ECOLE
POLYTECHNIQUE
DE BRUXELLES

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Randomized Quantum Approximate Optimization Algorithm

Mémoire présenté en vue de l'obtention du diplôme
d'Ingénieur Civil Physicien à finalité spécialisée

Jules Dumont

Promoteur
Professeur Jérémie Roland

Superviseur
Benoît Dubus

Service
Centre of Quantum Information and Quantum Communication

Année académique
2025 - 2026

ECOLE
POLYTECHNIQUE
DE BRUXELLES

DOCUMENT A INSERER DANS LE MÉMOIRE DE FIN D'ETUDES

CONSULTATION DU MÉMOIRE DE FIN D'ETUDES

Je soussigné

NOM ..... Dumont .....

PRENOM ..... Jules .....

MASTER ..... Physical Engineering .....

TITRE DU MÉMOIRE ..... Randomized Quantum Approximate Optimization Algorithm .....

.....

.....

.....

AUTORISE*

~~**REFUSE***~~

(EN CAS DE REFUS, CE DOCUMENT DEVRA ÊTRE CONTRESIGNÉ PAR LE PROMOTEUR DU MÉMOIRE DE FIN D'ÉTUDES)

(Biffer la mention inutile)*

La consultation du présent mémoire/travail de fin d'études par les utilisateurs des bibliothèques de l'Université libre de Bruxelles.

Si la consultation est autorisée, le soussigné concède par la présente à l'Université libre de Bruxelles, pour toute la durée légale de protection de l'œuvre, une licence gratuite et non exclusive de reproduction et de communication au public de son œuvre précisée ci-dessus, sur supports graphiques ou électroniques, afin d'en permettre la consultation par les utilisateurs des bibliothèques de l'ULB et d'autres institutions dans les limites du prêt inter-bibliothèques.

Bruxelles, le ..... 6 janvier 2026 .....

Signature de l'étudiant

Signature du promoteur
(uniquement en cas de refus de consultation)

Abstract

Randomized Quantum Approximate Optimization Algorithm
by **Jules Dumont**

for the degree of Master of Science in Physical Engineering, 2025-2026

The theoretical analysis of the Quantum Approximate Optimization Algorithm (QAOA) is currently hindered by a significant computational bottleneck: the exact evaluation of the expected cost function scales exponentially with the circuit depth, rendering the analysis of intermediate-depth regimes intractable. This thesis introduces a novel randomized framework to address this limitation, replacing the standard fixed-parameter paradigm with a continuous-time Markovian control process.

Our methodology frames the optimization task as a continuous-time control problem. We model the alternation between Hamiltonians using a Markov chain. The evolution of the system is driven by transition rates, which in turn define the statistical distribution of the Hamiltonian switching instants. To resolve the ill-posed nature of the resulting functional optimization, we employ entropy regularization. This maps the optimization onto a statistical mechanics problem, where the optimal parameter distribution corresponds to a Boltzmann distribution defined by a "multi-switch partition function".

For the single-switch case (depth $p = 1$), we derive exact analytical solutions for the expected cost in terms of modified Bessel functions, identifying resonant switching times that minimize it. We then attempt to generalize this to the multi-switch case (depth $p > 1$). By analyzing the specific case of Unstructured Search, we demonstrate that the multi-switch energy landscape becomes non-separable and recursively complex, limiting the feasibility of exact integration.

Leveraging a high-temperature expansion of our randomized framework, we establish that the initial return on optimization effort is driven entirely by the variance of the energy landscape. We operationalize this by linking thermodynamic variance to the spectral power of the expected cost function. This leads to the development of a polynomial-time classical Monte Carlo diagnostic. We prove that the survival probability of a random walk through the operator tree is mathematically equivalent to the Quantum Signal-to-Noise Ratio (SNR), enabling the a priori detection of Barren Plateaus in dense problem instances at polynomial classical cost.

Keywords: QAOA, Markov Chains, Entropy Regularization, Trainability Score, Barren Plateaus, Signal-to-Noise Ratio.

Acknowledgments

Before anything, I would like to express my gratitude to Prof. Roland for his guidance and support, for agreeing to be my supervisor and his inspiring passion. I would also like to thank Benoît Dubus for his continuous support as my advisor, for his patience and his availability throughout the year.

Next, I would like to thank my chosen family and many brothers. Through their constant support, believing in me and taking the time to listen to my never-ending rants about science, but mostly by being my biggest source of inspiration, through their respective work and simply as amazing human beings, they made this journey possible. I would like to thank in particular Nicholas B. for being the best friend and science buddy one could ever dream of and Merlin G. for his attentive listening and amazing hosting at the Kiosque during the last phase of writing. One last special thought goes to Théo H., whose determination continues to inspire me to this day, even though he is no longer with us.

Finally, I would like to thank my parents for their unconditional support and love, for never putting any pressure on me and for always encouraging me to follow my own path, while providing everything I needed to do so.

Reviewer Feedback & Disclaimer

This section compiles the feedback received during the evaluation of this thesis. The author acknowledges these critiques and presents them here for transparency.

1. General Assessment and Motivation

- **Unresolved Analytical Complexity:** The thesis is motivated by the need to overcome the "exponential resource barrier" of exact recursive analysis. However, the proposed randomized framework does not fully solve this. The author admits in Chapter 5 that the multi-switch energy landscape becomes "non-separable and recursively complex, limiting the feasibility of exact integration". Thus, the original goal of bypassing the bottleneck for intermediate depths remains largely unfulfilled.
- **Terminology (Trotterization vs. Discretization):** The author appears to conflate discretization with Trotterization, stating that to "utilize the insights of AQC... the continuous evolution must be discretized" and "The standard method to achieve this is Trotterization". This overlooks that discretization is the general concept, while Trotterization is a specific approximation for non-commuting Hamiltonians.
- **Terminology (Random vs. Probabilistic):** The thesis uses the term "random" throughout, whereas "probabilistic" might be more accurate for the intended meaning (like measurement).

2. Literature and References

- **Missing Citations:** Several key references are absent. While the term "NISQ" is used, the foundational Preskill paper is not in the bibliography. Similarly, regarding Adiabatic Quantum Computation (AQC), early authors like Nishimori, de Falco, or Apolloni are missing.
- **Search Speedup Claim:** In the introduction and abstract, the author claims to "demonstrate... a quadratic speedup at low depth". Later, the author asserts this result "hasn't been found in the QAOA literature to the author's knowledge". This overlooks existing work: Jiang et al. (2017) PRA, regarding QAOA for unstructured search.

3. Mathematical and Notation Issues

- **Inconsistent Notation:** The notation is difficult to follow. For example, the generator matrix notation switches conventions (noted as Q then A). Furthermore, Pauli matrices are inconsistently denoted, switching between σ_x in Chapter 4 and X_u in Chapter 6.
- **DC Term:** The concept of the "DC component" is employed in the spectral analysis without a proper introduction to the Fourier series context or a definition of the zero-frequency mode.

4. Technical Critique: Chapters 3 & 4 (Single-Switch)

- **Limited Generalizability:** While the development of the Markovian framework is interesting, the analytical results in Chapter 4 are explicitly derived for a "2-level system" (single qubit). The derivation of the partition function using Bessel functions and the subsequent thermodynamic analysis rely on this specific 1-qubit energy landscape, making the results "quite poor" in terms of generalization to N-qubit systems.
- **High-Temperature Limit:** In Section 4.3.6, the stochastic average converges to $1/4$ in the high-temperature limit. A deeper discussion on why this specific value is reached would be beneficial.
- **Visualization:** The thesis lacks plots to visualize key results, such as the asymptotic behavior of Eq. (4.64).

5. Technical Critique: Chapter 5 (Multi-Switch)

- **Invalid Scaling for Speedup:** The derivation of the quadratic speedup scaling $P_p \approx (2p+1)^2/N$ relies explicitly on the approximation that $p \ll n$. Therefore, extrapolating this to the regime where $p \approx \sqrt{N} = 2^{n/2}$ (where $p \gg n$) to claim a Grover-like speedup is mathematically inconsistent.
- **Precision Penalty:** Equation (5.32) suggests the stochastic average increases by a penalty term $k\alpha/2$. Since k scales with depth, this implies performance degrades as layers are added, which contradicts the goal of using deeper circuits.

6. Technical Critique: Chapter 6 (Spectral Analysis)

- **Variance Proxy:** The author argues that the "Trainability Score" (variance of $\mathcal{C}$) is a sufficient proxy for the variance of the gradient. This lacks rigorous justification regarding what is lost by ignoring the gradient variance, particularly in "narrow gorge" landscapes where the cost variance might be small but gradients are high.
- **Robustness:** The claim that the diagnostic is "extremely robust" is asserted without sufficient justification regarding potential false positives.
- **Chebyshev Inequality:** Equation (6.7) provides a bound based on SNR_Q. However, if the SNR_Q is large, the inequality provides no useful information (as probability is always ≤ 1), rendering it useful only in the specific regime where $SNR_Q < 1$.
- **Tree Visualization:** The "recursive branching algorithm" describes a complex operator tree structure. A schematic diagram is missing, which hinders visualization.
- **Summation Definition:** In Equation (6.16), it is ambiguous whether the sum iterates over all generated paths or strictly the leaves that survive the "Expectation Filtering" process described later.

Contents

Chapter 1

Introduction

The recognition that computation is fundamentally a physical process suggests that systems governed by quantum mechanics can offer intrinsic computational advantages over classical models. As originally noted by (Feynman 1982), classical simulation of quantum mechanical systems faces an exponential resource barrier. This observation suggests that a processor governed by quantum mechanics could offer a fundamental computational advantage (Deutsch 1985). While the field has largely standardized around the discrete *gate-based model* (also called circuit model), alternative paradigms have emerged specifically for optimization tasks. Notably, (Farhi, Goldstone, Gutmann, and Sipser 2000) introduced the framework of *Adiabatic Quantum Computation* (AQC). Unlike the discrete steps of the circuit model, AQC is a model of computation based on the *Adiabatic theorem* (Section 2.3.4). The theorem, according to (Born and Fock 1928), states that a quantum system remains in the ground state under the evolution of a slowly varying Hamiltonian (Section 2.1.2). AQC leverages the *Adiabatic theorem* by starting initially in the ground state of a Hamiltonian H_0 (a Hamiltonian and ground state easy to construct). The procedure then consists of evolving the system under a time-dependent Hamiltonian

$$H(t) = (1 - s(t))H_0 + s(t)H_C$$

with $s(t)$ the schedule, slowly varying from 0 to 1 and interpolating between H_0 and the problem Hamiltonian H_C (encoding the optimization problem). According to the *Adiabatic theorem*, the final state after a sufficiently slow evolution is the ground state of H_C, encoding the optimum (see Section 2.3). However, strictly adhering to the adiabatic condition requires coherence times often exceeding the capabilities of current hardware. Near-term quantum hardware (also called a NISQ device, for Noisy Intermediate-Scale Quantum) relies primarily on the standard circuit model and is subject to decoherence, noise, and limited coherence times. To utilize the insights of AQC on standard gate-based hardware, the continuous evolution must be discretized. The standard method to achieve this is *Trotterization* (based on the Lie-Trotter product formula) (Nielsen and Chuang 2010, Ch. 4), which approximates the evolution of a sum of non-commuting Hamiltonians by a sequence of discrete steps:

$$e^{-i(H_0+H_C)\delta t} \approx e^{-iH_0\delta t}e^{-iH_C\delta t}.$$

By slicing the total adiabatic time into small intervals, the continuous path is transformed into an alternating sequence of discrete unitary operators suitable for the circuit model. This discretized logic gave rise to the analysis subject of this thesis: The Quantum Approximate Optimization Algorithm (QAOA) (Section 2.4).

Introduced by (Farhi, Goldstone, and Gutmann 2014), QAOA is a hybrid quantum-classical algorithm targeting NP-hard combinatorial optimization problems (Section 2.3) such as MaxCut (Section 2.3.2). The algorithm starts in an easy-to-prepare initial state (usually the uniform

superposition) and applies p alternating layers of the following unitary operators: $e^{-i\beta H_C}$ (encoding the problem) and $e^{-i\gamma H_0}$ (a mixing operator). Respectively, those operators applied to a quantum state correspond to the evolution under (time-independent) Hamiltonians H_C and H_0. The durations associated with each alternation of Hamiltonians form a set of $2p$ real parameters $(\gamma_1, \beta_1, \ldots, \gamma_p, \beta_p)$ optimized by classical routines (Section 2.4); typically via gradient descent or gradient-free classical optimization loops. QAOA can be viewed as a variational relaxation of the Trotterized adiabatic path. Instead of fixing the time steps δt to be small (as required for strict adiabatic simulation), QAOA treats these durations as variational parameters to be classically optimized.

The analysis of QAOA has undergone significant evolution since its introduction. While early work focused on establishing worst-case performance guarantees (Barak and Marwaha 2022), recent attention has shifted toward understanding the *geometry* of the optimization landscape. A challenge arises from the "Barren Plateau" phenomenon (McClean, Boixo, et al. 2018) described in Section 2.4.5. In this regime, the expected cost function landscape concentrates exponentially around its mean across the parameter space, rendering the landscape effectively flat and untrainable (making optimization in the classical loop of QAOA effectively impossible).

Alongside these geometric challenges, analytical efforts to bound performance face a severe computational barrier. Theoretical analysis using exact recursive formulas scale exponentially with depth ($\mathcal{O}(p^2 4^p)$), rendering evaluation intractable beyond $p \approx 20$ (Basso et al. 2022). Although recent advances in tensor network simulations have pushed this boundary (Boulebnane et al. 2025), there remains a critical need for analytical frameworks capable of bypassing this exponential overhead to rigorously assess performance and trainability limits without resorting to intractable brute-force simulation.

Along with recent advances, several critical open questions remain for assessing QAOA's potential for quantum advantage.

- **Beating Classical Algorithms:** While performance guarantees improve with p, at what depth does QAOA actually outperform the best classical algorithms? For MaxCut on large-girth D-regular graphs, analysis by (Basso et al. 2022) found that $p = 11$ is required to beat the provable performance guarantee of the classical Goemans-Williamson algorithm (Goemans and Williamson 1995). Understanding this crossover point for different problems is a key research goal.

- **The Trainability Barrier:** Recent work has begun to link the "Barren Plateau" phenomenon to the spectral properties of the expected cost function (Okumura and Ohzeki 2025) and the complexity of the operator spread (Nemkov, Kiktenko, and Fedorov 2023). However, determining *a priori* whether a specific problem instance allows for effective gradient-based training, without incurring the exponential cost of running the circuit, remains an open diagnostic challenge.

To overcome the exponential complexity of exact recursive analysis, this thesis explores a novel probabilistic framework inspired by the random compiler formalism of (Dubus, Cunningham, and Roland 2025). Instead of treating QAOA as a sequence of fixed unitary gates, we model the algorithm as a continuous-time Markov process (Section 2.2), where the variational parameters $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ are realized as stochastic switching times between Hamiltonians.

This approach aims to potentially transfer the complexity from the combinatorial summation to the analysis of differential equations and asymptotic limits. Crucially, by viewing the parameters through a statistical lens, this formalism may offer a unifying framework for analyzing QAOA. This perspective might not only help derive new performance-bound techniques, but also provide new physical insights into the mechanisms of trainability and the emergence of Barren Plateaus.

The remainder of this thesis is organized as follows. Chapter 2 establishes the theoretical foundation required for this work, covering the density-matrix formalism, the relevant properties

of continuous-time Markov chains, the fundamental principles of QAOA and the combinatorial optimization background. Chapter 3 formalizes the randomized control framework, adapting the quantum-classical Markovian dynamics to the specific directed graph structure of the QAOA ansatz. Chapter 4 investigates single-switch strategies. We introduce entropy regularization to solve the variational problem, deriving exact analytical solutions for the transition density in terms of modified Bessel functions and providing a thermodynamic interpretation of the optimization landscape. Chapter 5 extends the analysis to multi-switch strategies. We discuss the analytical intractability arising from the nested recursive structure, but demonstrate, through the specific case of Unstructured Search, that this framework can theoretically validate a quadratic speedup at low depth. Chapter 6 pivots to a spectral analysis of trainability. By linking the high-temperature expansion of our randomized model to the Fourier spectrum of the expected cost function, we develop a polynomial-time Monte Carlo diagnostic that predicts the presence of Barren Plateaus by simulating the quantum signal-to-noise ratio. Finally, Chapter 7 concludes the thesis, summarizes the contributions, and outlines potential directions for future research.

Chapter 2

State of the Art

In this chapter, we present the theoretical background and the technical tools required throughout the manuscript. This foundation is necessary both to build the method described in Section 3 and correctly place the research question within its mathematical context. To begin with, we introduce the core mathematical notions used in the remainder of the thesis. This includes the formalism of quantum states and density operator (together with Hamiltonian and quantum evolution) relevant for the quantum computing context (2.1), as well as the Markov chain framework (2.2) necessary for the randomized approach. Next, we provide an overview of the computational theory relevant to this work. This overview clarifies where our journey begins, and which classes of problems we will focus on (2.3). This path naturally leads to the formal introduction and mathematical description of the Quantum Approximate Optimization Algorithm (2.4) followed by a discussion about recent theoretical results in the final part of this chapter.

2.1 Quantum Mechanics formalism

This section provides a compact, computation-oriented introduction to the density-matrix formalism and Hamiltonian-generated quantum dynamics. The goal is to precisely define the objects manipulated by quantum algorithms (states, Hamiltonians, measurements), how they compose, and how noise (or classical control) is modeled. This framework will be necessary in Sections 2.3.4, 2.4 and in Chapter 3 to describe respectively AQC, QAOA and the randomized scheme.

2.1.1 Basic principles

Quantum algorithms and more generally quantum computing, as their name suggests, are built upon the framework of quantum mechanics (QM). In QM, a closed system is described by a unit vector $|\psi\rangle$ called a quantum state and lives in a Hilbert space (Nielsen and Chuang 2010). Inherently, QM possesses a probabilistic nature. External interactions with the system, e.g. trying to extract information from it, will be modeled as a measurement process that yields random outcomes. This intrinsic probabilistic nature of QM is not a statement about ignorance but a feature of the theory itself.

At the level of a fixed orthonormal basis $\{|x\rangle\}$ (e.g., the computational basis), a state can be written as a superposition

$$|\psi\rangle = \sum_x \alpha_x |x\rangle, \qquad \sum_x |\alpha_x|^2 = 1. \tag{2.1}$$

Observables and measurement (Born rule). In quantum mechanics, an *observable* is a Hermitian operator $O = O^\dagger$ representing measurable properties. Observables admit a spectral decomposition $O = \sum_m o_m \Pi_m$, where the Π_m are orthogonal projectors that resolve the identity

($\Pi_m\Pi_n = \delta_{mn}\Pi_m$, $\sum_m \Pi_m = I$). A projective measurement of O yields eigenvalue o_m with probability

$$p(m) = \langle\psi|\,\Pi_m\,|\psi\rangle, \quad \text{and a post-measurement state} \quad |\psi_m\rangle = \frac{\Pi_m|\psi\rangle}{\sqrt{p(m)}}. \tag{2.2}$$

The expectation and variance follow from the same projectors:

$$\langle O\rangle_\psi = \sum_m o_m\, p(m) = \langle\psi|O|\psi\rangle, \qquad \mathrm{Var}_\psi(O) = \langle\psi|O^2|\psi\rangle - \langle\psi|O|\psi\rangle^2. \tag{2.3}$$

Measurement in the computational basis is the special case with $\Pi_x = |x\rangle\langle x|$, so that

$$p(x) = |\alpha_x|^2, \qquad \sum_x p(x) = 1. \tag{2.4}$$

Hamiltonian Operator. An example of observable (extensively used in this manuscript) is the Hamiltonian operator H, the Hermitian observable associated with the system's total energy. Its eigenvalues correspond to the allowed energy levels of the system. The eigenvector corresponding to the lowest eigenvalue is known as the *ground state*, while the others are *excited states*. Finding the ground state of a given Hamiltonian is a fundamental problem in physics and theoretical computer science. It is the main target of many quantum algorithms, like the quantum approximate optimization algorithm (QAOA) described in Section 2.4.

State evolution. For a closed system, time evolution is generated by the Hamiltonian, as described by the Schrödinger equation

$$i\,\tfrac{d}{dt}|\psi(t)\rangle = H(t)\,|\psi(t)\rangle, \qquad (\hbar = 1). \tag{2.5}$$

Its solution can be written using a unitary propagator $U(t, t_0)$ (also called the evolution operator): $|\psi(t)\rangle = U(t, t_0)|\psi(t_0)\rangle$ with $U(t, t_0) = \mathcal{T}\exp\big(-i\int_{t_0}^{t} H(s)\,ds\big)$ (time-ordering $\mathcal{T}$), and $U(t, t_0) = e^{-iH(t-t_0)}$ when H is time-independent. In that latter case, if $H|E_n\rangle = E_n|E_n\rangle$, components along energy eigenstates accumulate phases e^{-iE_nt} and the expectation $\langle H\rangle_\psi$ is conserved. The unitarity of $U(t)$ ensures normalization is preserved and closed-system dynamics are reversible.

2.1.2 Density Operators

In the state-vector picture, a closed system is described by $|\psi\rangle$, but for algorithms and realistic devices, it is convenient (and sometimes necessary) to use the *density operator* (also called density matrix) ρ. It generalizes pure states to statistical ensembles and open dynamics (Nielsen and Chuang 2010, Ch. 2). An ensemble $\{p_i, |\psi_i\rangle\}$ corresponds to the matrix

$$\rho \equiv \sum_i p_i\,|\psi_i\rangle\langle\psi_i|. \tag{2.6}$$

Valid density operators are precisely those matrices that are Hermitian, positive semidefinite, and have unit trace:

$$\rho = \rho^\dagger, \quad \rho \succeq 0, \quad \mathrm{tr}(\rho) = 1. \tag{2.7}$$

Pure states correspond to rank-1 projectors $\rho = |\psi\rangle\langle\psi|$ and satisfy $\mathrm{tr}(\rho^2) = 1$; mixed states satisfy $\mathrm{tr}(\rho^2) < 1$.

Time evolution For a closed system, the density operator obeys the quantum version of Liouville's Theorem (Huang 1987, Ch. 8)

$$\dot{\rho}(t) = -\,i\,[H(t), \rho(t)], \qquad (\hbar = 1), \tag{2.8}$$

(where $[\cdot, \cdot]$ denotes the commutator) and whose solution is $\rho(t) = U(t)\,\rho(0)\,U(t)^\dagger$.

Measurement As noted above and accordingly with (Nielsen and Chuang 2010, Ch. 2), Quantum mechanics is inherently probabilistic: even when the state ρ is fully known, the result of a measurement is random, with probabilities determined by the Born rule. Algorithmically, this has a consequence: to extract a classical answer, we must measure the quantum state (thereby collapsing it). We adopt the general POVM formalism below, which includes projective measurements as a special case and models the computational basis readout used later (e.g., to estimate QAOA expected cost value). Measurements are described by a collection of operators $\{M_m\}$ (POVM elements via $E_m = M_m^\dagger M_m$) satisfying $\sum_m E_m = I$. Given a pre-measurement state ρ, the probability of outcome m and the corresponding post-measurement state are respectively

$$p(m) = \mathrm{tr}(E_m\,\rho), \qquad \rho_m = \frac{M_m\,\rho\,M_m^\dagger}{p(m)}. \tag{2.9}$$

Expectation of observables. The expectation value of the measurement outcome of an observable with respect to a state ρ is

$$\langle O \rangle_\rho = \mathrm{tr}(\rho\,O), \tag{2.10}$$

which is the central quantity optimized or estimated in variational algorithms (e.g., take $O = H_C$ as a problem Hamiltonian) as described in Section 2.3.1.

Composite systems and Partial Trace. When dealing with composite quantum systems (e.g., a system A and an environment B) described by a joint Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, the state of a subsystem can be obtained via the *partial trace* (Nielsen and Chuang 2010, Ch. 8). If the joint state is ρ_{AB}, the reduced density matrix for system A is defined as:

$$\rho_A = \mathrm{tr}_B(\rho_{AB}), \tag{2.11}$$

where tr_B is the unique linear map satisfying $\mathrm{tr}_B(\rho_A \otimes \rho_B) = \rho_A\,\mathrm{tr}(\rho_B)$ for any product state. Mathematically, the operation is defined by tracing out the degrees of freedom of the unobserved system. In terms of basis vectors $\{|i\rangle_A\}$ and $\{|j\rangle_B\}$, the matrix elements are:

$$\langle i|\rho_A|i'\rangle = \sum_j \langle i \otimes j|\rho_{AB}|i' \otimes j\rangle. \tag{2.12}$$

The partial trace is the only operation that yields the correct statistics for measurements performed solely on subsystem A.

2.1.3 Open Systems

An open quantum system is a system that interacts with an external environment, resulting in non-unitary dynamics such as dissipation and decoherence.

Between gates or during evolution, the state undergoes a completely positive trace-preserving (CPTP) map:

$$\Phi(\rho) = \sum_k E_k\,\rho\,E_k^\dagger, \qquad \sum_k E_k^\dagger E_k = I. \tag{2.13}$$

This map describes the most general physical evolution of a quantum state. The trace-preserving condition, $\sum_k E_k^\dagger E_k = I$, ensures that total probability is conserved, while the completely positive property guarantees the map remains physically valid even if the system is entangled with an environment.

For continuous-time evolution, one considers a family of CPTP maps $\{\Phi_t\}_{t\geq 0}$ forming a dynamical semigroup:

$$\Phi_{t+s} = \Phi_t \circ \Phi_s.$$

This expresses the fact that the state at time t fully determines its future evolution. Such a semigroup always admits a linear generator $\mathcal{L}$ defined by

$$\Phi_t = e^{t\mathcal{L}}, \qquad \dot{\rho}(t) = \mathcal{L}\rho(t).$$

In continuous time under a Markovian assumption (the environment has no memory of past interactions with the system), the generator $\mathcal{L}$ is constrained by complete positivity and trace preservation to take the Gorini–Kossakowski–Sudarshan–Lindblad (GKSL) form, also known as the master equation (Nielsen and Chuang 2010, Ch. 8):

$$\dot{\rho} = -i[H,\rho] + \sum_k \Big(L_k \,\rho\, L_k^\dagger - \tfrac{1}{2}\{L_k^\dagger L_k, \rho\} \Big) \tag{2.14}$$

(where $\{\cdot,\cdot\}$ denotes the anti-commutator). Here, the Hamiltonian H governs the coherent, unitary part of the evolution. The operators L_j are known as the *Lindblad operators*, usually referred to as *jump operators*. They represent the incoherent processes arising from the coupling of the system to its environment (noise) or classical control as described in Section 3. Each index k in Eq. (2.14) corresponds to a different type of interaction with the environment or a possible transition from a state $|i\rangle$ to another state $|j\rangle$, defining a Lindblad operator as $L_{i\to j} = \sqrt{A_{ij}}\,|j\rangle\langle i|$ for $i \neq j$.

2.2 Markov Chains

As a base tool for this thesis, this section reintroduces Markov chains (MC), providing a foundational understanding of their principles and properties. Markov chains are a class of stochastic processes that model systems transitioning from one state to another with the characteristic that the future state of the chain only depends on the current state of the system (The so-called *Markov property*) (Lovric 2025). In other words, the next state will not depend on the sequence of events that preceded it. This "memoryless" property makes Markov chains a powerful and tractable tool for modeling a wide array of real-world phenomena (Norris 1997a), but not only. Their well-defined mathematical structure makes them a useful theoretical tool in computer science, as illustrated by the Metropolis Algorithm (Hastings 1970). MC helps to prove convergence conditions, showing how MC theory can underpin algorithmic guarantees. This framework allows for rigorous proofs of crucial properties such as convergence rates, long-term stability, and expected performance.

A Markov chain is characterized by a set of states, known as the *state space*, and the probabilities of transitioning between these states. These transitions can occur at discrete time intervals, defining a *Discrete-Time Markov Chain (DTMC)* (Lovric 2025), or at any point in time, which characterizes a *Continuous-Time Markov Chain (CTMC)* (Norris 1997b).

2.2.1 Discrete-Time Markov Chains (DTMCs)

Let $\{X_t\}_{t\in\mathbb{N}_0}$ be a stochastic process on a finite or countable state space S. We adopt the (row) probability vector convention: for a distribution μ on S, $\mu(i) = \mathbb{P}(X_0 = i)$ and the distribution at

time t is the row vector μ_t with entries $\mu_t(j) = \mathbb{P}(X_t = j)$. The *Markov property* (memorylessness) is

$$\mathbb{P}(X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{t+1} = j \mid X_t = i), \qquad i, i_0, \ldots, i_{t-1}, j \in S.$$

This says that conditionally on the present state, the future is independent of the past.

Time-homogeneous vs. inhomogeneous. The transition mechanism is specified by one-step conditional probabilities

$$P_{ij}(t) = \mathbb{P}(X_{t+1} = j \mid X_t = i), \quad i, j \in S,\ t \in \mathbb{N}_0.$$

The chain is *time-homogeneous* if $P_{ij}(t)$ is independent of t, i.e. $P_{ij}(t) = P_{ij}$ for all t; then a single transition matrix $\mathbf{P} = (P_{ij})_{i,j \in S}$ governs all steps. Otherwise we have a (time-)*inhomogeneous* chain with a sequence of transition matrices $\mathbf{P}^{(0)}, \mathbf{P}^{(1)}, \ldots$ where $\mathbf{P}^{(t)} = (P_{ij}(t))$.

