Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Empirical Quantum Advantage in Constrained Optimization from Encoded Unitary Designs

Chinonso Onah

conah@atom-computing.com

Volkswagen Group (Germany)

Kristel Michielsen

Forschungszentrum Jülich

Roman Firt

Volkswagen Group (Germany)

Research Article

Keywords:

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11396245/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Empirical Quantum Advantage in Constrained Optimization from Encoded Unitary Designs

Chinonso Onah[1,2], Roman Firt[1], and Kristel Michielsen[2,3]

[1]Volkswagen AG, Berliner Ring 2, Wolfsburg 38440, Germany
[2]Department of Physics, RWTH Aachen, Germany
[3]Forschungszentrum Jülich, Germany

We introduce the *Constraint–Enhanced Quantum Approximate Optimization Algorithm* (CE–QAOA), a shallow, constraint-aware ansatz that operates *inside* the one–hot product space $\mathcal{H}_{\mathrm{OH}} = [n]^m$, where m is the number of blocks and each block is initialized in an n-qubit W_n state. We give an ancilla-free, depth-optimal encoder that prepares W_n using $n-1$ two-qubit rotations per block, and a two-local block-XY mixer that preserves the one–hot manifold and has a constant spectral gap on the one-excitation sector. At the level of expressivity, we establish per-block controllability, implying approximate universality per block. At the level of distributional behavior, we show that (after natural block/symbol permutation twirls) shallow CE–QAOA realizes an encoded unitary 1-design and supports approximate second-moment (2-design) behavior; combined with a Paley–Zygmund argument, this yields finite-shot anticoncentration guarantees. Algorithmically, we wrap constant-depth sampling with a deterministic feasibility checker to obtain a polynomial-time hybrid quantum–classical solver (PHQC) that returns the best observed feasible solution in $O(Sn^2)$ time, where $S = \mathrm{poly}(n)$ is the shot budget. We obtain two advantages. First, when CE–QAOA fixes $r \geq 1$ locations different from the start city, we achieve a $\Theta(n^r)$ reduction in shot complexity even against a classical sampler that draws uniformly from the feasible set. Second, against a classical baseline restricted to raw bitstring sampling, we show an $\exp(\Theta(n^2))$ minimax separation. In noiseless circuit simulations of TSP instances with $n \in \{4, \ldots, 10\}$ locations from the QOPTLib benchmark library, we recover the global optimum at p=1 using polynomial shot budgets and coarse parameter grids defined by the problem size, as suggested by Theorem 13 and Algorithm 4.

[1]

1 Introduction

Combinatorial optimization problems (COP) underpin modern logistics, manufacturing, finance, and drug discovery; and it surfaces throughout artificial intelligence and machine-

[1]Data availability: A minimal Python implementation and integration into standard QAOA wrapper in Qiskit is made available https://doi.org/10.5281/zenodo.15725265

learning workflows [1, 2, 3, 4]. Classical methods like branch–and–bound and cutting planes [2, 5, 6] excel when they exploit problem structure. In contrast, many near-term quantum approaches—notably QAOA [7] and its variants [8, 9, 10]—tend to explore the full Hilbert space in search of the (near) optimal solution(s), paying in *measurement cost* [11, 12], *depth* [11, 12], and *barren plateaus* [13]. A growing line of work mitigates this by encoding directly into symmetry-rich feasible subspaces and using feasibility-preserving mixers [14, 15, 16, 17, 18, 19]. In this work we study *Constraint–Enhanced Quantum Approximate Optimization Algorithm* (CE–QAOA), a shallow, constraint-aware ansatz that *natively* operates on a block one-hot manifold. Our circuit constructions are based on an ancilla-free, depth-optimal encoder that prepares a block W_n state using exactly $n-1$ excitation preserving two qubit gates, and a two-local block–XY mixer that preserves the manifold.[14] The properties arising from these circuit primitives allow us to make contact with low-order moment analysis of random quantum circuits[20] in the encoded subspace and to subsequently study the emergence of approximate unitary $t-$ designs.

Why do unitary t-designs matter? In the context of variational quantum circuits, unitary designs are a principled proxy for *expressivity*, *anticoncentration*, and *typical-case trainability* [20]. Recall that a unitary t-design reproduces Haar moments up to degree t. Thus any degree-$\leq t$ observable (e.g., overlaps, energies, and many gradient statistics) computed under the design agrees with the Haar baseline up to ε [20]. Secondly, for $t = 2$, 2-designs suggests an *instance-agnostic typical-case behaviour* because second moments are insensitive to the fine details of the cost landscape. From the "landscape information" viewpoint [21], matching Haar second moments ensures that a nontrivial amount of task-relevant variance survives at shallow depth on the *encoded* space, which increases the effective information capacity of the ansatz [21, 22].

Our circuit design inherits from prior work in Refs. [14, 23]. Here we leverage additional problem specific structures through problem-algorithm codesign to capture a broad class of COPs (TSP/ATSP, QAP, CVRP, etc) via block one-hot encodings and shared constraint properties. Consequently, we propose a unifying kernel in Def. 1. Within this kernel, a block-permutation twirl yields a unitary 1-design on the encoded space, reproducing the Haar baseline $\mathbb{E}\,|\langle x|U|\phi\rangle|^2 = 1/D$ and a corresponding *existence* theorem for any target basis vector $x^\star$ (and thus for the optimum) in Thm. 13 and Cor. 16. Algorithmically, we introduce the the Polynomial time Hybrid Quantum-Classical Solver (PHQC) which wraps constant-depth quantum sampling over a coarse parameter grid with a deterministic classical checker that identifies the best sampled solutions, operationalizing these $t-$design bounds. PHQC identifies the optimal solution if it is sampled at least once.

Our running examples are derived from the Travelling Salesman Problem (TSP) where the n^2 qubits are imagined to be arranged in n blocks of size n and the double one-hot constraint restricts to a fixed-Hamming-weight manifold containing only $n!$ basis states out of the encoded n^n encoded basis vectors and 2^{n^2} vectors in the ambient Hilbert space [24]. Similar one-hot encoded problems arise naturally for assignment problems [25], matching problems [26], vehicle routing problems [3], graph coloring problems [24] and the recently introduced shared transportation framework [27]. CE–QAOA exploits this structure at the encoder, mixer layers, phase layers, and classical feasibility checker; extending problem–algorithm co-design from quantum to classical stage [28, 29]. The advertised quantum advanatge comes from direct utilization of problem structures at the classical and quantum levels of the algorithm.

We present noiseless circuit simulation results for TSP instances ranging from 4 to 10 locations drawn from the QOPTLib benchmark problems[30]. We demonstrate that PHQC

recovered the optimal solution in all instances considered while dramatically reducing the shot cost to a polynomial $O(n^k)$ at depth $p = 1$ as reported in Table 1. All plots presented in the following can be reproduced as demonstrated in [31].

1.1 Relation to Prior Work

Recently, Smith–Miles *et al.* [32] surveyed the obstacles to near-term quantum advantage on TSP—including resource inflation from generic encodings (e.g., QUBO), feasibility handling, noise/optimizer pathologies, and inconsistent benchmarking against strong classical solvers—and argue for structure-exploiting formulations and transparent, reproducible evaluations. Our CE–QAOA/PHQC approach aligns with this perspective but differs in emphasis because we treat global constraints as *useful structure* within an encoded manifold. Our formalism inherits from the preexisiting body of work in the *alternating-operator* literature (often called QAOA+) which develops mixers that commute with, or otherwise preserve, constraint projectors—e.g., one-hot, ring mixers, and null-swaps [14, 33]. For routing families (e.g., CVRP), feasibility is frequently enforced via Grover-style oracles [19, 16, 34], which introduce oracle/ancilla overhead and reflection-based subroutines. In contrast, our kernel uses a *two-local, ancilla-free* block–XY mixer that *preserves* all blockwise one-hot constraints, acts entirely within the encoded manifold, and is normalized to have a *constant* spectral gap on the one-excitation sector (Prop. 7). The initial state is the block-W product[35] prepared by an ancilla-free, depth-optimal cascade (Thm. 4), making the mixer's top eigenvector the initial state. Our mixer design ensures *invariance* (excitation conservation) and block factorization, enabling quditization $H_M^{(b)} \mapsto A(K_n)$[14] and explicit control of norms and gaps at the encoded per block level. This normalization is used quantitatively in our mixing/expressivity arguments and avoids hidden n-dependent angles that can plague unnormalized XY implementations.

Unlike prior feasibility-preserving approaches (e.g.QAOA+ [14, 16, 34]), our kernel *explicitly* requires that the *constraints* be exchangeable under block permutations S_m and symbol relabelings S_n. This symmetry is a co-design choice that is is enforced at the modeling level and then exploited algorithmically. Two consequences are central to our analysis: (i) H_{pen} is invariant under $S_m \times S_n$, so every penalty level set $L_t = \{x : H_{\text{pen}}(x) := \langle x \mid H_{\text{pen}} \mid x \rangle = t\}$ (in particular L_0) is preserved setwise; and (ii) a block-permutation twirl acts transitively on the encoded basis and yields an *exact unitary* 1*-design* on $\mathcal{H}_{\text{OH}}$ (Lemma 15). This exact first-moment identity directly powers our *existence* guarantee for the optimal basis vector $x^\star$ (Cor. 16). This symmetry-driven 1-design baseline, and the resulting existence results, differentiate our co-designed kernel from prior feasibility-preserving approaches.

Although our kernel uses familiar primitives, here we study them under the lens of norms, spectral properties, and typical case behaviors. By systematic use of *low-order moment* methods (L_1, L_2) together with symmetry and twirling, we expose and utilize: **(i)** Lie controllability and approximate universality within the one-excitation sectors induced by the block–XY mixer; **(ii)** *ε-approximate unitary 2-designs at shallow depth* arising from per-block design plus diagonal entanglers; **(iii)** *exponentially fast mixing* and *anticoncentration* on the encoded manifold; and **(iv)** expressivity that translates into provable lower tails at the c/D scale. Where D is the encoded dimension and $c > 0$ from Thm. 13. To our knowledge, prior feasibility-preserving or domain-specific QAOA variants do not provide *universal, instance-agnostic* guarantees of this kind.

Our analyses remain complementary to recent work on *landscape information content* [21]; and analysis of energy/gradient landscapes and alternating-operator structure [22]. Our ap-

proach equally remains orthogonal to warm-start schedules [36, 35]. Warm starts can yield strong numerics but provide limited worst-case control and sometimes require non-eigen initializations of the mixer, violating the adiabatic protocol and limiting performance[37]. We avoid classical initialization completely and focus on provable algorithmic performance.

2 Codesign and Circuit Primitives for CE-QAOA

2.1 Problem-Algorithm Codesign

Our goal is to identify potential opportunities for provable algorithmic enhancements by taking explicit advantage of *problem structure*. This aligns with the broader *co-design* perspective in quantum computing, where algorithms and platforms are jointly tailored to application constraints and device characteristics [28, 29]. Here, however, we omit hardware considerations and focus on problem-algorithm co-design. We look beyond *feasibility-preserving* ansätze and mixers [14, 38, 15, 19, 34] and impose additional symmetry on the constraint structure consistent with the various problem classes, bringing them under a single kernel defined in Def. 1. As a consequence, performance bounds obtained propagate across a broad class of COPs with minimal tweaks. The definition below fixes what it means to be "in the kernel" and will be the used henceforth for all results that follow.

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 $\mathsf{E}_{\mathrm{1hot}}$ that initializes the dynamics in the fixed–Hamming–weight space*

$$\mathcal{H}_{\mathrm{OH}} \;=\; (\mathcal{H}_1)^{\otimes m}, \qquad \mathcal{H}_1 \;=\; \mathrm{span}\{|e_0\rangle, \ldots, |e_{n-1}\rangle\} \quad \textit{(one excitation per block).}$$

The problem Hamiltonian splits as

$$H_C \;=\; H_{\mathrm{pen}} \;+\; H_{\mathrm{obj}},$$

Where H_{obj} is the Ising Hamiltonian representing the objective and only needs to be diagonal in the computational basis. H_{pen} is the penalty Hamiltonain and enjoys the symmetries specified in (a) and (b) in addition to being diagonal in the computational basis.

(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* $\mathrm{poly}(n)$*. Consequently,* $\mathrm{spec}(H_{\mathrm{pen}}) \subseteq \{0, 1, \ldots, t_{\max}\}$ *with* $t_{\max} = O(m) = \mathrm{poly}(n)$.

(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 $L_t = \{x : H_{\mathrm{pen}}(x) := \langle x \mid H_{\mathrm{pen}} \mid x\rangle = t\}$ that are preserved setwise.*

(c) *In addition, the initial state is the $+1$ eigenstate of the block–local normalized XY mixer Hamiltonian,*

$$\widetilde{H}_M^{(b)} \;=\; \frac{1}{n-1} \sum_{0 \le j < k \le n-1} (X_j^{(b)} X_k^{(b)} + Y_j^{(b)} Y_k^{(b)}),$$

with $\|\widetilde{H}_M^{(b)}\| = O(1)$ on each block. The initial state is the uniform one–hot product

$$|s_0\rangle \;=\; |s_{\mathrm{b}}\rangle^{\otimes m}, \qquad |s_{\mathrm{b}}\rangle \;=\; \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} |e_k\rangle \quad \textit{(a } W_n \textit{ state per block).}$$

Although the kernel above fixes one-hot blocks (1 excitation per block), many applications naturally use *k-hot* blocks (fixed Hamming weight of k per block). In that case, one replaces $\mathcal{H}_1$ by $\mathcal{H}_k = \mathrm{span}\{|x\rangle \in \{0,1\}^n : \|x\|_1 = k\}$, and chooses a suitable *number-conserving* block-local mixer (e.g. modified XY Hamiltonian, which preserves excitation

number) normalized to constant operator norm on $\mathcal{H}_k$. The corresponding initial state becomes a tensor product of Dicke states encoders [39],

$$\left|s_{\mathrm{b}}^{(k)}\right\rangle \;=\; \left|D_k^{(n)}\right\rangle \;=\; \binom{n}{k}^{-\frac{1}{2}} \sum_{\substack{x\in\{0,1\}^n \\ \|x\|_1=k}} |x\rangle\,, \qquad |s_0\rangle \;=\; \left(\left|D_k^{(n)}\right\rangle\right)^{\otimes m}.$$

This further enlarges the class of constrained combinatorial optimization problems that can fit into the kernel proposed above. Some identified problem classes include: Travelling Salesman (TSP/ATSP) [40, 41], Quadratic Assignment Problem (QAP) [42, 43], Vehicle Routing with Capacities (CVRP) [3], Generalized Assignment / Multiple–Knapsack [44, 45], kD Matching (NP-complete for $k \geq 3$) [46, 47], the shared transportation problems[27], and Job–Shop / Flow–Shop Scheduling [48, 49]. Some base cases (e.g., linear assignment) are polynomial-time solvable via the Hungarian algorithm [26, 50]; nevertheless, they fit the kernel structurally. Our interest is in the NP-hard variants (e.g., TSP/QAP/CVRP, etc.) for which the same one-hot encodings and symmetries apply.

Definition 2 (Feasible basis vectors)**.** *Let $\mathcal{H}_{\mathrm{OH}} = \mathcal{H}_1^{\otimes m}$ denote the encoded one–hot product space from Definition 1, and let*

$$H_C \;=\; H_{\mathrm{pen}} + H_{\mathrm{obj}}$$

be the problem Hamiltonian with H_{pen} satisfying the kernel symmetries. A computational basis vector $|x\rangle$ with block decomposition $x = (j_0, \ldots, j_{m-1}) \in [n]^m$ is called feasible *if and only if*

$$H_{\mathrm{pen}}(x) := \langle x \mid H_{\mathrm{pen}} \mid x\rangle \;=\; 0,$$

i.e. $|x\rangle$ lies in the zero–penalty level set

$$L_0(H_{\mathrm{pen}}) \;:=\; \{\, |x\rangle \in \mathcal{H}_{\mathrm{OH}} :\; H_{\mathrm{pen}}(x) = 0 \,\}.$$

Remark 3. In the standard one–hot encodings for permutation and routing problems, each basis vector $|x\rangle$ can be reshaped into an $n \times n$ binary matrix X with exactly one 1 in each row. A basis vector is feasible precisely when this reshaped matrix is a permutation matrix:

$$X \in S_n \qquad \Longleftrightarrow \qquad \sum_{k=0}^{n-1} X_{ik} = 1 \text{ for all rows } i \;\text{ and }\; \sum_{i=0}^{n-1} X_{ik} = 1 \text{ for all columns } k.$$

Thus the feasible basis vectors are exactly those $x \in \mathcal{H}_{\mathrm{OH}}$ whose blockwise one–hot assignments are globally collision–free and satisfy all penalty constraints, i.e. the elements of the permutation manifold $L_0(H_{\mathrm{pen}}) \subset \mathcal{H}_{\mathrm{OH}}$.

2.2 The Initial State Preparation Protocol

Consider some $n \times m$–qubit system where n can be considered of a fixed size repeated m times. Suppose each block has n qubits, and consider a single excitation subspace spanned by the equal superposition $|s\rangle = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} |e_k\rangle$. Over the m blocks, the final state is:

$$|s_0\rangle \;=\; \bigotimes_{b=0}^{m-1} \left(\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} |e_k\rangle \right), \tag{1}$$

realising block-wise tensor product of uniform W-states [51, 39]. Algorithm 1, a simple cascade-style algorithm can prepare the initial state described above in Eq 1; obtaining a single block of n qubits in a W-state. Then Algorithm 2 (`OneHotMultiBlockPrepare`(m)) uses the single-block procedure on each of the m blocks to create a W-state per block. The total Hamming weight per bitstring is fixed and the Hamming Geometry is constrained by

the location of single excitation contributed by each block. This circuit construction scales optimally in gate depth on a linear array.

