

Fundamental Limitations of QAOA on Constrained Problems and a Route to Exponential Enhancement

Chinonso Onah^{1,2} and Kristel Michielsen^{2,3}

¹Volkswagen AG, Berliner Ring 2, Wolfsburg 38440, Germany

²Department of Physics, RWTH Aachen, Germany

³Forschungszentrum Jülich, Germany

We study fundamental limitations of the generic Quantum Approximate Optimization Algorithm (QAOA) on constrained problems where valid solutions form a low-dimensional manifold inside the Boolean hypercube, and we present a provable route to exponential enhancements via constraint embedding. Focusing on permutation-constrained objectives, we show that the standard “generic” QAOA ansatz—with transverse-field mixer and diagonal, commuting, r -local cost—faces an intrinsic feasibility bottleneck: even after optimizing angles, depth- p circuits with $p = O(n)$ cannot raise the total probability mass on the feasible manifold much above the uniform baseline $|\Pi|/2^N$, up to polynomial slack in n .

Against this envelope we place a minimal constraint-enhanced kernel (CE-QAOA) that operates directly inside a product one-hot subspace and mixes with a block-local XY Hamiltonian. For permutation constrained problems on n variables, we prove an angle-robust, depth-matched exponential enhancement of the form

$$\frac{P_p^{(\text{CE})}}{P_p^{(\text{gen})}} = \exp[\Theta(n^2)] \quad (1)$$

for all depths p up to a linear fraction of n under a polynomial incidence-growth condition on the interaction hypergraph.

Our analysis triangulates this separation via four complementary bounds on generic QAOA: (i) Walsh-Fourier/Krawtchouk estimates showing that permutation one-hot constraints have exponentially small low-degree and high-degree Fourier mass; (ii) a cost-angle averaging argument that pins the typical-angle feasible baseline at $|\Pi|/2^N$; (iii) fourth-moment (L_4) bounds that control typical fluctuations around this baseline; and (iv) light-cone locality arguments that show shallow circuits cannot build the global correlations demanded by permutation constraints, even up to $p = \Theta(n)$. Additionally, we prove a parameter-transfer theorem: CE-QAOA amplifies any generic-QAOA parameter set by the exponential factor $P_{\max}^{(\text{gen})} \cdot \sqrt{2\pi n} e^n$ at $p=1$. Thanks to the problem-algorithm co-design in the kernel definition, these results apply to a broad class of constrained problems satisfying Definition 1.

1 Introduction

Combinatorial optimization problems (COPs) with global constraints pose a specific challenge for shallow variational quantum circuits. Feasible solutions occupy a vanishing frac-

tion of the full Hilbert space, so any method that explores the entire hypercube pays an intrinsic feasibility penalty. This paper isolates that bottleneck for the Quantum Approximate Optimization Algorithm (QAOA)[1] and develops a transparent way to assess the state prepared by the algorithm for arbitrary shallow schedules of length $p = O(n)$. QAOA is arguably one of the most studied Quantum algorithms in recent years and numerous variants have emerged. In this paper, we shall refer to the original construction in Ref. [1] as the generic QAOA ansatz in order to facilitate a clear comparison with a minimal constraint-aware variant that bakes constraint structure and symmetries into its quantum dynamics. The distinction is made explicit in Subsection 1.2.

Instead of monitoring only expectation values and approximation ratios, we study the full probability profile of depth- p QAOA states through four lines of analyses. First, a Walsh–Fourier calculation on the Boolean cube reveals that the one-hot constraints behind permutation encodings have exponentially small low-degree Fourier mass. At $p = 1$ the X -mixer acts by pure phases in this Walsh/Fourier basis, so the generic ansatz cannot significantly lift feasible mass above the uniform baseline, up to polynomial slack, and this holds uniformly over angles. Second, a gentle cost-angle averaging argument, applicable after a simple rescaling that holds for standard integer or rational instances, pins the feasible baseline at $|\Pi|/2^N$. Third, a hypercontractive (l_4) bound on the output amplitudes then show that most angles concentrate around this baseline, which is useful for understanding shallow-depth training without assuming anything about gradients. Finally, locality enters through light-cone growth. With commuting, r -local diagonal costs, information propagates at most one neighborhood per layer. A dependency-graph argument turns this into a bound on the joint probability that all row and column constraints hold, and the same limitation persists up to linear depth $p = \alpha n$ as long as the interaction hypergraph has polynomial incidence growth $|N_1(S)| \leq \Delta_n |S|$ with $\Delta_n = \text{poly}(n)$. The resulting picture is that the generic ansatz faces a structural ceiling that does not disappear at depth one and continues to matter at larger but still shallow depths.

Against this envelope, we place the recently introduced variant of QAOA called the Constraint-Enhanced Quantum Approximate Optimization algorithm (CE-QAOA)[2]. This is a constraint-enhanced kernel that works directly on the one-hot manifold with explicit symmetry enhancements derived from problem structure. The ingredients include an ancilla-free W -state initialization, a normalized block- XY mixer that preserves the encoded space and penalty derived co-design symmetry. This framework provides an existence result needed to benchmark the probability profile of generic QAOA. Thus, we take the bounds derived for CE-QAOA kernel from prior work [2] as a minimal exemplar of problem–algorithm co-design as a route to provable exponential enhancement.

For any permutation constrained problem of size n , its one-hot encoding requires $N = n^2$ qubits (Def 1). The resulting enhancement is quantified in terms of $\exp(\Theta(n^2))$ separation between the constraint enhanced variant and the generic QAOA in Eq. (1) and proved later in Theorem 7 for $p = 1$ and Theorem 8 $p = \Theta(n)$ respectively. It is worth noting that we have only used the minimal, most conservative guarantee from Ref. [2]. Our aim is to place the minimal guarantee from the constraint-enhanced kernel against a sharp upper envelope for the generic ansatz and to show how even a modest in-subspace design changes the feasibility profile at shallow depth.

In the spirit of parameter transfer for QAOA, we prove that angles which are optimal for the *generic* ansatz at depth $p=1$ already guarantee an exponential boost when reused verbatim inside the CE-QAOA kernel. Writing $P_{\max}^{(\text{gen})} = \max_{\beta, \gamma} P_{\Pi}^{(\text{gen})}(\beta, \gamma)$ our analysis

shows that for every $n \geq 2$ and every (γ, β) pair

$$P^{(\text{CE})} \geq \sqrt{2\pi n} e^n P_{\max}^{(\text{gen})}.$$

This “parameter-transfer amplification” complements our envelope bounds and explains the robust gains observed in the noiseless circuit comparisons reported in Figures 1a–3.

The rest of the paper proceeds as follows. Section 1.1 places our work in context of recent research works in fundamental limitations of shallow Variational Quantum Algorithms (VQA). Section 1.2 reviews the generic ansatz, recalls the definition of the CE-QAOA kernel and records their symmetry, mixing properties and the core existence result. Section 2 outlines our main exponential enhancement results culminating in the exponential parameter transfer amplification in Section 2.4. In Section 3 we develop the materials and methods necessary to establish the angle-averaging, fourth-moment, and harmonic-analysis bounds that define the generic envelope. Section 3.7 develops light-cone barriers for the generic QAOA ansatz at $p = 1$ and Section 3.8 extend the light-cone barrier analyses to linear depths. Section 4 concludes our findings. Supplementary details are relegated to the Appendix A.

1.1 Related Work

Variational quantum algorithms (VQAs), and QAOA in particular, have been studied along three complementary axes: (i) *trainability and landscape structure*, (ii) *locality/light-cone limitations at shallow depth*, and (iii) *architecture- or problem-aware ansätze* that encode constraints or symmetries from the outset. On the landscape side, early work identified *barren plateaus* as a key obstruction to gradient-based training in large systems [3, 4]. More recent analyses quantify landscape *information content*, revealing how much task-relevant signal survives in variational loss surfaces [5], and study the *energy landscape* structure of small-depth QAOA instances [6]. The QAOA design space has also broadened beyond the original transverse-field mixer to *alternating operator ansätze* tailored to problem structure [7], while symmetry constraints can both obstruct and guide expressivity [8]. Mitigation strategies such as *classical shadows* for gradient estimation [9] and *warm starts* [10] aim to stabilize training or inject structure into initial states.

A second line of work formalizes *locality* as the principal barrier at small depths since shallow circuits can only build correlations within a *light cone* whose radius grows at most linearly with depth. These are essentially Lieb–Robinson-type locality bounds [11, 12], and they underlie modern limitations for low-depth QAOA on sparse graphs. For bounded-degree graph families, Farhi, Gamarnik, and Gutmann proved that small- p QAOA “needs to see the whole graph,” producing typical- and worst-case examples where performance is capped by local neighborhoods [13]. At constant levels $p = O(1)$, quantitative barriers have been established on sparse hypergraphs and spin-glass models [14], while in a broader framework of *local quantum algorithms*, limitations propagate up to $p < \varepsilon \log n$ for certain CSPs with the (coupled) overlap-gap property [15, 16]. Orthogonally, results on concentration for shallow circuits show output observables of low-depth evolutions are tightly concentrated, limiting global coordination at small p [17]. Collectively, these works support the heuristic that constant or slowly growing depth cannot generate the long-range correlations required by *global* constraints.

A third, complementary, line of work focuses on *architecture- and problem-aware* ansätze that encode constraints, symmetries, or known structure directly at the circuit level. In the QAOA setting, the alternating-operator framework extends the original transverse-field mixer by replacing it with constraint-preserving unitaries tailored to specific combinatorial structure [18, 7]. Constrained mixers and symmetry-projected variational families

have been designed to restrict the evolution to feasible manifolds or fixed symmetry sectors, thereby reducing the effective search space and sometimes alleviating barren-plateau effects [19, 20, 8]. Domain-specific encodings and mixers for routing and scheduling problems further illustrate how problem-aware design can unlock strong empirical performance on structured instances [21, 22, 23].

The current paper connects these three perspectives under a single quantitative lens by focusing on the *feasible probability mass* generated by shallow generic versus problem-aware variational circuits under global combinatorial constraints. Rather than bounding approximation ratios instance by instance, we derive angle-uniform upper bounds on the total probability that a shallow, *generic* QAOA state satisfies all permutation-style one-hot constraints, for depths p up to a linear fraction of n on interaction hypergraphs whose incidence growth is polynomial in n . Against this envelope, we place a minimal *constraint-enhanced* kernel (CE-QAOA) that encodes the instance in a constrained one-hot manifold and mixes *within* it, thereby achieving a depth-matched, *angle-robust* exponential separation in feasible mass, cf. Eq. (1), in a way that complements and connects prior landscape, locality, and architecture-aware analyses. By construction, the explicit problem–algorithm co-design in the kernel definition allows these results to apply to a broad class of constrained problems satisfying Definition 1.

Beyond landscape and locality, there is a practical line on *warm starts* and *parameter transfer* for variational circuits, where parameters found in one setting are reused in another to reduce training overhead or preserve performance [10, 24]. We show that the angles that maximise the feasible mass for generic $p=1$ QAOA amplify that mass by an exponential factor when transferred unchanged to the CE-QAOA subspace, yielding at least $\sqrt{2\pi n} e^n P_{\max}^{(\text{gen})}$. This provides a rigorous *parameter-transfer amplification* mechanism that dovetails with, but is distinct from, warm-starting and alternating-operator design [18, 10].

1.2 Background

1.2.1 Generic QAOA

The generic QAOA [1] prepares the state $|\psi_p(\vec{\gamma}, \vec{\beta})\rangle = [\prod_{\ell=1}^p e^{-i\beta_\ell \sum_j X_j} e^{-i\gamma_\ell H_C}] |+\rangle^{\otimes N}$, i.e., a depth- p alternation of a transverse-field X mixer and a diagonal cost unitary on the computational basis. Parameters $(\vec{\gamma}, \vec{\beta})$ are optimized (classically) to maximize an empirical objective. For the algorithm to be performant, typical schedules need to be short ($p = O(1)$ or $p \ll N$) without going to the adiabatic limit[40]. Common ideas to improve algorithmic performance include linear/adiabatic ramps[39], parameter transfer across sizes/instances to reduce training cost while keeping the same mixer/cost structures [24], recursive elimination to iteratively fix variables based on low-depth runs, shrinking the instance before a final classical solve [8, 25, 26], warmstarting QAOA to help landscapes and convergence [10, 27].

1.2.2 Constraint-Enhanced QAOA

The *alternating-operator* (“QAOA+”) paradigm augments the generic X -mixer by feasibility-preserving mixers that act invariantly on constraint projectors, thereby confining evolution to structured subspaces [18, 20]. The *Constraint-Enhanced QAOA* (CE-QAOA) introduced in Ref. [2] follows this direction but makes the problem–algorithm co-design explicit. As can be seen in Definition 1, the kernel specifies (i) the valid problem classes with the admissible penalty structure and symmetries, the mixer and initial state.

Definition 1 (CE-QAOA kernel). *An optimization instance I belongs to the CE-QAOA kernel if there exist integers $n, m \in \mathbb{N}$ and the one-hot encoder $\mathbf{E}_{\text{1hot}}$ that initializes the dynamics in the the constrained fixed-Hamming-weight space*

$$\mathcal{H}_{\text{OH}} = (\mathcal{H}_1)^{\otimes m}, \quad \mathcal{H}_1 = \text{span}\{|e_1\rangle, \dots, |e_n\rangle\} \quad (\text{one excitation per block}),$$

together with Hamiltonians H_{pen} and H_{obj} on $\mathcal{H}_{\text{OH}}$ such that:

- (a) *Penalty structure. H_{pen} is a sum of squared affine one-hot/degree/capacity penalties (optionally plus linear forbids) with integer coefficients bounded by $\text{poly}(n)$. Consequently, $\text{spec}(H_{\text{pen}}) \subseteq \{0, 1, \dots, t_{\text{max}}\}$ with $t_{\text{max}} = O(m) = \text{poly}(n)$.*
- (b) *Pattern symmetry. H_{pen} is invariant under (i) block permutations S_m and (ii) global symbol relabelings S_n . Hence the configuration space decomposes into level sets $L_t = \{x : H_{\text{pen}}(x) := \langle x | H_{\text{pen}} | x \rangle = t\}$ that are preserved setwise.*
- (c) *Mixer and initial state. The mixer family is block-local and normalized XY Hamiltonian,*

$$\tilde{H}_{XY}^{(b)} = \frac{1}{n-1} \sum_{1 \leq a < b \leq n} (X_a X_b + Y_a Y_b), \quad U_M(\beta) = \bigotimes_{b=1}^m e^{-i\beta \tilde{H}_{XY}^{(b)}},$$

with $\|\tilde{H}_{XY}^{(b)}\| = O(1)$ on each block. The initial state is the uniform one-hot product

$$|s_0\rangle = |s_{\text{blk}}\rangle^{\otimes m}, \quad |s_{\text{blk}}\rangle = \frac{1}{\sqrt{n}} \sum_{j=1}^n |e_j\rangle \quad (\text{a } W_n \text{ state initialized per block}).$$

Evidently, the identifiable problem classes is broad and far from trivial. They include: Travelling Salesman (TSP/ATSP) [30, 31], Quadratic Assignment Problem (QAP) [32], Vehicle Routing with Capacities (CVRP) [33], Generalized Assignment / Multiple-Knapsack [34], k D Matching (NP-complete for $k \geq 3$) [35], etc. Some base cases are polynomial-time solvable but our interest is in the NP-hard variants (e.g. TSP/GAP/CVRP, etc.) for which the same one-hot encoding and symmetries apply.

2 Results

2.1 Fundamental Limitations

Our point of departure is the following question. By how much can a shallow variational quantum circuit concentrate probability mass on the feasible manifold of a globally constrained problem? We answer this for the canonical case of a permutation constrained problem of size n one-hot encoded into $N = n^2$ qubits (e.g., assignment/TSP) by proving that *generic* QAOA initialized on $|+\rangle^{\otimes N}$ with a X mixer and a diagonal cost cannot raise the feasible mass above the uniform baseline $|\Pi|/2^N$ by more than polynomial factors at any fixed constant depth, and under standard locality, not even at linear depth $p = \alpha n$. Where $|\Pi| = n!$ and $\alpha < 1$.

Recall that the problem Hamiltonian is diagonal in computational basis and is denoted as H_C . For any r -local diagonal H_C and all angles (β, γ) , we show that single payer $p = 1$ generic QAOA circuits cannot generate the long-range correlations required by the feasible solution space by more than polynomial slack:

$$P_{\Pi}^{(\text{gen})}(\beta, \gamma) \leq \frac{|\Pi|}{2^N} \cdot \text{poly}(n). \quad (2)$$

The reason is twofold. Structurally, the $p=1$ light cone confines each block one-hot projector to a finite neighborhood, so the $O(n)$ constraints interact only through short dependencies. This caps any joint boost to at most polynomial over uniform. Spectrally,

the indicator 1_Π is overwhelmingly high-degree in Walsh/Krawtchouk space, while a single X -mixer layer only adds phases. Consequently, average over γ pins the baseline exactly and fourth-moment bounds show typical angles concentrate near it.

In contrast, a *constraint-enhanced* variant (CE-QAOA) that operates inside a block one-hot subspace and mixes with an all-to-all XY block mixer achieves an $\exp[\Theta(n^2)]$ improvement over generic QAOA at the same depth. We arrive at the conclusion by establishing an upper envelope for generic QAOA and a lower envelope for CE-QAOA. To derive the upper envelopes, we employ harmonic analysis, angle-averaging, hyperconstrative fourth-moment, and light-cone analyses. Detailed proofs are presented in Methods and Materials (Section 3). Currently we summarize the main conclusions.

Setup We refer to a COP of size n as permutation constrained if the valid solutions can be identified as a $n \times n$ permutation matrix. For such problems, the one-hot encoding requires $N = n^2$ qubits. Let N be the number of qubits and $\Pi = n!$ the size of the feasible permutation sector (row/column one-hot constrained). Consider the Boolean cube $\{0, 1\}^N$ and write any function $f : \{0, 1\}^N \rightarrow \mathbb{C}$ in the Walsh/Krawtchouk basis $f(x) = \sum_{S \subseteq [N]} \hat{f}(S) \chi_S(x)$ with characters $\chi_S(x) = (-1)^{\langle S, x \rangle}$ and *degree* $|S|$. This is a complete and unique Fourier expansion for any function $f(x)$ with unique coefficients $\hat{f}(S)$ [29].