For a homogeneous chain, each row of $\mathbf{P}$ is a probability distribution:

$$P_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i), \qquad \sum_{j \in S} P_{ij} = 1, \quad P_{ij} \geq 0.$$

k-step transitions. In the homogeneous case, the *k-step transition matrix* is the k-th power $\mathbf{P}^k$ (with $\mathbf{P}^0 = \mathbf{I}$):

$$(\mathbf{P}^k)_{ij} = \mathbb{P}(X_{t+k} = j \mid X_t = i), \qquad k \in \mathbb{N}.$$

In the inhomogeneous case, by contrast,

$$\mathbb{P}(X_{t+k} = j \mid X_t = i) = \left(\mathbf{P}^{(t)} \mathbf{P}^{(t+1)} \cdots \mathbf{P}^{(t+k-1)}\right)_{ij}$$

Evolution of distributions. The pair $(\mu, \mathbf{P})$ (initial distribution + transition matrix) fully specifies a homogeneous DTMC. The distribution at time t satisfies

$$\mu_t = \mu \mathbf{P}^t, \qquad t \geq 0.$$

(Column-vector conventions appear in some texts; then one writes $\mu_t = \mathbf{P}^t \mu$.) Each entry $\mu_t(j)$ gives $\mathbb{P}(X_t = j)$. In the inhomogeneous case,

$$\mu_t = \mu\, \mathbf{P}^{(0)} \mathbf{P}^{(1)} \cdots \mathbf{P}^{(t-1)}.$$

2.2.2 Continuous-Time Markov Chains (CTMCs)

Continuous-Time Markov Chains (CTMCs) extend the concepts of a DTMC to a process $\{X_t\}_{t \in \mathbb{R}_{\geq 0}}$ evolving over continuous time. The fundamental principle remains the memoryless Markov property, but the transition mechanism is described by rates of change rather than fixed-step probabilities. A crucial distinction is whether these rates are constant over time (*time-homogeneous*) or vary with time (*time-inhomogeneous*) (Yin and Zhang 2012). For clarity, we will restrict our analysis to chains on a finite state space S. The CTMC will be the specific tool used in the randomized approach described in Chapter 3.

The Generator Matrix and Transition Probabilities. The dynamics of a CTMC are governed by a *generator matrix* (also called *rate matrix* or *Q-matrix*), often denoted $\mathbf{Q}$ or A in the literature. In the general, time-inhomogeneous case, this matrix is a function of time, $\mathbf{Q}(t) = (q_{ij}(t))_{i,j\in S}$. Its entries are transition rates satisfying the *q-Property* (Yin and Zhang 2012):

1. $q_{ij}(t) \geq 0$ for all $i \neq j$.
2. $\sum_{j\in S} q_{ij}(t) = 0$ for all $i \in S$, which implies $q_{ii}(t) = -\sum_{j\neq i} q_{ij}(t)$.

Here, $q_{ij}(t)$ for $i \neq j$ is the instantaneous rate of transition from state i to state j at time t. The diagonal entry $q_{ii}(t)$ is non-positive, and its negative, $q_i(t) = -q_{ii}(t)$, represents the total rate of leaving state i at time t. The probability of transitioning from state i at time s to state j at time t is given by the *transition probability matrix* $\mathbf{P}(t, s)$ such that $P_{ij}(t, s) = \mathbb{P}(X_t = j \mid X_s = i)$. This two-parameter family of matrices is the unique solution to the Kolmogorov forward and backward equations (Yin and Zhang 2012):

$$\frac{\partial}{\partial t}\mathbf{P}(t, s) = \mathbf{P}(t, s)\mathbf{Q}(t) \quad \text{(Forward Equation)},$$
$$\frac{\partial}{\partial s}\mathbf{P}(t, s) = -\mathbf{Q}(s)\mathbf{P}(t, s) \quad \text{(Backward Equation)},$$

with the initial condition $\mathbf{P}(s, s) = \mathbf{I}$.

In the simpler *time-homogeneous* case where $\mathbf{Q}(t) = \mathbf{Q}$ is a constant matrix, the transition probabilities depend only on the duration of the time interval, $t - s$. We can then define a one-parameter *transition semigroup* $\mathbf{P}(t) \equiv \mathbf{P}(t, 0)$. For a finite state space, this is given by the matrix exponential:

$$\mathbf{P}(t) = e^{t\mathbf{Q}} = \sum_{k=0}^{\infty} \frac{(t\mathbf{Q})^k}{k!}.$$

The distribution at time t evolves according to $\mu_t = \mu\mathbf{P}(t)$.

Remark (Notation conventions). We adopt the row-vector convention in this section, consistent with much of the probability literature. In Chapter 3, we switch to column vectors ($\partial_t\vec{p} = A^T\vec{p}$) to align with the density matrix formalism, and denote the generator by A to emphasize its role in the Lindblad equation 2.14.

Holding Times and Jumps. If the process enters state i at time t_0, it remains there for a random holding time. The probability that this holding time exceeds t is governed by the time-varying exit rate $q_i(s) = -q_{ii}(s)$:

$$\mathbb{P}(\text{Holding time} > t \mid \text{Entered state } i \text{ at } t_0) = \exp\left(-\int_{t_0}^{t_0+t} q_i(s)\, ds\right).$$

The quantity $\Lambda_i(t_0, t_0 + t) := \int_{t_0}^{t_0+t} q_i(s)\, ds$ is known as the *cumulative hazard* or *integrated rate*, and the exponential $S_i(t_0, t_0 + t) := e^{-\Lambda_i(t_0,t_0+t)}$ is the *survival probability*. These quantities will play a central role in Chapter 3.

After the holding time elapses at some time t_{jump}, the process jumps to a new state $j \neq i$ with probability proportional to the instantaneous rate:

$$\mathbb{P}(\text{Jump to } j \mid \text{Jump from } i \text{ at } t_{\text{jump}}) = \frac{q_{ij}(t_{\text{jump}})}{q_i(t_{\text{jump}})}, \quad \text{for } q_i(t_{\text{jump}}) > 0.$$

In the time-homogeneous case, the cumulative hazard simplifies to $\Lambda_i = q_i \cdot t$, so the holding time follows an exponential distribution with constant rate q_i.

2.3 Background of Optimization

The purpose of this section is to provide the general context and relevant concepts related to the problems addressed by the QAOA algorithm, with a focus on combinatorial optimization and how quantum models of computation tackle them. It also outlines the tools, algorithmic frameworks, and types of performance guarantees that will be relevant in the remainder of this work.

We start with general notions of optimization theory, placing the vocabulary and challenges encountered when dealing with combinatorial problems. Next, we introduce the canonical problem MaxCut and its mathematical formulation. Finally, we present the standard way of encoding optimization problems from a quantum computation perspective and how the adiabatic evolution framework naturally leads to gate-based quantum algorithms and QAOA.

2.3.1 Combinatorial Optimization Theory

Many important problems in science and industry can be framed as combinatorial optimization problems, where the goal is to find an optimal solution from a vast, discrete search space. A significant portion of these problems belongs to the complexity class of *NP-hard* problems[1]. For such problems, no known algorithm can find the exact optimal solution in a time that scales polynomially with the problem size; the required resources typically grow exponentially.

This computational intractability motivates the development of alternative strategies. Approaches generally fall into two categories: *approximation algorithms*, which run in polynomial time and provide a *provable performance guarantee* (quantified by the approximation ratio) regarding how far the solution is from the true optimum (Williamson and Shmoys 2011); and *heuristics*, which are practical methods designed to find good solutions quickly but critically *lack provable bounds* on their performance. Currently, QAOA is treated as a heuristic in practice, though in special cases it can be regarded as an approximation algorithm with provable performance guarantees.

Beyond worst-case analysis, the performance of an algorithm is also often evaluated on its *average-case* or *typical* behavior when applied to instances drawn from a specific random ensemble, such as random graphs.

2.3.2 Canonical Problem : MaxCut

To benchmark and analyze optimization algorithms, research often focuses on a standard, well-understood NP-hard problems: MaxCut.

Given an undirected graph $G = (V, E)$, the MaxCut problem seeks to partition the vertices V into two disjoint sets such that the number of edges connecting vertices in different sets is maximized. This combinatorial problem admits a natural mapping to physics via the *Ising formulation* (Lucas 2014). By assigning a spin variable $z_i \in \{+1, -1\}$ to each vertex, we can construct a classical cost function[2] that counts the number of cut edges:

$$C(\mathbf{z}) = \frac{1}{2} \sum_{(i,j)\in E} (1 - z_i z_j). \tag{2.15}$$

Here, the term $z_i z_j$ equals $+1$ if the spins are aligned (in the same set) and -1 if they are anti-aligned (in different sets). Consequently, the summand $(1 - z_i z_j)/2$ acts as an indicator

[1] The class **P** contains problems solvable in polynomial time. **NP** contains problems whose solutions can be *verified* in polynomial time. A problem is **NP-hard** if it is at least as hard as the hardest problems in NP; specifically, any problem in NP can be reduced to it. Solving an NP-hard problem in polynomial time would imply P = NP

[2] We insist on the '*classical*' to not confuse with the '*expected cost function*' $\mathcal{C}$ defined in Section 2.4

function, contributing 1 to the sum if and only if the edge (i, j) is cut. Maximizing this cut is mathematically equivalent to finding the ground state of an antiferromagnetic Ising Hamiltonian $H = \sum_{(i,j)\in E} Z_i Z_j$.

Theoretical analysis of MaxCut typically focuses on different classes of graph ensembles such as:

- **Sparse Regular Graphs:** A common benchmark for worst-case analysis is the family of D-regular graphs (Farhi, Goldstone, and Gutmann 2014), where every vertex has exactly D neighbors. For such a graph with n vertices, the total number of edges scales linearly with the system size, $|E| = nD/2$. The specific case of 3-regular graphs is frequently used to study the performance guarantees of QAOA at low depth.
- **Dense Random Graphs:** To analyze average-case performance and trainability issues (such as Barren Plateaus, see 2.4.5) in highly connected systems, we consider the Erdős-Rényi model $G(n, q)$ (Erdös and Rényi 1959). In this ensemble, each of the possible pairs of vertices are connected by an edge independently with probability q. For a constant q, these graphs are *dense*, with the expected number of edges scaling quadratically as $|E| \approx qn^2/2$.

2.3.3 Encoding Problems in Hamiltonians

To leverage quantum mechanics for solving these classical problems, we must first translate them into the language of quantum physics. This is achieved by encoding the classical cost function into a *Hamiltonian* (H). The central principle is to establish a correspondence between each possible classical solution and a quantum state, and between the classical cost of that solution and the energy of that state. A classical solution, represented by a bitstring or a spin configuration z, is mapped to a computational basis state $|z\rangle$ in a multi-qubit Hilbert space. The value of the classical cost function for that solution, $C(z)$, is then set as the energy (eigenvalue) associated with that state.

This procedure results in a *problem Hamiltonian* H_C (sometimes called cost Hamiltonian in the literature), which is diagonal in the computational basis. Its structure is given by:

$$H_C = \sum_z C(z)|z\rangle\langle z|. \tag{2.16}$$

By this construction, the quantum state associated with the lowest eigenvalue (the ground state of H_C) corresponds exactly to the classical bitstring z_{opt} that minimizes the classical cost function $C(z)$. The optimization problem is thus transformed into the "physical problem" of finding the ground state of a Hamiltonian.

2.3.4 From Continuous Evolution to Gate-Based Algorithms

Once a problem is encoded in a Hamiltonian H_C, the task becomes preparing its ground state. The theoretical foundation for this is the adiabatic theorem (Born and Fock 1928).

Adiabatic Quantum Computation (AQC) Described for the first time by (Farhi, Goldstone, Gutmann, and Sipser 2000), AQC begins by preparing a system in the easily created ground state of a simple initial Hamiltonian, H_0. This system is then evolved under a time-dependent Hamiltonian that slowly transforms from H_0 to the problem Hamiltonian H_C. A standard choice for this evolution path is a linear interpolation:

$$H(s) = (1 - s)H_0 + sH_C, \tag{2.17}$$

where $s = t/T$ varies smoothly from 0 to 1 over a total evolution time T. The adiabatic theorem guarantees that if the evolution is slow enough, the system will remain in the instantaneous ground state. The required time T is governed by the minimum energy gap $\Delta_{\min}$ between the ground state and the first excited state during the evolution:

$$T \gg \frac{\max_s ||dH/ds||}{\Delta_{\min}^2}. \tag{2.18}$$

This reveals the primary bottleneck: for NP-hard problems, the spectral gap might become exponentially small with the problem size, requiring an exponentially long evolution time and thus negating any quantum advantage.

Discretization via Trotterization Gate-based quantum computers (quantum circuit model of computation) cannot implement a continuous-time evolution directly. Instead, the evolution must be broken down into a sequence of discrete quantum gates. This process, known as Trotterization, approximates the continuous evolution using the *Lie-Trotter formula* (Nielsen and Chuang 2010, Ch. 4):

$$e^{-i(H_0+H_C)\Delta t} \approx e^{-iH_0\Delta t}e^{-iH_C\Delta t} + \mathcal{O}(\Delta t^2). \tag{2.19}$$

By repeatedly applying this approximation for small time steps Δt, one can simulate the overall evolution. This introduces a controllable *Trotter error* that decreases as the time steps become smaller.

The NISQ Context and Variational Algorithms Perfectly simulating a long adiabatic path with high precision requires very deep circuits (shorter time steps imply higher number of layers), which is infeasible on current *Noisy Intermediate-Scale Quantum (NISQ)* hardware due to limited coherence times and gate fidelities. This has motivated a shift towards *Variational Quantum Algorithms (VQAs)*. VQAs are hybrid quantum-classical methods that use a shallow, parameterized quantum circuit (also called an *ansatz*) to prepare a trial state. A classical optimizer then iteratively adjusts the circuit parameters to minimize the expected cost. The QAOA algorithm, described in the next section, is an example of VQA.

This approach offers a pragmatic compromise. Instead of a deep circuit that perfectly replicates the adiabatic path, QAOA uses a shallow, Trotter-inspired circuit as its variational ansatz, trading the rigorous guarantees of AQC for a structure better suited to the constraints of NISQ devices.

2.3.5 Dimensions of Complexity in Variational Optimization

Throughout this manuscript, the term "complexity" is used in distinct contexts that impose different constraints on the algorithm. To avoid ambiguity, we distinguish between four specific regimes:

- **Computational Complexity:** This refers to the intrinsic hardness of the combinatorial problem itself (e.g., MaxCut being NP-hard). Variational algorithms like QAOA function as heuristics here; they do not promise polynomial-time solutions for exact optima but aim for high approximation ratios.
- **Query and Circuit Complexity:** This relates to the physical resources required by the quantum device, specifically the circuit depth p. In Section 5.5, we analyze this in terms of the number of oracle calls[3] required to boost the success probability of finding a marked element in an unstructured search problem.

[3]The number of times we use unitary evolution $e^{-iH_C t}$

- **Analytical Complexity:** This constitutes the primary bottleneck for theoretical study and motivates the randomized framework in Chapters 3-5. It refers to the difficulty of *classically* evaluating the expected cost function $\mathcal{C}(\boldsymbol{\gamma}, \boldsymbol{\beta})$ (defined in the next section). Exact recursive formulas scale exponentially with depth ($\mathcal{O}(p^2 4^p)$), rendering exact analysis intractable for $p \gg 1$.
- **Statistical (Sample) Complexity:** Addressed in Chapter 6, this dictates the feasibility of training. Since quantum expectation values are estimated probabilistically, the classical optimizer requires a sufficient number of measurement shots (N_{shots}) on the quantum circuit to distinguish true gradient signals from statistical shot noise.

2.4 The Quantum Approximate Optimization Algorithm (QAOA)

Based on the discretized variational principles motivated in the previous section, this section formally introduces the Quantum Approximate Optimization Algorithm (QAOA). Introduced by (Farhi, Goldstone, and Gutmann 2014), QAOA is a hybrid quantum-classical algorithm designed for tackling combinatorial optimization problems on near-term quantum devices. We will define its structure, discuss its connection to adiabatic evolution, and review its performance characteristics and fundamental limitations, which together establish the context for the research questions addressed in this thesis.

2.4.1 Algorithm Definition

QAOA operates by preparing a parameterized quantum state on a register of qubits and classically optimizing those parameters to find an approximate solution to a problem encoded in a problem Hamiltonian, H_C as described in Section 2.3.3.

The QAOA Ansatz. The algorithm begins with the system in an easy-to-prepare initial state, typically the ground state of a *mixer Hamiltonian*, H_0. For combinatorial optimization problems, a standard choice is the transverse-field Hamiltonian $H_0 = -\sum_i \sigma_i^x$, whose ground state is the uniform superposition of all computational basis states, $|\psi_0\rangle = \frac{1}{\sqrt{2^n}} \sum_z |z\rangle$.

The core of the algorithm is the preparation of the variational state, or *ansatz*, by applying p alternating layers of unitary operators generated by the problem and mixer Hamiltonians. The state at depth p is given by:

$$|\psi_p(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = e^{-i\gamma_p H_0} e^{-i\beta_p H_C} \cdots e^{-i\gamma_1 H_0} e^{-i\beta_1 H_C} |\psi_0\rangle, \tag{2.20}$$

where $(\boldsymbol{\gamma}, \boldsymbol{\beta}) = (\gamma_1, \beta_1, \ldots, \gamma_p, \beta_p)$ is a set of $2p$ real-valued, variational parameters. In the density operator formalism, the final state is $\rho_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = |\psi_p\rangle\langle\psi_p|$.

The Hybrid Optimization Loop. The goal of QAOA is to find parameters $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ that minimize the expectation value of the problem Hamiltonian H_C with respect to the ansatz state. We denote this quantity by $\mathcal{C}_p$ and refer to it as the *expected cost* throughout the thesis:

$$\mathcal{C}_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \langle\psi_p(\boldsymbol{\gamma}, \boldsymbol{\beta})|H_C|\psi_p(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = \text{tr}(H_C \rho_p(\boldsymbol{\gamma}, \boldsymbol{\beta})), \tag{2.21}$$

where the subscript p emphasizes the dependence on the circuit depth. This expected cost $\mathcal{C}_p$ should not be confused with the classical cost $C(z)$ of a single bitstring; rather, it represents the average of $C(z)$ over all bitstrings weighted by their measurement probabilities in the state $|\psi_p\rangle$.

As a variational algorithm, QAOA operates through an iterative feedback loop between a quantum processor and a classical optimizer. The quantum processor cannot directly output $\mathcal{C}_p$: each measurement collapses the state to a single bitstring $|z\rangle$, yielding one sample of $C(z)$. To obtain an estimate, the circuit must be prepared and measured N_{shots} times, producing the sample average:

$$\widehat{\mathcal{C}}_p = \frac{1}{N_{\text{shots}}} \sum_{k=1}^{N_{\text{shots}}} C(z^{(k)}), \tag{2.22}$$

where $z^{(k)}$ is the bitstring from the k-th measurement. By the central limit theorem, this estimator $\widehat{\mathcal{C}}_p$ converges to the true expected cost $\mathcal{C}_p$ with a standard error of $\mathcal{O}(1/\sqrt{N_{\text{shots}}})$.

The classical optimizer receives this noisy estimate $\widehat{\mathcal{C}}_p$ and proposes updated angles $(\boldsymbol{\gamma}', \boldsymbol{\beta}')$. The quantum circuit is then re-executed with these new parameters, a fresh batch of N_{shots} measurements is collected, and the process repeats. This loop continues until convergence to (near-)optimal angles $(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)$. At that point, the final state $|\psi_p(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)\rangle$ is sampled to extract high-quality approximate solutions from the resulting distribution over bitstrings.

2.4.2 Connection to Adiabatic Quantum Computation

QAOA can be understood as a discretized and variational form of Adiabatic Quantum Computation (AQC). The continuous adiabatic evolution under $H(s) = (1-s)H_0 + sH_C$ can be approximated via Trotterization, as seen in the previous section. A first-order Trotter decomposition of the total evolution operator over a time T, broken into p steps, would yield a sequence of operators of the form $e^{-i\Delta t(1-s_k)H_0} e^{-i\Delta t s_k H_C}$.

The QAOA ansatz mirrors this structure but treats the evolution times, represented by γ_k and β_k, as free parameters to be optimized rather than being fixed by a predefined adiabatic schedule. This variational freedom allows the algorithm to find "shortcuts" or non-adiabatic pathways through the Hilbert space, making it more flexible and better suited for the shallow circuit depths available on NISQ hardware.

2.4.3 Theoretical Properties of QAOA

There are some notable theoretical results worth mentioning. First, the QAOA algorithm improves monotonically with the depth p:

$$\min_{\boldsymbol{\gamma},\boldsymbol{\beta}} \mathcal{C}_{p+1}(\boldsymbol{\gamma}, \boldsymbol{\beta}) \leq \min_{\boldsymbol{\gamma},\boldsymbol{\beta}} \mathcal{C}_p(\boldsymbol{\gamma}, \boldsymbol{\beta}). \tag{2.23}$$

This implies that, given an efficient classical optimization method to find the angles, the more layers we use, the better the performance of the algorithm. Unfortunately, without further analysis, nothing tells how much better the solution will get (the improvements could be exponentially small with p); more on this in Section 2.4.4.

Finally, in the limit of infinite layers ($p \to \infty$), Farhi et al. showed that QAOA can reproduce the dynamics of the adiabatic algorithm. Specifically, for any total time T and any error ϵ, there exist parameters $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ such that the QAOA state is arbitrarily close to the state evolved under the AQC Hamiltonian for time T. This implies:

$$\lim_{p\to\infty} \min_{\boldsymbol{\gamma},\boldsymbol{\beta}} \mathcal{C}_p(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \min_z C(z), \tag{2.24}$$

meaning that in this limit, QAOA can find the exact optimal solution.

Infinite layers are, however, impractical for NISQ devices, and the key challenge lies in understanding the performance of QAOA at finite, and especially small or intermediate depths. In other words, the field still lacks efficient, provable bounds on the approximation ratio achievable by QAOA.

2.4.4 Locality

The practical performance of QAOA is determined by its behavior at finite, and typically small, depth p. A key feature that governs its performance in this regime is its inherent *locality*. At depth p, the unitary operations effectively propagate information across the problem graph (where graph adjacencies are defined by the terms in H_C). The final state of a given qubit is only influenced by other qubits within a graph distance of p. This causal limitation is often referred to as the algorithm's "light cone." Consequently, for small p, QAOA is a fundamentally local algorithm and cannot "see" the global structure of the problem instance. This locality explains why its performance is often analyzed on high-girth graphs, which are locally tree-like and thus their global structure is not visible to a low-depth algorithm.

2.4.5 The Barren Plateau Phenomenon and Spectral Analysis

A fundamental bottleneck for the scalability of Variational Quantum Algorithms, particularly QAOA, is the "Barren Plateau" phenomenon (McClean, Boixo, et al. 2018). Formally, a Barren Plateau is defined by the exponential decay of the variance of the expected cost function gradients, $\mathrm{Var}(\partial_k \mathcal{C})$, as the system size increases. This renders the landscape[4] effectively flat. In such a regime, the magnitude of the gradient becomes comparable to the statistical shot noise inherent in quantum measurements. Consequently, a classical optimizer becomes unable to distinguish a valid descent direction from noise, rendering gradient-based training intractable regardless of the specific optimization routine employed. However, this phenomenon is inextricably linked to the global concentration of the expected cost function itself as we explain in Section 5.6.2 and 6.1.1. It is important to note that a concentration of the expected cost function values implies a concentration of the gradients. If the variance of the expected cost function $\mathrm{Var}(\mathcal{C})$ vanishes exponentially (the landscape becomes a flat plane), the capacity for the gradient to vary is necessarily suppressed. Therefore, while the literature often focuses on $\mathrm{Var}(\partial_k \mathcal{C})$, demonstrating a vanishing $\mathrm{Var}(\mathcal{C})$ is a sufficient condition to diagnose a Barren Plateau.

Recent theoretical advances have reframed the Barren Plateau problem through the lens of signal processing, leveraging the intrinsic periodicity of parameterized quantum circuits. Since the QAOA expected cost $\mathcal{C}(\gamma, \beta)$ depends on variational parameters via complex exponentials, it admits a natural Fourier series representation. In this context, the geometric "flatness" of the landscape corresponds to a spectral suppression of Fourier coefficients. As demonstrated by (Okumura and Ohzeki 2025) and (Nemkov, Kiktenko, and Fedorov 2023), Parseval's identity establishes a rigorous equivalence between the statistical variance of the expected cost function and its *total spectral power* (the sum of its squared Fourier magnitudes). (Nemkov, Kiktenko, and Fedorov 2023) further extended this relationship, showing that the variance of the gradient corresponds to the frequency-weighted spectral power of the ansatz.

This spectral perspective suggests a paradigm shift in how trainability can be diagnosed. Mathematically, the existence of a Barren Plateau is equivalent to the vanishing of the expected cost's spectral power. Therefore, if one can estimate the magnitude of these Fourier coefficients without executing the full quantum circuit, it becomes possible to determine *a priori* whether a specific QAOA instance is trainable. This insight motivates the development of efficient classical diagnostic tools capable of probing the spectral structure of the expected cost landscape, a challenge we address directly in Chapter 6.

[4] We will use the term "landscape" in this thesis when referring to a geometric interpretation of a function, interpreting its values as heights over a space of parameters.

Chapter 3

Problem Setting & The Randomized Approach

While standard QAOA relies on a deterministic sequence of unitary gates with fixed durations (or angles), the approach developed in this thesis adopts a stochastic perspective. We view the switching between Hamiltonians not as fixed time intervals, but as random events governed by a probabilistic process. This requires a shift from the standard circuit model to a formalism combining quantum dynamics with classical Markovian control.

This chapter establishes the theoretical foundation for this approach, adapting the *Random Compiler* formalism introduced by Dubus, Cunningham, and Roland (2025). We extend their density matrix description of Hamiltonian simulation to the specific context of variational optimization.

3.1 Classical Markovian Dynamics

Before introducing the quantum system, we first define the classical process governing the control. We consider a continuous-time Markov chain on a graph with Q nodes. In the context of a QAOA ansatz of depth p, the graph consists of $Q = 2p$ nodes, where each node i represents a specific step in the algorithm (associated with a specific Hamiltonian).

The state of this classical system at time t is described by a probability vector $\vec{p}(t) \in \mathbb{R}^Q$, where $p_i(t)$ denotes the probability of being in node i. The evolution of this vector is governed by the transition rate matrix A (of size $Q \times Q$):

$$\partial_t \vec{p} = A^T \vec{p}. \tag{3.1}$$

The coefficients A_{ij} represent the transition rate from node i to node j (for $i \neq j$), and diagonal elements are defined as $A_{ii} = -\sum_{j \neq i} A_{ij}$ to ensure conservation of probability.

To facilitate the coupling with a quantum system, it is useful to encode this classical probability distribution into a diagonal density matrix acting on a Hilbert space $\mathcal{H}_{mark} \cong \mathbb{C}^Q$. We define the classical density matrix P:

$$P(t) := \sum_{i=1}^{Q} p_i(t) |i\rangle\langle i|. \tag{3.2}$$

The classical evolution equation can then be rewritten in a Lindblad-like form acting on P. Using the jump operators $L_{i\to j} := |j\rangle\langle i|$ representing a transition from node i to j, the evolution is given by:

$$\partial_t P = \sum_{i \neq j} A_{ij} \left(L_{i\to j} P L_{i\to j}^\dagger - \frac{1}{2} \{ L_{i\to j}^\dagger L_{i\to j}, P \} \right). \tag{3.3}$$

This equation describes purely classical jumps between the diagonal elements of P, preserving its diagonality and trace.

3.2 The Coupled Quantum System

We now introduce the quantum hardware (the system of interest) living in a Hilbert space $\mathcal{H}_{sys} \cong \mathbb{C}^{2^n}$. The total system is described by the tensor product space $\mathcal{H}_{tot} = \mathcal{H}_{sys} \otimes \mathcal{H}_{mark}$. The state of this coupled system is represented by the total density matrix $\boldsymbol{\mathcal{P}}(t) \in \mathrm{End}(\mathcal{H}_{tot})$[1].

The control is applied via the *Global Hamiltonian* $\boldsymbol{\mathcal{H}}$, which applies a specific local Hamiltonian H_i to the system, depending on the state of the Markov chain:

$$\boldsymbol{\mathcal{H}} := \sum_{i=1}^{Q} H_i \otimes |i\rangle\langle i|. \tag{3.4}$$

For QAOA, the Hamiltonians H_i alternate between the problem Hamiltonian H_C and the mixer H_0.

The evolution of the total system $\boldsymbol{\mathcal{P}}$ is governed by a global Lindblad master equation (2.14). It contains the unitary evolution driven by $\boldsymbol{\mathcal{H}}$ and the incoherent transitions driven by the Markov chain. The jump operators are now extended to the total space as $\boldsymbol{\mathcal{L}}_{i\to j} := \mathbb{I}_{sys} \otimes |j\rangle\langle i|$. The master equation then reads:

$$\partial_t \boldsymbol{\mathcal{P}} = -i[\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{P}}] + \sum_{i\neq j} A_{ij} \left(\boldsymbol{\mathcal{L}}_{i\to j} \boldsymbol{\mathcal{P}} \boldsymbol{\mathcal{L}}_{i\to j}^\dagger - \frac{1}{2}\{\boldsymbol{\mathcal{L}}_{i\to j}^\dagger \boldsymbol{\mathcal{L}}_{i\to j}, \boldsymbol{\mathcal{P}}\} \right). \tag{3.5}$$

It is crucial to note that if we trace out the quantum system from Eq. (3.5), we exactly recover the classical evolution described in Eq. (3.3), ensuring the consistency of the framework (i.e., $\mathrm{Tr}_{sys}(\boldsymbol{\mathcal{P}}) = P$).

Furthermore, if $\boldsymbol{\mathcal{P}}$ is initially block-diagonal (representing a classical distribution of control states), this equation preserves that structure. We can therefore decompose the total state as:

$$\boldsymbol{\mathcal{P}}(t) = \sum_{i=1}^{Q} \rho_i(t) \otimes |i\rangle\langle i|. \tag{3.6}$$

Here, the operator $\rho_i(t) := \langle i|\boldsymbol{\mathcal{P}}(t)|i\rangle$ acts on $\mathcal{H}_{sys}$. It represents the sub-normalized state of the quantum system given that the Markov chain is in node i; this is why we will refer to it as the marginal density matrix. Its trace corresponds to the classical probability: $\mathrm{Tr}(\rho_i(t)) = p_i(t)$.

3.3 Marginal Density Matrix Equations

Tracing out the control register from Eq. (3.5) allows us to derive the coupled differential equations for the marginal matrices $\rho_i(t)$. The commutator term $-i[\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{P}}]$ acts block-wise, yielding the local coherent evolution $-i[H_i, \rho_i]$ for each node. The Lindblad part describes the classical probability flux. The term $\boldsymbol{\mathcal{L}}_{j\to i} \boldsymbol{\mathcal{P}} \boldsymbol{\mathcal{L}}_{j\to i}^\dagger$ transfers the population from the block ρ_j into the block ρ_i, while the anti-commutator term $-\frac{1}{2}\{L_{j\to i}^\dagger L_{j\to i}, \boldsymbol{\mathcal{P}}\}$ removes population from ρ_j.

