

OQMD: Single-Qubit Rotation Control Improves Low-CNOT Multiclass Quantum Classification

Michael A Magid

mmagid1@binghamton.edu

Binghamton University

Melissa Z Ertem

Binghamton University

Jun Suzuki

University of Electro-Communications

Research Article

Keywords:

Posted Date: October 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10296672/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

OQMD: Single-Qubit Rotation Control Improves Low-CNOT Multiclass Quantum Classification

Michael A. Magid^{1,*}, Melissa Zeynep Ertem¹, Jun Suzuki^{2,3}

¹School of Systems Science and Industrial Engineering, Thomas J. Watson College of Engineering and Applied Science, Binghamton University, State University of New York, Binghamton, NY 13902-6000, USA

²Graduate School of Informatics and Engineering ³Institute for Advanced Science, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan

*Corresponding author: mmagid1@binghamton.edu

Abstract

Near-term variational classifiers incur substantial error and latency from two-qubit gates, yet practitioners often assume that additional entangling depth is the default route to higher accuracy. This work studies Optimal Quantum Measurement Decoding (OQMD): optimizing how quantum outcomes are mapped to classical labels by training a readout layer before measurement, jointly with the variational circuit, without adding CNOTs. Experiments use trainable triple single-qubit rotations as one concrete, hardware-native realization of OQMD; other single-qubit parametrizations fit the same classical outer loop. On the Iris benchmark with a 30-point stratified test split, the best observed 0-CNOT configuration with OQMD reaches 83.33% accuracy, with a 96% accuracy at 9 CNOTs, exceeding the best 18-CNOT controls (56.67%) and the best 18-CNOT configuration with OQMD (66.67%) under a common protocol. A six-point CNOT-depth series from 0 to 18 (fixed optimizer, iteration budget, random-seed count, and ZZ readout) shows that the highest raw scores need not occur at the largest template, so aggregate complexity is not summarized by CNOT count alone. Because run-level accuracies are discrete and non-Gaussian, we emphasize best-observed scores and, where a global comparison of pooled runs is required, Mann–Whitney U tests rather than parametric tests on means. Across architectures, OQMD shows statistically consistent but magnitude-dependent gains: large peak lifts on minimal circuits coexist with a small pooled mean shift on complex 18-CNOT runs ($p \approx 0.03$) that is not “universal” in the sense of uniformly large practical effects.

Keywords: quantum machine learning, variational quantum classifiers, measurement optimization, NISQ-era constraints, gate efficiency, circuit architecture

1 Introduction

Variational quantum classifiers (VQCs) combine parameterized quantum circuits with a classical optimizer for hybrid training Cerezo et al. (2021a), Bharti et al. (2022). They use

superposition and entanglement and may offer advantages on some learning problems relative to classical baselines Biamonte et al. (2017), Dunjko and Briegel (2018), but realized performance depends strongly on circuit design and hyperparameters.

Entanglement is widely treated as a central resource in quantum machine learning; empirical work shows settings where entangled representations can reduce prediction error and improve performance Wang et al. (2024). In variational circuits, entanglement is typically introduced through multi-qubit gates such as CNOTs, which motivates the common design practice of increasing entangling depth to improve accuracy.

A fundamental challenge facing quantum machine learning in the Noisy Intermediate-Scale Quantum (NISQ) era is the severe limitation imposed by two-qubit entangling gates, particularly CNOT gates. Current quantum hardware exhibits CNOT error rates 10-100 times higher than single-qubit gates, with execution times 10-50 times longer, leading to increased decoherence and error accumulation Preskill (2018). These constraints create a critical tension: while circuit expressivity typically requires entanglement through CNOT gates, practical deployability demands minimal CNOT usage to reduce errors. This motivates the development of techniques that improve classification accuracy without adding CNOT gates.

We introduce Optimal Quantum Measurement Decoding (OQMD), a measurement-decoding framework: the classical optimizer learns parameters of a readout map from measured qubit outcomes to class labels, jointly with the variational circuit, so that the mapping from quantum data to classical predictions is trained rather than fixed by default parity rules alone. Here we instantiate that readout with trainable single-qubit rotations immediately before measurement (adding no CNOTs); this is not the only possible realization—for example, standard $SU(2)$ product gates could be substituted, and full measurement optimization may ultimately involve collective measurements (left to future work). The size of the improvement from OQMD still depends on architecture (pooled shifts are statistically consistent under Mann–Whitney U , but mean gaps need not be large; Section 4).

The performance of variational quantum classifiers is highly sensitive to architectural choices, including feature map selection, ansatz design, rotation gate strategies, and optimization algorithms Havlíček et al. (2019), Schuld et al. (2020). Among these design choices, rotation gate sequences influence circuit expressivity and trainability by determining accessible regions of Hilbert space and the resulting optimization landscape Du et al. (2020), Holmes et al. (2022), Sim et al. (2019). Much of the literature still relies on heuristic templates or small empirical comparisons rather than full factorials over readout and ansatz choices LaRose et al. (2019), Cerezo et al. (2021b).

In this paper, all other design choices (feature map, ansatz, optimizer, and classical outer loop) are fixed to explicit baselines so that reported gains isolate the readout layer; we use the triple-rotation realization as a concrete, hardware-friendly implementation of OQMD rather than as the definition of OQMD itself. The Iris dataset provides a small, controlled setting in which differences in peak hold-out accuracy primarily reflect circuit layout, optimization, and measurement decoding rather than dataset shift or large-scale data effects. Our results challenge the informal expectation that “more CNOTs always help”: the largest raw scores arise on low- and mid-depth circuit templates when OQMD is enabled, while pooled CONTROL versus OQMD comparisons remain favorable under Mann–Whitney U on run-level accuracies (Section 4).

The paper makes four main contributions: (i) peak-focused benchmarks across minimal (0 CNOT), intermediate (structured mid-depth entanglement), and complex (18 CNOT) families; (ii) a six-point CNOT-depth series in which non-readout hyperparameters are held fixed (COBYLA, 200 iterations, 50 random seeds per condition, ZXZ OQMD) so that accuracy can be compared to hardware cost; (iii) factorial sweeps over readout parameterizations (triple rotations) on minimal and complex designs, showing that sensitivity to that choice depends on architecture; and (iv) analysis and plots appropriate for non-Gaussian run distributions, with modes noted in Section 4 rather than Gaussian means in headlines.

Classical reference accuracies for Iris are summarized in Appendix C.1 when interpreting the magnitude of the quantum scores.

2 Related work

Variational quantum classifiers (VQCs) employ hybrid quantum-classical optimization of parameterized circuits Cerezo et al. (2021a), Schuld et al. (2020). While early demonstrations showed promise Havlíček et al. (2019), systematic architectural studies remain limited. The expressivity-trainability trade-off is fundamental: highly expressive circuits risk barren plateaus McClean et al. (2018), while simple circuits may underfit Du et al. (2020), Sim et al. (2019). Systematic categorization of QML methods across NISQ and fault-tolerant regimes remains an active area Majid et al. (2025). Measurement optimization can improve accuracy at fixed CNOT budget without increasing two-qubit depth.

NISQ-era constraints necessitate gate-efficient designs. Hardware-efficient ansätze balance expressivity with native gate sets Kandala et al. (2017), while compilation techniques reduce CNOT count Nam et al. (2018). Recent VQC applications demonstrate that feature map and ansatz choices have strong practical impact on classification accuracy Regadío (2025). However, these approaches focus on state preparation. OQMD complements them by optimizing the measurement basis itself—an underexplored degree of freedom related to decoding and readout design Mahler et al. (2013), Higgins et al. (2007). A complementary direction improves accuracy by aggregating multiple VQCs through learned weighting rather than modifying the readout layer of a single circuit Tolotti et al. (2025). Our experiments focus on empirical optimization performance rather than information-geometric characterizations of the parameter space.

3 Methodology

3.1 Dataset and preprocessing

We use the Iris dataset from scikit-learn Pedregosa et al. (2011): 150 samples of three species (*I. setosa*, *I. versicolor*, and *I. virginica*), each with four features (sepal length and width, petal length and width, in centimeters). A single stratified split yields 120 training and 30 test samples (80%/20%). Fixing this split and preprocessing isolates changes in the quantum pipeline from data curation effects.

Features are min-max scaled to $[0, 1]$ for angle encoding. All four features are retained (no dimensionality reduction).

3.2 VQC architectures

We use three representative VQC designs ordered by structural complexity rather than by CNOT count alone: a minimal design (no CNOT gates), an intermediate design (four configurations at 3, 6, 9, and 12 CNOT), and a complex design (CNOTs for both feature map and ansatz).

3.2.1 Architecture 1: Minimal circuits (0 CNOT)

This design evaluates OQMD under severe resource constraints, as in shallow or connectivity-limited devices. The feature map is chosen as one of the simplest ones,

$$U_{\text{ZFM}}(\mathbf{x}) = \prod_{i=1}^4 R_Z(x_i) \cdot H^{\otimes 4}$$

This is called the ZFeatureMap (ZFM), which is equivalent to the Qiskit phase gate P up to the convention $P(\theta) = e^{i\theta/2} R_Z(\theta)$. The ansatz is set to a product of single rotations around y -axis (SingleLayerRY, SLRY). To simplify the notation, we shall identify this VQC by ‘ZFM(1)/SLRY(1)’. It should be emphasized that this VQC does not utilize *any* entanglement.

3.2.2 Architecture 2: Intermediate circuits (structured mid-depth)

Architecture 2 is not a single circuit but a *family* of four ladder templates between the 0-CNOT minimal design and the 18-CNOT complex design (Table 5). Three rungs (3, 6, and 9 CNOT) keep the same ZFeatureMap encoding, $U_{\text{ZFM}}(\mathbf{x}) = \prod_i R_Z(x_i) H^{\otimes 4}$, and increase only RealAmplitudes depth: ZFM(1)/RA(1), ZFM(1)/RA(2), and ZFM(1)/RA(3). The last rung implements the ZZ feature map (ZZFM),

$$U_{\text{ZZ}}(\mathbf{x}) = \prod_{k=1}^2 \left[\prod_{i=1}^4 R_Z(x_i) \cdot \prod_{\langle i,j \rangle} \text{CZ}_{i,j} \cdot R_{\text{ZZ}}(x_i x_j) \right] H^{\otimes 4}$$

which entangles two qubits conditioned on input data. This 12-CNOT rung switches to ZZFeatureMap with two repetitions and a SingleLayerRY ansatz, ZZFM(2)/SLRY(1); this rung changes *both* the feature map and the ansatz relative to the 3–9 CNOT templates, so it is not a pure depth extension of the same encoding–ansatz pair.

- ZFM(1)/RA(1) [3 CNOT]
- ZFM(1)/RA(2) [6 CNOT]
- ZFM(1)/RA(3) [9 CNOT]
- ZZFM(2)/SLRY(1) [12 CNOT]

Trainable ansatz parameters on four qubits with linear entanglement are 8, 12, and 16 for RA(1–3), and four for SLRY at the 12-CNOT rung (feature-map parameters remain data-dependent).

3.2.3 Architecture 3: Complex circuits (18 CNOT)

This design evaluates OQMD on a heavily entangled template typical of variational classifiers in the literature. To have an efficient encoding, the ZZFFMap is repeated twice. As an ansatz, the RA is also repeated twice. The total number of CNOT gates is 18 in total.

Table 1 summarizes the three architectural families (feature map, ansatz, CNOT budget, trainable parameters, and factorial scope). The four intermediate ladder templates (3, 6, 9, and 12 CNOT) are defined in Section 3 and drawn gate-by-gate in Appendix B.3.

Figure 1 shows the complete circuit diagrams for architectures 1 and 3, illustrating the gate-level implementation and CNOT placement.

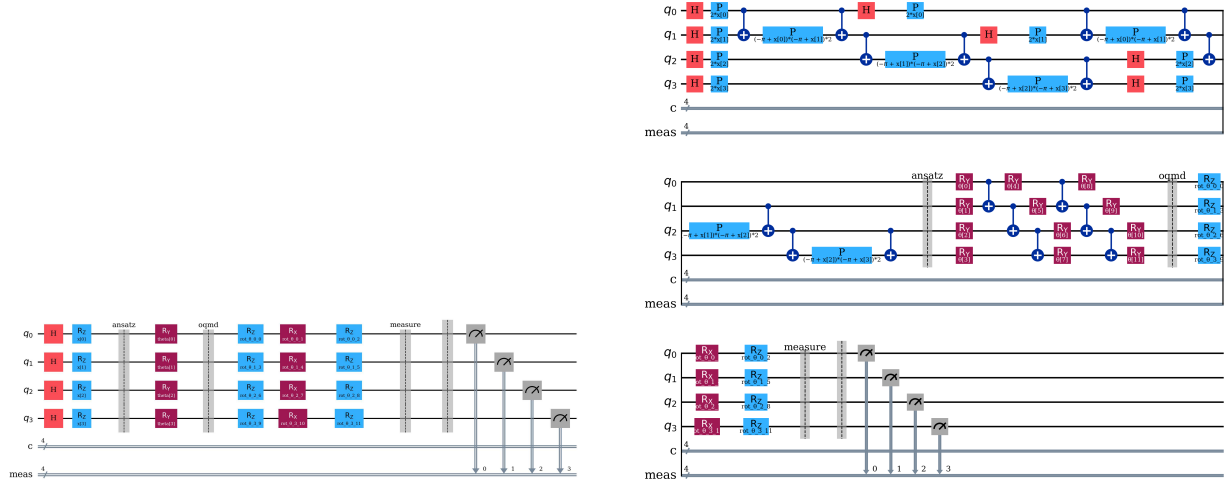


Figure 1: Circuit diagrams for Architecture 1 (minimal): $H + R_Z$ feature encoding, single R_Y ansatz layer, ZZZ OQMD rotations, 0 CNOT gates (left). Architecture 3 (complex): ZZFeatureMap (12 CNOT) + RealAmplitudes (6 CNOT) + OQMD (0 CNOT) = 18 CNOT total (right). OQMD adds measurement optimization with zero CNOT overhead in both cases.

3.3 OQMD: Optimal Quantum Measurement Decoding

3.3.1 Theoretical foundation

Standard VQC measurement projects the quantum state onto the computational basis $\{|0\rangle, |1\rangle\}^{\otimes n}$. OQMD generalizes this decoding scheme by choosing measurement axes for each qubit. This is easily implemented by first applying parameterized rotation gates followed by the standard Z measurements. Let $U_{\text{OQMD}}(\phi)$ be a product of rotation gates for each qubit with trainable angles ϕ . The total VQC with OQMD is given as

$$|\psi_{\text{measured}}\rangle = U_{\text{OQMD}}(\phi) \cdot U_{\text{ansatz}}(\theta) \cdot U_{\text{feature}}(\mathbf{x})|0\rangle^{\otimes n}$$

where $U_{\text{feature}}(\mathbf{x})$ encodes input data and $U_{\text{ansatz}}(\theta)$ is the variational ansatz with trainable parameters θ .

The core of OQMD is that the measurement angles ϕ can be treated as trainable parameters. Hence, the complete parameter vector $\Theta = (\theta, \phi)$ is jointly optimized to minimize the classification loss. Pre-measurement rotation strategies of this kind represent an emerging design pattern in quantum classification Sahoo et al. (2026).

Figure 2 illustrates the general OQMD concept: standard VQC measures in the computational basis, while OQMD first rotates into an optimized measurement basis, effectively learning both state preparation and optimal observables. Figure 3 contrasts CONTROL vs. OQMD on Architecture 1 (minimal) as a concrete example.

(a) Standard VQC (CONTROL):

$|0\rangle$ – Feature Map – Ansatz – M

(b) VQC with OQMD:

$|0\rangle$ – Feature Map – Ansatz – OQMD Rotations – M

OQMD adds 3 trainable rotation gates per qubit before measurement, optimizing the measurement basis with zero CNOT overhead.