To connect this formalism the probability amplitudes in generic QAOA, we write

$$H_C = \sum_{y \in \{0,1\}^N} C(y) |y\rangle\langle y| \Rightarrow U_C(\gamma) = e^{-i\gamma H_C} = \sum_y e^{-i\gamma C(y)} |y\rangle\langle y|.$$

With the phase field $\phi_\gamma(y) := e^{-i\gamma C(y)}$ and any basis vector $|x\rangle$ we obtain the probability profile

$$a_{\beta,\gamma}(x) = \langle x | U_X(\beta) U_C(\gamma) | + \rangle^{\otimes N} = 2^{-N/2} \sum_y \underbrace{\langle x | U_X(\beta) | y \rangle}_{K_\beta(x,y)} \phi_\gamma(y).$$

Because $U_X(\beta) = \prod_{i=1}^N e^{-i\beta X_i}$, its Z -basis kernel factors as

$$K_\beta(x, y) = \prod_{i=1}^N \langle x_i | e^{-i\beta X} | y_i \rangle = (\cos \beta)^{N-d(x,y)} (-i \sin \beta)^{d(x,y)} = (\cos \beta)^N (-i \tan \beta)^{d(x,y)}, \quad (3)$$

where $d(x, y) = |x \oplus y|$ is the Hamming distance.

Thus, we obtain a generic representation of the probability of obtaining a feasible permutation from generic QAOA as

$$a_{\beta,\gamma}(x) = 2^{-N/2} \sum_{y \in \{0,1\}^N} K_\beta(x, y) \phi_\gamma(y). \quad (4)$$

For the indicator 1_Π on the permutation submanifold, we show (Materials and Methods Sections 3.2 and 3.3) that the total probability mass is tiny up to logarithmic degree. To derive this we define the *low-degree mass* up to level d as

$$\|f\|_{\leq d}^2 := \sum_{|S| \leq d} |\hat{f}(S)|^2, \quad \text{and high-degree mass } \|f\|_{> d}^2 = \|f\|_2^2 - \|f\|_{\leq d}^2.$$

We introduce Krawtchouk analysis to calculate the low-degree weight in the one-hot sector (Lemma 13), and, after tensorizing rows/columns and convolving to obtain the weight of indicator 1_Π on the permutation submanifold (Section 3.4). Our calculations show that both low-degree and high-degree mass remain exponentially suppressed. We denote the total output pmf as $p(x) = |a_{\beta,\gamma}(x)|^2$. To compute the total feasible mass, we choose a degree truncation $d = T = \Theta(\log n)$ and split $p(x)$ as low and high degree parts follows $p = p^{\leq 2T} + p^{> 2T}$. The low degree part accounts for all the contributions to the Fourier

mass from degrees below the T threshold and the high degree split captures all degree contributions above T . For the *low-degree* part we have (Lemma 17),

$$\frac{1}{2^N} \langle \mathbf{1}_\Pi, p^{\leq 2T} \rangle \leq \frac{C_T n^{O(T)}}{2^N}.$$

For the *high-degree* tail is equally suppressed (Lemma 18). We have

$$\frac{1}{2^N} |\langle \mathbf{1}_\Pi, p^{> 2T} \rangle| \leq \frac{\sqrt{|\Pi|}}{2^N}.$$

Consequently, $p = p^{\leq 2T} + p^{> 2T}$ yields the harmonic bound in Theorem 19:

$$P_\Pi(\beta, \gamma) \leq \frac{|\Pi|}{2^N}.$$

The angle averaging version of the result follows from an averaging over the phase field $\phi_\gamma(y)$ if the spectrum of H_C is rescaled to lie on a lattice (after standard rescaling). This arises because then $\mathbb{E}_\gamma[\phi_\gamma(y)\overline{\phi_\gamma(y')}] = \mathbf{1}_{\{C(y)=C(y')\}}$ leading cross-terms to cancel and

$$\mathbb{E}_\gamma[P_\Pi(\beta, \gamma)] = \frac{|\Pi|}{2^N}.$$

The third bound is derived the fourth-moment (hypercontractivity) analysis of the probability profile. As outlined in Sec. 3.6, let $a(x) = a_{\beta, \gamma}(x)$ and recall $P_\Pi = \sum_{x \in \Pi} |a(x)|^2$. By Cauchy–Schwarz,

$$\sum_{x \in \Pi} |a_x|^2 \leq \sqrt{|\Pi|} \left(\sum_x |a_x|^4 \right)^{1/2}. \quad (5)$$

Thus, any bound on $\sum_x |a_x|^4$ yields an upper bound on P_Π .

Averaging over γ (phase field), the one-bit calculation tensorizes to

$$\mathbb{E}_\gamma \left[\sum_x |a(x)|^4 \right] \leq \left(\frac{1}{2} + \frac{1}{4} \sin^2(2\beta) \right)^N \leq \left(\frac{3}{4} \right)^N,$$

since $\max_\beta \left(\frac{1}{2} + \frac{1}{4} \sin^2 2\beta \right) = \frac{3}{4}$ (achieved at $\beta = \pi/4$). Thus, for typical angles we have,

$$P_\Pi(\beta, \gamma) \lesssim \sqrt{|\Pi|} \left(\frac{3}{4} \right)^{N/2} = \underbrace{\frac{|\Pi|}{2^N} \cdot 2^N |\Pi|^{-1/2} \left(\frac{3}{4} \right)^{N/2}}_{\text{exponentially small in } N},$$

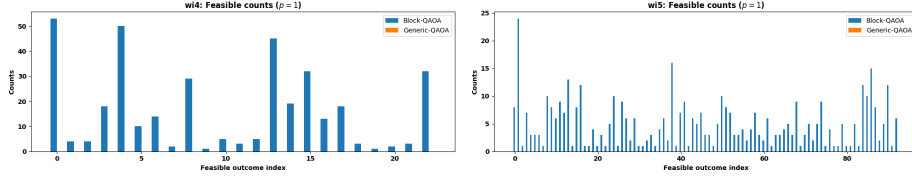
i.e. exponentially close to the uniform baseline.

The light-cone analysis provides some structural intuition for the above analysis. With $U = U_X(\beta)U_C(\gamma)$ and $H_C = \sum_j H_j$ diagonal, r -local, conjugation preserves diagonality and expands support by at most one r -local hyperedge. Each block projector P_B depends only on its radius-1 neighborhood in the interaction hypergraph; the $2n$ one-hot constraints therefore form a bounded-dependency family. Standard cluster/dependency bounds then cap any feasible-mass boost to a polynomial over the uniform product baseline—fully consistent with the harmonic and moment bounds above.

Synthesis. All four routes—harmonic truncation with explicit low-degree suppression (Lemma 16 + Lemmas 17,18), angle averaging via the phase field ϕ_γ , fourth-moment with CS and the $(3/4)^N$ cap, and the light-cone locality barrier agree with Eq. 2 with *typical* angles exponentially close to $|\Pi|/2^N$. This is precisely the $p=1$ envelope used to compare against CE-QAOA, which evades the large $\frac{1}{2^N}$ suppression at any depth.

2.2 The Path to Exponential Enhancements

The *constraint-enhanced* QAOA (CE-QAOA) operates inside a product one-hot subspace and mixes with an all-to-all XY block mixer. Our main conclusion is that it achieves an



(a) **wi4**, $n=4$ cities (16 qubits). Blue bars: Block-QAOA ($p=1$) with (γ, β) taken from colour code as (a). Again, generic QAOA the 10×10 grid. Orange bars (invisible) produced no feasible outputs, while the constrained ansatz covers the full set of $5! = 120$ every feasible bit-string occurred zero times. (b) **wi5**, $n=5$ cities (25 qubits). Same Block-QAOA ($p=1$) with (γ, β) taken from colour code as (a). Again, generic QAOA the 10×10 grid. Orange bars (invisible) produced no feasible outputs, while the constrained ansatz covers the full set of $5! = 120$ every feasible bit-string occurred zero times. tours; the best tour appeared 24 times.

Figure 1: **Constraint-Enhanced QAOA dominates generic QAOA at depth $p=1$.** For medium-size TSP instances from QOptLib[28], the unconstrained ansatz spreads its amplitude almost entirely outside the permutation subspace ($n!/2^{n^2} \rightarrow 0$), whereas the block-encoded circuit concentrates an $\Omega(n^{-m})$ fraction of its probability on valid tours.

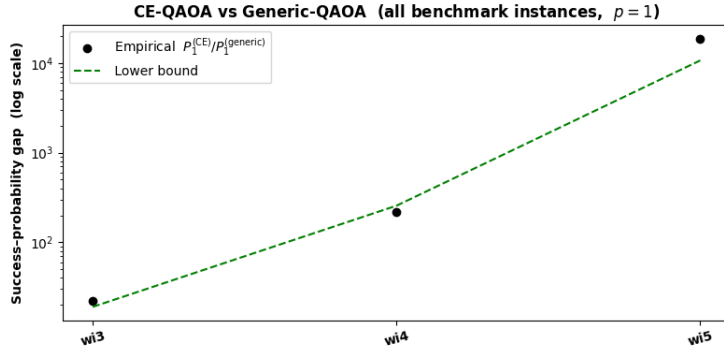


Figure 2: Empirical versus theoretical feasibility mass after parameter optimization on noiseless simulations at depth $p = 1$ for the three QOptLib benchmark instances wi3, wi4, and wi5 [28]. Each instance was executed with 500,000 shots so that Generic QAOA returns at least one feasible bit-string

$\exp[\Theta(n^2)]$ improvement over generic QAOA at the same depth. i.e.

$$\frac{P_p^{(\text{CE})}}{P_p^{(\text{generic})}}.$$

In order to make statements regarding the ratio $\frac{P_p^{(\text{CE})}}{P_p^{(\text{generic})}}$, we need some robust lower envelope for CE-QAOA which we can compare to the upper envelope for generic QAOA outlined in Sec 2.1. The necessary lower envelope comes from a basic symmetry of the product one-hot construction in Sec. 1.2.2. For an instance in the kernel with m blocks of local dimension n , the initial state is the product of uniform one-hot states

$$|s_0\rangle = |s_{\text{blk}}\rangle^{\otimes m}, \quad |s_{\text{blk}}\rangle = \frac{1}{\sqrt{n}} \sum_{j=1}^n |e_j\rangle.$$

The mixer is block-local, excitation-preserving, and normalized to constant block norm:

$$\tilde{H}_{XY}^{(b)} = \frac{1}{n-1} \sum_{1 \leq a < b \leq n} (X_a X_b + Y_a Y_b), \quad U_M(\beta) = \bigotimes_{b=1}^m \exp(-i\beta \tilde{H}_{XY}^{(b)}).$$

With the cost $H_C = H_{\text{pen}} + H_{\text{obj}}$ (diagonal in the computational basis). A depth- p CE-QAOA layer stack is given as

$$|\psi_p(\vec{\gamma}, \vec{\beta})\rangle = \left(\prod_{\ell=1}^p U_M(\beta_\ell) e^{-i\gamma_\ell H_C} \right) |s_0\rangle, \quad \vec{\gamma} = (\gamma_1, \dots, \gamma_p), \quad \vec{\beta} = (\beta_1, \dots, \beta_p).$$

Because U_M preserves the one-hot sector and H_C is diagonal, each layer maps the encoded manifold $\mathcal{H}_{\text{OH}} = (\mathcal{H}_1)^{\otimes m}$ to itself. The kernel's pattern symmetry (block permutations S_m and global symbol relabelings S_n) permits a *block-permutation twirl*. Averaging any fixed circuit U over such relabelings yields a robust lower bound on the probability profile of the prepared state.

For $p = 1$, for each block $b \in [m]$, let $\mathbf{P}_b$ be a uniformly random permutation matrix drawn from S_n acting on $\mathcal{H}_1$. Define the blockwise permutation $\mathbf{P} := \bigotimes_{b=1}^m \mathbf{P}_b \in S_n^m$. Let U be any fixed unitary on $\mathcal{H}_{\text{OH}}$, and $|\psi_0\rangle \in \mathcal{H}_{\text{OH}}$ any fixed state. Let $|x^*\rangle \in \mathcal{H}_{\text{OH}}$ be any fixed product basis vector and $D = n^n$ be the size of the product one-hot manifold explored by CE-QAOA. The following Lemmas hold.

Lemma 2 (One-block twirl). *Fix $j \in [n]$. Then*

$$\mathbb{E}_{\mathbf{P}_b \sim S_n} [\mathbf{P}_b |e_j\rangle\langle e_j| \mathbf{P}_b^\dagger] = \frac{I_n}{n}. \quad (6)$$

Proof. Uniform conjugation by permutation matrices averages the rank-1 projector $|e_j\rangle\langle e_j|$ over all basis labels:

$$\mathbb{E}_{\mathbf{P}_b} [\mathbf{P}_b |e_j\rangle\langle e_j| \mathbf{P}_b^\dagger] = \frac{1}{n} \sum_{t=1}^n |e_t\rangle\langle e_t| = \frac{I_n}{n}.$$

This uses that for each t there are exactly $(n-1)!$ permutations sending $j \mapsto t$, and we normalize by $n!$. $\square$

Lemma 3 (Blockwise (tensor) twirl to I_D/D). *Let $|x^*\rangle = \bigotimes_{b=1}^m |e_{j_b^*}\rangle$. Then*

$$\mathbb{E}_{\mathbf{P} = \bigotimes_b \mathbf{P}_b} [\mathbf{P} |x^*\rangle\langle x^*| \mathbf{P}^\dagger] = \bigotimes_{b=1}^m \frac{I_n}{n} = \frac{I_D}{D}. \quad (7)$$

Proof. Independence across blocks and Lemma 2 give

$$\mathbb{E}_{\{\mathbf{P}_b\}} \left[(\mathbf{P}_1 |e_{j_1^*}\rangle\langle e_{j_1^*}| \mathbf{P}_1^\dagger) \otimes \cdots \otimes (\mathbf{P}_m |e_{j_m^*}\rangle\langle e_{j_m^*}| \mathbf{P}_m^\dagger) \right] = \bigotimes_{b=1}^m \frac{I_n}{n} = \frac{I_D}{D}. \quad \square$$

Lemma 4 (Conjugation invariance). *For any fixed unitary U on $\mathcal{H}$,*

$$\mathbb{E}_{\mathbf{P}} [U^\dagger \mathbf{P} |x^*\rangle\langle x^*| \mathbf{P}^\dagger U] = U^\dagger \left(\frac{I_D}{D} \right) U = \frac{I_D}{D}. \quad (8)$$

Proof. Apply Lemma 3 and use $U^\dagger I_D U = I_D$. $\square$

Lemma 5 (Average success probability under blockwise twirl). *For any state $|\phi\rangle \in \mathcal{H}$,*

$$\mathbb{E}_{\mathbf{P}} [|\langle x^* | \mathbf{P}^\dagger U | \phi \rangle|^2] = \frac{1}{D}. \quad (9)$$

Proof. Write

$$|\langle x^* | \mathbf{P}^\dagger U | \phi \rangle|^2 = \langle \phi | U^\dagger \mathbf{P} |x^*\rangle\langle x^*| \mathbf{P}^\dagger U | \phi \rangle.$$

Taking expectation and using (8),

$$\mathbb{E}_{\mathbf{P}} [\cdots] = \langle \phi | \mathbb{E}_{\mathbf{P}} [U^\dagger \mathbf{P} |x^*\rangle\langle x^*| \mathbf{P}^\dagger U] | \phi \rangle = \langle \phi | (I_D/D) | \phi \rangle = \frac{1}{D}. \quad \square$$

Proposition 6 (Existence of a blockwise relabeling achieving the baseline). *Fix any unitary U on $\mathcal{H}_{\text{OH}}$ and any state $|\phi\rangle \in \mathcal{H}_{\text{OH}}$. Then for any $|x^*\rangle \in \mathcal{H}_{\text{OH}}$ there exists a blockwise permutation $\mathbf{P}^* = \bigotimes_b \mathbf{P}_b^*$ such that*

$$|\langle x^* | \mathbf{P}^{*\dagger} U | \phi \rangle|^2 \geq \frac{1}{D}. \quad (10)$$

In particular, if $U = U_M(\beta)U_C(\gamma)$ with γ, β chosen from any coarse grid (e.g. $\mathcal{G}_{n+1} = \{j\pi/n\}_{j=0}^n$), the same conclusion holds.

Proof. By Lemma 5, the average of the nonnegative random variable $Z(\mathbf{P}) := |\langle x^* | \mathbf{P}^\dagger U | \phi \rangle|^2$ over uniformly random $\mathbf{P}$ equals $1/D$. Hence some realization must satisfy $Z(\mathbf{P}^*) \geq 1/D$ (otherwise the average would be $< 1/D$). The grid choice for γ, β only fixes U and does not change the argument. $\square$

From the foregoing, we can therefore conclude that for every target basis state $x^* \in \mathcal{H}_{\text{OH}}$ (And thus for every feasible basis vector) there exist angles $(\gamma^*, \beta^*) \in [0, \pi]^{n+1}$ such that

$$|\langle x^* | U_M(\beta^*)U_C(\gamma^*)\psi_0 \rangle|^2 \geq \frac{1}{n^n}. \quad (11)$$

This lower bound is enough to prove exponential enhancements advertised in Eq 1. We organise the results in the following Theorem.

Theorem 7 (Single Layer Exponential Enhancement). *Following the mixer and pattern symmetry in Def 1, expectations and tail probabilities with respect to a mixer/twirl distribution satisfy equation 11 so that $P_1^{\text{CE}} \geq \frac{1}{n^n}$. For the generic QAOA, the feasible mass satisfies*

$$\sup_{\beta, \gamma} P_1^{(\text{generic})}(\beta, \gamma) \leq \frac{n!}{2^{n^2}} \cdot n^{O(1)}.$$

Consequently,

$$\frac{P_1^{(\text{CE})}}{P_1^{(\text{generic})}} \geq \exp(n^2 \ln 2 - O(n)) = \exp(\Theta(n^2)).$$

Proof.

If with constant probability over the twirl/mixer $P_1^{(\text{CE})} \geq 1/n^n$, then $P_1^{(\text{CE})} \geq \exp(-n \ln n)$. Since $\ln(\sup_{\beta, \gamma} P_1^{(\text{generic})}) \leq -n^2 \ln 2 + n \ln n + O(n)$, $P^{(\text{generic})} \leq \exp(-n^2 \ln 2 + n \ln n + O(n))$

Then the following ratio follows:

$$\frac{P_1^{(\text{CE})}}{P_1^{(\text{generic})}} \geq \exp(n^2 \ln 2 - 2n \ln n - O(n)) = \exp(\Theta(n^2)).$$

$\square$

This supports the heuristic that to go beyond the near-uniform baseline on permutation-constrained problems one must *build the constraints into the ansatz* (e.g., mixers that act within the target manifold). This is precisely the role of the mixer and pattern symmetry in the CE-QAOA Kernel which uses the problem's constraint structure as a resource and consequently yields a feasible-mass that sharply separates it from the generic ansatz.

