PARITY-CONSTRAINED QUANTUM RESERVOIR COMPUTING FOR IMAGE CLASSIFICATION: FORMAL GUARANTEES AND STAGED SIMULATION EVIDENCE

Anonymous authors
Paper under review

ABSTRACT

Quantum reservoir computing has recently reported encouraging image-classification performance, yet many claims remain sensitive to fairness controls, measurement-policy confounds, and benchmark selection. We study a parity-constrained evaluation program for PCA-encoded image inputs and transverse-Ising-style reservoir dynamics, and we combine formal analysis with staged simulation evidence under fixed CPU-only constraints. First, we formulate advantage assessment as a bi-level optimization problem with matched preprocessing, readout family, observable budget, and search budget across quantum and classical branches, and we prove that parity-controlled deltas cannot exceed naive deltas computed with asymmetric policy optimization. Second, we formalize a non-monotone entanglement-utility criterion, showing that boundary derivative sign changes and interior concavity imply a unique interior optimum in coupling strength. Third, we derive an operator-attribution framework based on balanced crossed random effects and prove range and unbiasedness properties for operator-share estimators. Using a staged validation run over MNIST, Fashion-MNIST, EMNIST Balanced, Kuzushiji-MNIST, and a grayscale PCA variant of CIFAR-10, we observe conditional practical advantage on one hard dataset under a pre-registered acceptance tuple, consistent interior-optimum signatures in the tested regime, and strong operator-share estimates with parity-audit closure. These results support a guarded conclusion: formal guarantees are strong, empirical gains are conditional, and attribution quality improves materially when parity is enforced at the row level.

1 INTRODUCTION

Reservoir computing reduces training complexity by fixing nonlinear dynamics and learning only a linear or ridge readout, a perspective that has remained influential from echo-state and liquid-state systems to contemporary quantum reservoir variants (Jaeger, 2001; Maass et al., 2002; Lukoševičius & Jaeger, 2009). In the quantum setting, this design is attractive because it avoids large variational parameter searches and can exploit hardware-native dynamics, while still supporting supervised prediction through classical post-processing. At the same time, recent quantum machine-learning literature repeatedly emphasizes that advantages depend on end-to-end assumptions, including data encoding, measurement design, and baseline parity (Biamonte et al., 2017; Havlíček et al., 2019; Schuld et al., 2020; Schuld, 2021). For image classification, these concerns are amplified when dimensionality reduction is applied upstream: if preprocessing already renders classes near-linearly separable, apparent gains can collapse under stronger controls.

This paper addresses a concrete open question: can PCA-encoded quantum reservoirs outperform strong classical reservoirs on image tasks once protocol parity is enforced, and what role does entanglement actually play in that outcome? Recent QELM and QRC studies report improved separability and accuracy in selected image and time-series settings (De Lorenzis et al., 2024; 2025; Prieto-Garcia et al., 2026; Antoncich et al., 2026; Liu et al., 2026). However, several papers also show that measurement policy optimization can shift performance substantially (Gross & Rieser, 2026a; Aaraba et al., 2026; Gross & Rieser, 2026b), and multiple entanglement-focused studies indicate regime dependence with possible saturation (De Lorenzis et al., 2025; Askari et al., 2025; Karimi et al., 2025). These observations motivate a methodology where advantage, mechanism, and attribution are evaluated in sequence rather than conflated.

Our study adopts that sequence explicitly. We first define a parity-constrained advantage objective with a pre-registered acceptance rule over a dataset ladder that includes MNIST and harder alternatives (LeCun et al., 1998; Xiao et al., 2017; Cohen et al., 2017; Clanuwat et al., 2018; Krizhevsky & Hinton, 2009). We then isolate entanglement effects via controlled coupling-time ablations at fixed readout and observable settings, and we finally quantify operator-policy

contribution using variance-component estimators under balanced design. This framing is deliberately hybrid: the paper is not purely theoretical and not purely benchmark-driven; instead, it develops formal guarantees and tests their operational implications in one integrated protocol.

The broader relevance extends beyond image classification. Fairness of comparator policies and identifiability of causal performance factors are recurring issues in reservoir and kernel methods, quantum and classical alike (Schuld, 2021; Gross & Rieser, 2026a; Lukoševičius & Jaeger, 2009). A parity-first methodology can therefore serve as a template for other domains where fixed-feature pipelines are evaluated under finite compute budgets, including molecular prediction and temporal forecasting (Beaulieu et al., 2024; Ahmed et al., 2025; Hamhoum et al., 2025; Di Bartolo et al., 2025; Kodali et al., 2025).