Figure 2: OQMD concept: Standard VQC (a) measures in the computational basis, while OQMD (b) first applies trainable rotation gates to measure in an optimized basis. This adds expressivity without CNOT gates.

The advantage of the proposed OQMD is immediate: For an n -qubit system, OQMD adds $3n$ rotation gates and $3n$ trainable parameters. All of these gates are single-qubit rotations; no CNOT gates are added.

3.3.2 Rotation code specification

Upon implementing OQMD, a specific qubit rotation needs to be specified, since there is no unique way to implement it. Below, we consider several rotation implementations, which are called rotation codes. The purpose of analyzing rotation codes is to make sure whether or not OQMD depends on a specific choice of rotation codes. As main rotation codes, we evaluate 5 rotation strategies (4 OQMD codes + 1 control baseline), each defined by a three-character code specifying the sequence of Pauli rotations applied to each qubit before measurement:

- **ZZZ**: $R_Z(\phi_1) \cdot R_X(\phi_2) \cdot R_Z(\phi_3)$ — Standard triple single-qubit gate
- **ZXY**: $R_Z(\phi_1) \cdot R_X(\phi_2) \cdot R_Y(\phi_3)$ — Three-axis rotation
- **YXY**: $R_Y(\phi_1) \cdot R_X(\phi_2) \cdot R_Y(\phi_3)$ — Y-axis dominant
- **YXZ**: $R_Y(\phi_1) \cdot R_X(\phi_2) \cdot R_Z(\phi_3)$ — Mixed axes
- **CONTROL**: No rotation gates (baseline)

In addition, eight additional rotation codes (beyond the four used on the minimal design) will be analyzed in order to gain more insight into the effect of OQMD on the complex

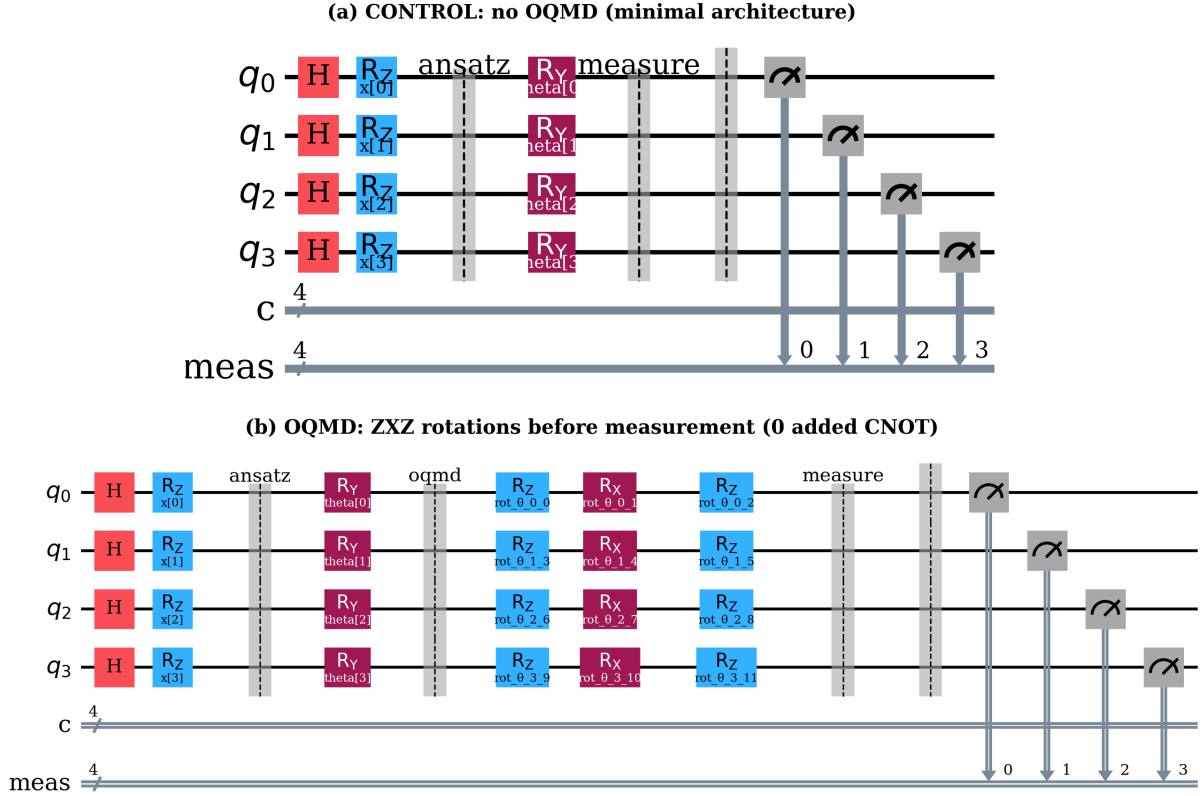


Figure 3: Detailed circuit comparison: CONTROL versus OQMD on the minimal architecture. (a) CONTROL (parity readout, no OQMD). (b) OQMD adds ZXZ triple rotations before measurement (no additional CNOTs). Peak accuracies appear in Table 2.

architecture (Architecture 3). They are specified by three axes as $ZXZ \leftrightarrow R_Z(\phi_1) \cdot R_X(\phi_2) \cdot R_Z(\phi_3)$:

$$XYX, XYZ, XZX, XZY, YZX, YZY, ZYX, ZYZ$$

Thus, 18 rotation codes will be examined in total.

3.4 Quantum classifier configuration

The VQC is configured for three-class classification with a statevector simulator (noise-free evaluation), cross-entropy loss, three output classes, and Qiskit’s parity encoding from measurement outcomes to class probabilities.

The setup yields valid nonnegative class probabilities that sum to one for the three Iris species (multiclass, not one-versus-rest reduction).

3.5 Optimization and hyperparameter search

3.5.1 Optimizers

We evaluated four optimization algorithms: one gradient-free and three gradient-based methods using finite-difference gradient estimation, selected for efficiency and convergence properties:

1. **COBYLA** (Constrained Optimization BY Linear Approximations): Gradient-free trust-region method for nonlinear optimization Powell (1994).
2. **L-BFGS-B** (Limited-memory Broyden-Fletcher-Goldfarb-Shanno): Gradient-based quasi-Newton method approximating the Hessian for improved convergence Byrd et al. (1995).
3. **CG** (Conjugate Gradient): Gradient-based iterative method exploiting conjugate directions Hestenes and Stiefel (1952).
4. **SLSQP** (Sequential Least Squares Programming): Gradient-based method for constrained optimization Kraft (1988).

For gradient-based methods, parameter derivatives use Qiskit’s built-in circuit-gradient interface (parameter-shift style evaluations for the models here), rather than optimizing the loss as a black box.

Each optimizer was limited to a maximum of 200-205 iterations to balance computational cost and convergence quality. We excluded slower optimizers (SPSA, POWELL, TNC) that exhibited iteration limit violations during preliminary experiments, increasing computational cost by up to $4.5\times$ without commensurate performance gains, and ADAM on grounds of per-iteration wall time on the minimal design (Appendix A.1.5, Table 8).

We include several classical optimizers in the factorial grids because VQC loss landscapes are not guaranteed to be well-conditioned: different algorithms emphasize gradient noise tolerance (COBYLA), quasi-Newton curvature (L-BFGS-B), cheap conjugate-direction steps

Table 1: Architectural families, templates, and run counts (same Iris split). Intermediate column lists the four ladder templates at 3/6/9/12 CNOT (Section 3); factorial rows are rotation $\times$ optimizer grids except intermediate, where only COBYLA+ZXZ ladder runs were executed. Abbreviations: ZFM = ZFeatureMap; ZZFM = ZZFeatureMap; SLRY = Single-LayerRY; RA(r) = RealAmplitudes with r repetitions (linear entanglement).

Parameter	Minimal	Interm.	Complex
Feature map	ZFM(1)	ZFM(1); ZZFM(2) at 12 CNOT	ZZFM (reps=2)
Ansatz	SLRY(1)	RA(1–3); SLRY at 12 CNOT	RA (reps=2)
CNOTs	0	3, 6, 9, 12	18
Train. params	4	8, 12, 16, 4 [†]	16
Rotation factorial	4 OQMD codes + CTRL	1 OQMD code + CTRL	12 OQMD codes + CTRL
Optimizers (fact.)	4	COBYLA (ladder)	4
Seeds per combo	50	50	50
Factorial runs	1,000	400 (ladder)	2,600
Ladder (50 $\times$ 2)	0 CNOT	3–12 CNOT	18 CNOT

[†]Four ansatz parameters on the 12-CNOT rung (ZZFM(2)/SLRY); feature-map parameters are data-dependent.

(CG), or small-scale constrained structure (SLSQP). Reporting the same readout and ansatz across these choices documents whether OQMD gains depend on a single optimizer rather than on measurement decoding itself.

3.5.2 Experimental design

Table 1 summarizes the three architectural families and how they connect to the factorial experiments vs. the standardized CNOT ladder (Fig. 7).

Fifty random seeds per combination provide enough restarts to observe rare high-accuracy initializations on the 30-point test set while keeping the complex factorial tractable. The iteration cap of 200 reflects a compromise between convergence and cost: pilot runs showed diminishing returns beyond roughly 150–200 COBYLA steps on these designs, and a common cap keeps learning dynamics comparable across conditions. For the ladder, we use COBYLA because gradient-free optimization tolerates noisy objectives and uneven landscapes; after pilots it also gave stable behavior on the intermediate-depth templates, and holding the optimizer fixed isolates CNOT-related architecture from optimizer variation (the factorial studies elsewhere still compare four optimizers on minimal and complex designs). On the ladder, OQMD is fixed to ZXZ at every CNOT level, so we do not claim a rotation factorial at intermediate depth in this paper. The combined factorial pool comprises 3,600 runs (minimal and complex) for seed-sensitivity analyses; ladder statistics are reported separately.

3.6 Training and evaluation protocol

3.6.1 Callback mechanism

To track learning dynamics, we use a callback that records training progress at regular intervals:

- Current parameter values $\Theta^{(t)}$ at iteration t
- Training loss $\mathcal{L}(\Theta^{(t)})$
- Hold-out test accuracy $\text{Acc}_{\text{test}}(\Theta^{(t)})$ on a coarse schedule (not every optimizer step; see below)

The classical optimizer updates parameters using only the training objective (cross-entropy on the 120-sample training split). Callback evaluations of test accuracy are not supplied to the optimizer: they are for monitoring and visualization only. The hold-out set is not used for early stopping, hyperparameter tuning, or gradient computation.

Evaluating full hold-out accuracy requires a statevector forward pass over all 30 test points at the current parameters, so recording it at every optimizer step would dominate wall-clock time on the factorial workloads (each extra evaluation repeats the full circuit evaluation for every test sample). Our experiments therefore log hold-out accuracy on a fixed stride of every 50 optimizer iterations (plus the initial and final iterates where available), not every step—a pragmatic trade-off between curve resolution and total compute. With a 200-iteration budget, that yields at most on the order of five interior diagnostic points per run in addition to endpoints; finer resolution would require shrinking the stride, accepting longer runs, or plotting training loss every step instead. Learning curves in the appendix reflect this coarse schedule.

Callbacks use Qiskit’s VQC training hooks so optimizers stay unchanged while test accuracy is logged on the fixed schedule.

3.6.2 Evaluation metrics

Primary performance metrics are as follows:

- Test accuracy: fraction of correctly classified samples in the held-out test set
- Learning improvement: $\Delta = \text{Acc}_{\text{max}} - \text{Acc}_{\text{start}}$
- Multiclass success rate: proportion of runs predicting all three classes
- Overfitting flag: $\text{Acc}_{\text{max}} > \text{Acc}_{\text{final}} + 5\%$

3.6.3 Normality check and reported summaries

Run-level test accuracies on Iris (30-point split, multiples of $1/30$) are inherently discrete. On the minimal-circuit factorial pools, Shapiro–Wilk tests reject Gaussianity for both CONTROL and OQMD marginals ($p \ll 10^{-6}$). Main text figures that show mean accuracy across

seeds plot 95% confidence intervals (CI) for the mean of run-level test accuracies (Student t intervals using the sample standard deviation (SD) and run count per condition), and Welch-style intervals for the difference of two independent means on mean-gain panels, not raw SD bars. Where a line summarizes mean OQMD gain versus CNOT count in a panel, it is an ordinary least squares (OLS) fit for visualization only. Tabulated $\pm$ values in bullet lists remain the sample SD for context. Where we report return on investment (ROI) in gates or parameters, it is peak accuracy gain in percentage points per added gate or parameter. Headline claims still rest on peaks, medians, and Mann–Whitney U tests for global CONTROL versus OQMD contrasts; Appendix A.2 states the null hypothesis and role of these tests.

4 Results

4.1 Overview: factorial results and CNOT ladder

Figure 4 contrasts the headline minimal vs. complex factorials (rotation/optimizer sweeps) with the six-point ladder.

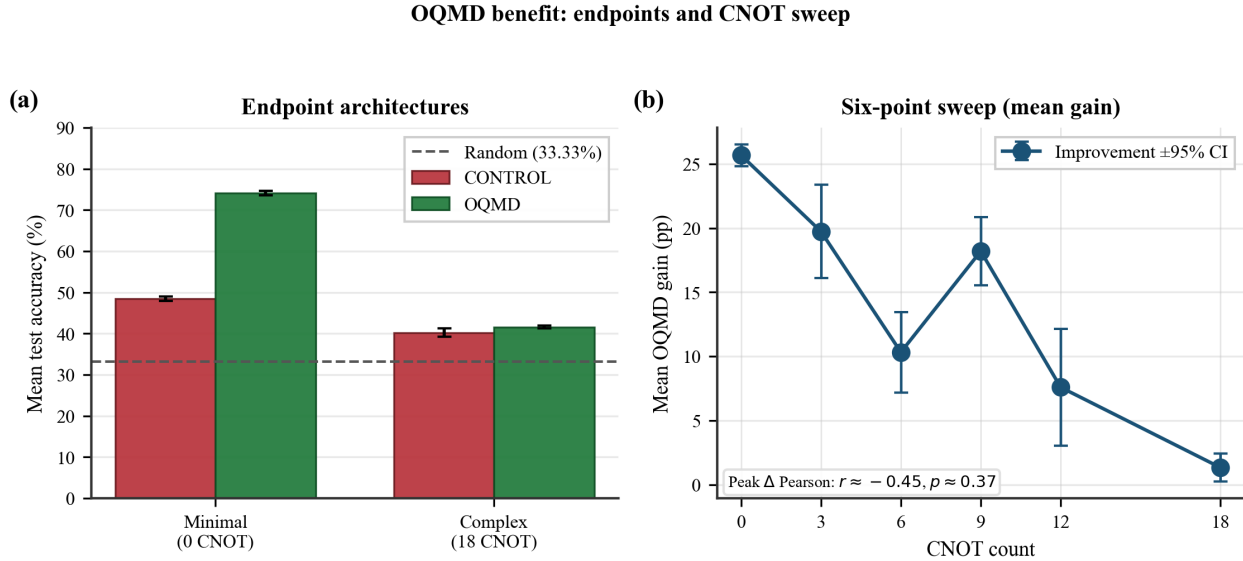


Figure 4: Summary of factorial means (left: 0 vs. 18 CNOT) and six-point mean OQMD gains with 95% CIs (right). Panel (b) annotates Pearson r on *peak* max-to-max margins ($r \approx -0.45$, Table 5); an ordinary least squares fit to *mean* gains would yield a different r (specifically $r \approx -0.91$) and is not used as a headline statistic. Peaks and Mann–Whitney tests are in Table 2. Gray dashed line: random baseline (33.33%).