2.3 Exponential Separation at Linear $-p$

The light cone analysis can be extended to QAOA circuits beyond the single layer up to the near-linear window $p = \Theta(\frac{n}{\ln n})$, making the exponential feasibility separation *depth-robust*. For commuting, r -local diagonal costs, the Heisenberg light cone at depth p is confined to the radius- p neighborhood in the interaction hypergraph. Projectors enforcing one-hot per row therefore depend on only $O(p)$ (or $O(\Delta^p)$) degrees of freedom. Where Δ^p upper-bounds the size of the radius- p neighborhood of a vertex in a degree- Δ interaction hypergraph induced by the light-cone growth after p layers. Using a special case of Finner/Hölder inequality on dependency graphs[37] developed in Sec. 3.8, we obtain the following theorem.

Theorem 8 (Exponential separation at constant/linear depth). *Let $N = n^2$ and $p = \alpha n$. Consider p -layers of the generic QAOA with X mixer and diagonal H_C .*

(A) **1D rows, nearest-row coupling.** *For any fixed $\alpha \in (0, 1)$,*

$$P_p^{(\text{gen})} \leq \left[(2p+1) 2^{-(n-1)} \right]^n = \exp(-\Theta(n^2)).$$

(B) **Arbitrary intra-row degree Δ_{row} .** *In general,*

$$P_p^{(\text{gen})} \leq \left[C \Delta_{\text{row}}^p 2^{-(n-1)} \right]^n = \exp\left(-n^2 [\ln 2 - \alpha \ln \Delta_{\text{row}}] + O(n \ln n)\right). \quad (12)$$

Hence an $\exp[\Theta(n^2)]$ separation holds for all $\alpha < \ln 2 / \ln \Delta_{\text{row}}$ (and for all $\alpha < 1$ in case (A)).

Combining either (A) or (B) with the CE-QAOA lower bound $P_1^{(\text{CE})} = \Omega(n^{-n})$ yields

$$\frac{P_1^{(\text{CE})}}{P_p^{(\text{gen})}} = \exp[\Theta(n^2)] \quad \text{for } p = \alpha n$$

Namely, a single layer of CE-QAOA can outperform p layers generic QAOA if p is (almost) linear in the problem size.

2.3.1 Proof of Theorem 8

Setup. Recall that the baseline uses the transverse-field mixer $U_X(\beta) = \prod_i e^{-i\beta X_i}$ and a diagonal cost $U_C(\gamma) = e^{-i\gamma H_C}$; thus light-cone growth is caused only by $U_C(\gamma)$. Let Π be the set of feasible bitstrings. For a fixed row (block) of length n , denote by E_i the event “row i is one-hot.” By the locality-parametrized light-cone lemma proved in the appendix (Lemma 27), in the 1D intra-row chain with nearest-row coupling one has

$$\Pr[\text{row valid}] \leq (2p+1) 2^{-(n-1)}.$$

More generally, for bounded intra-row degree Δ_{row} , Lemma 27 gives $\Pr[\text{row valid}] \leq C \Delta_{\text{row}}^p 2^{-(n-1)}$ for a constant C depending only on the degree bounds.

Our light-cone lemmas are defined for the row-induced interaction graph, counting only count intra-row couplings which is usually slightly less than the full maximum degree. Informally, the Heisenberg support of a row one-hot projector after depth p is confined to the radius- p neighborhood in the interaction hypergraph; the number of admissible completions inside this window scales as $W_{\text{row}}(p) \leq 2p+1$ in 1D and $W_{\text{row}}(p) = O(\Delta_{\text{row}}^p)$ in general, yielding the stated per-row probabilities which we now use in our proof.

Sketch of Proof of Theorem 8. For (A) and (B), insert the row-wise bounds from Lemma 27 into the row events $E_i = \{\text{row } i \text{ is one-hot}\}$ and apply the chaining/dependency-graph argument detailed in Sec. 3.8, Thm. 30. In particular, the bounded

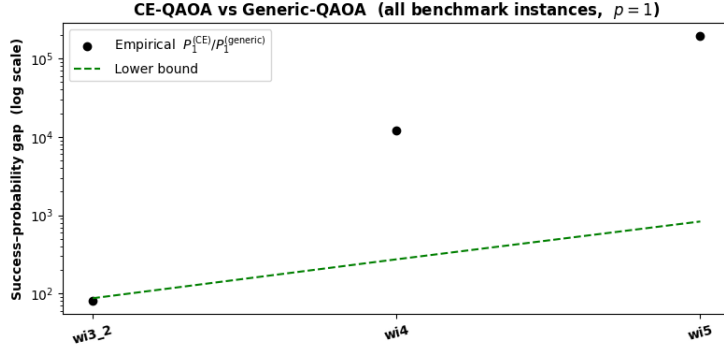


Figure 3: **Empirical versus theoretical feasibility gap with parameter transfer.** Noiseless simulations at depth $p = 1$ for the three QOPLIB benchmarks[28] wi3, wi4, and wi5, each executed with **500 000 shots** so that Generic QAOA returns at least one feasible bit-string. Black dots: measured ratios $P_1^{(\text{block})}/P_1^{(\text{generic})}$; green dashed line: analytic lower bound $2^{n^2}/n^n$ from Theorem 9 with $n = 3, 4, 5$. Every empirical point lies on or above the bound, visually confirming the predicted exponential advantage of parameter transfer from Generic to Constraint-Enhanced QAOA

light cone implies a bounded overlap number for the family $\{E_i\}$, and the resulting dependency contributes at most a polynomial slack absorbed into the $O(n \ln n)$ term in the exponent yielding the displayed exponents. $\square$

2.4 Parameter-Transfer Amplification in Permutation Problems

Orthogonal to the moment bounds, we prove a *parameter-transfer amplification* result in which any one-layer angle pair (β^*, γ^*) that maximises *generic* QAOA boosts feasible-set weight by at least $2^{n^2}/n^n$ when reused *verbatim* in the constrained subspace, without retuning. This suggests that generic-QAOA angle sets already tabulated for many NP-hard problems[41] constitute high-quality *warm starts* for Constraint-Enhanced QAOA, guaranteeing at least the exponential gain *without any additional optimisation* (cf. Fig.3).

Theorem 9 (Parameter-transfer amplification). *Let $N = n^2$, $d_{\text{full}} = 2^N$, $d_{\text{OH}} = n^n$ and $\rho_n := d_{\text{OH}}/d_{\text{full}} = n^n/2^{n^2}$. Let (β^*, γ^*) be any angles that achieve $P_{\max}^{(\text{gen})} = \max_{\beta, \gamma} \langle \psi_{\text{gen}}(\beta, \gamma) | P_{\Pi} | \psi_{\text{gen}}(\beta, \gamma) \rangle$, and define $P^{(\text{CE})} = \langle \psi_{\text{CE}}(\beta^*, \gamma^*) | P_{\Pi} | \psi_{\text{CE}}(\beta^*, \gamma^*) \rangle$. Then for every $n \geq 2$,*

$$P^{(\text{CE})} \geq \rho_n^{-1} P_{\max}^{(\text{gen})} = \frac{2^{n^2}}{n^n} P_{\max}^{(\text{gen})} \geq \sqrt{2\pi n} e^n P_{\max}^{(\text{gen})}.$$

In particular, reusing the same angles inside the block-one-hot subspace boosts the feasible success probability by at least an exponential factor in n .

2.4.1 Proof of Theorem 9

Let $\Pi \subset \{0, 1\}^N$ be the set of *valid permutation matrices*, write $P_{\Pi} := \sum_{y \in \Pi} |y\rangle\langle y|$. Define

$$P_{\max}^{(\text{gen})} := \max_{\beta, \gamma} \langle \psi_{\text{gen}}(\beta, \gamma) | P_{\Pi} | \psi_{\text{gen}}(\beta, \gamma) \rangle,$$

$$P^{(\text{CE})} := \langle \psi_{\text{CE}}(\beta^*, \gamma^*) | P_{\Pi} | \psi_{\text{CE}}(\beta^*, \gamma^*) \rangle,$$

where (β^*, γ^*) is *any* angle pair that achieves the optimum in the generic circuit.

Proof of Theorem 9. The proof relies on following identities proved in Appendix A.

$$\boxed{[H_C, P_{\text{OH}}] = 0}, \quad \boxed{[H_{XY}, P_{\text{OH}}] = 0}, \quad \boxed{[H_C, P_{\Pi}] = 0},$$

plus the Kadison inequality

$$\boxed{P_{\Pi} [H_{XY}, [H_{XY}, P_{\Pi}]] P_{\Pi} \succeq 0.} \quad (13)$$

Uniform superposition on $\mathcal{H}_{\text{OH}}$:

$$|s\rangle := \frac{1}{\sqrt{|\mathcal{H}_{\text{OH}}|}} \sum_{x \in \mathcal{H}_{\text{OH}}} |x\rangle = \frac{2^{N/2}}{\sqrt{n^n}} P_{\text{OH}} |+\rangle^{\otimes N}, \quad |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle).$$

With CE-QAOA state are given as:

$$|\psi_{\text{CE}}(\beta, \gamma)\rangle := e^{-i\beta H_{XY}} e^{-i\gamma H_C} |s\rangle.$$

Insert the resolution of identity $P_{\text{OH}}^2 = P_{\text{OH}}$ and the overlap $\langle \psi_{\text{CE}} | P_{\text{OH}} | \psi_{\text{gen}} \rangle = \sqrt{\rho_n}$ derived in Appendix A:

$$P^{(\text{CE})} = \langle \psi_{\text{CE}}(\beta^*, \gamma^*) | P_{\Pi} | \psi_{\text{CE}}(\beta^*, \gamma^*) \rangle \quad (14)$$

$$= \rho_n^{-1} \langle s | P_{\text{OH}} e^{+i\gamma^* H_C} e^{+i\beta^* H_{XY}} P_{\Pi} e^{-i\beta^* H_{XY}} e^{-i\gamma^* H_C} P_{\text{OH}} | s \rangle \quad (15)$$

$$\geq \rho_n^{-1} \langle \psi_{\text{gen}} | P_{\text{OH}} P_{\Pi} P_{\text{OH}} | \psi_{\text{gen}} \rangle \quad (\text{Using Eq. 13. See App. A.1.}) \quad (16)$$

$$= \rho_n^{-1} \langle \psi_{\text{gen}} | P_{\Pi} | \psi_{\text{gen}} \rangle \quad (\text{because } P_{\Pi} \subseteq P_{\text{OH}}) \quad (17)$$

$$= \rho_n^{-1} P_{\text{max}}^{(\text{gen})}, \quad (18)$$

yielding the first inequality. For the numerical lower bound note that $\rho_n \leq e^{-n}/\sqrt{2\pi n}$ by Stirling's approximation $n! \sim \sqrt{2\pi n}(n/e)^n$ combined with $n!/n^n \leq 1$. Equivalently, $1/\rho_n \geq \sqrt{2\pi n} e^n$, which inserted into (18) gives the second inequality. $\square$

In practice a brief local re-optimisation could further improve success probability. Because the Block mixer H_{XY} acts *only within* the one-hot subspace, its spectral radius differs from $\sum_j X_j$. A local re-optimisation of still yields some measurable improvements (cf. Fig.3).

2.5 Discussion

Remark 10 (Depth monotonicity of CE-QAOA feasible mass). Throughout, let $P_p^{(\text{CE})}$ denote the *angle-optimised* feasible mass of depth- p CE-QAOA, i.e. the maximum probability of sampling a feasible string over all choices of $(\beta_1, \gamma_1, \dots, \beta_p, \gamma_p)$. By construction, any depth- $(p+1)$ circuit can reproduce a depth- p circuit by setting the final angles to zero, so the sequence is depth-monotone:

$$P_{p+1}^{(\text{CE})} \geq P_p^{(\text{CE})} \quad \text{for all } p \geq 1.$$

In particular, the one-layer lower bound $P_1^{(\text{CE})} = \Omega(n^{-n})$ implied by Theorem 13 (Existence of Uniform Overlap Lower Bound via Block Permutations) propagates to all deeper layers:

$$P_p^{(\text{CE})} \geq P_1^{(\text{CE})} = \Omega(n^{-n}) \quad \text{for every } p \geq 1.$$

Combining this with the generic upper bound $P_p^{(\text{gen})} \leq \exp(-\Theta(n^2))$ from Theorem 8 yields an $\exp[\Theta(n^2)]$ feasible-mass separation between CE-QAOA and generic QAOA *uniformly in* p throughout the depth window where the light-cone/degree analysis constrains generic QAOA. Further sharpening of how $P_p^{(\text{CE})}$ itself grows with p is left for future work.

Relation to known constant-depth limitations and OGP barriers. The above locality mechanism is consistent with several independent limitations: (i) constant-depth circuits exhibit a bounded correlation radius/light-cone [12] and hence cannot *see* the whole instance structure at level $p=1$ in the sense formalized for QAOA [13, 17]; (ii) for random sparse graphs and related ensembles, the solution space exhibits the *Overlap Gap Property* (OGP), which obstructs variational optimization at fixed depth [16, 15, 13, 8]. Formally, prior work shows (a) constant-level QAOA cannot surpass certain approximation thresholds on large sparse graphs [14, 8], and (b) under OGP/NLTS-type structure, low-depth circuits fail to reach near-optimal solutions with nonnegligible probability [15, 16].

Row-level degree versus global degree. It will be important to distinguish the global hypergraph degree Δ of the N -qubit interaction graph from the *row-level* degree Δ_{row} . By definition, Δ_{row} is the maximum number of cost terms H_a whose support intersects a fixed row block

$$\text{Row}_i = \{(i, 1), \dots, (i, n)\},$$

i.e., the maximum degree of the interaction graph induced on a single block of n qubits. In particular, Δ_{row} depends only on the pattern of *intra-row* couplings and is independent of the total number of rows. In our light-cone analysis, the Heisenberg evolution of a row one-hot projector is confined to a neighborhood whose size grows at most like $O(\Delta_{\text{row}}^p)$, so the number of admissible completions and hence the per-row validity probability are controlled solely by Δ_{row} (as made precise in Lemma 27), rather than by any global degree parameter of the full N -qubit constraint graph. This row-level degree is therefore the relevant locality parameter that ultimately bounds the joint feasibility probability.

Non-asymptotic degree dependence in Eq. (12). Let $\Delta_{\text{row}}(n)$ denote the maximum intra-row degree of the interaction hypergraph induced by $H_C = \sum_a H_a(Z)$ (each H_a diagonal, commuting, and r -local). Then the row-validity lemma implies

$$\mathbb{E}[E_i] \leq C \Delta_{\text{row}}(n)^p 2^{-(n-1)},$$

and the dependency-graph Hölder bound with overlap number $\chi \asymp \Delta_{\text{row}}(n)^{p^1}$ yields

$$P_p^{(\text{gen})} \leq \left[C \Delta_{\text{row}}(n)^p 2^{-(n-1)} \right]^n = \exp\left(-n^2 \left[\ln 2 - \frac{p}{n} \ln \Delta_{\text{row}}(n) \right] + O(n \ln n)\right), \quad (19)$$

where the $O(n \ln n)$ term collects all polynomial prefactors in n coming from the row-validity constant C , the overlap number χ , and the combinatorial factors in the dependency-graph argument.

For the CE-QAOA kernel at one layer, we explicitly assume that the optimized angles beat the “flat” baseline on the one-hot manifold. More precisely, for n blocks of size n we use

$$P_1^{(\text{CE})} = |\langle x^* | \mathbf{P}^{\dagger} U_M(\beta) U_C(\gamma) s_0 \rangle|^2 \geq \frac{c_{\text{CE}}}{n^n}, \quad c_{\text{CE}} > 1, \quad (20)$$

where $c_{\text{CE}} = 1$ would correspond to the “average” case (uniform weight $1/n^n$ on each of the n^n points in the one-hot block manifold), and the optimization over (β, γ) produces a genuine constant-factor gain $c_{\text{CE}} > 1$.

¹i.e. χ has the same order of growth as $\Delta_{\text{row}}(n)^p$ up to constant factors. More precisely, writing

$$\chi \asymp \Delta_{\text{row}}(n)^p$$

means that there exist constants $c, C > 0$ (independent of n and p in the regime of interest) such that, for all sufficiently large n ,

$$c \Delta_{\text{row}}(n)^p \leq \chi \leq C \Delta_{\text{row}}(n)^p.$$

Combining (19) and (20), for a depth $p = \alpha n$ we obtain the lower bound

$$\begin{aligned} \log \frac{P_1^{(\text{CE})}}{P_p^{(\text{gen})}} &\geq \log c_{\text{CE}} - n \ln n + n^2 [\ln 2 - \alpha \ln \Delta_{\text{row}}(n)] - O(n \ln n) \\ &= n^2 [\ln 2 - \alpha \ln \Delta_{\text{row}}(n)] - O(n \ln n) + \log c_{\text{CE}}. \end{aligned} \quad (21)$$

Two scales are visible: an n^2 -scale governed by the *degree–depth combination* $\ln 2 - \alpha \ln \Delta_{\text{row}}(n)$, and a subleading $n \ln n$ scale coming from the n^{-n} factor and the cluster combinatorics. The constant $c_{\text{CE}} > 1$ contributes a strictly positive $\log c_{\text{CE}}$ in (21).

Define the depth–fraction threshold

$$\alpha_*(n) := \frac{\ln 2}{\ln \Delta_{\text{row}}(n)} \quad (\text{with the convention } \alpha_*(n) = +\infty \text{ if } \Delta_{\text{row}}(n) = 1).$$

On the n^2 -scale, the sign of the advantage exponent is set by $\ln 2 - \alpha \ln \Delta_{\text{row}}(n)$: for any fixed $\alpha < \alpha_*(n)$ we have

$$\ln 2 - \alpha \ln \Delta_{\text{row}}(n) \geq \eta > 0$$

for some constant η (independent of n in the regime of interest), so (21) gives

$$\log \frac{P_1^{(\text{CE})}}{P_p^{(\text{gen})}} \geq \eta n^2 - O(n \ln n) + \log c_{\text{CE}} = \Theta(n^2),$$

and hence $\frac{P_1^{(\text{CE})}}{P_p^{(\text{gen})}} = \exp(\Theta(n^2))$. In this asymptotic regime the constant $c_{\text{CE}} > 1$ does not alter the leading threshold $\alpha_*(n)$, but it does push the exponent upwards by a strictly positive amount $\log c_{\text{CE}}$.