Theorem 4 (Gate–depth optimality on a line)**.** *There exists an ancilla-free circuit that prepares* $|W_n\rangle$ *on a linear array using* $n-1$ *two-qubit* excitation-preserving rotations

$$U_M^{(k,k+1)}(\phi) \;:=\; \exp\Big[-\,i\,\frac{\phi}{2}\,(X_kX_{k+1}+Y_kY_{k+1})\Big],$$

on each adjacent pair $(k,k{+}1)$*, with suitable angles* $\{\phi_k\}_{k=0}^{n-1}$*. Moreover, any circuit over single-qubit gates and two-qubit gates on a linear array that prepares* $|W_n\rangle$ *must use at least* $n-1$ *two-qubit gates.*

Proof. Label qubits $0,1,2,\ldots,n-1$ on a line and, for each $k=0,\ldots,n-1$, consider the bipartition $\{0,\ldots,k\}\,|\,\{k+1,\ldots,n-1\}$. For the uniform W-state $|W_n\rangle = \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}|0\cdots1_k\cdots0\rangle$, the Schmidt rank across *every* such single-edge cut is $\mathrm{SR}_k(|W_n\rangle)=2$. The product state $|0\rangle^{\otimes n}$ has rank 1 across every cut. A two-qubit gate that does *not* cross a given cut cannot increase that cut's rank, while a two-qubit gate that *does* cross it can increase the rank by at most a factor of 2. Thus each of the $n-1$ adjacency cuts must be crossed at least once by a two-qubit gate, so any preparation requires $\geq n-1$ two-qubit gates. Alg. 1 gives a construction that achieves the bound. □

Remark 5 (Hardware-friendly alternatives). On platforms with native iSWAP$^\alpha$ or fSim(θ,φ), $U_M^{(k,k+1)}(\phi)$ is implementable in one entangling gate (up to single-qubit Z phases). If $XX{+}YY$ is not native, one can realize the same two-level rotation on the preparation path with the single 2Q primitive

$$\tilde{U}(\theta) \;:=\; \mathrm{CX}_{k+1\to k}\;\cdot\;\mathrm{CRY}_{\,k\to k+1}(\theta)\quad\text{with}\quad\theta=2\phi,$$

which has the block action $|10\rangle \mapsto \cos\!\left(\frac{\theta}{2}\right)|10\rangle + \sin\!\left(\frac{\theta}{2}\right)|01\rangle$ on $\mathrm{span}\{|10\rangle,|01\rangle\}$.

Algorithm 1 `OneHotBlockPrepare`(n) — exact single–excitation W_n

Require: number of qubits $n\geq 2$ labelled $0,\ldots,n{-}1$

1: For adjacent qubits $(k,k{+}1)$, let

$$\mathrm{U}^{(k,k+1)}(\theta) \;:=\; \exp\Big[-\,i\,\tfrac{\theta}{2}\,(X_kX_{k+1}+Y_kY_{k+1})\Big]$$

▷ acts as a real Givens rotation on $\mathrm{span}\{|10\rangle,|01\rangle\}$

(Hardware alternative: $\mathrm{U}^{(k,k+1)}(\theta)=\mathrm{CX}_{k+1\to k}\cdot\mathrm{CRY}_{\,k\to k+1}(\theta)$*)*

2: Apply X on qubit 0 ▷ $|100\ldots0\rangle$

3: **for** $k\leftarrow 0$ to $n-2$ **do**

4: $\quad\theta_k \leftarrow 2\arccos\!\Big(\frac{1}{\sqrt{n-k}}\Big)$

5: $\quad$Apply $\mathrm{U}^{(k,k+1)}(\theta_k)$ ▷ fixes amplitude on site k to $1/\sqrt{n}$

6: **end for**

Ensure: $|W_n\rangle = \frac{1}{\sqrt{n}}\displaystyle\sum_{k=0}^{n-1}|0\ldots1_k\ldots0\rangle$

Algorithm 2 `OneHotMultiBlockPrepare`(m) — tensor product of m W-blocks

Require: block size n, number of blocks m

1: **for** $b \leftarrow 0$ to $m-1$ **do**

2: Apply OneHotBlockPrepare(n) on qubits $bn,\ bn{+}1, \ldots, (b{+}1)n-1$

3: **end for**

Ensure: $(|W_n\rangle)^{\otimes m} = \dfrac{1}{\sqrt{n^m}} \displaystyle\sum_{\substack{k_0,\ldots,k_{m-1} \\ \in\{0,\ldots,n-1\}}} \bigotimes_{b=0}^{m-1} |0\ldots 1_{k_b} \ldots 0\rangle$

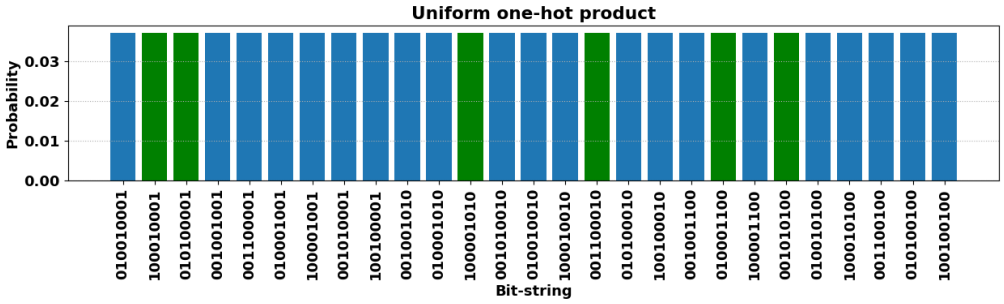


Figure 1: Histogram produced in Qiskit[52] for a TSP problem on 3 locations. Each of the $3^3 = 27$ block–one-hot bit-strings carries probability $1/27$ verifying that the state–preparation algorithm populates all basis states with equal amplitude while leaving probability mass strictly inside the one-hot subspace. The six strings whose block indices form a permutation are coloured green.

It follows from well known results in quantum circuit simulation that the initial state prepared by Algorithm 2 is non-stabilizer for $n \geq 3$[53] because each block contains a W_n-state. We cite this result in the following proposition.

Proposition 6 (W_n is non-Clifford [53])**.** *For $n \geq 3$ the n-qubit W state $|W_n\rangle = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} |e_k\rangle$ is not a stabilizer state. In particular, it cannot be prepared from $|0\rangle^{\otimes n}$ using Clifford operations alone. (The edge case $W_2 = (|01\rangle + |10\rangle)/\sqrt{2}$ is stabilizer.) Outside the stabilizer regime, strong simulation typically relies on non-Clifford techniques which are not known to run in polynomial time in general [53].*

It is thus reasonable to expect that our encoder already places the ansatz beyond the efficiently simulable stabilizer regime, strengthening the potential for quantum–classical separation discussed in Sections 3.3 and 4.3. Fig.1 shows the basis vectors forming a product of one-hot states in the reduced Hilbert space with bitstrings representing valid permutations marked in green. Fig. 2 is a single layer CE-QAOA circuit for a 12 qubit problem with 3 blocks[31]. [2]

2.3 The XY Mixer Operator

Recall that the task in QAOA+ is to construct, for the initial state, a suitable mixer Hamiltonian such that the initial state is its +1 eigenstate[18, 17]. Thus, we require an XY mixer Hamiltonian that preserves the one–hot subspace. For each n–qubit block b we

[2]Data availability: A minimal Python implementation and integration into standard QAOA wrapper in Qiskit is made available https://doi.org/10.5281/zenodo.15725265

define (up to constant prefactor)

$$H_M^{(b)} = \sum_{0 \le i < j \le n-1} \left(X_i^{(b)} X_j^{(b)} + Y_i^{(b)} Y_j^{(b)} \right), \tag{2}$$

where $X_i^{(b)}$ and $Y_i^{(b)}$ denote the Pauli X and Y operators acting on the ith qubit in block b. The overall mixer unitary is given by

$$U_M(\beta) = U_M^{(0)}(\beta) \otimes \cdots \otimes U_M^{(m-1)}(\beta). \tag{3}$$

Where $U_M^{(b)}(\beta) = e^{-i\beta H_M^{(b)}}$ for each block. The construction is outlined in Algorithm 3. See Fig. 2 for the blockwise construction depicted in a full circuit.

In terms of the dynamics, the XY Mixer continuously mixes the valid one–hot basis states among themselves thanks to its adjacency and ergodicity properties on the $\mathcal{H}_1$ subspace. To expose these properties, it is helpful to rewrite the mixer Hamiltonian in a form that is manifestly 2-local and preserves the excitation number. The mixer Hamiltonian is:

$$H_M^{(b)} := \sum_{0 \le i < j \le n-1} (X_i^{(b)} X_j^{(b)} + Y_i^{(b)} Y_j^{(b)}) = \sum_{i<j} (\sigma_{bi}^+ \sigma_{bj}^- + \sigma_{bi}^- \sigma_{bj}^+), \tag{4}$$

Restricting attention to the one-excitation subspace $\mathcal{H}_1 = \mathrm{span}\{|e_0\rangle, \ldots, |e_{n-1}\rangle\}$, and normalising by n gives $\widetilde{H}_M^{(b)} := H_M^{(b)}/n$ with *constant* spectral gap. (cf. App A.3).

Algorithm 3 BlockXYMixer (n, m, β) — two-local mixer for the one-hot subspace

Require: Block size n, number of blocks m, mixer angle β

for $b \leftarrow 0$ **to** $m-1$ **do**
  iterate over blocks
  for all (i, j) **with** $0 \le i < j < n-1$ **do**
    apply two-local mixer on qubit pair (i, j) *of block* b
    $\mathrm{RXX}_{bn+i,bn+j}(2\beta)$
    $\mathrm{RYY}_{bn+i,bn+j}(2\beta)$
  end for
end for

Ensure:

$$U_M(\beta) = \exp\Big[-i\beta \sum_{b=0}^{m-1} \sum_{0 \le i < j < n-1} \left(\frac{1}{n-1} \sum_{0 \le j < k \le n-1} (X_j^{(b)} X_k^{(b)} + Y_j^{(b)} Y_k^{(b)}) \right)\Big]$$

Proposition 7 (Spectral gap of one-block XY mixer)**.** *On* $\mathcal{H}_1$ *the operator* $H_M^{(b)}$ *acts as the adjacency matrix* $A(K_n)$ *of the complete graph on* n *vertices and has spectrum*

$$\mathrm{spec}(H_M^{(b)} \restriction_{\mathcal{H}_1}) = \{\, n-1, \underbrace{-1, \ldots, -1}_{n-1 \text{ times}} \}.$$

Hence the spectral gap is $\Delta(H_M^{(b)}) = n$. $\widetilde{H}_M^{(b)} = H_M^{(b)}/n$ *has constant gap* $\Delta(\widetilde{H}_M^{(b)}) = 1$.

Proof. Acting on a basis state with the excitation at site i with Eq. 4 (See App. A.3),

$$(X_i X_j + Y_i Y_j)\,|e_i\rangle = \sigma_i^- \sigma_j^+ \,|e_i\rangle = |e_j\rangle\,,$$

while the same operator annihilates $|e_k\rangle$ for $k \notin \{i, j\}$. Summing over all unordered pairs (i, j) therefore maps $|e_i\rangle \mapsto \sum_{j \ne i} |e_j\rangle$, which is exactly the action of the adjacency matrix $A(K_n)$ on the standard vertex basis. The spectrum of $A(K_n)$ is well known from spectral graph theory [54]. Largest eigenvalue $\lambda_{\max} = n-1$ (eigenvector $\frac{1}{\sqrt{n}} \sum_k |e_k\rangle$) and the remaining $n-1$ eigenvalues equal -1. Hence $\Delta(H_M^{(b)}) = (n-1) - (-1) = n$. Dividing

by n rescales both extremal eigenvalues by the same factor, so $\Delta(\widetilde{H}_M^{(b)}) = 1$. The claimed *constant* spectral gap and K_n denotes its complete adjacency graph. □

Proposition 8 (Ergodicity of the angle-averaged XY mixer on $\mathcal{H}_1$)**.** *Consider a single n-qubit one-hot block with one-excitation sector $\mathcal{H}_1 = \mathrm{span}\{|e_0\rangle, \ldots, |e_{n-1}\rangle\}$. Let $H_M^{(b)} \restriction_{\mathcal{H}_1} = A(K_n)$ be the restriction of the all-to-all unnomalised XY Hamiltonian to $\mathcal{H}_1$ (equivalently, the adjacency matrix of the complete graph on n vertices up to an overall scalar). For $\beta \in \mathbb{R}$ define $U(\beta) := e^{-i\beta H_M^{(b)}}$ and the transition matrix*

$$P_{ij} := \int_0^{2\pi} \frac{\mathrm{d}\beta}{2\pi} \, |\langle e_j \,|\, U(\beta) \,|\, e_i \rangle|^2 \,, \qquad 0 \le i, j \le n-1.$$

Then:

1. *P is* primitive *(all entries are strictly positive), hence the associated Markov chain is* ergodic *(irreducible and aperiodic).*
2. *P is* doubly stochastic*, and its unique stationary distribution is the uniform distribution $\pi^\star = (1/n, \ldots, 1/n)$.*
3. *Explicitly,*

$$P_{ii} \;=\; 1 - \frac{2}{n} + \frac{2}{n^2}, \qquad P_{ij} \;=\; \frac{2}{n^2} \quad (j \neq i),$$

so that $P^t \to \mathbf{1}\,\pi^{\star\top}$ as $t \to \infty$.

These conclusions are invariant under any nonzero rescaling $H_M^{(b)} \mapsto c\,H_M^{(b)}$, $c \in \mathbb{R} \setminus \{0\}$.

Proof. On $\mathcal{H}_1$, $H_M^{(b)}$ has spectral decomposition $H_M^{(b)} = (n-1)\Pi_s + (-1)\Pi_\perp$, where $\Pi_s = |s\rangle\langle s|$ with $|s\rangle = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} |e_k\rangle$ and $\Pi_\perp = I - \Pi_s$. Hence

$$U(\beta) = e^{-i\beta(n-1)}\Pi_s \;+\; e^{+i\beta}\Pi_\perp.$$

For basis states, $\langle e_j \,|\, U(\beta) \,|\, e_i \rangle = e^{-i\beta(n-1)}\frac{1}{n} + e^{i\beta}(\delta_{ij} - \frac{1}{n})$.

Averaging $|\cdot|^2$ over $\beta \in [0, 2\pi)$ removes cross terms (orthogonal Fourier modes), giving

$$P_{ij} = \frac{1}{n^2} + \Big(\delta_{ij} - \frac{1}{n}\Big)^2 = \begin{cases} 1 - \dfrac{2}{n} + \dfrac{2}{n^2}, & j = i, \\[2ex] \dfrac{2}{n^2}, & j \neq i. \end{cases}$$

All entries are strictly positive, so P is primitive (hence ergodic). Row and column sums are 1 by symmetry, so P is doubly stochastic and the uniform distribution is the unique stationary distribution. Convergence $P^t \to \mathbf{1}\,\pi^{\star\top}$ follows from primitivity. If $H_M^{(b)}$ is rescaled by a $c > 0$, the averaged cross terms still vanish and the same P formula holds. In particular it holds for $\widetilde{H}_M^{(b)}$. □

Local unitary designs make extensive use of qudit language. Since t-designs are central to our methods, following Ref. [14], we formulate the qudit ↔qubit isometry below.

Definition 9 (Encoding isometry (*qubit ↔ qudit*))**.** *Let $\mathbb{C}^n$ be the abstract qudit space with basis $\{|j\rangle\}_{j=0}^{n-1}$. Define the isometry*

$$V : \; \mathbb{C}^n \longrightarrow (\mathbb{C}^2)^{\otimes n}, \qquad V\,|j\rangle \;=\; |e_j\rangle\,.$$

Then $V^\dagger V = I_n$ and $VV^\dagger$ is the projector onto $\mathcal{H}_1$.

Proposition 10 (Invariance and quditization of the block–XY mixer)**.** *Let $H_M^{(\mathrm{b})} = \sum_{0 \le i < j \le n-1} (X_i^{(b)} X_j^{(b)} + Y_i^{(b)} Y_j^{(b)}) = \sum_{i \neq j} \sigma_{bi}^- \sigma_{bj}^+$ on the n qubits of a block. Then:*

1. *$\mathcal{H}_1$ is invariant under $H_M^{(\mathrm{b})}$ and $U^{(\mathrm{b})}(\beta) := e^{-i\beta H_M^{(\mathrm{b})}}$.*

2. *In the encoded qudit picture,*

$$V^\dagger H_M^{(\mathrm{b})} V \;=\; \sum_{i\neq j} |i\rangle\langle j| \;=\; A(K_n),$$

the adjacency matrix of the complete graph K_n. Consequently, $U^{(\mathrm{b})}(\beta) = V\, e^{-i\beta\, A(K_n)}\, V^\dagger$ on $\mathcal{H}_1$.

Proof. For any single–excitation basis vector $|e_j\rangle$, $\sigma_i^-\sigma_j^+ |e_j\rangle = |e_i\rangle$ and $\sigma_i^-\sigma_j^+ |e_k\rangle = 0$ if $k \notin \{i,j\}$. Thus $H_M^{(\mathrm{b})}$ maps $|e_j\rangle$ to a superposition of $\{|e_i\rangle\}$, so $\mathcal{H}_1$ and $U^{(\mathrm{b})}(\beta)$ are invariant. Moreover, $V^\dagger\sigma_i^-\sigma_j^+ V = |i\rangle\langle j|\,, \quad (i \neq j)$, hence the claimed identification with $A(K_n)$. □

2.4 Constraint-Enhanced QAOA Protocol

The next key ingredient is the cost Hamiltonian H_C that represents the optimization problem and constraints. This is usually the Ising Hamiltonian corresponding to the problem and thus diagonal in the computational basis. The full circuit with p layers alternates between the cost and mixer unitaries a set of parameters (respectively) $\vec{\gamma} = (\gamma_1, \ldots, \gamma_p)$ and $\vec{\beta} = (\beta_1, \ldots, \beta_p)$. Algebraically, the circuit is represented by the following:

$$|\psi_p(\vec{\gamma}, \vec{\beta})\rangle = \left(\prod_{l=1}^{p} e^{-i\beta_l H_M} e^{-i\gamma_l H_C} \right) |s_0\rangle. \tag{5}$$

The Alg. 2 prepares $|s_0\rangle$ and initializes the quantum dynamics in the constrained fixed–Hamming–weight space $\mathcal{H}_{\mathrm{OH}} = (\mathcal{H}_1)^{\otimes m}$. Because the cost Hamiltonian H_C is diagonal in the computational basis, we have $U_C(\gamma)|x\rangle \in \mathcal{H}_{\mathrm{OH}}$ for all product basis states $|x\rangle \in \mathcal{H}_{\mathrm{OH}}$ and it never mixes amplitudes or leaks outside the one-hot subspace. The overall depth increase from our construction, if any, is minimal. The following holds.