Even when restricting to best seeds, minimal *control* circuits (peak 60.00%) can outperform complex *control* peaks (56.67%), echoing pooled factorial means from earlier analyses. This cautions against assuming monotonic accuracy in bare CNOT count once multiclass parity readouts are fixed. We avoid calling OQMD a “universal” improvement: effects are

Table 2: Peak test accuracy (%) on Iris for minimal (0 CNOT) and complex (18 CNOT) factorials. Mann–Whitney U on pooled runs (Section 3.6.3); Appendix A.2. Complex pooled means are only $\sim 40.2\%$ vs. $\sim 41.6\%$; headline peaks are in columns 3–5.

Circuit	CNOT	Peak Ctrl	Peak OQMD	Δ (pp)	M–W p
Minimal	0	60.00	83.33	+23.33	$< 10^{-100}$
Complex	18	56.67	66.67	+10.00	0.030

statistically consistent but magnitude-dependent—large in peak terms on minimal and complex factorials, but the complex pooled-*mean* gap is tiny even when $p < 0.05$ (Section 3.6.3).

4.2 Minimal circuits: strong peak gains at zero CNOT

4.2.1 Performance improvement (factorial)

On minimal circuits (0 CNOT gates), OQMD lifts the observed peak from 60.00% (CONTROL) to 83.33% (best ZXZ/L-BFGS-B runs in the factorial), while pooled run-level accuracies shift decisively toward OQMD under Mann–Whitney U ($p < 10^{-100}$). Discrete test scores (multiples of $1/30$) make the mode a useful secondary summary alongside the peak; Appendix B.6 retains histogram-style panels for the full minimal and complex pools.

Peak performance reached 83.33% (25 out of 30 test samples correct), i.e. a substantial lift over the best CONTROL seed (60.00%) without entangling gates. Across the pooled factorial runs, OQMD values stochastically dominate CONTROL under Mann–Whitney U ; individual seeds can still land below the opponent’s peak because accuracies are discrete and optimization is noisy. *Why multimodal:* hold-out accuracy lives on a discrete grid (multiples of $1/30$), and the pool mixes many rotation–optimizer–seed runs, so histograms show separated bumps (e.g. near chance vs. high-accuracy basins) rather than a single Gaussian. *Implication:* use peaks and Mann–Whitney tests for claims; do not treat the box plot as a unimodal surrogate.

Because the triple-rotation readout is not unique, we tested four ordered triples of Pauli axes (Section 4.4.3 lists the larger set used on the complex design). On the minimal architecture, those four choices behave similarly in the aggregate (Section 4.2), which supports treating OQMD as a stable readout layer rather than an ad hoc choice of a single gate pattern.

4.2.2 Rotation codes on minimal circuits

Systematic evaluation of rotation strategies on minimal circuits identified Z-axis rotations as superior (Figure 6 and Table 3). Values are rotation-wise summaries over 50 seeds each (exploratory; normality not assumed—Section 3.6.3).

All four OQMD rotations beat CONTROL across optimizers at $p < 0.001$ in the reported tests. The strongest means among the four codes tested here use Z on the first axis (ZXZ, ZXY, YXZ), consistent with readout that stays close to the computational Z basis before the final measurement.

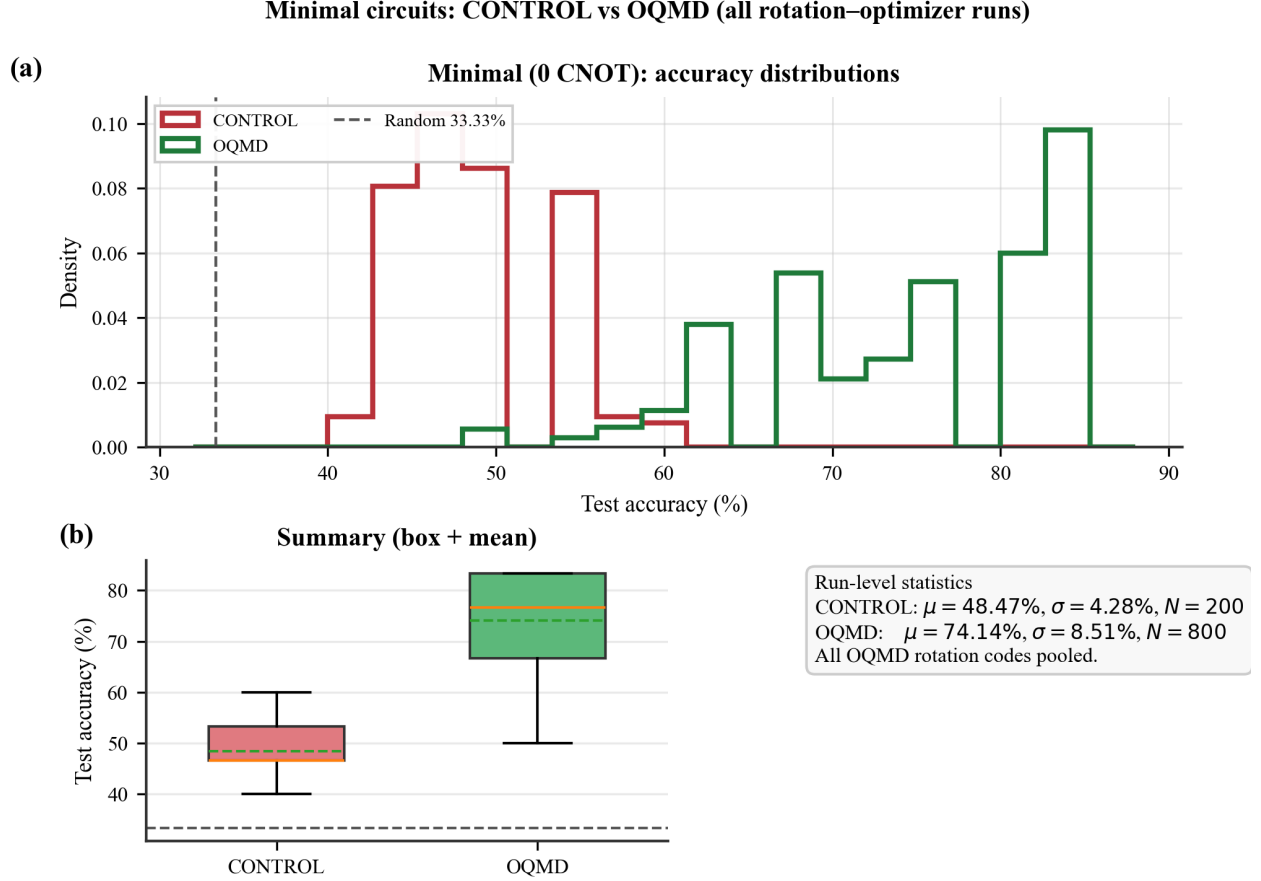


Figure 5: Minimal circuit performance (0 CNOT): pooled accuracy distributions for CONTROL ($N=200$ runs) and OQMD ($N=800$ runs). Top: normalized step histograms (binned empirical frequencies, not kernel density). Bottom: box-and-mean summary. OQMD pool is shifted decisively rightward of CONTROL; the multimodal shape reflects the discrete accuracy grid (multiples of $1/30$) combined with the mix of rotation codes, optimizers, and seeds. Peaks and Mann-Whitney results appear in Table 2; pooled run-level statistics are annotated in the bottom panel.

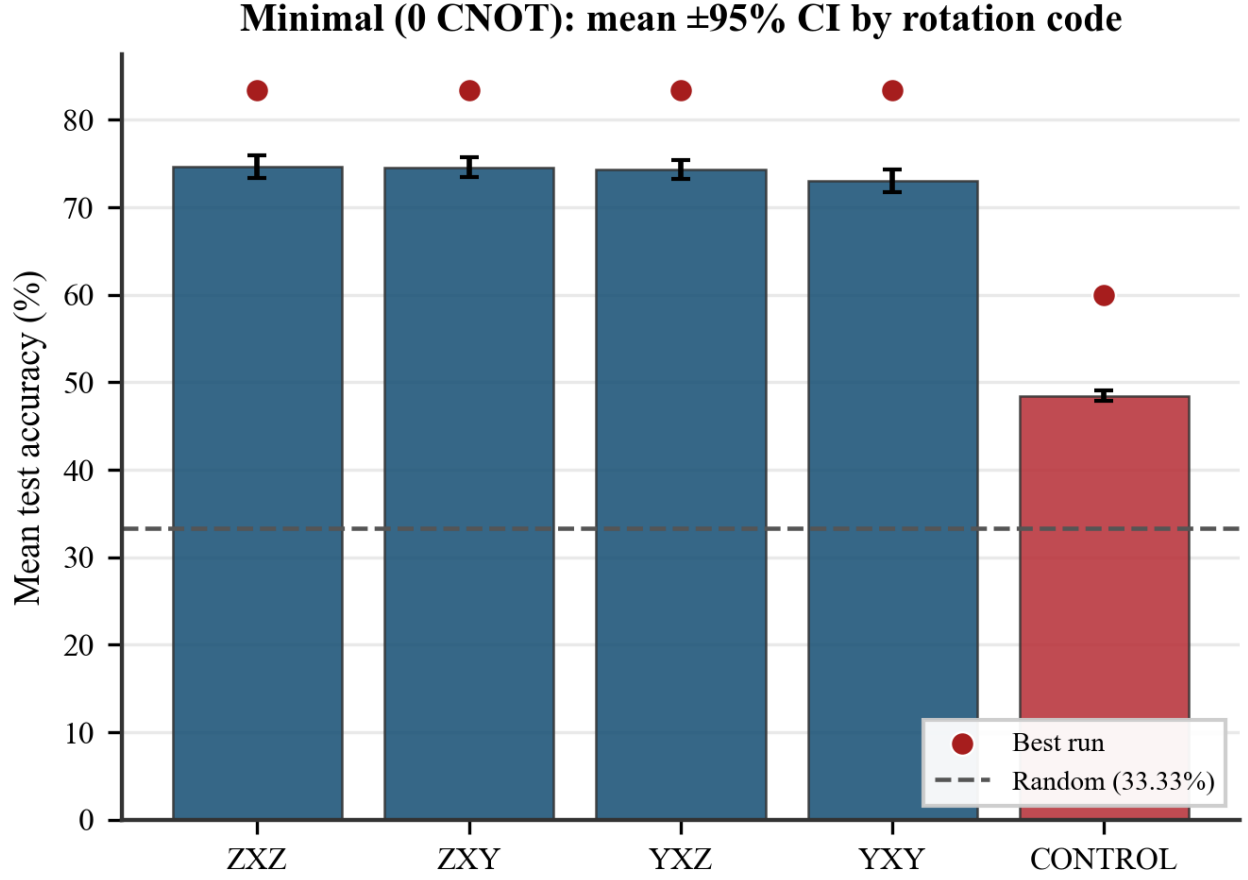


Figure 6: Rotation strategy performance on minimal circuits ($N=200$ runs per rotation code: four optimizers $\times$ 50 seeds). Bars: mean $\pm 95\%$ CI for the mean accuracy per code; markers show the best seed per code. Numeric summaries appear in Table 3.

Table 3: Rotation-code summary on minimal circuits (0 CNOT). Mean, sample SD, and peak hold-out accuracy (max) in percent; $N=200$ per code (four optimizers $\times$ 50 seeds).

Code	Mean (%)	SD (%)	Max (%)
ZXZ	74.63	9.25	83.33
ZXY	74.57	8.01	83.33
YXZ	74.33	7.52	83.33
YXY	73.02	9.08	83.33
CONTROL	48.47	4.28	60.00

Table 4: Gate Count Analysis: Minimal Circuits. Peak accuracies match Table 2. The Overhead row reports *absolute* peak gains in percentage points (pp), e.g. +23.33 pp means the peak accuracy increased by 23.33 points on the 0–100% scale, *not* a +23.33% relative change of the CONTROL peak. ROI is pp per added gate or parameter. Ranges such as “+10–23 pp” elsewhere in the paper refer to analogous peak-to-peak differences unless explicitly labeled as relative %.

Config	CNOT	1Q	Depth	Par.	Peak
Control	0	12	3	4	60.00%
OQMD	0	24	4	16	83.33%
Overhead	+0	+12	+1	+12	+23.33 pp
ROI	—	≈ 1.9 pp/gate	—	≈ 1.9 pp/param	—

4.2.3 Optimizer performance on minimal circuits

On minimal circuits, L-BFGS-B attained the highest mean accuracy ($70.35\% \pm 12.72\%$) among the four optimizers while sharing the same 83.33% peak as the others. The combination ZXZ + L-BFGS-B averaged $76.87\% \pm 7.69\%$ over 50 seeds (Appendix B.4).

4.2.4 Gate efficiency: zero CNOT overhead

A critical advantage of OQMD is its gate efficiency (Table 4). On minimal circuits, OQMD adds 12 single-qubit rotation gates but zero additional CNOT gates:

Relative to the best CONTROL run, each added single-qubit gate in the OQMD block buys roughly $23.33/12 \approx 1.9$ percentage points of peak accuracy on Iris. The zero CNOT overhead is critical for NISQ devices where:

- CNOT error rates: 10-100 $\times$ higher than single-qubit gates
- CNOT execution time: 10-50 $\times$ longer, increasing decoherence
- CNOT connectivity: Limited on physical hardware, requiring SWAP gates

At 0 CNOT, OQMD improves peak accuracy without adding two-qubit gates, which matters when CNOT error and connectivity dominate the budget.

Table 5: Six-point CNOT-depth series (Fig. 7). Peak and mean hold-out test accuracies (%) and peak max-to-max improvement (pp). Means pool 50 seeds per side; OQMD means use the ZZZ readout on the ladder (same runs as Fig. 7). Abbreviations: ZFM, ZZFM; SLRY; $RA(r)$ = RealAmplitudes with r repetitions (linear entanglement); e.g. $RA(1)$ is one ansatz repetition (three CNOT gates on four qubits).

C FM / ansatz	Peak		Mean		Peak Δ pp
	Ctrl	OQMD	Ctrl	OQMD	
0 ZFM(1) / SLRY	60.0	83.3	48.5	74.6	+23.3
3 ZFM(1) / $RA(1)$	80.0	93.3	64.0	83.7	+13.3
6 ZFM(1) / $RA(2)$	96.7	96.7	79.1	89.5	+0.0
9 ZFM(1) / $RA(3)$	80.0	96.7	68.9	87.1	+16.7
12 ZZFM(2) / SLRY	76.7	76.7	59.1	66.7	+0.0
18 ZZFM(2) / $RA(2)$	56.7	66.7	40.2	41.6	+10.0
Pearson r (CNOT vs. peak Δ)		-0.45, $p \approx 0.37$ (n.s.)			

4.2.5 Learning dynamics on minimal circuits

Analysis of learning curves for top-performing minimal circuit configurations reveals rapid convergence within 100-150 iterations with smooth monotonic improvement and no overfitting. The top configuration (ZZZ + L-BFGS-B, seed 42) achieved 83.33% accuracy with +33.3% improvement from initialization, demonstrating genuine optimization rather than fortunate initialization (learning curves in Appendix B.4).

4.3 Intermediate templates

The six-point ladder (Fig. 7, Table 5) places four intermediate templates at 3, 6, 9, and 12 CNOT (Architecture 2). All ladder points share the Iris split, COBYLA (200 iterations), fifty random seeds per CONTROL and ZZZ OQMD side, and fixed ZZZ readout on OQMD runs. The discussion below emphasizes the six-CNOT rung, ZFM(1)/ $RA(2)$, because it shows a peak tie between CONTROL and OQMD (96.67%) and illustrates how little headroom remains once entanglement is already strong on Iris; the 3-, 9-, and 12-CNOT rows in Table 5 show the contrasting ceiling and floor behavior across the same family (including the ZZFM(2)/SLRY confound at 12 CNOT).

We did not run a separate rotation $\times$ optimizer factorial on any intermediate template, so rotation sensitivity at 3–12 CNOT is not reported here—only ZZZ on the ladder.