Contributions. We make four contributions:

- We formalize practical-advantage testing as a parity-constrained bi-level program and prove a parity-shrink inequality showing that confound-controlled deltas are upper-bounded by naive asymmetric deltas.
- We provide an interior-optimum theorem for entanglement-conditioned utility and show how boundary derivative signs and concavity yield a unique operating point under explicit assumptions.
- We derive operator-share attribution estimators in a balanced crossed random-effects model, including range guarantees and unbiasedness identities used for interpretable causal decomposition.
- We report staged simulation evidence with strict parity auditing and pre-registered acceptance constants, yielding conditional advantage on a hard benchmark, non-monotone mechanism support in the tested regime, and high operator-share estimates after rerun-level parity closure.

2 Related Work and Motivation

2.1 Reservoir and Kernel Foundations

Classical reservoir computing established an enduring abstraction: rich, fixed, high-dimensional dynamics followed by convex readout fitting (Jaeger, 2001; Maass et al., 2002; Lukoševičius & Jaeger, 2009). Quantum supervised models can be interpreted through an analogous kernel lens, where data-encoding states define similarity geometry and measurement/readout specifies the final hypothesis class (Havlíček et al., 2019; Schuld, 2021). This perspective clarifies two practical points that are central to our design. First, expressivity and generalization are heavily conditioned by encoding choices and induced spectra (Schuld et al., 2020). Second, training quality comparisons are meaningful only when readout classes and optimization budgets are controlled symmetrically across alternatives.

These points are not merely conceptual. In finite-sample settings, subtle differences in preprocessing and regularization grids can create large apparent improvements that do not survive parity checks. Classical RC literature already warns that tuning policy can dominate architecture conclusions (Lukoševičius & Jaeger, 2009); recent QRC literature reinforces the same caution for quantum observables and kernelized readouts (Gross & Rieser, 2026a; Aaraba et al., 2026). Our protocol therefore treats parity not as a reporting footnote but as an explicit feasibility constraint in the optimization problem.

2.2 Image-Centric QELM and QRC Evidence

Image-focused quantum reservoir work has evolved rapidly. QELM studies combining PCA or learned compression with quantum dynamics and linear readouts have reported accuracy gains on MNIST-like and harder datasets, while often stressing computational efficiency from training only the readout layer (De Lorenzis et al., 2024; 2025). Additional hardware and simulation studies suggest that scalable reservoir-style workflows can be competitive in selected classification tasks (Kornjača et al., 2024; Prieto-Garcia et al., 2026; Antoncich et al., 2026; Liu et al., 2026; Carles et al., 2025; Świerczewski et al., 2026; Brusaschi et al., 2026; Maier et al., 2025; Assil et al., 2026).

Yet the literature remains heterogeneous in benchmark hardness, noise assumptions, and comparator strength. MNIST remains a useful sanity floor but frequently saturates under modern pipelines, including classical baselines with simple feature engineering (LeCun et al., 1998; Xiao et al., 2017; Cohen et al., 2017). For that reason, our study uses a ladder from MNIST to harder alternatives and explicitly conditions practical-advantage claims on hard-set performance. This design aligns with concerns in both QML surveys and recent QRC papers: isolated high scores on easy tasks do not establish robust advantage (Biamonte et al., 2017; De Lorenzis et al., 2025; Gross & Rieser, 2026b).

2.3 Entanglement, Simulability, and Operator Confounds

Recent studies increasingly agree that entanglement can improve embedding geometry in bounded regimes, but they also note saturation, regime dependence, and potential classical simulability of relevant dynamics (De Lorenzis et al., 2025; Askari et al., 2025; Karimi et al., 2025). This is an important contrast to simplistic monotone narratives. If moderate coupling improves discrimination while strong coupling degrades utility, then mechanism claims should be framed through interior-optimum criteria, not endpoint comparisons.

A second recurring issue is operator-policy confounding. Kernel-based optimization of observables can materially change QRC performance (Gross & Rieser, 2026a; Aaraba et al., 2026), and related interpretability analyses show that measurement selection can control effective representation even when dynamics are fixed (Gross & Rieser, 2026b). Without matched search budgets against classical feature-selection alternatives, architecture-level conclusions risk overstating quantum-dynamics effects. Our attribution model therefore decomposes variance contributions from operator policy and dynamics under balanced design, and reports shrinkage from naive to parity-controlled deltas as a first-class diagnostic.