Corollary 11 (Block size limits circuit depth)**.** *Let the block-structured QAOA ansatz consist of m disjoint blocks, each acting on n qubits. Then, under the assumption of parallel execution across blocks, the circuit depth of the full initial state preparation unitary is $\mathcal{D}_{\mathrm{prep}} = \mathcal{D}_{\text{block-prep}}$, and the depth of the full mixer unitary is $\mathcal{D}_{\mathrm{mix}} = \mathcal{D}_{\text{block-mix}}$. Hence, the overall depth overhead is independent of the number of blocks m.*

Resource savings from co-design. Suprisingly, the co-designed restriction to the one-hot manifold can *reduce* quantum resources relative to the standard QAOA formulation[7]. For permutation-constrained problems (e.g., TSP), feasibility imposes a double one-hot structure (row and column constraints) on the permutation matrix. In CE–QAOA, the chosen initial state and block-XY mixer preserve the row-wise one-hot condition by construction, so those constraint terms are redundant and can be dropped from the diagonal phase operator; only the complementary (e.g., column-wise) constraints need to be enforced. This eliminates an entire family of penalty interactions and yields an $O(n)$ reduction in the depth (and two-qubit gate count) of the phase-separation unitary compared to standard QAOA implementations. Corollary 11 already shows that the additional overhead from the proposed state preparation and mixing does not scale with the number of blocks m under parallel execution, so the net effect can be a *strict* depth reduction. For example, the $O(n)$ savings from removing redundant penalties can dominate the $O(n)$ preparation cost on architectures with native (or efficiently compiled) excitation-preserving rotations leading to an overall circuit depth smaller than that of a generic QAOA.

Crucially, these resource savings do not come from limiting the reachable set of feasible states. On the contrary, in the next proposition we show that the same block-local, constraint-preserving operations are controllable (and approximately universal) *within each block.*

Proposition 12 (Controllability and approximate universality on $\mathcal{H}_1$)**.** *Fix a single n-qubit one-hot block and the operator set*

$$\mathcal{G} := \{ H_{ij}^{(b)} = X_i^{(b)} X_j^{(b)} + Y_i^{(b)} Y_j^{(b)} : 0 \le i < j \le n-1, 0 \le b \le m-1 \} \cup \{ H_Z^{(1)} \},$$

where $H_Z^{(1)}$ is any diagonal excitation-preserving term whose restriction to $\mathcal{H}_1$ is not *proportional to the identity. Restricted to the one-excitation sector $\mathcal{H}_1 = \mathrm{span}\{|e_0\rangle, \ldots, |e_{n-1}\rangle\}$, the dynamical Lie algebra generated by $i\mathcal{G}$ is $\mathfrak{su}(n)$. Consequently, for any $V \in \mathrm{SU}(n)$ and any $\varepsilon > 0$ there exist angles and a finite product $\prod_{\ell=1}^{L} e^{-i\theta_\ell H_{\alpha_\ell}}$ with $H_{\alpha_\ell} \in \mathcal{G}$ such that*

$$\Big\| V - \prod_{\ell=1}^{L} e^{-i\theta_\ell H_{\alpha_\ell}} \Big\| \le \varepsilon.$$

Proof sketch(See App A.4). Work in the one-excitation basis $\{|e_k\rangle\}_{k=0}^{n-1}$ of $\mathcal{H}_1$. On this sector, each XY coupling acts as a symmetric matrix unit, $H_{ij}\,|_{\mathcal{H}_1} = E_{ij} + E_{ji}$, while the diagonal excitation-preserving term restricts to a non-scalar diagonal $D = H_Z^{(1)}\,|_{\mathcal{H}_1} = \sum_k d_k E_{kk}$ with $D \not\propto I$. The key observation is that commutators with D turn symmetric off-diagonals into skew-symmetric ones: whenever $d_i \neq d_j$,

$$[D,\, E_{ij} + E_{ji}] = (d_i - d_j)(E_{ij} - E_{ji}),$$

so the Lie closure contains both $E_{ij} + E_{ji}$ and $E_{ij} - E_{ji}$ for at least one pair. Commuting these then produces traceless diagonals,

$$[E_{ij} + E_{ji},\, E_{ij} - E_{ji}] = 2(E_{ii} - E_{jj}),$$

and iterating standard bracket identities yields the full space of traceless diagonal operators on $\mathcal{H}_1$.

Finally, using any traceless diagonal that separates k and ℓ,

$$[E_{kk} - E_{\ell\ell},\, E_{k\ell} + E_{\ell k}] = 2(E_{k\ell} - E_{\ell k}),$$

one obtains skew-symmetric off-diagonals for every (k, ℓ). Hence the Lie closure contains the usual spanning set of $\mathfrak{su}(n)$ given by symmetric off-diagonals, skew-symmetric off-diagonals, and traceless diagonals, so $\mathrm{Lie}(i\mathcal{G})\,|_{\mathcal{H}_1} = \mathfrak{su}(n)$. Standard controllability then implies that products of $\exp(-i\theta H)$ with $H \in \mathcal{G}$ can approximate any target $V \in \mathrm{SU}(n)$ to arbitrary precision. □

3 Encoded Unitary Designs

Roadmap. In §2.1 we set up the problem–algorithm co-design and formalized the *CE–QAOA kernel*, fixing the encoded space, symmetry assumptions, mixer normalization, and initial states. In §3.1 we show that blockwise label permutations induce an *exact unitary 1-design* on the encoded space, yielding a clean $1/D$ baseline and an averaging-to-existence guarantee for favorable angles (and explaining heavy-output peaks when no twirl is applied). In §3.2 we lift this to ε-approximate *2-design* control via per-block XY controllability combined with diagonal cross-block entanglers provides second-moment (anticoncentration) bounds at poly depth. In §3.3 we benchmark against classical procedures under two ambient domains, isolating the D/S vs. $2^{n^2}/S$ baselines

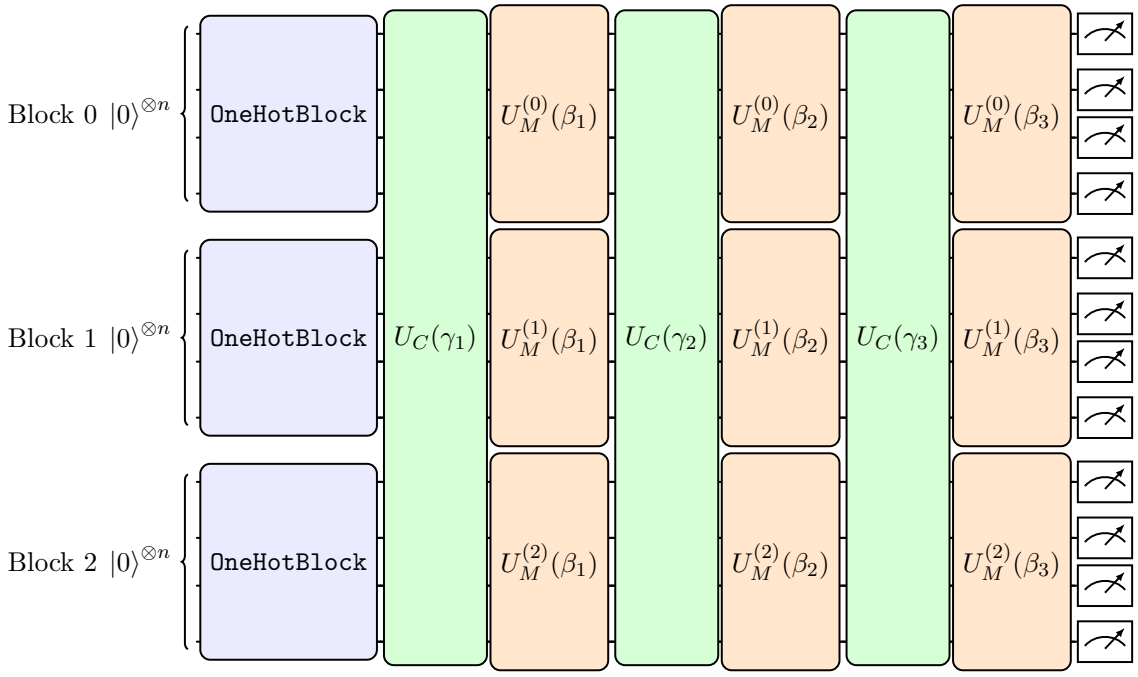


Figure 2: Depth-$p = 3$ CE-QAOA for $m = 3$ blocks of $n = 4$ qubits. Each layer applies a global cost $U_C(\gamma_\ell)$ over all mn wires, followed by parallel block-local XY mixers $U_M^{(j)}(\beta_\ell)$.

and the resulting conditional separation. Finally, §3.4 packages these ingredients into a *polytime hybrid quantum–classical* (PHQC) solver that couples a coarse (γ, β) grid with a deterministic checker, yielding Chernoff-style shot guarantees independent of empirical frequency.

3.1 Exact Unitary 1−design from block permutation twirl

We now turn to low order moment analysis [20] starting from the emergence of exact unitary 1 design in CE-QAOA. Our kernel (Def. 1) *explicitly* requires that the *constraints* be exchangeable under block permutations S_m and symbol relabelings S_n. This symmetry is a co-design choice that is enforced at the modeling level and then exploited algorithmically contributing a very rich dynamical structure to our formalism. For example, the block Level permutation–twirling action preserves the one-hot condition on the Hilbrt space $\mathcal{H}_{\mathrm{OH}}$ (See App. A.1 for a global feasibility preserving permutation). To see this, let

$$\mathcal{B} \equiv S_n^{\times m} = \left\{ \mathsf{P} = \bigotimes_{b=0}^{m-1} \mathsf{P}_b \ : \ \mathsf{P}_b \in S_n \right\} \subset U(\mathcal{H}_{\mathrm{OH}}),$$

which applies *independent* label permutations in each block. This action is *not* feasibility–preserving in general (it can break column one–hot), but it does implement a first–moment (1-design) average on the encoded space as an immediate consequence.

Consider a coarse uniform grid search with $n+1$ points:

$$\mathcal{G}_{n+1} := \left\{ j\frac{\pi}{n} : j = 0, 1, \ldots, n \right\}.$$

and take $\gamma \in \mathcal{G}_{n+1}$ and $\beta \in \mathcal{G}_{n+1}$. Let $\widetilde{H}_M^{(b)}$ be the normalized block-XY mixer on the one-excitation subspace. The following existence result holds:

Theorem 13 (Existence of Uniform Overlap Lower Bound via Block Permutations)**.** *For any product basis vector $|x^\star\rangle \in \mathcal{H}_{\mathrm{OH}}$, and any angles (γ, β), there exists a blockwise permutation $\mathsf{P}^\star$ such that*

$$\left|\langle x^\star \mid \mathsf{P}^{\star\dagger} U_M(\beta) U_C(\gamma) s_0\rangle\right|^2 \;\geq\; \frac{c}{n^m}.$$

To prove this, we shall need two Lemmas.

Lemma 14 (Existence from an average (pigeonhole/averaging principle))**.** *Let $\{O_t\}_{t\in T}$ be a finite family of Hermitian operators on any Hilbert space $\mathcal{H}$, and let ρ be any density operator (positive semidefinite, trace 1) on $\mathcal{H}$. Define the arithmetic mean operator*

$$\overline{O} \;:=\; \frac{1}{|T|}\sum_{t\in T} O_t.$$

Then there exists an index $t^\star \in T$ such that the expectation value of $O_{t^\star}$ in the state ρ is at least the expectation value of the mean operator:

$$\mathrm{Tr}(\rho\, O_{t^\star}) \;\geq\; \mathrm{Tr}\Big(\rho\,\overline{O}\Big).$$

i.e. among the set $\{\,\mathrm{Tr}(\rho\, O_t) \;:\; t \in T\,\}$ at least one is not smaller than their average.

Proof. By linearity of the trace,

$$\frac{1}{|T|}\sum_{t\in T}\mathrm{Tr}(\rho\, O_t) \;=\; \mathrm{Tr}\left(\rho\,\frac{1}{|T|}\sum_{t\in T} O_t\right) \;=\; \mathrm{Tr}\Big(\rho\,\overline{O}\Big).$$

If every $\mathrm{Tr}(\rho\, O_t)$ were strictly smaller than $\mathrm{Tr}\Big(\rho\,\overline{O}\Big)$, their average would also be strictly smaller—contradiction. Hence some $t^\star$ attains at least the average. □

Lemma 15 (Blockwise twirl ($\mathcal{B}$) yields the $1/D$ baseline)**.** *Let $\mathcal{H}_1 \cong \mathbb{C}^n$ be the one–hot subspace on a block $\mathcal{H}_1$. Let the product one-hot subspace be $\mathcal{H}_{\mathrm{OH}} = \mathcal{H}_1^{\otimes m}$. So $D = \dim \mathcal{H}_{\mathrm{OH}} = n^m$. For each block $b \in \{0,\ldots,m-1\}$, let P_b be a uniformly random permutation matrix acting on $\{|e_0\rangle,\ldots,|e_{n-1}\rangle\}$ and set $\mathsf{P} := \bigotimes_{b=0}^{m-1}\mathsf{P}_b$. Then, for any fixed unitary U on $\mathcal{H}_{\mathrm{OH}}$, any product basis vector $|x^\star\rangle$, and any state $|\phi\rangle \in \mathcal{H}_{\mathrm{OH}}$ (all independent of P),*

$$\mathbb{E}_{\mathsf{P}}\Big[\,\big|\langle x^\star \mid \mathsf{P}^\dagger U \mid \phi\rangle\big|^2\,\Big] \;=\; \frac{1}{D}. \tag{6}$$

Equivalently, at the operator level,

$$\mathbb{E}_{\mathsf{P}}\Big[\,U^\dagger\,\mathsf{P}\;|x^\star\rangle\langle x^\star|\;\mathsf{P}^\dagger\,U\,\Big] \;=\; \frac{I_D}{D}. \tag{7}$$

where I_D is the identity on $\mathcal{H}_{\mathrm{OH}}$.

Proof. Write $x^\star = (j_0^\star,\ldots,j_{m-1}^\star)$ with $|x^\star\rangle = \bigotimes_{b=0}^{m-1}\left|e_{j_b^\star}\right\rangle$. For a single block b, uniform conjugation by P_b averages a rank-one projector over all basis labels:

$$\mathbb{E}_{\mathsf{P}_b}\Big[\mathsf{P}_b\,\left|e_{j_b^\star}\right\rangle\!\left\langle e_{j_b^\star}\right|\,\mathsf{P}_b^\dagger\Big] = \frac{1}{n}\sum_{j=0}^{n-1}|e_j\rangle\langle e_j| = \frac{I_n}{n}.$$

Independence across blocks gives

$$\mathbb{E}_{\mathsf{P}}\Big[\mathsf{P}\;|x^\star\rangle\langle x^\star|\;\mathsf{P}^\dagger\Big] = \bigotimes_{b=0}^{m-1}\frac{I_n}{n} = \frac{I_D}{D}.$$

Conjugating by the fixed unitary U yields (7). Taking the expectation value of (7) in the state $|\phi\rangle$ gives the scalar identity (6):

$$\mathbb{E}_{\mathsf{P}}\Big[\,\big|\langle x^\star \mid \mathsf{P}^\dagger U \mid \phi\rangle\big|^2\,\Big] = \mathrm{Tr}\Big[U\,|\phi\rangle\langle\phi|\,U^\dagger\,\frac{I_D}{D}\Big] = \frac{1}{D}.$$

□

Proof of Thm. 13. Fix any angles (γ, β) and write $U := U_M(\beta)U_C(\gamma)$. Let $\rho := |s_0\rangle\langle s_0|$ and, for each blockwise permutation $\mathsf{P} = \bigotimes_{b=0}^{m-1} \mathsf{P}_b$, define the observable $O_{\mathsf{P}} := U^\dagger \mathsf{P} \, |x^\star\rangle\langle x^\star| \, \mathsf{P}^\dagger U$. By Lemma 15,

$$\mathbb{E}_{\mathsf{P}}[\, O_{\mathsf{P}} \,] = \frac{I_D}{D}\,, \qquad D = n^m,$$

and therefore

$$\mathbb{E}_{\mathsf{P}}\big[\, \mathrm{Tr}\big(\rho\, O_{\mathsf{P}}\big) \,\big] = \mathrm{Tr}\Big(\rho\, \frac{I_D}{D}\Big) = \frac{1}{D}.$$

Applying Lemma 14 (Existence via averaging) to the finite set $T = \{\mathsf{P}\}$ with observables $\{O_{\mathsf{P}}\}_{\mathsf{P}\in T}$ and state ρ yields the existence of some blockwise permutation $\mathsf{P}^\star$ (possibly depending on $\gamma, \beta, x^\star$) such that

$$\Big|\, \langle x^\star \mid \mathsf{P}^{\star\dagger} U \mid s_0\rangle \,\Big|^2 = \mathrm{Tr}(\rho\, O_{\mathsf{P}^\star}) \;\geq\; \frac{1}{D} = \frac{1}{n^m}.$$

□

Corollary 16 (Feasible–optimum specialization of Thm. 13)**.** *Fix any grid angles* $(\gamma, \beta) \in \mathcal{G}_{n+1} \times \mathcal{G}_{n+1}$ *and set* $U = U_M(\beta)U_C(\gamma)$ *on* $\mathcal{H}_{\mathrm{OH}}$ *with* $D = \dim \mathcal{H}_{\mathrm{OH}} = n^m$*. Let* $x^\star$ *be any* feasible optimal *label. Then there exists a blockwise permutation* $\mathsf{P}^\star$ *such that*

$$\big|\langle x^\star \mid \mathsf{P}^{\star\dagger} U \mid s_0\rangle\big|^2 \;\geq\; \frac{1}{D} \;=\; \frac{1}{n^m}.$$

In particular, the constant in Theorem 13 can be taken as $c = 1$*.*

Proof. Theorem 13 holds for *any* product basis vector $x^\star \in \mathcal{H}_{\mathrm{OH}}$. It also holds for any grid angles $(\gamma, \beta) \in \mathcal{G}_{n+1} \times \mathcal{G}_{n+1}$. Finally, choosing $x^\star$ to be any feasible optimum preserves the bound. Hence the inequality holds with $c = 1$. □

Lemmas 14–15 establish a permutation–twirl control: for any fixed U on $\mathcal{H}_{\mathrm{OH}}$ and any label $|x\rangle$, the averaged success probability equals the $1/D$ baseline. This identity certifies that random relabeling erases instance structure but offers a robust theoretical floor. For fixed angles $U := U_M(\beta)U_C(\gamma)$ and target $|x^\star\rangle$, define

$$p^{x^\star}_{\gamma,\beta}(\mathsf{P}) \;:=\; \big|\langle x^\star \mid \mathsf{P}^\dagger U \mid s_0\rangle\big|^2.$$

The blockwise permutation twirl gives the control identity $\mathbb{E}_{\mathsf{P}}[p^{x^\star}_{\gamma,\beta}(\mathsf{P})] = 1/D$. Hence the native labeling $\mathsf{P} = \mathrm{id}$ realizes one point in this distribution and, by averaging, there must exist labelings with $p^{x^\star}_{\gamma,\beta}(\mathsf{P}) > 1/D$ and others with $< 1/D$. Running the circuit *without* random per-shot relabelings (i.e., without implementing the twirl) samples $p^{x^\star}_{\gamma,\beta}(\mathrm{id})$, which can be *strictly larger* than $1/D$ whenever the angles and the instance structure produce constructive interference for the native labels. This arises when energetically favored blocks capture probability mass above $\frac{1}{n}$ average. Empirically this appears as peaks for the optimum/near-optimum bars that lie well above the dashed $1/D$ line (cf. Figure 4).