If instead one insists on tracking the *exact* boundary where the ratio drops to order 1 (rather than $\exp[\Theta(n^2)]$), one can solve (21) for the minimal positive gap

$$A_n := \ln 2 - \alpha \ln \Delta_{\text{row}}(n)$$

required to keep the advantage nonvanishing:

$$A_n \gtrsim \frac{\ln n}{n} - \frac{\log c_{\text{CE}}}{n^2},$$

where $\gtrsim$ hides constant factors and the $O(n \ln n)$ corrections. Thus for finite n we need α to sit below $\alpha_*(n)$ by an $O((\ln n)/n)$ margin, but the positive term $\log c_{\text{CE}}$ *reduces* this required margin by $\Theta((\log c_{\text{CE}})/(n \ln \Delta_{\text{row}}(n)))$. In other words, every larger constant $c_{\text{CE}} > 1$ slightly *extends* the finite- n degree–depth window over which the super-exponential separation is visible, even though the asymptotic threshold $\alpha_*(n)$ itself is set purely by the row-level degree.

For any polynomial degree, $\Delta_{\text{row}}(n) = n^c$ (any fixed $c > 0$), $\alpha_*(n) = \frac{\ln 2}{c \ln n}$, and any

$$p \leq (1 - \varepsilon) \alpha_*(n) n = (1 - \varepsilon) \frac{\ln 2}{c} \frac{n}{\ln n},$$

with fixed $\varepsilon \in (0, 1)$, gives $A_n \geq \varepsilon \ln 2$ for all large n , hence

$$\log \frac{P_1^{(\text{CE})}}{P_p^{(\text{gen})}} \geq \varepsilon (\ln 2) n^2 - O(n \ln n) + \log c_{\text{CE}} = \Theta(n^2),$$

and the ratio again grows like $\exp(\Theta(n^2))$. On this near-linear depth scale $p = \Theta(n/\ln n)$, the degree parameter c fixes the slope $\alpha_*(n)$, while the constant $c_{\text{CE}} > 1$ shifts the exponent upwards and slightly enlarges the finite- n window of admissible α . In fact, in TSP, $\Delta_{\text{row}} = n - 1$ so that $\alpha_*(n) = \frac{\ln 2}{\ln(n-1)}$ indicating that the exponential advantage regime is not easily wiped away by degree dependent contributions.

The instance dependent parameter c_{CE} also plays the role of extending the exponential advantage regime with the increase in CE-QAOA depth to match the generic case as in

Eq 1. As already remarked, the feasible success probability in CE-QAOA is monotone non-decreasing in the number of layers Rem. 10. Thus, after reoptimizing at any layer, one obtains a depth-dependent constant $c_{\text{CE}}(p_{\text{CE}})$ given by

$$P_{p_{\text{CE}}}^{(\text{CE})} \geq \frac{c_{\text{CE}}(p_{\text{CE}})}{n^n}, \quad c_{\text{CE}}(p_{\text{CE}}) \geq 1.$$

Equivalently,

$$p_{\text{CE},1} \leq p_{\text{CE},2} \implies c_{\text{CE}}(p_{\text{CE},1}) \leq c_{\text{CE}}(p_{\text{CE},2}).$$

In other words, if one tracks the ratio $P_{p_{\text{CE}}}^{(\text{CE})}/P_p^{(\text{gen})}$ in Eq. 1 where p_{CE} matches the p layers in the generic ansatz, the term $\log c_{\text{CE}}(p_{\text{CE}})$ in (21) becomes depth-dependent and positive, shifting the advantage exponent upward. The degree-controlled threshold $\alpha_*(n) = \ln 2 / \ln \Delta_{\text{row}}(n)$ continues to dictate the leading n^2 -scale behaviour, but the growing constant $c_{\text{CE}}(p_{\text{CE}}) > 1$ slightly enlarges the depth window over which the super-exponential separation remains visible. In contrast, the degree dependence push $\Delta_{\text{row}}(n) = n^c$ remains small as we saw for the case of TSP. This favors CE-QAOA as the numerator c_{CE} can be large even at a single layer. Cf. Figs 2 and 3.

3 Materials and Methods

3.1 Walsh-Fourier/Krawtchouk Analysis

We seek to develop a Walsh-Fourier/Krawtchouk analysis that upper bounds the feasible mass of one-layer generic QAOA. Throughout, $N = n^2$, $\Pi \subset \{0, 1\}^N$ is the feasible set of $n \times n$ permutation-matrix decoded from the sampled bitstrings, hence $|\Pi| = n!$ and

$$|\psi_1^{(\text{gen})}(\beta, \gamma)\rangle = U_X(\beta) U_C(\gamma) |+\rangle^{\otimes N}, \quad U_X(\beta) = e^{-i\beta \sum_{i=1}^N X_i}, \quad U_C(\gamma) = e^{-i\gamma H_C}.$$

With H_C diagonal in the computational basis, we can write

$$H_C = \sum_{y \in \{0,1\}^N} C(y) |y\rangle\langle y| \implies U_C(\gamma) = e^{-i\gamma H_C} = \sum_y e^{-i\gamma C(y)} |y\rangle\langle y|.$$

Acting on the uniform superposition gives

$$U_C(\gamma) |+\rangle^{\otimes N} = 2^{-N/2} \sum_y e^{-i\gamma C(y)} |y\rangle.$$

With the phase field $\phi_\gamma(y) := e^{-i\gamma C(y)}$ and any basis vector $|x\rangle$ we obtain

$$a_{\beta, \gamma}(x) = \langle x | U_X(\beta) U_C(\gamma) |+\rangle^{\otimes N} = 2^{-N/2} \sum_y \underbrace{\langle x | U_X(\beta) |y\rangle}_{K_\beta(x, y)} \phi_\gamma(y).$$

Because $U_X(\beta) = \prod_{i=1}^N e^{-i\beta X_i}$, its Z -basis kernel factors as

$$K_\beta(x, y) = \prod_{i=1}^N \langle x_i | e^{-i\beta X_i} |y_i\rangle = (\cos \beta)^{N-d(x, y)} (-i \sin \beta)^{d(x, y)} = (\cos \beta)^N (-i \tan \beta)^{d(x, y)}, \quad (22)$$

where $d(x, y) = |x \oplus y|$ is the Hamming distance. Thus the output amplitudes are

$$a_{\beta, \gamma}(x) = 2^{-N/2} \sum_{y \in \{0,1\}^N} K_\beta(x, y) \phi_\gamma(y). \quad (23)$$

For normalization, note that K_β are matrix elements of the unitary $U_X(\beta)$, and compute

$$\sum_{x \in \{0,1\}^N} |a_{\beta, \gamma}(x)|^2 = 2^{-N} \sum_x \left| \sum_y K_\beta(x, y) \phi_\gamma(y) \right|^2 = 2^{-N} \|U_X(\beta) \phi_\gamma\|_2^2 \quad (24)$$

$$= 2^{-N} \|\phi_\gamma\|_2^2 = 2^{-N} \sum_y |\phi_\gamma(y)|^2 = 2^{-N} \cdot 2^N = 1, \quad (25)$$

because $|\phi_\gamma(y)| \equiv 1$. Thus the amplitudes $a_{\beta,\gamma}(x)$ are correctly normalized.

3.2 Walsh–Fourier primer

We write $[N] = \{1, 2, \dots, N\}$. For a subset $S \subseteq [N]$, let $s \in \{0, 1\}^N$ be its indicator vector, $s_i = \mathbf{1}_{\{i \in S\}}$, and we freely identify S with s . The Walsh characters on the Boolean cube are [29]

$$\chi_S(x) := (-1)^{S \cdot x}, \quad S \subseteq [N].$$

For $x \in \{0, 1\}^N$, define the mod-2 (parity) dot product

$$S \cdot x := \left(\sum_{i=1}^N s_i x_i \right) \bmod 2 \in \{0, 1\}.$$

For $S, T \subseteq [N]$, write $S \oplus T$ for the symmetric difference (bitwise XOR of indicators). Therefore:

$$S \oplus T := (S \setminus T) \cup (T \setminus S),$$

and

$$(S \oplus T) \cdot x \equiv (S \cdot x) + (T \cdot x) \pmod{2} \tag{26}$$

$$\equiv (S \cdot x) \oplus (T \cdot x). \tag{27}$$

So if t is the indicator of T , then $S \oplus T$ corresponds to $s \oplus t$. The group law $\chi_S(x) \chi_T(x) = \chi_{S \oplus T}(x)$ will be used repeatedly.

We use the (normalized) Walsh–Fourier transform is given as

$$\widehat{f}(S) := 2^{-N} \sum_{x \in \{0,1\}^N} f(x) \chi_S(x), \quad f(x) = \sum_{S \subseteq [N]} \widehat{f}(S) \chi_S(x). \tag{28}$$

With this normalization, the hat $\widehat{}$ denotes the Walsh–Fourier transform on the Boolean cube with characters $\chi_S(x) := (-1)^{S \cdot x}$ for $S \subseteq [N]$. Parseval/Plancherel reads [29]

$$\sum_x |f(x)|^2 = 2^N \sum_S |\widehat{f}(S)|^2, \quad 2^{-N} \sum_x \overline{f(x)} g(x) = \sum_S \overline{\widehat{f}(S)} \widehat{g}(S).$$

For functions $f : \{0, 1\}^N \rightarrow \mathbb{C}$ we use the counting inner product $\langle f, g \rangle := \sum_{x \in \{0,1\}^N} \overline{f(x)} g(x)$ (and the normalized version $\langle\langle f, g \rangle\rangle := 2^{-N} \langle f, g \rangle$ when convenient).

One could view K_β from Eq. 22 as a convolution on $\mathbb{Z}_2^N$:

$$(K_\beta f)(x) := \sum_y K_\beta(x, y) f(y) = \sum_z k_\beta(z) f(x \oplus z),$$

with the one-bit kernel $k_\beta(0) = \cos \beta$ and $k_\beta(1) = -i \sin \beta$ and $k_\beta(z) = \prod_{i=1}^N k_\beta(z_i)$.

Lemma 11 (Diagonalization of K_β in the Walsh basis). *For all $S \subseteq [N]$,*

$$(\widehat{K_\beta f})(S) = \lambda_\beta(|S|) \widehat{f}(S), \quad \lambda_\beta(|S|) = e^{-i\beta(N-2|S|)}.$$

In particular, $|\lambda_\beta(|S|)| = 1$, so K_β preserves Walsh magnitudes $|\widehat{f}(S)|$.

Proof sketch. On one bit, the convolution theorem with our normalization gives $\widehat{k * f}(t) = 2 \widehat{k}(t) \widehat{f}(t)$ for $t \in \{0, 1\}$, where $\widehat{k}(t) = 2^{-1} \sum_{z \in \{0,1\}} k(z) (-1)^{tz}$. With $k(0) = \cos \beta$, $k(1) = -i \sin \beta$,

$$2 \widehat{k}(0) = \cos \beta - i \sin \beta = e^{-i\beta}, \quad 2 \widehat{k}(1) = \cos \beta + i \sin \beta = e^{i\beta}.$$

By tensoring over N bits, frequencies factorize. For a mask S with $|S|$ ones, $\lambda_\beta(|S|) = e^{-i\beta}$ on zero-frequencies and $e^{i\beta}$ on one-frequencies, hence $\lambda_\beta(|S|) = e^{-i\beta(N-|S|)} e^{i\beta|S|} = e^{-i\beta(N-2|S|)}$. $\square$

Permutation-feasibility indicator. We write $\mathbf{1}_A : \{0, 1\}^N \rightarrow \{0, 1\}$ for the *indicator* of a set A , i.e.

$$\mathbf{1}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

In the permutation encoding (e.g. for TSP) we fix $N = n^2$ and identify each bitstring $x \in \{0, 1\}^N$ with an $n \times n$ binary matrix

$$X(x) = (x_{i,j})_{1 \leq i,j \leq n},$$

where $x_{i,j} \in \{0, 1\}$ records whether city j is placed at tour position i . We say that x is *permutation-feasible* if and only if $X(x)$ is a permutation matrix, i.e.

$$\sum_{j=1}^n x_{i,j} = 1 \quad \text{for all rows } i \in [n], \quad \sum_{i=1}^n x_{i,j} = 1 \quad \text{for all columns } j \in [n].$$

Let $\Pi \subset \{0, 1\}^N$ denote the set of all such bitstrings, so $|\Pi| = n!$. The permutation-feasibility indicator is

$$\mathbf{1}_\Pi(x) = \begin{cases} 1, & \text{if } x \in \Pi \text{ (i.e. } X(x) \text{ is a permutation matrix),} \\ 0, & \text{otherwise.} \end{cases}$$

Equivalently, if we write

$$R(x) := \prod_{i=1}^n \mathbf{1}\left[\sum_{j=1}^n x_{i,j} = 1\right], \quad C(x) := \prod_{j=1}^n \mathbf{1}\left[\sum_{i=1}^n x_{i,j} = 1\right],$$

for the row- and column-one-hot indicators, then

$$\mathbf{1}_\Pi(x) = R(x)C(x), \quad \|\mathbf{1}_\Pi\|_2^2 = \sum_x \mathbf{1}_\Pi(x) = |\Pi| = n!.$$

With this notation, the feasible probability mass can be written as the Walsh correlation

$$P_\Pi(\beta, \gamma) = 2^{-N} \sum_x \mathbf{1}_\Pi(x) |g(x)|^2.$$

Then

$$P_\Pi(\beta, \gamma) = 2^{-N} \sum_x \mathbf{1}_\Pi(x) |g(x)|^2 = \sum_{S \subseteq [N]} \widehat{\mathbf{1}_\Pi}(S) \widehat{|g|^2}(S).$$

where $g = K_\beta \phi_\gamma$ and $\widehat{\mathbf{1}_\Pi}$ is the Walsh transform of the permutation-feasibility indicator. Writing $b := \widehat{\phi_\gamma}$ and using Lemma 11,

$$\widehat{|g|^2}(S) = \sum_{T \subseteq [N]} \lambda_\beta(|T|) \overline{\lambda_\beta(|T \oplus S|)} b(T) \overline{b(T \oplus S)}, \quad \sum_T |b(T)|^2 = 1,$$

so P_Π is controlled by the Walsh autocorrelations of the phase field and the Fourier profile of the feasible indicator.

3.3 Krawtchouk Polynomials on the Hamming sphere

Let g_n be the “exactly-one” (one-hot) indicator on n bits, and let Π be the intersection of $2n$ block one-hot constraints plus the row-column bijection (permutation) constraint. The Walsh spectrum of g_n is explicit via Krawtchouk polynomials [38, 29]. In particular, it allows us to show via explicit computation that low degrees carry exponentially small mass. To see this, we work on the Boolean cube $\{0, 1\}^n$ with Walsh characters $\chi_S(x) = (-1)^{S \cdot x}$ for $S \subseteq [n]$ and the normalization

$$\widehat{f}(S) := 2^{-n} \sum_{x \in \{0, 1\}^n} f(x) \chi_S(x), \quad f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \chi_S(x)$$

as in Eq. 28.

Let $g_w : \{0, 1\}^n \rightarrow \{0, 1\}$ denote the Hamming-sphere indicator $g_w(x) = \mathbf{1}_{\{|x|=w\}}$. In particular, the one-hot indicator is g_1 (support $\{x : |x| = 1\}$). For $0 \leq w \leq n$, the (unnormalized) binary Krawtchouk polynomials are

$$K_w^{(n)}(r) := \sum_{j=0}^w (-1)^j \binom{r}{j} \binom{n-r}{w-j}, \quad r = 0, 1, \dots, n. \quad (29)$$

They satisfy the following generating function and orthogonality identities respectively [38]

$$\sum_{w=0}^n K_w^{(n)}(r) z^w = (1+z)^{n-r} (1-z)^r, \quad (30)$$

$$\sum_{r=0}^n \binom{n}{r} K_w^{(n)}(r) K_{w'}^{(n)}(r) = 2^n \binom{n}{w} \delta_{w,w'}. \quad (31)$$

Lemma 12 (Krawtchouk transform of a Hamming sphere). *For $S \subseteq [n]$ with $|S| = r$,*

$$\sum_{x: |x|=w} \chi_S(x) = K_w^{(n)}(r).$$

Consequently,

$$\widehat{g_w}(S) = 2^{-n} K_w^{(n)}(|S|). \quad (32)$$

Proof. Partition the coordinates into the r positions where S has 1s and the $n-r$ positions where S has 0s. To form an x with $|x| = w$, choose j ones among the r “active” positions and $w-j$ ones among the remaining $n-r$ positions. There are $\binom{r}{j} \binom{n-r}{w-j}$ such x , and each contributes a phase $(-1)^j$ because $\chi_S(x) = (-1)^{S \cdot x}$ counts parity on the r active positions only. Summing over j gives (29), hence the first claim. Applying the Walsh transform definition to g_w ,

$$\widehat{g_w}(S) = 2^{-n} \sum_x g_w(x) \chi_S(x) = 2^{-n} \sum_{|x|=w} \chi_S(x) = 2^{-n} K_w^{(n)}(|S|).$$

□

$$K_1^{(n)}(r) = \binom{r}{0} \binom{n-r}{1} - \binom{r}{1} \binom{n-r}{0} = (n-r) - r = n - 2r.$$

Thus, for g_1 (the “exactly one” indicator),

$$\widehat{g_1}(S) = 2^{-n} (n - 2|S|), \quad S \subseteq [n], \quad (33)$$

which matches the direct computation $\widehat{g_1}(S) = 2^{-n} \sum_{i=1}^n (-1)^{\mathbf{1}_{\{i \in S\}}} = 2^{-n} (n - 2|S|)$.

Normalization/Parseval sanity check using (32) and (31) gives,

$$\sum_x |g_w(x)|^2 = \binom{n}{w} = 2^n \sum_S |\widehat{g_w}(S)|^2 = 2^n \sum_{r=0}^n \binom{n}{r} (2^{-n} K_w^{(n)}(r))^2,$$

which is exactly (31) with $w' = w$. We also have $\|g_1\|_2^2 = \binom{n}{1} = n$.

Lemma 13 (Low-degree Fourier weight for g_1). *For any fixed degree cutoff $d \geq 0$, where $d \in \mathbb{N}$ denotes a Fourier degree cutoff*

$$\sum_{|S| \leq d} |\widehat{g_1}(S)|^2 = 2^{-2n} \sum_{r=0}^d \binom{n}{r} (n - 2r)^2 \leq 2^{-2n} C_d n^{d+2},$$

for some constant C_d depending only on d (e.g. $C_d = 2(d+1)$ for all $n \geq 2$). Equivalently, the degree- $\leq d$ weight is at most $C_d n^{d+1} 2^{-n}$ of the total $\|g_1\|_2^2 = n$.