3 Problem Setting and Notation

We consider supervised image classification after deterministic preprocessing. Let raw images be mapped to PCA features $\mathbf{x} \in \mathbb{R}^d$ through a shared operator Π_d. For each candidate configuration, a reservoir pipeline produces fixed features $\mathbf{z}(\mathbf{x})$ followed by ridge readout. We compare quantum and classical branches under matched preprocessing, matched readout class, matched observable or feature dimension, and matched search budget.

Definition 3.1 (Parity-Constrained Feasible Set)**.** *Let $\boldsymbol{\theta}_Q$ and $\boldsymbol{\theta}_C$ denote quantum and classical policy parameters, respectively. The feasible set is*

$$\Theta_{\text{parity}} := \left\{ (\boldsymbol{\theta}_Q, \boldsymbol{\theta}_C) : \Pi_d^Q = \Pi_d^C,\ M_Q = M_C,\ \mathcal{R}_Q = \mathcal{R}_C,\ B_Q = B_C \right\}, \tag{1}$$

where M is observable/feature count, $\mathcal{R}$ is readout family and regularization grid, and B is the search-budget accounting rule.

For a dataset d_s, denote parity-controlled accuracy gap by $\Delta_{\text{par}}^{(d_s)}$ and naive gap by $\Delta_{\text{naive}}^{(d_s)}$. We use a pre-registered acceptance predicate

$$A_{\text{adv}}^{(d_s)} = \mathbf{1}\left[\Delta_{\text{par}}^{(d_s)} \geq \delta_{\min} \ \wedge\ p_{\text{BH}}^{(d_s)} \leq p_{\max} \ \wedge\ r^{(d_s)} \geq r_{\min} \right], \tag{2}$$

where p_{BH} is Benjamini-Hochberg corrected significance and r is a robustness score across seeds.

The inner readout problem for either branch is ridge regression over fixed features:

$$\mathcal{L}(\boldsymbol{W}) = \|\boldsymbol{W}\boldsymbol{Z} - \boldsymbol{Y}\|_F^2 + \lambda \|\boldsymbol{W}\|_F^2, \qquad \lambda > 0. \tag{3}$$

Its closed-form optimum is

$$\boldsymbol{W}^* = \boldsymbol{Y}\boldsymbol{Z}^\top (\boldsymbol{Z}\boldsymbol{Z}^\top + \lambda \boldsymbol{I})^{-1}. \tag{4}$$

The outer objective maximizes hard-set mean parity gap:

$$\boldsymbol{\theta}^* \in \arg\max_{\boldsymbol{\theta} \in \Theta_{\text{parity}}} \frac{1}{|\mathcal{D}_{\text{hard}}|} \sum_{d_s \in \mathcal{D}_{\text{hard}}} \Delta_{\text{par}}^{(d_s)}(\boldsymbol{\theta}). \tag{5}$$

In this study the hard set is pre-registered as Fashion-MNIST, EMNIST Balanced, Kuzushiji-MNIST, and grayscale PCA CIFAR-10.

3.1 Notation Summary

Table 1 summarizes symbols that recur in the method and proofs.

4 Methodology

4.1 Parity-Constrained Bi-Level Advantage Model

The first stage asks whether any claimed performance gain survives strict protocol parity. Using equation 3–equation 5, we solve ridge readout exactly and compare branch-level performance only within Θ_{parity}. The acceptance criterion in equation 2 is fixed before model selection to avoid threshold leakage.

Table 1: Notation used in the formal development. The table groups optimization, mechanism, and attribution symbols that are referenced repeatedly in section 4 and section 6. Symbols are defined in-line at first use and repeated here for readability during equation-heavy sections.

