in } \mathbf{z}^1}, \tag{5.15}$$

and,

$$\Phi_{jk} (\mathbf{z}^1, \mathbf{z}^2) = \gamma [z_j^1 z_k^1 - z_j^2 z_k^2]. \tag{5.16}$$

Now we take the averaging over instances,  $\mathbb{E}_{\mathbf{J}} [\dots]$ . Note that  $\mathbb{E}_{J_{jk}} [e^{iJ_{jk}x}] = e^{-x^2/2}$ , substituting the result, we obtain,

$$\begin{aligned}
 F(t) &= \sum_{\mathbf{z}^1, \mathbf{z}^2} \prod_{j < k} \exp \left\{ \frac{-1}{2n} \left( \frac{t}{n} + \Phi_{jk} \right)^2 \right\} g_{\beta}^* (\mathbf{z}^1) \cdot g_{\beta} (\mathbf{z}^2) \\
 &= \sum_{\mathbf{z}^1, \mathbf{z}^2} \exp \left\{ \frac{-1}{2n} \sum_{j < k} \Phi_{jk}^2 - \frac{t}{n} \sum_{j < k} \Phi_{jk} - \frac{t^2}{n^3} \binom{n}{2} \right\} g_{\beta}^* (\mathbf{z}^1) \cdot g_{\beta} (\mathbf{z}^2). \tag{5.17}
 \end{aligned}$$

To evaluate the sum in Eq. (5.17), it is convenient to introduce a new basis. The authors in [33] call this a configuration basis. The motivation for this choice of basis is as follows:

Given 2 spin configurations  $\mathbf{z}^1$  and  $\mathbf{z}^2$ , consider the sum,

$$\begin{aligned}
 \sum_{j < k} \Phi_{jk}(\mathbf{z}^1, \mathbf{z}^2) &= \frac{\gamma}{2} \left( \sum_{j,k} z_j^1 z_k^1 - \sum_{j,k} z_j^2 z_k^2 \right) \\
 &= \frac{\gamma}{2} \left[ \left( \sum_j z_j^1 \right)^2 - \left( \sum_j z_j^2 \right)^2 \right] \\
 &= \frac{\gamma}{2} \left[ \left( \sum_j z_j^1 + z_j^2 \right) \left( \sum_j z_j^1 - z_j^2 \right) \right].
 \end{aligned} \tag{5.18}$$

From the form of this function, we only require knowledge of 4 numbers:  $\{n_{++}, n_{+-}, n_{-+}, n_{--}\}$ , to evaluate  $\sum_{j < k} \Phi_{jk}(\mathbf{z}^1, \mathbf{z}^2)$ . Here,  $n_{++}$  corresponds to the number of positions (taken in order) along  $\mathbf{z}^1$  and  $\mathbf{z}^2$  whenever  $z_j^1 = z_j^2 = +1$  and similarly for  $n_{+-}$  where we count positions along  $\mathbf{z}^1$  and  $\mathbf{z}^2$  whenever  $z_j^1 = +1$  and  $z_j^2 = -1$ . Therefore, instead of summing all possible spin configurations for  $\mathbf{z}^1$  and  $\mathbf{z}^2$ , we sum over  $\{n_{++}, n_{+-}, n_{-+}, n_{--}\}$  accounting for multiplicities due to permutations. Note that a valid configuration,  $\{\mathbf{n}_a\}$ ,  $a \in \{++, +-, -+, --\}$  necessarily satisfies  $\sum_a n_a = n$ . Eq. (5.17) then becomes,

$$F(t) = \sum_{\{\mathbf{n}_a\}} \binom{n}{\{\mathbf{n}_a\}} \exp \left\{ \frac{-1}{2n} \sum_{j < k} \Phi_{jk}^2(\{\mathbf{n}_a\}) - \frac{t}{n} \sum_{j < k} \Phi_{jk}(\{\mathbf{n}_a\}) - \frac{t^2}{n^3} \binom{n}{2} \right\} \prod_a G_a^{n_a}, \tag{5.19}$$

where,  $G_{++} = \cos^2 \beta$ ,  $G_{+-} = G_{-+} = i \cos \beta \sin \beta$  and  $G_{--} = \sin^2 \beta$ .

### 5.3.2 Evaluation of moments

To arrive at the explicit expression, we evaluate  $-i \frac{d}{dt} F(t) \Big|_{t=0}$  and  $-\frac{d^2}{dt^2} F(t) \Big|_{t=0}$ .

$$\begin{aligned}
 \mathbb{E}_{\mathbf{J}} \left[ \langle \psi(\gamma, \beta) | \frac{H}{n} | \psi(\gamma, \beta) \rangle \right] &= \frac{i2\gamma}{n^2} \sum_{\{\mathbf{n}_a\}} \binom{n}{\{\mathbf{n}_a\}} (n_{++} - n_{--}) (n_{+-} - n_{-+}) \times \\
 &\quad \exp \left[ \frac{-2\gamma^2}{n} (n_{++} + n_{--}) (n_{+-} + n_{-+}) \right] \prod_a G_a^{n_a},
 \end{aligned} \tag{5.20}$$

which can be further simplified by grouping  $n_{+-} + n_{-+} = t$  and  $n_{++} + n_{--} = n - t$ ,

$$\begin{aligned}
 \mathbb{E}_{\mathbf{J}} \left[ \langle \psi(\gamma, \beta) | \frac{H}{n} | \psi(\gamma, \beta) \rangle \right] &= \frac{i2\gamma}{n^2} \sum_{t=0}^n \binom{n}{t} \exp \left[ -2\gamma^2 \frac{t(n-t)}{n} \right] \times \\
 &\quad \sum_{n_{+-} + n_{-+} = t} \binom{n}{n_{+-}, n_{-+}} (n_{+-} - n_{-+}) G_{+-}^{n_{+-}} G_{-+}^{n_{-+}} \times \\
 &\quad \sum_{n_{++} + n_{--} = n-t} \binom{n-t}{n_{++}, n_{--}} (n_{++} - n_{--}) G_{++}^{n_{++}} G_{--}^{n_{--}}.
 \end{aligned} \tag{5.21}$$

For the inside summands we make use of the following identity,

$$\sum_{a+b=n} \binom{n}{a, b} (a-b) x^a y^b = n(a-b)(a+b)^{n-1}, \tag{5.22}$$

therefore upon simplification we arrive at the final expression as,

$$\mathbb{E}_{\mathbf{J}} \left[ \langle \psi(\gamma, \beta) | \frac{H}{n} | \psi(\gamma, \beta) \rangle \right] = \frac{n-1}{n} \gamma \exp \left( -2\gamma^2 \frac{n-1}{n} \right) \sin(4\beta), \tag{5.23}$$

when taking the thermodynamic limit  $n \rightarrow \infty$ ,

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbf{J}} \left[ \langle \psi(\gamma, \beta) | \frac{H}{n} | \psi(\gamma, \beta) \rangle \right] = \gamma \exp(-2\gamma^2) \sin(4\beta). \tag{5.24}$$

A similar calculation can be done for the second moment [33] which results in,

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbf{J}} \left[ \langle \psi(\gamma, \beta) | \left( \frac{H}{n} \right)^2 | \psi(\gamma, \beta) \rangle \right] = \gamma^2 \exp(-4\gamma^2) \sin^2(4\beta), \tag{5.25}$$

which is the squared first moment. This then establishes instance concentration for  $p = 1$  QAOA.

## 5.4 Parameter concentrations in variational state preparation

Variational state preparation can be stated as follows: let  $|t\rangle$  be a  $n$ -qubit target state in the computational basis. The task is to variationally prepare a candidate state  $|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle$  with high overlap with  $|t\rangle$ . Note that this problem can be viewed as an optimization problem:

**Definition 5.4** (Variational state preparation). *The optimization task in variational state preparation is to maximize the overlap between the candidate state  $|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle$  and the target state  $|t\rangle$  given by  $|\langle t|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle|^2$ . This is equivalent to the minimization of the problem Hamiltonian  $H_t = \mathbb{1} - |t\rangle\langle t|$ ,*

$$\min_{\boldsymbol{\gamma}, \boldsymbol{\beta}} \langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | H_t | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle = 1 - \max_{\boldsymbol{\gamma}, \boldsymbol{\beta}} |\langle t | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle|^2. \quad (5.26)$$

**Theorem 5.1** (Parameter concentration in variational state preparation). *Let  $|t\rangle \in [\mathbb{C}_2]^{\otimes n}$  be an  $n$  qubit target state in the computational basis. For depth  $p = 1, 2$ , parameters concentrate as:*

$$|\beta_{n+1}^* - \beta_n^*|^2 + |\gamma_{n+1}^* - \gamma_n^*|^2 = \mathcal{O}\left(\frac{1}{n^4}\right). \quad (5.27)$$

*Proof.* We start by calculating the parameters that deliver maximum for the overlap function  $|\langle t|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle|^2$ .