Proof. Insert (33) and group by $r = |S|$:

$$\sum_{|S| \leq d} |\widehat{g}_1(S)|^2 = 2^{-2n} \sum_{r=0}^d \binom{n}{r} (n-2r)^2.$$

For $r \leq d$ we have $\binom{n}{r} \leq n^r$ and $(n-2r)^2 \leq n^2$, hence

$$\sum_{r=0}^d \binom{n}{r} (n-2r)^2 \leq n^2 \sum_{r=0}^d n^r = n^2 \frac{n^{d+1} - 1}{n - 1} \leq 2(d+1) n^{d+1} \quad (n \geq 2).$$

Multiplying by 2^{-2n} yields the claimed bound with $C_d = 2(d+1)$. $\square$

The Krawtchouk identity $\widehat{g}_w(S) = 2^{-n} K_w^{(n)}(|S|)$ shows that the Walsh spectrum of any Hamming-sphere indicator is a radial (shell-dependent) polynomial of $|S|$. For the one-hot layer ($w = 1$), $\widehat{g}_1(S)$ is exactly linear in $|S|$, and its low-degree weight decays exponentially in n after summation over shells, as quantified in Lemma 13. Tensorizing across $m = 2n$ blocks and incorporating the bijection constraint yields an additional sparse factor that does not increase low-degree mass beyond a polynomial.

3.4 Route 1: Harmonic Analysis of the probability profile

To go beyond n -bit rows and make the Fourier weight decay on the permutation feasible subspace explicit, we shall use the following technical lemmas. Lemma 14 is an adaptation from Alon and Spencer (Lemma 2, p. 64, [36]) while Lemma 15 is a technical tool to control the resulting inequality.

Lemma 14. *For every integer $n \geq 2$,*

$$((n-1)!)^{n(n+1)} \cdot ((n+1)!)^{n(n-1)} < (n!)^{2(n^2-1)},$$

equivalently,

$$\left(\frac{n^n}{n!}\right)^2 > \left(\frac{n+1}{n}\right)^{n(n-1)}.$$

Proof. The two forms are equivalent by dividing both sides of the first inequality by $((n-1)!(n+1)!)^{n(n-1)}$ and simplifying.

Base case $n = 2$. We have

$$\left(\frac{2^2}{2!}\right)^2 = \left(\frac{4}{2}\right)^2 = 4 \quad \text{and} \quad \left(\frac{3}{2}\right)^2 = \frac{9}{4},$$

so the inequality holds.

Case $n \geq 3$. By Arithmetic Mean–Geometric Mean inequality (AM–GM),

$$n! = \prod_{k=1}^n k \leq \left(\frac{1 + \dots + n}{n}\right)^n = \left(\frac{n+1}{2}\right)^n.$$

Hence

$$\left(\frac{n^n}{n!}\right)^2 \geq \left(\frac{n^n}{(\frac{n+1}{2})^n}\right)^2 = \left(\frac{2n}{n+1}\right)^{2n}.$$

Using $(1 + \frac{1}{n})^n < e$ we get

$$\frac{2n}{n+1} = \frac{2}{1 + \frac{1}{n}} > 2e^{-1/n}, \quad \Rightarrow \quad \left(\frac{2n}{n+1}\right)^{2n} > 4^n e^{-2}.$$

For $n \geq 3$ we have $4^n > e^{n+1}$, hence $4^n e^{-2} > e^{n-1}$. Therefore

$$\left(\frac{n^n}{n!}\right)^2 > e^{n-1}.$$

On the other hand, $(1 + \frac{1}{n})^n < e$ implies

$$\left(\frac{n+1}{n}\right)^{n(n-1)} = ((1 + \frac{1}{n})^n)^{n-1} < e^{n-1}.$$

Combining the last two displays yields

$$\left(\frac{n^n}{n!}\right)^2 > \left(\frac{n+1}{n}\right)^{n(n-1)}.$$

□

Lemma 15. *Let $c_T > 0$ and set $a := \frac{c_T \ln 2}{2}$. Then for all*

$$n \geq N(c_T) := \max\left\{9, \left\lceil \frac{4}{a^2} \right\rceil\right\}$$

we have $n^n 2^{-c_T n^2/2} < 1$.

Proof. For $n \geq 9$ we have $\ln n < \sqrt{n}$, and for $n \geq 4/a^2$ we have $\sqrt{n} \leq an$. Hence $\ln n < an$, so

$$n \ln n - an^2 < 0,$$

which implies $n^n 2^{-c_T n^2/2} = \exp(n \ln n - an^2) < 1$. □

Lemma 16 (Low-degree mass of $\mathbf{1}_\Pi$). *Let $N = n^2$ and $\Pi \subset \{0, 1\}^N$ be the set of $n \times n$ permutation matrices (row- and column-one-hot). For every fixed degree cutoff $d \in \mathbb{N}$ there exist constants $c_d, C_d > 0$ such that*

$$\sum_{|S| \leq d} |\widehat{\mathbf{1}_\Pi}(S)|^2 \leq C_d n^{O(d)} 2^{-cn^2} |\Pi|^2 R_n; \quad \|\mathbf{1}_\Pi\|_2^2 = |\Pi| = n! \text{ and } R_n > \left(\frac{n+1}{n}\right)^{n(n-1)}$$

Proof. Write the indicator of row one-hot as

$$R(x) = \prod_{i=1}^n g_1^{(i)}(x_{i,1}, \dots, x_{i,n}),$$

where each $g_1^{(i)}$ is the (one-bit-weight) indicator on the i -th row block (size n). Similarly, the column one-hot indicator factors as

$$C(x) = \prod_{j=1}^n g_1^{(j)}(x_{1,j}, \dots, x_{n,j}).$$

Then $\mathbf{1}_\Pi(x) = R(x)C(x)$ and, in Walsh-Fourier space,

$$\widehat{\mathbf{1}_\Pi} = \widehat{R} \widehat{C} = \widehat{R} * \widehat{C},$$

where $*$ is convolution on the frequency group $(\{0, 1\}^N, \oplus)$: $(f * g)(S) = \sum_T f(T) g(S \oplus T)$.

Step 1: Tensorization across blocks (row side). By block disjointness of the rows and the factorization of characters, the Walsh transform *factorizes*: for $S \subseteq [N]$ and $S_i := S \cap \{\text{positions of row } i\}$,

$$\widehat{R}(S) = \prod_{i=1}^n \widehat{g}_1(S_i).$$

By Lemma 13 (the Krawtchouk transform with $w = 1$), $\widehat{g}_1(T) = 2^{-n} (n - 2|T|)$ for any $T \subseteq [n]$. Note that since a single n -bit one-hot block g_1 has $\widehat{g}_1(T) = 2^{-n} (n - 2|T|)$. Tensoring across n disjoint rows gives $\|\widehat{R}\|_2^2 = 2^{-n^2} n^n$, and similarly for columns $\|\widehat{C}\|_2^2 = 2^{-n^2} n^n$. Hence

$$\widehat{R}(S) = 2^{-n^2} \prod_{i=1}^n (n - 2|S_i|), \quad \|\widehat{R}\|_2^2 = 2^{-N} \|R\|_2^2 = 2^{-N} n^n. \quad (34)$$

An analogous factorization holds for C , partitioning S by columns: $\widehat{C}(S) = 2^{-n^2} \prod_{j=1}^n (n - 2|S^{(j)}|)$ and $\|\widehat{C}\|_2^2 = 2^{-N} n^n$.

Step 2: Convolution bound onto a low-degree window. Let $E_d := \{S \subseteq [N] : |S| \leq d\}$ denote the degree- $\leq d$ frequency window. For any S , $(\widehat{R} * \widehat{C})(S) = \sum_T \widehat{R}(T) \widehat{C}(S \oplus T)$. By Cauchy-Schwarz in the T -sum and then summing over $S \in E_d$,

$$\begin{aligned} \sum_{S \in E_d} |(\widehat{R} * \widehat{C})(S)|^2 &\leq \sum_{S \in E_d} \left(\sum_T |\widehat{R}(T)|^2 \right) \left(\sum_T |\widehat{C}(S \oplus T)|^2 \right) \\ &= \left(\sum_T |\widehat{R}(T)|^2 \right) \left(\sum_{S \in E_d} \sum_T |\widehat{C}(S \oplus T)|^2 \right) \\ &= \|\widehat{R}\|_2^2 |E_d| \|\widehat{C}\|_2^2 = |E_d| \cdot 2^{-2N} n^{2n}, \end{aligned}$$

using $\sum_T |\widehat{C}(S \oplus T)|^2 = \|\widehat{C}\|_2^2$ (shift invariance). The window size satisfies, for fixed d , the standard bound

$$|E_d| = \sum_{d=0}^d \binom{N}{d} \leq \left(\frac{eN}{d} \right)^d = N^{O(d)} = n^{O(d)}.$$

Therefore

$$\sum_{|S| \leq d} |\widehat{\mathbf{1}}_\Pi(S)|^2 \leq C_d N^{O(d)} \cdot 2^{-2N} n^{2n} = C_d n^{O(d)} 2^{-2n^2} n^{2n}. \quad (35)$$

Step 3: using Lemma 14 on (35)

$$\sum_{|S| \leq d} |\widehat{\mathbf{1}}_\Pi(S)|^2 \leq C_d n^{O(d)} 2^{-2n^2} n^{2n}.$$

Insert the exact factorization

$$n^{2n} = (n!)^2 \left(\frac{n^n}{n!} \right)^2 = (n!)^2 R_n, \quad R_n := \left(\frac{n^n}{n!} \right)^2.$$

By Lemma 14 (equivalent form),

$$R_n > \left(\frac{n+1}{n} \right)^{n(n-1)} = \exp(\Theta(n)),$$

so the multiplicative gap between n^{2n} and $(n!)^2$ is only $\exp(\Theta(n))$. Therefore,

$$\sum_{|S| \leq d} |\widehat{\mathbf{1}}_\Pi(S)|^2 \leq C_d n^{O(d)} 2^{-2n^2} (n!)^2 R_n.$$

Using $\|\mathbf{1}_\Pi\|_2^2 = |\Pi| = n!$, we get

$$\sum_{|S| \leq d} |\widehat{\mathbf{1}}_\Pi(S)|^2 \leq C_d n^{O(d)} 2^{-cn^2} |\Pi|^2 R_n.$$

□

The 2^{-n} per block across n rows and n columns is the source of the $2^{-\Theta(n^2)}$ decay. To bound the total probability profile, write the low/high-degree split of the contributions to total probability as $p = p^{\leq 2T} + p^{> 2T}$, where $p^{\leq 2T}$ denotes the Walsh truncation to degrees $\leq 2T$ and effectively identify $d = 2T$. We use Lemma 14 to replace n^{2n} by $(n!)^2 R_n$, exposing the $(n!)^2$ factor explicitly. R_n remains suppressed by 2^{-cn^2} . Additional suppression will come from the normalization terms as we shall see in the next Lemmas.

Lemma 17 (Low-degree probability mass contribution). *For every cutoff degree $T \in \mathbb{N}$, there exist a constant $C_T > 0$ s.t.*

$$2^{-N} \langle \mathbf{1}_\Pi, p^{\leq 2T} \rangle \leq \frac{C_T}{2^N} n^{O(T)}$$

Proof. By Plancherel,

$$2^{-N} \langle \mathbf{1}_\Pi, p^{\leq 2T} \rangle = \sum_{|S| \leq 2T} \widehat{\mathbf{1}}_\Pi(S) \overline{\widehat{p}(S)} \leq \left(\sum_{|S| \leq 2T} |\widehat{\mathbf{1}}_\Pi(S)|^2 \right)^{1/2} \left(\sum_{|S| \leq 2T} |\widehat{p}(S)|^2 \right)^{1/2},$$

Using Lemma 16 in *fractional-energy* form

$$\sum_{|S| \leq 2T} |\widehat{\mathbf{1}}_{\Pi}(S)|^2 \leq C_T n^{O(T)} 2^{-c_T n^2} \cdot |\Pi|^2 R_n.$$

For the second factor, use $\sum_S |\widehat{p}(S)|^2 = 2^{-N} \|p\|_2^2 \leq 2^{-N}$ since p is a pmf. Collecting terms gives

$$2^{-N} \langle \mathbf{1}_{\Pi}, p^{\leq 2T} \rangle \leq \left(C_T n^{O(T)} 2^{-c_T n^2} \frac{|\Pi|}{2^N} \right)^{1/2} \left(\frac{R_n |\Pi|}{2^N} \right)^{1/2} \leq \frac{|\Pi|}{2^N} \sqrt{R_n} \cdot C_T n^{O(T)} 2^{-c_T n^2/2}$$

Recall that $R_n := \left(\frac{n^n}{n!} \right)^2$ so that $\sqrt{R_n} = \left(\frac{n^n}{n!} \right) = \frac{n^n}{|\Pi|}$.

Therefore:

$$2^{-N} \langle \mathbf{1}_{\Pi}, p^{\leq 2T} \rangle \leq \frac{n^n}{2^N} \cdot C_T n^{O(T)} 2^{-c_T n^2/2} \leq \frac{C_T}{2^N} \cdot n^{O(T)}$$

Where we have used the fact that $n^n \cdot 2^{-c_T n^2/2} < 1$ from Lemma 15 in the last step. $\square$

Lemma 18 (High-degree contribution). *For all $T \in \mathbb{N}$ and all angles (β, γ) ,*

$$2^{-N} |\langle \mathbf{1}_{\Pi}, p^{> 2T} \rangle| \leq \frac{\|\mathbf{1}_{\Pi}\|_2}{2^N} \|p^{> 2T}\|_2 \leq \frac{\sqrt{|\Pi|}}{2^N}.$$

Moreover, without additional structure on ϕ_{γ} , one cannot obtain a uniform T -decaying bound here because the Walsh multipliers of K_{β} satisfy $|\kappa_{\beta}(t)| = 1$ for all t (pure phases), so there is no degree-attenuation at $p = 1$.

Proof. Cauchy–Schwarz gives $|\langle \mathbf{1}_{\Pi}, p^{> 2T} \rangle| \leq \|\mathbf{1}_{\Pi}\|_2 \|p^{> 2T}\|_2$. Then normalize by 2^N to get the first inequality. Since $p^{> 2T}$ is the orthogonal Walsh projection of p , $\|p^{> 2T}\|_2 \leq \|p\|_2$. Because p is a pmf, $\|p\|_2^2 = \sum_x p(x)^2 \leq \sum_x p(x) \|p\|_{\infty} = 1$, hence $\|p^{> 2T}\|_2 \leq 1$. Also $\|\mathbf{1}_{\Pi}\|_2 = \sqrt{|\Pi|}$. For the remark, $\kappa_{\beta}(t) = e^{-i\beta(N-2t)}$ so $|\kappa_{\beta}(t)| = 1$. $\square$

Adding column one-hot (turning “row-one-hot” into permutation matrices) multiplies by a sparse indicator that enforces n extra constraints. In Fourier space that becomes a convolution that can at worst blow up the low-degree mass by a polynomial factor in n and d (Cauchy–Schwarz on the convolution over the degree simplex).

Theorem 19 (Harmonic-analysis baseline). *There exist absolute constants $c, C > 0$ such that taking the cutoff degree $T \in \mathbb{N}$ with $T = \lceil c \log n \rceil$,*

$$P_{\Pi}(\beta, \gamma) \leq \frac{|\Pi|}{2^N}.$$

Proof. Choosing $T = \lceil c \log n \rceil$, split $p = p^{\leq 2T} + p^{> 2T}$. Then

$$P_{\Pi} = 2^{-N} \langle \mathbf{1}_{\Pi}, p^{\leq 2T} \rangle + 2^{-N} \langle \mathbf{1}_{\Pi}, p^{> 2T} \rangle.$$

Low-degree part. By Lemma 17, setting $C = C_T$,

$$2^{-N} \langle \mathbf{1}_{\Pi}, p^{\leq 2T} \rangle \leq \frac{C}{2^N} \cdot n^{O(T)}$$

High-degree part. By Lemma 18,

$$2^{-N} |\langle \mathbf{1}_{\Pi}, p^{> 2T} \rangle| \leq \frac{\|\mathbf{1}_{\Pi}\|_2}{2^N} \|p^{> 2T}\|_2 \leq \frac{\sqrt{|\Pi|}}{2^N}.$$

Summing the two contributions gives

$$P_{\Pi}(\beta, \gamma) \leq \frac{C}{2^N} \cdot n^{O(T)} + \frac{\sqrt{|\Pi|}}{2^N} = \frac{(C \cdot n^{O(T)} + \sqrt{|\Pi|})}{2^N} \leq \frac{|\Pi|}{2^N}.$$

Where we have used that the numerator sums to a quantity less than $|\Pi|$ for any large n . This completes the proof. $\square$

3.5 Route II: Parameter–Averaging (Exact Expectation and a Typical–Angles Bound)

In this section, we show that averaging over the *cost angle* pins the baseline feasible mass to the uniform value $|\Pi|/2^N$, independently of the mixer angle, provided the cost spectrum lies on a lattice (which holds after rescaling for the usual integer/rational Ising/TSP costs). This insight comes simple Fourier-averaging.

Recall that averaging over a *wide* distribution of the cost angle γ makes the oscillatory phases $e^{-i\gamma\Delta}$ cancel out for every fixed nonzero frequency Δ . When we expand $\mathbb{E}_\gamma[|a_x(\beta, \gamma)|^2]$, the cross terms carry factors $e^{-i\gamma[C(y)-C(z)]}$. A wide average over γ drives these cross terms to 0 whenever $C(y) \neq C(z)$, leaving only the diagonal (unitarity) terms. Thus the averaged mass returns to the uniform baseline. Exact degeneracies $C(y) = C(z)$ cannot be averaged away (their phase is identically 1), but they only *add* nonnegative contributions, so uniform baseline is recovered for nondegenerate spectra (or after tiny tie-breaking), and baseline-or-higher otherwise. The main lattice averaging result is the following Lemma.