Collecting these terms, we arrive at the set of coupled ordinary differential equations governing the randomized ansatz:

$$\partial_t \rho_i(t) = -i[H_i, \rho_i(t)] + \sum_{j\neq i} \left(A_{ji}(t)\rho_j(t) - A_{ij}(t)\rho_i(t) \right). \tag{3.7}$$

Equation (3.7) describes a system that evolves coherently under H_i while continuously exchanging probability mass with other configurations according to the transition matrix A.

[1] $\mathrm{End}(\mathcal{H})$ denotes the set of all linear operators from $\mathcal{H}$ to itself (endomorphisms)

3.4 Application to QAOA: The Directed Regime

While Dubus, Cunningham, and Roland (2025) utilize this framework for Hamiltonian simulation (approximating a target $H = \sum w_i H_i$), this thesis applies the formalism to the *optimization* of the QAOA ansatz in order to reach a specific state (the ground state of H_C).

3.4.1 Directed Chain Structure

The key structural feature for QAOA is that the Markov graph becomes a directed line graph ($1 \to 2 \to \cdots \to 2p$). Transitions are unidirectional: nodes with odd indices are associated with evolution under H_C and even indices with evolution under H_0. Consequently, the transition matrix elements A_{ij} are time-dependent functions $\lambda(t)$ such that the only non-zero off-diagonal entries are $A_{k-1,k}(t) = \lambda_{k-1}(t)$ (diagonal elements are defined to enforce conservation of probability). Figure 3.1 illustrates this structure for $p = 2$.

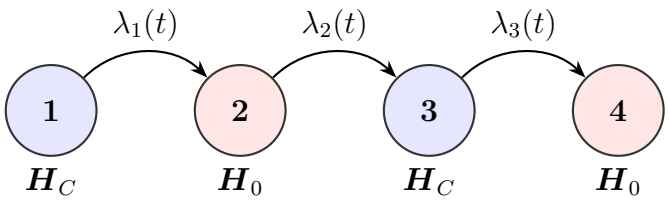


Figure 3.1: Markov chain graph for a QAOA ansatz of depth $p = 2$. Blue nodes represent evolution under the problem Hamiltonian $\boldsymbol{H}_C$, and red nodes represent the mixer $\boldsymbol{H}_0$. Transitions are governed by time-dependent rates $\lambda_k(t)$.

The transition matrix A^T takes the form:

$$A^T(t) = \begin{pmatrix} -\lambda_1(t) & 0 & 0 & \cdots & 0 \\ \lambda_1(t) & -\lambda_2(t) & 0 & \cdots & 0 \\ 0 & \lambda_2(t) & -\lambda_3(t) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \lambda_{2p-1}(t) & 0 \end{pmatrix}, \tag{3.8}$$

and the general system (3.7) simplifies to a cascading set of equations:

$$\begin{cases} \partial_t \rho_1 = -i[H_1, \rho_1] - \lambda_1(t)\rho_1 \\ \partial_t \rho_k = -i[H_k, \rho_k] - \lambda_k(t)\rho_k + \lambda_{k-1}(t)\rho_{k-1} \quad \text{for } 1 < k < 2p \\ \partial_t \rho_{2p} = -i[H_{2p}, \rho_{2p}] + \lambda_{2p-1}(t)\rho_{2p-1} \end{cases} \tag{3.9}$$

The variational parameters of the algorithm are now encoded in the rate functions $\lambda_k(t)$.

3.4.2 Relationship to Standard QAOA

To connect this continuous-time framework to the standard gate-based QAOA, we introduce an alternative parameterization. A single realization of the stochastic process corresponds to a specific sequence of switching instants $\vec{\tau} = (\tau_1, \ldots, \tau_{2p-1})$, where τ_k is the time at which the system transitions from node k to node $k+1$. The duration spent in the k-th layer is then:

$$\Delta_k = \tau_k - \tau_{k-1}, \quad (\text{with } \tau_0 = 0). \tag{3.10}$$

These durations Δ_k are precisely the QAOA angles $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ in the standard formulation. A deterministic choice of switching instants $\vec{\tau}$ thus defines a deterministic QAOA circuit. Optimizing the rate functions $\lambda_k(t)$ in our framework is therefore equivalent to optimizing a *probability distribution* over the QAOA angles.

3.4.3 The Optimization Objective

We now define the quantities to be optimized. We distinguish between the energy associated with a single trajectory and the average energy over all trajectories.

Single-trajectory energy. Each realization of the stochastic process produces a specific sequence of switching instants $\vec{\tau} = (\tau_1, \ldots, \tau_{2p-1})$, sampled according to the rates $\lambda_k(t)$. Conditioned on this realization, the system follows a deterministic pure-state evolution $|\psi(\vec{\tau})\rangle$, and its energy is:

$$\mathcal{E}(\vec{\tau}) := \langle\psi(\vec{\tau})|H_C|\psi(\vec{\tau})\rangle. \tag{3.11}$$

We refer to this as the *single-trajectory energy*. It is mathematically identical to the standard QAOA expected cost $\mathcal{C}_p(\boldsymbol{\gamma}, \boldsymbol{\beta})$ under the coordinate transformation described above. The function $\mathcal{E}(\vec{\tau})$ defines the energy landscape (the geometric surface over the parameter space determining the difficulty of the optimization).

Ensemble state and stochastic average. The stochastic process induces a mixture over all possible trajectories. The density matrix of the quantum system, obtained by tracing out the Markov register, is:

$$\rho(t) := \mathrm{Tr}_{mark}(\boldsymbol{\mathcal{P}}(t)) = \sum_{i=1}^{2p} \rho_i(t). \tag{3.12}$$

The optimization target is the *stochastic average* of the energy at final time T:

$$\bar{\mathcal{E}}(T) := \mathrm{Tr}(H_C\,\rho(T)) = \mathbb{E}_{\vec{\tau}}\big[\mathcal{E}(\vec{\tau})\big]. \tag{3.13}$$

This involves a double averaging: each $\mathcal{E}(\vec{\tau})$ is itself the quantum expectation $\langle\psi(\vec{\tau})|H_C|\psi(\vec{\tau})\rangle$ (average over measurement outcomes), and $\bar{\mathcal{E}}$ then averages over the distribution of switching instants induced by the rates $\lambda_k(t)$.

The optimization problem. The variational problem we study is:

$$\min_{\{\lambda_k(t)\}} \bar{\mathcal{E}}(T) = \min_{\{\lambda_k(t)\}} \mathrm{Tr}(H_C\,\rho(T)), \tag{3.14}$$

subject to the dynamics (3.9) and the initial condition $\rho(0) = \rho_1(0) = |\psi_0\rangle\langle\psi_0|$, where $|\psi_0\rangle$ is the uniform superposition (Section 2.4.1).

This formulation transforms the discrete gate-based optimization into a continuous control problem. Instead of searching directly over angles, we search over rate functions that induce distributions on angles. As we will see, this perspective enables analytical tools such as differential equations and spectral methods.

3.4.4 Analytic solution of classical Markovian System

For developments to come, solving the classical Markovian system will bring us more insight into the form and type of solution we can expect for the coupled quantum system developed in the next chapters.

Starting from equation (3.1) with the matrix $A^T(t)$ described in (3.8), the system of differential equations for the classical probabilities $p_k(t)$ reads:

$$\begin{cases} \partial_t p_1 = -\lambda_1(t) p_1 \\ \partial_t p_k = -\lambda_k(t) p_k + \lambda_{k-1}(t) p_{k-1} \quad \text{for } 1 < k < 2p \\ \partial_t p_{2p} = \lambda_{2p-1}(t) p_{2p-1} \end{cases} \tag{3.15}$$

with the initial condition $p_1(0) = 1$ and $p_k(0) = 0$ for all $k > 1$ (the system starts deterministically in node 1).

General solution. The cascading structure of this system allows for a recursive solution. We first introduce several quantities from survival analysis that will prove useful throughout this thesis.

The *integrated rate* (or cumulative hazard) associated with node k is defined as:

$$\Lambda_k(t) := \int_0^t \lambda_k(s)\, ds. \tag{3.16}$$

The *survival function* at node k represents the probability that a system arriving at node k at time 0 has not yet transitioned to node $k+1$ by time t:

$$S_k(t) := e^{-\Lambda_k(t)}. \tag{3.17}$$

More generally, for a system arriving at node k at time s, the conditional survival probability until time $t \geq s$ is given by the ratio $S_k(t)/S_k(s) = e^{-\int_s^t \lambda_k(u)\, du}$.

Finally, the *transition density* is defined as:

$$\eta_k(t) := \lambda_k(t)\, S_k(t) = \lambda_k(t)\, e^{-\Lambda_k(t)}. \tag{3.18}$$

This quantity represents the probability density of leaving node k at time t, assuming arrival at time 0. It satisfies $\int_0^\infty \eta_k(t)\, dt = 1$ when $\Lambda_k(\infty) = \infty$, confirming its interpretation as a probability density.

With these definitions, the first equation of system (3.15) yields a simple exponential decay:

$$p_1(t) = S_1(t) = e^{-\Lambda_1(t)}. \tag{3.19}$$

For subsequent nodes, each equation is a first-order linear ODE that can be solved using an integrating factor. The solution for p_k with $k > 1$ is:

$$p_k(t) = e^{-\Lambda_k(t)} \int_0^t \lambda_{k-1}(s)\, p_{k-1}(s)\, e^{\Lambda_k(s)}\, ds. \tag{3.20}$$

For the final node $k = 2p$, where $\lambda_{2p}(t) = 0$ (hence $S_{2p}(t) = 1$), this simplifies to:

$$p_{2p}(t) = \int_0^t \lambda_{2p-1}(s)\, p_{2p-1}(s)\, ds. \tag{3.21}$$

Interpretation. The recursive formula (3.20) can be understood by marginalizing over arrival times. To occupy node k at time t, the system must have arrived from node $k-1$ at some earlier time $s \in [0, t]$, and then survived in node k until time t. The integrand decomposes accordingly: $p_{k-1}(s)$ is the probability of being in node $k-1$ at time s, $\lambda_{k-1}(s)\, ds$ is the probability of transitioning during $[s, s+ds]$, and $S_k(t)/S_k(s)$ is the conditional survival probability from s to t. This structure reflects the sequential nature of the directed chain: probability must flow through each intermediate node before reaching subsequent ones. For the final node, the absorbing condition $\lambda_{2p} = 0$ makes $p_{2p}(t)$ purely accumulative (the probability that enters node $2p$ remains there indefinitely, as expected for the terminal state of the protocol).

Chapter 4

Analysis of Single-Switch Strategies

This chapter focuses on the single-switch regime within the randomized QAOA framework, corresponding to the standard QAOA at depth $p = 1$. In this setting, a continuous-time Markov process governs the switching dynamics between the cost and mixer Hamiltonians.

We begin with the simplest non-trivial case: a 2-level quantum system controlled by a 2-state Markov chain. This toy model allows us to derive explicit analytical expressions for the marginal density matrices and the stochastic average $\bar{\mathcal{E}}$ as a functional of the transition rate $\lambda(t)$. We then study the deterministic optimal switch on the Bloch sphere to build geometric intuition and establish energy bounds. Building on this, we introduce an *entropy regularization framework* to resolve the ill-posedness of the functional optimization, allowing us to derive a unique, smooth transition rate. Finally, we show that in certain asymptotic regimes, the randomized approach achieves the minimum of the single-trajectory energy $\mathcal{E}(\tau)$.

4.1 2-level system

To initiate our analysis of single-switch strategies, we consider the most elementary non-trivial case: a single qubit system ($\mathcal{H}_{\text{sys}} \cong \mathbb{C}^2$) controlled by a Markov chain with two states ($Q = 2$). This setup serves as a toy model to investigate the fundamental dynamics of the randomized approach before generalizing to multi-level systems.

Classical Control Dynamics

We define the control process as a directed, time-inhomogeneous Markov chain consisting of two nodes. The system is initialized in node 1 and transitions to node 2 with a time-dependent rate $\lambda(t)$. Once in node 2, the transition process from a Hamiltonian to another terminates (but the quantum system continues to evolve under the last Hamiltonian). The classical probability distribution $\vec{p}(t) = (p_1(t), p_2(t))^T$ evolves according to the transition matrix derived in Section 3.1:

$$\partial_t \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = A^T \vec{p} = \begin{pmatrix} -\lambda(t) & 0 \\ \lambda(t) & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}, \tag{4.1}$$

yielding the coupled equations:

$$\begin{aligned} \partial_t p_1(t) &= -\lambda(t) p_1(t), \\ \partial_t p_2(t) &= \lambda(t) p_1(t). \end{aligned} \tag{4.2}$$

This classical process is illustrated in Figure 4.1.

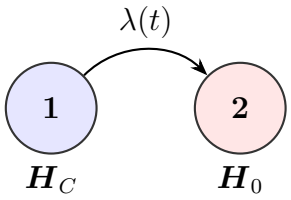


Figure 4.1: Schematic of the directed Markov chain for the 2-level system. The system evolves under H_C in node 1 and transitions with rate $\lambda(t)$ to node 2, where it evolves under H_0.

Hamiltonian Encoding

In accordance with the QAOA ansatz structure described in Section 2.4.1, we assign the problem (cost) Hamiltonian to the first node and the mixer Hamiltonian to the second node. For this 2-level illustrative example, we define:

1. **Node 1 (Problem Hamiltonian H_C):** We consider a Hamiltonian diagonal in the computational basis, energetically penalizing the state $|1\rangle$:

$$H_C = E_0|0\rangle\langle 0| + E_1|1\rangle\langle 1| = |1\rangle\langle 1|. \tag{4.3}$$

   Here we have set $E_0 = 0$ and $E_1 = 1$. This acts as a simple cost function where the associated minimum is $z = 0$ and the corresponding ground state is $|0\rangle$.

2. **Node 2 (Mixer Hamiltonian H_0):** We select the standard transverse field mixer, typically used to generate a superposition:

$$H_0 = -\frac{1}{2}\sigma_x = -\frac{1}{2}\big(|0\rangle\langle 1| + |1\rangle\langle 0|\big). \tag{4.4}$$

We artificially add a factor of $\frac{1}{2}$ to the mixer Hamiltonian to simplify later expressions. The system is initialized in the standard QAOA starting state, the uniform superposition:

$$\rho(0) = |+\rangle\langle +|, \quad \text{where } |+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}. \tag{4.5}$$

This state is the ground state of the mixer Hamiltonian H_0.

Marginal Density Matrix Equations

Applying the formalism developed in (Dubus, Cunningham, and Roland 2025), specifically the directed regime equations (Eq. 3.9), the evolution of the quantum state is described by the marginal density matrices ρ_1 and ρ_2. The total system state is $\rho(t) = \rho_1(t) + \rho_2(t)$, where ρ_i is the subnormalized state conditioned on the Markov chain being in node i.

The master equation governing this evolution is:

$$\partial_t \rho_i = -i[H_i, \rho_i] + \sum_{j \neq i} \left(A_{ji}\rho_j - A_{ij}\rho_i\right). \tag{4.6}$$

Substituting the specific transition rates and Hamiltonians for our 2-level system, we obtain the explicit system of ordinary differential equations:

$$\begin{cases} \partial_t \rho_1(t) = -i[H_C, \rho_1(t)] - \lambda(t)\rho_1(t), \\ \partial_t \rho_2(t) = -i[H_0, \rho_2(t)] + \lambda(t)\rho_1(t). \end{cases} \tag{4.7}$$

Here, the term $-\lambda(t)\rho_1$ represents the incoherent loss of population from node 1, which acts as a source term $+\lambda(t)\rho_1$ for node 2.

Optimization Objective

The goal is to control the transition rate $\lambda(t)$ such that the stochastic average of the energy is minimized at the final time T. Following the notation established in Chapter 3, this is:

$$\min_{\lambda(t)} \bar{\mathcal{E}}(T) = \min_{\lambda(t)} \mathrm{Tr}\left(H_C \rho(T)\right) = \min_{\lambda(t)} \mathrm{Tr}\left(H_C(\rho_1(T) + \rho_2(T))\right). \tag{4.8}$$

This transforms the variational problem from finding discrete angles (γ, β) into a continuous optimal control problem over the function $\lambda(t)$.

4.1.1 Analytical Derivation of the State Evolution

To gain further insight into the dynamics, we can derive analytical expressions for the sub-normalized density matrices $\rho_1(t)$ and $\rho_2(t)$.

First, we solve for $\rho_1(t)$. We propose a density matrix of the form:

$$\rho_1(t) = e^{-iH_C t}\tilde{\rho}_1(t)e^{iH_C t}. \tag{4.9}$$

To verify this ansatz, we differentiate it with respect to time using the product rule:

$$\begin{aligned}
\partial_t \rho_1(t) &= (\partial_t e^{-iH_C t})\tilde{\rho}_1(t)e^{iH_C t} + e^{-iH_C t}(\partial_t \tilde{\rho}_1(t))e^{iH_C t} + e^{-iH_C t}\tilde{\rho}_1(t)(\partial_t e^{iH_C t}) \\
&= -iH_C\rho_1(t) + e^{-iH_C t}\dot{\tilde{\rho}}_1(t)e^{iH_C t} + i\rho_1(t)H_C \\
&= -i[H_C, \rho_1(t)] + e^{-iH_C t}\dot{\tilde{\rho}}_1(t)e^{iH_C t}.
\end{aligned} \tag{4.10}$$

We now compare this expression with the master equation for the first node, given in (4.7):

$$\partial_t \rho_1(t) = -i[H_C, \rho_1(t)] - \lambda(t)\rho_1(t). \tag{4.11}$$

By equating the two expressions for $\partial_t \rho_1(t)$, the commutator term $-i[H_C, \rho_1(t)]$ cancels out on both sides:

$$e^{-iH_C t}\dot{\tilde{\rho}}_1(t)e^{iH_C t} = -\lambda(t)e^{-iH_C t}\tilde{\rho}_1(t)e^{iH_C t}. \tag{4.12}$$

Multiplying by $e^{iH_C t}$ from the left and $e^{-iH_C t}$ from the right removes the unitary operators, leaving a simple scalar differential equation for the envelope $\tilde{\rho}_1(t)$:

$$\partial_t \tilde{\rho}_1(t) = -\lambda(t)\tilde{\rho}_1(t). \tag{4.13}$$

Integrating this scalar differential equation yields:

$$\tilde{\rho}_1(t) = \tilde{\rho}_1(0)e^{-\int_0^t \lambda(\tau)\,d\tau}. \tag{4.14}$$

Using the initial condition $\tilde{\rho}_1(0) = \rho_1(0) = \rho(0)$, we obtain the solution for the first node:

$$\rho_1(t) = e^{-\Lambda(t)}e^{-iH_C t}\rho(0)e^{iH_C t}, \tag{4.15}$$

where we used the definition of the cumulative rate $\Lambda(t) = \int_0^t \lambda(\tau)\,d\tau$.

Next, we address the evolution of $\rho_2(t)$. Equation (4.7) for the second node can be written:

$$\partial_t \rho_2(t) + i[H_0, \rho_2(t)] = \lambda(t)\rho_1(t). \tag{4.16}$$

Substituting the solution we found for $\rho_1(t)$, we identify the right hand side $R(t)$:

$$R(t) \equiv \lambda(t)\rho_1(t) = \lambda(t)e^{-\Lambda(t)}e^{-iH_C t}\rho(0)e^{iH_C t}. \tag{4.17}$$

To solve this inhomogeneous differential equation, we move to the interaction picture with respect to H_0. We define the transformed state:

$$\tilde{\rho}_2(t) = e^{iH_0 t}\rho_2(t)e^{-iH_0 t}. \tag{4.18}$$

We differentiate this expression with respect to time to see how it evolves:

$$\begin{aligned}
\partial_t\tilde{\rho}_2(t) &= (\partial_t e^{iH_0 t})\rho_2(t)e^{-iH_0 t} + e^{iH_0 t}(\partial_t\rho_2(t))e^{-iH_0 t} + e^{iH_0 t}\rho_2(t)(\partial_t e^{-iH_0 t}) \\
&= iH_0 e^{iH_0 t}\rho_2(t)e^{-iH_0 t} + e^{iH_0 t}\dot{\rho}_2(t)e^{-iH_0 t} - ie^{iH_0 t}\rho_2(t)H_0 e^{-iH_0 t} \\
&= e^{iH_0 t}\left(\dot{\rho}_2(t) + iH_0\rho_2(t) - i\rho_2(t)H_0\right)e^{-iH_0 t} \\
&= e^{iH_0 t}\left(\partial_t\rho_2(t) + i[H_0, \rho_2(t)]\right)e^{-iH_0 t}.
\end{aligned} \tag{4.19}$$

The term in the parentheses is exactly the left-hand side of the equations 4.16, which equals $R(t)$. Thus, the evolution in the interaction picture simplifies to:

$$\partial_t\tilde{\rho}_2(t) = e^{iH_0 t}R(t)e^{-iH_0 t}. \tag{4.20}$$

Integrating from 0 to t and using the initial condition $\rho_2(0) = 0$ (implies $\tilde{\rho}_2(0) = 0$):

$$\tilde{\rho}_2(t) = \int_0^t e^{iH_0\tau}R(\tau)e^{-iH_0\tau}\,d\tau. \tag{4.21}$$

Finally, transforming back to the Schrödinger picture gives:

$$\rho_2(t) = e^{-iH_0 t}\tilde{\rho}_2(t)e^{iH_0 t} = \int_0^t e^{-iH_0(t-\tau)}R(\tau)e^{iH_0(t-\tau)}\,d\tau. \tag{4.22}$$

Substituting the explicit form of $R(\tau)$, we arrive at the full analytical expression for the state in the second node:

$$\rho_2(t) = \int_0^t \lambda(\tau)e^{-\int_0^\tau \lambda(s)\,ds}\,e^{-iH_0(t-\tau)}e^{-iH_C\tau}\rho(0)e^{iH_C\tau}e^{iH_0(t-\tau)}\,d\tau. \tag{4.23}$$

This equation has a clear physical interpretation: at any time τ, a portion of the population transitions from node 1 to node 2 with rate $\lambda(\tau)$. Before the transition, the system evolves under H_C for duration τ. After the transition, it evolves under H_0 for the remaining duration $t-\tau$. The total state is the coherent sum (integral) of these histories, weighted by the factor $\lambda(\tau)e^{-\int_0^\tau \lambda(s)\,ds}$. This is the probability density $\eta(\tau)$ of the transition occurring at τ as defined in Section 3.4.4.

4.1.2 Explicit Calculation of the Stochastic Average

With the analytical expressions for the sub-normalized density matrices in hand, we can now evaluate the stochastic average $\bar{\mathcal{E}}(T)$ defined in Equation (4.8). For the problem Hamiltonian $H_C = |1\rangle\langle 1|$:

$$\bar{\mathcal{E}}(T) = \mathrm{Tr}(H_C\rho(T)) = \langle 1|\rho_1(T)|1\rangle + \langle 1|\rho_2(T)|1\rangle. \tag{4.24}$$

First, we compute the contribution from the first node. Using the solution for $\rho_1(t)$:

$$\rho_1(T) = e^{-\Lambda(T)}e^{-iH_C T}\rho(0)e^{iH_C T}. \tag{4.25}$$

Given the initial state $\rho(0) = |+\rangle\langle +|$, the population in $|1\rangle$ evolves as:

$$\langle 1|e^{-iH_C T}|+\rangle\langle +|e^{iH_C T}|1\rangle = |\langle 1|e^{-iH_C T}|+\rangle|^2. \tag{4.26}$$

Since $H_C|1\rangle = |1\rangle$ and $H_C|0\rangle = 0$, we have:

$$e^{-iH_C T}|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{-iT}|1\rangle\right). \tag{4.27}$$

The probability of measuring $|1\rangle$ is simply $|\frac{1}{\sqrt{2}}e^{-iT}|^2 = \frac{1}{2}$. Thus:

$$\langle 1|\rho_1(T)|1\rangle = \frac{1}{2}e^{-\Lambda(T)}. \tag{4.28}$$

Next, we compute the contribution from the second node. The state $\rho_2(T)$ is given by the integral over transition times τ:

$$\rho_2(T) = \int_0^T \eta(\tau)\,\mathcal{U}(\tau,T)\rho(0)\mathcal{U}^\dagger(\tau,T)\,d\tau, \tag{4.29}$$

where $\mathcal{U}(\tau,T) = e^{-iH_0(T-\tau)}e^{-iH_C\tau}$ represents the unitary evolution of a trajectory that switches at time τ. The contribution to the stochastic average is:

$$\langle 1|\rho_2(T)|1\rangle = \int_0^T \eta(\tau)\,P_1(\tau,T)\,d\tau, \tag{4.30}$$

where $P_1(\tau,T) = |\langle 1|\mathcal{U}(\tau,T)|+\rangle|^2$ is the probability of finding the system in state $|1\rangle$ for a specific switching time τ.

We evaluate the state vector $|\psi(\tau,T)\rangle = \mathcal{U}(\tau,T)|+\rangle$:

$$|\psi(\tau,T)\rangle = e^{-iH_0(T-\tau)}\frac{1}{\sqrt{2}}\left(|0\rangle + e^{-i\tau}|1\rangle\right). \tag{4.31}$$

Recalling that $H_0 = -\frac{1}{2}\sigma_x$, we can expand the evolution operator using the Taylor series definition of the matrix exponential. Let $\theta = \frac{T-\tau}{2}$. Using the property that $\sigma_x^2 = \mathbb{I}$ (and thus $\sigma_x^{2k} = \mathbb{I}$ and $\sigma_x^{2k+1} = \sigma_x$), we have:

$$\begin{aligned} e^{i\theta\sigma_x} &= \sum_{k=0}^{\infty}\frac{(i\theta\sigma_x)^k}{k!} \\ &= \sum_{k=0}^{\infty}\frac{(i\theta)^{2k}\sigma_x^{2k}}{(2k)!} + \sum_{k=0}^{\infty}\frac{(i\theta)^{2k+1}\sigma_x^{2k+1}}{(2k+1)!} \\ &= \left(\sum_{k=0}^{\infty}\frac{(-1)^k\theta^{2k}}{(2k)!}\right)\mathbb{I} + \left(i\sum_{k=0}^{\infty}\frac{(-1)^k\theta^{2k+1}}{(2k+1)!}\right)\sigma_x \\ &= \cos(\theta)\mathbb{I} + i\sin(\theta)\sigma_x. \end{aligned} \tag{4.32}$$

Substituting back $\theta = \frac{T-\tau}{2}$, the evolution operator is:

$$e^{-iH_0(T-\tau)} = e^{i\frac{\sigma_x}{2}(T-\tau)} = \cos\left(\frac{T-\tau}{2}\right)\mathbb{I} + i\sin\left(\frac{T-\tau}{2}\right)\sigma_x. \tag{4.33}$$

Applying this to the state (and noting $\sigma_x|0\rangle = |1\rangle, \sigma_x|1\rangle = |0\rangle$):

$$|\psi(\tau,T)\rangle = \frac{1}{\sqrt{2}}\left[\cos\left(\frac{T-\tau}{2}\right)\begin{pmatrix}1\\e^{-i\tau}\end{pmatrix} + i\sin\left(\frac{T-\tau}{2}\right)\begin{pmatrix}e^{-i\tau}\\1\end{pmatrix}\right]. \tag{4.34}$$

The amplitude of the $|1\rangle$ component is:

$$\langle 1|\psi(\tau,T)\rangle = \frac{1}{\sqrt{2}}\left(e^{-i\tau}\cos\left(\frac{T-\tau}{2}\right) + i\sin\left(\frac{T-\tau}{2}\right)\right). \tag{4.35}$$

The probability $P_1(\tau, T)$ is the squared modulus of this amplitude:

$$\begin{aligned} P_1(\tau,T) &= \frac{1}{2}\left| e^{-i\tau}\cos\left(\frac{T-\tau}{2}\right) + i\sin\left(\frac{T-\tau}{2}\right)\right|^2 \\ &= \frac{1}{2}\left(\cos^2\left(\frac{T-\tau}{2}\right) + \sin^2\left(\frac{T-\tau}{2}\right) + 2\mathrm{Re}\left\{-ie^{i\tau}\cos\left(\frac{T-\tau}{2}\right)\sin\left(\frac{T-\tau}{2}\right)\right\}\right) \\ &= \frac{1}{2}\left(1-\sin(T-\tau)\sin(\tau)\right). \end{aligned} \tag{4.36}$$

Substituting this back into the expression for $\langle 1|\rho_2(T)|1\rangle$:

$$\langle 1|\rho_2(T)|1\rangle = \frac{1}{2}\int_0^T \eta(\tau)\,d\tau - \frac{1}{2}\int_0^T \eta(\tau)\sin(T-\tau)\sin(\tau)\,d\tau. \tag{4.37}$$

We recognize the first integral as the total probability of switching to node 2:

$$\int_0^T \eta(\tau)\,d\tau = 1 - e^{-\Lambda(T)}. \tag{4.38}$$

Combining the contributions from ρ_1 and ρ_2:

$$\begin{aligned} \bar{\mathcal{E}}(T) &= \frac{1}{2}e^{-\Lambda(T)} + \frac{1}{2}\left(1-e^{-\Lambda(T)}\right) - \frac{1}{2}\int_0^T \eta(\tau)\sin(T-\tau)\sin(\tau)\,d\tau \\ &= \frac{1}{2}\left(1 - \int_0^T \eta(\tau)\sin(T-\tau)\sin(\tau)\,d\tau\right). \end{aligned} \tag{4.39}$$

This result reduces the control problem to minimizing a scalar integral functional over the transition rate $\lambda(t)$ encoded in $\eta(t)$.

Anticipating the normalization constraint. We can already observe that the optimal strategy will require the switch to happen with certainty during the interval $[0, T]$. Indeed, the term $\sin(T-\tau)\sin(\tau)$ can be made positive for appropriate choices of τ, meaning that concentrating probability mass at such times reduces the stochastic average. If the system were allowed to *not* switch (i.e., $\int_0^T \eta < 1$), we would be "wasting" probability that could otherwise contribute to lowering $\bar{\mathcal{E}}(T)$.