### 5.4.1 Parameter concentration at $p=1$

Consider  $p = 1$  QAOA ansatz:

$$|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = e^{-i\beta H_x} e^{-i\gamma |t\rangle\langle t|} |+\rangle^{\otimes n}. \quad (5.28)$$

The required amplitude can be calculated as,

$$\begin{aligned} g_1(\gamma, \beta) &= \langle t | \psi(\gamma, \beta) \rangle \\ &= \frac{1}{\sqrt{2^n}} [e^{-i\beta n} + \cos^n \beta (e^{-i\gamma} - 1)]. \end{aligned} \quad (5.29)$$

which then gives the explicit overlap function:

$$\begin{aligned} F_1(\gamma, \beta) &= \frac{1}{2^n} [1 + 2 \cos^n(\beta) (\cos(\gamma - n\beta) - \cos(n\beta)) \\ &\quad + 2 \cos^{2n}(\beta) (1 - \cos \gamma)]. \end{aligned} \quad (5.30)$$

We are now concerned with parameters that maximize Eq. (5.30), and therefore conditions  $\partial_\gamma F_1(\gamma, \beta) = \partial_\beta F_1(\gamma, \beta) = 0$  must be satisfied. From these conditions, we can arrive at two equations:

$$\tan \gamma = \frac{\sin n\beta}{\cos n\beta - \cos^n \beta}, \quad (5.31)$$

$$\tan \frac{\gamma}{2} = \frac{\cos(n\beta + \beta)}{2 \cos^n \beta \sin \beta - \sin(n\beta + \beta)}. \quad (5.32)$$

Merging Eq. (5.31) and Eq. (5.32) we get:

$$\cos^n \beta \sin 2\beta - \sin(n+2)\beta = 0. \quad (5.33)$$

Substituting Eq. (5.33) into Eq. (5.32), one can relate optimal parameters as  $\gamma = \pi - 2\beta$ .

We now consider solutions for Eq. (5.33). For a given  $n$ , it can be rewritten as a polynomial equation of power  $n$  which does not have solutions in radicals. However, we see that for  $n \rightarrow \infty$ ,  $\beta = \frac{\pi}{n} + \mathcal{O}(n^{-2})$ . We find the correction to this asymptotic solution as  $\beta = \frac{\pi}{n} + \delta\beta$ , and calculate  $\delta\beta = -\frac{4\pi}{n^2}$  up to second order in  $1/n$ . Thus,

for the optimal parameters, we get:

$$\beta^* = \frac{\pi}{n} - \frac{4\pi}{n^2} + \mathcal{O}(n^{-3}), \quad (5.34)$$

$$\gamma^* = \pi - \frac{2\pi}{n} + \frac{8\pi}{n^2} + \mathcal{O}(n^{-3}). \quad (5.35)$$

Therefore,

$$\begin{aligned} & |\beta_{n+1}^* - \beta_n^*|^2 + |\gamma_{n+1}^* - \gamma_n^*|^2 \approx \\ & \approx \frac{5\pi^2}{(n+4)^2(n+5)^2} = \mathcal{O}\left(\frac{1}{n^4}\right), \end{aligned} \quad (5.36)$$

which establishes parameter concentration for  $p = 1$ .

### 5.4.2 Parameter concentration for $p=2$

We repeat the process for  $p = 2$ . The ansatz in this case is:

$$\begin{aligned} |\psi(\gamma_1, \beta_1, \gamma_2, \beta_2)\rangle = \\ e^{-i\beta_2\mathcal{H}_x} e^{-i\gamma_2|t\rangle\langle t|} e^{-i\beta_1\mathcal{H}_x} e^{-i\gamma_1|t\rangle\langle t|} |+\rangle^{\otimes n}. \end{aligned} \quad (5.37)$$

The corresponding amplitude can be expressed in terms of the amplitude at  $p = 1$  from Eq. (5.29):

$$\begin{aligned} g_2(\gamma_1, \beta_1, \gamma_2, \beta_2) = \\ g_1(\gamma_1, \beta_1 + \beta_2) + g_1(\gamma_1, \beta_1) \cos^n \beta_2 (e^{-i\gamma_2} - 1). \end{aligned} \quad (5.38)$$

To find the parameters that maximize overlap  $F_2(\gamma_1, \beta_1, \gamma_2, \beta_2)$ , we set the gradients to zero and obtain a set of four equations. Even though in this case the variables do not separate, it can be seen that for  $n \rightarrow \infty$ , solutions behave as:

$$n\beta_i^* \rightarrow \pi, \gamma_i^* \rightarrow \pi. \quad (5.39)$$

Assuming  $n \gg 1$  and corrections to be of the next order in  $1/n$ , we again search

for parameters as  $\beta_i = \frac{\pi}{n} + \delta\beta_i$  and  $\gamma = \pi + \delta\gamma_i$  and obtain optimal parameters as:

$$\beta_2^* = \frac{\pi}{n} - \frac{4\pi}{n^2} + \mathcal{O}(n^{-3}), \quad (5.40)$$

$$\gamma_2^* = \pi - \frac{2\pi}{n} + \mathcal{O}(n^{-2}), \quad (5.41)$$

$$\beta_1^* = \frac{\pi}{n} + \mathcal{O}(n^{-3}), \quad (5.42)$$

$$\gamma_1^* = \pi + \mathcal{O}(n^{-2}). \quad (5.43)$$

Therefore,

$$|\beta_{n+1}^* - \beta_n^*|^2 + |\gamma_{n+1}^* - \gamma_n^*|^2 = \mathcal{O}\left(\frac{1}{n^4}\right), \quad (5.44)$$

which establishes parameter concentration for  $p = 2$ . ■

### 5.4.3 Numerical results for $p \geq 3$

Although the method described for obtaining optimal parameters can be extended for higher depths  $p \geq 3$ , the sets of equations that describe zero gradients may not be trivially simplified. A strategy then is to consider the limit  $n \rightarrow \infty$  and find corrections to the first-order solutions, as we did in the case of  $p = 2$ . Interestingly, we numerically see that for up to 17 qubits and  $p = 5$ , the optimal solutions behave as:

$$n\beta_i \rightarrow \pi, \gamma_i \rightarrow \pi. \quad (5.45)$$

We expect it to be a general feature for this problem, and therefore to obtain corrections one can Taylor-expand the overlap function around low-order solutions as in Eq. (5.39) up to second order in  $1/n$ . This simplified expression is of a quadratic form and thus can be maximized to obtain corrections. For any finite region in  $n$ , we see that such optimal parameters can be well approximated by the functions,

$$\beta = \frac{\pi}{a_1 n + a_2}, \quad \gamma = b_1 \pi - b_2 \beta, \quad (5.46)$$

where the fitting constants  $a_{1,2}$  and  $b_{1,2}$  are region specific.

In the following, we present our numerical results for  $p = 5$ . Figure 5-1 shows the optimal parameters calculated numerically for the last layer. The optimal parameters at depth 1 – 5 are fit according to Eq. (5.46) and the corresponding fit constants appear in Table 5.1.

According to fitting curves which accurately describe the numerical data, parameter concentration is evident and is the same as in Eq. (5.44). We also plot our numerical data in Figure 5-2 to visually illustrate the phenomenon of parameter concentration.

	$a_1$	$a_2$	$b_1$	$b_2$
$\beta_1, \gamma_1$	1.04	0.92	1.06	2.07
$\beta_2, \gamma_2$	0.98	1.23	1.05	2.04
$\beta_3, \gamma_3$	0.94	1.58	1.05	1.96
$\beta_4, \gamma_4$	0.88	2.32	1.03	1.83
$\beta_5, \gamma_5$	1.09	5.25	1.0	2.0

Table 5.1: Parameters  $a_{1,2}$  and  $b_{1,2}$  according to fitting functions (5.46).

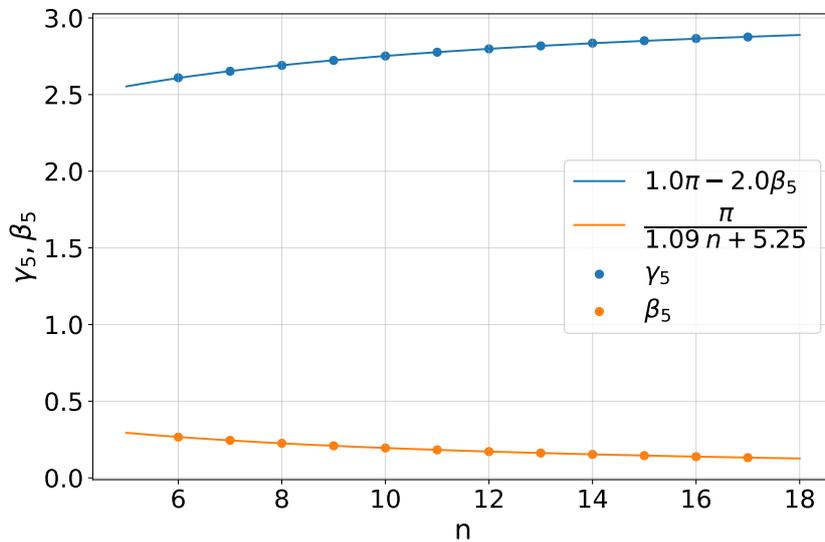


Figure 5-1: Comparison of numerical results (dots) and our fitting for optimal parameters (lines)  $\beta_5$  (bottom; orange),  $\gamma_5$  (top; blue) with respect to the number of qubits  $n$ . The dots represent the parameters obtained numerically via maximization of the overlap function. Figure reprocessed from [3].