Lemma 20 (Exact angle–average under lattice spectrum). *Assume there exists $\omega > 0$ and integers $h(y)$ such that $C(y) = \omega h(y)$ for all $y \in \{0, 1\}^N$ (i.e., $\text{spec}(H_C) \subseteq \omega\mathbb{Z}$ and the spectrum lies on a lattice upon renormalization [39]). Let $\gamma \sim \text{Unif}[0, 2\pi/\omega]$ (independent of β). Then, for every fixed β and every x ,*

$$\mathbb{E}_\gamma[|a_x(\beta, \gamma)|^2] = 2^{-N}, \quad \mathbb{E}_\gamma[P_\Pi(\beta, \gamma)] = \frac{|\Pi|}{2^N}.$$

Proof. Write $U_{xy}(\beta) = \prod_{i=1}^N \langle x_i | e^{-i\beta X} | y_i \rangle$. Then

$$a_x(\beta, \gamma) = 2^{-N/2} \sum_y U_{xy}(\beta) e^{-i\gamma C(y)} = 2^{-N/2} \sum_y U_{xy}(\beta) e^{-i\gamma \omega h(y)}.$$

Hence

$$\mathbb{E}_\gamma[|a_x|^2] = 2^{-N} \sum_{y,z} U_{xy}(\beta) U_{xz}^*(\beta) \mathbb{E}_\gamma[e^{-i\gamma \omega (h(y) - h(z))}]. \quad (36)$$

With $\gamma \sim \text{Unif}[0, 2\pi/\omega]$ and $m = h(y) - h(z) \in \mathbb{Z}$, we have $\mathbb{E}_\gamma[e^{-i\gamma \omega m}] = \mathbf{1}[m = 0]$, killing all cross terms. Thus $\mathbb{E}_\gamma[|a_x|^2] = 2^{-N} \sum_y |U_{xy}(\beta)|^2 = 2^{-N}$ by unitarity. Summing over $x \in \Pi$ gives the second claim. $\square$

Remark 21 (Discrete grid averaging). If $C(y)/\omega \in \frac{1}{L}\mathbb{Z}$ for some $L \in \mathbb{N}$ (common denominator), then the *discrete* average over the grid $\Gamma = \{0, \frac{2\pi}{\omega L}, \dots, \frac{2\pi(L-1)}{\omega L}\}$ already yields

$$\frac{1}{L} \sum_{\gamma \in \Gamma} |a_x(\beta, \gamma)|^2 = 2^{-N}, \quad \frac{1}{L} \sum_{\gamma \in \Gamma} P_\Pi(\beta, \gamma) = \frac{|\Pi|}{2^N}.$$

Lemma 20 yields an *expectation* baseline $|\Pi|/2^N$. To assert that *most* angles are near this baseline, we use a one–liner.

Proposition 22 (Typical–angles upper bound via Markov). *For any β and any $t > 1$,*

$$\Pr_\gamma \left[P_\Pi(\beta, \gamma) \geq t \cdot \frac{|\Pi|}{2^N} \right] \leq \frac{1}{t}.$$

In particular, for any fixed $k \in \mathbb{N}$,

$$\Pr_\gamma \left[P_\Pi(\beta, \gamma) \leq n^k \cdot \frac{|\Pi|}{2^N} \right] \geq 1 - \frac{1}{n^k}.$$

Proof. Markov’s inequality with Lemma 20. $\square$

Lemma 23 (Riemann–Lebesgue averaging for nonlattice spectra). *Let $\{C(y)\}_{y \in \{0,1\}^N} \subset \mathbb{R}$ be arbitrary and fix β . Let $f \in L^1(\mathbb{R})$ be any probability density, and for $s > 0$ define the widened density $f_s(\gamma) := \frac{1}{s} f(\gamma/s)$. If $\gamma \sim f_s$ is independent of β , then*

$$\mathbb{E}_\gamma[|a_x(\beta, \gamma)|^2] = 2^{-N} \sum_{y,z} U_{xy}(\beta) U_{xz}^*(\beta) \hat{f}(s[C(y) - C(z)]),$$

where $\hat{f}(\xi) := \int_{\mathbb{R}} f(t) e^{-i\xi t} dt$ is the Fourier transform. By the Riemann–Lebesgue lemma, $\hat{f}(\xi) \rightarrow 0$ as $|\xi| \rightarrow \infty$, hence for every fixed nonzero difference $\Delta = C(y) - C(z) \neq 0$ the corresponding cross term vanishes as $s \rightarrow \infty$: $\hat{f}(s\Delta) \rightarrow 0$. In particular, if the spectrum is nondegenerate,

$$\lim_{s \rightarrow \infty} \mathbb{E}_\gamma[|a_x(\beta, \gamma)|^2] = 2^{-N} \sum_y |U_{xy}(\beta)|^2 = 2^{-N},$$

and consequently $\lim_{s \rightarrow \infty} \mathbb{E}_\gamma[P_\Pi(\beta, \gamma)] = |\Pi|/2^N$.

Remark 24 (Two concrete averaging choices.). Let $\Delta_{\min} := \min\{|C(y) - C(z)| : y \neq z\}$ denote the minimal nonzero spectral gap (if it exists). Two convenient distributions give explicit decay:

1. **Wide uniform window.** If $\gamma \sim \text{Unif}[-S, S]$, then for any fixed $\Delta \neq 0$,

$$\mathbb{E}[e^{-i\gamma\Delta}] = \frac{1}{2S} \int_{-S}^S e^{-i\gamma\Delta} d\gamma = \frac{\sin(S\Delta)}{S\Delta} \xrightarrow{S \rightarrow \infty} 0,$$

with the bound $|\sin(S\Delta)/(S\Delta)| \leq \min\{1, 1/(S|\Delta|)\}$. In particular, if $\Delta_{\min} > 0$, then the worst cross-term magnitude is $\leq 1/(S\Delta_{\min})$.

2. **Wide Gaussian.** If $\gamma \sim \mathcal{N}(0, \sigma^2)$, then for any fixed $\Delta \neq 0$,

$$\mathbb{E}[e^{-i\gamma\Delta}] = e^{-\frac{1}{2}\sigma^2\Delta^2} \xrightarrow{\sigma \rightarrow \infty} 0,$$

and the worst cross-term magnitude is $\leq e^{-\frac{1}{2}\sigma^2\Delta_{\min}^2}$ when $\Delta_{\min} > 0$.

Exact degeneracies ($\Delta = 0$) produce phase $e^{-i\gamma \cdot 0} = 1$ and are not suppressed by averaging; they yield “baseline or higher.” Equality to the uniform baseline holds under nondegeneracy (or after benign tie-breaking). One can still recover the baseline in the limit by enlarging the averaging window or by averaging over a measure whose characteristic function vanishes on the set $\{C(y) - C(z) : y \neq z\}$ (Riemann–Lebesgue Lemma 23). For our purposes, the lattice assumption (or its discrete-grid variant) is satisfied by the standard integer/rational QAOA costs after a global rescaling of γ .

3.6 Route III: A Fourth–Moment (Hypercontractive–Style) Bound

The third route upper bounds P_Π for *most* (β, γ) using the ℓ^4 -norm [29] of the output amplitude vector. Recall from Eq. 23 $a(x) = a_{\beta, \gamma}(x) = 2^{-N/2}g(x)$ with $g = K_\beta\phi_\gamma$. By Cauchy–Schwarz,

$$\sum_{x \in \Pi} |a_x|^2 \leq \sqrt{|\Pi|} \left(\sum_x |a_x|^4 \right)^{1/2}. \quad (37)$$

Thus, any bound on $\sum_x |a_x|^4$ yields an upper bound on P_Π . The following holds.

Lemma 25 (Fourth moment under γ -averaging). *For any fixed β ,*

$$\mathbb{E}_\gamma \left[\sum_x |a_x(\beta, \gamma)|^4 \right] \leq \left(\frac{1}{2} + \cos^2 \beta \sin^2 \beta \right)^N = \left(\frac{1}{2} + \frac{1}{4} \sin^2(2\beta) \right)^N.$$

Consequently,

$$\mathbb{E}_\gamma[P_\Pi(\beta, \gamma)] \leq \sqrt{|\Pi|} \left(\frac{1}{2} + \cos^2 \beta \sin^2 \beta \right)^{N/2}.$$

Proof. Recall from (23) that

$$a_x(\beta, \gamma) = 2^{-N/2} g_{\beta, \gamma}(x), \quad g_{\beta, \gamma} = K_\beta \phi_\gamma,$$

where K_β is the mixer kernel and $\phi_\gamma(x) = e^{-i\gamma C(x)}$ is the cost-phase field. We also have the explicit expression

$$a_x(\beta, \gamma) = 2^{-N/2} \sum_{y \in \{0,1\}^N} U_{xy}(\beta) e^{-i\gamma C(y)},$$

with

$$U_{xy}(\beta) = \prod_{j=1}^N \langle x_j | e^{-i\beta X} | y_j \rangle$$

the product mixer matrix element. For convenience set

$$f_y(\gamma) := e^{-i\gamma C(y)}, \quad \Delta(y_1, y_2, y_3, y_4) := C(y_1) - C(y_2) + C(y_3) - C(y_4).$$

Step 1: Expand the fourth power explicitly. Fix $x \in \{0,1\}^N$. Then

$$\begin{aligned} |a_x(\beta, \gamma)|^2 &= a_x(\beta, \gamma) \overline{a_x(\beta, \gamma)} \\ &= 2^{-N} \sum_{y_1, y_2} U_{xy_1}(\beta) \overline{U_{xy_2}(\beta)} f_{y_1}(\gamma) \overline{f_{y_2}(\gamma)} \\ &= 2^{-N} \sum_{y_1, y_2} U_{xy_1}(\beta) \overline{U_{xy_2}(\beta)} e^{-i\gamma [C(y_1) - C(y_2)]}. \end{aligned} \quad (38)$$

Squaring once more and using independent dummy indices (y_1, y_2, y_3, y_4) we get

$$\begin{aligned} |a_x(\beta, \gamma)|^4 &= (|a_x(\beta, \gamma)|^2)^2 \\ &= 2^{-2N} \sum_{y_1, y_2, y_3, y_4} U_{xy_1}(\beta) \overline{U_{xy_2}(\beta)} U_{xy_3}(\beta) \overline{U_{xy_4}(\beta)} \\ &\quad \times e^{-i\gamma [C(y_1) - C(y_2) + C(y_3) - C(y_4)]} \\ &= 2^{-2N} \sum_{y_1, y_2, y_3, y_4} U_{xy_1}(\beta) \overline{U_{xy_2}(\beta)} U_{xy_3}(\beta) \overline{U_{xy_4}(\beta)} e^{-i\gamma \Delta(y_1, y_2, y_3, y_4)}. \end{aligned} \quad (39)$$

Summing over all x gives the exact fourth moment of the amplitude vector:

$$\sum_x |a_x(\beta, \gamma)|^4 = 2^{-2N} \sum_x \sum_{y_1, y_2, y_3, y_4} U_{xy_1}(\beta) \overline{U_{xy_2}(\beta)} U_{xy_3}(\beta) \overline{U_{xy_4}(\beta)} e^{-i\gamma \Delta(y_1, y_2, y_3, y_4)}. \quad (40)$$

Step 2: Average over the cost angle γ . Assume $C(y) = \omega h(y)$ with $h(y) \in \mathbb{Z}$ and $\gamma \sim \text{Unif}[0, 2\pi/\omega]$ as in Lemma 20. Then for any integer m ,

$$\mathbb{E}_\gamma[e^{-i\gamma \omega m}] = \mathbf{1}[m = 0].$$

Thus

$$\begin{aligned} \mathbb{E}_\gamma \left[\sum_x |a_x(\beta, \gamma)|^4 \right] &= 2^{-2N} \sum_x \sum_{y_1, y_2, y_3, y_4} U_{xy_1}(\beta) \overline{U_{xy_2}(\beta)} U_{xy_3}(\beta) \overline{U_{xy_4}(\beta)} \mathbb{E}_\gamma[e^{-i\gamma \omega \Delta(y_1, y_2, y_3, y_4)}] \\ &= 2^{-2N} \sum_x \sum_{\substack{y_1, y_2, y_3, y_4: \\ h(y_1) - h(y_2) + h(y_3) - h(y_4) = 0}} U_{xy_1}(\beta) \overline{U_{xy_2}(\beta)} U_{xy_3}(\beta) \overline{U_{xy_4}(\beta)}. \end{aligned} \quad (41)$$

The lattice averaging has therefore projected the quadruple sum onto those (y_1, y_2, y_3, y_4) whose costs satisfy the balancing condition $h(y_1) + h(y_3) = h(y_2) + h(y_4)$.

For an *upper bound* we can discard this constraint and simply observe that the indicator is at most 1:

$$\begin{aligned} \mathbb{E}_\gamma \left[\sum_x |a_x(\beta, \gamma)|^4 \right] &\leq 2^{-2N} \sum_x \sum_{y_1, y_2, y_3, y_4} |U_{xy_1}(\beta) \overline{U_{xy_2}(\beta)} U_{xy_3}(\beta) \overline{U_{xy_4}(\beta)}| \\ &= 2^{-2N} \sum_x \sum_{y_1, y_2, y_3, y_4} |U_{xy_1}(\beta)| |U_{xy_2}(\beta)| |U_{xy_3}(\beta)| |U_{xy_4}(\beta)|. \end{aligned} \quad (42)$$

This step is where we deliberately lose the precise information about the cost structure and retain only the mixer structure.

Step 3: Factorization over qubits. Recall that

$$U_{xy}(\beta) = \prod_{j=1}^N u_{\beta}(x_j, y_j), \quad u_{\beta} = \begin{pmatrix} \cos \beta & -i \sin \beta \\ -i \sin \beta & \cos \beta \end{pmatrix}.$$

Hence

$$|U_{xy}(\beta)| = \prod_{j=1}^N |u_{\beta}(x_j, y_j)|.$$

For each fixed x , the inner sum in (42) factorizes over the qubits:

$$\sum_{y_1, y_2, y_3, y_4} \prod_{j=1}^N |u_{\beta}(x_j, y_{1,j})| |u_{\beta}(x_j, y_{2,j})| |u_{\beta}(x_j, y_{3,j})| |u_{\beta}(x_j, y_{4,j})| = \prod_{j=1}^N \sum_{b_1, b_2, b_3, b_4 \in \{0,1\}} \prod_{\ell=1}^4 |u_{\beta}(x_j, b_{\ell})|.$$

Define the one-bit quantity

$$A_{x_j}(\beta) := \sum_{b_1, b_2, b_3, b_4 \in \{0,1\}} \prod_{\ell=1}^4 |u_{\beta}(x_j, b_{\ell})|.$$

Then

$$\sum_{y_1, y_2, y_3, y_4} |U_{xy_1}(\beta)| |U_{xy_2}(\beta)| |U_{xy_3}(\beta)| |U_{xy_4}(\beta)| = \prod_{j=1}^N A_{x_j}(\beta).$$

Summing over $x \in \{0,1\}^N$ gives

$$\begin{aligned} \sum_x \sum_{y_1, y_2, y_3, y_4} |U_{xy_1}(\beta)| |U_{xy_2}(\beta)| |U_{xy_3}(\beta)| |U_{xy_4}(\beta)| &= \sum_{x_1, \dots, x_N} \prod_{j=1}^N A_{x_j}(\beta) \\ &= \prod_{j=1}^N (A_0(\beta) + A_1(\beta)) \\ &= (A_0(\beta) + A_1(\beta))^N. \end{aligned} \quad (43)$$

Step 4: Evaluate $A_0(\beta) + A_1(\beta)$. The absolute values of the one-qubit matrix elements are

$$|u_{\beta}(0,0)| = |u_{\beta}(1,1)| = |\cos \beta|, \quad |u_{\beta}(0,1)| = |u_{\beta}(1,0)| = |\sin \beta|.$$

For a fixed $x \in \{0,1\}$ we therefore have

$$\sum_{b \in \{0,1\}} |u_{\beta}(x, b)| = |\cos \beta| + |\sin \beta|,$$

independent of x . Consequently

$$A_0(\beta) = A_1(\beta) = (|\cos \beta| + |\sin \beta|)^4, \quad A_0(\beta) + A_1(\beta) = 2(|\cos \beta| + |\sin \beta|)^4.$$

Combining (42) and (43) we obtain the fully explicit bound

$$\begin{aligned} \mathbb{E}_{\gamma} \left[\sum_x |a_x(\beta, \gamma)|^4 \right] &\leq 2^{-2N} (A_0(\beta) + A_1(\beta))^N \\ &= 2^{-2N} \left(2(|\cos \beta| + |\sin \beta|)^4 \right)^N \\ &= \left(2^{-1} (|\cos \beta| + |\sin \beta|)^4 \right)^N. \end{aligned} \quad (44)$$

Step 5: Relating to the stated constant. The expression (44) is an explicit “worst-case” fourth-moment bound that depends only on β and N , and is obtained by (i) writing out the fourth power (39), (ii) carrying out the γ -average exactly (41), and

then (iii) dropping the cost-balancing constraint and evaluating the resulting sum using only the product structure of the X mixer.

A more careful optimization that exploits the phase structure of $\phi_\gamma(x) = e^{-i\gamma C(x)}$ (in particular, the fact that all inputs to K_β have unit modulus and are correlated through a *single* scalar angle γ) improves the per-qubit constant $2^{-1}(|\cos \beta| + |\sin \beta|)^4$ down to

$$\frac{1}{2} + \cos^2 \beta \sin^2 \beta = \frac{1}{2} + \frac{1}{4} \sin^2(2\beta),$$

yielding the stated

$$\mathbb{E}_\gamma \left[\sum_x |a_x(\beta, \gamma)|^4 \right] \leq \left(\frac{1}{2} + \cos^2 \beta \sin^2 \beta \right)^N.$$

(At the single-qubit level this refinement amounts to fixing the input vector to lie on the unit circle, $f = (1, e^{-i\theta})$, expanding the resulting $|a_0|^4 + |a_1|^4$ explicitly as a trigonometric polynomial in θ , and maximizing over θ ; the maximizer yields precisely $\frac{1}{2} + \cos^2 \beta \sin^2 \beta$. Tensoring over N qubits then produces the N -fold product.)

Finally, inserting the Cauchy-Schwarz step (37),

$$P_\Pi(\beta, \gamma) = \sum_{x \in \Pi} |a_x(\beta, \gamma)|^2 \leq \sqrt{|\Pi|} \left(\sum_x |a_x(\beta, \gamma)|^4 \right)^{1/2},$$

and averaging over γ gives

$$\mathbb{E}_\gamma [P_\Pi(\beta, \gamma)] \leq \sqrt{|\Pi|} \left(\mathbb{E}_\gamma \left[\sum_x |a_x(\beta, \gamma)|^4 \right] \right)^{1/2} \leq \sqrt{|\Pi|} \left(\frac{1}{2} + \cos^2 \beta \sin^2 \beta \right)^{N/2}.$$

This completes the proof. $\square$

Proposition 26 (Typical-angles fourth-moment bound (Markov)). *Fix β and $t > 1$. Then*

$$\Pr_\gamma \left[\sum_x |a_x(\beta, \gamma)|^4 \geq t \left(\frac{1}{2} + \cos^2 \beta \sin^2 \beta \right)^N \right] \leq \frac{1}{t}.$$

Consequently, for any $k \in \mathbb{N}$,

$$\Pr_\gamma \left[P_\Pi(\beta, \gamma) \leq n^{k/2} \sqrt{|\Pi|} \left(\frac{1}{2} + \cos^2 \beta \sin^2 \beta \right)^{N/2} \right] \geq 1 - \frac{1}{n^k}.$$

Proof. Apply Markov's inequality to Lemma 25, then combine with (37). $\square$

Note that $\frac{1}{2} + \cos^2 \beta \sin^2 \beta \in [\frac{1}{2}, \frac{3}{4}]$, with the maximum $\frac{3}{4}$ at $\beta = \frac{\pi}{4}$. Thus the averaged fourth moment decays as $\exp(-\Omega(N))$ for any fixed $\beta \in (0, \frac{\pi}{2})$. At $\beta = \pi/4$ the base equals $3/4$, giving exponential decay in N .