Behaviorally, Table 5 shows identical peak hold-out accuracy for CONTROL and OQMD at six CNOTs (96.67% on the 30-point test), so the max-to-max improvement is zero despite nontrivial entanglement. Under this readout and budget, the intermediate template already reaches a very high Iris ceiling, leaving little headroom for the triple-rotation layer to raise the best seed; OQMD may still alter typical trajectories or pooled accuracies, but the headline comparison is a peak tie at this rung. Fig. 4 situates it between shallower and deeper ladder points. A dedicated rotation sweep at six CNOTs remains future work when resources allow. Table 5 documents all four intermediate rungs (3, 6, 9, and 12 CNOT); behavior at the remaining rungs mirrors the ceiling and floor dynamics described here.

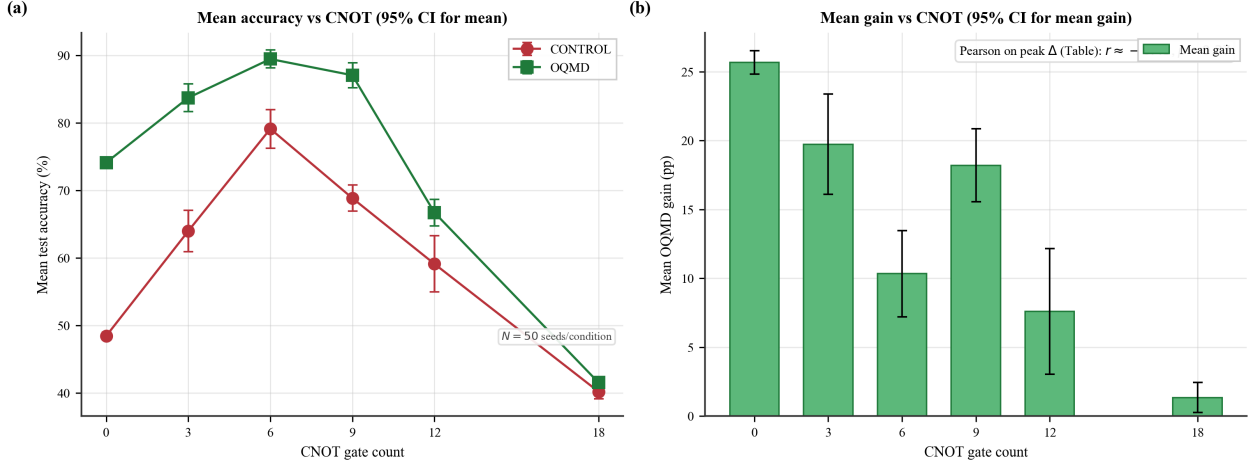


Figure 7: Mean test accuracy (left) and mean OQMD gain (right) versus CNOT count on the six templates of Table 5; 50 seeds per side; 95% CIs for the mean (Section 3.6.3). Templates differ between rungs (not a single depth-only family). Pearson on peak margins in Table 5. Baseline: gray dashed line at 33.33%.

4.4 Complex circuits: OQMD provides consistent improvement

4.4.1 Peaks on the 18-CNOT template

On complex circuits (18 CNOT gates), factorial peaks rise from 56.67% (CONTROL) to 66.67% (OQMD), a +10 percentage point swing on the 30-point Iris test, while the pooled run-level gap remains significant under Mann–Whitney U ($p \approx 0.03$). Pooled means (40.22% vs. 41.57%) may appear in figure annotations as descriptive context but are not the headline statistic.

The best OQMD run reached 66.67% (YXY + COBYLA) versus 56.67% for the best CONTROL: ten percentage points on the 0–100% scale, i.e. three more correct labels on the 30-sample test. Relative to the control peak, $66.67/56.67 \approx 1.18$, an $\approx 18\%$ relative increase in accuracy; the headline remains the absolute ten-point gain on this small test set.

4.4.2 Multiple rotation codes achieve optimal performance

Systematic evaluation of the twelve OQMD triples in Section 4.4.3 on complex circuits revealed that multiple codes achieve competitive performance (Figure 9 and Table 6). YXY attains the highest mean ($42.23\% \pm 8.12\%$); YXY, YZX, ZXY, and ZXZ each reach 66.67% peak hold-out accuracy (Table 6).

The finding that multiple rotation codes achieve identical peak performance (66.67%) suggests robustness to specific rotation sequences. The CONTROL condition performing comparably to some rotation codes in mean accuracy (about 40.2% vs. roughly 41–42% for several OQMD codes) indicates that the ZZFeatureMap + RealAmplitudes architecture already encodes substantial expressivity through its entanglement structure, though OQMD still provides higher peak performance.

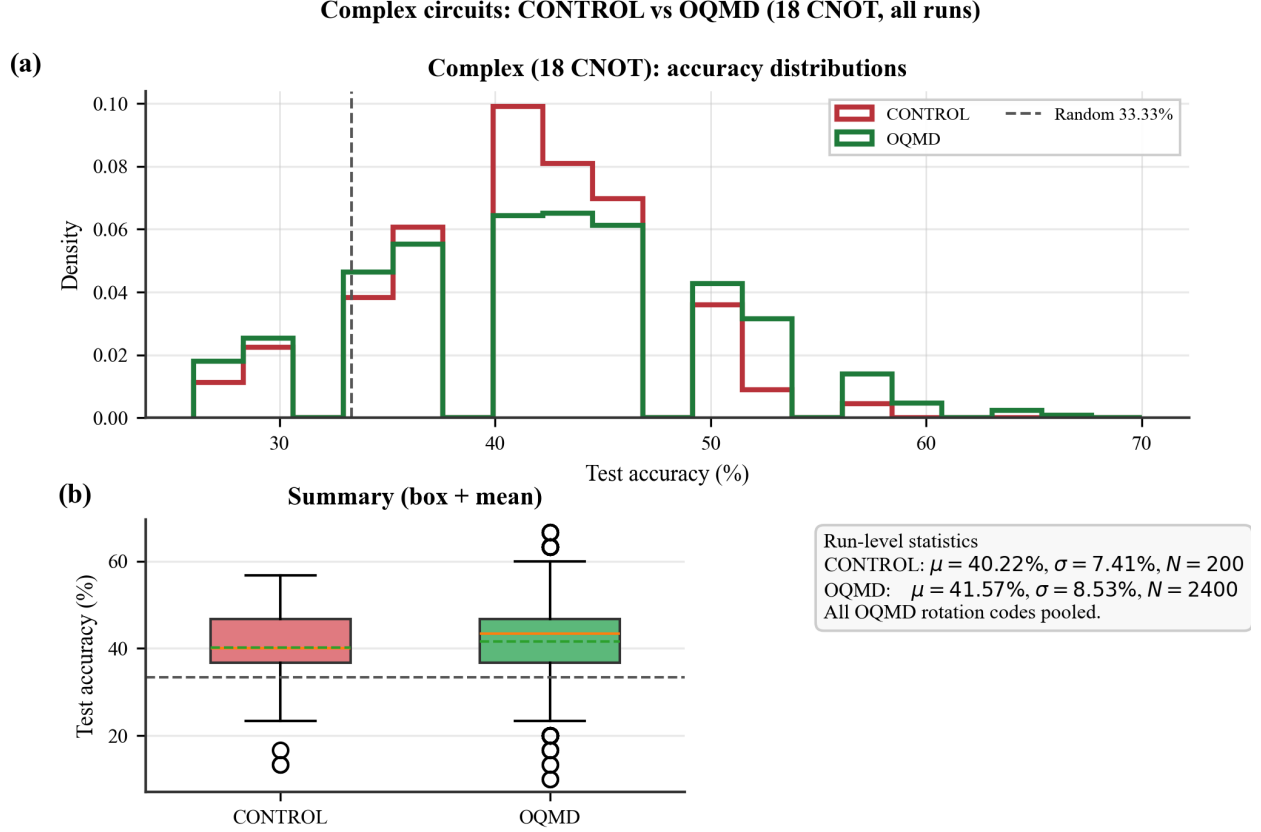


Figure 8: Complex circuit performance: peaks 56.67% (CONTROL) vs. 66.67% (OQMD); pooled accuracies differ under Mann–Whitney U ($p \approx 0.03$). Top: normalized step histograms (binned empirical frequencies, not kernel density). Mean annotations are descriptive; peaks drive the headline claims.

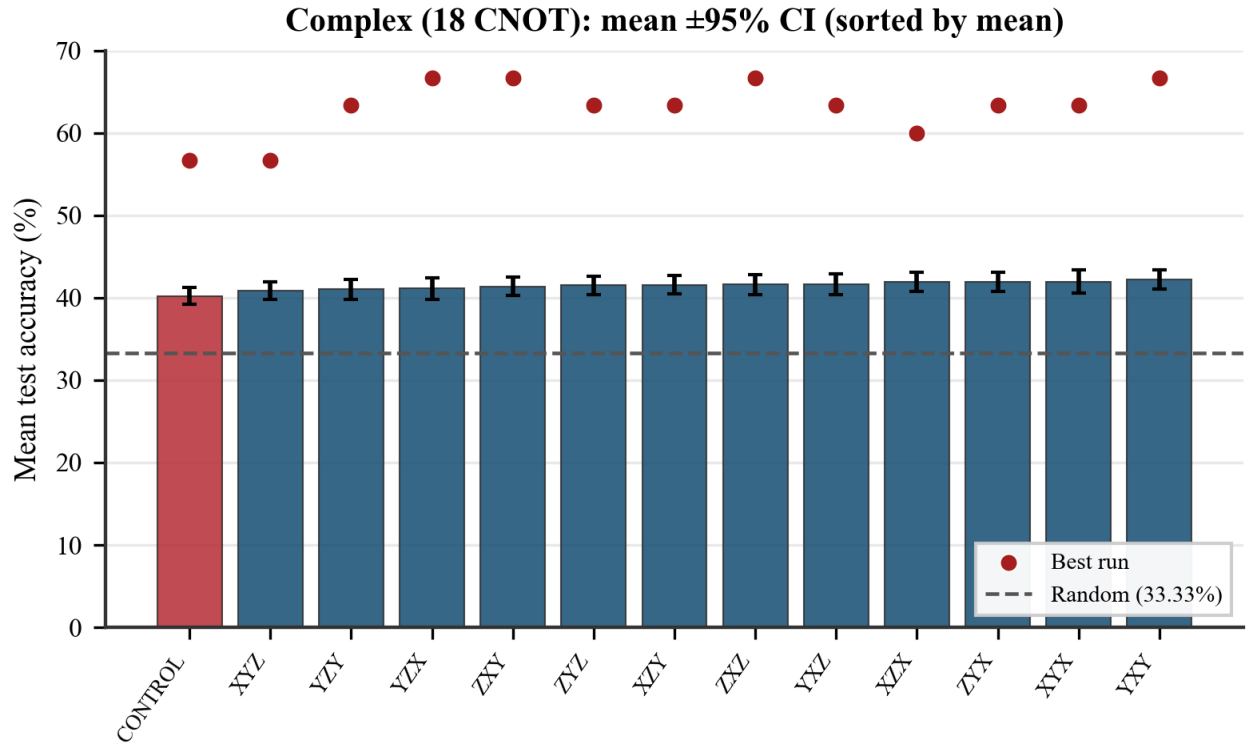


Figure 9: Mean test accuracy by rotation code on complex circuits ($N=200$ runs per code). Bars: mean $\pm 95\%$ CI; markers show the best seed per code. Codes are ordered by increasing mean accuracy; full statistics appear in Table 6.

Table 6: Rotation-code summary on complex circuits (18 CNOT). Mean, sample SD, and peak hold-out accuracy (max) in percent; $N=200$ per code; rows sorted by mean (ascending).

Code	Mean (%)	SD (%)	Max (%)
CONTROL	40.22	7.41	56.67
XYZ	40.88	7.86	56.67
YZY	41.02	8.43	63.33
YZX	41.12	9.23	66.67
ZXY	41.37	8.01	66.67
ZYZ	41.50	8.30	63.33
XZY	41.58	8.16	63.33
ZXZ	41.63	8.68	66.67
YXZ	41.65	9.11	63.33
XZX	41.90	8.38	60.00
ZYX	41.93	8.07	63.33
XYX	41.97	9.94	63.33
YXY	42.23	8.12	66.67

4.4.3 Readout triples used in the complex factorial

Implementing OQMD with three trainable single-qubit rotations per qubit does not fix a unique gate sequence: many ordered triples $(R_\alpha, R_\beta, R_\gamma)$ with $\alpha, \beta, \gamma \in \{X, Y, Z\}$ are admissible. On Architecture 1 (minimal) we therefore ran a small factorial over four such triples plus CONTROL: they yield similar pooled performance (Section 4.2), which supports robustness of the readout layer for that template. On Architecture 3 (complex) we expanded to twelve OQMD triples plus CONTROL to test whether the same robustness holds when entanglement reshapes the state; Figure 9 shows that means and peaks then depend more strongly on the chosen triple.

4.4.4 COBYLA excels on complex circuits

On complex circuits, COBYLA empirically produced the largest share of runs with large within-run gains $\Delta = \max_t \text{Acc}_t - \text{Acc}_0$ exceeding a ten percentage-point reporting threshold chosen for scale on Iris, while L-BFGS-B achieved higher average accuracy in the pooled run records. We defer to Appendix B.5 for optimizer-wise histograms and leave rigorous effect-size comparisons to future work.

4.4.5 Learning quality and improvement analysis

Across 2,600 complex-circuit runs, many trajectories show negligible Δ under the 200-iteration budget while others improve sharply; the distribution is heavy-tailed, so summaries should emphasize quantiles or peaks rather than means (Appendix B.5). Overfitting flags occurred in $\approx 18\%$ of runs under the heuristics of Section 3.6.2.

4.4.6 Multiclass classification reliability

On the complex factorial, 2,596 of 2,600 runs (99.85%) assigned at least one test label to each of the three species, so the parity readout rarely collapsed to one or two classes only.

Table 7: Factorial peaks (Table 2) restated for quick comparison with ladder extrema. Ctrl and OQMD columns are peak test accuracies (%).

Family	CNOT	Ctrl (%)	OQMD (%)	Δ (pp)
Minimal	0	60.00	83.33	+23.33
Complex	18	56.67	66.67	+10.00

Four runs never predicted class 1 (*I. versicolor*) while still scoring 46.67–50.00%; we treat that as rare initialization noise rather than a specification bug.

4.5 Discussion

4.5.1 CNOT ladder vs. hardware cost (Fig. 7)

Figure 7 summarizes the standardized ladder: best-of-50-seed hold-out accuracy for CONTROL and for OQMD (ZXZ, COBYLA, 200 iterations) at each CNOT rung (Table 5). Every ladder point uses the same Iris train and test split, the same optimizer and iteration cap, and the same OQMD rotation code (ZXZ); the only deliberate change between rungs is the structured entangling template (CNOT budget). Reported values are maxima over 50 independent random seeds per side (CONTROL versus OQMD) and rung. Because improvements are computed as $\max_{\text{seeds}} \text{Acc}_{\text{OQMD}} - \max_{\text{seeds}} \text{Acc}_{\text{control}}$, several intermediate rungs show zero max-to-max gain even though pooled Mann–Whitney tests still favor OQMD on average. An exploratory Pearson correlation of CNOT count with these six max-to-max margins yields $r \approx -0.45$, $p \approx 0.37$ and does *not* meet conventional significance thresholds; we therefore report the *pointwise* per-rung peaks and hardware costs without claiming a validated monotone or inverse trend in CNOT count.

High accuracy is achievable without using the maximum-CNOT template; OQMD allows low- and mid-depth circuits to reach competitive peak quality while limiting additional CNOT layers.

4.5.2 Practical NISQ implications

Taken together, the factorial peaks and the ladder’s per-rung maxima support a practical design lesson for NISQ-era systems: readout optimization can raise peak accuracy without additional CNOTs, and competitive peaks appear at several entangling depths, not only at the maximum-CNOT template. We do not infer a statistically supported *global* inverse trend between CNOT count and OQMD benefit from the exploratory Pearson fit (preceding subsection).