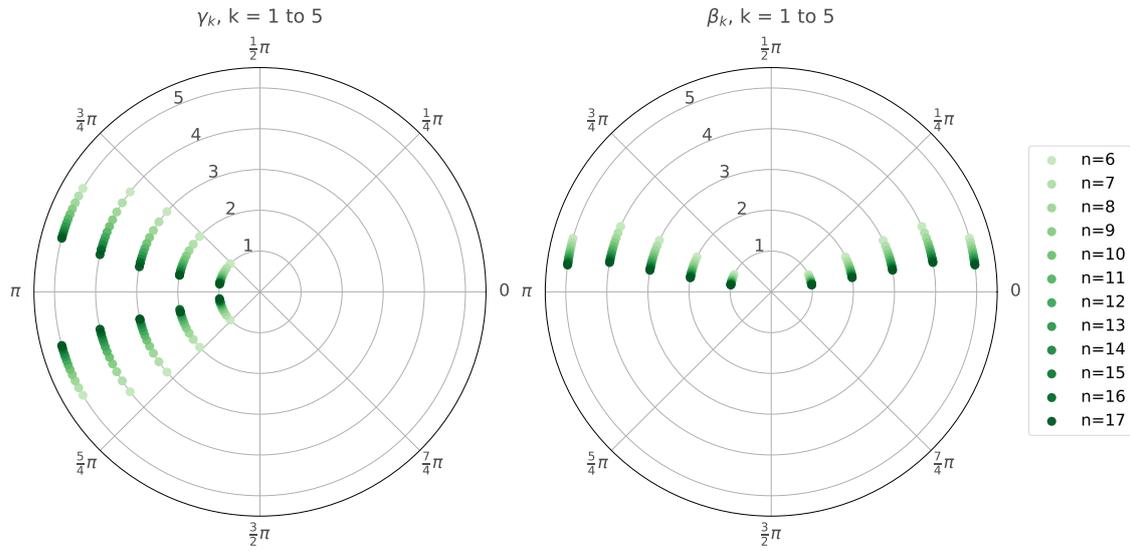


Figure 5-2: Parameter concentrations visualized for  $p = 5$  depth QAOA and varying number of qubits  $n$ . Two symmetric branches of optimal parameters are well distinguished. Parameter concentration is seen within each of the branches. Figure reprocessed from [3].

## 5.5 Discussion

Our work on concentrations begins with a rigorous definition of parameter concentrations. This definition is motivated by how the effect can be leveraged for efficient training. Different approaches claim concentrations in the QAOA literature, yet our results have a clear distinction. Specifically, [33] analytically addresses what we call *instance concentration* in the case of the Sherrington-Kirkpatrick model. Here, the variance in the objective function value vanishes in the infinite size limit ( $n \rightarrow \infty$ ), and therefore QAOA becomes instance independent. However, the result alone does not predict nor address the behavior of optimal parameters.

Numerically in [99] it is seen that the optimal parameters at each depth  $p$ , are distributed over a small range when considering randomly generated MAX-CUT instances on 3-regular graphs. Furthermore, this distribution became narrower as the system size increased. The authors in [99] explain their numerical observation via *reverse causal cone* i.e. the subgraphs effectively contributing to the objective

function when taken over a particular edge. Since in the limit  $n \rightarrow \infty$ , the likely subgraphs are trees, the optimal parameters for QAOA on such subgraphs become optimal for the entire graph. However, the scaling behavior (if any) for the optimal parameters towards the infinite  $n$  limit still remain lacking.

Although our results on parameter concentration apply to the specific case of variational state preparation, wherein only a unique set of optimal parameters is addressed at each  $n$ , their scaling behavior is fully understood due to our analytical result. Furthermore, we expect our results to be applicable in more general settings which would imply that the distribution of optimal parameters with respect to instances are only slightly sensitive if  $n$  is large. This implication can be leveraged to reduce the training cost of finding optimal parameters. In particular, our observed concentrations scale as  $\mathcal{O}(n^{-4})$ , which implies that optimal parameters also have a limit as  $n \rightarrow \infty$ . Therefore, one can train on a finite fraction of  $w \ll n$  qubits and perform a polynomially restricted training over optimal parameters at  $w$  qubits to recover optimal parameters for  $n$  qubits. However, to fully exploit this approach, further investigation is needed.

# Chapter 6

## Circuit depth scaling in QAOA

In this chapter, we propose a predictive model motivated by numerous empirical findings that aims to ascertain the critical depth required for QAOA to succeed in recovering  $\epsilon$  tolerant performance. The chapter begins with a conjecture called *logistic saturation* that postulates the behavior of critical QAOA depth to scale logarithmically with problem density for MAX-2-SAT. The chapter then describes the numerical study, which we use to test the accuracy of the predictive model. Finally, we close with a brief discussion on how such empirical models may expose critical depth scaling with respect to problem size for general combinatorial optimization problems which is currently out of scope beyond specific cases which are statistically underrepresentative.

### 6.1 Background

The circuit structure of QAOA resembles a discretized version of adiabatic evolution and can approximate adiabatic evolution with  $\mathcal{O}(1)$  errors when the depth  $p \sim (T^2 \cdot \text{poly}(n))$  (seen in Chapter 3) [67]. However, when considering variational minimization, this result on depth only provides an upper bound to approximate ground states of some objective Hamiltonian. Shorter depth circuits have indeed been shown to offer some benefits such as: recovering solutions to low-density instances of combinatorial optimization problems [1, 2], guarantying cut that is at least 0.6924 of the optimal cut for specific MAX-CUT instances [26], and recov-

ering near-optimal query complexity for Grover’s search [24, 86]. Beyond combinatorial optimization and short depths, promising results such as prospects for quantum supremacy and universality have been proven in the framework of QAOA [27, 28, 32, 79], making the approach a promising model for NISQ era quantum computing.

Although several advantages have been reported [24, 26, 32, 40, 86, 87, 108–110], effects that limit the performance of QAOA have also been discovered [1, 2, 103]. Taking into account the energy error in approximation as a performance metric, we [1, 2] demonstrated *reachability deficits*. Deeper circuits therefore become a necessity and hence come with overheads in classical outer-loop optimization. Although approaches such as layerwise training aim to reduce this classical computational cost, such strategies have been shown to be suboptimal [103]. Other approaches aimed at reducing the classical computation cost exploit what are known as *concentration effects* [3, 33, 99, 111]. Although these effects allow one to heuristically guess near-optimal parameters at fixed depth and for increasing number of qubits, they fail to address the depth required to guarantee fixed performance. Even while neglecting the computational cost in outer loop optimization, a general theory that addresses the required depth needed for QAOA to succeed remains open.

## 6.2 Logistic saturation

As a brief background, we return to the variational state space available for a  $p$ -depth QAOA circuit. Given some problem Hamiltonian  $H$ , on  $n$  qubits and problem density  $\alpha$ , the variational state space available for QAOA is described by:

$$\Omega = \bigcup_{\gamma, \beta} \{ |\psi(\gamma, \beta)\rangle \}, \quad (6.1)$$

with  $|\psi(\gamma, \beta)\rangle$  given by Eq. (3.4). We have explored that density induces underparameterization in QAOA. For the case of MAX-2-SAT, it was shown that for instances beyond  $\alpha_c \sim 1$ , QAOA suffers reachability deficits. Although better performance is achieved at the cost of increasing QAOA depth, a similar trend is re-

flected for deficits. Here, we investigate the depth required to guarantee an  $\epsilon$  tolerant performance.

**Definition 6.1** (Critical depth). *Given some fixed tolerance on performance,  $\epsilon > 0$ . For QAOA on instances characterized by density  $\alpha$ , and problem size  $n$ , the critical depth  $p^*$  is defined as:*

$$p^* = \min\{p \mid f(\alpha, n) \leq \epsilon\}. \quad (6.2)$$

**Conjecture 6.1** (Logistic Saturation). *Critical depth  $p^*$ , depends on clause density  $\alpha$ , for MAX-2-SAT instances as:*

$$p^*(\alpha) \approx \frac{p_{max}}{1 + e^{-\kappa(\alpha - \alpha_c)}}, \quad (6.3)$$

where  $\alpha_c$  is the critical density,  $\kappa$  the logistic growth rate, and  $p_{max}$  the saturation value.

### 6.3 Numerical study on the predictive model

Here, we briefly review the numerical details of the study. For an elaborate description of the software code emulating QAOA and the optimization heuristics used, we ask the reader to refer to the appendix.

1. To study the performance of QAOA, we initially generate random instances of 2-SAT for problem sizes  $n = 5$  up to 15 qubits according to Eq. (2.7). The problem Hamiltonians are then constructed according to the embedding schemes as described in Section 2.4.1:

$$H = \sum_{j < k} J_{jk} Z_j Z_k + \sum_j h_j Z_j, \quad (6.4)$$

for appropriate choice of coefficients  $J_{jk}$  and  $h_j$ .

2. In generating instances, we fix the number of variables and increment the clause density  $\alpha$  by  $\frac{1}{n}$  for the range  $\alpha \in [1/n, 4.0]$ .