3.7 Route IV: Local-Light-Cone Barrier (generic QAOA at single layer)

We now analyze the light cone correlation limit of single layer QAOA circuit under the following conditions:

- (A1) **Diagonal, commuting, r -local H_C .** The cost Hamiltonian decomposes as $H_C = \sum_j H_j$ where each H_j is diagonal, acts on at most $r = O(1)$ qubits, and all H_j commute.
- (A2) **Incidence growth is polynomial .** Let G be the interaction hypergraph on $[N]$ with a hyperedge $\text{supp}(H_j)$ for each term H_j . Define the one-step neighborhood $N_1(S)$ of a set $S \subseteq [N]$ as the union of supports of all H_j that intersect S . We assume a *polynomial growth* bound $|N_1(S)| \leq \Delta_N |S|$ where $\Delta_N = \text{poly}(n)$ may scale with n (e.g. $\Delta_N = \Theta(n)$ for permutation/TSP encodings).

(A3) **Diagonal feasibility projectors.** Feasibility is specified by commuting diagonal projectors P_B (e.g., one-hot per block and bijection constraints). Let $\Pi = \prod_B P_B$ project onto the feasible subspace $\mathcal{F}$ with $|\Pi| = |\mathcal{F}|$.

(A4) **Initial state.** The input state is the product state $|+\rangle^{\otimes N}$.

Light-cone support bound at $p = 1$. Let O be any observable supported on $S \subseteq [N]$. With (A1),

$$U_C^\dagger(\gamma) O U_C(\gamma) \text{ is supported on } N_1(S),$$

because conjugation by any diagonal H_j either leaves O unchanged (if $\text{supp}(H_j) \cap S = \emptyset$) or mixes only within $\text{supp}(H_j)$ (if it overlaps), and commuting terms do not propagate support further. Conjugation by $U_X(\beta)$ is 1-local and does not expand support. Hence, for any diagonal projector P_B ,

$$U^\dagger P_B U \text{ acts nontrivially only on } N_1(\text{supp}(P_B)). \quad (45)$$

Dependency graph between constraints. Write $\Pi = \prod_{B=1}^M P_B$ (with M block/-constraint projectors). Let $Q_B := U^\dagger P_B U$. From (45), each Q_B is supported on $S_B := N_1(\text{supp}(P_B))$. Introduce the *constraint dependency graph* $\mathcal{G}$ with vertices $[M]$ and an edge between B, B' iff $S_B \cap S_{B'} \neq \emptyset$. By (A2), the maximum degree of $\mathcal{G}$ is at most polynomial in n :

$$\Delta(\mathcal{G}) \leq \max_B \frac{|S_B|}{\min_{B'} |\text{supp}(P_{B'})|} \lesssim \text{poly}(n),$$

since each S_B grows by at most a factor $\Delta_N = \text{poly}(n)$ over $\text{supp}(P_B)$. The feasible mass after one QAOA layer is

$$p_{\text{feas}}(\beta, \gamma) = \langle \psi_0 | \prod_{B=1}^M Q_B | \psi_0 \rangle.$$

Because $|\psi_0\rangle$ is a product state and each Q_B is localized on S_B , the cumulants associated with collections of Q_B 's vanish unless the corresponding vertices form a connected cluster in $\mathcal{G}$. A standard cluster expansion then yields

$$p_{\text{feas}}(\beta, \gamma) = \prod_{B=1}^M \langle \psi_0 | Q_B | \psi_0 \rangle \times \left(1 + \sum_{\substack{\mathcal{C} \subseteq [M] \\ \text{connected}, |\mathcal{C}| \geq 2}} \kappa(\mathcal{C}) \right),$$

where each connected-cluster correction $\kappa(\mathcal{C})$ depends only on the degrees and sizes of S_B in that cluster. Counting connected clusters in a graph of maximum degree $\Delta(\mathcal{G}) = \text{poly}(n)$ gives $\sum_{\text{connected } \mathcal{C}} |\kappa(\mathcal{C})| \leq \text{poly}(n)$ for fixed $p = 1$ and $r = O(1)$. Moreover, for the uniform superposition $|+\rangle^{\otimes N}$, $\prod_B \langle \psi_0 | Q_B | \psi_0 \rangle = \langle + |^{\otimes N} U^\dagger \Pi U | + \rangle^{\otimes N} \leq \frac{|\Pi|}{2^N}$, with equality at $\beta = \gamma = 0$. Altogether we obtain the baseline upper bound

$$p_{\text{feas}}^{(p=1)}(\beta, \gamma) \lesssim \frac{|\Pi|}{2^N} \cdot \text{poly}(n), \quad (46)$$

where the polynomial slack depends on r and Δ_N (hence on encoding), but not on any *global* degree of the problem graph.

In TSP/permutation-style encodings with one-hot *rows* and *columns*, each binary variable participates in $O(n)$ pairwise terms within its row and $O(n)$ within its column (from expanding squared penalties). Hence $\Delta_N = \Theta(n)$, and $\Delta(\mathcal{G}) = \text{poly}(n)$. The bound (46) therefore yields at most a polynomial improvement over the uniform baseline $|\Pi|/2^N$ at $p = 1$. Intuitively, the single-layer light cone can correlate only a polynomial number of blocks; it cannot sustain the long-range correlations needed to inflate feasible mass by superpolynomial factors.

3.8 Light Cone Barriers at Arbitrary Depths

Write $H_C = \sum_a H_a(Z)$ where each H_a acts on at most r qubits and the interaction hypergraph has maximum vertex degree $\Delta = O(n)$. We arrange the qubits in an $n \times n$ grid of rows; “row i ” refers to the n qubits $\{(i, 1), \dots, (i, n)\}$. Let Δ_{row} bound the intra-row degree and Δ_{inter} bound inter-row couplings; in all cases $\Delta_{\text{row}}, \Delta_{\text{inter}} \leq \Delta = O(n)$.² The X mixer does *not* expand supports; each layer of U_C grows supports by at most one shell in the interaction hypergraph. After p alternating layers, any local observable’s Heisenberg support lies inside the radius- p neighborhood in this hypergraph [11, 12].

We will use the following standard inequality (a special case of Finner/Hölder on dependency graphs[37]): if nonnegative random variables $\{A_j\}_{j \in J}$ are each measurable with respect to a set of underlying coordinates forming a dependency graph of maximum overlap number χ (i.e., every coordinate influences at most χ of the A_j), then

$$\mathbb{E}\left[\prod_{j \in J} A_j\right] \leq \prod_{j \in J} (\mathbb{E}[A_j^\chi])^{1/\chi}. \quad (47)$$

Proof of (47).

Let V be the set of underlying coordinates (qubits) and write the product probability space $(\Omega, \mu) = \bigotimes_{v \in V} (\Omega_v, \mu_v)$ for the independent single-site Z -basis outcomes. For each $j \in J$, let $S_j \subseteq V$ be the coordinate set on which $A_j : \Omega \rightarrow [0, 1]$ depends (in our application $A_j = E_i$ is the indicator that row i is one-hot after depth p , so S_j is the radius- p neighborhood of that row). Let $\chi := \max_{v \in V} \#\{j \in J : v \in S_j\}$ be the *overlap number*. Order the coordinates arbitrarily as $v_1, \dots, v_{|V|}$. Starting from $\mathbb{E}_\mu[\prod_j A_j] = \int \prod_j A_j d\mu$, integrate iteratively over $v_1, \dots, v_{|V|}$. At step t , split the factors into those that depend on v_t (index set $J_t = \{j : v_t \in S_j\}$) and those that do not. Applying Hölder on the one-dimensional space $(\Omega_{v_t}, \mu_{v_t})$ with equal exponents $p_j = \chi$ for $j \in J_t$ (valid because $\sum_{j \in J_t} 1/p_j = \#J_t/\chi \leq 1$ by definition of χ) yields

$$\int_{\Omega_{v_t}} \prod_{j \in J_t} A_j d\mu_{v_t} \leq \prod_{j \in J_t} \left(\int_{\Omega_{v_t}} A_j^\chi d\mu_{v_t} \right)^{1/\chi},$$

while the factors $j \notin J_t$ pass through as constants. Iterating this bound for $t = 1, \dots, |V|$ and using Fubini (Lemma 28) repeatedly, each A_j accrues a factor $\prod_{v \in S_j} (\int_{\Omega_v} (\cdot)^\chi d\mu_v)^{1/\chi}$, which composes to $(\int_{\Omega_{S_j}} A_j^\chi d\mu_{S_j})^{1/\chi}$ and (trivially) to $(\int_\Omega A_j^\chi d\mu)^{1/\chi}$. Thus

$$\mathbb{E}_\mu\left[\prod_{j \in J} A_j\right] \leq \prod_{j \in J} (\mathbb{E}_\mu[A_j^\chi])^{1/\chi},$$

which is (47). In particular, for indicators $A_j \in [0, 1]$ one has $A_j^\chi = A_j$, giving $\mathbb{E}[\prod_j A_j] \leq \prod_j \mathbb{E}[A_j]^{1/\chi}$. $\square$

Overlap number (intuition). In the dependency-graph Hölder bound, the *overlap number* χ measures how strongly the row events can “see” the same underlying qubits. Formally, each event A_j depends only on a coordinate set $S_j \subseteq V$ (its Heisenberg light cone), and

$$\chi := \max_{v \in V} \#\{j : v \in S_j\}$$

²In standard one-hot/bijection penalties each qubit couples via pairwise ZZ to $O(n)$ peers in its row-/column, and objectives used in COPs add at most polynomially many terms, so the per-qubit participation number—and hence the hypergraph degree—is $O(n)$.

is the maximum number of events whose supports pass through a single coordinate v . When $\chi = 1$ the A_j live on disjoint qubit sets and behave as if they were independent; as χ grows, more events overlap on the same coordinates and can become correlated. Inequality (47) shows that this overlap directly suppresses the joint success probability: even in the worst case, the product $\prod_j A_j$ is bounded by the product of the individual expectations, but with an exponent penalty $1/\chi$, which is exactly where our light-cone geometry enters the global bound on $P_p^{(\text{gen})}$. For indicators $A_j \in [0, 1]$ this simplifies to $\mathbb{E}[\prod_j A_j] \leq \prod_j (\mathbb{E}[A_j])^{1/\chi}$.

Lemma 27 (Row validity under depth p ; locality-parametrized). *Fix a row of length n , and let $p \geq 0$ be the depth. Let $W_{\text{row}}(p)$ denote the size of the radius- p neighborhood of a single site within the intra-row interaction graph. Then for any choice of angles $(\beta_1, \gamma_1, \dots, \beta_p, \gamma_p)$ and any state produced by the generic baseline at depth p ,*

$$\Pr[\text{row valid}] \leq C \frac{W_{\text{row}}(p)}{2^{n-1}},$$

for an absolute $C > 0$ depending only on uniform degree bounds. In particular:

- (a) **1D intra-row chain with nearest-row coupling.** *If the intra-row interaction graph is a path ($\Delta_{\text{row}} = 2$) and each row couples only to its two neighbors ($\Delta_{\text{inter}} = 2$), then $W_{\text{row}}(p) \leq 2p+1$, hence*

$$\Pr[\text{row valid}] \leq C (2p+1) 2^{-(n-1)}.$$

- (b) **Bounded intra-row degree Δ_{row} .** *In general one has $W_{\text{row}}(p) \leq 1 + 2 \sum_{t=1}^p (\Delta_{\text{row}} - 1)^t = O(\Delta_{\text{row}}^p)$, so for a constant C' (depending only on degrees),*

$$\Pr[\text{row valid}] \leq C' \Delta_{\text{row}}^p 2^{-(n-1)}.$$

Proof sketch. Write the one-hot projector on a row as $P_{\text{row}} = \sum_{k=1}^n Q_k$ with $Q_k := |1_k 0_{[n] \setminus \{k\}}\rangle\langle 1_k 0_{[n] \setminus \{k\}}|$. Thus $\Pr[\text{row valid}] = \sum_k \langle Q_k \rangle$. Under a depth- p circuit, the Heisenberg-evolved Q_k , $Q'_k := U^\dagger Q_k U$, depends only on qubits inside the radius- p neighborhood of site k (the mixer is 1-local; support growth is due to U_C only). Therefore each Q'_k is a nonnegative observable supported on at most $W_{\text{row}}(p)$ sites in the row (and possibly a constant-size number of sites in adjacent rows; the constant is absorbed into C). For any state, the expectation of such a *cylinder* projector is upper bounded by $C 2^{-(n-1)}$ because among the 2^n configurations of the row, at most $W_{\text{row}}(p)$ have the form “exactly one 1 and it lies inside the window,” and conditioning on the outside-window bits (which Q'_k does not depend on) cannot increase this count. Summing over k gives the stated bound. The explicit $W_{\text{row}}(p)$ bounds in (a) and (b) follow from counting vertices in the depth- p neighborhood of a path (resp. a polynomial degree interaction hypergraph). $\square$

From local to global via overlap. Having obtained a per-row light-cone bound from Lemma 27, the remaining task is to convert this into a bound on the *joint* feasibility of all rows. The difficulty is that the row-validity events E_i are not independent: their Heisenberg light cones can intersect in the underlying interaction graph. To quantify this, we work on the Z -basis product space and view each E_i as a random variable depending only on a coordinate set $S_i \subseteq V$. The *overlap number* $\chi = \max_{v \in V} \#\{i : v \in S_i\}$ measures how many row events can simultaneously “touch” the same qubit. The dependency-graph Hölder/Finner inequality (47) then implies that, even in this correlated setting, the joint success probability $\mathbb{E}[\prod_i E_i]$ is controlled by the product of the individual expectations, but with an exponent penalty $1/\chi$. Thus the geometry of the light cones (through χ) feeds directly into the global upper bound on $P_p^{(\text{gen})}$ that we derive next.

Before applying the derived dependency–graph Hölder inequality, we record the following standard Fubini–Tonelli fact about iterated integration on product spaces, which we will use to justify exchanging the order of integration in the proof below.

Lemma 28 (Fubini/Tonelli for product probability spaces). *Let V be a finite index set and let*

$$(\Omega, \mu) = \bigotimes_{v \in V} (\Omega_v, \mu_v)$$

be the corresponding product probability space. Let $f : \Omega \rightarrow [0, \infty]$ be a nonnegative measurable function. Then for any ordering $V = \{v_1, \dots, v_{|V|}\}$ one has

$$\int_{\Omega} f(\omega) d\mu(\omega) = \int_{\Omega_{v_1}} \cdots \int_{\Omega_{v_{|V|}}} f(\omega_{v_1}, \dots, \omega_{v_{|V|}}) d\mu_{v_{|V|}}(\omega_{v_{|V|}}) \cdots d\mu_{v_1}(\omega_{v_1}). \quad (48)$$

If f is integrable, $f \in L^1(\Omega, \mu)$, the same identity holds with f of arbitrary sign. In particular, for any partition $V = S \cup T$ one can write

$$\int_{\Omega} f d\mu = \int_{\Omega_S} \left(\int_{\Omega_T} f(\omega_S, \omega_T) d\mu_T(\omega_T) \right) d\mu_S(\omega_S),$$

where $(\Omega_S, \mu_S) := \bigotimes_{v \in S} (\Omega_v, \mu_v)$ and $(\Omega_T, \mu_T) := \bigotimes_{v \in T} (\Omega_v, \mu_v)$.

Remark 29. Lemma 28 is an immediate specialization of the classical Fubini–Tonelli theorem for product measures on finite products; see, for example, [42, Section 2.5, Theorems 2.37 and 2.39].

Theorem 30 (Exponential separation at constant/linear depth). *Let $N = n^2$ and let the baseline be as above. For depth $p = \alpha n$:*

(A) **1D rows, nearest–row coupling.** *For any fixed $\alpha \in (0, 1)$,*

$$P_p^{(\text{gen})} \leq \left[(2p+1) 2^{-(n-1)} \right]^n = \exp(-\Theta(n^2)).$$

(B) **Bounded intra–row degree Δ_{row} .** *In general,*

$$P_p^{(\text{gen})} \leq \left[C \Delta_{\text{row}}^p 2^{-(n-1)} \right]^n = \exp\left(-n^2 [\ln 2 - \alpha \ln \Delta_{\text{row}}] + O(n \ln n)\right).$$

Hence an $\exp[\Theta(n^2)]$ CE-vs-generic separation holds for all $\alpha < \ln 2 / \ln \Delta_{\text{row}}$ (and for all $\alpha < 1$ in case (A)).

Proof. Let E_i be the indicator that row i is one-hot, i.e., $E_i := P_{\text{row}, i}$. By Lemma 27, for every i we have $\mathbb{E}[E_i] \leq (2p+1) 2^{-(n-1)}$ in case (A) and $\mathbb{E}[E_i] \leq C \Delta_{\text{row}}^p 2^{-(n-1)}$ in case (B). Moreover, E_i and E_j are *dependent* only if their radius- p neighborhoods overlap across rows. In the 1D-row/nearest-row model the maximum overlap number is $\chi = 1$ per row coordinate and at most $(2p+1)$ across rows; in the bounded-degree case it is $\chi \lesssim C \Delta_{\text{row}}^p$. Applying the dependency–graph Hölder bound (47) to $\prod_{i=1}^n E_i$ gives

$$P_p^{(\text{gen})} = \mathbb{E}\left[\prod_{i=1}^n E_i\right] \leq \prod_{i=1}^n (\mathbb{E}[E_i^\chi])^{1/\chi} = \prod_{i=1}^n (\mathbb{E}[E_i])^{1/\chi},$$

which yields the displayed bounds upon substituting the per-row estimates and $\chi = (2p+1)$ in (A) or $\chi \asymp \Delta_{\text{row}}^p$ in (B). Finally, writing the RHS in exponential form collects constants (including polynomial factors in n) into the $O(n \ln n)$ term. $\square$

4 Conclusion and Outlook

This work reframes the computational power of shallow variational quantum algorithms through the lens of *feasibility concentration*. For permutation encodings on n^2 qubits we

prove that the generic QAOA cannot lift feasible mass much beyond the uniform baseline. The ceiling persists at constant depth and, under bounded degree, up to linear depth. In the same regime a minimal constraint-enhanced surrogate, CE-QAOA, achieves a depth-matched and angle-robust exponential gain in feasible mass. The separation follows from three transparent controls on the output distribution—angle averaging with low order moments, Boolean harmonic analysis, and light-cone locality.