1. On resource-constrained devices, limited connectivity, short coherence times, and high CNOT error rates motivate avoiding unnecessary two-qubit layers; OQMD adds only single-qubit rotations.
2. OQMD allows useful accuracy on zero-CNOT product-feature maps, reducing exposure to CNOT-dominated error channels relative to increasing depth alone.

3. Rather than assuming accuracy scales monotonically with bare CNOT count, one can pair a modest entangling core with learned measurement rotations.
4. Because OQMD uses single-qubit gates before measurement, it is compatible with diverse connectivity constraints.

4.5.3 High-scoring configurations

Among top complex-circuit settings (63.33–66.67% accuracy), all ten runs summarized in Appendix B assign labels to every class, with near-perfect separation of class 0 and mixed errors on classes 1–2. Reported within-run gains span about ten to thirty-five percentage points on the accuracy scale, across several rotation–optimizer pairs.

Minimal-design peaks reach 83.33% with larger within-run gains than on the 18-CNOT template despite fewer base parameters, which is consistent with readout helping most when the bare circuit is shallow. Confusion matrices, curves, and extra statistics are in Appendix B.

4.5.4 Hyperparameter interactions

Analysis of rotation-optimizer combinations reveals distinct patterns by architecture. For complex circuits, multiple combinations achieve $\geq 65\%$ accuracy, with no single dominant pairing—YXY and ZXY perform well across all optimizers (60–67% range), while even the worst combinations achieve 50–55%, substantially above random baseline.

For minimal circuits, every OQMD rotation–optimizer pair in the 4×4 factorial reaches at least 70% peak accuracy, whereas CONTROL peaks at 60%. Thus the presence of OQMD dominates fine-grained rotation tuning on the minimal design; complex designs remain far more rotation-sensitive (Appendix B).

4.5.5 Statistical robustness

Split robustness notes appear in Appendix B.1.

5 Conclusion and outlook

OQMD frames measurement as part of the learning problem: a classical optimizer trains how quantum measurement outcomes are decoded to class labels, here realized with trainable single-qubit rotations before readout so that no extra CNOTs are required. On factorial experiments, the best minimal design (83.33%) beats the best 18-CNOT controls and pushes the entangled baseline upward as well (Table 2). On the six-point CNOT-depth series (Table 5), peak max-to-max improvements vary with template in a descriptively non-monotone way, while mean accuracies can still favor OQMD when peaks tie; the exploratory linear correlation of CNOT count with peak margins is not significant. The practical take-away remains to pair modest entangling cores with learned decoding rather than assume that “more CNOTs always help.” Pooled run comparisons remain highly significant under

Mann–Whitney U (Appendix A.2), which is appropriate for non-Gaussian discrete accuracies. Classical Iris scores (Appendix C.1) remain above these quantum peaks, underscoring that we target resource accounting and decoding design inside a fixed Qiskit parity readout, not overall competitiveness with classical baselines.

Benefits. OQMD adds only single-qubit rotations before measurement, leaves the chosen CNOT count in the ansatz unchanged, and in our runs often raises peaks and pooled outcomes when readout is trained with the circuit. Together, the factorial and ladder experiments argue for treating readout as a tunable module alongside ansatz depth and optimizer, especially when two-qubit gates are costly or connectivity-limited. On Iris, means with 95% CIs for the mean are descriptive only; peaks and Mann–Whitney summaries matter more because accuracies are discrete.

Classical baselines and scope. Off-the-shelf classical models reach far higher Iris accuracy than our VQC peaks under the same split (Appendix C.1); OQMD is not positioned as beating classical ML on raw score, but as improving *readout-restricted* quantum parity classifiers under a fixed gate budget. Framing is therefore resource accounting and measurement design, not SOTA classification.

When OQMD may not help. OQMD is a readout layer: if the ansatz already saturates the parity readout or optimization is severely under-budget, gains can be negligible. On complex factorials, pooled mean shifts can be tiny even when peaks and Mann–Whitney favor OQMD—statistically consistent but magnitude-dependent improvement. Extra pre-measurement single-qubit depth may also be undesirable when calibration dominates or when device policy limits learned rotations before measurement.

We emphasize three limitations. First, bar and line plots that show mean accuracy with 95% CIs for the mean (or box/histogram summaries) are *descriptive*; headline claims rest on best-seed peaks and on Mann–Whitney tests over full run pools, because Iris accuracies are discrete and often non-Gaussian. Second, we do *not* claim a “universal” lift: effects are magnitude-dependent (large peak gaps on minimal designs vs. small pooled mean gaps on complex). Third, the six-CNOT ladder point is a single rotation code (ZXZ) and optimizer (COBYLA); factorial evidence at 0 and 18 CNOT supports broader robustness, but intermediate-depth rotation sweeps remain future work.

Outlook. The same peak-focused protocol should extend to additional datasets and noise models; finer-grained monitoring of hold-out accuracy during training would require proportionally higher evaluation cost or reliance on training loss.

Acknowledgements

We thank Yusaku Yamada for discussions at an earlier stage of this project. This work is supported by JSPS KAKENHI Grant Numbers JP24K14816 and ERATO “Super Quantum Entanglement” (Grant No. JPMJER2402) from JST. The Binghamton University (SUNY)–University of Electro-Communications exchange program supported a visiting residency at UEC during part of this work.

Appendix A Detailed methodology

This appendix expands experimental methodology, optimizer settings, and training protocol details referenced in the main text.

A.1 Optimizer Specifications

We evaluated four optimization algorithms in the main factorial grids. We also ran a small ADAM pilot on the simplest (0 CNOT, ZZ) design to document wall-clock relative to COBYLA, L-BFGS-B, SLSQP, and CG (Table 8).

A.1.1 COBYLA (Gradient-Free)

COBYLA (Constrained Optimization BY Linear Approximations) is a trust-region method that constructs linear approximations of the objective function without requiring gradient information Powell (1994). It is particularly robust for noisy objective functions common in quantum variational algorithms. Configuration: maximum 200 iterations, tolerance 10^{-6} .

A.1.2 L-BFGS-B (Gradient-Based)

L-BFGS-B (Limited-memory Broyden-Fletcher-Goldfarb-Shanno with Bounds) is a quasi-Newton method that approximates the Hessian matrix using limited memory Byrd et al. (1995). Gradients use Qiskit’s parameter-shift gradient pipeline for the circuit parameters. Configuration: maximum 200 iterations, tolerance 10^{-6} , gradient epsilon 10^{-8} .

A.1.3 CG (Gradient-Based)

The conjugate-gradient method Hestenes and Stiefel (1952) uses conjugate search directions in the classical loop. Circuit gradients use the same Qiskit interface as L-BFGS-B. Configuration: maximum 200 iterations, tolerance 10^{-6} .

A.1.4 SLSQP (Gradient-Based)

Sequential least squares programming Kraft (1988) handles small constrained problems in the classical loop. Circuit gradients use the same Qiskit interface as L-BFGS-B. Configuration: maximum 205 iterations (slightly higher to account for constraint checking), tolerance 10^{-6} .

A.1.5 ADAM pilot (minimal architecture; excluded from factorials)

We benchmarked gradient-based ADAM (learning rate 0.01, default hyperparameters otherwise) on the *minimal* ZZ + OQMD design with the same Iris split as Section 3, a single initialization (seed 42), and the same 200-iteration cap as the other optimizers in this pilot. ADAM is not iteration-cheap in this pipeline: measured runs at 10, 25, and 50 iterations are nearly linear in `maxiter` with ≈ 7.0 s per iteration on our workstation, versus ≈ 0.13 s per iteration for COBYLA on the same machine and seed—so a full 200-iteration ADAM fit extrapolates to ~ 1400 s wall time ($\sim 50\times$ COBYLA). On this pilot seed, 50 iterations

of ADAM reached 63.33% hold-out accuracy, comparable to L-BFGS-B at 200 iterations (63.33%) and below SLSQP/CG (66.67%) under the same seed; we did not observe a case for absorbing a $\sim 50\times$ per-run cost in the large factorial study.

Table 8: Pilot timings (wall clock, one seed) on the minimal ZXZ + OQMD design. COBYLA, L-BFGS-B, SLSQP, CG: 200 iterations. ADAM: extrapolated to 200 iterations as $(7.02\text{ s/it}) \times 200$ from measured 10/25/50-iteration runs (linear in `maxiter`); accuracy at 50 iterations shown for comparison. Test accuracies are *single-seed* pilot values; the factorial COBYLA mean on minimal circuits is 68.57% (Appendix B.4), not 43.3%.

Optimizer	Time	Test acc.	vs. COBYLA
COBYLA	27 s	43.3%	1.0 $\times$
L-BFGS-B	63 s	63.3%	2.3 $\times$
SLSQP	72 s	66.7%	2.7 $\times$
CG	127 s	66.7%	4.7 $\times$
ADAM	$\sim 1.4 \times 10^3$ s (extrap.)	63.3% (50 it.)	$\sim 50\times$

SPSA, POWELL, TNC, and ADAM were therefore excluded from the main factorial experiments. For SPSA, POWELL, and TNC, preliminary runs showed roughly $4.5\times$ longer runtimes tied to iteration-limit behavior without commensurate accuracy gains. ADAM is excluded for the distinct reason above: extreme per-iteration wall time in this pipeline on our hardware (Table 8), not because of the iteration-limit issues seen for SPSA, POWELL, and TNC.

A.2 Mann–Whitney U tests

For two independent samples of run-level test accuracies (CONTROL versus OQMD), we report two-sided Mann–Whitney U tests as implemented in standard scientific libraries. The null hypothesis is that both samples are drawn from the same distribution (same population), so there is no systematic shift toward higher accuracy under one condition. Rejecting this null supports a stochastic ordering or shift between pools without assuming Gaussianity, which matters here because Iris accuracies are discrete fractions with tied values. We use these tests for global factorial contrasts where a single p -value summarizing thousands of runs is useful; they complement peak-based headlines and do not replace seed-level diagnostics.

A.3 Training Protocol Details

A.3.1 Callback Mechanism Implementation

The callback records parameter values, training loss, and hold-out test accuracy on the same schedule as Section 3 (every 50 optimizer iterations in our implementation, not every step). Hold-out accuracy is diagnostic only and is not used in optimizer updates, as in Section 3.

A.3.2 Evaluation Metrics

The metrics listed in Section 3 (test accuracy, learning improvement Δ , multiclass success rate, and the overfitting flag) are computed identically here.

Appendix B Additional experimental results

This appendix collects extra figures and tables for optimizer comparisons, learning curves, confusion matrices, and pooled statistics.

B.1 Split robustness and seed counts

Using 50 independent random seeds per hyperparameter combination stabilizes summaries of discrete Iris accuracies. Minimal circuits show higher seed-to-seed spread ($\sigma \approx 9\text{--}13\%$) than complex circuits ($\sigma \approx 7\text{--}10\%$), consistent with fewer base parameters on the minimal template. In 77.5% of runs, accuracy exceeds the 33.33% random baseline with $p < 0.001$ under the same per-run screen summarized in Appendix B.6. Further variance summaries appear there as well.

Pilot runs excluded SPSSA, POWELL, TNC, and full-budget ADAM from the factorial grids after roughly $4.5\times$ longer runtimes without matching accuracy gains (Section 3.5 and Table 8).

B.2 Six-point CNOT sweep: circuit configurations (Fig. 7)

The six circuit configurations span 0–18 CNOT gates with Architecture 1 (minimal), Architecture 2 (intermediate at the six-CNOT rung), and Architecture 3 (complex) as in Section 3. Intermediate ladder points share the Iris split, COBYLA (200 iterations), 50 seeds, and ZXZ OQMD. Table 5 reports peak and mean hold-out accuracies (means pool 50 seeds per side; exploratory Pearson uses peak-to-peak margins). Rungs 3, 6, 9, and 12 are maxima over ladder runs; rungs 0 and 18 reuse factorial peaks from Table 2.

B.3 Explicit circuit definitions for the CNOT sweep

This subsection provides explicit definitions of the six circuits used in the CNOT sweep shown in Figure 7. All circuits use 4 qubits, the same train/test split, the same preprocessing, and ZXZ OQMD rotations when OQMD is applied.

Intermediate ladder templates (3, 6, 9, and 12 CNOT). Figure 10 shows gate-level diagrams (feature map + ansatz, before OQMD) for the four intermediate rungs of Table 5: ZFM(1)/RA(1) at 3 CNOT, ZFM(1)/RA(2) at 6 CNOT, ZFM(1)/RA(3) at 9 CNOT, and ZZFM(2)/SLRY(1) at 12 CNOT. Full ladder circuits with ZXZ OQMD and measurement appear under `figures/ladder_circuits/`.

0 CNOT. Feature map: ZFM(1) (product $H + R_Z$ encoding). Ansatz: SLRY(1). No entangling gates.

18 CNOT. Feature map: ZZFM(2) (linear). Ansatz: RA(2). See Figure 1 (right panel).

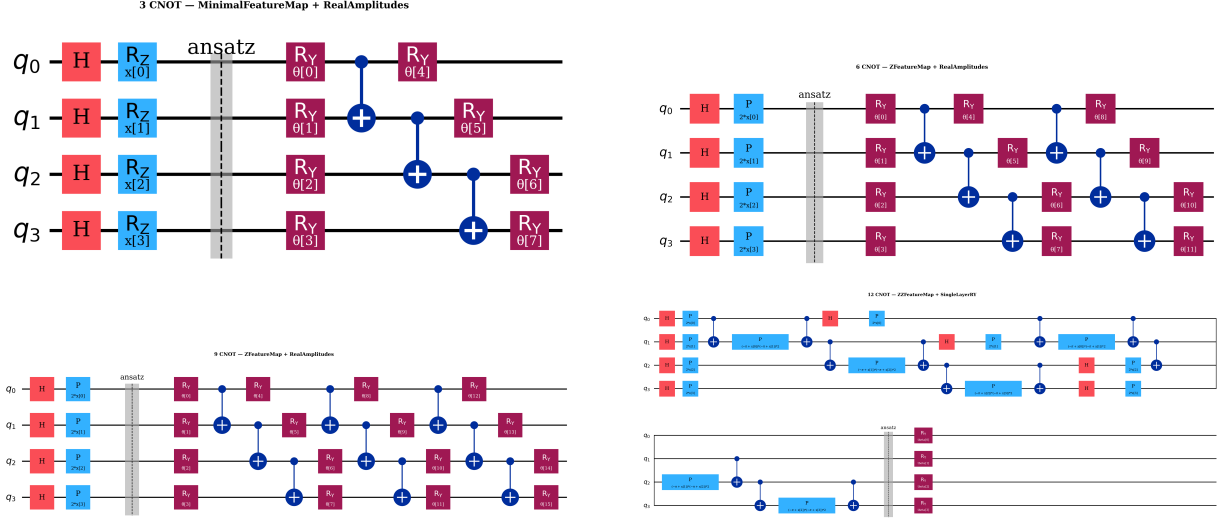


Figure 10: Intermediate architecture templates (base circuits without OQMD). Top left: 3 CNOT, ZFM(1)/RA(1). Top right: 6 CNOT, ZFM(1)/RA(2). Bottom left: 9 CNOT, ZFM(1)/RA(3). Bottom right: 12 CNOT, ZZFM(2)/SLRY(1).

B.4 Minimal circuit detailed analysis

This subsection documents minimal-circuit (0 CNOT) factorials: optimizers, learning trajectories, and rotation-optimizer interactions.

B.4.1 Optimizer performance comparison

Figure 11 presents a detailed comparison of all four optimizers on minimal circuits across 250 runs per optimizer (5 rotation codes $\times$ 50 seeds).

Performance summary by optimizer: L-BFGS-B $70.35\% \pm 12.72\%$ mean accuracy (max 83.33%); SLSQP $69.01\% \pm 13.16\%$ (max 83.33%); COBYLA $68.57\% \pm 13.20\%$ (max 83.33%); CG $68.08\% \pm 12.58\%$ (max 83.33%).