Symbol	Meaning
$\mathbf{x} \in \mathbb{R}^d$	PCA-compressed input vector
Π_d	Shared preprocessing operator
$\boldsymbol{Z}$	Feature matrix from reservoir measurements
$\boldsymbol{W}$	Readout weight matrix
λ	Ridge regularization parameter
Θ_{parity}	Feasible set with matched protocol controls
Δ_{par}	Parity-controlled quantum-classical gap
Δ_{naive}	Naive gap under asymmetric policy optimization
A_{adv}	Practical-advantage acceptance predicate
$g_t(J)$	Entanglement-geometry utility at time t and coupling J
R_{op}	Operator variance-share ratio

Two practical motivations justify this design. First, fixed-feature models are sensitive to readout regularization and feature count, so unconstrained comparison can conflate capacity and dynamics effects (Lukoševičius & Jaeger, 2009; Schuld, 2021). Second, recent operator-optimization work shows that asymmetric policy search can inflate apparent quantum gains (Gross & Rieser, 2026a; Aaraba et al., 2026). By embedding parity into the feasible set rather than post hoc filtering, we make fairness a structural property of the optimization.

Theorem 4.1 (Unique Inner Optimum)**.** *For $\lambda > 0$, the ridge objective in equation 3 is strictly convex in $\boldsymbol{W}$ and has unique minimizer given by equation 4.*

Proof. Vectorize $\boldsymbol{W}$ and write the Hessian of equation 3 as $2(\boldsymbol{Z}\boldsymbol{Z}^\top + \lambda \boldsymbol{I}) \otimes \boldsymbol{I}_{\text{out}}$. Because $\boldsymbol{Z}\boldsymbol{Z}^\top$ is positive semidefinite and $\lambda > 0$, the matrix $\boldsymbol{Z}\boldsymbol{Z}^\top + \lambda \boldsymbol{I}$ is positive definite. The Kronecker product with an identity matrix preserves positive definiteness, so the Hessian is positive definite and the objective is strictly convex. Setting the gradient to zero yields $2(\boldsymbol{W}\boldsymbol{Z}\boldsymbol{Z}^\top - \boldsymbol{Y}\boldsymbol{Z}^\top + \lambda \boldsymbol{W}) = 0$, which rearranges to $\boldsymbol{W}(\boldsymbol{Z}\boldsymbol{Z}^\top + \lambda \boldsymbol{I}) = \boldsymbol{Y}\boldsymbol{Z}^\top$. Since the factor in parentheses is invertible, right-multiplication gives $\boldsymbol{W}^* = \boldsymbol{Y}\boldsymbol{Z}^\top(\boldsymbol{Z}\boldsymbol{Z}^\top + \lambda \boldsymbol{I})^{-1}$, uniquely. □

Theorem 4.2 (Parity-Shrink Inequality)**.** *Define*

$$\Delta_{\text{naive}} := \sup_{o_Q \in \mathcal{O}_Q^B} \text{Acc}_Q(o_Q) - \text{Acc}_C(o_C^0), \quad \Delta_{\text{par}} := \sup_{o_Q \in \mathcal{O}_Q^B} \text{Acc}_Q(o_Q) - \sup_{o_C \in \mathcal{O}_C^B} \text{Acc}_C(o_C), \tag{6}$$

with $o_C^0 \in \mathcal{O}_C^B$. Then $\Delta_{\text{par}} \leq \Delta_{\text{naive}}$.

Proof. Because $o_C^0 \in \mathcal{O}_C^B$, we have $\sup_{o_C \in \mathcal{O}_C^B} \text{Acc}_C(o_C) \geq \text{Acc}_C(o_C^0)$. Subtracting the larger classical term from the same quantum supremum can only reduce (or leave unchanged) the resulting gap, hence

$$\begin{aligned} \Delta_{\text{par}} &= \sup_{o_Q \in \mathcal{O}_Q^B} \text{Acc}_Q(o_Q) - \sup_{o_C \in \mathcal{O}_C^B} \text{Acc}_C(o_C) \\ &\leq \sup_{o_Q \in \mathcal{O}_Q^B} \text{Acc}_Q(o_Q) - \text{Acc}_C(o_C^0) = \Delta_{\text{naive}}. \end{aligned}$$

□

4.2 Entanglement-Geometry Mechanism Model

For each fixed evolution time t, we define a scalar utility over coupling strength J:

$$g_t(J) := \alpha\, \text{Sil}(\mathbf{z}_{t,J}) + \beta\, \text{Margin}(\mathbf{z}_{t,J}) - \gamma\, \text{Err}(t, J), \tag{7}$$

with $J \in [0, J_{\max}]$ and $(\alpha, \beta, \gamma) = (1, 1, 1)$ in our validation run. This functional bridges representation geometry and predictive error, allowing mechanism claims to be tested through sign and curvature conditions rather than monotone trend assumptions.