3. To study average performance, we make use of error in approximation,  $f$  as defined in Def. 3.10 and consider a statistic of 300 random instances for densities  $\alpha \in [1/n, 4.0]$  and for qubits  $n = 5$  up to 15.
4. Corresponding to each instance, a problem Hamiltonian  $H_{\text{SAT}}$ , is constructed, and a standard QAOA circuit with depths  $p = 1$  is initialized with random initial parameters  $\boldsymbol{\gamma}$  and  $\boldsymbol{\beta}$  as:

$$|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = \prod_{k=1}^p \exp\{-i\beta_k H_x\} \cdot \exp\{-i\gamma_k H_{\text{SAT}}\} |+\rangle^{\otimes n}, \quad (6.5)$$

where we use the standard driver and ranges,  $H_x = \sum_{j=1}^n X_j$ ,  $\gamma_k \in [0, 2\pi)$  and  $\beta_k \in [0, \pi)$ .

5. As a threshold for performance  $\epsilon = 0.3$  is chosen. By Theorem 3.1, this threshold guarantees a minimum success probability of 0.7.
6. Considering each instance, the algorithm is run with incremental depths of 1 until  $p^*$  as defined in Eq. (6.2) is attained.
7. To find the optimal angles at each depth  $p$ , in the routine,

$$(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*) \in \arg \min_{(\boldsymbol{\gamma}, \boldsymbol{\beta})} \langle \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) | H | \psi(\boldsymbol{\gamma}, \boldsymbol{\beta}) \rangle, \quad (6.6)$$

we use a heuristic optimization strategy motivated by layerwise training and employ the Limited Broyden–Fletcher–Goldfarb–Shanno method (L-BFGS-B) to find optimal parameters during each step of the heuristic strategy.

8. Once  $p = p^*$  is attained the entire process is repeated for the next instance in the statistic.
9. The average  $p^*$  is calculated once the statistics are collected and the procedure is repeated for different densities and qubit counts.
10. The data for average  $p^*$  is then fitted against the logistic function as described in Eq. (6.3). We calculate the  $3\sigma$  confidence interval for the estimated mean and compare the predictive model against the numerical data.

### 6.3.1 Empirical observations

1. Considering up to 15 variables and clause density up to 4 we observe that our proposed model as in Eq. (6.3) describes the data within a  $3\sigma$  confidence interval (see Fig. 6-1).
2. We recovered the scaling of critical depth with respect to problem size, which illustrates a linear trend in our considered range of problem sizes (see Fig. 6-2).

## 6.4 Discussion

QAOA is among the most studied gate-based approaches for combinatorial optimization in NISQ era devices. Although analytical techniques address QAOA at low depth or in the adiabatic limit,  $p \rightarrow \infty$ , understanding performance at intermediate depths remains largely open. Furthermore, problem density induces underparameterization and limits the performance of fixed depth QAOA. An open question in this regard is the depth required for QAOA to succeed. Additionally, by treating the circuit depth of QAOA as a computational resource, it becomes paramount in complexity studies to know the scaling of depth with respect to problem size. Understanding these open problems paves the way towards identifying quantum advantage in QAOA.

We introduced a methodology based on a predictive model that aims to address these open problems. Furthermore, we recovered, for the first time, the scaling of critical depth with respect to problem size, which illustrates a linear trend in our considered range of problem sizes. Although this finite range is insufficient to assert quantum advantage, we anticipate future work to test the accuracy of the presented model over comprehensive ranges of densities and problem sizes.

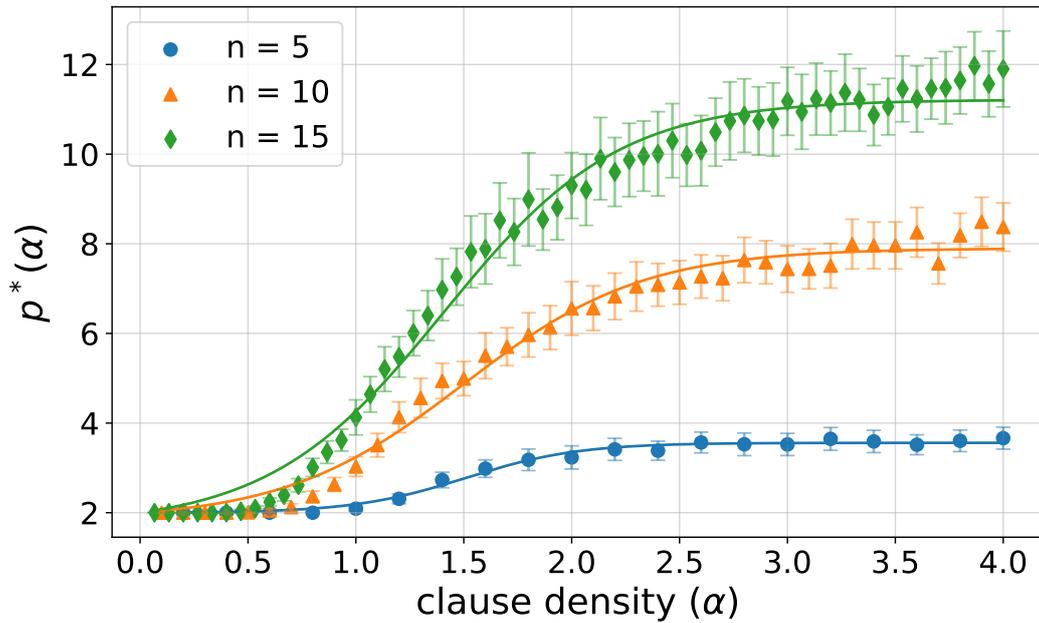


Figure 6-1: Average critical depth  $p^*$  versus clause density. Data points represent the average  $p^*$  calculated according to Eq. (6.2) across problem density for MAX-2-SAT with tolerance  $\epsilon = 0.3$ . Colors correspond to different problem sizes  $n = \{5, 10, 15\}$  with errorbars representing  $3\sigma$  for the estimated mean. Solid curves represent the least squares fit of the data to the predictive model as described in Eq. (6.3).

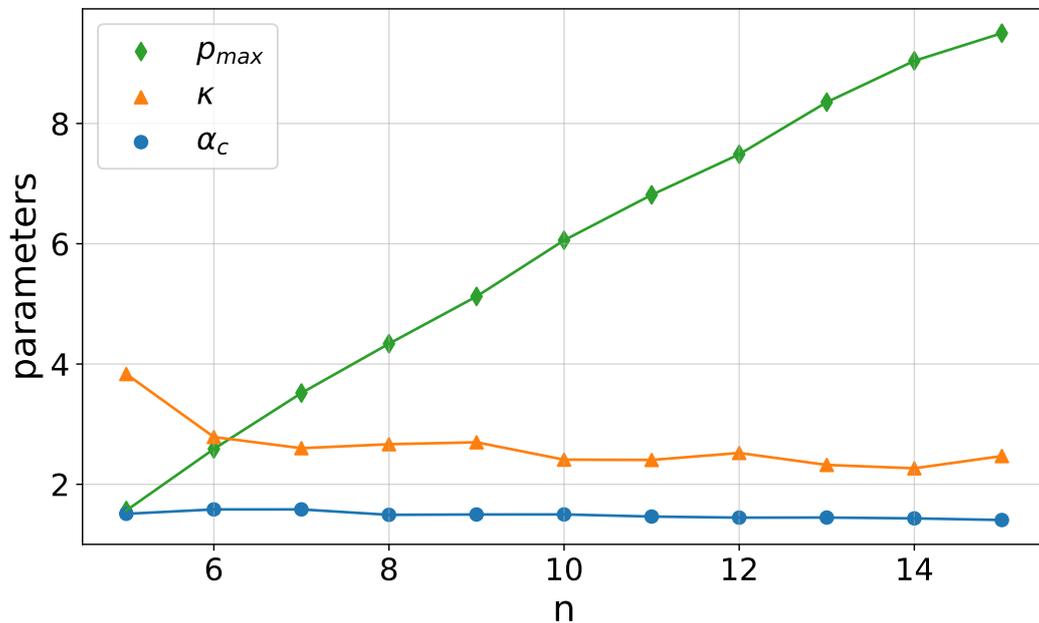


Figure 6-2: Fitting parameters versus number of qubits for QAOA on MAX-2-SAT with tolerance  $\epsilon = 0.3$ . Note that the recovered critical density  $\alpha_c \approx 1$ , and growth rate  $\kappa$ , show very little variability with respect to number of qubits. In contrast, the the saturation value  $p_{max}$  illustrates a linear trend in the considered range of 5 to 15 qubits.

# Conclusion

The modern approach in the development of quantum algorithms is centered on the variational model of quantum computation, wherein a hybrid architecture, composed of a classical coprocessor and a near-term quantum device, is exploited to realize ground state quantum computing. In this framework, for classical combinatorial optimization tasks, QAOA [26] is among the most studied with promising prospects in near-term applications [1, 19, 20, 25–35]. Although recent milestones such as experimental demonstration [36, 37, 112, 113], universality results [27, 28], and prospects for quantum supremacy [32] add to this active interest, many open questions remain regarding the applicability of this approach. In this thesis, we unravel some of its prospects in the NISQ era of quantum computing.