Anticipating this, suppose we impose $\int_0^T \eta(\tau)\,d\tau = 1$. Then the constant $\frac{1}{2}$ can be absorbed into the integral:

$$\bar{\mathcal{E}}(T) = \frac{1}{2} - \frac{1}{2}\int_0^T \eta(\tau)\sin(T-\tau)\sin(\tau)\,d\tau = \int_0^T \eta(\tau)\cdot\mathcal{E}(\tau,T)\,d\tau, \tag{4.40}$$

where we identify the *single-trajectory energy*

$$\mathcal{E}(\tau,T) = \frac{1}{2}\Big(1-\sin(T-\tau)\sin(\tau)\Big), \tag{4.41}$$

as the energy of a deterministic switch occurring at time τ. Under the normalization constraint, the stochastic average $\bar{\mathcal{E}}(T)$ is simply the expectation of the single-trajectory energy over the distribution η. The optimization problem becomes: find the probability distribution η over $[0, T]$ that minimizes $\mathbb{E}_{\tau\sim\eta}[\mathcal{E}(\tau, T)]$.

Lower bound. Since $H_C = |1\rangle\langle 1|$ is a positive semi-definite operator (a projector with eigenvalues 0 and 1), its expectation value with respect to any valid density matrix $\rho(T)$ must be non-negative. Therefore, regardless of the specific form of the transition density $\eta(t)$ or the duration T:

$$\bar{\mathcal{E}}(T) \geq 0. \tag{4.42}$$

This physical constraint provides a trivial lower bound for the optimization problem.

4.2 The Energy Landscape and Deterministic Bounds

Before exploring the entropy regularization framework for the randomized approach, we first analyze the deterministic limit to build geometric intuition and establish the optimal conditions for the problem. Recall that minimizing the energy functional $\mathcal{E}(\tau, T)$ with respect to the switching time τ is equivalent to optimizing the variational parameters to reach the target ground state $|0\rangle$ (the solution with the lowest classical cost). This analysis will reveal the "energy landscape" and justify the normalization constraint that we will later impose on the transition density.

In the deterministic limit, the transition density approaches a Dirac delta function $\eta(t) = \delta(t - \tau)$, representing a single sharp switch at time τ. By finding the optimal deterministic parameters, we identify the global minima of the underlying energy landscape $\mathcal{E}(\tau, T)$.

4.2.1 Geometrical Interpretation

The evolution of the single qubit can be visualized on the Bloch sphere. We seek a trajectory that transforms the initial state $|+\rangle$ into the target ground state $|0\rangle$.

1. **Phase 1 (H_C):** The system evolves under $H_C = |1\rangle\langle 1|$. Since $|+\rangle$ lies on the equator of the Bloch sphere, H_C (which generates rotations around the Z-axis) rotates the state along the equator. The goal is to rotate $|+\rangle$ to a position where the subsequent Hamiltonian can act most effectively.

2. **Phase 2 (H_0):** At time τ, the system switches to $H_0 = -\frac{1}{2}\sigma_x$. This generates rotations around the X-axis. To reach the north pole ($|0\rangle$), the state at time τ must lie on the circle defined by the intersection of the sphere and the YZ-plane (specifically at $\pm i|+\rangle$ or similar variants).

This geometric intuition suggests that specific "resonant" switching times τ and total durations T are required to perfectly navigate from $|+\rangle$ to $|0\rangle$. This is illustrated in Figures 4.2 and 4.3 for two distinct optimal switching families.

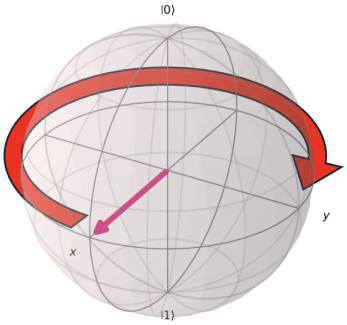
4.2.2 Algebraic Derivation of Optimal Times

We can rigorously derive these times by minimizing the deterministic single-trajectory energy. Let the system evolve with H_C for duration τ and H_0 for duration $T - \tau$. From Eq. (4.41), the single-trajectory energy is:

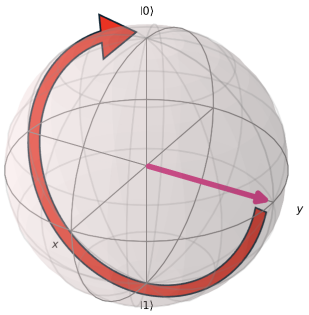
$$\mathcal{E}(\tau, T) = \frac{1}{2}\Big(1 - \sin(T - \tau)\,\sin(\tau)\Big). \tag{4.43}$$

We seek the global minimum of this function with respect to the switching time τ for a fixed total time T. Since $|\sin(T - \tau)\sin(\tau)| \leq 1$, the theoretical lower bound is 0. This minimum is achieved if and only if:

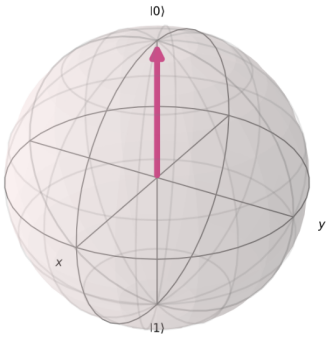
$$\sin(T - \tau)\sin(\tau) = 1. \tag{4.44}$$



(a) Initial state $|+\rangle$, evolving under H_C.

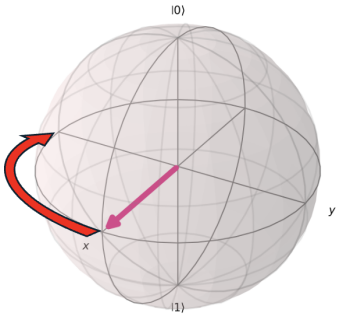


(b) Optimal switch point: state rotates to $i|+\rangle$ at $\tau = \frac{3\pi}{2}$.

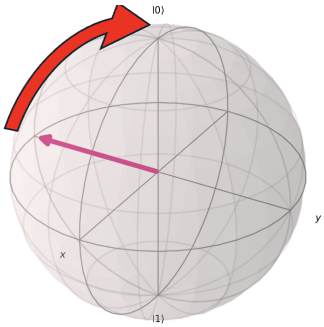


(c) Final rotation under H_0 reaches $|0\rangle$.

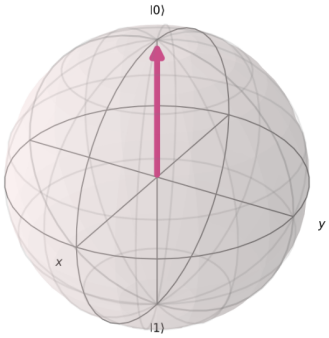
Figure 4.2: State evolution for the first solution family ($\tau = 3\pi/2$). The system rotates around Z, switches, and then rotates around X to the ground state.



(a) Initial state $|+\rangle$, evolving under H_C.



(b) Optimal switch point: state rotates to $i|+\rangle$ at $\tau = \frac{\pi}{2}$.



(c) Final rotation under H_0 reaches $|0\rangle$.

Figure 4.3: State evolution for the second solution family ($\tau = \pi/2$).

This condition splits into two coupled requirements:

$$\sin(T - \tau) = \pm 1 \quad \text{and} \quad \sin(\tau) = \pm 1. \tag{4.45}$$

Solving these trigonometric equations yields two families of optimal switching times τ:

$$\tau = \frac{\pi}{2} + 2\pi m \quad \text{or} \quad \tau = \frac{3\pi}{2} + 2\pi m, \qquad m \in \mathbb{Z}_{\geq 0}. \tag{4.46}$$

Furthermore, satisfying the condition for the second duration implies that the total time T must be an odd integer multiple of π:

$$\boxed{T = (2k+1)\pi, \qquad k \in \mathbb{Z}_{\geq 0}.} \tag{4.47}$$

This derivation provides insight for the randomized optimization problem. It reveals that the single-trajectory energy $\mathcal{E}(\tau, T)$ possesses global minima at zero energy only if we select the total runtime T to be an odd multiple of π.

Simplified single-trajectory energy for optimal T. When $T = (2k+1)\pi$, a useful simplification occurs. Using the identity $\sin(T - \tau) = \sin((2k+1)\pi - \tau) = (-1)^{2k+1}\sin(-\tau) = \sin(\tau)$, the single-trajectory energy becomes:

$$\mathcal{E}(\tau, T) = \frac{1}{2}\Big(1 - \sin^2(\tau)\Big) = \frac{1}{2}\cos^2(\tau). \tag{4.48}$$

This reveals a simple "potential well" structure: the energy vanishes at $\tau = \frac{\pi}{2} + n\pi$ for integer n, corresponding to the optimal switching times derived above.

In the general analysis that follows, we retain the full expression $\mathcal{E}(\tau, T) = \frac{1}{2}(1 - \sin(T - \tau)\sin(\tau))$ to handle arbitrary T.

4.3 Entropy Regularization Framework

Having established the geometric intuition and optimal deterministic bounds in the previous section, we now turn to the randomized approach. The key challenge is to find a well-posed optimization problem that yields a continuous, smooth transition rate $\lambda(\tau)$ rather than a degenerate Dirac delta solution. This will then allow us to leverage asymptotic analysis techniques to study the behavior of the stochastic average.

4.3.1 Functional Optimization Strategies

The optimization problem can now be reformulated as minimizing the stochastic average $\bar{\mathcal{E}}(T)$ with respect to the transition rate $\lambda(\tau)$:

$$\min_{\lambda(\tau)} \bar{\mathcal{E}}(T) = \min_{\lambda(\tau)} \int_0^T \eta(\tau)\, \mathcal{E}(\tau, T)\, d\tau, \tag{4.49}$$

where $\eta(\tau) = \lambda(\tau) e^{-\int_0^\tau \lambda(s)\, ds}$ is the transition density, subject to the constraints $\lambda(\tau) \geq 0$ for all $\tau \in [0, T]$ and $\int_0^T \eta(\tau)\, d\tau = 1$.

The direct minimization of this functional is ill-posed: without additional constraints, the optimal solution is a Dirac delta function $\eta(\tau) = \delta(\tau - \tau^*)$ concentrated at the minimum of $\mathcal{E}(\tau, T)$, corresponding to the deterministic switching analyzed in Section 4.2. Standard approaches such as direct calculus of variations or L^2 norm regularization lead to either non-smooth solutions or intractable integro-differential equations (see Appendix A for details).

Regularization strategy. To obtain a well-posed optimization problem with a unique, smooth solution, we employ a regularization strategy. Instead of minimizing the raw stochastic average $\bar{\mathcal{E}}(T)$ directly, we introduce a parameterized auxiliary functional $J_\alpha[\eta]$, which includes a penalty term weighted by a regularization parameter $\alpha > 0$. This parameter acts as a "button": as $\alpha \to 0$, we recover the original problem; for $\alpha > 0$, the problem becomes strictly convex with a unique minimizer. This indirect approach is visualized in Figure 4.4.

The natural choice for the regularizer $\mathcal{R}[\eta]$ is the negative entropy functional $\int_0^T \eta(\tau) \ln \eta(\tau)\, d\tau$. This choice is justified by three key properties:

- **Barrier Property:** It implicitly enforces the physical constraint $\eta(\tau) \geq 0$ by diverging as $\eta \to 0$, naturally confining the solution to the probability simplex.
- **Uniqueness:** It is 1-strongly convex with respect to the L_1-norm (Kakade and Shalev-Shwartz 2009), which guarantees the existence of a unique global minimizer.[1]
- **Geometric Consistency:** In information geometry, negative entropy is the canonical choice for optimizations over probability measures (Amari and Cichocki 2010).

[1] A function f is μ-strongly convex w.r.t. a norm $\|\cdot\|$ if $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|^2$ for all x, y in the domain. This ensures the function curves upward at least quadratically.

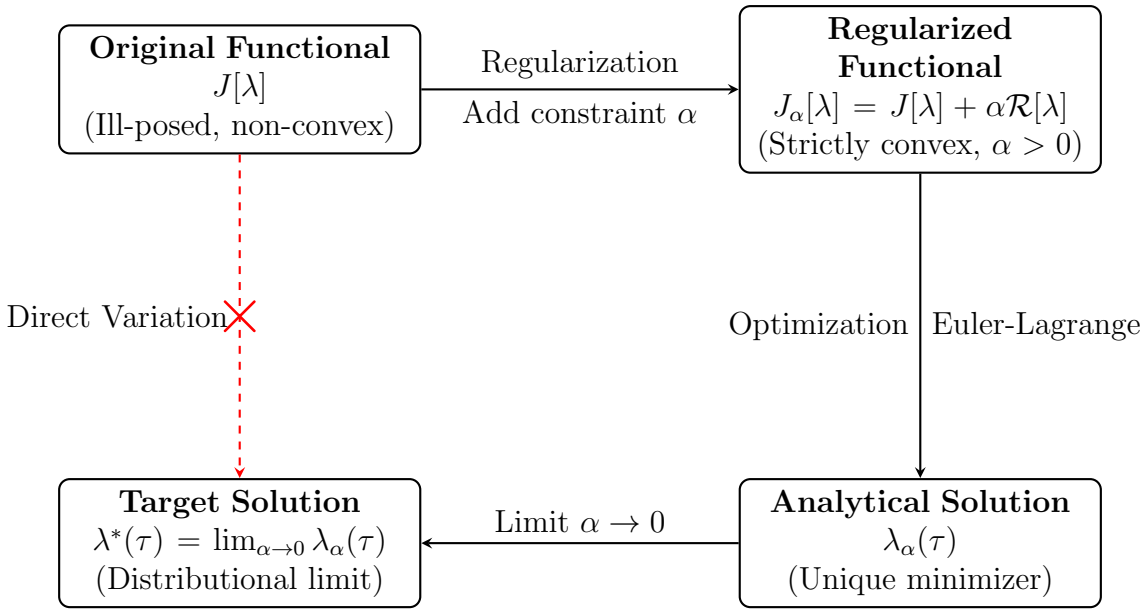


Figure 4.4: Schematic of the regularization method. The direct optimization path (dashed red) is obstructed by the ill-posed nature of the functional. We bypass this by introducing a convex constraint α, solving the convex regularized problem for η_α, and taking the limit $\alpha \to 0$ to find the target solution.

We also reformulate the problem in terms of the transition density $\eta(\tau)$ rather than the transition rate $\lambda(\tau)$. These are related by the bijective mapping:

$$\lambda(\tau) = \frac{\eta(\tau)}{1 - \int_0^\tau \eta(s)\, ds}, \tag{4.50}$$

which is valid provided $\eta(\tau) \geq 0$ and $\int_0^T \eta(\tau)\, d\tau \leq 1$.

4.3.2 Formulation of the Entropy-Regularized Problem

Having motivated the use of entropy regularization, we now formulate the complete optimization problem. We minimize the regularized functional:

$$J_\alpha[\eta] = \int_0^T \eta(\tau)\, \mathcal{E}(\tau, T)\, d\tau \; + \; \alpha \int_0^T \eta(\tau) \ln \eta(\tau)\, d\tau, \tag{4.51}$$

subject to the normalization constraint $\int_0^T \eta(\tau)\, d\tau = 1$. The term $\mathcal{R}[\eta] = \int \eta \ln \eta$ acts as a barrier function, enforcing positivity ($\eta(\tau) \geq 0$) and ensuring strong convexity (Kakade and Shalev-Shwartz 2009). This guarantees the existence of a unique, smooth minimizer $\eta_\alpha(\tau)$.

As discussed, we optimize the functional with respect to the transition density $\eta(\tau)$ directly. The corresponding transition rate $\lambda(\tau)$ can be recovered via the mapping $\lambda(\tau) = \eta(\tau)/(1 - \int_0^\tau \eta)$. We now have all the ingredients to solve the regularized problem explicitly.

Since our single-switching model in a 2-level system is simple enough, we can analyze it directly on the qubit. This analysis will allow us to understand also the optimal angles or switching times are appropriate in order to minimize the cost function, i.e. to minimize the energy system's expectation value at time T.

4.3.3 Optimal transition rate

Now that we justified the normalization constraint, we can proceed to solve the regularized optimization problem (4.51) with standard calculus of variations techniques using the Euler-Lagrange

equations (Kielhöfer 2018):

Euler–Lagrange principle (recap). For a functional $J[\eta] = \int_0^T L\big(\tau, \eta(\tau), \dot\eta(\tau)\big)\, d\tau$ subject to appropriate boundary conditions ($\eta(0) = \eta(T) = 0$), any smooth minimizer η^* must satisfy the Euler–Lagrange equation

$$\frac{d}{d\tau}\frac{\partial L}{\partial \dot\eta}\big(\tau, \eta, \dot\eta\big) \;-\; \frac{\partial L}{\partial \eta}\big(\tau, \eta, \dot\eta\big) = 0. \tag{4.52}$$

When a constraint such as normalization $\int_0^T \eta(\tau)\, d\tau = 1$ is imposed, one introduces another Lagrange multiplier $\zeta \in \mathbb{R}$ and works with the augmented Lagrangian $\tilde L = L - \zeta\, \eta$. Stationarity then requires the Euler–Lagrange equation for $\tilde L$, together with the constraint itself.

We recall the Lagrangian integrand of our problem:

$$L(\tau, \eta, \dot\eta) = \eta(\tau)\, \mathcal{E}(\tau, T) + \alpha\, \eta(\tau)\, \ln \eta(\tau). \tag{4.53}$$

In our setting, the integrand depends only on $\eta(\tau)$ (no time derivative), so the Euler–Lagrange condition reduces to the first-order stationarity equation

$$\frac{\partial \tilde L}{\partial \eta}(\tau) = \mathcal{E}(\tau, T) + \alpha\big(\ln \eta(\tau) + 1\big) - \zeta = 0, \tag{4.54}$$

which is the principle we apply in the remainder to obtain our optimal solution.

$$\mathcal{E}(\tau, T) + \alpha(\ln \eta(\tau) + 1) - \zeta = 0 \quad \Longleftrightarrow \quad \ln \eta(\tau) = -\frac{\mathcal{E}(\tau, T)}{\alpha} + \frac{\zeta}{\alpha} - 1.$$

Exponentiating,

$$\eta(\tau) = \exp\Big(-\tfrac{\mathcal{E}(\tau,T)}{\alpha}\Big) \cdot \exp\Big(\tfrac{\zeta}{\alpha} - 1\Big) =: A\, e^{-\mathcal{E}(\tau,T)/\alpha},$$

where we have set $A := \exp(\zeta/\alpha - 1) > 0$.

We now eliminate ζ via the imposed normalization $\int_0^T \eta(\tau)\, d\tau = 1$:

$$1 = \int_0^T \eta(\tau)\, d\tau = A \int_0^T e^{-\mathcal{E}(\tau,T)/\alpha}\, d\tau.$$

Thus

$$A = \frac{1}{\displaystyle\int_0^T e^{-\mathcal{E}(\tau,T)/\alpha}\, d\tau} =: \frac{1}{Z_\alpha}, \qquad Z_\alpha := \int_0^T e^{-\mathcal{E}(\tau,T)/\alpha}\, d\tau.$$

Therefore, the unique minimizer is the density:

$$\boxed{\eta_\alpha(\tau) = \frac{e^{-\mathcal{E}(\tau,T)/\alpha}}{Z_\alpha}.} \tag{4.55}$$

If one wants ζ explicitly, from $A = e^{\zeta/\alpha - 1} = 1/Z_\alpha$ we get

$$\zeta = \alpha\big(1 - \ln Z_\alpha\big).$$

Recovering survival and hazard. Once η_α is fixed, we can recover survival and hazard.

$$S_\alpha(\tau) = 1 - \int_0^\tau \eta_\alpha(u)\, du, \qquad \lambda_\alpha(\tau) = \frac{\eta_\alpha(\tau)}{S_\alpha(\tau)} = \frac{e^{-\mathcal{E}(\tau,T)/\alpha}}{Z_\alpha\, S_\alpha(\tau)}.$$

4.3.4 A Statistical Mechanics Interpretation

The exponential form of the optimal density derived in Eq. (4.55) is similar to the Boltzmann distribution found in statistical mechanics (Gibbs 1902). This analogy is no coincidence; it provides a powerful framework for interpreting the regularization parameter α and the resulting stochastic average $\bar{\mathcal{E}}_\alpha$. By the end, it will allow us to derive an analytical form for $\bar{\mathcal{E}}_\alpha$.

Free energy and partition function. We can identify the single-trajectory energy $\mathcal{E}(\tau, T)$ as the "energy", and the regularization parameter α as a "pseudo-temperature." It is therefore natural to introduce the inverse temperature $\beta := \alpha^{-1} > 0$. The normalization constant derived previously, Z_α, becomes the canonical partition function:

$$Z(T, \alpha) = \int_0^T e^{-\mathcal{E}(\tau,T)/\alpha}\, d\tau \qquad \Longleftrightarrow \qquad Z(T, \beta) = \int_0^T e^{-\beta\, \mathcal{E}(\tau,T)}\, d\tau. \tag{4.56}$$

Consequently, the optimal transition density $\eta_\alpha(\tau)$ describes a canonical ensemble:

$$\eta_\alpha(\tau) = \frac{e^{-\mathcal{E}(\tau,T)/\alpha}}{Z(T, \alpha)} \quad \Longleftrightarrow \quad \eta_\beta(\tau) = \frac{e^{-\beta\, \mathcal{E}(\tau,T)}}{Z(T, \beta)}. \tag{4.57}$$

In this context, the transition time τ is a random variable distributed according to the Boltzmann weight determined by the single-trajectory energy $\mathcal{E}(\tau, T)$.

Interpretation of the entropic constraint. This statistical perspective sheds light on the role of the entropic constraint introduced in Section 4.3.2. The original functional $J_\alpha[\eta]$ defined in Eq. (4.51) can be rewritten as:

$$J_\beta[\eta] = \underbrace{\mathbb{E}[\mathcal{E}]}_{\text{Internal Energy } U} + \underbrace{\frac{1}{\beta}\int \eta \ln \eta}_{\text{- Temperature}\times\text{Entropy } S}.$$

Minimizing J_β is therefore mathematically equivalent to minimizing the *Helmholtz Free Energy* ($F = U - TS$) in thermodynamics.

The parameter β controls the trade-off between energy minimization and entropy maximization (Uncertainty).

- As $\beta \to \infty$ ($\alpha \to 0$), the system freezes into the ground state: the distribution $\eta(t)$ becomes a Dirac delta function centered at the global minimum of $\mathcal{E}(t)$ (we are in a minimal entropy regime, the switching times are known with certainty, the system becomes deterministic).
- As $\beta \to 0$ ($\alpha \to \infty$), the entropy term dominates: the single-trajectory energy is ignored, and $\eta(t)$ approaches a uniform distribution over $[0, T]$ (we are in a maximal entropy regime, maximal uncertainty over the switching times).

Constraining the entropy is therefore equivalent to fixing the temperature of the distribution, forcing the switching times to concentrate around the optimal theoretical times with a specific variance determined by β.

Stochastic average as a Gibbs expectation. The regularized stochastic average $\bar{\mathcal{E}}_\beta(T)$ corresponds to the expected energy of the system at thermal equilibrium. It is given by the Gibbs average:

$$\bar{\mathcal{E}}_\beta(T) \;=\; \int_0^T \eta_\beta(\tau)\, \mathcal{E}(\tau, T)\, d\tau \;=\; \mathbb{E}_\beta\big[\mathcal{E}(\tau, T)\big].$$

Key identity: computing $\bar{\mathcal{E}}_\beta(T)$ from Z. A standard result in statistical physics allows us to compute expectation values directly from the partition function. Differentiating $Z(T,\beta)$ with respect to the inverse temperature β yields:

$$\frac{\partial Z}{\partial \beta} \;=\; \int_0^T \big(-\mathcal{E}(\tau,T)\big)\, e^{-\beta \mathcal{E}(\tau,T)}\, d\tau \;=\; -\,Z(T,\beta)\,\mathbb{E}_\beta[\mathcal{E}(\tau,T)].$$

Dividing by $Z(T,\beta)$, we obtain the logarithmic derivative:

$$\frac{1}{Z}\frac{\partial Z}{\partial \beta} \;=\; \frac{\partial}{\partial \beta}\ln Z(T,\beta) \;=\; -\,\mathbb{E}_\beta[\mathcal{E}].$$

This provides the fundamental identity linking the stochastic average to the partition function:

$$\boxed{\bar{\mathcal{E}}_\beta(T) \;=\; -\frac{\partial}{\partial \beta}\ln Z(T,\beta), \qquad \text{with } \beta = \alpha^{-1}.} \tag{4.58}$$

This relation might simplify the problem or at least offer a new perspective: rather than computing the expectation integral directly, we can integrate the Boltzmann factor to find Z, and then differentiate to find the stochastic average.

4.3.5 Analytical Results from Partition Function Analysis

We now apply the statistical framework to derive an explicit expression for the regularized stochastic average $\bar{\mathcal{E}}_\beta(T)$. The calculation involves evaluating the partition function $Z(T,\beta)$ using Bessel function identities. Due to the technical nature of this derivation, the complete details are provided in Appendix C. Here we present the key results.

Partition function and analytic solution. Through trigonometric linearization and application of the Jacobi-Anger expansion[2] for modified Bessel functions $I_n(\kappa)$, the partition function can be expressed as:

$$Z(T,\beta) \;=\; \frac{1}{2} e^{-\kappa(\cos T+2)}\, Q(T,\kappa), \tag{4.59}$$

where $\kappa := \frac{\beta}{4}$ and the auxiliary integral $Q(T,\kappa)$ is given by:

$$Q(T,\kappa) = 2T\, I_0(\kappa) + 4\sum_{n=1}^{\infty} I_n(\kappa)\,\frac{\sin(nT)}{n}. \tag{4.60}$$

Using the fundamental thermodynamic identity $\bar{\mathcal{E}}_\beta(T) = -\frac{\partial}{\partial\beta}\ln Z$ (Eq. (4.58)) and the chain rule with $\beta = 4\kappa$, we obtain the general analytic solution:

$$\boxed{\bar{\mathcal{E}}_\beta(T) = \frac{1}{2} + \frac{1}{4}\cos T - \frac{1}{4}\,\frac{Q_\kappa(T,\kappa)}{Q(T,\kappa)}, \qquad \kappa = \frac{\beta}{4}.} \tag{4.61}$$

Here, $Q_\kappa := \partial Q/\partial\kappa$ is computed using Bessel function recurrence relations:

$$Q_\kappa(T,\kappa) = 2T\, I_1(\kappa) + 2\sum_{n=1}^{\infty}\big(I_{n-1}(\kappa) + I_{n+1}(\kappa)\big)\,\frac{\sin(nT)}{n}. \tag{4.62}$$

This exact analytical expression, while compact, requires further analysis to extract physical insight. We now examine its behavior in several important limiting regimes.

[2]The Jacobi-Anger expansion decomposes the exponential of a cosine into a series of modified Bessel functions: $e^{z\cos\theta} = I_0(z) + 2\sum_{n=1}^{\infty} I_n(z)\cos(n\theta)$.

4.3.6 Asymptotic Analysis and Special Cases

With the exact analytical form of the regularized stochastic average $\bar{\mathcal{E}}_\beta(T)$ established, we now turn to its physical interpretation. While Eq. (4.61) provides a precise mathematical description; the term with the Bessel functions remains ambiguous. To elucidate the dynamics of the randomized protocol, we verify the solution in critical limiting regimes. First, we examine the long-time behavior ($T \to \infty$) to determine if the system naturally relaxes to the optimum or if active timing control is required. Second, we focus on the specific duration $T = (2k+1)\pi$, which was identified in the deterministic geometric analysis (Section 4.2) as a necessary condition for reaching the ground state. Finally, we scrutinize the high- and low-temperature limits of the regularization parameter α to confirm that our probabilistic framework recovers the expected physical bounds.

Asymptotic behavior for $T \to \infty$. While the series solution is exact, the behavior for large T reveals the limitations of arbitrary durations. We can separate the linear growth in T from the oscillating remainder terms:

$$Q(T,\kappa) = 2T\, I_0(\kappa) + R(T,\kappa), \qquad Q_\kappa(T,\kappa) = 2T\, I_1(\kappa) + R_\kappa(T,\kappa),$$

where the remainders R and R_κ are bounded series (uniformly in T). As $T \to \infty$, the ratio simplifies:

$$\frac{Q_\kappa(T,\kappa)}{Q(T,\kappa)} = \frac{2T\, I_1(\kappa) + O(1)}{2T\, I_0(\kappa) + O(1)} = \frac{I_1(\kappa)}{I_0(\kappa)} + O\left(\frac{1}{T}\right). \tag{4.63}$$

Combining this with Eq. (4.61), we find the asymptotic stochastic average:

$$\boxed{\bar{\mathcal{E}}_\beta(T) = \underbrace{\frac{1}{2} - \frac{1}{4}\frac{I_1(\kappa)}{I_0(\kappa)}}_{\text{Bias}} + \underbrace{\frac{1}{4}\cos T}_{\text{Oscillatory}} + O\left(\frac{1}{T}\right), \qquad T \to \infty.} \tag{4.64}$$

The stochastic average consists of a constant bias term and an oscillatory term $+\frac{1}{4}\cos T$ that does not decay. This result highlights the importance of selecting the correct duration T: for arbitrary T, the oscillatory term prevents the system from reaching the theoretical minimum. We must synchronize the total time with the natural period of the system.

The Resonant Case: $T = (2k+1)\pi$. We consider the specific duration $T = (2k+1)\pi$, identified previously as a condition for optimal deterministic switching. In this case, $\cos T = -1$, and the sine terms $\sin(nT)$ in the series expansions for $Q(T,\kappa)$ and $Q_\kappa(T,\kappa)$ vanish. Consequently, the ratio of integrals simplifies to the ratio of the zeroth and first-order Bessel functions:

$$\frac{Q_\kappa((2k+1)\pi,\kappa)}{Q((2k+1)\pi,\kappa)} = \frac{2(2k+1)\pi I_1(\kappa)}{2(2k+1)\pi I_0(\kappa)} = \frac{I_1(\kappa)}{I_0(\kappa)}.$$

Substituting these values into the general solution (4.61), the stochastic average reduces to a compact closed form:

$$\bar{\mathcal{E}}_\beta((2k+1)\pi) = \frac{1}{2} + \frac{1}{4}(-1) - \frac{1}{4}\frac{I_1(\kappa)}{I_0(\kappa)}.$$

Simplifying the constants yields:

$$\boxed{\bar{\mathcal{E}}_\beta((2k+1)\pi) = \frac{1}{4}\left(1 - \frac{I_1(\kappa)}{I_0(\kappa)}\right).}$$

Limiting cases and physical interpretation. We analyze the behavior of the optimal stochastic average in the high and low regularization limits by applying standard expansions of the modified Bessel functions.