By contrast, applying random blockwise permutations per shot *flattens* the histogram toward the $1/D$ control and typically reduces such peaks. Thus non-twirled sampling is the appropriate mode for optimization (preserves useful bias), while twirling serves as a null/control for analytical bounds rather than a performance booster. Consequently, we optimize in the native labeling (no twirl) and interpret any deviation $p_{\mathrm{id}}(x) - 1/D$ as certified problem-induced bias. The practical value of such bias is quantified by shot complexity: if the optimal label satisfies $p_\star = \alpha/D$ with $\alpha > 1$, the shots required to observe it with high probability shrink by the same factor. The following corollary holds.

Corollary 17 (Shot complexity gain from heavy outputs)**.** *Let* $p_\star := p_{\mathrm{id}}(x^\star)$ *for the optimal label* $x^\star$ *at some grid point* (γ, β)*. With* S *shots, the failure probability to observe* $x^\star$ *at least once is* $(1-p_\star)^S \le e^{-p_\star S}$*. To achieve success probability* $\ge 1-\delta$ *it suffices that*

$$S \;\ge\; \frac{\ln(1/\delta)}{p_\star}.$$

At the permutation twirling baseline $p_\star = 1/D$*, this is* $S = \Theta(D \log(1/\delta))$*. Any heavy-output bias* $p_\star = c/D$ *with* $c > 1$ *yields an* c*-fold reduction in shots. If* $p_\star \ge n^{-k}$ *for some fixed* k *while* $D = n^m$ *(with* $m = n$ *in TSP encodings), then* $S = \tilde{O}(n^k)$ *versus* $\tilde{\Omega}(n^m)$ *at baseline. See Lemma 24, Thm. 25 and Alg. 4. The two flavors of computational advantages arising from our proposals are discussed in Sec. 3.3 and 4.3. An illustration of the effects of heavy outputs on success probability is shown in Fig. 3[31].* [3]

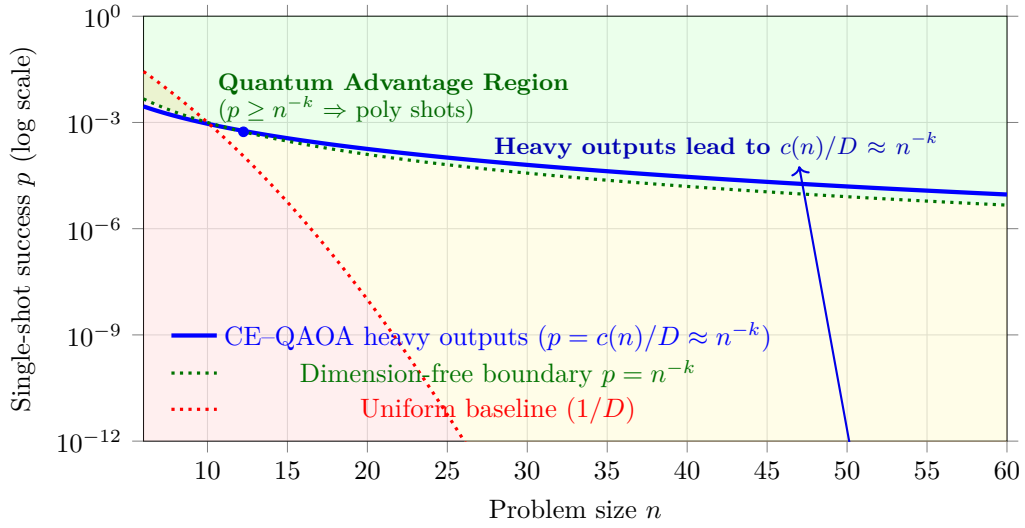
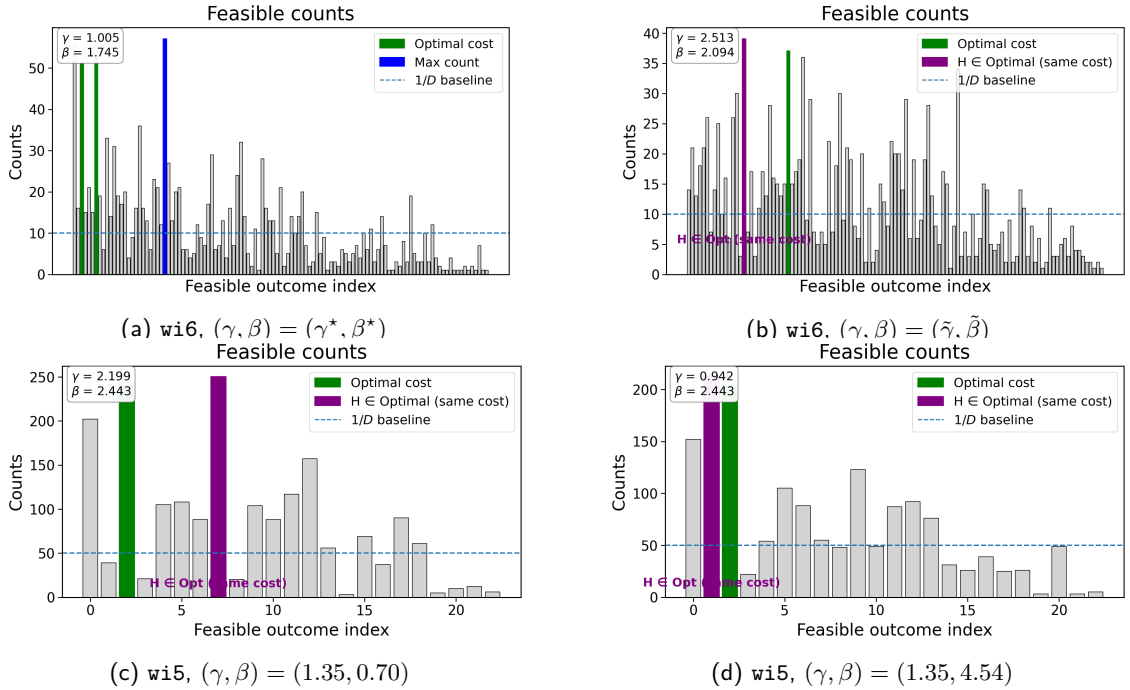


Figure 3: Heavy outputs vs. baselines. The red dotted line is the twirling baseline of $\mathbb{E}\,|\langle x|U|\phi\rangle|^2 = 1/D$. The light yellow region denotes $\mathbb{E}\,|\langle x|U|\phi\rangle|^2 > c/D$ existence results from Thm. 13 and Cor. 16. The green region indicate $\mathbb{E}\,|\langle x|U|\phi\rangle|^2 \approx n^{-k}$ for a fixed constant k. Here we write the constant numerator in Thm. 13 c as $c(n)$ to emphasize that it is instance dependent. It becomes large when constructive interference arises in absence of the permutation twirling. These higher values overlap with n^{-k} region shaded in green. If the probability of the optimum approaches n^{-k}, then to achieve success probability $\ge 1-\delta$ it suffices that $S \ge \frac{\ln(1/\delta)}{p_\star}$ as outline in Cor. 17. Our simulation results with $\ln(1/\delta) = 10$ is reported in Fig. 5 where the instance dependent heavy outputs lead to overlap probabilities in the shaded green region.

Remarks on the $[0,\pi]^2$ window. We use a π-sized search box in Thm. 13 and the accompanying twirling analyses purely for *coverage*. The $t \in \{1,2\}$ moment guarantees (perm-twirl and shallow anticoncentration) only require angles drawn from a constant-size set and do not strengthen by widening the range. Empirically, $[0,\pi]^2$ balances exploration of distinct phase patterns with modest grid sizes. Noiseless simulation results on Qiskit Aer are reported in Figure 4 to demonstrate this.

To conclude on the emergence of exact unitary 1-design in our formalism, recall that an

[3] Data availability: Minimal Python implementation and integration into the standard QAOA wrapper in Qiskit is made available https://doi.org/10.5281/zenodo.15725265.



(a) `wi6`, $(\gamma, \beta) = (\gamma^\star, \beta^\star)$

(b) `wi6`, $(\gamma, \beta) = (\tilde{\gamma}, \tilde{\beta})$

(c) `wi5`, $(\gamma, \beta) = (1.35, 0.70)$

(d) `wi5`, $(\gamma, \beta) = (1.35, 4.54)$

Figure 4: **Single Layer CE-QAOA circuits on noiseless Aer simulator [52] on two TSP instances (`wi5` (25 qubits) and `wi6` (36 qubits)) taken from QOptLib benchmark set [30].** Histograms show the counts of *feasible* bit-strings (permutation indices) out of $50 \times n^4$ shots for a problem of size n locations. The green bars mark two degenerate best tours. Several grid points already identify the optimum with nontrivial probability well above the lower bound. The dashed line marks the $1/D$ baseline with $D = n^n$. The purple bars show when the best bitsring is also the most frequent.

ensemble $\{p_i, U_i\} \subset U(d)$ is a unitary 1-design if its first-moment (twirling) channel

$$\Phi_{\text{Ist}}(X) \;:=\; \sum_i p_i \, (U_i)\, X\, (U_i^\dagger)$$

agrees with the Haar first moment $\Phi^{(1)}_{\text{Haar}}(X) := \int_{\text{Haar}} (U)\, X\, (U^\dagger)\, dU$ for all X on $\mathbb{C}^d$. It is an ε-*approximate* 1-design if $\|\Phi_{\text{Ist}} - \Phi^{(1)}_{\text{Haar}}\| \le \varepsilon$ in diamond (or operator) norm. Thus, the identity proved in Lemma 15 is an exact 1-design identity with $\varepsilon = 0$. This is a consequence of the block–XY mixer acting as coherent *hops* between labels inside the $\mathcal{H}_1$ subspace. In the encoded n-level (qudit) picture, these hops are realized coherently by the unitary operator, $e^{-i\beta A(K_n)}$, where $A(K_n)$ is the adjacency of the complete graph on n nodes (See Def. 9 and Prop. 10) and any blockwise permutation operator respects this symmetry. Further t-unitary design details are available for example in Ref. [20].

3.2 ε-approximate unitary 2- designs

Similarly to the 1-design notion, an ensemble $\{p_i, U_i\} \subset U(d)$ is a unitary 2-design if its second-moment (twirling) channel

$$\Phi_{\text{2nd}}(X) \;:=\; \sum_i p_i \, (U_i \otimes U_i)\, X\, (U_i^\dagger \otimes U_i^\dagger)$$

agrees with the Haar second moment $\Phi^{(2)}_{\text{Haar}}(X) := \int_{\text{Haar}} (U \otimes U)\, X\, (U^\dagger \otimes U^\dagger)\, dU$ for all X on $\mathbb{C}^d \otimes \mathbb{C}^d$. It is an ε-*approximate* 2-design if $\|\Phi_{\text{2nd}} - \Phi^{(2)}_{\text{Haar}}\| \le \varepsilon$ in diamond (or operator) norm. In our formalism, if on each block's one–excitation space $\mathcal{H}_1 \cong \mathbb{C}^n$, we form the local ensemble by applying random XY edge pulses, i.e., layers that choose pairs (i, j) and angles

θ and apply $\exp(-i\theta(X_iX_j + Y_iY_j))$. Varying the edges and angles yields a (controllable) gate set whose dynamical Lie algebra is $\mathfrak{su}(n)$ (Prop. 12). So by local random–circuit design results[20], a circuit of length $L = \mathrm{poly}(n) \cdot \mathrm{polylog}(1/\varepsilon)$ forms an ε-approximate unitary 2-design on $U(n)$. We capture this result in the following proposition.

Proposition 18 (Per–block ε-approximate 2–design)**.** *Recall controllability and approximate universality on $\mathcal{H}_1$ (Prop. 12). Let $\mathcal{U}$ be generated by XY pulses $\exp(-i\theta(X_iX_j + Y_iY_j))$ and Z phases $\exp(-i\phi Z_k)$, with (θ, ϕ) drawn from bounded continuous densities. On $\mathcal{H}_1$, a random circuit of length $\mathrm{poly}(n) \cdot \mathrm{polylog}(1/\varepsilon)$ sampled from $\mathcal{U}$ forms an ε-approximate unitary 2–design [20].*

Interleaving such per–block layers with the diagonal, instance–dependent phase $U_C(\gamma)$ which is entangling across blocks is needed to promote the product of per–block designs to a global $\varepsilon_{\mathrm{glob}}$-approximate unitary 2-design on the encoded space in depth $\mathrm{poly}(m, n, \log(1/\varepsilon_{\mathrm{glob}}))$.

In the CE-QAOA kernel (Def. 1), each phase layer is generated by a *diagonal* Hamiltonian in the computational basis which is *entangling across blocks* due to the global pattern symmetries. With block labels $z = (j_0, \ldots, j_{m-1})$, $U_C(\gamma)$ factorizes over blocks (and thus non-entangling across blocks) iff

$$C(j_0, \ldots, j_{m-1}) \;=\; \sum_{b=0}^{m-1} c_b(j_b) \quad \text{for some} \quad c_b : [n] \to \mathbb{R}. \tag{8}$$

Proposition 19 (Diagonal entangler criterion)**.** *For two blocks $b \neq b'$, restrict C to $C_{b,b'}(j,k)$ with other labels fixed and consider*

$$U_{b,b'}(\gamma) \;=\; \sum_{j,k=0}^{n-1} e^{-i\gamma C_{b,b'}(j,k)}\, |j\rangle\langle j| \otimes |k\rangle\langle k|.$$

The operator–Schmidt rank of $U_{b,b'}(\gamma)$ exceeds 1 (hence the gate is entangling) unless there exist functions $\alpha(j), \beta(k)$ and a phase θ with $e^{-i\gamma C_{b,b'}(j,k)} = e^{-i\theta}\, a(j)b(k)$ for all j, k, i.e. unless $C_{b,b'}(j,k) = \alpha(j) + \beta(k)$ modulo $2\pi/\gamma$. Equivalently, the diagonal two–qudit gate has operator–Schmidt rank > 1 iff the matrix $M_{jk} = e^{-i\gamma C_{b,b'}(j,k)}$ has rank > 1.

Proof. For a diagonal two–qudit operator, the operator–Schmidt rank equals the rank of the matrix $M_{jk} = e^{-i\gamma C_{b,b'}(j,k)}$. Rank 1 holds iff $M_{jk} = a(j)b(k)$. □

Global constraints in typical COPs (TSP/VRP, assignment, capacity/balance) introduce cross–block couplings, so $U_C(\gamma)$ is *generically* entangling on $(\mathbb{C}^n)^{\otimes m}$. We can therefore combine it with short, per–block 2–designs to obtain strong second–moment mixing on the encoded space. To that end, we partition the diagonal generator as

$$H_C \;=\; H_{\mathrm{obj}} + H_{\mathrm{pen}},$$

with the following roles. (i) H_{obj} is diagonal and, upon restriction to a single block, provides sufficiently rich on–block phases so that, together with XY hops inside the block, the Lie algebra on $\mathcal{H}_1 \cong \mathbb{C}^n$ is $\mathfrak{u}(n)$. (These on–block phases may be present explicitly or synthesized via standard refocusing.) (ii) H_{pen} is diagonal and *cross–block*. It contains at least one non–additive term across some pair $b \neq b'$, so that for fixed spectators the induced table $C_{b,b'}(j,k)$ fails $C_{b,b'}(j,k) = \alpha(j) + \beta(k)$. By Prop. 19, $e^{-i\gamma H_{\mathrm{pen}}}$ is then entangling across blocks. (Equivalently, the matrix $M_{jk} = e^{-i\gamma C_{b,b'}(j,k)}$ has rank > 1.)

If each block ensemble is an ε-approximate unitary 2–design on $U(n)$, then the product ensemble on $U(n)^{\otimes m}$ reproduces the Haar second moments of degree–≤ 2 polynomials with error $O(m\varepsilon)$.