Theorem 4.3 (Interior-Optimum Criterion). *Assume* $g_t \in C^1([0, J_{\max}]) \cap C^2((0, J_{\max}))$ *and*

$$\frac{dg_t}{dJ}(0) > 0, \qquad \frac{dg_t}{dJ}(J_{\max}) < 0, \qquad \frac{d^2 g_t}{dJ^2}(J) < 0 \; \forall J \in (0, J_{\max}). \tag{8}$$

Then there exists a unique $J_t^* \in (0, J_{\max})$ *maximizing* g_t *on* $[0, J_{\max}]$.

Proof. Continuity of dg_t/dJ and opposite endpoint signs in equation 8 imply, by the intermediate value theorem, existence of at least one $J_t^* \in (0, J_{\max})$ such that $dg_t/dJ(J_t^*) = 0$. The strict negativity of the second derivative on the interior implies strict concavity of g_t over $(0, J_{\max})$. A strictly concave differentiable function has at most one stationary point, so the stationary point is unique. For strictly concave functions on compact intervals, the unique stationary interior point is the global maximizer over the full interval. Therefore J_t^* exists, is unique, and is globally optimal. □

4.3 Operator-Policy Attribution Model

To identify whether performance shifts are primarily due to dynamics or measurement policy, we adopt a balanced crossed random-effects model on cell-level outcomes:

$$y_{ijr} = \mu + a_i + b_j + (ab)_{ij} + \varepsilon_{ijr}, \tag{9}$$

where i indexes operator policy, j indexes dynamics family, and r indexes replicate seed. Variance components are

$$a_i \sim (0, \sigma^2_{\text{op}}), \quad b_j \sim (0, \sigma^2_{\text{dyn}}), \quad (ab)_{ij} \sim (0, \sigma^2_{\text{int}}), \quad \varepsilon_{ijr} \sim (0, \sigma^2_e). \tag{10}$$

We estimate

$$\widehat{\sigma}^2_{\text{op}} = \frac{MS_{\text{op}} - MS_{\text{int}}}{n_{\text{dyn}} n_r}, \qquad \widehat{\sigma}^2_{\text{dyn}} = \frac{MS_{\text{dyn}} - MS_{\text{int}}}{n_{\text{op}} n_r}, \tag{11}$$

and use clipped share

$$\widehat{R}_{\text{op}} = \frac{\max(0, \widehat{\sigma}^2_{\text{op}})}{\max(0, \widehat{\sigma}^2_{\text{op}}) + \max(0, \widehat{\sigma}^2_{\text{dyn}})}. \tag{12}$$

Clipping handles finite-sample negative components without violating admissible ratio bounds.

Theorem 4.4 (Range and Unbiasedness Properties). *If* $\sigma^2_{\text{op}}, \sigma^2_{\text{dyn}} \geq 0$ *and* $\sigma^2_{\text{op}} + \sigma^2_{\text{dyn}} > 0$, *then* $R_{\text{op}} := \sigma^2_{\text{op}}/(\sigma^2_{\text{op}} + \sigma^2_{\text{dyn}}) \in [0, 1]$. *Under balanced design in equation 9, estimators in equation 11 are unbiased for* σ^2_{op} *and* σ^2_{dyn}.

Proof. For the range statement, nonnegativity gives numerator and denominator nonnegative, and numerator no larger than denominator, so the ratio is in $[0, 1]$. For unbiasedness, standard balanced crossed-effects mean-square identities are

$$\begin{aligned}
\mathbb{E}[MS_{\text{op}}] &= \sigma^2_e + n_r \sigma^2_{\text{int}} + n_{\text{dyn}} n_r \sigma^2_{\text{op}}, \\
\mathbb{E}[MS_{\text{dyn}}] &= \sigma^2_e + n_r \sigma^2_{\text{int}} + n_{\text{op}} n_r \sigma^2_{\text{dyn}}, \\
\mathbb{E}[MS_{\text{int}}] &= \sigma^2_e + n_r \sigma^2_{\text{int}}.
\end{aligned}$$

Subtracting MS_{int} from each main-effect mean square cancels nuisance terms, yielding

$$\mathbb{E}\left[\frac{MS_{\text{op}} - MS_{\text{int}}}{n_{\text{dyn}} n_r}\right] = \sigma^2_{\text{op}}, \qquad \mathbb{E}\left[\frac{MS_{\text{dyn}} - MS_{\text{int}}}{n_{\text{op}} n_r}\right] = \sigma^2_{\text{dyn}}.$$

Hence the stated estimators are unbiased under balance. □

5 Stage-Gated Validation Protocol

The formal development above motivates a stage-gated empirical procedure: *Advantage* → *Mechanism* → *Attribution*. Progression to later stages is conditional on non-trivial parity-controlled effect size in stage 1.