To begin with, one of the open questions within the framework of QAOA is whether quantum advantage over classical algorithms can be attained by near-term devices? By considering a single depth setting and for specific instances, initial results indeed report an optimistic prospect for QAOA. However, the hope of achieving quantum advantage via shallow-depth circuits were quickly questioned when classical algorithms capable of outperforming low-depth QAOA were developed [90]. Furthermore, owing to the black-box nature of the classical outer-loop optimization in variational algorithms, analytical techniques quickly become out of scope, leaving open the question of performance benefits (if any) for QAOA with higher depths and on general problem instances such as in the case of constraint satisfiability.

Here, we identify and reveal a fundamental limitation for QAOA caused by a density induced underparameterization, an effect we term – *reachability deficits*. Specifically, when considering the ratio of problem instances constraints to variables, clause density in the case of satisfiability instances, and graph density in the

case of random graphs, we observe the limiting performance of fixed depth QAOA to correlate inversely with problem density [1, 2]. For relatively low problem density instances, a shallow-depth QAOA circuit is able to recover exact solutions to problem instances. We suspect that it is this low-density subset of problem instances that could be addressed by current NISQ devices [36]. Reachability deficits necessitate the need for employing deeper circuits to address real-world problem instances, typically belonging to the high-density subset. Therefore, practical advantage over classical optimization algorithms may only be achieved via QAOA with hundreds of qubits and, in the worst case, exponentially increasing circuit depth [38]. Such resource requirements are definitely beyond the current scope of NISQ, even excluding the need for error correction and hardware-specific circuit compilations [38].

Another interesting area of consideration in variational algorithms and, by extension, in QAOA, is training parameterized quantum circuits using a classical outer-loop optimization routine. As such, gradient-based and gradient-free methods are employed to iteratively adjust the circuit parameters to attain convergence to the optimum of some objective function. However, this step is challenging beyond low depth, because of an exponentially increasing parameter space to search for an optimum. Furthermore, it is known that for random initialization of circuit parameters and for certain objective functions, variational algorithms may suffer from barren plateaus [88, 89]. Therefore, gradient-based methods may end up failing at the task. Owing to this, several strategies have already been explored in hopes of reducing the training complexity associated with variational algorithms. Some of which include circuit re-parameterization, layer-wise training, and even noise-assisted training [25, 29, 31, 101–103, 114]. Specifically, in the case of QAOA, strategies that employ machine learning models are of particular interest [99]. The motivation behind such strategies is the ample empirical evidence of non-trivial co-relations and/or behaviors for optimal circuit parameters. A milestone result in this regard is on *concentrations* [33, 35, 98], which may be at the heart of these observed non-trivialities.

Here, we identify a subtle effect of concentrations in optimal circuit parameters, an effect we term as *parameter concentrations*. Specifically, considering QAOA on

---

the problem of variational state preparation, we analytically demonstrate that for  $p = 1, 2$ , given a set of optimal parameters for  $n$  qubits, a set of optimal parameters for  $n + 1$  qubits can always be found within a distance bounded by some inverse polynomial in  $n$  [3]. As a consequence, one can leverage parameter concentrations to perform polynomially restrictive training over optimal parameters found at small qubit counts to find optimal parameters for instances with large number of qubits. Such a strategy reduces the training cost and is therefore of practical importance.

Finally, one of the most tantalizing open problems in QAOA is estimating the circuit depth needed for the algorithm to succeed, i.e. recovering approximations close to the global optimum of some problem Hamiltonian and the scaling of this critical depth as a function of problem size. One could argue that the required circuit depth is related to the adiabatic time, since QAOA can be thought of as a discretized adiabatic evolution [67]. However, it is not necessary that QAOA follows an adiabatic pathway due to the freedom of arbitrarily changing circuit parameters via the classical outer loop optimization step. Forcing adiabaticity results in optimal circuit parameters to follow a smooth schedule; however, evidence in violation of adiabatic schedules has been reported [115]. At most, the adiabatic time provides a loose upper bound on the required circuit depth given some tolerance on approximations. Moreover, even in the case of adiabatic quantum computation, it remains unclear how one would estimate the adiabatic time for general NP-Hard optimization problems such as in constraint satisfiability.

Here, we attempt to address part of this open problem for the case of MAX-2-SAT by proposing a predictive model based on the conjecture that the critical depth required for QAOA to guarantee an  $\epsilon$ -tolerant performance grows as a logistic function on the problem density [4]. We observe that our predictive model is well within the confidence of the simulated data. Also, based on the predictive model, we recovered a linear scaling for critical depth with respect to problem size in our considered dataset. Although our results are not enough to assert any quantum advantage, it provides an empirical means to be tested on larger problem sizes to recover and verify this linear scaling. This will pave a path towards identifying the algorithmic complexity associated with QAOA, a problem of paramount interest

which is currently out of scope.

In the thesis, we explored QAOA in a noiseless setting; however, NISQ devices are prone to systematic errors and control limitations [16–18, 58, 116]. Since the classical outer-loop optimization is operationally a black-box, one might expect the variational method to overcome some of the systematic limitations of NISQ. Therefore, it is interesting to perform similar studies and extend our inquiry to noisy regimes. Unsurprisingly, progress in this direction has started to appear in the literature [117]. On the limitation of reachability deficits, an important insight gained is on the expressiveness of the constructed variational state space itself. Unlike variational algorithms such as Variational Quantum Eigensolver (or VQE), where the ansatz construction is problem agnostic [19, 20], the variational state space generated by QAOA is problem specific, potentially hindering the approach. In fact, deviating from typical construction and making the QAOA problem agnostic as VQE, recent results suggest the possibility of improvement [118]. Whether such improvements can free alternating ansatz designs from reachability deficits? is another interesting future scope. From an information theory perspective, recent advances have been made using ideas borrowed from tensor networks [119]. An interesting tool to study variational ansatz is to consider their most rank-efficient matrix product state approximations [120]. This technique allows for a straightforward inquiry into the potential bipartite entanglement that can be generated by variational circuits. Such studies enable and provide a means to probe the role of entanglement in the variational method, an area that is yet to be explored.

## Results

1. QAOA on random MAX-3-SAT and random MAX-2-SAT instances exhibits reachability deficits [1].
2. Onset of deficits for QAOA on random MAX-3-SAT and random MAX-2-SAT instances numerically coincide with the computational phase transition point at problem density 1 [1].

3. QAOA on graph minimization problems – corresponding to minimizing 2-local Ising model with uniform random couplings drawn from  $\{-1, +1\}$ , exhibits reachability deficits [2].
4. Parameters concentrate for  $p = 1, 2$  QAOA on variational state preparation with concentration scaling [3]:

$$|\beta_{n+1}^* - \beta_n^*|^2 + |\gamma_{n+1}^* - \gamma_n^*|^2 = \mathcal{O}(n^{-4}),$$

5. Parameters concentrate for  $p \geq 3$  depth QAOA on variational state preparation for up to  $p = 5$  and 17 qubits with observed scaling being the same as for  $p = 1, 2$  [3].
6. The circuit depth required for QAOA to recover  $\epsilon$  tolerant performance on random MAX-2-SAT instances on up to 15 qubits can empirically be described by a logistic function on problem density [4].

# Bibliography

- [1] V. Akshay, H. Philathong, M.E.S. Morales, and J. Biamonte. Reachability deficits in quantum approximate optimization. *Physical Review Letters*, 124(9):090504, 2020.
- [2] V. Akshay, H. Philathong, I. Zacharov, and J. Biamonte. Reachability deficits in quantum approximate optimization of graph problems. *Quantum*, 5:532, 2021.
- [3] V. Akshay, D. Rabinovich, E. Campos, and J. Biamonte. Parameter concentrations in quantum approximate optimization. *Physical Review A*, 104(1):L010401, 2021.
- [4] V. Akshay, H. Philathong, E. Campos, D. Rabinovich, I. Zacharov, X. Zhang, and J. Biamonte. Circuit depth scaling for quantum approximate optimization. *Physical Review A*, 106:042438, 2022.
- [5] John S Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.
- [6] John F Clauser and Abner Shimony. Bell’s theorem. experimental tests and implications. *Reports on Progress in Physics*, 41(12):1881, 1978.
- [7] Richard P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6-7):467–488, June 1982.
- [8] S. Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, August 1996.
- [9] David Deutsch. Quantum theory, the Church–Turing principle and the universal quantum computer. *Proceedings of the Royal Society of London Series A, Mathematical and Physical Sciences*, 400(1818):97–117, 1985.
- [10] Richard P Feynman. Quantum mechanical computers. *Found. Phys.*, 16(6):507–532, 1986.
- [11] James V Rauff. The golden ticket: P, np, and the search for the impossible. *Mathematics and Computer Education*, 49(2):147, 2015.
- [12] David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907):553–558, December 1992.