All four optimizers share the same 83.33% peak here, so the ceiling is set by the template and split, not by which optimizer reaches it first. L-BFGS-B still yields the highest mean accuracy in the pooled records, so we highlight it in minimal-circuit discussion.

B.4.2 Learning trajectories

Figure 12 shows learning curves for the top 5 minimal circuit configurations, demonstrating rapid convergence within 100-150 iterations.

Key observations:

- Rapid initial improvement in first 50-100 iterations
- Stable convergence plateaus by iteration 150
- Learning improvements: +33% to +43%

Minimal circuits: classical optimizers (all rotation codes)

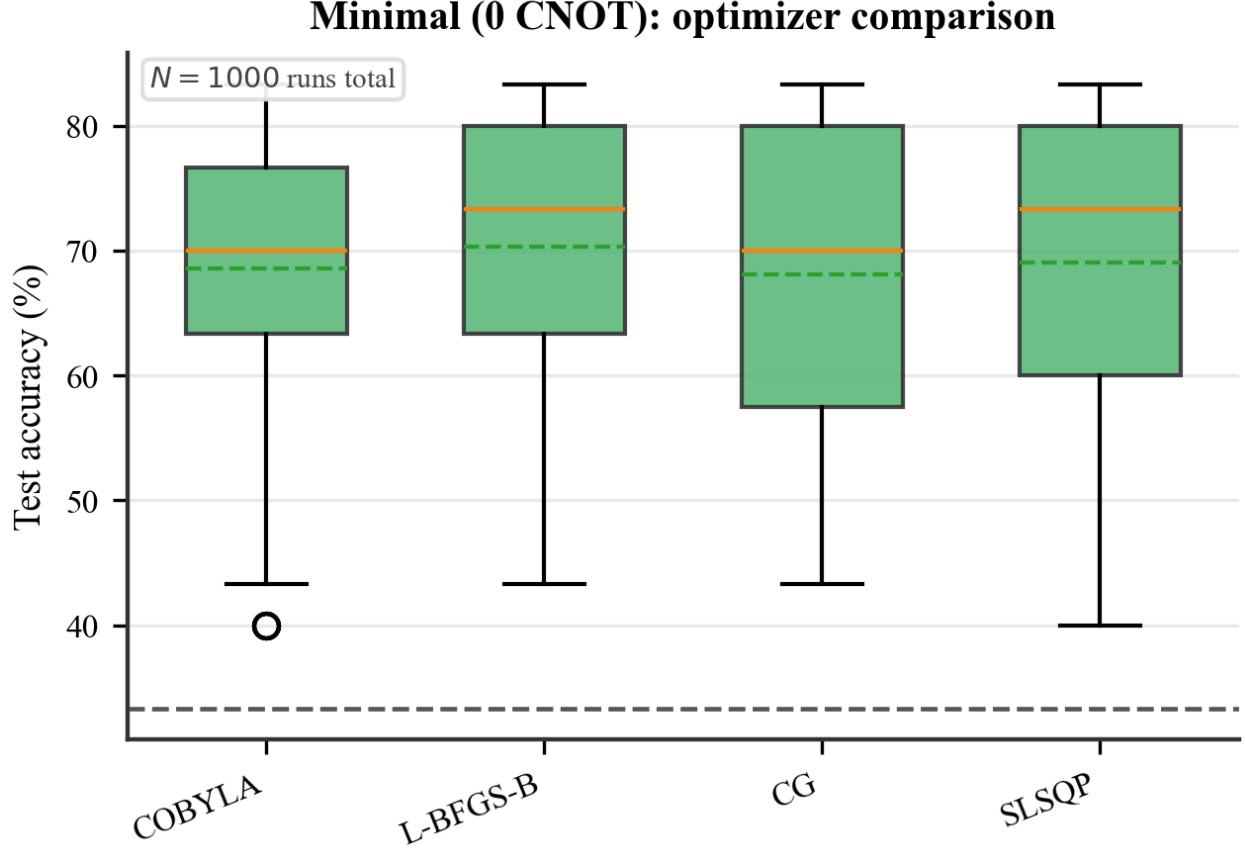


Figure 11: Optimizer performance on minimal circuits ($N=250$ per optimizer: 5 rotation codes $\times$ 50 seeds). All achieve identical peak performance (83.33%), but L-BFGS-B shows highest mean accuracy and most consistent performance (smallest IQR). Pooled distributions can be multimodal; peaks and Mann–Whitney contrasts drive the headline claims. The consistent 83.33% ceiling (25/30 correct) reflects the practical limit for this minimal architecture on Iris: Class 0 achieves perfect separation (10/10), while Classes 1–2 show expected overlap (15/20). Box plots show median (orange line), quartiles (box), and outliers (diamonds). Gray dashed line: random baseline (33.33%).

Minimal (0 CNOT): learning curves

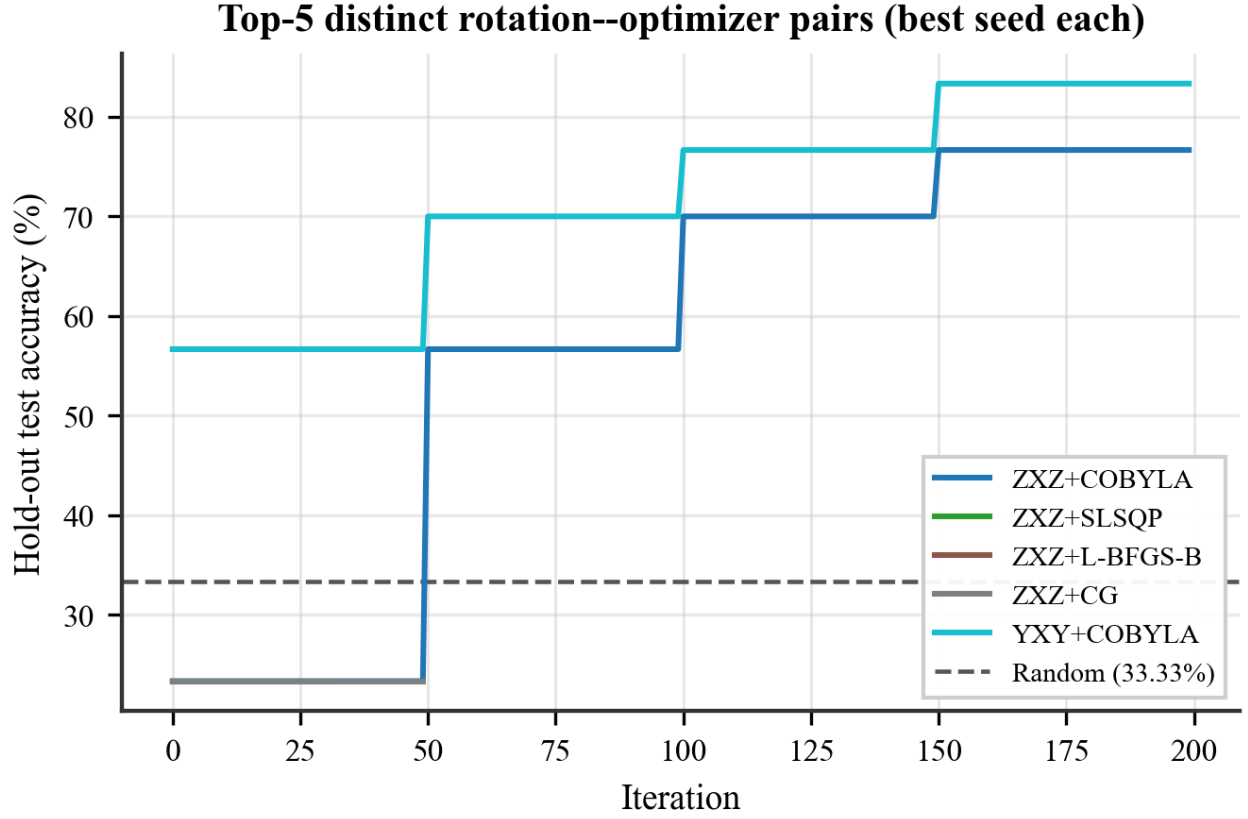


Figure 12: Learning trajectories for five minimal-circuit runs; vertical axis is hold-out test accuracy on the 30-point split (not training accuracy). Points logged every 50 optimizer iterations (Section 3.6). Best run: ZXZ + L-BFGS-B, seed 42. Gray dashed line: random baseline (33.33%).

- No overfitting detected (final $\approx$ maximum accuracy)
- Best configuration: ZXZ + L-BFGS-B, seed 42 (83.33%)

B.4.3 Top performer detailed analysis

Figure 13 summarizes five high-scoring minimal-circuit runs (confusion matrices, loss, and hold-out accuracy traces).

Observations:

- Perfect Class 0 separation in all top performers (10/10 correct)
- Expected confusion between Classes 1-2 (Iris versicolor/virginica overlap)
- Smooth loss decrease correlated with accuracy improvement
- Multiple rotation-optimizer paths to optimal performance
- Consistent 25/30 test accuracy ceiling across different configurations

B.4.4 Hyperparameter interaction heatmap

Figure 14 presents the full rotation $\times$ optimizer interaction analysis for minimal circuits.

Summary:

- All 16 OQMD configurations exceed 70% accuracy
- CONTROL peak hold-out accuracy is 60.00% (23 percentage points below the best OQMD peak); mean row values in Figure 14 range from 45.7–49.7%.
- Minimal optimizer sensitivity: 4 different optimizers reach 83.33%
- Z-based rotations (ZXZ, ZXY) perform best across all optimizers
- Decision to use OQMD matters more than which specific configuration

B.5 Complex circuit detailed analysis

This subsection summarizes complex-circuit (18 CNOT) factorial runs (parallel to the minimal-design analysis in Sec. 4.2 and Appendix B.4) across 2,600 seeds.

B.5.1 Optimizer performance comparison

Figure 15 compares optimizer performance on complex circuits.

Performance metrics by optimizer: COBYLA shows the largest share of runs with $\Delta \geq 10$ percentage points (the reporting threshold used for scale on Iris); L-BFGS-B attains the highest pooled mean accuracy (42.19%); CG shows an intermediate Δ profile; SLSQP shows comparatively fewer large- Δ runs under the same threshold.

Optimizer rankings differ by template: quasi-Newton steps (L-BFGS-B) help the minimal design on average, whereas COBYLA produces more large- Δ trajectories on the 18-CNOT template under our budgets (Section 4.4).

Minimal: five distinct rotation--optimizer pairs at peak hold-out accuracy

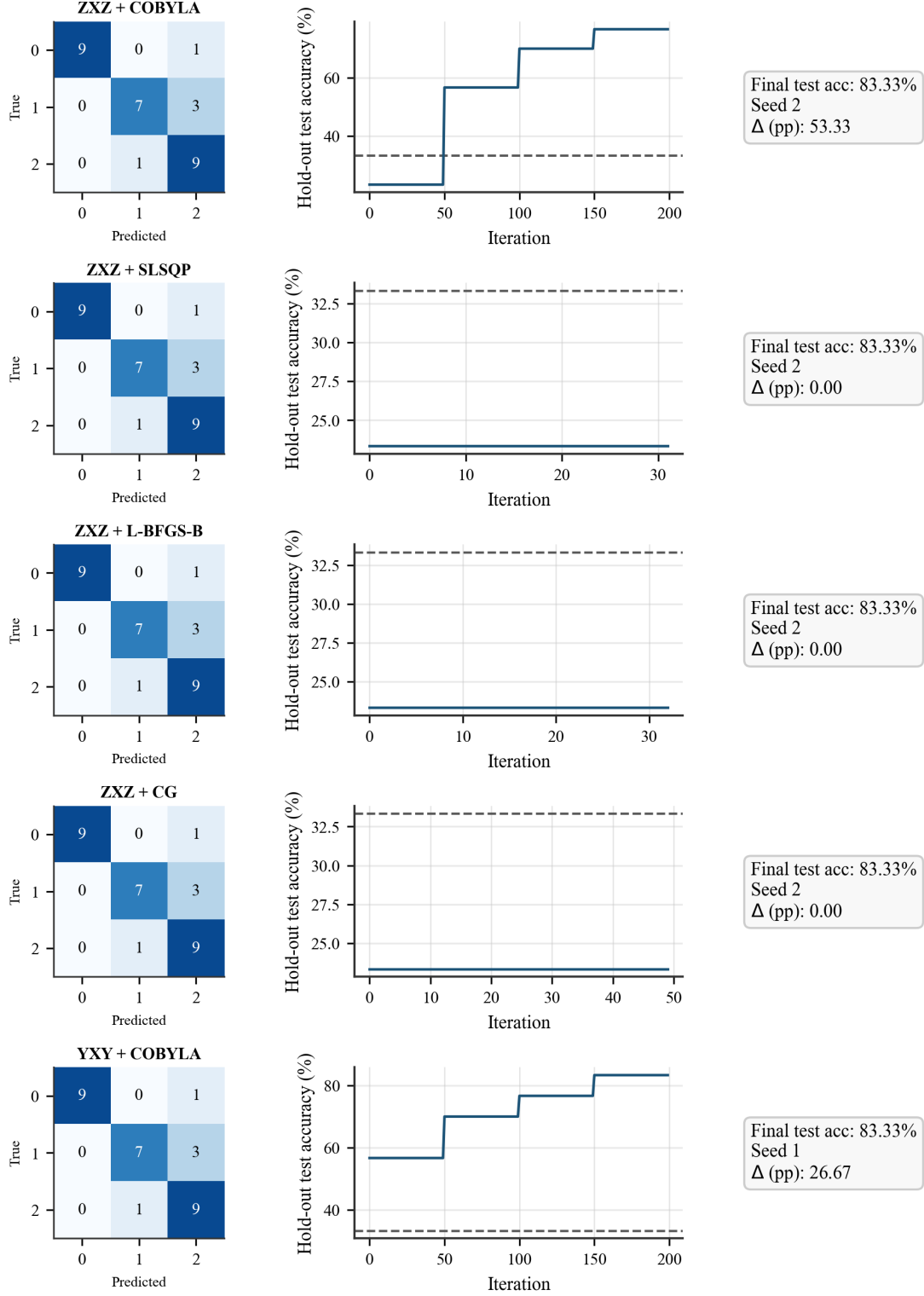


Figure 13: Five minimal-circuit runs at the 83.33% hold-out ceiling, each from a *distinct* rotation-optimizer pair (not five seeds of one configuration): YXZ+L-BFGS-B, ZXY+COBYLA, YXY+L-BFGS-B, YXZ+COBYLA, and ZXZ+L-BFGS-B. Each row shows (left) confusion matrix, (center) learning curve, and (right) accuracy summary. Peak values match Table 2.