Theorem 20 (Global $\varepsilon_{\text{glob}}$-approximate unitary 2–design (mild randomness))**.** *Assume the block–interaction graph induced by H_{pen} is connected and that either (i) the phases γ_ℓ are i.i.d. from a bounded continuous density, or (ii) the set of coupled block pairs in H_{pen} is chosen i.i.d. each layer from a distribution with full support on the edges of the graph. Then there exists*

$$L \;=\; \text{poly}(m, n, \log(1/\varepsilon_{\text{glob}}))$$

s.t. the distribution of $U_L = \mathcal{L}_L \cdots \mathcal{L}_1$ is an $\varepsilon_{\text{glob}}$-approximate unitary 2–design on $U(n^m)$.

Proof. Setup: Consider alternating layers on $(\mathbb{C}^n)^{\otimes m}$

$$\mathcal{L}_\ell \;=\; \left(\bigotimes_{b=0}^{m-1} R_b^{(\ell)}\right) e^{-i\gamma_\ell H_{\text{pen}}}, \qquad \ell = 1, \ldots, L, \tag{9}$$

where each $R_b^{(\ell)}$ is drawn independently from a per–block ε-approximate 2–design and $e^{-i\gamma_\ell H_{\text{pen}}}$ is an entangling diagonal gate across blocks (Prop. 19). Assume a mild randomness on the penalty layer and fix one layer $\mathcal{L}_\ell = \bigotimes_b R_b^{(\ell)}\; e^{-i\gamma_\ell H_{\text{pen}}}$. For $t = 2$ moments, analyze the superoperator $\mathbb{E}\left[U^{\otimes 2} \otimes \overline{U}^{\otimes 2}\right]$ of the depth–L circuit.

(1) *Local twirl to* $\{I, \text{SWAP}\}$. Each $R_b^{(\ell)}$ is drawn from a per–block ε–approximate 2–design, so its second–moment twirl projects (up to $O(\varepsilon)$) any operator on the two–copy space of block b onto the commutant of $U(n)$, i.e. the span of $\{I_b, \text{SWAP}_b\}$. Thus, after averaging over the $R_b^{(\ell)}$, the state space relevant to second moments reduces to a classical spin system on m sites with local alphabet $\{I, \text{SWAP}\}$.

(2) *Effect of the entangling diagonal.* Consider an edge (b, b') coupled by a term in H_{pen}. Conjugation by $e^{-i\gamma_\ell H_{\text{pen}}}$, restricted to this edge and then followed by the local twirls from step (1), induces a stochastic update on the pair $(\{I, \text{SWAP}\}_b, \{I, \text{SWAP}\}_{b'})$ that is *nontrivial* whenever the two–block phase table $C_{b,b'}(j, k)$ is non–additive (Prop. 19). Hence each coupled edge implements a mixing step on the reduced alphabet.

(3) *Ergodicity and spectral gap (Cf. Prop. 8).* Assume the block–interaction graph of H_{pen} is connected and that, per layer, either the coupled edge or the phase parameter is drawn from a distribution with full support. Then the induced Markov chain on $\{I, \text{SWAP}\}^m$ is irreducible and aperiodic, with a spectral gap bounded below by $\Omega(1/\text{poly}(m))$. Consequently the product of $L = \text{poly}(m, n)\,\text{polylog}(1/\varepsilon_{\text{glob}})$ such layers converges to the unique fixed point, which coincides with the global Haar second moment on $(\mathbb{C}^n)^{\otimes m}$ (up to the usual $O(\varepsilon)$ error from the local designs).

(4) *In conclusion,* this is exactly the t=2 instance of the Brandão–Harrow–Horodecki random–circuit mixing framework[20], with the qudit case handled by the same moment–operator gap argument[55].[4] □

Corollary 21 (Constant-measure good angles under a 2-design)**.** *Let* U *be an ε-approximate unitary 2-design on the encoded space (e.g., as in Thm. 20). For fixed $x^\star$ and input $|\phi\rangle$, with $X = |\langle x^\star|U|\phi\rangle|^2$ where $U \sim \mathsf{U}$,*

$$\mathbb{E}[X] = \tfrac{1}{D} \pm O(\varepsilon), \quad \mathbb{E}[X^2] = \tfrac{2}{D(D+1)} \pm O(\varepsilon)^5,$$

[4] Here, the moment-operator gap refers to the spectral gap of the one-layer second-moment operator

$$M_2 \;=\; \mathbb{E}\left[U^{\otimes 2} \otimes \overline{U}^{\otimes 2}\right],$$

or equivalently, after local twirling, to the spectral gap of the induced Markov kernel on the reduced commutant basis $\{I, \text{SWAP}\}^m$; this gap controls the exponential convergence of M_2^L to the Haar second moment [20, 55]. See App. A.5.

[5] See derivation in App. A.6

and Paley–Zygmund gives $\Pr[X \geq \frac{1}{2D}] \geq (D{+}1)/(8D) - O(\varepsilon) = \Omega(1)$. *Thus, Paley–Zygmund gives, for sufficiently small* $\varepsilon_{\text{glob}}$,

$$\Pr\Big[X \geq \tfrac{1}{2D}\Big] \geq \frac{\big(\mathbb{E}[X] - \frac{1}{2D}\big)^2}{\mathbb{E}[X^2]} \gtrsim \frac{(1/2D)^2}{2/(D(D{+}1))} = \frac{D{+}1}{8D} = \Omega(1). \tag{10}$$

Design bounds as controls: Unitary 1-designs vs Unitary 2-designs. Unitary 1-designs give *means* while unitary 2-designs add *dispersion* and *correlation* control and calibrate pairwise covariances of quadratic observables. In combinatorial optimization this means we can quantify anticoncentration *within* the encoded space; and separate genuine instance–parameter structure (like persistent peaks above $1/D$) from labeling artifacts. In variational training, 2-design control on second moments stabilizes the statistics of loss estimates (and common gradient/finite-difference estimators that depend on degree-≤ 2 unitary moments), yielding instance–size predictable variance and thus shot requirements. For our constraint-enhanced QAOA, the 1-design perspective explains *why* favorable parameters must exist (and provides a clean null/control), whereas the 2-design refinement explains *how reliably* we can *find and certify* them.

Finally, a caveat. Unitary t–design bounds (including our 1– and per–block 2–design controls) are *conservative*: they certify correct means/second moments under structured randomization, but they do not translate into tight, instance–specific shot complexity. In particular, the 1–design baseline $1/D$ is an analytic *null*—useful for existence and for flattening controls via twirls—but it grossly overestimates the shots needed when constructive interference lifts the success probability well above $1/D$. Likewise, second–moment control ($O(m\varepsilon)$ product moments) stabilizes estimators and separates signal from labeling artifacts, yet it is not a substitute for the binomial tails of the *actual* success rate $p(\gamma, \beta)$ attained by CE–QAOA at good angles. For shot complexity we therefore appeal to the realized Bernoulli scaling $N = \tilde{O}(1/p_\star)$ (via Chernoff (Lemma 24)), where in our setting $p_\star$ empirically behaves like n^{-k} once only $k \ll m$ blocks remain unaligned. This yields polynomial shot budgets, in stark contrast to the exponential $\Theta(D)$ implicit in the 1–design null (Cf. Classical competitveness in Sec 3.3 vs. Sec 4.3).

3.3 Classical Competitiveness I: The Unitary t-Design Baseline

The comparison with classical procedures hinges on *what domain the classical sampler can efficiently prepare*. Our CE–QAOA acts *inside* the encoded one–hot product space $\mathcal{H}_{\text{OH}} = [n]^m$ of size $D = n^m$. A classical algorithm, however, may or may not have an efficient generator for $\mathcal{H}_{\text{OH}}$. We therefore separate two information models and stay strictly within first/second moment reasoning where permutation twirls achieve overlap at least $1/D$. Throughout, let $F \subseteq \mathcal{H}_{\text{OH}}$ be the feasible set, $|F| =: S$, and let U be the depth-p CE–QAOA circuit restricted to the encoded space. We now make the "ambient domain" explicit.

Model A (structured domain). Both quantum and classical procedures can sample arbitrarily from the encoded domain $\mathcal{H}_{\text{OH}} = [n]^m$ (size $D = n^m$), and both have an oracle answering feasibility membership in F.

Model B (raw bitstrings only). The classical procedure sees only raw bitstrings of length N (for TSP/assignment, $N = n^2$) and has access to a feasibility predicate $\mathsf{feas}(z) \in \{0, 1\}$ that recognizes whether a bitstring encodes a feasible element of $F \subseteq \mathcal{H}_{\text{OH}}$. No

efficient generator for $\mathcal{H}_{\mathrm{OH}}$ is assumed. The quantum procedure, by construction, prepares superpositions *inside* $\mathcal{H}_{\mathrm{OH}}$.

Proposition 22 (Model A: matching $\Theta(D/S)$ baselines)**.** *Suppose both sides can sample uniformly from* $\mathcal{H}_{\mathrm{OH}}$ *and query* feas*. Then any classical algorithm that draws i.i.d. proposals from* $\mathcal{H}_{\mathrm{OH}}$ *has success probability* S/D *per trial, so the expected number of trials to hit* F *is* D/S*. For CE–QAOA with a random twirl, Lemma 15 implies the same expected success* S/D *for "hit any feasible", hence the same* D/S *expected trials.*

Proposition 23 (Model B: raw-bitstring baseline)**.** *Let the classical ambient space be* $\{0,1\}^N$ *with* $N = n^2$ *(for n=m), of size* 2^N*, and assume only the feasibility predicate* feas *is available. Under the minimax distribution in which* F *is uniform among all* S*–subsets of* $\{0,1\}^N$*, the expected number of feasibility queries until the first hit is*

$$\mathbb{E}[\#\mathit{queries}] \;=\; \frac{2^N+1}{S+1} \;=\; \Theta\!\left(\frac{2^{n^2}}{S}\right).$$

In contrast, CE–QAOA prepares states inside $\mathcal{H}_{\mathrm{OH}}$ *and, with only the first–moment guarantee, has expected trials* $D/S = \Theta(n^n/S)$*. Thus the ratio of classical (Model B) to quantum expected trials scales like*

$$\frac{2^{n^2}/S}{n^n/S} \;=\; \frac{2^{n^2}}{n^n} \;=\; \Big(\frac{2^n}{n}\Big)^n \;=\; \exp\!\big(\Theta(n^2)\big).$$

Thus, CE–QAOA already achieves a conditional quantum-classical separation from design only bounds. However, we as we have already discussed, this is a mere theoretical floor independent of constructive interference present when the twirls are not implemented. As we saw in Fig 2, when the twirling is omitted, heavy outputs appear due to instance dependent constructive interference. As illustrated in Fig 3, these instance dependent heavy outputs pushes the probability of sampling the optimum from the design baseline into the finite probability regime. Additional contributions from parameter optimization are discussed in Sec. 4.3.

3.4 Polytime Hybrid Quantum–Classical Solver (PHQC)

We now extend the algorithmic pipeline to include a *coarse grid* search over (γ, β) parameters as used in Sec. 3.1 and an *exact classical checker* that returns the lowest energy feasible bitstring regardless of its empirical frequency. Let $\beta \in [0,\pi]$, $\gamma \in [0,\pi]$. For a chosen depth p, we build a rectangular grid

$$\mathcal{G} \;=\; \{(\beta_i,\gamma_j):\; \beta_i = i\,\Delta_\beta,\; \gamma_j = j\,\Delta_\gamma,\; 0 \le i \le N_\beta,\; 0 \le j \le N_\gamma\},$$

with spacings $\Delta_\beta, \Delta_\gamma \in [0,\pi]^{n+1}$ as in Theorem 13. Given a multiset of measured bitstrings $\mathcal{S}$, the checker evaluates *exact* objective values $E(b) = \langle b|H_{\mathrm{obj}}|b\rangle$ for all $b \in \mathcal{S}$ and returns the minimal element. Consequently, *it is irrelevant whether the optimal bitstring is the most frequent sample.* Any single appearance of an optimal $b^\star$ suffices for the checker to output the optimum. This strengthens the overall protocol and minimizes the shot budget. The deterministic post-processing performs scoring $b \mapsto \langle b|H_C|b\rangle$ in $O(n^2)$ with total classical work $O(S\,n^2)$.

Lemma 24 (Chernoff bound for optimum hits)**.** *Let* $N_\star$ *be the number of* optimal *bitstrings observed in* S *independent samples, each occurring with probability* $p_{\min}(\epsilon)$*. Then*

$$\Pr[N_\star = 0] \;\le\; \exp\!\big(-p_{\min}(\epsilon)\,S\big), \qquad S \;\ge\; \left\lceil \tfrac{\ln(1/\delta)}{p_{\min}(\epsilon)} \right\rceil \;\Longrightarrow\; \Pr[N_\star \ge 1] \;\ge\; 1-\delta.$$

Proof. Let $X_i \sim \text{Bernoulli}(p_{\min}(\epsilon))$ be the indicator for an optimum hit in shot i. Then $N_\star = \sum_{i=1}^{S} X_i$ with $\mathbb{E}[N_\star] = S\, p_{\min}(\epsilon)$. Using $\Pr[N_\star = 0] = (1 - p_{\min})^S \le e^{-p_{\min} S}$ proves the first inequality; the shot bound follows by inversion. □

Theorem 25 (Sample complexity for perfect recovery)**.** *Let $p_{\min} > 0$ be the total probability mass on the optimal bit-strings after p Constraint-Enhanced QAOA layers. With $S \ge \lceil \ln(\delta^{-1})/p_{\min} \rceil$ shots, the PHQC post-processor returns a* globally optimal *solution with probability at least $1 - \delta$.*

Sketch. Apply Lemma 24 with $p_{\min}(\epsilon) = p_{\min}$. □

Algorithm 4 PHQC — constant-depth CE–QAOA + deterministic checker

Require: size n; depth p; grid $\mathcal{G}$ over (β, γ); $H_C = H_{\text{pen}} + H_{\text{obj}}$; mixer H_M; shots $S = \lceil n^k \ln(1/\delta) \rceil$.
Ensure: optimal feasible $(b^\star, c^\star)$.

$|s_0\rangle \leftarrow |s_{blk}\rangle^{\otimes m}$.
$\mathcal{S} \leftarrow \emptyset$.
for each $(\beta, \gamma) \in \mathcal{G}$ **do**
    prepare $|s_0\rangle$; apply $U_C(\gamma)$; apply $U_M(\beta)$; measure S shots into $\mathcal{S}$.
end for
$(b^\star, c^\star) \leftarrow (\text{null}, +\infty)$.
for each $b \in \mathcal{S}$ **do**
    if FEASIBLE(b) **then**
        $c \leftarrow \langle b | H_{\text{obj}} | b \rangle$ ▷ or full H_C if you score penalties
        if $c < c^\star$ **then**
            $(b^\star, c^\star) \leftarrow (b, c)$
        end if
    end if
end for
return $(b^\star, c^\star)$.

4 Circuit Simulation Results

4.1 Traveling Salesman Problem

The Traveling Salesman Problem (TSP) is defined on a complete weighted graph $G = (V, E)$, $|V| = n$, with distance matrix $C_{ab} \in \mathbb{R}^{n \times n}_{\ge 0}$. A tour is a Hamiltonian cycle visiting each city exactly once. We represent a tour by an $n \times n$ binary assignment matrix $X = (x_{i,a})$ where row i indicates the city visited at *position* i in the tour and column a indicates *which* city. Each position is assigned exactly one city and each city appears exactly once,

$$\sum_{a=1}^{n} x_{i,a} = 1 \quad (\forall i), \qquad \sum_{i=1}^{n} x_{i,a} = 1 \quad (\forall a), \qquad x_{i,a} \in \{0, 1\}. \tag{11}$$

This is the double one-hot encoding. It matches our block structure defined in Def. 1 with each of the $m = n$ rows defined as a one-hot block of size n.
On the computational basis $x \mapsto X = (x_{i,a})$[6] and the *objective* is the cyclic pairwise cost

[6] We write a computational-basis string $x \in \{0, 1\}^{n^2}$ in matrix form $X = (x_{i,a})$ only to make the block structure explicit. One may simply treat x itself as the $n \times n$ assignment matrix with entries $x_{i,a}$. The

between consecutive positions:

$$H_{\text{obj}}(x) \;=\; \sum_{i=1}^{n} \sum_{a,b=1}^{n} C_{ab}\, x_{i,a}\, x_{i+1,b} \qquad (i{+}1 \equiv 1). \tag{12}$$

Feasibility is enforced by quadratic *penalties* for row/column sums,

$$H_{\text{pen}}(x) \;=\; \lambda_{\text{row}} \sum_{i=1}^{n} \Big(\sum_{a=1}^{n} x_{i,a} - 1 \Big)^2 \;+\; \lambda_{\text{col}} \sum_{a=1}^{n} \Big(\sum_{i=1}^{n} x_{i,a} - 1 \Big)^2, \tag{13}$$

optionally augmented by linear forbids (e.g., disallowing self-loops or precluded edges). The *double one-hot* constraints enforce a permutation of the visited locations. The first term in Eq 13 is now redundant and can be dropped. The *cost Hamiltonian* is $H_C = H_{\text{pen}} + H_{\text{obj}}$, which is diagonal in the computational basis and fits our kernel requirements in Def 1. See [24] for detailed QUBO derivations and discussions.

4.2 Numerical validation on approximate circuit simulators

We benchmark the CE-QAOA on QOPTLib instances[30] by setting a fixed start city ("anchored" reduction), which yields $(n{-}1)^2$ logical qubits (blocks $= n{-}1$, block size $= n{-}1$) for a problem defined on n locations. We work with a single layer composed as $U(\gamma, \beta) = U_M(\beta)\, U_C(\gamma)$. We sweep a grid $\mathcal{G} \subset [0, \pi]^2$ with $|\mathcal{G}| = (n{+}1)^2$ angle pairs, and use a shot budget of $S \;=\; 10\, n^k$ with k ranging from $3 - 5$ which comfortably absorbs finite-precision and rounding effects of the grid search and finite sampling since the number of qubits scales as $O(n^2)$. For each grid point we measure in the computational basis, classically filter feasibile solutions, and evaluate cyclic tour costs relative to the fixed start. We report (i) the best measured feasible solution, (ii) the fraction of feasible outcomes, and (iii) the minimum tour cost. In Table 1, the PHQC recovered the optimum in all cases. The fact that Algorithm 4 guarantees any sampled optimal bitstring to be identified regardless of its frequency, greatly reduces the shot burden in our implementation. [7]

Table 1: Single–layer CE–QAOA sampling with design–based angles and shot budget $S = 10\, n^{3-5}$ used to recover the global optimum. For a fair comparison, we assume a classical sampler with access to the block one-hot space so that the classical shot cost formula $S_{\text{cl}} \;=\; \Theta\big(n^{\,m-1} \log(1/\delta)\big)$ holds. We set $\log(1/\delta) = 10$ for both quantum and classical models and n^r refers to the ratio in Eq. 15.