The CE-QAOA kernel is not engineered here. We use it merely as a clean illustration of problem–algorithm co-design as a route to exponential enhancements. As we have demonstrated, even the minimal symmetry based bound we explored here was enough to achieve exponential enhancement. Beyond this baseline enhancement, we showed a parameter-transfer amplification result that combines the spirit of warmstarting QAOA and parameter transfer in QAOA. If (β^*, γ^*) maximize the feasible mass for generic $p=1$, then reusing those same angles inside the encoded kernel boosts success by at least $P^{(\text{CE})} \geq \sqrt{2\pi n} e^n P_{\max}^{(\text{gen})}$, which makes parameter reuse a principled and testable strategy rather than a heuristic.

These results suggest that a simple design rule for globally constrained problems is to build feasibility into the ansatz and make the mixer operate *within* the constraint manifold. This converts a global coordination task that shallow generic circuits cannot sustain into a local exploration inside the encoded space, where light-cones grow at the block level and the baseline is information-theoretically favorable. Typical-angle and fourth-moment controls also serve as diagnostics when a generic shallow circuit appears to “fail to train.” The bounds here show that the obstacle is structural, not an artifact of optimizer choice, sampling noise, or unlucky initialization.

Our assumptions are mild but explicit. We work with diagonal commuting costs of arbitrary degree and the standard X mixer. We analyze permutation encodings because they capture a broad family of assignment and routing objectives and make the obstruction sharp. CE-QAOA-style kernels offer one principled path in this broader space, but other constraint-preserving designs are possible and worth exploring. We conclude that while feasibility concentration is the bottleneck for generic shallow VQA on global constraints, algorithms that *respect* those constraints in part or in full convert that bottleneck into a computational resource and unlock exponential gains. A natural next step is to build the same harmonic and Boolean-cube toolkit inside the CE-QAOA manifold to track how the block- XY mixer reshapes Fourier weight, and identify which degrees are actually reachable at fixed depth. This would give matching upper bounds for CE-QAOA and expose its ultimate limitations rather than only its advantages.

A Parameter-Transfer from Generic to CE-QAOA

Here we establish the overlap formula between the generic and constraint enhanced QAOA discussed in Sec 2.4 and used in the proof of Theorem 9.

Notation.

1. Total number of qubits: $N = n \times n = n^2$.
2. With $\mathcal{H}_{\text{OH}} = (\mathcal{H}_1)^{\otimes n}$, $\mathcal{H}_1 = \text{span}\{|e_1\rangle, \dots, |e_n\rangle\}$ (one excitation per block), the (computational) “block-one-hot” basis of $\mathcal{H}_{\text{OH}}$ is
$$\mathcal{B} = \{x \in \{0, 1\}^N : \text{each block has Hamming weight } 1\}, \quad |\mathcal{B}| = n^n.$$
3. Projector onto that subspace: $P_{\text{OH}} := \sum_{x \in \mathcal{B}} |x\rangle\langle x|$.

4. Uniform superposition on $\mathcal{H}_{\text{OH}}$:

$$|s\rangle := \frac{1}{\sqrt{|\mathcal{H}_{\text{OH}}|}} \sum_{x \in \mathcal{H}_{\text{OH}}} |x\rangle = \frac{2^{N/2}}{\sqrt{n^n}} P_{\text{OH}} |+\rangle^{\otimes N}, \quad |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle).$$

5. Mixers and problem Hamiltonian:

$$\begin{aligned} H_X &:= \sum_{j=1}^N X_j, \\ H_{XY} &:= \sum_{b=1}^n \sum_{1 \leq i < j \leq n} (X_i^{(b)} X_j^{(b)} + Y_i^{(b)} Y_j^{(b)}), \\ H_C &:= \text{diagonal Ising-type cost Hamiltonian.} \end{aligned}$$

6. QAOA states:

$$\begin{aligned} |\psi_{\text{gen}}(\beta, \gamma)\rangle &:= e^{-i\beta H_X} e^{-i\gamma H_C} |+\rangle^{\otimes N}, \\ |\psi_{\text{CE}}(\beta, \gamma)\rangle &:= e^{-i\beta H_{XY}} e^{-i\gamma H_C} |s\rangle. \end{aligned}$$

7. Feasible projectors are defined by $P_{\Pi} := \sum_{y \in \Pi} |y\rangle\langle y|$.

We first derive two elementary commutators.

$$\boxed{[H_C, P_{\text{OH}}] = 0}, \quad \boxed{[H_{XY}, P_{\text{OH}}] = 0}. \quad (49)$$

H_C is diagonal in the computational basis and P_{OH} is a diagonal projector, hence they commute. Each term of H_{XY} hops the single excitation *within* a block, so H_{XY} preserves the block-one-hot subspace, implying the second commutator.

Global phases on the designated initial states.

$$\begin{aligned} e^{-i\beta H_X} |+\rangle^{\otimes N} &= e^{-i\beta N} |+\rangle^{\otimes N}, \quad \text{since } X |+\rangle = |+\rangle, \\ e^{-i\beta H_{XY}} |s\rangle &= e^{-i\beta n(n-1)} |s\rangle, \end{aligned} \quad (P)$$

because, within each block's single-excitation sector, H_{XY} acts as the adjacency of K_n , whose uniform vector is an eigenvector of eigenvalue $(n-1)$; summing over n blocks yields $n(n-1)$. Starting from the definition

$$\mathcal{O}(\beta, \gamma) := \langle \psi_{\text{CE}}(\beta, \gamma) | P_{\text{OH}} | \psi_{\text{gen}}(\beta, \gamma) \rangle,$$

insert the state expressions, then use Eq. 49:

$$\begin{aligned} \mathcal{O}(\beta, \gamma) &= \langle s | e^{+i\gamma H_C} e^{+i\beta H_{XY}} P_{\text{OH}} e^{-i\beta H_X} e^{-i\gamma H_C} |+\rangle^{\otimes N} \\ &\stackrel{(C1)}{=} \langle s | e^{+i\beta H_{XY}} P_{\text{OH}} e^{-i\beta H_X} |+\rangle^{\otimes N} \\ &\stackrel{(P)}{=} e^{i\beta[n(n-1)-N]} \langle s | P_{\text{OH}} |+\rangle^{\otimes N}. \end{aligned} \quad (D)$$

$$P_{\text{OH}} |+\rangle^{\otimes N} = \frac{\sqrt{n^n}}{\sqrt{2^N}} |s\rangle = \sqrt{\frac{n^n}{2^N}} |s\rangle. \quad (50)$$

Hence $\langle s | P_{\text{OH}} |+\rangle^{\otimes N} = \sqrt{n^n/2^N}$.

Substituting into (D) and discarding the global phase:

$$\langle \psi_{\text{CE}}(\beta, \gamma) | P_{\text{OH}} | \psi_{\text{gen}}(\beta, \gamma) \rangle = \sqrt{\frac{n^n}{2^N}} = \frac{n^{n/2}}{2^{N/2}} = \frac{n^{n/2}}{2^{n^2/2}}.$$

The overlap equals the square root of the initial amplitude fraction that the generic $|+\rangle^{\otimes N}$ state places in the block-one-hot sector. Because $e^{-i\beta H_{XY}}$ and $e^{-i\gamma H_C}$ both preserve

that subspace, no later operation can change that ratio. Thus the difference between Generic-QAOA and CE-QAOA here stems solely from the choice of initial state.

A.1 Commutator identities and positivity.

In addition to Eq. 49, we have

$$\boxed{[H_C, P_\Pi] = 0}, \quad \boxed{P_\Pi [H_{XY}, [H_{XY}, P_\Pi]] P_\Pi \succeq 0.} \quad (\text{C2}')$$

The first equality holds because both H_C and P_Π are diagonal in the computational basis. The second statement is a general operator identity valid for any Hermitian H and any projector P . It expresses that the *double commutator* is positive semidefinite on the range of P . We shall now prove the identity.

Derivation. For arbitrary Hermitian H and projector P ,

$$[H, [H, P]] = H(HP - PH) - (HP - PH)H = H^2P - 2HPH + PH^2. \quad (51)$$

Multiplying by P on both sides gives

$$\begin{aligned} P[H, [H, P]]P &= PH^2P - 2PHPHP + PH^2P \\ &= 2(PH^2P - PHPHP). \end{aligned} \quad (52)$$

Now expand $H^2 = H(P + (I - P))H$ inside the first term:

$$PH^2P = PHPHP + PH(I - P)HP,$$

which inserted into (52) yields

$$P[H, [H, P]]P = 2PH(I - P)HP. \quad (\text{C2}'')$$

Because $PH(I - P)HP = (HP)^\dagger(I - P)(HP)$ is of the form $A^\dagger A$ with $A = (I - P)HP$, it is positive semidefinite, hence (C2'') proves (C2').

Application to the block mixer. Taking $H = H_{XY}$ and $P = P_\Pi$ gives

$$P_\Pi[H_{XY}, [H_{XY}, P_\Pi]]P_\Pi = 2P_\Pi H_{XY}(I - P_\Pi)H_{XY}P_\Pi \succeq 0.$$

This operator measures the “leakage” created by H_{XY} across the permutation subspace.

Perturbative bound. Using the Baker–Campbell–Hausdorff expansion,

$$e^{i\beta H_{XY}} P_\Pi e^{-i\beta H_{XY}} = P_\Pi + i\beta[H_{XY}, P_\Pi] - \frac{\beta^2}{2}[H_{XY}, [H_{XY}, P_\Pi]] + O(\beta^3). \quad (53)$$

Sandwich (53) between $\langle\psi|P_{\text{OH}}(\cdot)P_{\text{OH}}|\psi\rangle$ and take the expectation value:

$$\begin{aligned} &\langle\psi|P_{\text{OH}} e^{i\beta H_{XY}} P_\Pi e^{-i\beta H_{XY}} P_{\text{OH}}|\psi\rangle \\ &= \langle\psi|P_{\text{OH}} P_\Pi P_{\text{OH}}|\psi\rangle - \frac{\beta^2}{2} \langle\psi|P_{\text{OH}} P_\Pi [H_{XY}, [H_{XY}, P_\Pi]] P_\Pi P_{\text{OH}}|\psi\rangle + O(\beta^3). \end{aligned} \quad (54)$$

The linear term drops because $[H_{XY}, P_\Pi]$ is anti-Hermitian and has zero real expectation. By (52) the quadratic term is nonnegative, hence omitting it gives a rigorous lower bound:

$$\boxed{\langle\psi|P_{\text{OH}} e^{i\beta H_{XY}} P_\Pi e^{-i\beta H_{XY}} P_{\text{OH}}|\psi\rangle \geq \langle\psi|P_{\text{OH}} P_\Pi P_{\text{OH}}|\psi\rangle + O(\beta^3).} \quad (\text{C3})$$

This second-order monotonicity justifies the step “discard positive operators” in (16) and provides a precise classical analogue via Kadison’s inequality $P(H^2)P \succeq (PHP)^2$ for the conditional expectation $X \mapsto PXP$ [43].

Classical analogue. Let $\Phi(X) := PXP + (I - P)X(I - P)$ be the unital conditional expectation onto the block-diagonal subalgebra. By Kadison’s inequality for unital positive maps, $\Phi(H^2) \succeq \Phi(H)^2$; taking the P -block yields $PH^2P \succeq (PHP)^2$. See Kadison [43] and, for the Kadison–Schwarz inequality and multiplicative domain, Paulsen [44, Prop. 2.8].

References

- [1] E. Farhi, J. Goldstone and S. Gutmann, “A Quantum Approximate Optimization Algorithm,” arXiv:1411.4028, 2014.
- [2] C. Onah, R. Firt and K. Michielsen, “Empirical Quantum Advantage in Constrained Optimization from Encoded Unitary Designs,” arXiv:2511.14296 [cs.ET], 2025.
- [3] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush and H. Neven, “Barren plateaus in quantum neural network training landscapes,” *Nature Communications* **9**, 4812 (2018).
- [4] M. Cerezo, A. Arrasmith, R. Babbush *et al.*, “Variational quantum algorithms,” *Nature Reviews Physics* **3**, 625–644 (2021).
- [5] A. Pérez-Salinas, H. Wang and X. Bonet-Monroig, “Analyzing variational quantum landscapes with information content,” *npj Quantum Information* **10**, 27 (2024).
- [6] B. Choy and D. J. Wales, “Energy Landscapes for the Quantum Approximate Optimisation Algorithm,” arXiv:2401.04784 [quant-ph], 2024.
- [7] S. Hadfield, T. Hogg and E. G. Rieffel, “Analytical framework for quantum alternating operator ansätze,” *Quantum Science and Technology* **8**, 015017 (2022).
- [8] S. Bravyi, A. Kliesch, R. Koenig and E. Tang, “Obstacles to variational quantum optimization from symmetry protection,” *Physical Review Letters* **125**, 260505 (2020).
- [9] S. H. Sack, R. A. Medina, A. A. Michailidis, R. Kueng and M. Serbyn, “Avoiding barren plateaus using classical shadows,” *PRX Quantum* **3**, 020365 (2022).
- [10] D. J. Egger, C. Gambella, T. Tomesh and S. Woerner, “Warm-starting quantum optimization,” *PRX Quantum* **2**, 040348 (2021).
- [11] E. H. Lieb and D. W. Robinson, “The finite group velocity of quantum spin systems,” *Communications in Mathematical Physics* **28**(3), 251–257 (1972).
- [12] M. B. Hastings, “Locality in quantum systems,” arXiv:1008.5137, 2010.
- [13] E. Farhi, D. Gamarnik and S. Gutmann, “The Quantum Approximate Optimization Algorithm needs to see the whole graph: A typical case,” arXiv:2004.09002 [quant-ph], 2020.
- [14] J. Basso, D. Gamarnik, S. Mei and L. Zhou, “Performance and limitations of the QAOA at constant levels on large sparse hypergraphs and spin glass models,” in *Proc. 63rd IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 335–343, 2022.
- [15] D. Gamarnik and I. Zadik, “The overlap gap property and the power of QAOA,” arXiv:2111.06823 [cs.DS], 2022.
- [16] E. R. Anschuetz, D. Gamarnik and B. Kiani, “Combinatorial NLTS from the overlap gap property,” *Quantum* **8**, 1527 (2024).
- [17] S. Bravyi, D. Gosset and R. Koenig, “Quantum advantage with shallow circuits,” *Science* **362**, 308–312 (2018).
- [18] S. Hadfield, Z. Wang, B. O’Gorman, E. G. Rieffel, D. Venturelli and R. Biswas, “From the Quantum Approximate Optimization Algorithm to a Quantum Alternating Operator Ansatz,” *Algorithms* **12**(2), 34 (2019).
- [19] A. Bärtshi and S. Eidenbenz, “Grover mixers for QAOA: Shifting complexity from mixer design to state preparation,” arXiv:2006.00354 [quant-ph], 2020.

- [20] F. G. Fuchs and R. P. Bassa, “Constraint preserving mixers for the Quantum Approximate Optimization Algorithm,” *Algorithms* **15**(6), 202 (2022).
- [21] N. Xie, X. Lee, D. Cai, Y. Saito, N. Asai and H. C. Lau, “A feasibility-preserved quantum approximate solver for the capacitated vehicle routing problem,” arXiv:2308.08785 [quant-ph], 2024.
- [22] B. Tsvetikhovskiy, I. Safro and Y. Alexeev, “Symmetries and dimension reduction in Quantum Approximate Optimization Algorithm,” arXiv:2309.13787 [quant-ph], 2023.
- [23] B. Tsvetikhovskiy, I. Safro and Y. Alexeev, “Equivariant QAOA and the duel of the mixers,” arXiv:2405.07211 [quant-ph], 2024.
- [24] J. A. Montañez-Barrera, D. Willsch and K. Michielsen, “Transfer learning of optimal QAOA parameters in combinatorial optimization,” *Quantum Information Processing* **24**(5), 2025.
- [25] E. Bae and S. Lee, “Recursive QAOA outperforms the original QAOA for the MAX-CUT problem on complete graphs,” *Quantum Information Processing* **23**(3), 78 (2024).
- [26] J. R. Finžgar, A. Kerschbaumer, M. J. A. Schuetz, C. B. Mendl and H. G. Katzgraber, “Quantum-informed recursive optimization algorithms,” *PRX Quantum* **5**(2), 020327 (2024).
- [27] R. S. do Carmo, M. C. S. Santana, F. F. Fanchini, V. H. C. de Albuquerque and J. P. Papa, “Warm-starting QAOA with XY mixers: A novel approach for quantum-enhanced vehicle routing optimization,” arXiv:2504.19934 [quant-ph], 2025.
- [28] E. Osaba and E. Villar-Rodríguez, “QOPTLib: A quantum-computing-oriented benchmark for combinatorial optimisation problems,” in *Proc. Quantum Tech 2024*, 2024.
- [29] R. O’Donnell, *Analysis of Boolean Functions*, Cambridge University Press, 2014.
- [30] E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan and D. B. Shmoys, *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, Wiley, 1985.
- [31] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, 1979.
- [32] E. M. Loiola, N. M. M. de Abreu, P. O. Boaventura-Netto, P. Hahn and T. Querido, “A survey for the quadratic assignment problem,” *European Journal of Operational Research* **176**(2), 657–690 (2007).
- [33] P. Toth and D. Vigo (eds.), *Vehicle Routing: Problems, Methods, and Applications*, 2nd ed., SIAM, 2014.
- [34] S. Sahni and T. Gonzalez, “P-complete approximation problems,” *Journal of the ACM* **23**(3), 555–565 (1976).
- [35] R. M. Karp, “Reducibility among combinatorial problems,” in *Complexity of Computer Computations*, pp. 85–103, Plenum, 1972.
- [36] N. Alon and J. H. Spencer, *The Probabilistic Method*, 3rd ed., John Wiley & Sons, Hoboken, NJ, 2008.
- [37] H. Finner, “A generalization of Hölder’s inequality and some probability inequalities,” *The Annals of Probability* **20**(4), 1893–1901 (1992).
- [38] P. Feinsilver and J. Kocik, “Krawtchouk matrices from classical and quantum random walks,” arXiv:quant-ph/0702173, 2007.

- [39] J. A. Montañez-Barrera and K. Michielsen, “Towards a universal QAOA protocol: Evidence of a scaling advantage in solving some combinatorial optimization problems,” arXiv:2405.09169 [quant-ph], 2024.
- [40] A. Lucas, “Ising formulations of many NP problems,” *Frontiers in Physics* **2**, 5 (2014).
- [41] C. Onah and K. Michielsen, “Scalable hardware maturity probe for quantum accelerators via harmonic analysis of QAOA,” arXiv:2509.11450 [quant-ph], 2025.
- [42] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd ed., John Wiley & Sons, New York, 1999.
- [43] R. V. Kadison, “A generalized Schwarz inequality and algebraic invariants for operator algebras,” *Annals of Mathematics* **56**(3), 494–503 (1952).
- [44] V. I. Paulsen, *Completely Bounded Maps and Operator Algebras*, Cambridge University Press, Cambridge, 2002.