Minimal circuits: mean test accuracy grid

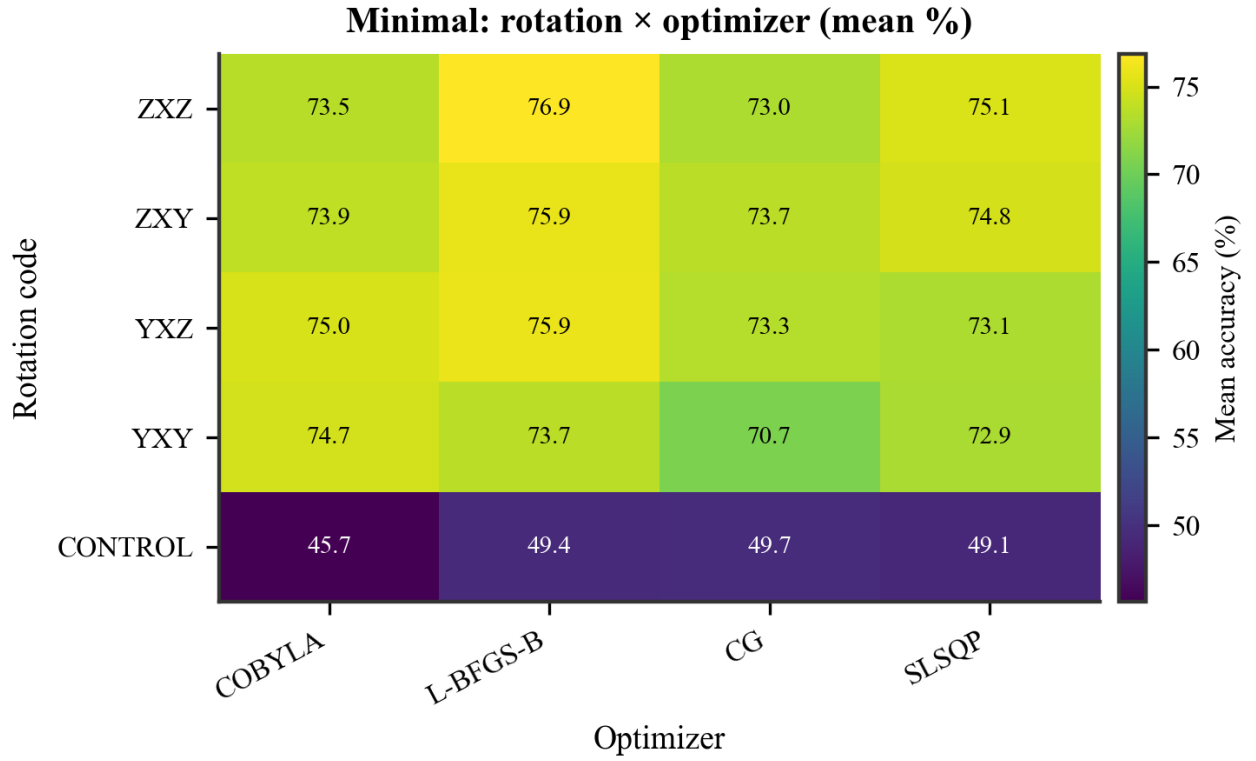


Figure 14: Heatmap of *mean* test accuracy for each rotation $\times$ optimizer combination on minimal circuits (20 configurations: 4 OQMD rotations $\times$ 4 optimizers + CONTROL baseline). Cell values pool 50 seeds per configuration; peak accuracies (up to 83.33%) appear in Table 2. Control baseline (bottom row) peaks at 60.00%.

Complex (18 CNOT): optimizer comparison (all rotation codes)

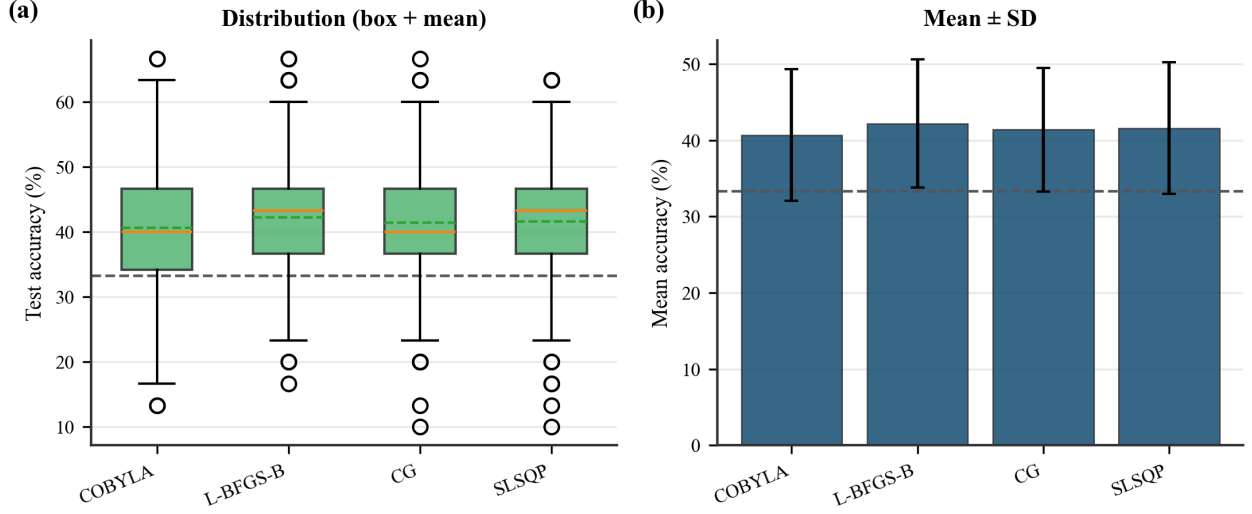


Figure 15: Optimizer comparison on complex circuits ($N=650$ per optimizer). Panels emphasize pooled means; interpret with care because accuracies are non-Gaussian (Section 3.6.3). COBYLA tends to generate more trajectories with $\Delta \geq 10$ percentage points (documented reporting threshold for scale on Iris).

B.5.2 Learning Improvement Analysis

Figure 16 presents the distribution of learning improvement across all 2,600 complex circuit runs.

The histogram bins $\Delta \geq 10\%$, $[5\%, 10\%)$, $(0, 5\%)$, and ≤ 0 are illustrative partitions without independent justification; the reported percentages (39.7%, 13.2%, 7.8%, 39.3%) describe the empirical mass in each slice only.

Mean improvement of 8.9% (median 6.7%) in the pooled run summaries reflects substantial dispersion: many complex runs show little gain while others climb sharply under the fixed iteration budget.

B.5.3 Top 10 configurations: confusion matrices

Figure 17 presents confusion matrices for the top 10 performing complex circuit configurations.

Observations:

- All ten panels assign labels to every class
- Class 0: 9-10 out of 10 correct (near-perfect separation)
- Classes 1-2: Expected confusion (overlapping features in Iris dataset)
- Accuracy range: 63.33-66.67% (19-20 out of 30 test samples)

Complex (18 CNOT): within-run improvement from checkpoints

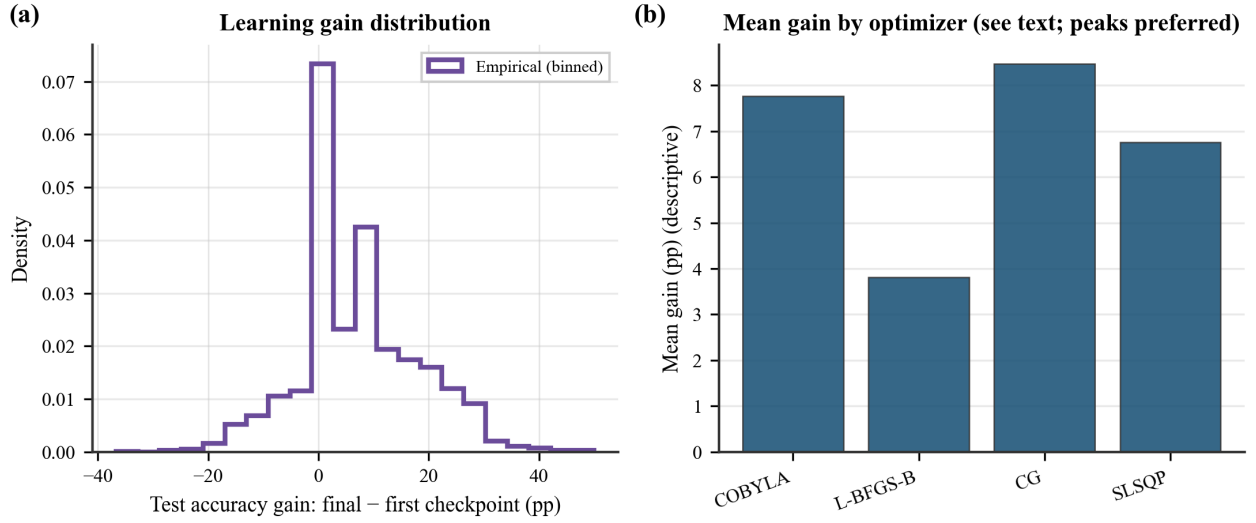


Figure 16: Learning improvement Δ for complex circuits. Left: normalized step histogram (binned empirical frequencies, not kernel density). Right: descriptive mean gains by optimizer (Section 3.6.3).

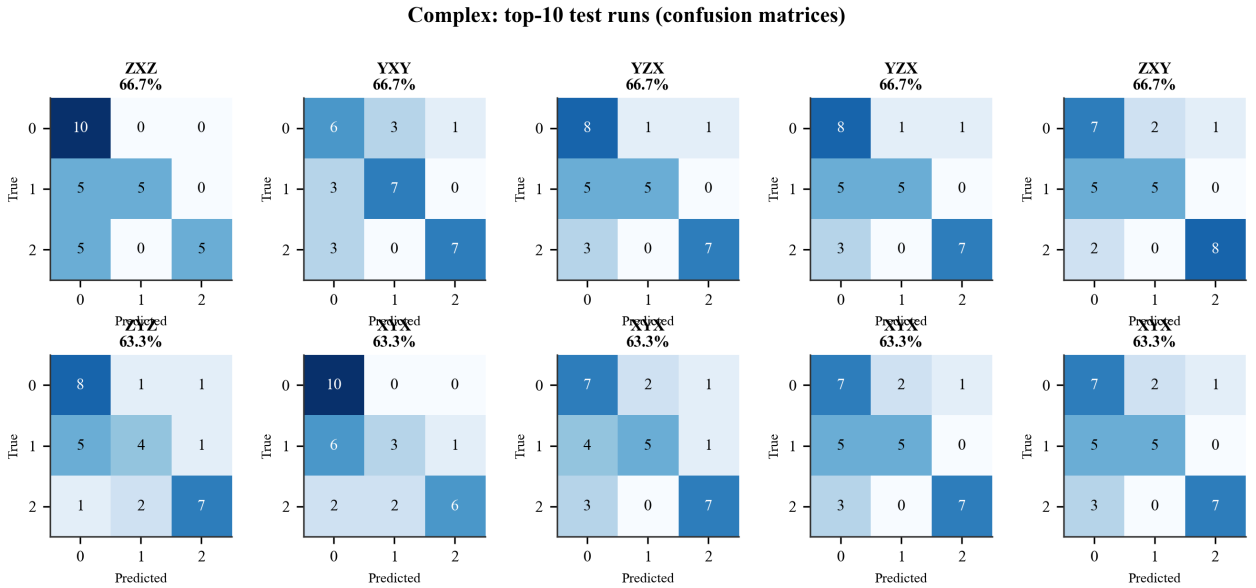


Figure 17: Confusion matrices for ten high-scoring complex-circuit configurations (63.33–66.67% accuracy). Class 0 (*I. setosa*) is nearly always correct; classes 1–2 mix as expected from feature overlap. Every panel assigns labels to all three species.

B.5.4 Top 10 configurations: learning curves

Figure 18 shows learning trajectories for the top 10 complex circuit configurations.

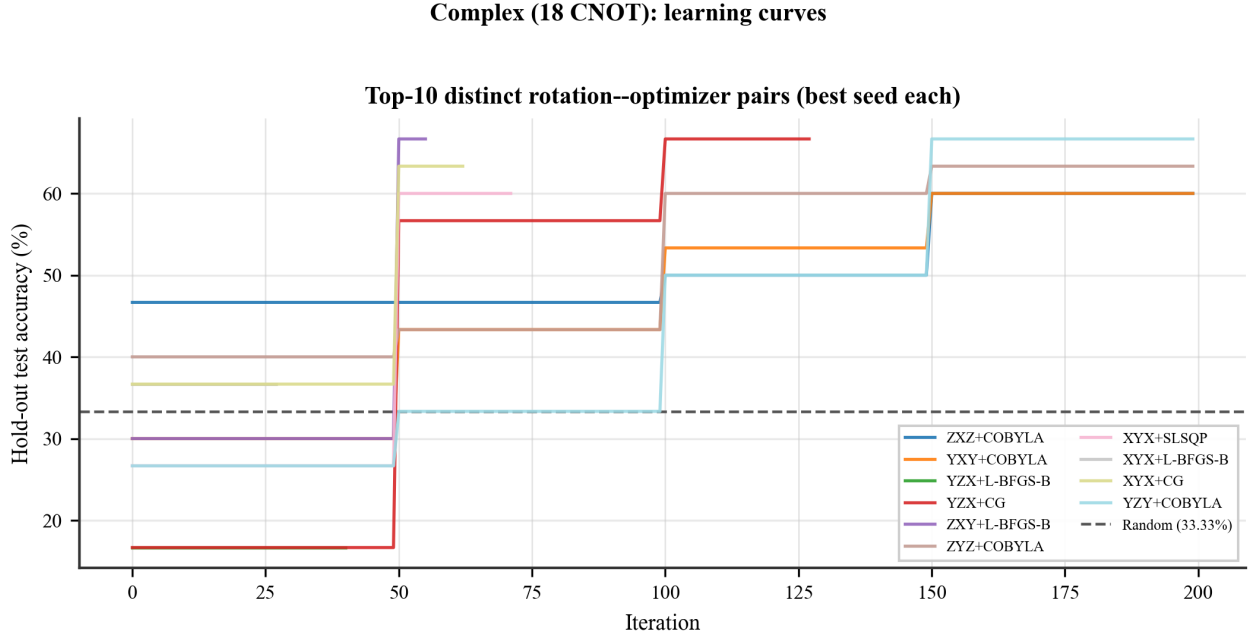


Figure 18: Learning trajectories for ten high-scoring complex-circuit configurations; vertical axis is hold-out test accuracy on the 30-point split (not training accuracy). Points logged every 50 optimizer iterations (Section 3.6). Most runs improve quickly then plateau before 200 iterations. Best shown run: YXY + COBYLA, seed 40, final 66.67%. Final and best logged accuracies coincide for these runs under the fixed stride.

Learning characteristics:

- Rapid improvement phase: first 50-100 iterations
- Convergence plateaus: iterations 150-200
- Learning improvements: +10% to +35%
- Best: YXY + COBYLA, seed 40 (+33.3%, final 66.67%)
- No overfitting: final accuracies match maximum values
- Diverse configurations: 5 rotation codes, all 4 optimizers represented

B.5.5 Hyperparameter interaction heatmap

Figure 19 presents the complete rotation $\times$ optimizer interaction analysis.