Instance	Cities n	Shots S	Advantage (n^r)	Tour cost	Optimal	(γ, β)
`wi4`	4	160	n^1	6,700	Yes	$(1.57, 2.36)$
`wi5`	5	250	n^2	6,786	Yes	$(1.57, 2.36)$
`wi6`	6	360	n^3	9,815	Yes	$(2.62, 2.09)$
`wi7`	7	733	$> n^3$	7,245	Yes	$(3.14, 2.62)$
`dj8`	8	$2.56{\times}10^4$	$> n^3$	2,762	Yes	$(1.35, 2.24)$
`dj9`	9	$3.65{\times}10^4$	$> n^4$	2,134	Yes	$(0.30, 1.60)$
`dj10`	10	$1.00{\times}10^6$	$> n^4$	2,822	Yes	$(1.01, 1.57)$

[8]

reshape $x \mapsto X$ merely clarifies the row/column one–hot constraints and the cyclic pairwise objective $H_{\text{obj}}(x) = \sum_{i,a,b} C_{ab}\, x_{i,a} x_{i+1,b}$, which naturally refer to this matrix structure.

[7] For the larger system sizes (dj8-dj10) the per grid shot budget had to be raised because the circuit approximation struggled to capture the full dynamics. Similarly, for our smaller instances where full circuit simulation was possible, a lower $10n^2$ shots was found sufficient without increasing layer or grid size.

[8] A note on classically sampling from feasible permutations. Even if a classical sampler draws *directly*

4.3 Classical Competitiveness II: $\Theta(n^r)$ Sampling Advantage

The ε-approximate 2-design on the *encoded* manifold gives Haar-like second-moment anticoncentration at scale $1/D$ with $D = n^m$ (Thm. 20), which is a *conservative* worst-case sampling floor and a predictor of typical-case overlap statistics (Cor. 21). However, there are two empirical observations to be made at depth p=1. (i) Polynomial-shot recovery across instances with a basic coarse grid search over the (γ, β) parameter space. (ii) Systematic *peaks above* $1/D$ when the block-permutation twirl is *omitted* (Fig. 4) suggesting that the native labeling preserves instance structure and constructive interference. This is expected to remain the case at higher layers with $p > 1$. These *empirical* observations therefore *motivate* a targeted search over angle pairs and shallow depths to find constructive-interference settings that *achieve* per-shot success at the heuristic scale

$$p_{\min} \gtrsim \frac{1}{n^k} \quad \text{for } k < m. \tag{14}$$

To see how this might arise under finer angle and layer optimization, we can think of each fixed block in m as systematically reducing the effective degrees of freedom from m blocks down to k. We start in a uniform superposition of block one-hot basis vectors and assume that at least one block can be fixed at the beginning of the optimization process. This is no different from fixing the starting position in the TSP. The parameter optimization process fixes further block(s) to minimize the objective function. Thanks to the global constraint, each layer can be further assumed to slightly favor configurations that agree with a target pattern on a few blocks with a suitable angle choice. These blocks are then assumed *locked*. Meaning that finer parameter optimization at current layer or next can yield a new set of parameters that improves the number of locks but not degrade it. We can think of a block as *locked* once the target symbol in that block gains a fixed margin over the uniform $1/n$ baseline and k is interpreted as residual "undecided" blocks. Because of the global constraint, subsequent layers leverage already–locked blocks to bias additional ones, producing a cascade that leaves only k blocks unfixed.

At that point the success probability depends only on these k degrees of freedom and is therefore bounded below by $1/n^k$ independent of the ambient dimension $D = n^m$. The success probability in Eq. 14 follows. PHQC *then* converts this into *polynomial* shot complexity, $S \geq \lceil \log(1/\delta)/p_{\min} \rceil = O(n^k \log(1/\delta))$ (Thm. 25). With the classical shot cost given as $S_{\text{cl}} = \Theta(n^{m-1} \log(1/\delta))$ shots, the ratio is

$$\frac{S_{\text{cl}}}{S_{\text{q}}} = \Theta(n^r), \tag{15}$$

Where $r = m - 1 - k$. Taking into account the clear upward trend in Table 1; with $r = 6$ for $n = 15$ problem, the quantitative advantage exceeds 10^7. It goes without saying that better circuit approximations would improve accuracy and amplify the advantage.

4.3.1 Near-Optimality of Grid Parameters

It is also worth noting that our coarse parameter grid is far from optimal. We use it here as a sufficient condition for the existence of parameters that can yield the design based

and uniformly from feasible bitstrings, the success probability for a unique optimum is

$$p_{\text{succ}} = \frac{1}{n!} \quad \Rightarrow \quad S_{\text{cl}} = \Theta\big(n! \log(1/\delta)\big).$$

Given the rapid growth of $n!$, the advantage persists, with tweaks in prefactors and orders of magnitude.

results and subsequent large reduction in shot cost expected when the twirling averaging is bypassed. The approximate 2-design results suggest inherent trainability and anticoncentration in the proposed construction. Empirically, increasing the the grid points yields more frequent near-optimal interference within the same shot budget. For example, at $n = 8$, using 20×20 grids for (γ, β) (instead of $(n+1) \times (n+1) = 9 \times 9$) and the same shot budget of $10n^3 = 5120$ produced 12 angle pairs attaining the optimum, and 40 angle pairs within 1% of the optimum and 229 points within 5% of the optimum. Whereas the coarser grid of size $n + 1$ found only 2 pairs of grid points attaining the optimum. Similar phenomenon was observed for problems with $n = 9$ and 10 respectively. This supports our choice to prioritize *density over range* and treat $[0, \pi]^2$ as a sufficient compact tile for practical scans, while leaving systematic parameter optimization for future work.

Approximate circuit simulation as probes. All circuit evaluations are performed using a *matrix–product–state* (MPS) simulator in the sense of Vidal's tensor–network formulation [56]. Specifically, we used the `matrix_product_state` backend of the IBM Qiskit Aer simulator[52]. For reproducibility, we configured the backend in a deliberately *aggressive truncation* regime with maximum bond dimension of 128, truncation and validation thresholds of 10^{-3}, and an amplitude–chopping threshold of 10^{-3}. Even under these approximations, the CE–QAOA circuit consistently recovers optimal tours on instances of size 4 to 10 nodes. This strongly suggests that the diagonal entangler (Prop. 19) and constant-gap block–XY mixer (Prop. 7) induce useful interference patterns already at modest entanglement, aligning with the encoded 2-design baseline (Thm. 20) as a conservative control. Because truncation damps constructive interference, we expect the non-twirled peaks to sharpen under higher-fidelity simulation, increasing the single-shot success $p_{\min}$ and directly reducing the required shots at fixed confidence. Larger instances (15–25 nodes) were probed only at zero angles due to the prohibitive wall–time for full MPS simulation on commodity hardware.

The truncations we employ motivate near-term *hardware probes* on devices that admit the circuit width/depth of these instances or their compressed variants. Our grid over (γ, β), define a calibration surface for *algorithm–hardware co-design*. Compilers can target mixer locality; error-mitigation can prioritize preserving diagonal-phase coherence; and hardware can be tuned to maintain the short-depth interference that creates the observed peaks. The encoded 2-design baseline (Thm. 20) supplies a Haar-like *null model* at scale $1/D$ against which hardware-induced deviations can be quantitatively assessed. We therefore propose our CE-QAOA kernel with Tables 1 as references for *probing* future hardwares.

PHQC vs. classical exact/heuristic solvers. High-quality heuristics (like LKH implementation) routinely find near-optimal tours quickly [6], and exact methods (e.g., branch-and-cut as in Concorde) provide provably optimal solutions in practice on large benchmarks [57]. Dynamic programming à la Bellman–Held–Karp offers a systematic exact approach with worst-case complexity $T_{\mathrm{HK}} = \Theta(n^2 2^n)$ time and $M_{\mathrm{HK}} = \Theta(n2^n)$ memory [58]. By contrast, our PHQC protocol gives a *probabilistic* recovery guarantee where a per-shot optimal-hit probability $p_{\min}$ at selected shallow angles, $S \geq \lceil \ln(1/\delta)/p_{\min} \rceil$ shots suffice to obtain the optimal tour with probability at least $1 - \delta$ (Thm. 25). This is incomparable to the worst-case exponential bound of dynamic programming. While PHQC provides randomized *anytime* certificates (via repeats to boost $1 - \delta$), DP provides deterministic exactness with exponential resources. We leave a rigorous head-to-head comparison between exact solvers and PHQC to future work. However, we collect the first indication of the ability of the instance dependent

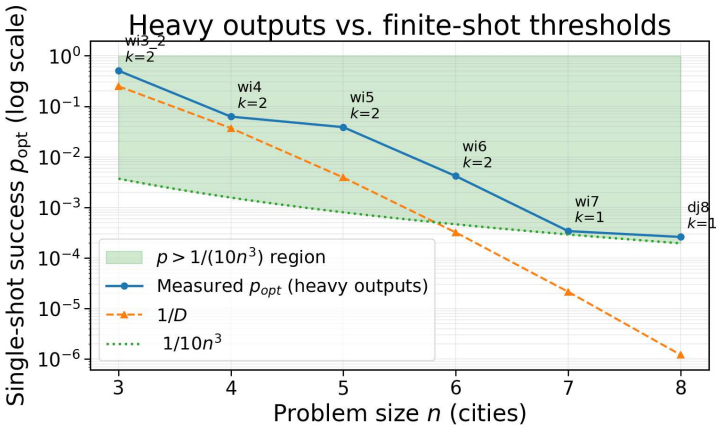


Figure 5: **Heavy outputs vs. finite-shot thresholds.** Shaded green shows the finite-shot region $p \geq \frac{1}{10n^3}$. Theorem-level guarantees (yellow boundary in schematic figures) place us above this curve, while instance-dependent heavy outputs (blue points/curve) push into the finite-shot regime Cf. 3.

Inst.	n	n_{red}	Shots	k	(γ, β)	Opt	Ver.	p_{opt}
`wi3_2`	3	2	540	2	(0.57, 0.36)	5456	PASS	5.1×10^{-1}
`wi4`	4	3	960	2	(1.57, 2.36)	6700	PASS	6.3×10^{-2}
`wi5`	5	4	1500	2	(2.20, 2.44)	6786	PASS	3.9×10^{-2}
`wi6`	6	5	2160	2	(1.01, 1.75)	9815	PASS	4.2×10^{-3}
`wi7`	7	6	2940	1	(1.35, 2.69)	7245	PASS	3.4×10^{-4}
`dj8`	8	7	3840	1	(3.14, 1.35)	2762	PASS	2.6×10^{-4}

Table 2: Empirical verification of polynomial single shot success probability from non-twirled circuit outputs. k is the degeneracy observed in the sampled bitstrings while n_{red} is teh number of blocks left after fixing the starting position. The twirling baseline $= 1/D$. The choice of $\ln(1/\delta) = 10$ corresponds to $\delta = e^{-10} \approx 4.5 \times 10^{-5}$, a very stringent failure probability ($\approx 0.005\%$). p_{opt} sums probabilities over all degenerate optimal bitstrings.

heavy outputs to keep the per-shot success probability in the finite shot region in the Table 2 and Fig. 5.

5 Conclusion

We introduced CE–QAOA, a constraint–aware kernel that (i) prepares and preserves a block one–hot manifold with an ancilla–free, depth–optimal encoder and a constant–gap two–local block–XY mixer; (ii) establishes low–order unitary moments—an exact 1–design via block–permutation twirling and an approximate 2–design from short per–block XY evolutions combined with a diagonal inter–block entangler; (iii) yields a Paley–Zygmund anticoncentration and typical–case overlap guarantees at the $1/D$ scale for encoded dimension $D = n^m$; (iv) couples naturally to a deterministic classical post–processor (PHQC) that converts overlap into an end–to–end solver with explicit shot complexity; and (v) frames a quantum–classical separation based on access models that make explicit what the classical sampler can generate efficiently. Algorithmically, PHQC identifies the best feasible sample in $O(Sn^2)$ classical time. This implies that obtaining a shot complexity independent of the encoded dimension D is the path to an unconditional quantum advantage with a polynomial time quantum-classical solver. We show empirically that in absence of permutation twirling, instance dependent interference leads to heavy ouputs in

the probability profile far above the design baseline into the polynomial shot regime.

We subsequently develop an inductive explanation to suggest that this arises in CE-QAOA when only a fraction of blocks remain unaligned on the target solution during parameter optimization. Our observation suggests that when CE–QAOA is seeded with a fixed starting block, it concentrates weight so that success depends on the number k of *unaligned* blocks at the end of state preparation. Typically $k \ll m$, and the effective scale shifts from $1/n^m$ to about $1/n^k$, yielding orders–of–magnitude gains. Thus, we obtain a $\Theta(n^r)$ shot–complexity advantage whenever CE–QAOA locks $r \geq 1$ blocks. This advantage would persist even if a classical sampler drew uniformly from the feasible space itself because classical sampling offers no mechanism to overcome the exponential suppression of success probability. We leave a rigorous proof of this phenomenon to future work.

However, when a classical competitor only samples raw $N = n^2$ bitstrings and relies on a feasibility predicate, while CE–QAOA prepares states directly in $\mathcal{H}_{\mathrm{OH}}$, the expected trial counts differ by an $\exp[\Theta(n^2)]$ factor in a minimax sense. We view these results as a step toward symmetry–inspired, shallow quantum heuristics with analyzable typical–case behavior and clean interfaces to efficient classical checking. Finally, our paper illustrates the practical impact of encoded designs in optimization problems. We observe that due to the typical behaviour results that emerge from this method, the bounds always tend to be conservative. We hope that future works are able to go beyond unitary designs to prove a worst case shot complexity independent of the encoded dimension D, paving the way for unconditional quantum advantage.

A CE-QAOA Circuit Diagrams and Block-Constraint Example

In a typical QAOA formulation with N qubits, the full Hilbert space is

$$\mathcal{H}^{\otimes N} \quad \text{with dimension} \quad 2^N.$$

For the case where $N = n^2$ (i.e., n^2 qubits), the full space has dimension

$$\dim(\mathcal{H}_{\text{full}}) = 2^{n^2}.$$

Suppose we partition the n^2 qubits into n blocks of n qubits each. We require that each block is restricted to states of the form

$$|0 \cdots 1_i \cdots 0\rangle, \quad i = 0, \ldots, n-1.$$

The projector onto the one-hot subspace for block j is defined as

$$P_n^{(j)} = \sum_{i=0}^{n-1} |0 \cdots 1_i \cdots 0\rangle_j \langle 0 \cdots 1_i \cdots 0|_j.$$

The overall projector on the full system is then

$$P = \bigotimes_{j=0}^{n-1} P_n^{(j)},$$

which restricts the Hilbert space to the tensor product of one-hot subspaces. Since each block now has dimension n, the dimension of the one-hot subspace is

$$\dim(\mathcal{H}_{\text{one-hot}}) = n^n.$$

Here we illustrate the Kernel structure of Definition 1 for the Traveling Salesman Problem[24]. Suppose we have a 3-city TSP-like problem requiring a 3×3 binary matrix encoding. In the block-wise viewpoint we need 9 qubits qrouped into 3 blocks, representing the roes of the permutation matrix (each block represents which city is in the 1^{st}, 2^{nd}, or 3^{rd} position). Each block has 3 qubits (exactly one of the 3 qubits is $|1\rangle$, the others $|0\rangle$). A valid bitstring in the computational basis then takes the form

$$(\, b_0^0 \; b_0^1 \; b_0^2) \mid (\, b_1^0 \; b_1^1 \; b_1^2) \mid (\, b_2^0 \; b_2^1 \; b_2^2),$$

where each parenthetical group has exactly one $|1\rangle$. For example:

$$\underbrace{100}_{\text{Block 0}} \; \underbrace{010}_{\text{Block 1}} \; \underbrace{001}_{\text{Block 2}},$$

$$\underbrace{010}_{\text{Block 0}} \; \underbrace{100}_{\text{Block 1}} \; \underbrace{001}_{\text{Block 2}},$$

$$\underbrace{001}_{\text{Block 0}} \; \underbrace{100}_{\text{Block 1}} \; \underbrace{010}_{\text{Block 2}},$$

etc. Each triple of bits has exactly one $|1\rangle$ and so on.

Suppose we have a 4-city TSP-like problem requiring a 4×4 binary matrix encoding. The block structure in the computational basis then takes the form

$$(b_0^0 \; b_0^1 \; b_0^2 \; b_0^3) \mid (b_1^0 \; b_1^1 \; b_1^2 \; b_1^3) \mid (b_2^0 \; b_2^1 \; b_2^2 \; b_2^3) \mid (b_3^0 \; b_3^1 \; b_3^2 \; b_3^3),$$

where each parenthetical group has exactly one $|1\rangle$. It is worth emphasizing again that not all the bitstrings satisfying this blockwise constraint are valid tours, however, all valid tours satisfy this structure. So for a TSP, the constraint is only partially encoded.

A.1 Permutation Group Actions

Global feasibility–preserving action. Let $\mathcal{A} \equiv S_m \times S_n \subset U(\mathcal{H}_{\text{OH}})$, act by

$$\mathsf{P}_{\sigma,\tau} \left(|e_{j_0}\rangle \otimes \cdots \otimes |e_{j_{m-1}}\rangle \right) = \left| e_{\tau(j_{\sigma^{-1}(0)})} \right\rangle \otimes \cdots \otimes \left| e_{\tau(j_{\sigma^{-1}(m-1)})} \right\rangle, \tag{16}$$

i.e., a row permutation $\sigma \in S_m$ and the *same* column relabeling $\tau \in S_n$ applied to every block. When H_{pen} is $S_m \times S_n$–invariant (Def. 1), the feasible set

$$\mathcal{X} \;:=\; L_0(H_{\text{pen}}) = \{\, x : \langle x|H_{\text{pen}}|x\rangle = 0 \,\}$$

is preserved setwise by $\mathcal{A}$.