- 
- [13] D.R. Simon. On the power of quantum computation. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*. IEEE Comput. Soc. Press, 1994.
- [14] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, October 1997.
- [15] Lov K Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219, 1996.
- [16] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, Aug 2018.
- [17] Peter W. Shor. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A*, 52:R2493–R2496, Oct 1995.
- [18] Iulia Georgescu. 25 years of quantum error correction. *Nature Reviews Physics*, 2(10):519–519, September 2020.
- [19] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.
- [20] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.*, 94:015004, Feb 2022.
- [21] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.
- [22] A Kardashin, A Uvarov, and J Biamonte. Quantum machine learning tensor network states. *Frontiers in Physics*, 8:586374, 2021.
- [23] A V Uvarov, A S Kardashin, and J D Biamonte. Machine learning phase transitions with a quantum processor. *Physical Review A*, 102:012415, 2020.
- [24] M E S Morales, T Tlyachev, and J Biamonte. Variational learning of grover’s quantum search algorithm. *Physical Review A*, 98:062333, 2018.
- [25] Murphy Yuezhen Niu, Sirui Lu, and Isaac L Chuang. Optimizing qaoa: Success probability and runtime dependence on circuit depth. *arXiv preprint arXiv:1905.12134*, May 2019.
- [26] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014.

- 
- [27] Seth Lloyd. Quantum approximate optimization is computationally universal. *arXiv preprint arXiv:1812.11075*, 2018.
- [28] Mauro ES Morales, JD Biamonte, and Zoltán Zimborás. On the universality of the quantum approximate optimization algorithm. *Quantum Information Processing*, 19(9):1–26, 2020.
- [29] Leo Zhou, Sheng-Tao Wang, Soonwon Choi, Hannes Pichler, and Mikhail D Lukin. Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices. *Physical Review X*, 10(2):021067, 2020.
- [30] Zhihui Wang, Nicholas C Rubin, Jason M Dominy, and Eleanor G Rieffel. X y mixers: Analytical and numerical results for the quantum alternating operator ansatz. *Physical Review A*, 101(1):012320, 2020.
- [31] Lucas T. Brady, Christopher L. Baldwin, Aniruddha Bapat, Yaroslav Kharkov, and Alexey V. Gorshkov. Optimal Protocols in Quantum Annealing and Quantum Approximate Optimization Algorithm Problems. *Physical Review Letters*, 126(7):070505, Feb 2021.
- [32] Edward Farhi and Aram W Harrow. Quantum supremacy through the quantum approximate optimization algorithm. *arXiv preprint arXiv:1602.07674*, 2016.
- [33] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Leo Zhou. The quantum approximate optimization algorithm and the sherrington-kirkpatrick model at infinite size. *arXiv preprint arXiv:1910.08187*, Oct 2019.
- [34] Matteo M. Wauters, Glen Bigan Mbeng, and Giuseppe E. Santoro. Polynomial scaling of QAOA for ground-state preparation of the fully-connected p-spin ferromagnet. *arXiv e-prints*, page arXiv:2003.07419, March 2020.
- [35] Jahan Claes and Wim van Dam. Instance Independence of Single Layer Quantum Approximate Optimization Algorithm on Mixed-Spin Models at Infinite Size. *arXiv e-prints*, page arXiv:2102.12043, February 2021.
- [36] Matthew P Harrigan, Kevin J Sung, Matthew Neeley, Kevin J Satzinger, Frank Arute, Kunal Arya, Juan Atalaya, Joseph C Bardin, Rami Barends, Sergio Boixo, et al. Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. *Nature Physics*, 17(3):332–336, 2021.
- [37] Guido Pagano, Aniruddha Bapat, Patrick Becker, Katherine S. Collins, Arinjoy De, Paul W. Hess, Harvey B. Kaplan, Antonis Kyprianidis, Wen Lin Tan, Christopher Baldwin, Lucas T. Brady, Abhinav Deshpande, Fangli Liu, Stephen Jordan, Alexey V. Gorshkov, and Christopher Monroe. Quantum approximate optimization of the long-range ising model with a trapped-ion quantum simulator. *Proceedings of the National Academy of Sciences*, 117(41):25396–25401, 2020.

- 
- [38] Gian Giacomo Guerreschi and Anne Y Matsuura. Qaoa for max-cut requires hundreds of qubits for quantum speed-up. *Scientific reports*, 9(1):1–7, 2019.
- [39] Anastasiia Butko, George Micheliogiannakis, Samuel Williams, Costin Iancu, David Donofrio, John Shalf, Jonathan Carter, and Irfan Siddiqi. Understanding quantum control processor capabilities and limitations through circuit characterization. In *2020 International Conference on Rebooting Computing (ICRC)*, pages 66–75. IEEE, 2020.
- [40] Zhihui Wang, Stuart Hadfield, Zhang Jiang, and Eleanor G Rieffel. Quantum approximate optimization algorithm for maxcut: A fermionic view. *Physical Review A*, 97(2):022304, 2018.
- [41] Alexander K Hartmann and Martin Weigt. *Phase transitions in combinatorial optimization problems: basics, algorithms and statistical mechanics*. John Wiley & Sons, 2006.
- [42] Michael Sipser. Introduction to the theory of computation. *ACM Sigact News*, 27(1):27–29, 1996.
- [43] Lance Fortnow. The status of the p versus np problem. *Communications of the ACM*, 52(9):78–86, 2009.
- [44] Joseph D Bryngelson and Peter G Wolynes. Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of sciences*, 84(21):7524–7528, 1987.
- [45] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.
- [46] Jean-Philippe Bouchaud. Crises and collective socio-economic phenomena: simple models and challenges. *Journal of Statistical Physics*, 151(3):567–606, 2013.
- [47] Andrew Lucas and Ching Hua Lee. Multistable binary decision making on networks. *Physical Review E*, 87(3):032806, 2013.
- [48] Niles A Pierce and Erik Winfree. Protein design is np-hard. *Protein engineering*, 15(10):779–782, 2002.
- [49] J.D. Ullman. Np-complete scheduling problems. *Journal of Computer and System Sciences*, 10(3):384–393, 1975.
- [50] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972.
- [51] Francisco Barahona. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):3241, 1982.

- 
- [52] Scott Kirkpatrick, Dg jr., and mp vecchi. *Optimization by simmulated annealing. science*, 220(4598):671–680, 1983.
- [53] Feng Xia, Jiaying Liu, Hansong Nie, Yonghao Fu, Liangtian Wan, and Xiangjie Kong. Random walks: A review of algorithms and applications. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 4(2):95–107, 2019.
- [54] Shoko Utsunomiya, Kenta Takata, and Yoshihisa Yamamoto. Mapping of ising models onto injection-locked laser systems. *Optics express*, 19(19):18091–18108, 2011.
- [55] Takahiro Inagaki, Yoshitaka Haribara, Koji Igarashi, Tomohiro Sonobe, Shuhei Tamate, Toshimori Honjo, Alireza Marandi, Peter L McMahan, Takeshi Umeki, Koji Enbutsu, et al. A coherent ising machine for 2000-node optimization problems. *Science*, 354(6312):603–606, 2016.
- [56] D Pierangeli, G Marcucci, and C Conti. Large-scale photonic ising machine by spatial light modulation. *Physical review letters*, 122(21):213902, 2019.
- [57] Micha Nixon, Eitan Ronen, Asher A Friesem, and Nir Davidson. Observing geometric frustration with thousands of coupled lasers. *Physical review letters*, 110(18):184102, 2013.
- [58] Rami Barends, Alireza Shabani, Lucas Lamata, Julian Kelly, Antonio Mezzacapo, U Las Heras, Ryan Babbush, Austin G Fowler, Brooks Campbell, Yu Chen, et al. Digitized adiabatic quantum computing with a superconducting circuit. *Nature*, 534(7606):222–226, 2016.
- [59] James Daniel Whitfield, Mauro Faccin, and JD Biamonte. Ground-state spin logic. *EPL (Europhysics Letters)*, 99(5):57004, 2012.
- [60] Stephen A Cook. The complexity of theorem-proving procedures. In *Proceedings of the third annual ACM symposium on Theory of computing*, pages 151–158, 1971.
- [61] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM (JACM)*, 48(4):798–859, 2001.
- [62] Andrew Lucas. Ising formulations of many np problems. *Frontiers in physics*, page 5, 2014.
- [63] James M Crawford and Larry D Auton. Experimental results on the crossover point in random 3-sat. *Artificial intelligence*, 81(1-2):31–57, 1996.
- [64] Don Coppersmith, David Gamarnik, MohammadTaghi Hajiaghayi, and Gregory B Sorkin. Random max sat, random max cut, and their phase transitions. *Random Structures & Algorithms*, 24(4):502–545, 2004.
- [65] Ernst Ising. Contribution to the theory of ferromagnetism. *Z. Phys*, 31(1):253–258, 1925.