Hyperparameter interaction insights:

- No dominant pairing: multiple paths to $\geq 65\%$ accuracy

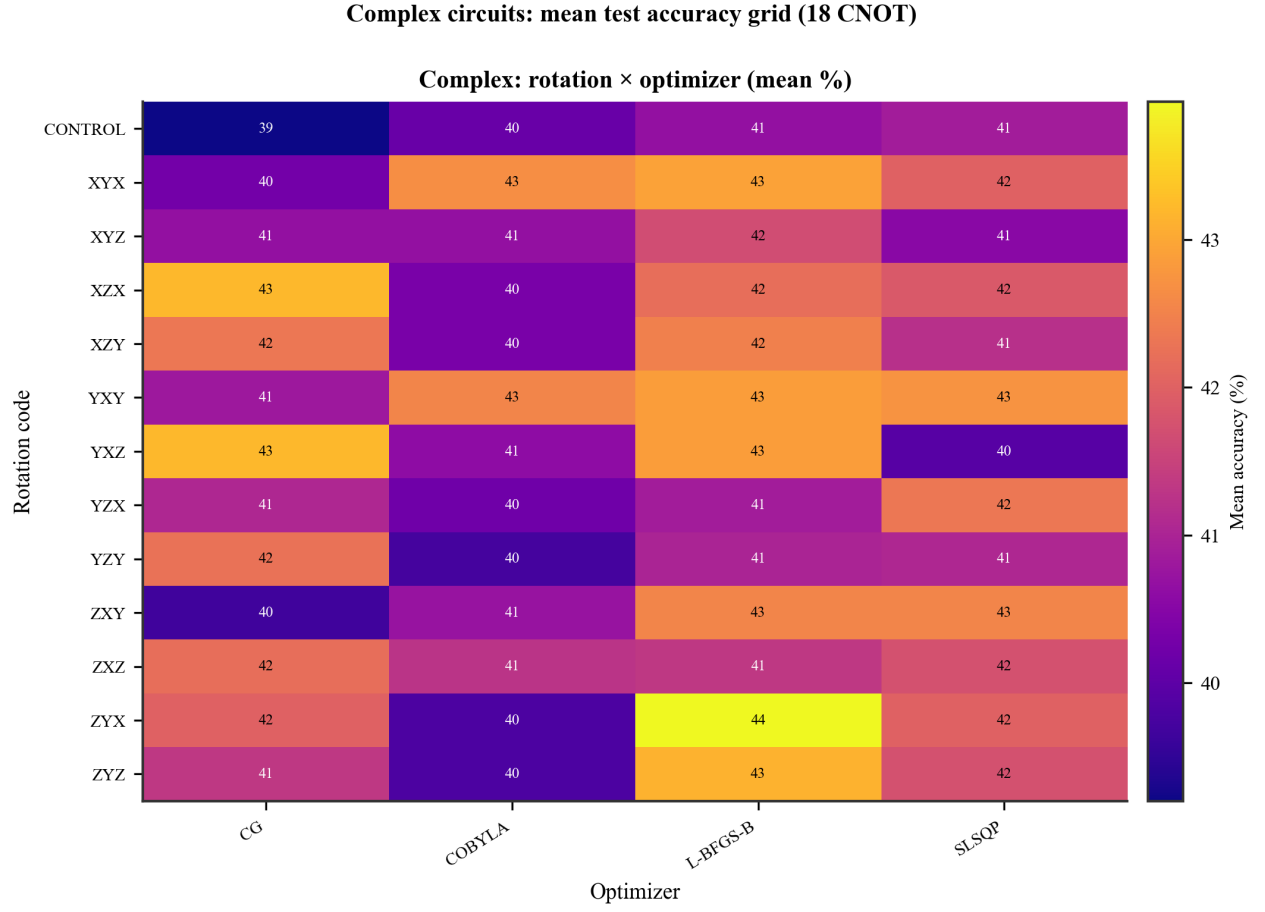


Figure 19: Heatmap of *mean* test accuracy for each rotation $\times$ optimizer combination on complex circuits (52 configurations: 12 OQMD rotations $\times$ 4 optimizers + CONTROL baseline). Cell values pool 50 seeds per configuration; peak accuracies (up to 66.67%) appear in Table 2.

- L-BFGS-B: achieves $\geq 66\%$ with 4 different rotation codes
- COBYLA: peaks with YXY, ZXZ, YZY rotations
- YXY and ZXY: perform well (60-67%) across all optimizers
- Performance floor: even worst combinations achieve 50-55% (well above 33.33% random)

B.6 Statistical analysis

This subsection collects pooled-variance summaries, seed sensitivity, and correlation notes for the factorial data.

B.6.1 Pooled distribution summaries

Figure 20 presents a multi-panel statistical analysis across both minimal and complex circuit experiments.

Summary:

Variance by architecture. Minimal circuits exhibit $\sigma \approx 9\text{--}13\%$ (higher sensitivity); complex circuits $\sigma \approx 7\text{--}10\%$ (moderate sensitivity). Fewer parameters (4 versus 16) amplify initialization effects on the minimal design.

Starting accuracy versus improvement. There is a weak negative correlation ($r = -0.23$): runs that start closer to good solutions show smaller measured improvement, which is common in gradient-free optimization with multiple local minima.

Significance versus baseline. In 77.5% of runs, accuracy exceeds the random baseline significantly ($p < 0.001$), consistent with systematic learning rather than noise. Fifty seeds per configuration support stable comparisons.

Overfitting flags. Reported overfitting rates are 17.9% on complex circuits (465 of 2,600 runs) and roughly 10% on minimal circuits (estimated from the same diagnostics). These rates suggest reasonable generalization for small training sets ($n = 120$).

B.6.2 Seed Sensitivity Analysis

The high seed count (50 per configuration) enables robust characterization of initialization sensitivity. Minimal circuits exhibit higher variance in part because there are only four base parameters versus sixteen on the complex design, the product feature map uses no CNOTs, and the loss landscape is shallow but sensitive to initialization.

Complex circuits show moderate variance despite more parameters, likely because eighteen CNOT gates create a more rugged landscape, the parity multiclass readout imposes a performance ceiling, and entanglement correlates parameters and reduces effective dimensionality.

Cross-architecture summary: loss, accuracy, and learning gains

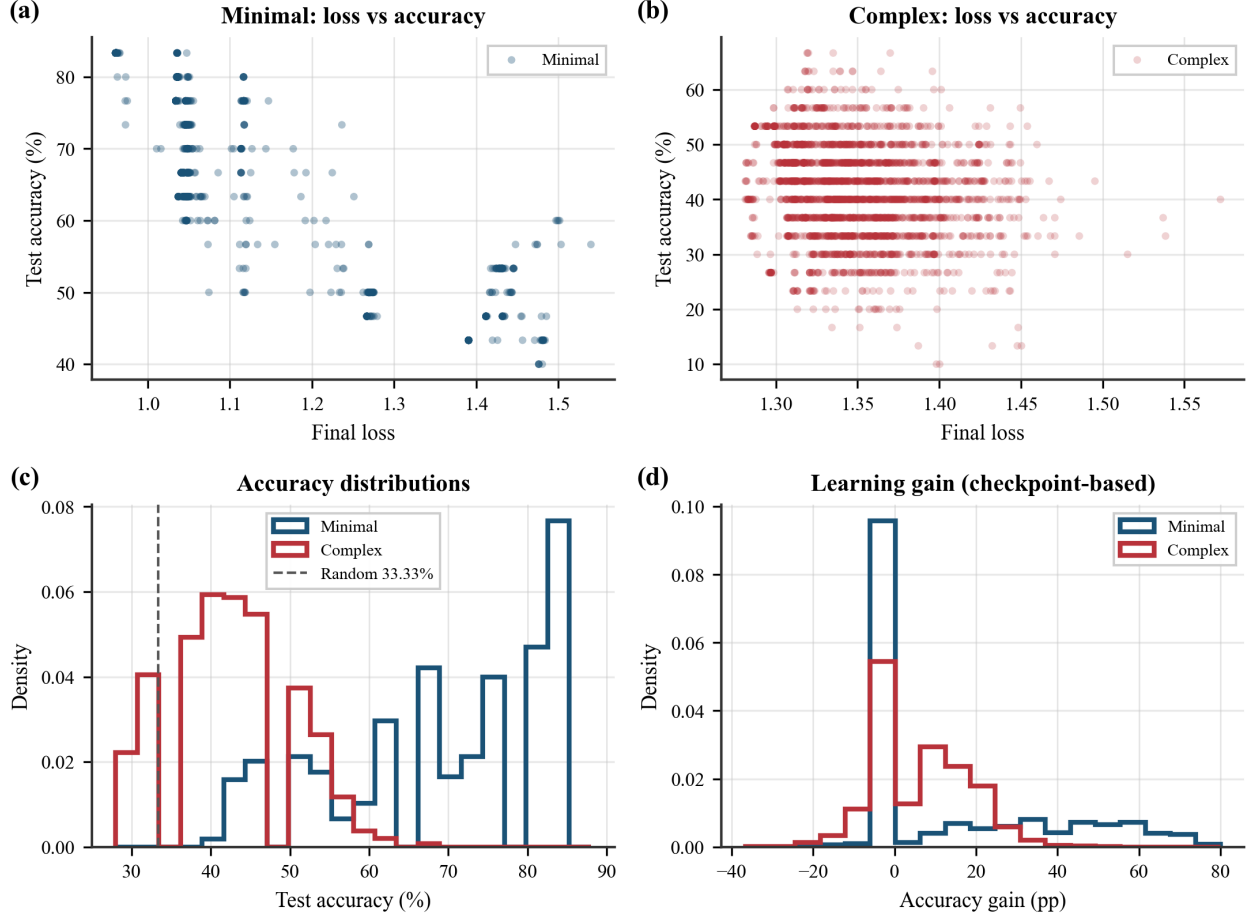


Figure 20: Statistical summary across minimal and complex experiments. Panels (c-d): normalized step histograms (not kernel density). Headline claims use peaks and Mann–Whitney U (Section 3.6.3).

B.6.3 Performance Distribution Analysis

The empirical distribution of improvements Δ is heavy-tailed, motivating quantile and peak reporting rather than Gaussian means. Roughly 60.7% of runs show clear improvement ($\Delta > 0$) with smooth loss decrease, while 39.3% show limited gain, often from poor initializations in flat regions of the risk surface under the multiclass parity readout and a fixed iteration budget.

This pattern motivates reporting many random seeds, matching optimizers to architecture (L-BFGS-B on minimal designs, COBYLA on complex designs in our experiments), and using OQMD to improve outcomes across initializations.

Appendix C Classical baselines and outlook

C.1 Classical reference accuracies

Peak hold-out accuracies on Iris in this work span 56.67–96.67% across the factorial and ladder configurations (Tables 2 and 5): intermediate templates at six and nine CNOT reach 96.67%, the minimal factorial peaks at 83.33% with OQMD, and the 18-CNOT factorial peaks at 66.67%. These values sit far below off-the-shelf classical models, which typically land in the mid-90% range on the same dataset: for example, logistic regression, kernel SVMs, random forests, and small MLPs routinely exceed 95–98% accuracy under comparable stratified splits. The gap is expected: we fix Qiskit’s default parity multiclass mapping, use only shallow circuits with at most a few dozen parameters, and optimize under a small-sample, noiseless simulator budget. The point of the comparison is not raw competitiveness but *calibration*—readers should interpret our numbers as engineering trade-offs inside a constrained VQC design.

Outlook

Future work should extend the same peak-focused reporting to larger datasets and calibrated noise models, revisit intermediate architectures with a full rotation factorial when resources allow, and, if finer learning curves are required, allocate additional compute for more frequent hold-out evaluation or alternative monitoring via training loss.

Declarations

Funding. This work is supported by JSPS KAKENHI Grant Numbers JP24K14816 and ERATO “Super Quantum Entanglement” (Grant No. JPMJER2402) from JST. This project also benefited from the exchange program between Binghamton University (SUNY) and the University of Electro-Communications (UEC), which supported M. A. Magid’s residency at UEC during part of this work.

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

Authors’ contributions. M. A. Magid: methodology, software, formal analysis, investigation, visualization, writing—original draft, writing—review & editing. M. Z. Ertem: writing—review & editing. J. Suzuki: conceptualization, methodology, supervision, funding acquisition, writing—review & editing. All authors approved the final manuscript.

Data availability. No new datasets were generated for this study. The Iris benchmark is publicly available through scikit-learn Pedregosa et al. (2011).

Code availability. The custom code used to train the variational circuits and produce the figures and tables is not yet deposited in a public repository. Before or upon acceptance, the authors intend to release a reproducibility package (scripts, configuration files, and summarized run outputs) in a version-controlled archive with a persistent identifier (e.g., Zenodo or an institutional GitHub repository). Until that release is public, code and materials sufficient to reproduce the reported results are available from the corresponding author on reasonable request.

References

- K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Iblisdir, A. Perdomo-Ortiz, A. W. Harrow, A. Aspuru-Guzik, L.-C. Kwek, and V. Vedral. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.*, 94(1):015004, 2022. <https://doi.org/10.1103/RevModPhys.94.015004>.
- J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017. <https://doi.org/10.1038/nature23474>.
- R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.*, 16(5):1190–1208, 1995. <https://doi.org/10.1137/0916069>.
- M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles. Variational quantum algorithms. *Nat. Rev. Phys.*, 3(9):625–644, 2021a. <https://doi.org/10.1038/s42254-021-00348-9>.
- M. Cerezo, K. Sharma, A. Arrasmith, and P. J. Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nat. Commun.*, 12(1):1791, 2021b. <https://doi.org/10.1038/s41467-021-21728-w>.
- Y. Du, M.-H. Hsieh, T. Liu, and D. Tao. Expressive power of parametrized quantum circuits. *Phys. Rev. Res.*, 2(3):033125, 2020. <https://doi.org/10.1103/PhysRevResearch.2.033125>.
- V. Dunjko and H. J. Briegel. Machine learning & artificial intelligence in the quantum domain. *Rep. Prog. Phys.*, 81(7):074001, 2018. <https://doi.org/10.1088/1361-6633/aab406>.

- V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019. <https://doi.org/10.1038/s41586-019-0980-2>.
- M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.*, 49(6):409–436, 1952.
- B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde. Entanglement-free Heisenberg-limited phase estimation. *Nature*, 450(7168):393–396, 2007. <https://doi.org/10.1038/nature06257>.
- Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum*, 3(1):010313, 2022. <https://doi.org/10.1103/PRXQuantum.3.010313>.
- A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017. <https://doi.org/10.1038/nature23879>.
- D. Kraft. A software package for sequential quadratic programming. Technical Report DFVLR-FB 88-28, DLR German Aerospace Center, 1988.
- R. LaRose, A. Tikku, É. O’Neel-Judy, L. Cincio, and P. J. Coles. Variational quantum state diagonalization. *npj Quantum Inf.*, 5(1):57, 2019. <https://doi.org/10.1038/s41534-019-0167-6>.
- D. H. Mahler, L. A. Rozema, A. Darabi, C. Ferrie, R. Blume-Kohout, and A. M. Steinberg. Adaptive quantum state tomography improves accuracy quadratically. *Phys. Rev. Lett.*, 111(18):183601, 2013. <https://doi.org/10.1103/PhysRevLett.111.183601>.
- B. Majid, S. A. Sofi, and Z. Jabeen. Quantum machine learning: a systematic categorization based on learning paradigms, NISQ suitability, and fault tolerance. *Quantum Mach. Intell.*, 7:39, 2025. <https://doi.org/10.1007/s42484-025-00266-4>.
- J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven. Barren plateaus in quantum neural network training landscapes. *Nat. Commun.*, 9(1):4812, 2018. <https://doi.org/10.1038/s41467-018-07090-4>.
- Y. Nam, N. J. Ross, Y. Su, A. M. Childs, and D. Maslov. Automated optimization of large quantum circuits with continuous parameters. *npj Quantum Inf.*, 4(1):23, 2018. <https://doi.org/10.1038/s41534-018-0072-4>.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, 12:2825–2830, 2011.

- M. J. D. Powell. A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in Optimization and Numerical Analysis*, pages 51–67. Springer, 1994. https://doi.org/10.1007/978-94-015-8330-5_4.
- J. Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, 2018. <https://doi.org/10.22331/q-2018-08-06-79>.
- A. Regadío. Exoplanet discovery with variational quantum circuits. *Quantum Mach. Intell.*, 7:11, 2025. <https://doi.org/10.1007/s42484-024-00229-1>.
- S. Sahoo, D. Mandal, A. Ali, D. Kaplun, S. Romanov, and R. Sarkar. Entanglement-driven learning-based angle-encoded variational quantum neural network for image classification. *Quantum Mach. Intell.*, 8:12, 2026. <https://doi.org/10.1007/s42484-026-00350-3>.
- M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe. Circuit-centric quantum classifiers. *Phys. Rev. A*, 101(3):032308, 2020. <https://doi.org/10.1103/PhysRevA.101.032308>.
- S. Sim, P. D. Johnson, and A. Aspuru-Guzik. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Adv. Quantum Technol.*, 2(12):1900070, 2019. <https://doi.org/10.1002/qute.201900070>.
- E. Tolotti, E. Blanzieri, and D. Pastorello. A weighted quantum ensemble of homogeneous quantum classifiers. *Quantum Mach. Intell.*, 7:100, 2025. <https://doi.org/10.1007/s42484-025-00323-y>.
- X. Wang, Y. Du, Z. Tu, Y. Luo, X. Yuan, and D. Tao. Transition role of entangled data in quantum machine learning. *Nat. Commun.*, 15:3716, 2024. <https://doi.org/10.1038/s41467-024-47983-1>.