Block Level feasibility–preserving action.

Let

$$\mathcal{B} \;\equiv\; S_n^{\times m} \;=\; \Big\{\, \mathsf{P} = \bigotimes_{b=0}^{m-1} \mathsf{P}_b \;:\; \mathsf{P}_b \in S_n \,\Big\} \;\subset\; U(\mathcal{H}_{\text{OH}}),$$

which applies *independent* label permutations in each block. This action is *not* feasibility–preserving in general (it can break column one–hot), but it does implement a first–moment (1-design) average on the encoded space. By a *permutation twirl* we mean the feasibility–preserving group action

$$\mathcal{A} \;=\; S_m^{(\text{rows})} \times S_n^{(\text{global cols})},$$

which (i) permutes the *rows/blocks* uniformly via a single $\pi \in S_m$ and (ii) applies the *same* column relabeling $\tau \in S_n$ to *every* row. This action preserves the set of permutation matrices (feasible states).

Recall. If we encode an $m \times n$ assignment (rows $b \in [m]$, columns $j \in [n]$) by an m-block one–hot pattern with n qubits per block. A feasible bitstring is a *permutation matrix* $X \in \{0,1\}^{m\times n}$ with exactly one "1" in each row and each column. Let S_m act on rows (blocks) and S_n act on columns (symbols) with action on X given by

$$(R_\pi X)_{b,j} = X_{\pi^{-1}(b),\,j}, \qquad \pi \in S_m, \tag{17}$$

$$(C_\tau X)_{b,j} = X_{b,\,\tau^{-1}(j)}, \qquad \tau \in S_n, \tag{18}$$

and combined action $X \mapsto R_\pi C_\tau X$.

Proposition 26 (Global Feasibility is preserved by $\mathcal{A}$)**.** *If X is a permutation matrix and $(\pi,\tau) \in \mathcal{A}$, then $X' := R_\pi C_\tau X$ is also a permutation matrix.*

Proof. Any feasible X encodes a permutation $\sigma : [m] \to [n]$ via $X_{b,j} = 1 \iff j = \sigma(b)$. Under the group action,

$(R_\pi C_\tau X)_{b,j} = 1 \iff X_{\pi^{-1}(b),\,\tau^{-1}(j)} = 1 \iff \tau^{-1}(j) = \sigma(\pi^{-1}(b)) \iff j = \tau\big(\sigma(\pi^{-1}(b))\big)$.

Thus $R_\pi C_\tau X$ encodes the permutation $\tau \circ \sigma \circ \pi^{-1}$. It remains one-hot in each row and each column, i.e. feasible. □

For a qubit-level description, label qubits as (b,j) with $b \in [m]$ (block/row) and $j \in [n]$ (column). Then

$$(b,j) \xmapsto{R_\pi} (\pi(b),\, j), \tag{19}$$

$$(b,j) \xmapsto{C_\tau} (b,\, \tau(j)), \tag{20}$$

and the unitary implementing $R_\pi C_\tau$ is a permutation matrix over the full computational basis. Because H_{pen} is invariant under $\mathcal{A}$, its zero-level set is stable under these unitaries. This is the precise sense in which "the permutation preserves feasibility" in our kernel.

3×3 **(9-qubit) example** Take $m = n = 3$, so a feasible state is a 3×3 permutation matrix. Start from $\sigma = \mathrm{id}$,

$$X \;=\; \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Apply a row cycle $\pi = (1\,2\,3)$: $R_\pi X$ is $\left(\begin{smallmatrix} 0&1&0\\ 0&0&1\\ 1&0&0 \end{smallmatrix}\right)$, still feasible. Apply a *global* column swap $\tau = (1\,3)$ to all rows: $C_\tau X = \left(\begin{smallmatrix} 0&0&1\\ 0&1&0\\ 1&0&0 \end{smallmatrix}\right)$, still feasible. The combined action $R_\pi C_\tau X$ remains a permutation matrix, encoding $\tau \circ \sigma \circ \pi^{-1}$.

A.2 A Quick Schmidt–Rank Primer

Definition 27 (Schmidt decomposition and Schmidt rank)**.** *Let $|\psi\rangle \in \mathcal{H}_L \otimes \mathcal{H}_R$ be a pure state. There exist orthonormal sets $\{|u_k\rangle\}_{k=1}^r \subset \mathcal{H}_L$, $\{|v_k\rangle\}_{k=1}^r \subset \mathcal{H}_R$, and strictly positive numbers $\lambda_1, \ldots, \lambda_r$ such that*

$$|\psi\rangle \;=\; \sum_{k=1}^{r} \lambda_k \; |u_k\rangle_L \otimes |v_k\rangle_R \,.$$

The integer r is the Schmidt rank*, denoted* $\mathrm{SR}(|\psi\rangle)$ *(with respect to the bipartition $L|R$).*

Matrix-reshaping. Fix product bases on L and R. If one reshapes the amplitude tensor of $|\psi\rangle$ into a matrix $M \in \mathbb{C}^{(\dim \mathcal{H}_L)\times(\dim \mathcal{H}_R)}$ by grouping all "left" indices as rows and all "right" indices as columns, then

$$\mathrm{SR}(|\psi\rangle) \;=\; \mathrm{rank}(M).$$

Thus, the Schmidt rank of a state equals the usual linear-algebra rank of its reshaped coefficient matrix.

Example: $|0\rangle^{\otimes n}$ has Schmidt rank 1 across any single-edge cut. For any cut $\{0,\ldots,i-1\}\,|\,\{i,\ldots,n-1\}$ one has the factorization

$$|0\rangle^{\otimes n} \;=\; \left(|0\rangle^{\otimes i}\right) \otimes \left(|0\rangle^{\otimes (n-i)}\right),$$

i.e. a *single* product term. Hence SR = 1. In the matrix view, the reshaped matrix has one nonzero entry (at the all-zero row/column), so its rank is 1.

Example: $|W_n\rangle$ has Schmidt rank 2 across any single-edge cut. Recall

$$|W_n\rangle \;=\; \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} |0\cdots 1_j \cdots 0\rangle \,.$$

Across the cut $\{0,\ldots,i-1\}\,|\,\{i,\ldots,n\}$ split the excitation "on the left" vs. "on the right":

$$|W_n\rangle \;=\; \sqrt{\frac{i}{n}}\; |L_1\rangle \otimes |0\cdots 0\rangle_R \;+\; \sqrt{\frac{n-i}{n}}\; |0\cdots 0\rangle_L \otimes |R_1\rangle \,,$$

$$|L_1\rangle = \frac{1}{\sqrt{i}} \sum_{j=0}^{i-1} |0\cdots 1_j \cdots 0\rangle_L \,, \qquad |R_1\rangle = \frac{1}{\sqrt{n-i}} \sum_{j=i}^{n-1} |0\cdots 1_j \cdots 0\rangle_R$$

are orthonormal to the respective all-zero vectors. The two product terms are orthogonal and both coefficients are nonzero for $1 \le i \le n-1$, hence $\mathrm{SR}(|W_n\rangle) = 2$. In the matrix view, the reshaped matrix has two nonzero singular values $\sqrt{i/n}$ and $\sqrt{(n-i)/n}$.

A.3 Two–local and number–conserving structure

On a given block b with sites $i \in \{0,\ldots,n-1\}$, define the Pauli ladder operators

$$\sigma_i^+ \;=\; |1\rangle\langle 0|_i \;=\; \tfrac{1}{2}(X_i - iY_i), \qquad \sigma_i^- \;=\; |0\rangle\langle 1|_i \;=\; \tfrac{1}{2}(X_i + iY_i).$$

A straightforward calculation gives

$$X_iX_j + Y_iY_j \;=\; 2(\sigma_i^+\sigma_j^- + \sigma_i^-\sigma_j^+), \qquad (i \neq j). \tag{21}$$

Hence the block mixer can be written (up to an overall factor 2 absorbable into the angle)

$$H_M^{(b)} := \sum_{0\le i<j\le n-1} (X_i^{(b)} X_j^{(b)} + Y_i^{(b)} Y_j^{(b)}) = 2 \sum_{0\le i<j\le n-1} (\sigma_{bi}^+ \sigma_{bj}^- + \sigma_{bi}^- \sigma_{bj}^+). \tag{22}$$

Action on the one–excitation basis. Let $\mathcal{H}_1 = \mathrm{span}\{|e_0\rangle, \ldots, |e_{n-1}\rangle\}$ denote the one–hot subspace on the block, where $|e_k\rangle$ has a single '1' at position k. For $i \neq j$,

$$\sigma_j^+ |e_i\rangle = |\ldots 1_j, 1_i, \ldots\rangle, \qquad \sigma_i^- \sigma_j^+ |e_i\rangle = |e_j\rangle, \qquad \sigma_i^+ \sigma_j^- |e_i\rangle = 0.$$

Using (21),

$$(X_i X_j + Y_i Y_j) |e_i\rangle = 2\, \sigma_i^- \sigma_j^+ |e_i\rangle = 2\, |e_j\rangle. \tag{23}$$

By symmetry, $(X_i X_j + Y_i Y_j) |e_j\rangle = 2\, |e_i\rangle$, and for any $k \notin \{i, j\}$, $(X_i X_j + Y_i Y_j) |e_k\rangle = 0$. Thus, *restricted to* $\mathcal{H}_1$, $H_M^{(b)}$ acts as the (weighted) adjacency of the complete graph K_n:

$$H_M^{(b)}|_{\mathcal{H}_1} = 2 \sum_{0\le i<j\le n-1} (|e_i\rangle\langle e_j| + |e_j\rangle\langle e_i|) = 2\, A(K_n). \tag{24}$$

Excitation number preservation is straight forward. Let $\hat{N} = \sum_{i=0}^{n-1} \sigma_i^+ \sigma_i^-$ be the excitation–number operator on the block. Each term $\sigma_i^+ \sigma_j^-$ lowers at j and raises at i, keeping the total count unchanged; likewise for its Hermitian conjugate. Hence

$$[\, H_M^{(b)}, \hat{N}\,] = 0,$$

so $H_M^{(b)}$ preserves all fixed–Hamming–weight sectors, in particular $\mathcal{H}_1$.

A.4 Proof of Proposition 12

Proof. **Step 1 (Representation on $\mathcal{H}_1$).** With basis $\{|e_k\rangle\}_{k=0}^{n-1}$, the XY terms act as

$$H_{ij}^{(b)}|_{\mathcal{H}_1} = E_{ij} + E_{ji}, \qquad E_{ij} := |e_i\rangle\langle e_j|,$$

and $H_Z^{(1)}|_{\mathcal{H}_1} = D := \sum_{k=0}^{n-1} d_k E_{kk}$ is diagonal. "Nontrivial" means $D \not\propto I$.

Step 2 (Skew-symmetric off-diagonals). For any pair (i, j) with $d_i \neq d_j$,

$$[D,\ E_{ij} + E_{ji}] = (d_i - d_j)\,(E_{ij} - E_{ji}),$$

so $E_{ij} - E_{ji}$ lies in the Lie closure $\mathfrak{L} := \mathrm{Lie}\{\, i\mathcal{G}\}$.

Step 3 (Traceless diagonals). With $S_{ij} := E_{ij} + E_{ji}$ and $A_{ij} := E_{ij} - E_{ji}$ obtained as above for some pair (i, j) with $d_i \neq d_j$, compute

$$[S_{ij},\ A_{ij}] = 2\,(E_{ii} - E_{jj}) \in \mathfrak{L}.$$

Hence $\Delta_{ij} := E_{ii} - E_{jj} \in \mathfrak{L}$ for at least one (i, j). Since commutators of such Δ's generate the linear span of all traceless diagonals, we can obtain

$$\mathrm{span}\{\, E_{ii} - E_{jj} \,:\, i \neq j\} \subseteq \mathfrak{L}.$$

Step 4 (All skew-symmetric off-diagonals). Given any $k \neq \ell$, use the available XY generator $S_{k\ell} = E_{k\ell} + E_{\ell k} \in \mathfrak{L}$ and a traceless diagonal that separates k and ℓ, namely $\Delta_{k\ell} = E_{kk} - E_{\ell\ell} \in \mathfrak{L}$ from Step 3, to get

$$[\Delta_{k\ell},\ S_{k\ell}] = 2\,(E_{k\ell} - E_{\ell k}) \in \mathfrak{L}.$$

Thus for every $k \neq \ell$ we have both $S_{k\ell}$ and $A_{k\ell}$.

Step 5 (Lie algebra identification). The set

$$\{\, i(E_{k\ell} + E_{\ell k}),\ i(E_{k\ell} - E_{\ell k})\ (k < \ell);\ i(E_{kk} - E_{\ell\ell})\ (k < \ell)\,\}$$

spans $\mathfrak{su}(n)$. Steps 3–4 show all these elements lie in $\mathfrak{L}$, hence $\mathfrak{L} = \mathfrak{su}(n)$.

Step 6 (Synthesis). By standard Lie–algebraic controllability (e.g., [59]), the connected Lie group generated by exponentials of elements in $i\mathcal{G}$ has Lie algebra $\mathfrak{su}(n)$, hence equals $\mathrm{SU}(n)$. Density plus compactness of $\mathrm{SU}(n)$ yields the stated approximation of any $V \in \mathrm{SU}(n)$ by a finite product of $\exp(-i\theta H)$ with $H \in \mathcal{G}$ to arbitrary precision $\varepsilon > 0$. □

A.5 On the Moment-operator gap used in the proof

For one random layer U, define the second-moment operator

$$M_2 \;:=\; \mathbb{E}\Big[U^{\otimes 2} \otimes \overline{U}^{\otimes 2}\Big].$$

This is the averaging operator that governs the evolution of second moments under the circuit ensemble. The corresponding Haar benchmark is

$$M_2^{\mathrm{Haar}} \;:=\; \int_{U(d)} V^{\otimes 2} \otimes \overline{V}^{\otimes 2}\, d\mu_{\mathrm{Haar}}(V), \qquad d = n^m,$$

which is the projector onto the invariant subspace of the two-copy representation.

In this language, the *moment-operator gap* is the spectral gap of M_2 away from the Haar-fixed subspace. More precisely, if λ_* denotes the largest nontrivial eigenvalue of M_2 in magnitude on the orthogonal complement of the Haar-invariant space, then

$$\Delta_2 \;:=\; 1 - \lambda_*.$$

Equivalently, when the reduced dynamics is represented by a reversible Markov kernel, one may write

$$\Delta_2 = 1 - \lambda_2,$$

where λ_2 is the second-largest eigenvalue. A positive lower bound on Δ_2 implies exponential convergence of iterated moments:

$$\|M_2^L - M_2^{\mathrm{Haar}}\| \;\le\; (1 - \Delta_2)^L,$$

up to the perturbative error inherited from the per-block approximate 2-designs.

In the present construction, the local block twirls reduce the relevant second-moment space to the commutant basis

$$\{I, \mathrm{SWAP}\}^m.$$

After this reduction, the action of one layer is encoded by an induced Markov kernel on $\{I, \mathrm{SWAP}\}^m$. Thus, in the proof, the moment-operator gap can be identified with the spectral gap of this reduced kernel, provided the reduction preserves the Haar fixed space and the nontrivial sector. Under connectivity of the block-interaction graph and mild randomness in the entangling layer, this is the standard mechanism by which one obtains inverse-polynomial mixing and hence convergence to an approximate global unitary 2-design [20, 55]. Therefore, more precisely, the proof uses a *second-moment spectral-gap argument* as follows: first identify the fixed space with the Haar second moment, then show that the remaining spectrum is bounded away from 1 by at least an inverse polynomial. This gap is exactly what yields the depth bound

$$L = O\Big(\Delta_2^{-1} \log(1/\varepsilon_{\mathrm{glob}})\Big),$$

up to the approximation error of the local block designs.

A.6 A derivation of the single-target second moment under a unitary 2-design

Let D be the dimension of the encoded Hilbert space. Fix a normalized input state

$$|\phi\rangle \in \mathcal{H}_{\mathrm{OH}}$$

and a fixed target basis state

$$|x^\star\rangle \in \mathcal{H}_{\mathrm{OH}}.$$

Let

$$X := \big|\langle x^\star \mid U \mid \phi\rangle\big|^2,$$

where U is drawn either from Haar measure on $\mathrm{U}(D)$ or from an exact unitary 2-design.

Define the rank-one projectors

$$\rho := |\phi\rangle\langle\phi|\,, \qquad \Pi_{x^\star} := |x^\star\rangle\langle x^\star|\,.$$

Then

$$X = \mathrm{Tr}(\Pi_{x^\star} U\rho U^\dagger). \tag{25}$$

First moment. For Haar measure, or any exact unitary 1-design,

$$\mathbb{E}_U[U\rho U^\dagger] = \frac{\mathbb{I}}{D}.$$

Hence

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}_U\Big[\mathrm{Tr}(\Pi_{x^\star} U\rho U^\dagger)\Big] = \mathrm{Tr}\Big(\Pi_{x^\star}\,\mathbb{E}_U[U\rho U^\dagger]\Big)\\ &= \mathrm{Tr}\left(\Pi_{x^\star}\frac{\mathbb{I}}{D}\right) = \frac{\mathrm{Tr}(\Pi_{x^\star})}{D} = \frac{1}{D}.\end{aligned} \tag{26}$$

Second moment. Starting from (25),

$$X^2 = \mathrm{Tr}(\Pi_{x^\star} U\rho U^\dagger)^2.$$

Using the tensor-product trace identity

$$\mathrm{Tr}(A)\,\mathrm{Tr}(B) = \mathrm{Tr}(A\otimes B),$$

we obtain

$$\begin{aligned}X^2 &= \mathrm{Tr}(\Pi_{x^\star} U\rho U^\dagger)\ \mathrm{Tr}(\Pi_{x^\star} U\rho U^\dagger)\\ &= \mathrm{Tr}\Big((\Pi_{x^\star}\otimes\Pi_{x^\star})(U\rho U^\dagger)^{\otimes 2}\Big).\end{aligned} \tag{27}$$

Therefore

$$\mathbb{E}[X^2] = \mathrm{Tr}\Big((\Pi_{x^\star}\otimes\Pi_{x^\star})\,\mathbb{E}_U\big[(U\rho U^\dagger)^{\otimes 2}\big]\Big)\,. \tag{28}$$

For Haar measure, or any exact unitary 2-design,

$$\mathbb{E}_U\big[(U\rho U^\dagger)^{\otimes 2}\big] = \frac{\mathbb{I}+\mathbb{F}}{D(D+1)}, \tag{29}$$

where $\mathbb{F}$ is the swap operator,

$$\mathbb{F}(|a\rangle\otimes|b\rangle) = |b\rangle\otimes|a\rangle\,.$$

Substituting (29) into (28) gives

$$\begin{aligned}\mathbb{E}[X^2] &= \frac{1}{D(D+1)}\,\mathrm{Tr}\Big((\Pi_{x^\star}\otimes\Pi_{x^\star})(\mathbb{I}+\mathbb{F})\Big)\\ &= \frac{1}{D(D+1)}\left[\mathrm{Tr}(\Pi_{x^\star}\otimes\Pi_{x^\star}) + \mathrm{Tr}\big((\Pi_{x^\star}\otimes\Pi_{x^\star})\mathbb{F}\big)\right].\end{aligned} \tag{30}$$

Now evaluate the two traces separately.