- 
- [66] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum computation by adiabatic evolution. *arXiv preprint quant-ph/0001106*, 2000.
- [67] Wim Van Dam, Michele Mosca, and Umesh Vazirani. How powerful is adiabatic quantum computation? In *Proceedings 42nd IEEE symposium on foundations of computer science*, pages 279–287. IEEE, 2001.
- [68] Max Born and Vladimir Fock. Beweis des adiabatenatzes. *Zeitschrift für Physik*, 51(3-4):165–180, 1928.
- [69] John A Mydosh. *Spin glasses: an experimental introduction*. CRC Press, 1993.
- [70] PA Joy, PS Anil Kumar, and SK Date. The relationship between field-cooled and zero-field-cooled susceptibilities of some ordered magnetic systems. *Journal of physics: condensed matter*, 10(48):11049, 1998.
- [71] P Nordblad, L Lundgren, and L Sandlund. A link between the relaxation of the zero field cooled and the thermoremanent magnetizations in spin glasses. *Journal of magnetism and magnetic materials*, 54:185–186, 1986.
- [72] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754, 1979.
- [73] Naeimeh Mohseni, Peter L McMahon, and Tim Byrnes. Ising machines as hardware solvers of combinatorial optimization problems. *Nature Reviews Physics*, 4(6):363–379, 2022.
- [74] Alan Mathison Turing. On computable numbers, with an application to the entscheidungsproblem. a correction. *Proceedings of the London Mathematical Society*, 2(1):544–546, 1938.
- [75] Martin Davis. *The universal computer: The road from Leibniz to Turing*. AK Peters/CRC Press, 2018.
- [76] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.
- [77] A Yu Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191, 1997.
- [78] Dorit Aharonov, Wim Van Dam, Julia Kempe, Zeph Landau, Seth Lloyd, and Oded Regev. Adiabatic quantum computation is equivalent to standard quantum computation. *SIAM review*, 50(4):755–787, 2008.
- [79] Jacob Biamonte. Universal variational quantum computation. *Physical Review A*, 103(3):L030401, 2021.
- [80] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, 2016.

- 
- [81] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O’Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5(1):1–7, 2014.
- [82] Guillaume Verdon, Michael Broughton, and Jacob Biamonte. A quantum algorithm to train neural networks using low-depth circuits. *arXiv preprint arXiv:1712.05304*, 2017.
- [83] Stuart Hadfield, Zihui Wang, Bryan O’gorman, Eleanor G Rieffel, Davide Venturelli, and Rupak Biswas. From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. *Algorithms*, 12(2):34, 2019.
- [84] Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.
- [85] Tosio Kato. On the adiabatic theorem of quantum mechanics. *Journal of the Physical Society of Japan*, 5(6):435–439, 1950.
- [86] Zhang Jiang, Eleanor G. Rieffel, and Zihui Wang. Near-optimal quantum circuit for grover’s unstructured search using a transverse field. *Phys. Rev. A*, 95:062317, Jun 2017.
- [87] 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(2):021027, 2017.
- [88] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9(1):4812, 2018.
- [89] AV Uvarov and Jacob D Biamonte. On barren plateaus and cost function locality in variational quantum algorithms. *Journal of Physics A: Mathematical and Theoretical*, 54(24):245301, 2021.
- [90] Matthew B Hastings. Classical and quantum bounded depth approximation algorithms. *arXiv preprint arXiv:1905.07047*, 2019.
- [91] Lennart Bittel and Martin Kliesch. Training variational quantum algorithms is np-hard. *Physical Review Letters*, 127(12):120502, 2021.
- [92] Ernesto Campos, Aly Nasrallah, and Jacob Biamonte. Abrupt transitions in variational quantum circuit training. *Phys. Rev. A*, 103:032607, Mar 2021.
- [93] Roger Fletcher. *Practical methods of optimization*. John Wiley & Sons, 2013.
- [94] Weixiong Zhang. Phase transitions and backbones of 3-sat and maximum 3-sat. In *Principles and Practice of Constraint Programming—CP 2001: 7th International Conference, CP 2001 Paphos, Cyprus, November 26–December 1, 2001 Proceedings 7*, pages 153–167. Springer, 2001.

- 
- [95] Olivier C Martin, Rémi Monasson, and Riccardo Zecchina. Statistical mechanics methods and phase transitions in optimization problems. *Theoretical computer science*, 265(1-2):3–67, 2001.
- [96] Juan J de Pablo, Qiliang Yan, and Fernando A Escobedo. Simulation of phase transitions in fluids. *Annual review of physical chemistry*, 50(1):377–411, 1999.
- [97] Hector Gomez, Miguel Bures, and Adrian Moure. A review on computational modelling of phase-transition problems. *Philosophical Transactions of the Royal Society A*, 377(2143):20180203, 2019.
- [98] Fernando GSL 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. *arXiv preprint arXiv:1812.04170*, Dec 2018.
- [99] Michael Streif and Martin Leib. Training the quantum approximate optimization algorithm without access to a quantum processing unit. *Quantum Science and Technology*, 5(3):34008, Jul 2020.
- [100] Xuchen You and Xiaodi Wu. Exponentially many local minima in quantum neural networks. In *International Conference on Machine Learning*, pages 12144–12155. PMLR, 2021.
- [101] Stefan H Sack and Maksym Serbyn. Quantum annealing initialization of the quantum approximate optimization algorithm. *arXiv preprint arXiv:2101.05742*, Jan 2021.
- [102] Jun Li, Xiaodong Yang, Xinhua Peng, and Chang Pu Sun. Hybrid Quantum-Classical Approach to Quantum Optimal Control. *Physical Review Letters*, 118(15):150503, Apr 2017.
- [103] E. Campos, D. Rabinovich, V. Akshay, and J. Biamonte. Training saturation in layerwise quantum approximate optimization. *Physical Review A*, 104(3):L030401, 2021.
- [104] Gavin E Crooks. Performance of the quantum approximate optimization algorithm on the maximum cut problem. *arXiv preprint arXiv:1811.08419*, Nov 2018.
- [105] Dmitry Panchenko. *The sherrington-kirkpatrick model*. Springer Science & Business Media, 2013.
- [106] G Parisi and Travis Lange. private communication; c. *Rebbi et al., private communication*, 1986.
- [107] Michael Aizenman, Joel L Lebowitz, and David Ruelle. Some rigorous results on the sherrington-kirkpatrick spin glass model. *Communications in mathematical physics*, 112(1):3–20, 1987.

- 
- [108] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm applied to a bounded occurrence constraint problem. *arXiv preprint arXiv:1412.6062*, 2014.
- [109] Eric Anschuetz, Jonathan Olson, Alán Aspuru-Guzik, and Yudong Cao. Variational quantum factoring. In *International Workshop on Quantum Technology and Optimization Problems*, pages 74–85. Springer, 2019.
- [110] Wen Wei Ho and Timothy H Hsieh. Efficient variational simulation of non-trivial quantum states. *SciPost Phys*, 6:29, 2019.
- [111] Dave Wecker, Matthew B Hastings, and Matthias Troyer. Training a quantum optimizer. *Physical Review A*, 94(2):022309, 2016.
- [112] Christopher Monroe, Wes C Campbell, L-M Duan, Z-X Gong, Alexey V Gorshkov, Paul W Hess, Rajibul Islam, Kihwan Kim, Norbert M Linke, Guido Pagano, et al. Programmable quantum simulations of spin systems with trapped ions. *Reviews of Modern Physics*, 93(2):025001, 2021.
- [113] Anders Sørensen and Klaus Mølmer. Quantum computation with ions in thermal motion. *Physical review letters*, 82(9):1971, 1999.
- [114] Ernesto Campos, Aly Nasrallah, and Jacob Biamonte. Abrupt transitions in variational quantum circuit training. *Phys. Rev. A*, 103:032607, Mar 2021.
- [115] Lucas T Brady, Christopher L Baldwin, Aniruddha Bapat, Yaroslav Kharkov, and Alexey V Gorshkov. Optimal protocols in quantum annealing and qaoa problems, 2020. URL <https://arxiv.org/abs/2003.08952>. *arXiv*, 2003.
- [116] Shantanu Debnath, Norbert M Linke, Caroline Figgatt, Kevin A Landsman, Kevin Wright, and Christopher Monroe. Demonstration of a small programmable quantum computer with atomic qubits. *Nature*, 536(7614):63–66, 2016.
- [117] Enrico Fontana, Nathan Fitzpatrick, David Muñoz Ramo, Ross Duncan, and Ivan Rungger. Evaluating the noise resilience of variational quantum algorithms. *Physical Review A*, 104(2):022403, 2021.
- [118] D. Rabinovich, S. Adhikary, E. Campos, V. Akshay, E. Anikin, R. Sengupta, O. Lakhmanskaya, K. Lakhmanskiy, and J. Biamonte. Ion-native variational ansatz for quantum approximate optimization. *Physical Review A*, 106:032418, 2022.
- [119] Jacob Biamonte and Ville Bergholm. Tensor networks in a nutshell. *arXiv preprint arXiv:1708.00006*, 2017.
- [120] Tai-Danae Bradley, E Miles Stoudenmire, and John Terilla. Modeling sequences with quantum states: a look under the hood. *Machine Learning: Science and Technology*, 1(3):035008, 2020.

# Appendix A

## QAOA software implementation

The numerical results compiled in the dissertation are based on data generated by an in-house implementation of QAOA. We perform a ideal state-vector emulation using the standard packages and libraries of Python, specifically NumPy and SciPy.