- **High Temperature / Strong Regularization ($\alpha \to \infty \implies \kappa, \beta \to 0$):**
  In this regime, entropy dominates, and the distribution $\eta(t)$ becomes uniform. We utilize the power series definition of the modified Bessel functions for small arguments:

$$I_\nu(\kappa) = \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(\nu+k+1)} \left(\frac{\kappa}{2}\right)^{2k+\nu}.$$

  Expanding the first few terms for $\nu = 0$ and $\nu = 1$:

$$I_0(\kappa) \approx 1 + \frac{\kappa^2}{4} + O(\kappa^4), \qquad I_1(\kappa) \approx \frac{\kappa}{2} + O(\kappa^3).$$

  The ratio can be approximated using the binomial expansion for small x $(1+x)^{-1} \approx 1 - x$:

$$\frac{I_1(\kappa)}{I_0(\kappa)} \approx \frac{\kappa/2}{1+\kappa^2/4} \approx \frac{\kappa}{2}\left(1 - \frac{\kappa^2}{4}\right) \approx \frac{\kappa}{2}.$$

  Substituting this into the stochastic average and recalling $\kappa = \frac{\beta}{4}$:

$$\bar{\mathcal{E}}_\beta \approx \frac{1}{4}\left(1 - \frac{\kappa}{2}\right) = \frac{1}{4} - \frac{\kappa}{8} = \frac{1}{4} - \frac{\beta}{32}.$$

  We see for $\beta \to 0$ that the stochastic average approaches $1/4$. To contextualize this, we note that the *spectral average* of the problem Hamiltonian is $\mathrm{Tr}(H_C)/2 = 1/2$. This corresponds to the expected cost of a state sampled uniformly at random from the Hilbert space, as well as the energy of our unperturbed initial state $|+\rangle$. Therefore, this result implies that a randomized QAOA scheme using a uniform distribution over switching times performs better than a random guess (or doing nothing), provided the total runtime T is fixed to an optimal value.

- **Low Temperature / Weak Regularization ($\alpha \to 0^+ \implies \kappa, \beta \to \infty$):**
  In this regime, the system concentrates on the ground state. We utilize the standard asymptotic expansions for large arguments (often referred to as Hankel expansions):

$$I_\nu(\kappa) \sim \frac{e^\kappa}{\sqrt{2\pi\kappa}}\left(1 - \frac{4\nu^2 - 1}{8\kappa} + \dots\right).$$

  Applying this for $\nu = 0$ and $\nu = 1$:

$$I_0(\kappa) \sim \frac{e^\kappa}{\sqrt{2\pi\kappa}}\left(1 + \frac{1}{8\kappa}\right), \qquad I_1(\kappa) \sim \frac{e^\kappa}{\sqrt{2\pi\kappa}}\left(1 - \frac{3}{8\kappa}\right).$$

  The ratio is approximated by performing a binomial expansion of the denominator:

$$\frac{I_1(\kappa)}{I_0(\kappa)} \approx \frac{1 - \frac{3}{8\kappa}}{1 + \frac{1}{8\kappa}} \approx \left(1 - \frac{3}{8\kappa}\right)\left(1 - \frac{1}{8\kappa}\right) \approx 1 - \frac{4}{8\kappa} = 1 - \frac{1}{2\kappa}.$$

  Substituting this back into the stochastic average expression:

$$\bar{\mathcal{E}}_\beta \approx \frac{1}{4}\left[1 - \left(1 - \frac{1}{2\kappa}\right)\right] = \frac{1}{8\kappa}.$$

Finally, using $\kappa = \frac{\beta}{4}$, we find back the linear dependence on temperature:

$$\bar{\mathcal{E}}_\beta \approx \frac{1}{8(\beta/4)} = \frac{1}{2\beta}.$$

Remarkably, as $\beta \to \infty$, the stochastic average converges to 0. This value implies the system perfectly prepares the ground state $|0\rangle$. The linear term $1/2\beta$ represents the thermal fluctuation penalty preventing perfect convergence.

4.4 Interpretation in the Classical Optimization Loop Perspective

The regularization parameter α offers a compelling physical interpretation within the context of the classical optimization loop used in VQA. As established, high values of α correspond to a high-temperature, high-entropy regime. In this limit, the distribution of switching times $\eta_\alpha(\tau)$ becomes increasingly spread out and uniform over the interval $[0, T]$.

In the context of the QAOA optimization loop, distinct phases can be identified:

1. **Exploration (High Entropy):** At the beginning of the optimization, the classical optimizer (e.g., gradient descent) has no prior information about the location of the optimal parameters. It essentially samples the landscape broadly. This mimics the high-α regime, where the switching times are distributed with high uncertainty (entropy) across the entire duration.

2. **Exploitation (Low Entropy):** As the optimization progresses and the optimizer converges towards a minimum, the distribution of effective parameters narrows. This corresponds to decreasing α (lowering the "temperature"), where the system "freezes" into the optimal deterministic configuration. The distribution $\eta(\tau)$ concentrates around the optimal switching times derived in Section 4.2.

This perspective suggests that the randomized regularized approach can be viewed as an effective description of the uncertainty present during the classical optimization process. The result that $\bar{\mathcal{E}}_\alpha \propto \alpha$ (derived in the previous section) implies that the expected energy error decreases linearly with the reduction of this uncertainty.

Chapter 5

Multi-Switch Strategies

In this chapter, we extend the randomized framework from the single-switch event analyzed in Chapter 4 to a sequence of L potential transitions. This generalization allows the system to alternate between the cost and mixer Hamiltonians multiple times, effectively mimicking a QAOA circuit of depth p where the number of layers is related to the number of switches. By treating the sequence of switching times as a joint random variable, we establish a multi-variable generalization of the entropy regularization technique and define the corresponding multi-switch partition function.

We discuss the analytical intractability arising from the nested recursive structure of this partition function, illustrated by the specific case of Unstructured Search. Nevertheless, we take advantage of this derivation to validate a quadratic speedup at low depth. Finally, we apply asymptotic approximation techniques to extract physical insights from the regularized model. We analyze the low-temperature limit to establish hardware precision requirements and leverage a high-temperature expansion to link the landscape variance to the onset of Barren Plateaus, setting the stage for the spectral analysis in Chapter 6.

5.1 Formulation of the Multi-Switch Problem

We consider a directed Markov chain consisting of $Q = L + 1$ nodes. In this setup, the system undergoes a sequence of L switches, resulting in $L + 1$ distinct nodes. Following the structure of the QAOA ansatz described in Section 2.4, the Hamiltonian alternates as:

$$H^{(k)} = \begin{cases} H_C & \text{if } k \text{ is odd,} \\ H_0 & \text{if } k \text{ is even.} \end{cases} \tag{5.1}$$

The total state density matrix $\rho(T)$ decomposes into sub-normalized components corresponding to the probability of being in each node:

$$\rho(T) = \sum_{k=1}^{L+1} \rho_k(T) \tag{5.2}$$

where $\rho_k(T)/p_k$ represents the average state of the quantum system over all realizations of the Markov chain, given that the Markov chain is in node k at time T. Equivalently, $\mathrm{Tr}(\rho_k(T)) = p_k$ gives the probability mass of the sub-ensemble that has switched exactly $k-1$ times by the final time T. Here for simplicity we will consider that $\mathrm{Tr}(\rho_{2p}(T)) = 1$ i.e. that the system has switched $L = 2p - 1$ times with probability 1 by time T, completing all p QAOA layers.

5.2 Recursive Path Integral Representation

Using the cascading logic of the integrals derived in Section 3.4.1, we can express the component $\rho_k(T)$ as an integral over the ordered switching times. Let $\vec{\tau} = (\tau_1, \tau_2, \ldots, \tau_{k-1})$ be the vector of switch times. The state at time T is given by:

$$\rho_k(T) = \int_0^T d\tau_{k-1} \cdots \int_0^{\tau_2} d\tau_1 \, \mathcal{W}_k(\vec{\tau}, T) \, \mathcal{U}^{(k)}(\vec{\tau}, T) \rho(0) \mathcal{U}^{(k)\dagger}(\vec{\tau}, T) \tag{5.3}$$

where the *unitary evolution* operator follows the specific path:

$$\mathcal{U}^{(k)}(\vec{\tau}, T) = e^{-iH^{(k)}(T-\tau_{k-1})} \ldots e^{-iH^{(2)}(\tau_2-\tau_1)} e^{-iH^{(1)}(\tau_1)} \tag{5.4}$$

The *path weight* $\mathcal{W}_k(\vec{\tau}, T)$ is the probability density of this specific trajectory. Following the causal factorization of Markovian holding times:

$$\mathcal{W}_k(\vec{\tau}, T) = \left[\prod_{j=1}^{k-1} \frac{\eta_j(\tau_j)}{S_j(\tau_{j-1})} \right] \frac{S_k(T)}{S_k(\tau_{k-1})} \quad (\text{with } \tau_0 = 0) \tag{5.5}$$

where $\eta_j(t) = \lambda_j(t) S_j(t)$ is the transition density and $S_j(t) = e^{-\Lambda_j(t)}$ is the survival function for node j, as defined in Section 3.4.4.

Interpretation of the path weight. The structure of $\mathcal{W}_k$ directly generalizes the recursive solution (3.20) of the classical Markov chain. Each factor in the product corresponds to a single transition event:

- The ratio $\eta_j(\tau_j)/S_j(\tau_{j-1})$ represents the *conditional* transition density at time τ_j, given that the system arrived at node j at time τ_{j-1}. Using the definition $\eta_j = \lambda_j S_j$, this simplifies to $\lambda_j(\tau_j) \cdot S_j(\tau_j)/S_j(\tau_{j-1})$, which is the instantaneous rate $\lambda_j(\tau_j)$ multiplied by the conditional survival probability from τ_{j-1} to τ_j.
- The final factor $S_k(T)/S_k(\tau_{k-1})$ is the probability of remaining in node k from the last switch time τ_{k-1} until the final time T, without transitioning further.

This causal factorization reflects the memoryless property of the Markov chain: the probability of each transition depends only on the time elapsed since arriving at the current node, not on the earlier history of the trajectory.

5.3 The Directed Regime and Stochastic Average

In the "directed regime," we assume the system completes all L potential switches to reach the final state $L+1$ by time T. This implies that no probability mass remains in the intermediate states $1, \ldots, L$. If the final state is stable (i.e., $\lambda_{L+1} = 0$, thus $S_{L+1}(T)/S_{L+1}(\tau_L) = 1$), the total density is dominated by $\rho_{L+1}(T)$.

The weight function then simplifies to the joint probability density $p(\vec{\tau})$ over the ordered L-dimensional simplex $\mathcal{D}_L = \{(\tau_1, \ldots, \tau_L) \in \mathbb{R}^L \mid 0 \le \tau_1 \le \cdots \le \tau_L \le T\}$:

$$p(\tau_1, \ldots, \tau_L) = \prod_{j=1}^{L} \lambda_j(\tau_j) \frac{S_j(\tau_j)}{S_j(\tau_{j-1})} \tag{5.6}$$

Taking the trace with the problem Hamiltonian H_C, the stochastic average $\bar{\mathcal{E}}(T)$ becomes:

$$\bar{\mathcal{E}}(T) = \int_{\mathcal{D}_L} p(\vec{\tau}) \, \mathcal{E}(\vec{\tau}) \, d\vec{\tau} \tag{5.7}$$

where $\mathcal{E}(\vec{\tau})$ is the *single-trajectory energy*, representing the energy of a single deterministic path:

$$\mathcal{E}(\vec{\tau}) = \mathrm{Tr}\left(H_C\, \mathcal{U}^{(L+1)}(\vec{\tau}, T)\rho(0)\mathcal{U}^{(L+1)^\dagger}(\vec{\tau}, T)\right) \tag{5.8}$$

5.4 Multi-Switch Entropy Regularization

As seen in Chapter 4, direct optimization of the switching distribution $p(\vec{\tau})$ is ill-posed. To find a smooth, continuous distribution that represents a randomized QAOA protocol, we introduce the multivariate entropy regularizer. We define the regularized functional J_β:

$$J_\beta[p] = \int_{\mathcal{D}_L} p(\vec{\tau})\mathcal{E}(\vec{\tau})d\vec{\tau} + \frac{1}{\beta}\int_{\mathcal{D}_L} p(\vec{\tau})\ln p(\vec{\tau})d\vec{\tau} \tag{5.9}$$

Following the same variational logic as the 2-level system, the minimizer of J_β subject to the normalization constraint $\int p d\vec{\tau} = 1$ is the *Multivariate Boltzmann Distribution*:

$$p_\beta(\vec{\tau}) = \frac{e^{-\beta\mathcal{E}(\vec{\tau})}}{Z(T,\beta)} \tag{5.10}$$

5.4.1 The Multi-Switch Partition Function

The normalization constant $Z(T,\beta)$ is the *Multi-Switch Partition Function*, defined as the integral of the Boltzmann weight over the ordered simplex:

$$Z(T,\beta) = \int_0^T d\tau_L \cdots \int_0^{\tau_2} d\tau_1\ \exp\left(-\beta\mathcal{E}(\tau_1,\ldots,\tau_L)\right) \tag{5.11}$$

By identifying $\beta = \alpha^{-1}$, we recover the fundamental identity linking the stochastic average of the randomized multi-switch protocol to the logarithmic derivative of the partition function:

$$\bar{\mathcal{E}}_\beta(T) = -\frac{\partial}{\partial\beta}\ln Z(T,\beta) \tag{5.12}$$

This formulation transfers the complexity of analyzing deep QAOA circuits into the evaluation of a high-dimensional partition function. This provides a bridge between variational quantum algorithms and statistical mechanics, allowing for the potential use of asymptotic integration techniques to find performance bounds.

5.5 Case Study: Unstructured Search

While the identity $\bar{\mathcal{E}}_\beta(T) = -\frac{\partial}{\partial\beta}\ln Z(T,\beta)$ provides an elegant theoretical framework, its practical application faces a significant challenge: the "curse of dimensionality." To evaluate the partition function $Z(T,\beta)$, one must perform an L-dimensional integration of the Boltzmann factor over the ordered simplex $\mathcal{D}_L$.

In the single-switch case ($L = 1$), we utilized Bessel function identities to solve this integral. However, for $L > 1$, the single-trajectory energy $\mathcal{E}(\vec{\tau})$ becomes a highly non-linear nested function. To illustrate this analytical impasse, we examine the unstructured search problem.

5.5.1 Hamiltonian Setup and Coordinate Mapping

We focus on the problem of an unstructured search problem, which serves as a standard benchmark for quantum optimization. The goal is to find a single marked item z^* in a search space of size $N = 2^n$. Without loss of generality, we set the marked item to be the all-zero string $z^* = 0^n$. The classical cost function is defined as:

$$C(z) = \begin{cases} -1 & \text{if } z = 0^n, \\ 0 & \text{otherwise.} \end{cases} \tag{5.13}$$

This problem is encoded into the *cost Hamiltonian* H_C, which acts as a projector assigning energy -1 to the target state and 0 elsewhere:

$$H_C = -|0^n\rangle\langle 0^n|. \tag{5.14}$$

The *mixer Hamiltonian* is the standard[1] transverse field acting on all qubits, $H_0 = \sum_{j=1}^{n} \sigma_x^{(j)}$.

To compare our continuous-time results with standard gate-based QAOA papers, we must move from absolute time coordinates (τ) to duration coordinates (Δ). In the standard QAOA ansatz of depth p, the unitary evolution consists of p alternating layers of cost and mixer evolutions:

$$\mathcal{U}^{(p)} = \prod_{k=1}^{p} U_0(\Delta_{0,k})\, U_C(\Delta_{C,k}), \tag{5.15}$$

where $\Delta_{C,k}$ and $\Delta_{0,k}$ are the durations (or angles (γ_k, β_k)) of the k-th layer. In our multi-switch framework, these durations correspond directly to the time intervals between consecutive switches. Let $\vec{\tau} = (\tau_1, \ldots, \tau_L)$ be the ordered switching times, with $L = 2p - 1$, $\tau_0 = 0$ and $\tau_{2p} = T$. The mapping is given by:

$$\Delta_{C,k} = \tau_{2k-1} - \tau_{2k-2} \quad \text{(odd intervals)}, \tag{5.16}$$

$$\Delta_{0,k} = \tau_{2k} - \tau_{2k-1} \quad \text{(even intervals)}. \tag{5.17}$$

This linear transformation has a unit Jacobian determinant, ensuring that optimization over the simplex of switching times is equivalent to optimization over the QAOA angles.

5.5.2 The Recursive Landscape

The single-trajectory energy $\mathcal{E}(\vec{\tau})$ is defined as the expectation value of the problem Hamiltonian at the final time T; this function defines an energy landscape. For the projection Hamiltonian $H_C = -|0^n\rangle\langle 0^n|$, this simplifies to the negative probability of finding the system in the target state:

$$\mathcal{E}(\vec{\tau}) = \langle\psi(\vec{\tau})|H_C|\psi(\vec{\tau})\rangle = -|\langle 0^n|\mathcal{U}^{(p)}(\vec{\tau})|\psi_0\rangle|^2 = -|a_p(\vec{\tau})|^2. \tag{5.18}$$

Here, $a_p = \langle 0^n|\mathcal{U}^{(p)}(\vec{\tau})|\psi_0\rangle$ is the *transition amplitude* from the initial uniform superposition $|\psi_0\rangle$ to the target state $|0^n\rangle$. Note that $|a_p|^2$ corresponds precisely to the *success probability* of observing the solution. The complexity of the landscape is thus entirely contained in this amplitude.

For the unstructured search, a_p satisfies a nested recurrence relation (see Appendix B.1 for the proof by recurrence):

$$a_p = \frac{1}{\sqrt{2^n}} e^{-in\Theta_{p,0}} + \sum_{r=1}^{p} (e^{i\Delta_{C,r}} - 1)\cos^n(\Theta_{p,r}) a_{r-1}, \tag{5.19}$$

[1]The mixer always associated to QAOA in the literature (VQA using transverse field mixer = QAOA roughly speaking)

where $\Theta_{p,r} = \sum_{k=r}^{p} \Delta_{0,k}$ represents the total 'memory' or accumulated rotation angle from the mixer Hamiltonian between layer r and the end of the circuit.

This recurrence highlights the analytical bottleneck: the energy of the system depends on the entire history of phase accumulations, preventing the factorization of the partition function $Z(T, \alpha)$. Unlike the 2-level system, where the landscape was a simple trigonometric product, here the integrand involves the exponential of this nested recurrence, rendering the integral analytically intractable for general p.

5.5.3 Lower Bound on Performance

While the randomized landscape is generally intractable, the recurrence relation allows us to analyze specific deterministic trajectories. By analyzing the constructive interference between terms in Eq. (5.19), we identify a candidate parameter schedule:

- *Oracle Angles:* $\Delta_{C,r} = \pi$ (phase flip).
- *Mixer Angles:* $\Delta_{0,k} = \pi/n$ (small rotation).

As proven in Appendix B, this schedule prevents the exponential decay of the history kernel. In the shallow depth limit ($p \ll n$), the success probability $P_p = |a_p|^2$ scales as:

$$P_p \approx \frac{(2p+1)^2}{N}. \tag{5.20}$$

This confirms that the ansatz supports a quadratic speedup ($\mathcal{O}(p^2/N)$) at low depth, establishing a rigorous lower bound on the maximum performance achievable by the system. This result hasn't been found in the QAOA literature to the author's knowledge. This finding is particularly relevant for NISQ devices, where coherence times and gate error rates severely limit circuit depth, making the performance of shallow circuits ($p \ll n$) critical.

5.6 Approximations

Now that we have established in Section 5.5.2 that the multi-switch partition function is generally non-separable, and provided a deterministic lower bound in Section 5.5, we turn to approximation techniques to extract meaningful bounds for the randomized case. We will start by examining the low-temperature (small α) regime, where the integral is dominated by the global minimum of the energy landscape. In this limit, it will be interesting to see the factor α as an actual temperature or at least as a parameter linked to the precision of the optimization. Then, we will explore the high-temperature (large α) regime, where a perturbative expansion around the uniform distribution may yield insights on the a priori performances of the classical optimization loop.

5.6.1 The Low-temperature Limit

In the limit of low temperature or weak regularization ($\alpha \to 0$ or $\beta \to \infty$), the Boltzmann distribution $p_\beta(\vec{\tau})$ becomes highly peaked. The probability mass concentrates almost entirely in the immediate neighborhood of the global minimum of the energy landscape $\mathcal{E}(\vec{\tau}, T)$. In this regime, the global geometry of the landscape becomes less relevant, and the thermodynamics are dominated by the local curvature around the optimum. We evaluate the partition function using the multidimensional generalization of *Laplace's Method* (Bender and Orszag 1978).

Notation. For clarity, we denote the full vector of parameters as $\mathbf{x} = (\vec{\tau}, T)$, combining the L internal switching times with the total evolution time. The dimension of the optimization domain is thus $k = L + 1$.

Recall of the Partition Function and Stochastic Average. The multi-switch partition function and its connection to the stochastic average were established above:

$$Z(\beta) = \int_{\mathcal{D}_L} e^{-\beta \mathcal{E}(\mathbf{x})} d\mathbf{x}, \quad \text{and} \quad \bar{\mathcal{E}}_\beta = -\frac{\partial}{\partial \beta} \ln Z(\beta). \tag{5.21}$$

Local Expansion of the Energy Landscape. Let $\mathbf{x}^*$ denote the global minimum of the energy landscape inside the simplex $\mathcal{D}_L$, and let $\mathcal{E}_{\min} = \mathcal{E}(\mathbf{x}^*)$ be the minimum energy. We assume that $\mathcal{E}(\mathbf{x})$ is smooth and that the minimum is non-degenerate. We expand $\mathcal{E}(\mathbf{x})$ in a Taylor series around $\mathbf{x}^*$:

$$\mathcal{E}(\mathbf{x}) \approx \mathcal{E}(\mathbf{x}^*) + \nabla \mathcal{E}(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^*)^T \mathcal{H} (\mathbf{x} - \mathbf{x}^*). \tag{5.22}$$

Since $\nabla \mathcal{E}(\mathbf{x}^*) = \mathbf{0}$ at the minimum, this simplifies to:

$$\mathcal{E}(\mathbf{x}) \approx \mathcal{E}_{\min} + \frac{1}{2} \delta \mathbf{x}^T \mathcal{H}\, \delta \mathbf{x}, \tag{5.23}$$

where $\delta \mathbf{x} = \mathbf{x} - \mathbf{x}^*$ represents the fluctuation vector. The matrix $\mathcal{H}$ is the $k \times k$ *Hessian matrix* of the energy landscape, defined element-wise as:

$$\mathcal{H}_{ij} = \left. \frac{\partial^2 \mathcal{E}}{\partial x_i \partial x_j} \right|_{\mathbf{x} = \mathbf{x}^*}. \tag{5.24}$$

Physically, the eigenvalues of $\mathcal{H}$ correspond to the "stiffness" or "curvature" of the energy landscape in different directions. Large eigenvalues correspond to steep valley walls (where deviations are costly).

Evaluation of the Partition Function. We substitute the quadratic approximation into the integral for $Z(\beta)$. As $\beta \to \infty$, the integrand $e^{-\beta \mathcal{E}(\mathbf{x})}$ decays exponentially away from $\mathbf{x}^*$. This localization allows us to extend the integration limits from the simplex $\mathcal{D}_L$ to the entire space $\mathbb{R}^k$ with exponentially small error:

$$Z(\beta) \approx e^{-\beta \mathcal{E}_{\min}} \int_{\mathbb{R}^k} e^{-\frac{\beta}{2} \delta \mathbf{x}^T \mathcal{H}\, \delta \mathbf{x}} d(\delta \mathbf{x}). \tag{5.25}$$

The remaining integral is a standard multidimensional Gaussian. Since the Hessian $\mathcal{H}$ is a real symmetric matrix (by the equality of mixed partial derivatives), the Spectral Theorem guarantees the existence of an orthogonal matrix Q such that $\mathcal{H} = Q \Lambda Q^T$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$ contains the eigenvalues and the columns of Q are the corresponding orthonormal eigenvectors.

We introduce the rotated coordinates $\mathbf{y} = Q^T \delta \mathbf{x}$, which align with the principal axes of the quadratic form. Since Q is orthogonal, the Jacobian of this transformation is $|\det Q| = 1$, and the quadratic form diagonalizes:

$$\delta \mathbf{x}^T \mathcal{H}\, \delta \mathbf{x} = \delta \mathbf{x}^T Q \Lambda Q^T \delta \mathbf{x} = \mathbf{y}^T \Lambda \mathbf{y} = \sum_{i=1}^{k} \lambda_i y_i^2. \tag{5.26}$$

In these coordinates, the k-dimensional integral factorizes into a product of k independent one-dimensional Gaussian integrals:

$$\int_{\mathbb{R}^k} e^{-\frac{\beta}{2} \delta \mathbf{x}^T \mathcal{H}\, \delta \mathbf{x}} d(\delta \mathbf{x}) = \int_{\mathbb{R}^k} e^{-\frac{\beta}{2} \sum_i \lambda_i y_i^2} d\mathbf{y} = \prod_{i=1}^{k} \int_{-\infty}^{\infty} e^{-\frac{\beta \lambda_i}{2} y_i^2} dy_i. \tag{5.27}$$

Each factor is a standard Gaussian integral $\int_{-\infty}^{\infty} e^{-ay_i^2} dy_i = \sqrt{\pi/a}$, yielding:

$$\prod_{i=1}^{k} \sqrt{\frac{2\pi}{\beta\lambda_i}} = \sqrt{\frac{(2\pi)^k}{\beta^k \prod_i \lambda_i}} = \sqrt{\frac{(2\pi)^k}{\beta^k \det(\mathcal{H})}}, \tag{5.28}$$

where we used the identity $\det(\mathcal{H}) = \prod_i \lambda_i$. Thus, the partition function becomes:

$$Z(\beta) \approx e^{-\beta \mathcal{E}_{\min}} (2\pi)^{k/2} (\det \mathcal{H})^{-1/2} \beta^{-k/2}. \tag{5.29}$$

Derivation of the Stochastic Average. Taking the logarithm:

$$\ln Z(\beta) \approx -\beta \mathcal{E}_{\min} + \frac{k}{2} \ln(2\pi) - \frac{1}{2} \ln(\det \mathcal{H}) - \frac{k}{2} \ln \beta. \tag{5.30}$$

The geometric term $\det \mathcal{H}$ is independent of β and vanishes upon differentiation. Applying the identity $\bar{\mathcal{E}}_\beta = -\partial_\beta \ln Z$:

$$\bar{\mathcal{E}}_\beta \approx \mathcal{E}_{\min} + \frac{k}{2\beta} = \mathcal{E}_{\min} + \frac{k\alpha}{2}. \tag{5.31}$$

Remark (The Equipartition Theorem). It is notable that the Hessian determinant $\det \mathcal{H}$ cancels out in the final expression for the expected cost. This is a manifestation of the *Equipartition Theorem* in statistical mechanics (Gibbs 1902): for a system in a harmonic potential, the average thermal energy depends only on the temperature (α) and the number of degrees of freedom (k), not on the frequency (curvature) of the trap. However, as discussed below, the Hessian spectrum remains physically relevant as it dictates the *validity region* of this approximation.

Interpretation: Linear Error Accumulation and Landscape Stiffness. The result above admits a natural interpretation when α is viewed as a measure of the *angular precision* available to the optimizer. Since α controls the variance of the Boltzmann distribution around the optimal angles, it physically corresponds to the level of noise or control error in the hardware. In this picture, α quantifies the uncertainty or noise in each switching time (equivalently, each QAOA angle). The stochastic average then decomposes as:

$$\bar{\mathcal{E}}_\alpha = \underbrace{\mathcal{E}_{\min}}_{\text{Optimal energy}} + \underbrace{\frac{k}{2} \cdot \alpha}_{\text{Precision penalty}}. \tag{5.32}$$

Since the number of parameters scales as $k = L + 1 = 2p$ for a p-layer QAOA circuit (with $L = 2p - 1$ switches), the total penalty grows *linearly* with the circuit depth:

$$\text{Precision penalty} = p \cdot \alpha. \tag{5.33}$$

This linear scaling reflects the fact that, within the low temperature limit, each degree of freedom contributes independently to the total variance. The errors from different layers do not cancel but rather accumulate additively.

However, a critical distinction must be made between the *magnitude* of the penalty and the *validity* of the approximation. Recent analyses of the QAOA energy landscape (Stechły et al. 2024; Nemkov, Kiktenko, and Fedorov 2023) indicate that the landscape becomes "rougher" (characterized by higher Fourier frequencies) as the depth p increases. In our formalism, this roughness corresponds to large eigenvalues of the Hessian $\mathcal{H}$, implying a high "stiffness" of the potential well.

This leads to a *Precision Bottleneck*. While a steeper potential does not change the expected thermal energy (due to equipartition), it narrows the basin of attraction. Therefore, to remain in the quadratic regime where this linear penalty holds, the hardware precision must satisfy:

$$\alpha \ll \frac{1}{\lambda_{\max}}, \tag{5.34}$$

where $\lambda_{\max}$ is the largest eigenvalue of the Hessian. As circuit depth increases, $\lambda_{\max}$ grows rapidly, creating a "narrow gorge" geometry. This requires exponentially higher precision ($\alpha \to 0$) just to stabilize the system within the optimal basin, representing a control theoretic limit distinct from the trainability issues discussed next.