First, since $\Pi_{x^\star}$ is rank one,

$$\mathrm{Tr}(\Pi_{x^\star}\otimes\Pi_{x^\star}) = \mathrm{Tr}(\Pi_{x^\star})^2 = 1. \tag{31}$$

Second, use the standard identity

$$\mathrm{Tr}((A\otimes B)\mathbb{F}) = \mathrm{Tr}(AB).$$

With $A = B = \Pi_{x^\star}$, this gives

$$\mathrm{Tr}\big((\Pi_{x^\star}\otimes\Pi_{x^\star})\mathbb{F}\big) = \mathrm{Tr}\left(\Pi_{x^\star}^2\right) = \mathrm{Tr}(\Pi_{x^\star}) = 1. \tag{32}$$

Substituting (31) and (32) into (30), we obtain

$$\boxed{\mathbb{E}[X^2] = \frac{2}{D(D+1)}.} \tag{33}$$

Paley–Zygmund consequence. Using (26) and (33), Paley–Zygmund gives

$$\Pr\left[X \geq \frac{1}{2D}\right] \geq \frac{\left(\mathbb{E}[X] - \frac{1}{2D}\right)^2}{\mathbb{E}[X^2]} = \frac{(1/2D)^2}{2/(D(D+1))} = \frac{D+1}{8D} = \Omega(1). \tag{34}$$

Approximate 2-design overlap. If U is an ε-approximate unitary 2-design on the encoded space, then the second-moment operator differs from the Haar value by an error term Δ_ε:

$$\mathbb{E}_{U\sim\mathsf{U}}[(U\rho U^\dagger)^{\otimes 2}] = \frac{\mathbb{I}+\mathbb{F}}{D(D+1)} + \Delta_\varepsilon,$$

with $\|\Delta_\varepsilon\|$ controlled by the approximation parameter in the norm specified by the design theorem. In that case,

$$\begin{aligned}\mathbb{E}[X^2] &= \mathrm{Tr}\left((\Pi_{x^\star}\otimes\Pi_{x^\star})\left[\frac{\mathbb{I}+\mathbb{F}}{D(D+1)} + \Delta_\varepsilon\right]\right) \\ &= \frac{2}{D(D+1)} + \mathrm{Tr}\big((\Pi_{x^\star}\otimes\Pi_{x^\star})\Delta_\varepsilon\big) \\ &= \frac{2}{D(D+1)} \pm O(\varepsilon),\end{aligned} \tag{35}$$

where the implicit constant depends on the precise definition of approximate 2-design and the norm in which the deviation is controlled. Similarly,

$$\mathbb{E}[X] = \frac{1}{D} \pm O(\varepsilon).$$

A.7 Rank-S projector version

The same calculation applies when the rank-one target projector $\Pi_{x^\star}$ is replaced by a projector P of rank S, for example the feasible projector P_Π.

Define

$$Y := \langle\phi|\, U^\dagger P U\, |\phi\rangle = \mathrm{Tr}\Big(PU\rho U^\dagger\Big).$$

Then

$$\mathbb{E}[Y] = \frac{S}{D},$$

and

$$\begin{aligned}\mathbb{E}[Y^2] &= \mathrm{Tr}\left((P\otimes P)\frac{\mathbb{I}+\mathbb{F}}{D(D+1)}\right) \\ &= \frac{\mathrm{Tr}(P\otimes P) + \mathrm{Tr}((P\otimes P)\mathbb{F})}{D(D+1)} \\ &= \frac{(\mathrm{Tr}\,P)^2 + \mathrm{Tr}(P^2)}{D(D+1)}.\end{aligned} \tag{36}$$

If P is a projector of rank S, then

$$\mathrm{Tr}(P) = S, \qquad \mathrm{Tr}\Big(P^2\Big) = S,$$

so

$$\boxed{\mathbb{E}[Y^2] = \frac{S^2+S}{D(D+1)}.} \tag{37}$$

In particular,

$$\mathbb{E}[Y^2] = \Big(\frac{S}{D}\Big)^2 \frac{1+1/S}{1+1/D} \leq 2\Big(\frac{S}{D}\Big)^2$$

References

[1] Christos H. Papadimitriou and Kenneth Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Prentice Hall, 1982.

[2] Manfred Padberg and Giovanni Rinaldi. "A Branch-and-Cut Algorithm for the Resolution of Large-Scale Symmetric Travelling Salesman Problems". In: *SIAM Review* 33.1 (1991), pp. 60–100. DOI: 10.1137/1033004.

[3] Paolo Toth and Daniele Vigo. *Vehicle Routing: Problems, Methods, and Applications*. 2nd ed. MOS-SIAM Series on Optimization. SIAM, 2014. DOI: 10.1137/1.9781611973594.

[4] Yoshua Bengio, Andrea Lodi, and Antoine Prouvost. "Machine Learning for Combinatorial Optimization: A Methodological Tour d'Horizon". In: *European Journal of Operational Research* 290.2 (2021), pp. 405–421. DOI: 10.1016/j.ejor.2020.07.063.

[5] Eugene L. Lawler and David E. Wood. "Branch-and-Bound Methods: A Survey". In: *Operations Research* 14.4 (1966), pp. 699–719. DOI: 10.1287/opre.14.4.699.

[6] Shen Lin and Brian W. Kernighan. "An Effective Heuristic Algorithm for the Travelling-Salesman Problem". In: *Operations Research* 21.2 (1973), pp. 498–516. DOI: 10.1287/opre.21.2.498.

[7] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. "A Quantum Approximate Optimization Algorithm". In: *arXiv preprint arXiv:1411.4028* (2014).

[8] Juan A. Montañez-Barrera and Kristel Michielsen. *Towards a Universal QAOA Protocol: Evidence of a Scaling Advantage in Solving Some Combinatorial Optimization Problems*. https://arxiv.org/abs/2405.09169. 2024. arXiv: 2405.09169 [quant-ph].

[9] Eunok Bae and Soojoon Lee. "Recursive QAOA outperforms the original QAOA for the MAX-CUT problem on complete graphs". In: *Quantum Information Processing* 23.3 (2024), p. 78. DOI: 10.1007/s11128-024-04286-0.

[10] Jernej Rudi Finžgar et al. "Quantum-Informed Recursive Optimization Algorithms". In: *PRX Quantum* 5.2 (May 2024), p. 020327. DOI: 10.1103/PRXQuantum.5.020327. URL: https://link.aps.org/doi/10.1103/PRXQuantum.5.020327.

[11] M. Cerezo et al. "Variational quantum algorithms". In: *Nature Reviews Physics* 3.9 (2021), pp. 625–644. DOI: 10.1038/s42254-021-00348-9.

[12] J. Tilly et al. "The Variational Quantum Eigensolver: A review of methods and best practices". In: *PRX Quantum* 3.3 (2022), p. 030204. DOI: 10.1103/PRXQuantum.3.030204.

[13] Jarrod R. McClean et al. "Barren plateaus in quantum neural network training landscapes". In: *Nature Communications* 9.1 (2018), p. 4812. DOI: 10.1038/s41467-018-07090-4.

[14] Stuart Hadfield et al. "From the Quantum Approximate Optimization Algorithm to a Quantum Alternating Operator Ansatz". In: *Algorithms* 12.2 (2019), p. 34. DOI: 10.3390/a12020034.

[15] Franz G. Fuchs and Ruben Pariente Bassa. "Constraint Preserving Mixers for the Quantum Approximate Optimization Algorithm". In: *Algorithms* 15.6 (2022), p. 202. DOI: 10.3390/a15060202.

[16] Andreas Bärtschi and Stephan Eidenbenz. *Grover Mixers for QAOA: Shifting Complexity from Mixer Design to State Preparation*. arXiv:2006.00354. 2020. DOI: 10.48550/arXiv.2006.00354.

[17] B. Tsvelikhovskiy, I. Safro, and Y. Alexeev. "Equivariant QAOA and the Duel of the Mixers". In: *arXiv:2405.07211* (2024). arXiv: 2405.07211 [quant-ph].

[18] B. Tsvelikhovskiy, I. Safro, and Y. Alexeev. "Symmetries and Dimension Reduction in Quantum Approximate Optimization Algorithm". In: *arXiv preprint arXiv:2309.13787* (2023). arXiv: 2309.13787 [quant-ph].

[19] Ningyi Xie and Hoong Chuin Lau. *A Feasibility-Preserved Quantum Approximate Solver for the Capacitated Vehicle Routing Problem*. arXiv:2308.08785. 2024. URL: https://arxiv.org/abs/2308.08785.

[20] Fernando G. S. L. Brandão and Aram W. Harrow. "Local random quantum circuits are approximate polynomial-designs". In: *Communications in Mathematical Physics* 346.2 (2016), pp. 397–434.

[21] A. Pérez-Salinas, H. Wang, and X. Bonet-Monroig. "Analyzing variational quantum landscapes with information content". In: *npj Quantum Information* 10 (Feb. 2024), p. 27. DOI: 10.1038/s41534-024-00819-8.

[22] Stuart Hadfield, Tad Hogg, and Eleanor G. Rieffel. "Analytical framework for quantum alternating operator ansätze". In: *Quantum Science and Technology* 8.1 (Dec. 2022), p. 015017. ISSN: 2058-9565. DOI: 10.1088/2058-9565/aca3ce.

[23] Rafael S. do Carmo et al. *Warm-Starting QAOA with XY Mixers: A Novel Approach for Quantum-Enhanced Vehicle Routing Optimization*. arXiv preprint arXiv:2504.19934. 2025. arXiv: 2504.19934 [quant-ph].

[24] Andrew Lucas. "Ising formulations of many NP problems". In: *Frontiers in Physics* 2 (2014), p. 5. DOI: 10.3389/fphy.2014.00005.

[25] Jack Edmonds. "Paths, Trees, and Flowers". In: *Canadian Journal of Mathematics* 17 (1965), pp. 449–467. DOI: 10.4153/CJM-1965-045-4.

[26] Harold W. Kuhn. "The Hungarian Method for the Assignment Problem". In: *Naval Research Logistics Quarterly* 2.1-2 (1955), pp. 83–97. DOI: 10.1002/nav.3800020109.

[27] Chinonso Onah et al. "QUEST: QUantum-Enhanced Shared Transportation". In: *2025 IEEE International Conference on Quantum Computing and Engineering (QCE)*. Vol. 01. 2025, pp. 2149–2160. DOI: 10.1109/QCE65121.2025.00235.

[28] Gushu Li, Yufei Ding, and Yuan Xie. "On the Co-Design of Quantum Software and Hardware". In: *ICCAD '21: IEEE/ACM International Conference on Computer-Aided Design*. 2021. DOI: 10.1145/3477206.3477464.

[29] Teague Tomesh and Margaret Martonosi. "Quantum Codesign". In: *IEEE Micro* 41.5 (2021), pp. 33–40. DOI: 10.1109/MM.2021.3094461.

[30] Eneko Osaba and Erika Villar-Rodríguez. "QOPTLib: a Quantum-Computing-Oriented Benchmark for Combinatorial Optimisation Problems". In: *Proc. Quantum Tech. 2024*. 2024.

[31] Chinonso Onah, Roman Firt, and Kristel Michielsen. *Dataset: Empirical Quantum Advantage in Constrained Optimization from Encoded Unitary Designs (0.1)*. Version 0.1. 2025. DOI: 10.5281/zenodo.15725265. URL: https://doi.org/10.5281/zenodo.15725265.

[32] Kate A. Smith-Miles et al. “The travelling salesperson problem and the challenges of near-term quantum advantage”. In: *Quantum Science and Technology* 10.3 (May 2025), p. 033001. DOI: 10.1088/2058-9565/add61d.

[33] Nicholas Chancellor. “Circuit design for multi-body interactions in superconducting quantum annealing systems with applications to a scalable architecture”. In: *npj Quantum Information* (2017).

[34] Ryan LaRose, M. Cerezo, Piotr Czarnik, et al. “Mixer-phaser Ansätze for quantum optimization with hard constraints”. In: *Quantum Machine Intelligence* 4.2 (2022), p. 12. DOI: 10.1007/s42484-022-00069-x.

[35] Rafael S. do Carmo et al. *Warm-Starting QAOA with XY Mixers: A Novel Approach for Quantum-Enhanced Vehicle Routing Optimization*. 2025. arXiv: 2504.19934 [quant-ph].

[36] Daniel J. Egger et al. “Warm-starting quantum optimization”. In: *PRX Quantum* 2 (2021), p. 040348.

[37] Zichang He et al. “Alignment between initial state and mixer improves QAOA performance for constrained optimization”. In: *npj Quantum Information* 9.1 (Nov. 2023). ISSN: 2056-6387. DOI: 10.1038/s41534-023-00787-5. URL: https://doi.org/10.1038/s41534-023-00787-5.

[38] Zhihui Wang et al. “XY mixers: Analytical and numerical results for the quantum approximate optimization algorithm”. In: *Physical Review A* 101.1 (2020), p. 012320. DOI: 10.1103/PhysRevA.101.012320.

[39] Andreas Bärtschi and Stephan Eidenbenz. “Deterministic Preparation of Dicke States”. In: *Fundamentals of Computation Theory*. Springer International Publishing, 2019, pp. 126–139. ISBN: 9783030250270. DOI: 10.1007/978-3-030-25027-0_9.

[40] Eugene L. Lawler et al. *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*. Wiley, 1985.

[41] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.

[42] Tjalling C. Koopmans and Martin Beckmann. “Assignment Problems and the Location of Economic Activities”. In: *Econometrica* 25.1 (1957), pp. 53–76. DOI: 10.2307/1907742.

[43] E. M. Loiola et al. “A Survey for the Quadratic Assignment Problem”. In: *European Journal of Operational Research* 176.2 (2007), pp. 657–690. DOI: 10.1016/j.ejor.2005.09.032.

[44] Sartaj Sahni and Teofilo Gonzalez. “P-Complete Approximation Problems”. In: *Journal of the ACM* 23.3 (1976), pp. 555–565. DOI: 10.1145/321958.321975.

[45] Silvano Martello and Paolo Toth. *Knapsack Problems: Algorithms and Computer Implementations*. Wiley, 1990.

[46] Richard M. Karp. “Reducibility among Combinatorial Problems”. In: *Complexity of Computer Computations*. Plenum, 1972, pp. 85–103.

[47] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York: W. H. Freeman and Company, 1979.

[48] Michael R. Garey, David S. Johnson, and Ravi Sethi. “The Complexity of Flowshop and Jobshop Scheduling”. In: *Mathematics of Operations Research* 1.2 (1976), pp. 117–129. DOI: 10.1287/moor.1.2.117.

[49] Jan Karel Lenstra, Alexander H. G. Rinnooy Kan, and Peter Brucker. “Complexity of Machine Scheduling Problems”. In: *Annals of Discrete Mathematics* 1 (1977), pp. 343–362. DOI: 10.1016/S0167-5060(08)70743-X.

[50] James Munkres. “Algorithms for the Assignment and Transportation Problems”. In: *Journal of the Society for Industrial and Applied Mathematics* 5.1 (1957), pp. 32–38. DOI: 10.1137/0105003.

[51] Firat Diker. *Deterministic construction of arbitrary W states with quadratically increasing number of two-qubit gates*. 2022. arXiv: 1606.09290 [quant-ph]. URL: https://arxiv.org/abs/1606.09290.

[52] M. Ayrton et al. *Qiskit: An Open-source Framework for Quantum Computing*. Version 0.47. 2023. URL: https://qiskit.org.

[53] S. Bravyi and D. Gosset. “Improved classical simulation of quantum circuits dominated by Clifford gates”. In: *Phys. Rev. Lett.* 116 (2016), p. 250501.

[54] Chris Godsil and Gordon Royle. *Algebraic Graph Theory*. Vol. 207. Graduate Texts in Mathematics. New York: Springer, 2001.

[55] Jean Bourgain and Alex Gamburd. “A Spectral Gap Theorem in SU(d)”. In: *J. Eur. Math. Soc.* 14.5 (2012), pp. 1455–1511.

[56] Guifré Vidal. “Efficient Classical Simulation of Slightly Entangled Quantum Computations”. In: *Phys. Rev. Lett.* 91.14 (2003), p. 147902. DOI: 10.1103/PhysRevLett.91.147902.

[57] David L. Applegate et al. *The Traveling Salesman Problem: A Computational Study*. Princeton, NJ: Princeton University Press, 2007.

[58] Richard Bellman. “Dynamic Programming Treatment of the Traveling Salesman Problem”. In: *Journal of the ACM* 9.1 (1962), pp. 61–63. DOI: 10.1145/321105.321111.

[59] Domenico D’Alessandro. *Introduction to Quantum Control and Dynamics*. Boca Raton, FL: CRC Press, 2007.