### A.1 Ansatz implementation

Let us recall a depth  $p$ , QAOA ansatz  $|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle$ , we wish to implement:

$$|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = \prod_{k=1}^p U(\gamma_k, \beta_k) |+\rangle^{\otimes n}, \quad (\text{A.1})$$

where

$$U(\gamma_k, \beta_k) = \exp(-i\beta_k H_x) \cdot \exp(-i\gamma_k H). \quad (\text{A.2})$$

The Hamiltonians  $H_x = \sum_{k=1}^n X_k$ , is the mixer Hamiltonian and  $H$  the objective Hamiltonian to be minimized; typically diagonal in the computational basis.

1. **Plus state:** The initial state in Eq. (A.1) is trivially implemented as a NumPy 1D array of ones of size  $2^n$  and shape  $(2^n, 1)$ . The array is then multiplied by the normalization factor,  $\frac{1}{\sqrt{2^n}}$  and stored in the memory.
2. **Action of  $\exp(-i\gamma_k H)$ :** Since  $H$  is diagonal, the matrix exponential is trivially calculated and stored in the memory as a NumPy 1D array of size  $2^n$  and shape

$(2^n, 1)$ . The action of  $\exp(-i\gamma_k H)$  on a state can then be implemented as an element-wise multiplication of 2 1D arrays.

3. **Action of  $\exp(-i\beta_k H_x)$ :** Observe that:

$$\exp(-i\beta_k H_x) = \prod_{j=1}^n (\cos(\beta_k) \mathbf{1} - i \sin(\beta_k) X_j), \quad (\text{A.3})$$

where

$$X_j = \mathbf{1}_1 \otimes \cdots \otimes X_j \cdots \otimes \mathbf{1}_n. \quad (\text{A.4})$$

Matrices of the kind in Eq. (A.4) are permutation matrices and therefore the action of such matrices on a state can easily be realized using array reorderings. Action of  $\exp(-i\beta_k H_x)$  can therefore be realized sequentially as in Eq. (A.3) using appropriate scalar multiplications, array additions and array reorderings without the need for matrix multiplication or matrix exponentiation.

For a given set of parameters  $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ ,  $|\boldsymbol{\gamma}| = |\boldsymbol{\beta}| = p$ , steps 2 and 3 can repeatedly be applied to create ansatz state as in Eq. (A.1) which is then stored in the memory as a 1D array of size  $2^n$  and shape  $(2^n, 1)$ .

## A.2 Layerwise heuristic optimization strategy

The classical step in variational algorithms and thus for QAOA is the outer-loop optimization routine. This classical-quantum feedback loop iteratively adjusts the circuit parameters thereby allowing for a solution search process over the variational state space (see Eq. (3.3)). In QAOA, a  $p$ -depth ansatz is prescribed by a total of  $2p$  variational parameters. By increasing depth, one encounters an exponentially large parameter space in searching for a candidate optimum. Therefore naively searching for global optima quickly becomes challenging. Furthermore, due to the phenomenon of *barren plateaus* [88, 89], gradient based optimizers may end up failing at the task. Owing to these factors, in our software implementation of QAOA, we explore a heuristic optimization strategy motivated by layer-wise training. Consider a fixed depth  $p \in \mathbb{N}$ :

1. Start with layer  $l = 1$ . Using best out of 50 random initial guesses  $\{(\bar{\gamma}_1, \bar{\beta}_1)\}$ , the standard L-BFGS-B optimizer (a finite-difference based gradient optimizer) from SciPy finds optimal parameters:

$$(\gamma_1^*, \beta_1^*) \in \arg \min_{(\gamma_1, \beta_1)} \langle \psi(\gamma_1, \beta_1) | \mathbf{H} | \psi(\gamma_1, \beta_1) \rangle. \quad (\text{A.5})$$

2. Considering the next layer  $l = 2$ , and using 25 random guesses  $\{(\bar{\gamma}_2, \bar{\beta}_2)\}$ , we again search for optimal parameters:

$$(\gamma_2^*, \beta_2^*) \in \arg \min_{(\gamma_2, \beta_2)} \langle \psi(\gamma_1^*, \gamma_2, \beta_1^*, \beta_2) | \mathbf{H} | \psi(\gamma_1^*, \gamma_2, \beta_1^*, \beta_2) \rangle. \quad (\text{A.6})$$

3. Using the layerwise trained parameters,  $(\gamma_1^*, \gamma_2^*, \beta_1^*, \beta_2^*)$  as an initial guess  $(\bar{\gamma}_{1,2}, \bar{\beta}_{1,2})$ , we perform a round of simultaneous optimization to yield new parameters:

$$(\gamma_{1,2}^*, \beta_{1,2}^*) \in \arg \min_{(\gamma_{1,2}, \beta_{1,2})} \langle \psi(\gamma_{1,2}, \beta_{1,2}) | \mathbf{H} | \psi(\gamma_{1,2}, \beta_{1,2}) \rangle. \quad (\text{A.7})$$

4. Steps 2 and 3 are then repeated for subsequent layers, following the procedure of concatenating simultaneously optimized parameters and layerwise trained parameters as initial guesses until the required depth  $l = p$  is attained.

### A.3 Software benchmarks

Here we benchmark the time to construct the full state-vector by our implementation of QAOA. We consider 5 upto 15 qubits and QAOA depths of upto 50 to assess the run-time of the implementation. Since the implementation avoids matrix multiplication and exponentiation, the run-times scales linearly with QAOA depth,  $p$  and exponentially with problem sizes,  $n$ . The exponential behaviour is expected due to array sizes scaling as  $2^n$  in memory.

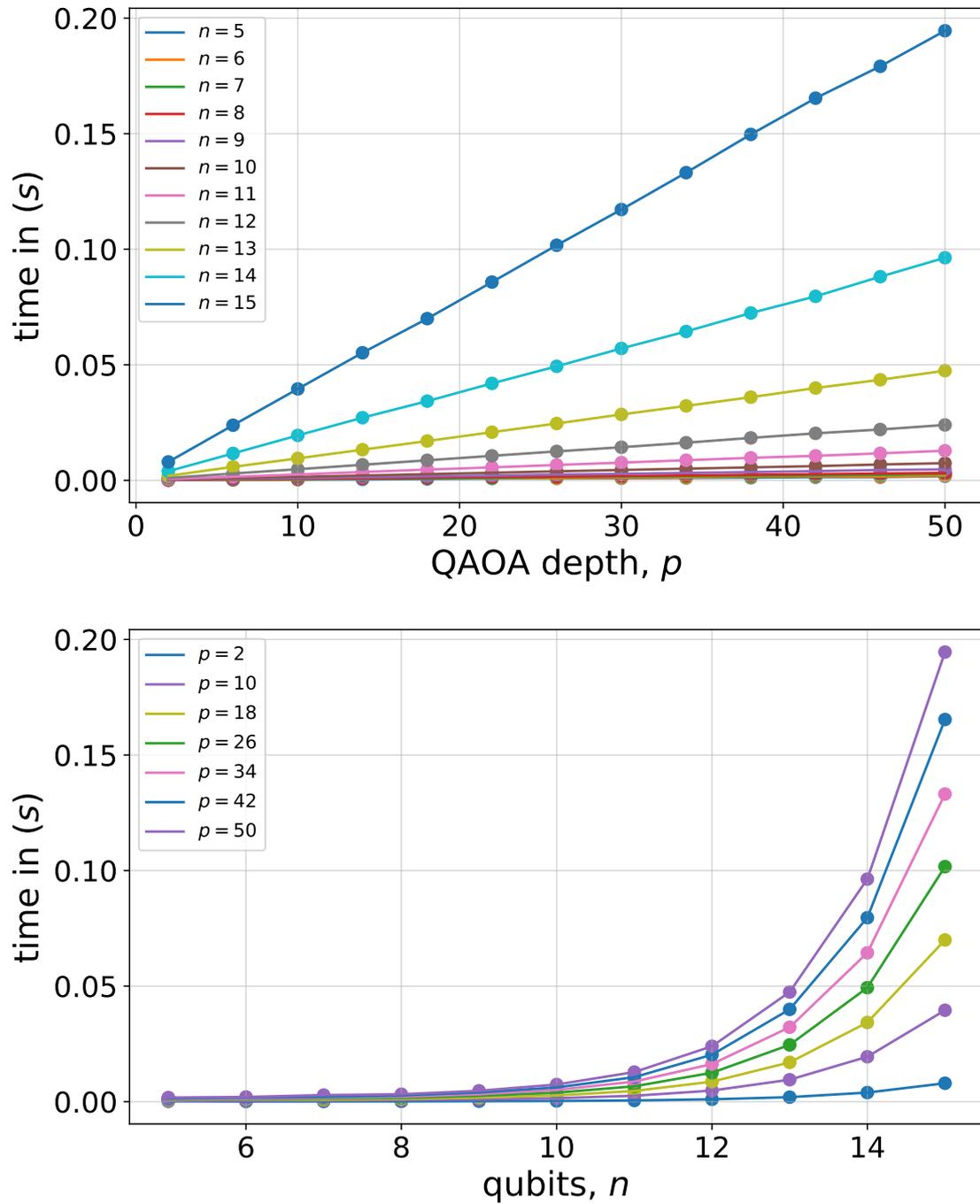


Figure A-1: Run-time (in seconds) for generating the full state-vector as a function of QAOA depth,  $p$  and problem size,  $n$ . Dots represent average times computed for a statistic of 100 cases at each depth and problem size.