5.6.2 High-temperature Expansion

The previous analysis assumes we have already found the basin of attraction. However, this offers no guarantee that the basin is visible to an optimizer starting from a random guess. To understand the *trainability* of the landscape (i.e., the ability to distinguish good parameters from bad ones initially), we must look at the opposite limit: the High-Temperature expansion ($\beta \to 0$ or $\alpha \to \infty$).

This corresponds to strong regularization, where the Boltzmann distribution $p_\alpha(\mathbf{x})$ is nearly uniform over the simplex $\mathcal{D}_L$. Physically, this represents the *a priori* performance of the algorithm before significant optimization has occurred.

The Uniform Average. We define the uniform expectation operator $\mathbb{E}_0[\cdot]$ as the average over the simplex with respect to the flat measure:

$$\mathbb{E}_0[f(\mathbf{x})] \equiv \frac{1}{V_{\mathcal{D}}} \int_{\mathcal{D}_L} f(\mathbf{x})\, d\mathbf{x}, \tag{5.35}$$

where $V_{\mathcal{D}} = \int_{\mathcal{D}_L} d\mathbf{x}$ is the volume of the ordered simplex. Applied to the energy landscape $\mathcal{E}(\mathbf{x})$, $\mathbb{E}_0[\mathcal{E}(\mathbf{x})]$ represents the statistics of the energy landscape when the switching times are chosen purely at random.

Cumulant Generating Function Expansion. To analyze the behavior for small β, we work with the free energy $\ln Z(\beta)$. The logarithm of the moment-generating function is recognized in probability theory as the *Cumulant Generating Function* of the random variable $\mathcal{E}(\mathbf{x})$ evaluated at $-\beta$ (Billingsley 1995). Its Taylor expansion yields:

$$\ln \mathbb{E}_0\left[e^{-\beta\mathcal{E}}\right] = -\beta\kappa_1 + \frac{\beta^2}{2}\kappa_2 - \frac{\beta^3}{6}\kappa_3 + \mathcal{O}(\beta^4), \tag{5.36}$$

where the first two cumulants κ_n correspond to the mean and variance:

$$\kappa_1 = \mathbb{E}_0[\mathcal{E}], \tag{5.37}$$
$$\kappa_2 = \mathbb{E}_0[\mathcal{E}^2] - (\mathbb{E}_0[\mathcal{E}])^2 = \mathrm{Var}_0(\mathcal{E}). \tag{5.38}$$

Substituting this into the free energy $\ln Z \approx \ln V_{\mathcal{D}} + \ln \mathbb{E}_0[e^{-\beta\mathcal{E}}]$, we obtain the expansion of the partition function.

Derivation of the Stochastic Average. Applying the thermodynamic identity $\bar{\mathcal{E}}_\beta = -\partial_\beta \ln Z$:

$$\bar{\mathcal{E}}_\beta = \mathbb{E}_0[\mathcal{E}] - \beta\,\mathrm{Var}_0(\mathcal{E}) + \mathcal{O}(\beta^2). \tag{5.39}$$

Interpretation: Variance as Thermodynamic Susceptibility. We can interpret the inverse temperature β as a measure of the *optimization effort* expended to refine the parameter distribution away from uniformity. The leading term $\mathbb{E}_0[\mathcal{E}]$ is simply the average baseline performance of a random circuit.

The first correction, $-\beta \operatorname{Var}_0(\mathcal{E})$, reveals that the initial return on optimization effort is driven entirely by the *variance* of the energy landscape. The variance acts as the *thermodynamic susceptibility* of the system: it quantifies how easily the system can be "polarized" towards lower energy states by the optimization pressure β.

A landscape with high variance allows the optimizer to quickly distinguish high-performing regions from poor ones. Conversely, if $\operatorname{Var}_0(\mathcal{E}) \to 0$ (as the system size N increases), the landscape becomes statistically featureless. In this regime, the initial improvement vanishes, and the optimizer perceives a flat landscape where all parameter choices perform identically. This statistical suppression of variance is the signature of the *Barren Plateau* phenomenon (see 2.4.5), which we analyze spectrally in the next chapter.

5.6.3 Summary: The Two Asymptotic Regimes

Our asymptotic analysis of the randomized QAOA framework reveals that the algorithm is constrained by two distinct physical mechanisms at opposite ends of the optimization process:

1. **The Precision Bottleneck (Low α):** At the end of optimization, the challenge is *stability*. As depth p increases, the potential well becomes a "Narrow Gorge" (high stiffness $\lambda_{\max}$). Maintaining the optimal energy requires hardware precision that scales as $\alpha \ll 1/\lambda_{\max}$.

2. **The Trainability Bottleneck (High α):** At the start of optimization, the challenge is *visibility*. As system size N increases, the variance $\operatorname{Var}_0(\mathcal{E})$ may vanish exponentially. This corresponds to a Barren Plateau, where the signal-to-noise ratio for any gradient-based search drops below the shot-noise limit.

While the Low-Temperature analysis informs us about the requisite control quality, it is the High-Temperature analysis that determines whether the optimization can start at all. This motivates the shift in Chapter 6 from a geometric analysis to a spectral one, specifically designing a diagnostic tool to estimate $\operatorname{Var}_0(\mathcal{E})$ efficiently.

Chapter 6

Spectral Analysis of Trainability and the Barren Plateau

In the previous chapter, we utilized the entropy regularization framework to derive an analytical approximation for the stochastic average of the multi-switch QAOA. An interesting insight emerged from the high-temperature expansion (Section 5.6.2): the initial improvement of the stochastic average is not driven by the global geometry of the landscape, but by its local statistical properties. Specifically, the system's response to the optimization parameter β (the inverse temperature) was shown to be proportional to the variance of the energy landscape, $\mathrm{Var}_0(\mathcal{E})$.

This mathematical result has a physical implication: the variance acts as the *thermodynamic susceptibility* of the system. If this variance vanishes, the system becomes effectively "stiff," resisting any attempt to lower the expected cost via variational optimization. In the context of Variational Quantum Algorithms (VQAs), this vanishing variance is the statistical hallmark of the *Barren Plateau* phenomenon (McClean, Boixo, et al. 2018), where the optimization landscape concentrates exponentially around its mean, rendering the optimization landscape flat and untrainable (see 2.4.5).

In this chapter, we operationalize this insight by shifting our perspective from statistical mechanics to signal processing. We exploit the periodicity of the QAOA ansatz to solve MaxCut problems and analyze its energy landscape in the frequency domain. By leveraging fundamental relations between the variance and the Fourier spectrum, we propose a rigorous, *a priori* diagnostic tool: the *Trainability Score* $\mathcal{T}$.

We begin by establishing the theoretical link between the thermodynamic variance and the spectral power of the expected cost function. We then define physical bounds for trainability based on the inevitable shot noise present in quantum measurements. Finally, we repurpose the recursive operator formalism of (Nemkov, Kiktenko, and Fedorov 2023). While originally designed to estimate the complexity of the Fourier series, we demonstrate that the survival probability of a classical random walk through the operator tree is mathematically equivalent to the signal-to-noise ratio of a quantum experiment. By aligning the classical sampling budget with the experimental shot budget, we transform the Monte Carlo estimator into an efficient binary diagnostic for trainability, allowing us to detect Barren Plateaus in dense graphs with polynomial classical resources.

6.1 The Thermodynamic-Spectral Connection

Notation: From Switching Times to QAOA Angles. In the previous chapters, we analyzed the randomized QAOA framework where the system evolves according to a stochastic sequence of switching times $\vec{\tau} = (\tau_1, \ldots, \tau_{2p-1})$. The single-trajectory energy was denoted $\mathcal{E}(\vec{\tau})$, and the stochastic average $\bar{\mathcal{E}}(T)$ represented the expectation over the Boltzmann distribution of switching times.

In this chapter, we shift to the standard deterministic QAOA perspective, where the circuit is parameterized by fixed angle vectors $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$. These parameterizations are related by the coordinate transformation described in Section 3: the durations $\Delta_k = \tau_k - \tau_{k-1}$ correspond to the QAOA angles. Note that while Chapter 3-5 used switching times τ, the spectral analysis in this chapter adopts the standard QAOA angle notation $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ to align with the signal processing literature. The relationship is $\gamma_k = \tau_{2k} - \tau_{2k-1}$. To avoid notational confusion, we adopt the standard QAOA notation:

$$\mathcal{C}(\boldsymbol{\gamma}, \boldsymbol{\beta}) := \langle \psi_0 | U^\dagger(\boldsymbol{\gamma}, \boldsymbol{\beta}) H_C U(\boldsymbol{\gamma}, \boldsymbol{\beta}) | \psi_0 \rangle, \tag{6.1}$$

where $\mathcal{C}$ denotes the *expected cost* as a function of the QAOA angles. This is equivalent to the single-trajectory energy $\mathcal{E}(\vec{\tau})$ under the appropriate coordinate transformation, but the distinct notation emphasizes that we are now analyzing the variance over the *uniform distribution of angles*, not the Boltzmann distribution of switching times.

To understand the mechanics of trainability, we revisit the high-temperature expansion derived in Eq. (5.39). Translating to the angle parameterization, the uniform average $\mathbb{E}_0[\mathcal{C}]$ (the expected cost when angles are sampled uniformly from $[0, 2\pi]^{2p}$) evolves under optimization according to:

$$\mathbb{E}_0[\mathcal{C}] - \beta \mathrm{Var}_0(\mathcal{C}) + \mathcal{O}(\beta^2). \tag{6.2}$$

This relation implies that the *return on optimization effort* (how much the expected cost improves per unit increase in β) is entirely determined by $\mathrm{Var}_0(\mathcal{C})$. A high variance means the optimizer can quickly distinguish promising parameter regions from poor ones, yielding rapid improvement as we move away from a uniform (random) distribution.

Critically, the scaling of $\mathrm{Var}_0(\mathcal{C})$ with the system size n (or circuit depth p) determines whether this discriminative power persists at scale. If the variance decays only polynomially, the landscape retains sufficient structure to guide optimization. However, if $\mathrm{Var}_0(\mathcal{C})$ decays *exponentially*, the correction term $\beta \,\mathrm{Var}_0(\mathcal{C})$ becomes negligible compared to measurement noise for any practical β, and optimization effort yields no measurable return.

To analyze this variance without executing the circuit on a quantum device, we turn to Fourier analysis. The cost function $\mathcal{C}(\boldsymbol{\gamma}, \boldsymbol{\beta})$ is a function of the QAOA angles, which enter the unitary operators via complex exponentials $e^{-iH\gamma}$ and $e^{-iH\beta}$. Consequently, $\mathcal{C}(\boldsymbol{\gamma}, \boldsymbol{\beta})$ is a periodic signal and can be expanded as a multidimensional Fourier series:

$$\mathcal{C}(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \sum_{\vec{k}} c_{\vec{k}} e^{i\vec{k}\cdot(\boldsymbol{\gamma}, \boldsymbol{\beta})}, \tag{6.3}$$

where $\vec{k}$ represents the frequency vector associated with the spectral gaps of the Hamiltonian.

6.1.1 Variance as Spectral Power: The Trainability Score

The connection between the statistical properties of the landscape and its frequency components is established via *Parseval's identity*. As highlighted in the recent work by (Okumura and Ohzeki 2025), the variance of a periodic function is equivalent to its total spectral power, excluding the DC component (the mean). Physically, the spectral power corresponds to the sum of the squared magnitudes of the Fourier coefficients, which quantifies the 'amount of variation' or 'feature richness' in the landscape. Their work provides a rigorous theoretical framework linking Barren Plateaus to the Fourier structure of VQA landscapes. However, directly computing these Fourier coefficients for large or dense problem instances remains computationally challenging, motivating the adoption of the recursive operator methods developed by (Nemkov, Kiktenko, and Fedorov 2023), which we describe in Section 6.3.

The expectation value $\mathbb{E}_0[\mathcal{C}]$ corresponds to the zero-frequency coefficient $c_{\vec{0}}$. The second moment is given by the sum of the squared magnitudes of all coefficients: $\mathbb{E}_0[\mathcal{C}^2] = \sum_{\vec{k}} |c_{\vec{k}}|^2$. Therefore, the variance can be expressed purely in terms of the non-zero Fourier coefficients. We define this quantity as the *Trainability Score* $\mathcal{T}$:

$$\mathcal{T} := \mathrm{Var}_0(\mathcal{C}) = \sum_{\vec{k}\neq\vec{0}} |c_{\vec{k}}|^2. \tag{6.4}$$

This spectral perspective transforms the abstract statistical problem into a concrete one: characterizing the Barren Plateau is equivalent to characterizing the decay of the Fourier coefficients $|c_{\vec{k}}|$. If the sum of these coefficients vanishes exponentially, the signal has no power, and the landscape is flat. The renaming from "variance" to "Trainability Score" highlights its role as a resource-dependent threshold for optimization feasibility; this score serves as a direct proxy for the "feature richness" of the optimization landscape.

It is important to distinguish this metric from the variance of the *gradient*, which is the standard metric used in the literature to diagnose Barren Plateaus. (Nemkov, Kiktenko, and Fedorov 2023) demonstrated that the variance of the gradient is related to the frequency-weighted spectral power:

$$\mathrm{Var}(\nabla\mathcal{C}) = \sum_{\vec{k}\neq\vec{0}} |\vec{k}|^2 |c_{\vec{k}}|^2. \tag{6.5}$$

While gradient-based optimizers technically depend on $\mathrm{Var}(\nabla\mathcal{C})$, the unweighted cost variance (our Trainability Score $\mathcal{T}$) is a sufficient and robust indicator for the following reason: in a Barren Plateau regime, the exponential suppression of the coefficients $|c_{\vec{k}}|^2$ is the dominant factor. This suppression overwhelms the polynomial growth of the frequency norm $|\vec{k}|^2$, causing both the cost variance and the gradient variance to vanish simultaneously.

Therefore, $\mathcal{T}$ provides a necessary condition for trainability. If $\mathcal{T}$ is exponentially small, no gradient information is accessible, regardless of the optimization strategy. This definition allows us to assess the feasibility of QAOA instances *a priori*, using the classical algorithms we will describe in Section 6.3.

6.2 The Physical Limit of Optimization

Having defined the Trainability Score $\mathcal{T}$ as the spectral power of the cost function, we must establish a quantitative criterion for its interpretation. A vanishing variance is not merely a geometric feature of a "flat" landscape; in a quantum experiment, it represents a fundamental barrier to observability imposed by finite sampling noise.

In any Variational Quantum Algorithm, the cost function $\mathcal{C}(\boldsymbol{\gamma}, \boldsymbol{\beta})$ is estimated by averaging over a finite number of measurement shots, N_{shots}. This process induces a statistical error floor (shot noise) scaling as $\epsilon_{\text{shot}} \approx 1/\sqrt{N_{\text{shots}}}$ (McClean, Romero, et al. 2016). For an optimizer to identify a descent direction, the landscape's variations (its "features") must be resolvable against this thermal background. If the landscape features (signal) are smaller than the statistical fluctuations of the measurement (noise), the optimizer sees only random noise.

We quantify this resolvability by defining the *Quantum Signal-to-Noise Ratio* (SNR_Q). This metric compares the signal power (the variance $\mathcal{T}$) to the noise power of the measurement estimator ($1/N_{\text{shots}}$):

$$\mathrm{SNR}_Q := \frac{\text{Signal Power}}{\text{Noise Power}} = \frac{\mathcal{T}}{\epsilon_{\text{shot}}^2} = N_{\text{shots}} \cdot \mathcal{T}. \tag{6.6}$$

The significance of this ratio is rigorous. Using Chebyshev's inequality (Arrasmith et al. 2021), we can bound the probability of observing a cost value that deviates from the random mean by

at least the noise floor ϵ_{shot}:

$$\Pr\left(|\mathcal{C} - \mathbb{E}[\mathcal{C}]| \geq \epsilon_{\text{shot}}\right) \leq \frac{\text{Var}(\mathcal{C})}{\epsilon_{\text{shot}}^2} = \text{SNR}_Q. \tag{6.7}$$

This inequality dictates the physical limit of optimization. In the Barren Plateau regime where $\mathcal{T}$ is exponentially suppressed, we find $\text{SNR}_Q \ll 1$. Under these conditions, the probability of measuring any value distinguishable from the random baseline approaches zero. The optimizer effectively perceives a featureless landscape buried under statistical noise, rendering training impossible regardless of the optimization strategy.

Thus, the condition for trainability is not simply $\mathcal{T} > 0$, but $\text{SNR}_Q \gtrsim 1$. This dimensionless metric is the central quantity of our analysis: as we demonstrate in Section 6.4, the expected number of surviving paths in our classical Monte Carlo diagnostic is mathematically equivalent to SNR_Q. This equivalence allows us to classically predict the solvability of a quantum instance by simulating its signal-to-noise ratio.

6.3 Classical Computation of Spectral Power

To evaluate the Trainability Score $\mathcal{T}$ defined in Eq. (6.4), one must compute the sum of the squared Fourier coefficients of the cost function. Attempting to estimate this variance by sampling the cost function on a quantum device is logically circular: inside a Barren Plateau, the variance is so small that an exponential number of shots would be required just to measure it.

Instead, we compute these coefficients classically using the recursive "Dressed Hamiltonian" method developed by (Nemkov, Kiktenko, and Fedorov 2023). We focus our application on the QAOA for the MaxCut problem, which serves as the canonical benchmark in the literature, and the primary case study in (Nemkov, Kiktenko, and Fedorov 2023). While the algorithm's worst-case complexity scales exponentially with the number of non-commuting gates, for shallow depths or specific graph structures, it provides an exact evaluation of the spectral power without instantiating the full 2^n-dimensional state vector.

6.3.1 The Dressed Hamiltonian Formalism

The central idea is to evaluate the expected cost in the Heisenberg picture[1]. Rather than evolving the quantum state vector $|\psi\rangle$ forward in time, we evolve the problem Hamiltonian H_C backwards through the circuit ansatz. The expectation value is given by:

$$\mathcal{C}(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \langle\psi_0|U^\dagger(\boldsymbol{\gamma}, \boldsymbol{\beta}) H_C U(\boldsymbol{\gamma}, \boldsymbol{\beta})|\psi_0\rangle = \langle\psi_0|H_{\text{dressed}}(\boldsymbol{\gamma}, \boldsymbol{\beta})|\psi_0\rangle. \tag{6.8}$$

For MaxCut (see 2.3.2), the problem Hamiltonian is a linear sum of local edge terms $H_C = \sum_{(u,v)\in E} Z_u Z_v$. Because the conjugation operation is linear, we can compute the spectral power contribution for each edge term independently and sum the results:

$$H_{\text{dressed}} = U^\dagger \left(\sum_{(u,v)\in E} Z_u Z_v \right) U = \sum_{(u,v)\in E} \underbrace{U^\dagger Z_u Z_v U}_{H_{\text{dressed}}^{(u,v)}}. \tag{6.9}$$

This term-by-term decomposition is crucial for memory management: instead of storing a massive sum of Pauli strings for the entire Hamiltonian, we only need to store the operator tree[2] for one edge at a time. This structure is also perfect for parallelization, as each edge can be processed independently, which makes it even more interesting for large-scale classical simulations.

[1] In the Heisenberg picture, operators evolve in time ($U^\dagger H U$) while states remain fixed, as opposed to the Schrödinger picture where states evolve.

[2] The operator tree is the graphical data structure representing the recursive expansion of the dressed Hamiltonian in the Heisenberg picture.

6.3.2 The Recursive Branching Algorithm

The QAOA ansatz consists of a sequence of Pauli rotations $R_G(\phi) = e^{-i\frac{\phi}{2}G}$, where G is a Pauli string, either Z_uZ_v for the problem layer or X_u (previously denoted as σ_x) for the mixer layer. The algorithm recursively computes the conjugated operator $O' = R_G^\dagger(\phi)OR_G(\phi)$.

To understand why the recursion proceeds *backwards* through the circuit, let us expand the dressed Hamiltonian explicitly. The QAOA unitary for depth p is:

$$U(\boldsymbol{\gamma}, \boldsymbol{\beta}) = U_B(\beta_p)U_C(\gamma_p)\cdots U_B(\beta_1)U_C(\gamma_1), \tag{6.10}$$

where $U_C(\gamma) = e^{-i\gamma H_C}$ and $U_B(\beta) = e^{-i\beta H_B}$. The fully conjugation reads:

$$H_{\text{dressed}} = \underbrace{U_C^\dagger(\gamma_1)U_B^\dagger(\beta_1)\cdots U_C^\dagger(\gamma_p)U_B^\dagger(\beta_p)}_{\text{reverse order}}\ H_C\ \underbrace{U_B(\beta_p)U_C(\gamma_p)\cdots U_B(\beta_1)U_C(\gamma_1)}_{\text{original order}}. \tag{6.11}$$

To peel off layers one at a time, we start from the *innermost* operators (those adjacent to H_C) which correspond to layer p:

$$H_{\text{dressed}} = U_C^\dagger(\gamma_1)\cdots\underbrace{U_C^\dagger(\gamma_p)U_B^\dagger(\beta_p)\ H_C\ U_B(\beta_p)U_C(\gamma_p)}_{\text{first iteration: conjugate by layer } p}\cdots U_C(\gamma_1). \tag{6.12}$$

Each iteration conjugates the current operator by the next layer inward (i.e., layer $p-1$, then $p-2$, etc.), until layer 1 is processed and the full dressed operator is obtained.

Due to the algebraic properties of Pauli operators, for any term O in the Hamiltonian and any gate generator G, only two outcomes are possible:

1. **Commutation ($[O, G] = 0$):** The operator commutes with the gate generator. The conjugation leaves the operator unchanged:

$$R_G^\dagger(\phi)\, O\, R_G(\phi) = O. \tag{6.13}$$

2. **Anti-Commutation ($\{O, G\} = 0$):** The operator anti-commutes[3] with the generator. The conjugation splits into two branches:

$$R_G^\dagger(\phi)\, O\, R_G(\phi) = O\cos(\phi) + i(G \cdot O)\sin(\phi). \tag{6.14}$$

Note that the exponential structure of the rotation gate $R_G(\phi) = e^{-i\frac{\phi}{2}G}$ is entirely absorbed into the trigonometric coefficients $\cos\phi$ and $\sin\phi$. This is the key insight: we never need to store the evolution operator explicitly (only the resulting Pauli strings and the number of trigonometric factors they carry).

This logic defines a binary tree of operators. As mentioned previously, we initialize the algorithm at the end of the circuit (layer p) with a single target edge $O = Z_uZ_v$, since this term sits at the center of the nested conjugation. We then iterate backwards through the circuit layers, progressively "unwrapping" the conjugation from layer p down to layer 1.

The algorithm maintains a collection of *Operator Terms*. Each term represents a specific path through the recursive expansion tree and is defined by a tuple:

$$\text{Term}_k = (P_k, A_k, M_k) \tag{6.15}$$

where:

[3]Recall Pauli algebra: $\{X, Z\} = XZ + ZX = 0$ (anti-commute), while $[Z, Z] = ZZ - ZZ = 0$ (commute).

- $P_k \in \{I, X, Y, Z\}^{\otimes n}$ is the Pauli string acting on the n qubits. Although formally defined on all qubits, most entries are the identity I. For instance, the initial edge term $Z_u Z_v$ corresponds to a string with Z on qubits u and v, and I on all others (very sparse).
- $A_k \in \mathbb{C}$ is the complex amplitude accumulated along the path (including factors of i from anti-commutations).
- $M_k \in \mathbb{N}$ is the *trigonometric degree*, counting how many times a Pauli string 'branched' or 'split' during the backward evolution. Each split adds a factor of $1/2$ to the magnitude, reducing the weight of that path.

Whenever an anti-commutation occurs, the current term (P, A, M) is removed and replaced by two new terms:

- **Cosine Branch:** $(P, A, M + 1)$. The Pauli string is unchanged, but the degree increases.
- **Sine Branch:** $(G \cdot P, A \cdot i, M + 1)$. The Pauli string is updated by the generator G, the amplitude picks up a phase i, and the degree increases.

Each path through the binary tree is kept distinct, indexed by the unique sequence of branching choices (its *trigonometric signature*). Two paths may arrive at the same Pauli string P but with different trigonometric signatures; these are treated as separate contributions to the spectral power.

Structure of the Final Expansion. After processing all $2p$ layers (alternating cost and mixer), each surviving path through the binary tree corresponds to a specific sequence of branching choices (cosine or sine) at each anti-commutation encountered. The dressed Hamiltonian, and hence the cost function, is therefore expressed as a sum over all such paths:

$$\mathcal{C}(\boldsymbol{\gamma}, \boldsymbol{\beta}) = A_0 + \sum_{k \neq 0} A_k \, t_k(\boldsymbol{\gamma}, \boldsymbol{\beta}), \tag{6.16}$$

where A_0 represents the parameter-independent mean (the term associated to the zero-frequency Fourier coefficient c_0), A_k is the accumulated amplitude (including phases i from sine branches) and $t_k(\boldsymbol{\gamma}, \boldsymbol{\beta})$ denotes the trigonometric monomial, a product of M_k factors of cos or sin applied to the circuit parameters $\{\gamma_1, \ldots, \gamma_p, \beta_1, \ldots, \beta_p\}$ encountered along path k. This is precisely a *multivariate trigonometric polynomial*, the natural Fourier representation of parametrised quantum circuits. The tuple (P_k, A_k, M_k) encodes everything needed: the Pauli string P_k determines whether the path survives expectation filtering, the amplitude A_k gives the coefficient, and the degree M_k counts exactly how many trigonometric factors appear in this term.

6.3.3 Expectation Filtering and Scoring

The recursion continues until it reaches the beginning of the circuit (Layer 0). At this point, we calculate the expectation value of the surviving terms with respect to the initial state $|\psi_0\rangle = |+\rangle^{\otimes n}$. This step acts as a rigorous filter. Since $|+\rangle$ is an eigenstate of X and I but orthogonal to eigenstates of Z and Y, the expectation value $\langle +|P|+\rangle$ is non-zero *if and only if* the string P consists exclusively of I and X operators:

$$\langle +|I|+\rangle = \langle +|X|+\rangle = 1, \qquad \langle +|Z|+\rangle = \langle +|Y|+\rangle = 0. \tag{6.17}$$

Since the full expectation factorizes as $\langle +|^{\otimes n} P |+\rangle^{\otimes n} = \prod_i \langle +|P_i|+\rangle$, any branch resulting in a term containing a Z or Y operator yields zero and is discarded.

Finally, we compute the contribution of the surviving valid terms to the Trainability Score.

From Operator Trees to Spectral Power. Let us recall our goal: we seek to compute the Trainability Score $\mathcal{T} = \mathrm{Var}_0(\mathcal{C}) = \sum_{\vec{k}\neq 0} |c_{\vec{k}}|^2$, i.e., the total spectral power of the cost function excluding the DC component. As established in Eq. (6.16), the cost function takes the form $\mathcal{C}(\boldsymbol{\gamma}, \boldsymbol{\beta}) = A_0 + \sum_{k\neq 0} A_k\, t_k(\boldsymbol{\gamma}, \boldsymbol{\beta})$, where each t_k is a product of M_k trigonometric factors (cosines and sines of the circuit parameters), and A_0 is equal to the uniform average $\mathbb{E}_0[\mathcal{C}]$.

The variance $\mathrm{Var}_0(\mathcal{C})$ is computed by integrating the squared deviation over all parameters uniformly. Expanding the square reveals diagonal and cross-terms:

$$\mathrm{Var}_0(\mathcal{C}) = \int \frac{d\boldsymbol{\gamma}\, d\boldsymbol{\beta}}{(2\pi)^{2p}} \left|\mathcal{C} - \mathbb{E}_0[\mathcal{C}]\right|^2 = \underbrace{\sum_k |A_k|^2 \int |t_k|^2 \frac{d\boldsymbol{\gamma}\, d\boldsymbol{\beta}}{(2\pi)^{2p}}}_{\text{diagonal terms}} + \underbrace{\sum_{k\neq \ell} A_k A_\ell^* \int t_k\, t_\ell^* \frac{d\boldsymbol{\gamma}\, d\boldsymbol{\beta}}{(2\pi)^{2p}}}_{\text{cross-terms}}. \tag{6.18}$$

The key simplification comes from the *orthogonality of trigonometric monomials*. As established by (Nemkov, Kiktenko, and Fedorov 2023), distinct trigonometric products satisfy:

$$\int_0^{2\pi} t_k(\boldsymbol{\gamma}, \boldsymbol{\beta})\, t_\ell(\boldsymbol{\gamma}, \boldsymbol{\beta}) \frac{d\boldsymbol{\gamma}\, d\boldsymbol{\beta}}{(2\pi)^{2p}} = 2^{-M_k} \delta_{k\ell}. \tag{6.19}$$

Indeed, if two branches k and ℓ differ in *any* parameter, i.e. say branch k involves $\cos\gamma_i$ while branch ℓ involves $\sin\gamma_i$, the integral over γ_i yields $\int_0^{2\pi} \cos\gamma \sin\gamma\, d\gamma = 0$, then it will kill the entire cross-term. By construction of the algorithm, each path through the tree has a unique trigonometric signature (determined by its sequence of cos/sin choices), so all cross-terms vanish.

For the diagonal terms, we evaluate the integral of squared trigonometric functions. Using $\int_0^{2\pi} \cos^2\phi \frac{d\phi}{2\pi} = \int_0^{2\pi} \sin^2\phi \frac{d\phi}{2\pi} = \frac{1}{2}$, the contribution factorizes:

$$\text{Contribution}_k = |A_k|^2 \cdot \underbrace{\frac{1}{2} \cdot \frac{1}{2} \cdots \frac{1}{2}}_{M_k \text{ times}} = |A_k|^2 \cdot \left(\frac{1}{2}\right)^{M_k}. \tag{6.20}$$

This is precisely the statement of *Parseval's identity* in the trigonometric basis: each branch contributes independently to the total spectral power, weighted by its trigonometric degree. Summing over all surviving paths for all edges yields the exact Trainability Score $\mathcal{T} = \sum_{\vec{k}\neq 0} |c_{\vec{k}}|^2$.

Remark on the DC Component. For the standard MaxCut Hamiltonian ($H_C = \sum Z_u Z_v$), the operator trace is zero. Since unitary evolution preserves the trace, the dressed Hamiltonian cannot contain a term proportional to the Identity operator. This implies that the DC component (the mean over all angles) vanishes: $c_{\vec{0}} = 0$. Consequently, the total spectral power computed by the algorithm corresponds exactly to the variance $\mathrm{Var}_0(\mathcal{C})$.