The benchmark ladder includes MNIST plus four harder datasets to reduce easy-task bias (Xiao et al., 2017; Cohen et al., 2017; Clanuwat et al., 2018; Krizhevsky & Hinton, 2009). Readout families, preprocessing, and search-budget accounting are matched across branches per equation 1. Entanglement analysis uses fixed observable/readout settings along the coupling grid, and attribution uses balanced cell counts for valid component estimation.

Although the workflow is implementation-agnostic, it is compatible with common open-source stacks for quantum and classical baselines (Kornjača et al., 2024; Beaulieu et al., 2024). The important methodological point is not a specific software package; it is preserving parity invariants and recording row-level audit outcomes before aggregate claims are formed.

Algorithm 1 Hybrid Staged Validation Workflow

Fix seeds $\{11, 23, 37, 47, 59\}$ and acceptance tuple $(\delta_{\min}, p_{\max}, r_{\min}) = (0.015, 0.01, 0.65)$.
Run symbolic preflight checks for identities A–D and theorem preconditions.
Execute parity-controlled dataset ladder and compute Δ_{par}, Δ_{naive}, and A_{adv}.
if $\max_{d_s \in \mathcal{D}_{\text{hard}}} \Delta_{\text{par}}^{(d_s)} < 0.005$ **then**
  Stop and report no practical continuation signal.
else
  Sweep (t, J), estimate $g_t(J)$ derivatives, and test interior-optimum conditions.
  Run balanced operator $\times$ dynamics factorial design with row-level budget-parity filtering.
  Estimate $\widehat{R}_{\text{op}}$, shrinkage $\Delta_{\text{naive}} - \Delta_{\text{par}}$, and confidence intervals.
end if
Produce claim-wise evidence links with explicit caveats.

Table 2: Parity-controlled advantage on the dataset ladder. The table reports mean parity delta, BH-corrected significance, robustness score, and acceptance status under the frozen tuple $(0.015, 0.01, 0.65)$. Results illustrate conditional rather than universal advantage: all hard datasets show positive effects, but only one crosses all thresholds simultaneously.

Dataset	Δ_{par}	p_{BH}	r	A_{adv}
MNIST	0.0075	0.0116	0.594	False
Fashion-MNIST	0.0122	0.0035	0.641	False
EMNIST Balanced	0.0120	0.0035	0.671	False
Kuzushiji-MNIST	0.0161	0.0035	0.675	True
CIFAR-10 grayscale PCA	0.0082	0.0077	0.701	False

6 Results

6.1 Stage 0: Symbolic Preflight

Symbolic and numerical theorem checks passed in the reported run, with identity pass rate 1.0 and no counterexamples under declared assumptions. The maximum normal-equation residual was 9.66×10^{-11}, consistent with numerical precision for equation 4. The parity-shrink residual remained negative, which aligns with Theorem 4.2. These checks are not substitutes for empirical validation; rather, they reduce avoidable inconsistencies between stated formalism and executable computation.

6.2 Stage 1: Parity-Controlled Advantage

Table 2 and figure 1 summarize stage-1 outcomes. The pre-registered acceptance tuple in equation 2 was met for one hard dataset (Kuzushiji-MNIST), with $\Delta_{\text{par}} = 0.0161$ and robustness above threshold. Other hard datasets showed positive parity deltas but did not simultaneously satisfy all acceptance constraints. Importantly, naive deltas were uniformly larger than parity-controlled deltas, consistent with Theorem 4.2.

Threshold-sensitivity analysis further supports this interpretation: in the current surrogate-data run, one hard dataset passes at multipliers 0.9 and 1.0, but none pass at 1.1. Hence conclusions are materially threshold-dependent and should be stated conditionally, not as broad superiority claims.