6.4 Monte Carlo Estimation for Dense Graphs

The recursive algorithm described in the previous section provides an exact calculation of the spectral power. However, its computational cost is directly tied to the branching factor of the operator tree. For sparse graphs (like 3-regular graphs) at shallow depths, the number of anti-commutations remains manageable. In contrast, for dense problem instances such as Erdős-Rényi graphs, the average degree scales linearly with the system size n. Consequently, the number of non-commuting gates M grows rapidly, and the tree size $\mathcal{O}(2^M)$ explodes, rendering exact summation intractable even for small systems.

To overcome this barrier, we employ a stochastic approach. We adopt the Monte Carlo sampling strategy originally introduced by (Nemkov, Kiktenko, and Fedorov 2023). In their work, this method was developed both to characterize the complexity of classical simulation and to estimate

the cost variance (spectral power) by sampling paths through the operator tree. Crucially, they established that the Monte Carlo estimator is unbiased for the total spectral power, providing an efficient probe of the Fourier structure without exhaustive enumeration.

Here, we operationalize their technique for a different purpose: rather than analyzing complexity scaling or coefficient distributions, we use the sampling as a binary diagnostic tool to probe a "trainability boundary".

6.4.1 The Random Walk Protocol and Unbiasedness

Instead of traversing every branch of the operator tree (which splits at every anti-commutation), we perform a random walk down the structure. The procedure for a single sample is as follows:

1. Start with the target edge operator $P = Z_u Z_v$ at layer p.
2. Iterate backwards through the circuit. When the current operator P anti-commutes with a gate G:
   - Do not split the tree.
   - Flip a fair coin (probability 0.5).
   - If Heads: Take the *Cosine Branch* (keep P, update amplitude).
   - If Tails: Take the *Sine Branch* (update $P \to G \cdot P$, update amplitude).
3. Continue until Layer 0.
4. Perform the *Expectation Filtering*. Let X be the random variable representing the output of this walk:
   - If the final string contains any Z or Y, the path "dies" and we set $X = 0$.
   - If the final string contains only I and X, the path "survives" and we set $X = |A|^2$.

To estimate the spectral power, we perform N_{MC} independent random walks to obtain samples $\{X_1, X_2, \ldots, X_{N_{\mathrm{MC}}}\}$ and compute their empirical mean:

$$\hat{\mathcal{T}} = \frac{1}{N_{\mathrm{MC}}} \sum_{i=1}^{N_{\mathrm{MC}}} X_i. \tag{6.21}$$

We can rigorously show that this estimator is unbiased (i.e., the expected value of the estimator over many samples converges exactly to the true spectral power $\mathcal{T}$) by evaluating its expected value. By the linearity of expectation, and since all X_i are independent and identically distributed, we have:

$$\mathbb{E}[\hat{\mathcal{T}}] = \frac{1}{N_{\mathrm{MC}}} \sum_{i=1}^{N_{\mathrm{MC}}} \mathbb{E}[X_i] = \mathbb{E}[X]. \tag{6.22}$$

The expectation of a single walk $\mathbb{E}[X]$ is defined as the sum over all possible paths k in the full operator tree, weighted by their probability of occurrence. A path k with M_k branching points is selected with probability $P(\text{path } k) = (1/2)^{M_k}$. The value of the random variable for this path is $X_k = |A_k|^2$ if the path survives, and $X_k = 0$ otherwise. Thus:

$$\mathbb{E}[X] = \sum_{\text{all paths } k} P(\text{path } k) \cdot X_k = \sum_{k \in \text{surviving}} \left(\frac{1}{2}\right)^{M_k} |A_k|^2. \tag{6.23}$$

Comparison with the exact definition of the spectral power derived via Parseval's identity (Section 6.3.3) reveals that this sum is exactly $\mathcal{T}$. Therefore:

$$\mathbb{E}[\hat{\mathcal{T}}] = \mathcal{T}. \tag{6.24}$$

The estimator is unbiased: the geometric weighting factor $(1/2)^{M_k}$ required for the exact sum is inherently provided by the probability of sampling the path.

Crucially, for QAOA MaxCut circuits, the amplitude A_k accumulates only phase factors of i (one for each sine branch), so $|A_k|^2 = 1$ for every surviving path. This simplifies the random variable X to a Bernoulli variable (returning 1 for survival, 0 for death). Consequently, the Trainability Score $\mathcal{T}$ is strictly equivalent to the *survival probability* of the random walk:

$$\mathcal{T} = \sum_{k \in \text{surviving}} P(\text{path } k) \cdot 1 = P(\text{survival}). \tag{6.25}$$

This confirms that estimating the trainability of a QAOA instance is formally equivalent to estimating the probability that a random operator path successfully traverses the circuit without "dying"; a property we can probe efficiently even when the exact variance is vanishingly small.

6.4.2 Operationalizing the Estimator as a Diagnostic Tool

While (Nemkov, Kiktenko, and Fedorov 2023) developed this recursive sampling method to analyze the *combinatorial complexity* of Fourier series (specifically, the distribution of terms across different levels), we repurpose it here as a binary probe for physical trainability.

The distinction is crucial. A complexity analysis seeks to characterize the full structure of the Fourier expansion, requiring weighted sampling to estimate coefficients at all levels. In contrast, our goal is simply to determine if the total spectral power $\mathcal{T}$ exceeds the physical noise floor. As established above, for QAOA, $\mathcal{T}$ is equivalent to the raw survival probability of the random walk. This reorients the algorithm from a tool for numerical estimation to a direct simulation of physical observability: we subject the problem instance to a classical sampling regime that mirrors the statistical constraints of a quantum device.

In the Barren Plateau regime, the variance is exponentially suppressed (e.g., $\mathcal{T} \sim 10^{-10}$). An exact algorithm would be intractable, and a standard Monte Carlo sampling would effectively fail, returning exactly 0 for any realistic number of samples because the "volume" of valid paths is so sparse. From a numerical analysis perspective, returning 0 when the true value is non-zero is a convergence failure. However, from our *diagnostic perspective*, this "failure" is the result.

We posit that the computational difficulty of finding a surviving path on a classical computer mirrors the physical difficulty of finding a gradient on a quantum computer:

- If the classical simulation with N_{MC} samples yields zero valid paths, it implies the spectral power is diluted below the resolution of the sampling budget.
- Analogously, a quantum experiment with N_{shots} measurement shots would fail to resolve any landscape features above the shot noise.

The SNR Equivalence. We can formalize this correspondence by aligning our classical resources with the experimental constraints. By setting the classical sampling budget equal to the experimental shot budget ($N_{\text{MC}} = N_{\text{shots}}$), we create a direct link to the Quantum Signal-to-Noise Ratio defined in Section 6.2.

Since the Trainability Score $\mathcal{T}$ is the survival probability of a single walk, the expected number of surviving paths in our simulation is:

$$\mathbb{E}[N_{\text{survivors}}] = N_{\text{MC}} \cdot P(\text{survival}) = N_{\text{shots}} \cdot \mathcal{T}. \tag{6.26}$$

Comparing this to Eq. (6.6), we arrive at a fundamental identity:

$$\mathbb{E}[N_{\text{survivors}}] \equiv \text{SNR}_Q. \tag{6.27}$$

This identity establishes the central claim of our operationalization: finding a single surviving path in the classical simulation is mathematically equivalent to resolving a signal above the noise floor in a quantum experiment.

Both the classical diagnostic and the quantum experiment are governed by the same dimensionless ratio. When $\text{SNR}_Q \ll 1$, the quantum landscape is buried in shot noise, and the classical simulation yields zero survivors with high probability. When $\text{SNR}_Q \gtrsim 1$, the landscape becomes resolvable, and the classical simulation reliably returns non-zero counts. Thus, the Monte Carlo method provides a rigorous *a priori* test for trainability: it does not merely estimate the variance, but simulates the observability of the optimization landscape itself.

6.4.3 Diagnostic Reliability and the Classical Advantage

The reliability of our diagnostic is determined by the stochastic nature of the random walk. Because of the equivalence $\mathbb{E}[N_{\text{survivors}}] = \text{SNR}_Q$, the probability of the classical simulation "failing" (finding zero paths) is governed by the same Poissonian statistics that dictate quantum resolvability.

We analyze the two failure modes in terms of the simulated signal-to-noise ratio:

Deep Barren Plateaus ($\text{SNR}_Q \ll 1$). If an instance lies deep in a Barren Plateau, the signal is exponentially suppressed. The expected number of survivors is $\text{SNR}_Q \approx 0$. The probability of incorrectly flagging such an instance as trainable (a false positive) is upper bounded by SNR_Q, which vanishes exponentially. Thus, the diagnostic is extremely robust: if the landscape is truly flat, the simulation will almost certainly return zero.

Threshold Trainability ($\text{SNR}_Q \approx 1$). The critical regime occurs near the trainability boundary, where the signal power roughly equals the noise floor. Here, a "false negative" can occur: the instance is theoretically trainable, but the random walk simply fails to stumble upon a valid path due to bad luck. The probability of this error decays exponentially with the sampling budget. Let c be the ratio of classical samples to quantum shots ($N_{\text{MC}} = c \cdot N_{\text{shots}}$). The probability of finding zero survivors is:

$$\Pr(\text{False Negative}) = (1 - \mathcal{T})^{N_{\text{MC}}} \approx e^{-N_{\text{MC}} \cdot \mathcal{T}} = e^{-c \cdot \text{SNR}_Q}. \tag{6.28}$$

At the precise trainability threshold ($\text{SNR}_Q = 1$) and with equal budgets ($c = 1$), the error rate is $e^{-1} \approx 37\%$. This confirms that a single-shot simulation is uncertain, exactly where the quantum experiment is uncertain.

The Classical Advantage. However, a crucial asymmetry exists between the two regimes. Increasing the quantum shot budget N_{shots} is experimentally expensive. In contrast, classical coin flips are computationally negligible. We can therefore choose an "over-sampling factor" $c > 1$ (e.g., $c = 10$). This effectively allows us to simulate a quantum experiment with a boosted signal-to-noise ratio. By setting $N_{\text{MC}} = 10 \cdot N_{\text{shots}}$, the false negative rate at the threshold drops from 37% to $e^{-10} < 0.005\%$. This ensures that if an instance is physically solvable, our diagnostic will detect it with near certainty, filtering out only those instances that are truly hopeless.

6.4.4 Computational Complexity: Exact vs. Monte Carlo

Having established the Monte Carlo diagnostic as a reliable probe of trainability, we now rigorously compare its computational cost to that of the exact algorithm. This analysis demonstrates that the MC method transforms an exponentially hard summation into a polynomially tractable decision problem, justifying its use for dense graph instances where exact computation is infeasible.

Setup: Dense Graphs. We consider Erdős-Rényi random graphs $G(n, q)$, where each of the $\binom{n}{2}$ possible edges is included independently with probability q. The expected number of edges is:

$$|E| = q\binom{n}{2} = \frac{qn(n-1)}{2} \sim \frac{qn^2}{2}. \tag{6.29}$$

For dense graphs, $q = \mathcal{O}(1)$ (i.e., constant edge probability), so $|E| = \mathcal{O}(n^2)$. The average vertex degree is $\bar{d} = q(n-1) \sim qn$.

Exact Algorithm Complexity. The exact algorithm processes each edge independently, computing the contribution of the dressed operator $U^\dagger Z_u Z_v U$ to the total spectral power. For a single edge (u, v), the algorithm traverses the $2p$ circuit layers (alternating between p cost layers and p mixer layers), branching at each anti-commutation.

To count the number of anti-commutations M encountered by a single edge, we examine each layer type:

- **Mixer layers:** Each mixer layer applies n single-qubit rotations $e^{-i\beta X_j}$. The edge operator $Z_u Z_v$ (or its Pauli descendants) anti-commutes with X_u and X_v, contributing at most 2 anti-commutations per mixer layer. Over p mixer layers, this yields at most $2p$ anti-commutations from the mixer.
- **Cost layers:** Each cost layer applies $|E|$ two-qubit rotations $e^{-i\gamma Z_a Z_b}$ for each edge $(a, b) \in E$. The current Pauli string anti-commutes with $Z_a Z_b$ if and only if the string has a non-Z, non-I operator on exactly one of $\{a, b\}$. As the Pauli string evolves, it spreads to neighbouring vertices. For a vertex of degree d, the string can interact with d adjacent edges. In dense graphs, $d \sim qn$, so each cost layer can contribute $\mathcal{O}(qn)$ anti-commutations in the worst case.

Combining these contributions, the total number of anti-commutations per edge scales as:

$$M(n, p) \sim 2p + p \cdot qn = p(2 + qn) \sim p \cdot qn \quad \text{(for dense graphs with } qn \gg 1\text{)}. \tag{6.30}$$

Since each anti-commutation doubles the number of terms in the operator tree, the number of paths (and hence the memory and time cost) for processing a single edge is:

$$\text{Paths per edge} = 2^M \sim 2^{p \cdot qn}. \tag{6.31}$$

Each path requires $\mathcal{O}(n)$ operations to store and update the Pauli string (array of size n). Summing over all $|E| \sim qn^2/2$ edges, the total complexity of the exact algorithm is:

$$\boxed{\text{Exact Complexity: } \mathcal{O}\left(n^3 \cdot 2^{p \cdot qn}\right).} \tag{6.32}$$

This is *exponential in* n for any fixed $p \geq 1$ and $q = \mathcal{O}(1)$. Even at depth $p = 1$, a dense graph with $n = 50$ qubits and $q = 0.5$ would require enumerating $\sim 2^{25} \approx 10^7$ paths *per edge*, totalling $\sim 10^{10}$ operations (already at the limit of practical computation). For $n = 100$, the cost becomes astronomical ($\sim 2^{50}$ paths per edge).

Monte Carlo Complexity. The Monte Carlo method replaces exhaustive enumeration with stochastic sampling. Each sample consists of a single random walk through the operator tree, making one coin flip per anti-commutation instead of exploring both branches.

For a single random walk on a single edge:

- The walk traverses $2p$ layers, encountering $M \sim p \cdot qn$ anti-commutations.
- At each anti-commutation, we perform $\mathcal{O}(1)$ operations: a coin flip and (possibly) a Pauli string update.
- Updating the Pauli string after a sine branch requires multiplying two Pauli operators on a subset of qubits. Since the string remains sparse (supported on $\mathcal{O}(p \cdot qn)$ qubits in the worst case), each update costs $\mathcal{O}(1)$ to $\mathcal{O}(n)$ depending on implementation.

Taking the conservative bound of $\mathcal{O}(n)$ per update, the cost of a single random walk on a single edge is:

$$\text{Cost per walk per edge} = \mathcal{O}(M \cdot n) = \mathcal{O}(p \cdot qn \cdot n) = \mathcal{O}(p \cdot qn^2). \tag{6.33}$$

To estimate $\mathcal{T}$ with N_{MC} samples, we perform N_{MC} walks for each of the $|E| \sim qn^2/2$ edges. The total complexity is:

$$\boxed{\text{MC Complexity: } \mathcal{O}\left(N_{\text{MC}} \cdot |E| \cdot p \cdot qn^2\right) = \mathcal{O}\left(N_{\text{MC}} \cdot p \cdot q^2 n^4\right).} \tag{6.34}$$

For constant q and p, this simplifies to $\mathcal{O}(N_{\text{MC}} \cdot n^4)$, *polynomial in* n.

The Diagnostic Trade-off. The dramatic difference in complexity reflects a fundamental distinction in what each method computes:

Method	**Complexity**	**Output**
Exact	$\mathcal{O}(n^3 \cdot 2^{pqn})$	Precise value of $\mathcal{T}$
Monte Carlo	$\mathcal{O}(N_{\text{MC}} \cdot p \cdot q^2 n^4)$	Detection Prob. $\approx 1 - e^{-N_{\text{MC}}\mathcal{T}}$

We have effectively transformed the problem from "calculating the exact value of a vanishing sum" (exponentially hard when the contributions are spread across 2^M terms) to "detecting whether the signal exists" (easy when the signal is sufficiently strong). The exact algorithm computes the full spectral power to arbitrary precision, but this precision is unnecessary. Indeed, it is physically meaningless when $\mathcal{T}$ is exponentially small.

The Monte Carlo method instead answers a *detection problem*. Since the estimator $\hat{\mathcal{T}}$ is derived from discrete paths, it acts as a binary flag: finding $\hat{\mathcal{T}} > 0$ implies we have found at least one surviving path. This is precisely the question relevant to quantum optimization. A quantum device does not compute $\mathcal{T}$; it either succeeds in finding gradients (when $\mathcal{T} \gtrsim 1/N_{\text{shots}}$) or fails (when $\mathcal{T} \ll 1/N_{\text{shots}}$). The MC diagnostic simulates this success/failure dichotomy at polynomial classical cost.

Scaling with Trainability. If the Trainability Score decays polynomially with system size, $\mathcal{T}(n, p) \sim n^{-\xi}$ for some exponent $\xi > 0$, then maintaining a fixed false negative rate requires scaling the MC budget as $N_{\text{MC}} \sim c/\mathcal{T} \sim c \cdot n^{\xi}$. Substituting into the MC complexity:

$$\text{Total MC cost} = \mathcal{O}(c \cdot n^{\xi} \cdot p \cdot q^2 n^4) = \mathcal{O}(c \cdot p \cdot q^2 \cdot n^{4+\xi}). \tag{6.35}$$

Even for quadratic decay ($\xi = 2$), this yields $\mathcal{O}(n^6)$ (still polynomial). The MC method remains efficient as long as $\mathcal{T}$ does not decay faster than polynomially, which is precisely the regime of trainable instances.

Conversely, in the Barren Plateau regime where $\mathcal{T} \sim 2^{-n}$, the MC method with any polynomial budget $N_{\mathrm{MC}} = \mathrm{poly}(n)$ will return zero with overwhelming probability. This is not a failure of the algorithm; it is the correct diagnosis: the instance is untrainable, and the classical difficulty of finding a surviving path mirrors the quantum difficulty of finding a gradient.

Physical Interpretation. The exponential hardness of exact computation reflects the exponential number of interfering quantum amplitudes in the dressed Hamiltonian. Each path through the operator tree corresponds to a distinct interference pattern, and the exact algorithm must track all of them.

The Monte Carlo method succeeds because trainability is a *threshold property*: we need not resolve the fine-grained interference structure, only determine whether the total constructive interference (the spectral power) exceeds the noise floor. By sampling paths with probability proportional to their contribution to $\mathcal{T}$, the MC estimator automatically concentrates on the physically relevant regime.

This establishes the Monte Carlo diagnostic as an efficient classical tool: for any polynomially trainable instance, the diagnostic runs in polynomial time; for exponentially untrainable instances, it correctly returns zero. The method thus provides a complete characterization of the trainability boundary at a cost scaling polynomially with system size.

Chapter 7

Discussion and Conclusion

This thesis set out to explore the landscape of the Quantum Approximate Optimization Algorithm (QAOA) through a new perspective: that of randomized, continuous-time control. Motivated by the analytical bottlenecks inherent in exact recursive evaluations of QAOA at moderate depths, we replaced the standard fixed-angle paradigm with a probabilistic framework governed by Markovian dynamics. This approach allowed us to reframe the variational optimization problem as a statistical mechanics problem, ultimately leading to a new classical diagnostic for quantum trainability.

7.1 Summary of Contributions

Our investigation followed a progression from fundamental control dynamics to practical algorithmic diagnostics.

The Randomized Framework and Entropy Regularization. In Chapters 3 and 4, we established the formalism for controlling a quantum system via a continuous-time Markov chain. By analyzing the single-switch case on a 2-level system, we demonstrated that direct functional optimization of the transition rate is ill-posed. This necessitated the introduction of *Entropy Regularization*, physically interpreted as introducing a non-zero temperature to the optimization landscape. This regularization proved to be more than a mathematical convenience; it enforced a physically meaningful distribution of switching times, preventing the collapse into trivial deterministic solutions. We derived explicit analytical solutions for the optimal transition density in terms of modified Bessel functions, revealing how the system concentrates around optimal "resonant" trajectories as the regularization parameter $\alpha \to 0$.

Generalization and Analytical Limits. In Chapter 5, we extended this framework to the multi-switch setting, mimicking a QAOA circuit of arbitrary depth p. We derived a general *Multi-Switch Partition Function* $Z(T, \alpha)$, linking the expected cost of the algorithm to the logarithmic derivative of Z. However, the analysis of the Unstructured Search problem as an example highlighted the "curse of dimensionality" inherent in this approach. As the depth increases, the single-trajectory energy $\mathcal{E}(\vec{\tau}, T)$ becomes highly non-separable, rendering exact analytical integration intractable. This confirmed that while the randomized framework provides a powerful conceptual bridge to statistical mechanics, it does not essentially bypass the combinatorial complexity of the underlying optimization problem.

The shift to Spectral Analysis. A conceptual shift occurred in Chapter 6, where we utilized the high-temperature expansion of our randomized framework. This expansion revealed that the initial improvement in the cost function is driven entirely by the *variance* of the energy landscape. By using Parseval's identity, we reinterpreted this variance as the total *Spectral Power* (or Trainability Score $\mathcal{T}$) of the Fourier series of the cost function.

7.2 The Operational Link: From Complexity to Diagnosticity

A central contribution of this work lies in the operationalization of the recursive operator formalism initially proposed by (Nemkov, Kiktenko, and Fedorov 2023). While previous works utilized this Monte Carlo method primarily to estimate the computational complexity of the Fourier series (i.e., counting how many terms are non-zero), we identified a deeper physical link.

We demonstrated that for QAOA, the *raw survival probability* of a classical random walk through the operator tree is mathematically equivalent to the *Signal-to-Noise Ratio (SNR)* of a quantum experiment.

This realization reorients the algorithm from a tool for numerical estimation to a direct simulation of physical observability. We move beyond the goal of converging to a precise value for the variance; instead, we determine whether the landscape's features are statistically resolvable against the inevitable shot noise of a quantum device.

- If the classical simulation fails to find a surviving path, it implies $\mathrm{SNR}_Q \ll 1$: the instance lies in a Barren Plateau.
- If the classical simulation succeeds, it implies $\mathrm{SNR}_Q \gtrsim 1$: the landscape features are resolvable.

This provides a polynomial-time classical diagnostic that can predict the solvability of a quantum instance without ever running the quantum circuit. This result is particularly impactful for dense graphs, where exact simulation is impossible, but where our diagnostic can efficiently demarcate the "trainability boundary."

7.3 Future Directions

The framework established in this thesis opens several promising avenues for future research, ranging from analytical theory to empirical classification of NP-hard instances.

Spectral Decomposition of the Partition Function. While the multi-switch partition function $Z(T, \alpha)$ proved difficult to integrate directly in the time domain, its structure suggests that a Fourier-based analysis could yield a new path to solvability. Just as we used the Jacobi-Anger expansion to solve the single-switch integral using Bessel functions, future work could attempt to decompose the multi-switch landscape into its frequency components. Since we established the link between the stochastic average and the partition function ($\bar{\mathcal{E}}_\beta = -\partial_\beta \ln Z$) and that $\lim_{\beta\to\infty} \bar{\mathcal{E}}_\beta = \mathcal{E}_{\min} = \mathcal{C}_{\min}$ recovers the global minimum of the expected cost, such a decomposition could lead to efficient new methods for computing approximation ratios analytically, bypassing the need for expensive brute-force simulations. Furthermore, this spectral perspective would provide a theoretical framework to rigorously study how specific frequency components and Fourier coefficients influence both the minimum achievable energy and the trainability of an instance. By isolating which spectral features drive performance, one could unravel the precise characteristics that render a problem instance trainable or untrainable.

Diagnosing Phase Transitions in Hard Problems. The Monte Carlo diagnostic tool developed in Chapter 6 is ready for deployment on large-scale problem instances. A natural next step is to use this tool to map the "Phase Diagram of Trainability" for combinatorial problems like MaxCut. By varying parameters such as graph density, regularity, or the ratio of constraints to variables, one could identify phase transitions where the problem shifts from trainable to a Barren Plateau regime. Understanding the mechanism behind these transitions (how the volume of surviving operator paths collapses) could provide insights into why certain instances are hard

for quantum algorithms, potentially linking the onset of Barren Plateaus to the connectivity properties of the underlying graph.

In conclusion, the randomized perspective establishes itself as a versatile and general framework, capable of unifying concepts that are often treated in isolation within the literature. By mapping the discrete parameters of QAOA onto a continuous statistical manifold, we can interpret asymptotic performance, trainability, and control precision not as separate challenges, but as distinct regimes within a single thermodynamic model. This synthesis offers a cohesive vantage point: whether analyzing the low-temperature limit to extract performance bounds or exploiting high-temperature expansions to diagnose Barren Plateaus, the framework provides the theoretical connective tissue needed to fully characterize the potential and limitations of variational quantum algorithms.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the author used ChatGPT for its Wolfram plugin in order to perform symbolic computations and verify some algebraic derivations. Additionally, Gemini was used for its long-context capabilities to provide constructive remarks on the overall structure and coherence of the manuscript. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Bibliography

Abramowitz, Milton and Irene A Stegun (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. See Chapter 9 for Modified Bessel Functions. New York: Dover.

Amari, Shun-ichi and Andrzej Cichocki (2010). "Information geometry of divergence functions". In: *Bulletin of The Polish Academy of Sciences-technical Sciences* 58, pp. 183–195.

Arrasmith, Andrew et al. (Oct. 2021). "Effect of barren plateaus on gradient-free optimization". In: *Quantum* 5, p. 558. ISSN: 2521-327X. DOI: 10.22331/q-2021-10-05-558.

Barak, Boaz and Kunal Marwaha (2022). "Classical Algorithms and Quantum Limitations for Maximum Cut on High-Girth Graphs". In: Schloss Dagstuhl – Leibniz-Zentrum für Informatik. DOI: 10.4230/LIPICS.ITCS.2022.14.

Basso, Joao et al. (2022). "The Quantum Approximate Optimization Algorithm at High Depth for MaxCut on Large-Girth Regular Graphs and the Sherrington-Kirkpatrick Model". In: Schloss Dagstuhl – Leibniz-Zentrum für Informatik. DOI: 10.4230/LIPICS.TQC.2022.7.

Bender, Carl M. and Steven A. Orszag (1978). *Advanced Mathematical Methods for Scientists and Engineers*. International Series in Pure and Applied Mathematics. New York: McGraw-Hill Book Company. ISBN: 0-07-004452-X.

Billingsley, Patrick (1995). *Probability and Measure*. 3rd ed. New York: John Wiley & Sons. ISBN: 9780471007104.

Born, M. and V. A. Fock (1928). "Beweis des Adiabatensatzes". In: *Zeitschrift für Physik* 51.3–4, pp. 165–180. DOI: 10.1007/BF01343193.

Boulebnane, Sami et al. (2025). *Evidence that the Quantum Approximate Optimization Algorithm Optimizes the Sherrington-Kirkpatrick Model Efficiently in the Average Case*. arXiv: 2505.07929 [quant-ph].

Deutsch, David (1985). "Quantum theory, the Church-Turing principle and the universal quantum computer". In: *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 400.1818, pp. 97–117. DOI: 10.1098/rspa.1985.0070.

Dubus, Benoît, Joseph Cunningham, and Jérémie Roland (2025). *New random compiler for Hamiltonians via Markov Chains*. arXiv: 2411.06485 [quant-ph].

Erdös, P. and A. Rényi (1959). "On Random Graphs I". In: *Publicationes Mathematicae Debrecen* 6, pp. 290–297.

Farhi, Edward, Jeffrey Goldstone, and Sam Gutmann (2014). *A Quantum Approximate Optimization Algorithm*. arXiv: 1411.4028 [quant-ph].

Farhi, Edward, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser (2000). *Quantum Computation by Adiabatic Evolution*. arXiv: quant-ph/0001106 [quant-ph].

Feynman, R. P. (1982). "Simulating Physics with Computers". In: *International Journal of Theoretical Physics* 21.6–7, pp. 467–488. DOI: 10.1007/BF02650179.

Gibbs, Josiah Willard (1902). *Elementary Principles in Statistical Mechanics, Developed with Especial Reference to the Rational Foundation of Thermodynamics*. New York: Charles Scribner's Sons.

Goemans, Michel X. and David P. Williamson (1995). "Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming". In: *Journal of the ACM* 42.6, pp. 1115–1145.

Hastings, W. K. (1970). "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1, pp. 97–109. DOI: 10.1093/biomet/57.1.97.

Huang, Kerson (1987). *Statistical Mechanics*. 2nd ed. New York: John Wiley & Sons. ISBN: 0-471-81518-7.

Kakade, Sham M. and Shai Shalev-Shwartz (2009). "On the duality of strong convexity and strong smoothness : Learning applications and matrix regularization". In: *Proceedings of the 22nd Annual Conference on Learning Theory (COLT)*, pp. 35–65.

Kielhöfer, Hansjörg (2018). *Calculus of Variations: An Introduction to the One-Dimensional Theory with Examples and Exercises*. Texts in Applied Mathematics. Cham: Springer. ISBN: 978-3-319-71122-5.

Lovric, Miodrag. (2025). *International Encyclopedia of Statistical Science*. 2nd ed. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 9783662693599.

Lucas, Andrew (2014). "Ising formulations of many NP problems". In: *Frontiers in Physics* 2. ISSN: 2296-424X. DOI: 10.3389/fphy.2014.00005.

McClean, Jarrod R., Sergio Boixo, et al. (Nov. 2018). "Barren plateaus in quantum neural network training landscapes". In: *Nature Communications* 9.1. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07090-4.

McClean, Jarrod R., Jonathan Romero, et al. (Feb. 2016). "The theory of variational hybrid quantum-classical algorithms". In: *New Journal of Physics* 18.2, p. 023023. ISSN: 1367-2630. DOI: 10.1088/1367-2630/18/2/023023.

Nemkov, Nikita A., Evgeniy O. Kiktenko, and Aleksey K. Fedorov (Sept. 2023). "Fourier expansion in variational quantum algorithms". In: *Physical Review A* 108.3. ISSN: 2469-9934. DOI: 10.1103/physreva.108.032406.

Nielsen, Michael A. and Isaac L. Chuang (2010). *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge: Cambridge University Press.