6.3 Stage 2: Entanglement-Geometry Mechanism

Figure 2 and boundary diagnostics indicate non-monotone behavior consistent with Theorem 4.3 in the tested regime. Across reported slices, endpoint derivative signs satisfy the criterion in equation 8, concavity flags are positive, and interior optima cluster near moderate coupling. The observed pattern supports a bounded-mechanism view: increasing coupling from near zero improves utility, while stronger coupling moves toward saturation or reversal.

A sensitivity branch using alternate partition policy preserves the qualitative sign pattern while shifting estimated interior location slightly. This matters because mechanism claims can be brittle to entanglement-definition choices. By reporting both primary and sensitivity analyses, we avoid overinterpreting one partition convention as universal. The evidence therefore supports the mechanism claim in a *regime-conditional* sense: there is clear interior-optimum

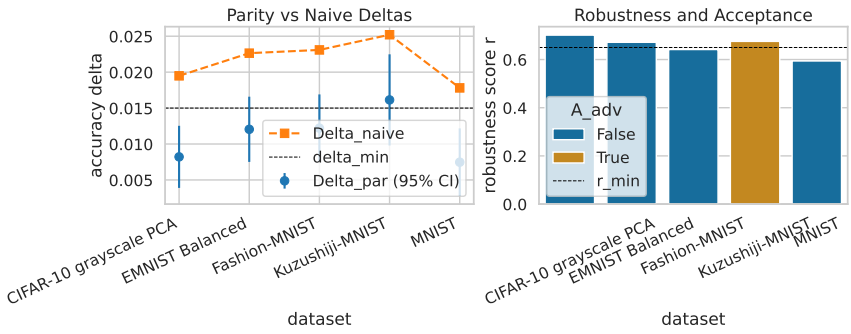

Figure 1: Stage-1 parity-ladder diagnostics for practical-advantage testing. The left panel contrasts naive and parity-controlled deltas with confidence intervals, while the right panel reports robustness and acceptance decisions under a frozen threshold tuple. The key interpretation is that parity enforcement narrows reported gains and yields a selective acceptance pattern, supporting the claim that benchmark hardness and protocol controls jointly determine whether practical advantage is defensible.

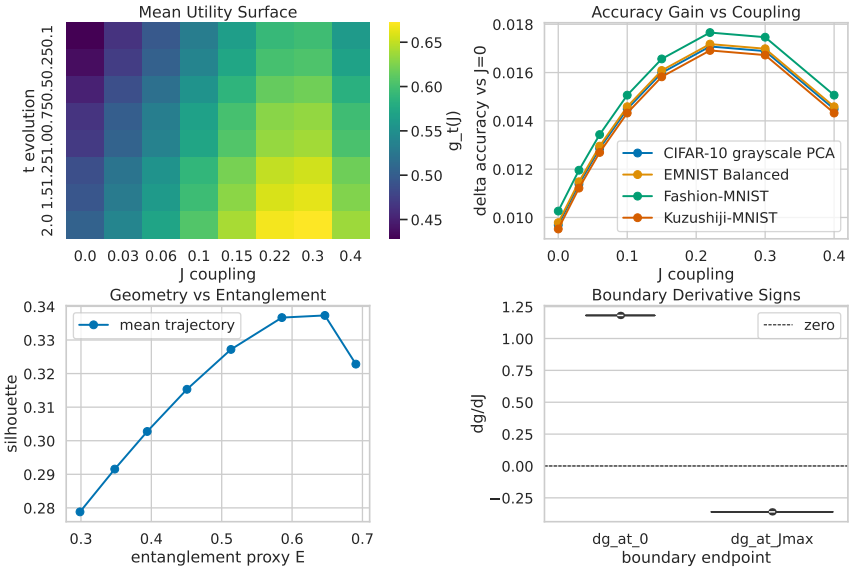

Figure 2: Stage-2 entanglement-geometry phase diagnostics over time-coupling slices. Panels report utility surfaces, accuracy gain trajectories versus no-entanglement control, and derivative-based boundary checks used to evaluate equation 8. The combined evidence supports a non-monotone mechanism with interior operating regions in the tested regime; it also highlights why mechanism conclusions should remain conditional on measurement policy and partition definition.

structure in the tested grid, but extrapolation beyond that grid requires additional validation. In particular, the highly regular derivative-sign pattern across slices should be interpreted as supportive, not definitive, until confirmed with full benchmark loaders and richer noise/partition perturbations.