Norris, J. R. (1997a). "Applications". In: *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, pp. 170–216.

— (1997b). *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Okumura, Shun and Masayuki Ohzeki (2025). *Fourier Analysis of Parameterized Quantum Circuits and the Barren Plateau Problem*. arXiv: 2309.06740 [quant-ph].

Stechły, Michał et al. (2024). "Connecting the Hamiltonian structure to the QAOA energy and Fourier landscape structure". In: *Quantum* 8, p. 1335. DOI: 10.22331/q-2024-05-09-1335.

Williamson, David P. and David B. Shmoys (2011). *The Design of Approximation Algorithms*. Cambridge: Cambridge University Press. ISBN: 978-0-521-19527-0.

Yin, G. George and Qing Zhang (2012). *Continuous-Time Markov Chains and Applications: A Two-Time-Scale Approach*. 2nd ed. Vol. 37. Stochastic Modelling and Applied Probability. New York, NY: Springer Nature. ISBN: 9781461443469.

Appendix A

Squared Constraint

We study the functional arising from the expectation value derived in Chapter 4. Under the normalization constraint $\int_0^T \eta(\tau)\,d\tau = 1$, the stochastic average takes the form

$$\bar{\mathcal{E}}[\lambda] = \int_0^T \eta(\tau)\,\mathcal{E}(\tau)\,d\tau, \tag{A.1}$$

where $\eta(\tau) = \lambda(\tau)\,e^{-\int_0^\tau \lambda(\tau')\,d\tau'}$ is the transition density and the single-trajectory energy is

$$\mathcal{E}(\tau) := \tfrac{1}{2}\big(1 - \sin(T-\tau)\sin\tau\big). \tag{A.2}$$

Throughout we assume a measurable rate $\lambda\colon [0,T] \to [0,\infty[$.

Kernel and basic bound. Note that $\sin(T-\tau)\sin\tau = \frac{1}{2}\big(\cos(T-2\tau) - \cos T\big)$, so $|\sin(T-\tau)\sin\tau| \le 1$, which implies $0 \le \mathcal{E}(\tau) \le 1$. Since the stochastic average is an expectation value of the projector $H_C = |1\rangle\langle 1|$, we have the immediate bounds

$$0 \le \bar{\mathcal{E}}[\lambda] \le 1. \tag{Bound}$$

Exponential survival notation. Set the cumulative rate and survival factor

$$\Lambda(\tau) := \int_0^\tau \lambda(u)\,du, \qquad S(\tau) := e^{-\Lambda(\tau)}.$$

Then $S'(\tau) = -\lambda(\tau)S(\tau)$ and $\eta(\tau) = -S'(\tau)$, so (A.1) becomes

$$\bar{\mathcal{E}}[\lambda] = \int_0^T -\mathcal{E}(\tau)\,dS(\tau).$$

Integrating by parts (the boundary terms involve $\mathcal{E}(0) = \mathcal{E}(T) = \frac{1}{2}(1-0) = \frac{1}{2}$, but since we work under normalization $S(0) - S(T) = 1$, these cancel) gives the alternative representation

$$\bar{\mathcal{E}}[\lambda] = \tfrac{1}{2} + \int_0^T S(\tau)\,\mathcal{E}'(\tau)\,d\tau = \tfrac{1}{2} + \int_0^T e^{-\Lambda(\tau)}\mathcal{E}'(\tau)\,d\tau, \tag{A.3}$$

where $\mathcal{E}'(\tau) = \frac{d}{d\tau}\mathcal{E}(\tau) = \frac{1}{2}\big(\sin(T-\tau)\cos\tau - \cos(T-\tau)\sin\tau\big) = -\frac{1}{2}\sin(T-2\tau)$.

Unconstrained first variation. For a perturbation $\lambda_\epsilon = \lambda + \epsilon\,\delta\lambda$ with sufficiently regular $\delta\lambda$, we differentiate (A.1):

$$\bar{\mathcal{E}}[\lambda_\epsilon] = \int_0^T (\lambda + \epsilon\,\delta\lambda)(\tau)\,\exp\Big(-\int_0^\tau (\lambda+\epsilon\,\delta\lambda)(u)\,du\Big)\mathcal{E}(\tau)\,d\tau.$$

Expanding to first order in ϵ yields

$$\frac{d}{d\epsilon}\bar{\mathcal{E}}[\lambda_\epsilon]\Big|_{\epsilon=0} = \int_0^T \delta\lambda(\tau)e^{-\Lambda(\tau)}\mathcal{E}(\tau)\,d\tau - \int_0^T \lambda(\tau)e^{-\Lambda(\tau)}\mathcal{E}(\tau)\Big(\int_0^\tau \delta\lambda(u)\,du\Big)d\tau. \tag{A.4}$$

Introduce the weight $\Phi(\tau) := \mathcal{E}(\tau)e^{-\Lambda(\tau)}$. Then the second integral in (A.4) can be rewritten by exchanging the order of integration over the triangle $0 \le u \le \tau \le T$:

$$\int_0^T \delta\lambda(\tau)\Phi(\tau)d\tau = \int_0^T \delta\lambda(u)\Big(\int_u^T \lambda(\tau)\Phi(\tau)d\tau\Big)du.$$

Stationarity for arbitrary $\delta\lambda$ would require

$$\Phi(\tau) = \int_\tau^T \lambda(t)\Phi(t)\,dt \qquad (0 \le \tau \le T). \tag{A.5}$$

Differentiating (A.5) with respect to τ and using Leibniz' rule gives $\mathcal{E}'(\tau)e^{-\Lambda(\tau)} = 0$ for all τ, i.e. $\mathcal{E}' \equiv 0$, which is false. Therefore *no smooth, interior, unconstrained minimizer exists.* Heuristically, minimizing sequences concentrate the hazard mass at the minimum of $\mathcal{E}(\tau)$.

Quadratically constrained problem

To obtain a well-posed optimization problem we introduce an energy constraint

$$\int_0^T \lambda(\tau)^2 d\tau = K, \qquad K > 0. \tag{A.6}$$

Consider the augmented functional

$$\mathcal{J}[\lambda] = \bar{\mathcal{E}}[\lambda] + \alpha\Big(\int_0^T \lambda(\tau)^2 d\tau - K\Big), \quad \alpha \in \mathbb{R}. \tag{A.7}$$

Using (A.4), the first variation of $\mathcal{J}$ yields the Euler–Lagrange condition

$$\Phi(\tau) - \int_\tau^T \lambda(t)\Phi(t)dt + 2\alpha\lambda(\tau) = 0. \tag{A.8}$$

Substituting $\Phi(t) = \mathcal{E}(t)e^{-\Lambda(t)}$ gives

$$\mathcal{E}(\tau)e^{-\Lambda(\tau)} - \int_\tau^T \lambda(t)\mathcal{E}(t)e^{-\Lambda(t)}dt + 2\alpha\lambda(\tau) = 0. \tag{A.9}$$

Differentiating (A.9) w.r.t. τ (and using $\Lambda' = \lambda$) produces

$$\mathcal{E}'(\tau)e^{-\Lambda(\tau)} + 2\alpha\lambda'(\tau) = 0 \implies \lambda'(\tau) = -\frac{1}{2\alpha}\mathcal{E}'(\tau)e^{-\Lambda(\tau)}. \tag{A.10}$$

Since $\lambda = \Lambda'$, we arrive at the nonlinear boundary value problem

$$\Lambda''(\tau) + \frac{1}{2\alpha}\mathcal{E}'(\tau)e^{-\Lambda(\tau)} = 0, \qquad \Lambda(0) = 0, \quad \Lambda'(T) = \lambda(T) = 0. \tag{A.11}$$

This ODE encodes the trade-off between concentrating λ at the minimum of $\mathcal{E}$ and the quadratic energy penalty.

Remark. The ill-posedness of the unconstrained problem motivates recasting the optimization in alternative variables (hazard, survival, or induced density), which we now develop.

Alternative change of variables approach

We start again from the stochastic average functional

$$\bar{\mathcal{E}}[\lambda] = \int_0^T \lambda(\tau) e^{-\int_0^\tau \lambda(u)\,du}\, \mathcal{E}(\tau)\, d\tau, \tag{A.12}$$

where $\mathcal{E}(\tau) = \frac{1}{2}(1 - \sin(T-\tau)\sin\tau)$, and impose the (general) constraint

$$\int_0^T \lambda(\tau)^k\, d\tau = M, \qquad k > 0,\ M > 0, \tag{A.13}$$

with Lagrange multiplier α. Set the survival-like variable

$$\mu(\tau) := e^{-\int_0^\tau \lambda(u)\,du}, \qquad \mu(0) = 1,\ \mu(\tau) > 0,\ \mu'(\tau) = -\lambda(\tau)\mu(\tau) \le 0. \tag{A.14}$$

Then $\lambda(\tau) = -\mu'(\tau)/\mu(\tau)$ and

$$\bar{\mathcal{E}}[\mu] = \int_0^T \Big(-\frac{\mu'(\tau)}{\mu(\tau)}\Big)\mu(\tau)\mathcal{E}(\tau)\, d\tau = -\int_0^T \mu'(\tau)\mathcal{E}(\tau)\, d\tau. \tag{A.15}$$

Integrating by parts gives

$$\bar{\mathcal{E}}[\mu] = -\big[\mu(\tau)\mathcal{E}(\tau)\big]_0^T + \int_0^T \mu(\tau)\mathcal{E}'(\tau)\, d\tau = \tfrac{1}{2}(1-\mu(T)) + \int_0^T \mu(\tau)\mathcal{E}'(\tau)\, d\tau, \tag{A.16}$$

using $\mathcal{E}(0) = \mathcal{E}(T) = \frac{1}{2}$ and $\mu(0) = 1$. The constraint in μ becomes

$$\int_0^T \Big(-\frac{\mu'(\tau)}{\mu(\tau)}\Big)^k d\tau = M. \tag{A.17}$$

Form the augmented functional (dropping the constant $-\alpha M$):

$$\mathcal{J}[\mu] = \int_0^T \Big[\mu(\tau)\mathcal{E}'(\tau) + \alpha\Big(-\frac{\mu'(\tau)}{\mu(\tau)}\Big)^k\Big] d\tau. \tag{A.18}$$

Define $L(\tau,\mu,\mu') = \mu\mathcal{E}'(\tau) + \alpha(-1)^k\mu'(\tau)^k\mu(\tau)^{-k}$. The Euler–Lagrange equation $\partial L/\partial\mu - \frac{d}{d\tau}(\partial L/\partial\mu') = 0$ yields

$$\mathcal{E}'(\tau) + \alpha(-1)^k(-k)\mu'(\tau)^k\mu(\tau)^{-k-1} - \frac{d}{d\tau}\Big(\alpha(-1)^k k\mu'(\tau)^{k-1}\mu(\tau)^{-k}\Big) = 0. \tag{A.19}$$

Equivalently, factoring $\alpha(-1)^k k$,

$$\mathcal{E}'(\tau) - \alpha(-1)^k k\Big[\mu'(\tau)^k\mu(\tau)^{-k-1} + \frac{d}{d\tau}\big(\mu'(\tau)^{k-1}\mu(\tau)^{-k}\big)\Big] = 0. \tag{A.20}$$

Expanding the τ-derivative explicitly for clarity (for $k \ge 1$):

$$\frac{d}{d\tau}\big(\mu'^{k-1}\mu^{-k}\big) = (k-1)\mu'^{k-2}\mu''\mu^{-k} - k\mu'^{k-1}\mu^{-k-1}\mu'. \tag{A.21}$$

Substituting this into (A.20) and collecting terms gives,

$$\mathcal{E}'(\tau) - \alpha(-1)^k k(k-1)\mu'^{k-2}\mu^{-k-1}\big(\mu\mu'' - \mu'^2\big) = 0. \tag{A.22}$$

Case $k = 1$. When $k = 1$ $\mathcal{E}'(\tau) = 0$, which is impossible (as previously noted).

Case $k = 2$. Equation (A.22) becomes

$$\mathcal{E}'(\tau) - 2\alpha\,\mu^{-3}(\mu\mu'' - \mu'^2) = 0. \tag{A.23}$$

Let $y(\tau) = \ln \mu(\tau)$ so $\mu = e^y$, $\mu' = y'e^y$ and $\mu'' = (y'' + y'^2)e^y$. Then

$$\mu\mu'' - \mu'^2 = e^y(y'' + y'^2)e^y - y'^2e^{2y} = y''e^{2y}. \tag{A.24}$$

Substituting gives

$$y''(\tau) = \frac{1}{2\alpha}\mathcal{E}'(\tau)e^{-y(\tau)}, \qquad y(0) = 0,\ y'(T) = 0 \text{ (if } \lambda(T) = 0). \tag{A.25}$$

This is the same ODE as (A.11) and remains unsolvable analytically.

Appendix B

Derivations for Unstructured Search via QAOA

This appendix provides the rigorous derivation for the performance of the Quantum Approximate Optimization Algorithm (QAOA) applied to the unstructured search problem using the standard transverse field mixer $H_0 = \sum \sigma_x^{(j)}$. We first establish the recurrence relation for the transition amplitude and then justify the parameter schedule used to demonstrate the quadratic speedup.

B.1 Proof of the Operator Recurrence Relation

We derive the recursive decomposition of the unitaries used in Section 5.5.2. We define $\beta_k = e^{i\Delta_{C,k}} - 1$ and let $W_r^{(p)}$ be the cumulative mixer operator from layer r to p:

$$W_r^{(p)} = \prod_{k=r}^{p} U_{0,k}, \quad \text{with } W_0^{(0)} = \mathbb{I}. \tag{B.1}$$

We aim to prove by induction that the total p-layer unitary $\mathcal{U}^{(p)}$ satisfies:

$$\mathcal{U}^{(p)} = \sum_{r=1}^{p} W_r^{(p)} (\beta_r |0^n\rangle\langle 0^n|) \mathcal{U}^{(r-1)} + W_0^{(p)}. \tag{B.2}$$

Induction Base Case. For $p = 1$, using the definition of a single QAOA layer $\mathcal{U}_1 = U_{0,1} U_{C,1}$ and the expansion $U_{C,1} = \mathbb{I} + \beta_1 |0^n\rangle\langle 0^n|$, we have:

$$\mathcal{U}^{(1)} = U_{0,1}(I + \beta_1 |0^n\rangle\langle 0^n|) = U_{0,1}\beta_1 |0^n\rangle\langle 0^n| \mathbb{I} + U_{0,1}, \tag{B.3}$$

which matches Eq. (B.2) for $r = 1$ and $\mathcal{U}^{(0)} = \mathbb{I}$.

Inductive Step. Assume the relation holds for p. We calculate the operator for $p + 1$ layers:

$$\begin{aligned} \mathcal{U}^{(p+1)} &= \mathcal{U}_{p+1} \mathcal{U}^{(p)} = U_{0,p+1}(\mathbb{I} + \beta_{p+1} |0^n\rangle\langle 0^n|) \mathcal{U}^{(p)} \\ &= \beta_{p+1} U_{0,p+1} |0^n\rangle\langle 0^n| \mathcal{U}^{(p)} + U_{0,p+1} \mathcal{U}^{(p)}. \end{aligned} \tag{B.4}$$

Substituting the inductive hypothesis for $\mathcal{U}^{(p)}$ into the second term:

$$\begin{aligned}\mathcal{U}^{(p+1)} &= \beta_{p+1}U_{0,p+1}|0^n\rangle\langle 0^n|\mathcal{U}^{(p)} + U_{0,p+1}\left(\sum_{r=1}^{p} W_r^{(p)}\beta_r|0^n\rangle\langle 0^n|\mathcal{U}^{(r-1)} + W_0^{(p)}\right) \\ &= W_{p+1}^{(p+1)}\beta_{p+1}|0^n\rangle\langle 0^n|\mathcal{U}^{(p)} + \sum_{r=1}^{p} W_r^{(p+1)}\beta_r|0^n\rangle\langle 0^n|\mathcal{U}^{(r-1)} + W_0^{(p+1)} \\ &= \sum_{r=1}^{p+1} W_r^{(p+1)}\beta_r|0^n\rangle\langle 0^n|\mathcal{U}^{(r-1)} + W_0^{(p+1)}. \end{aligned} \tag{B.5}$$

The operator recurrence is thus proved for all p.

Derivation of the Transition Amplitude

With the operator expansion established, we define the complex amplitude as the overlap between the target state and the evolved initial state: $a_p = \langle 0^n|\mathcal{U}^{(p)}|\psi_0\rangle$. By projecting Eq. (B.2) onto $\langle 0^n|$ and $|\psi_0\rangle$, we obtain:

$$\begin{aligned} a_p &= \sum_{r=1}^{p} \beta_r\langle 0^n|W_r^{(p)}|0^n\rangle\langle 0^n|\mathcal{U}^{(r-1)}|\psi_0\rangle + \langle 0^n|W_0^{(p)}|\psi_0\rangle \\ &= \sum_{r=1}^{p} \beta_r\langle 0^n|W_r^{(p)}|0^n\rangle a_{r-1} + \langle 0^n|W_0^{(p)}|\psi_0\rangle. \end{aligned} \tag{B.6}$$

Substituting the matrix elements derived in the main text, we recover the recurrence relation used in Section 5.5.2.

B.2 Combinatorial Interpretation/ Original Derivation

The recurrence relation derived in Section B.1 can be intuitively understood through a combinatorial expansion of the total unitary evolution. The total operator $\mathcal{U}^{(p)}$ is the ordered product of p layers:

$$\mathcal{U}^{(p)} = \prod_{k=p}^{1}\mathcal{U}_k = \prod_{k=p}^{1}\left(U_{0,k} + B_k\right), \tag{B.7}$$

where we have separated the single-layer operator into a "free" propagation term $U_{0,k}$ and an "interaction" term $B_k = U_{0,k}\beta_k|0^n\rangle\langle 0^n|$. Expanding this product yields 2^p terms, corresponding to all possible binary sequences $\varepsilon \in \{0,1\}^p$ where, at each step k, we either select the free evolution $U_{0,k}$ (if $\varepsilon_k = 0$) or the interaction term B_k (if $\varepsilon_k = 1$).

To structure this sum, we group terms based on the *latest* layer index r where an interaction occurred (i.e., the largest r such that $\varepsilon_r = 1$). This partitions the sum into:

$$\mathcal{U}^{(p)} = \underbrace{\prod_{k=p}^{1} U_{0,k}}_{\text{No interactions}} + \sum_{r=1}^{p} \sum_{\substack{\varepsilon \\ \text{last interaction at } r}} \prod_{k=p}^{1} \cdots \tag{B.8}$$

For a fixed r, the evolution after layer r is purely free (composed of $U_{0,k}$ for $k > r$), while the evolution before layer r (indices $k < r$) includes all possible histories. This allows us to factorize

the terms:

$$\text{Term}_r = \underbrace{\left(\prod_{k=p}^{r+1} U_{0,k}\right)}_{\text{Future (Free)}} B_r \underbrace{\mathcal{U}^{(r-1)}}_{\text{Past (All histories)}}. \tag{B.9}$$

Substituting $B_r = U_{0,r}\beta_r|0^n\rangle\langle 0^n|$ and identifying the product of mixers as $W_r^{(p)} = \prod_{k=p}^{r} U_{0,k}$, we recover the recurrence relation:

$$\mathcal{U}^{(p)} = W_0^{(p)} + \sum_{r=1}^{p} W_r^{(p)} \beta_r |0^n\rangle\langle 0^n| \mathcal{U}^{(r-1)}. \tag{B.10}$$

This confirms that the total unitary is the sum of a background trajectory (no interactions) and a superposition of trajectories that last "scattered" off the target state at layer r.

B.3 Justification of the Parameter Schedule

The exact recurrence for the transition amplitude $a_p = \langle 0^n|\mathcal{U}^{(p)}|\psi_0\rangle$ derived in Chapter 5 is:

$$a_p = \frac{1}{\sqrt{2^n}} e^{-in\Theta_{p,0}} + \sum_{r=1}^{p} (e^{i\Delta_{C,r}} - 1)\cos^n(\Theta_{p,r}) a_{r-1}. \tag{B.11}$$

Oracle Angles. The contribution of each layer r is scaled by the oracle coefficient $\beta_r = e^{i\Delta_{C,r}} - 1$. To maximize the magnitude of the amplitude update at each step, we maximize:

$$|\beta_r|^2 = |e^{i\Delta_{C,r}} - 1|^2 = 2 - 2\cos(\Delta_{C,r}). \tag{B.12}$$

The global maximum is achieved when $\cos(\Delta_{C,r}) = -1$, implying $\Delta_C = \pi$ and $\beta_r = -2$. This choice corresponds to the standard phase flip on the target state, ensuring the maximum possible probability transfer.

Mixer Angles. The accumulated history is weighted by the propagator kernel $\cos^n(\Theta_{p,r})$. To prevent the exponential suppression of previous amplitudes ($|\cos|^n \to 0$ as $n \to \infty$), we require $|\cos(\Theta_{p,r})| \approx 1$.

Simultaneously, we must ensure constructive interference between the "source" term (the first term in Eq. (B.11)) and the summation. For a constant angle Δ_0, the source phase is $e^{-in(p\Delta_0)}$. To align this phase with the updates (scaled by $\beta = -2$), we set $n\Delta_0 = \pi$ (i.e., $\Delta_0 = \pi/n$). In the small depth limit $p \ll n$, the cumulative angles satisfy:

$$\cos^n\left(k\frac{\pi}{n}\right) \approx \left(1 - \frac{k^2\pi^2}{2n^2}\right)^n \approx e^{-k^2\pi^2/2n} \approx 1. \tag{B.13}$$

Under this schedule, the source term simplifies to $(-1)^p/\sqrt{N}$, where $N = 2^n$, and the history is preserved without decay.

B.4 Inductive Proof of the Success Probability

In the limit $p \ll n$, with $\Delta_{C,r} = \pi$ and $\Delta_0 = \pi/n$, the recurrence for a_p simplifies to a linear form with constant coefficients:

$$a_p = \frac{(-1)^p}{\sqrt{N}} - 2\sum_{r=1}^{p} a_{r-1}, \quad \text{with } a_0 = \frac{1}{\sqrt{N}}. \tag{B.14}$$

We propose the ansatz $a_p = \frac{(-1)^p(2p+1)}{\sqrt{N}}$.

Algebraic Lemma. For any integer $m \geq 0$, the following identity holds:

$$\sum_{k=0}^{m}(-1)^k(2k+1) = (-1)^m(m+1). \tag{B.15}$$

Proof of Lemma: We proceed by induction on m.

Base case ($m = 0$): The left-hand side evaluates to

$$\sum_{k=0}^{0}(-1)^k(2k+1) = (-1)^0(2\cdot 0+1) = 1,$$

while the right-hand side gives $(-1)^0(0+1) = 1$. Both sides agree.

Inductive step: Assume the identity holds for some $m \geq 0$. For $m+1$:

$$\begin{aligned}\sum_{k=0}^{m+1}(-1)^k(2k+1) &= (-1)^m(m+1) + (-1)^{m+1}(2m+3)\\ &= (-1)^{m+1}\big[-(m+1)+2m+3\big] = (-1)^{m+1}(m+2).\end{aligned}$$

Proof of the Amplitude Scaling.

Proof. **Base Case ($p = 0$):** $a_0 = \frac{(-1)^0(2(0)+1)}{\sqrt{N}} = \frac{1}{\sqrt{N}}$, which matches the initial overlap.

Inductive Step: Assume the formula holds for all $k < p$. We compute a_p using Eq. (B.14):

$$\begin{aligned}a_p &= \frac{(-1)^p}{\sqrt{N}} - 2\sum_{k=0}^{p-1}a_k\\ &= \frac{1}{\sqrt{N}}\left[(-1)^p - 2\sum_{k=0}^{p-1}(-1)^k(2k+1)\right].\end{aligned} \tag{B.16}$$

Applying the Algebraic Lemma (Eq. (B.15)) with $m = p-1$:

$$\begin{aligned}a_p &= \frac{1}{\sqrt{N}}\left[(-1)^p - 2\left((-1)^{p-1}p\right)\right]\\ &= \frac{1}{\sqrt{N}}\left[(-1)^p - 2\left(-(-1)^p p\right)\right]\\ &= \frac{1}{\sqrt{N}}\left[(-1)^p + 2p(-1)^p\right] = \frac{(-1)^p(2p+1)}{\sqrt{N}}.\end{aligned} \tag{B.17}$$

The formula is thus proven for all $p \geq 0$. □

Success Probability. The probability of success P_p is the squared modulus of the amplitude:

$$P_p = |a_p|^2 = \frac{(2p+1)^2}{N}. \tag{B.18}$$

This confirms that the success probability grows quadratically with the depth p in the $p \ll n$ limit, identifying a $p \sim \sqrt{N}$ scaling to reach a constant success probability.

Appendix C

Derivation of the Partition Function using Bessel Functions

This appendix provides the detailed technical derivation of the partition function $Z(T, \alpha)$ for the entropy-regularized single-switch problem analyzed in Chapter 4. The calculation relies on identities from Bessel function theory found in (Abramowitz and Stegun 1972).

We recall from Eq. (4.44) that the single-trajectory energy is:

$$\mathcal{E}(t) = \frac{1}{2}\big(1 - \sin t \, \sin(T - t)\big). \tag{C.1}$$

Trigonometric Linearization and Change of Variables

To evaluate the partition function $Z(T, \alpha)$, we first linearize the trigonometric product using the identity $\sin A \sin B = \frac{1}{2}[\cos(A - B) - \cos(A + B)]$:

$$\mathcal{E}(t) \;=\; \frac{1}{2}\left(1 - \frac{1}{2}\big(\cos(2t - T) - \cos T\big)\right) \;=\; \frac{1}{2} - \frac{1}{4}\cos(2t - T) + \frac{1}{4}\cos T.$$

Substituting this into the Boltzmann weight $e^{-\beta\mathcal{E}(t)}$:

$$e^{-\beta\mathcal{E}(t)} \;=\; \exp\left\{-\frac{\beta}{2} - \frac{\beta}{4}\cos T + \frac{\beta}{4}\cos(2t - T)\right\}.$$

To simplify notation, we introduce the parameter κ related to the inverse temperature. Note that due to the factors arising from the linearization, the natural scaling is:

$$\kappa \;:=\; \frac{\beta}{4} \;=\; \frac{1}{4\alpha}.$$

The Boltzmann weight becomes:

$$e^{-\beta\mathcal{E}(t)} \;=\; e^{-2\kappa - \kappa\cos T}\, e^{\kappa\cos(2t - T)}.$$

The partition function integral is then:

$$Z(T, \alpha) \;=\; e^{-2\kappa - \kappa\cos T} \int_0^T e^{\kappa\cos(2t - T)}\, dt. \tag{C.2}$$

We perform the change of variables $u = 2t - T$, which implies $dt = \frac{1}{2}\, du$. The integration limits transform from $[0, T]$ to $[-T, T]$:

$$Z(T, \alpha) \;=\; \frac{1}{2} e^{-2\kappa - \kappa\cos T} \int_{-T}^{T} e^{\kappa\cos u}\, du. \tag{C.3}$$

Bessel Expansion of the Angular Integral

We define the auxiliary integral $Q(T,\kappa) := \int_{-T}^{T} e^{\kappa\cos u}\,du$. This can be evaluated analytically using the Jacobi-Anger expansion:

$$e^{\kappa\cos u} \;=\; I_0(\kappa) + 2\sum_{n=1}^{\infty} I_n(\kappa)\,\cos(nu),$$

where $I_n(\kappa)$ are the modified Bessel functions of the first kind. Integrating term-by-term over the symmetric interval $u \in [-T,T]$ eliminates the sine terms (which are odd), yielding:

$$Q(T,\kappa) = 2T\,I_0(\kappa) + 4\sum_{n=1}^{\infty} I_n(\kappa)\,\frac{\sin(nT)}{n}. \tag{C.4}$$

Substituting this back into Eq. (C.3), the full partition function is:

$$Z(T,\alpha) \;=\; \frac{1}{2}e^{-\kappa(\cos T+2)}\,Q(T,\kappa). \tag{C.5}$$

Analytic Expression for the Stochastic Average

We now invoke the fundamental identity $\bar{\mathcal{E}}_\alpha(T) = -\frac{\partial}{\partial\beta}\ln Z$. Using the chain rule with $\beta = 4\kappa$:

$$\frac{\partial}{\partial\beta} \;=\; \frac{\partial\kappa}{\partial\beta}\frac{\partial}{\partial\kappa} \;=\; \frac{1}{4}\frac{\partial}{\partial\kappa}.$$

First, we compute the logarithmic derivative of Z with respect to κ:

$$\frac{\partial}{\partial\kappa}\ln Z(T,\alpha) \;=\; \frac{\partial}{\partial\kappa}\left(\text{const} - 2\kappa - \kappa\cos T + \ln Q(T,\kappa)\right) \;=\; -2 - \cos T + \frac{Q_\kappa(T,\kappa)}{Q(T,\kappa)},$$

where $Q_\kappa := \partial Q/\partial\kappa$. The stochastic average is therefore:

$$\bar{\mathcal{E}}_\alpha(T) \;=\; -\frac{1}{4}\left(-2 - \cos T + \frac{Q_\kappa(T,\kappa)}{Q(T,\kappa)}\right).$$

Rearranging terms yields the general analytic solution:

$$\boxed{\bar{\mathcal{E}}_\alpha(T) = \frac{1}{2} + \frac{1}{4}\cos T - \frac{1}{4}\,\frac{Q_\kappa(T,\kappa)}{Q(T,\kappa)}, \qquad \kappa = \frac{1}{4\alpha}.} \tag{C.6}$$

The derivative Q_κ is computed using the recurrence relations $I_0'(\kappa) = I_1(\kappa)$ and $I_n'(\kappa) = \frac{1}{2}(I_{n-1}(\kappa) + I_{n+1}(\kappa))$:

$$Q_\kappa(T,\kappa) = 2T\,I_1(\kappa) + 2\sum_{n=1}^{\infty}\big(I_{n-1}(\kappa) + I_{n+1}(\kappa)\big)\,\frac{\sin(nT)}{n}. \tag{C.7}$$

This completes the derivation of the analytic expression for the regularized stochastic average in terms of modified Bessel functions. The asymptotic analysis and physical interpretation of this result are presented in Section 4.4 of the main text.