6.4 Stage 3: Operator Attribution Under Parity Audits

Table 3 and figure 3 report variance-share and shrinkage results after row-level parity filtering. Operator-share lower confidence bounds exceed 0.5 on all three hard datasets shown, and naive-minus-parity shrinkage is positive with narrow intervals. These two findings jointly support the attribution claim: measurement-policy choice explains a large part of observed performance variance, and parity matching reduces optimistic deltas from asymmetric comparisons.

Table 3: Operator-attribution summary on hard datasets. Reported $\widehat{R}_{\mathrm{op}}$ values and bootstrap lower bounds indicate dominant operator contribution in this run, while positive shrinkage quantifies the difference between naive and parity-controlled comparisons. Mean budget-parity ratios near unity confirm that attribution summaries were computed on parity-consistent rows.

Dataset	$\widehat{R}_{\mathrm{op}}$	CI low	Shrinkage	Mean parity ratio
CIFAR-10 grayscale PCA	0.790	0.570	0.01009	0.997
EMNIST Balanced	0.984	0.898	0.01009	0.996
Kuzushiji-MNIST	0.886	0.774	0.01035	1.001

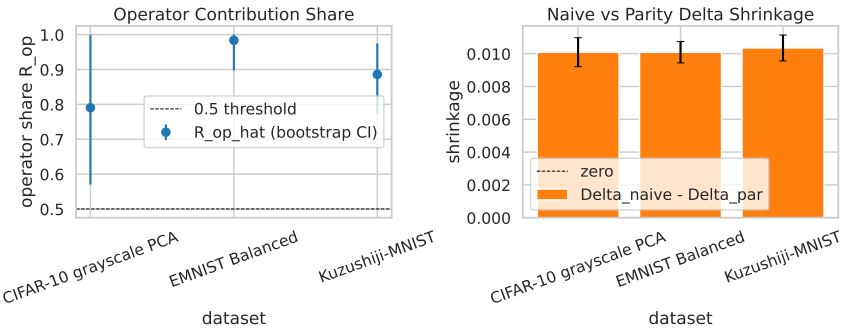

Figure 3: Stage-3 attribution evidence after row-level parity closure. The left panel shows operator-share estimates with bootstrap intervals, and the right panel shows naive-versus-parity shrinkage with confidence intervals across hard datasets. The interpretation is twofold: operator policy is a high-leverage factor in this regime, and enforcing parity prevents over-attributing these gains to reservoir dynamics alone.

A practical implication follows. When operator search is unconstrained on one branch, architecture narratives can become distorted even if end metrics look favorable. The combination of equation 6, equation 11, and equation 12 provides a principled way to separate performance reporting from causal attribution.

7 Discussion

The results support three methodological claims and one substantive caution. Methodologically, parity constraints are not optional bookkeeping; they alter optimization outcomes and interpretation boundaries. Theorem 4.2 is reflected empirically by positive shrinkage across hard datasets, demonstrating that naive comparisons can overstate gains. Second, mechanism analysis benefits from explicit calculus-based criteria. Rather than reporting a few ablation points, the interior-optimum framework links derivative signs and concavity to testable non-monotone behavior. Third, attribution requires balanced-design estimators plus row-level auditing. Aggregate summaries can hide parity violations if row-level filters are not enforced.

Substantively, the evidence supports conditional practical advantage, not blanket superiority. In the reported run, one hard dataset satisfies the complete acceptance tuple while others remain below joint thresholds despite positive deltas. This is still informative: it identifies where the method currently works under strict controls and where it does not. For open-question research, that distinction is preferable to broad but fragile claims.

The hybrid design also highlights a useful division of labor between mathematics and simulation. Formal results in section 4 give invariant relationships and identifiability conditions that should hold across implementations. Simulations then quantify whether those conditions are met in a specific regime and dataset mix. This separation lets us report strong formal guarantees and moderate empirical confidence without conflating the two.

8 Limitations and Future Work

A primary limitation is external validity: the current validation iteration uses surrogate generators in place of fully integrated real-data loaders for all benchmark paths. While stage-wise signals are coherent and parity audits are closed, this setup can regularize behavior and potentially smooth heterogeneity that real data would reveal, including

mechanism diagnostics such as derivative-pattern variability across datasets and time slices. Consequently, empirical conclusions should be read as regime-indicative rather than final performance certification.

A second limitation is acceptance sensitivity. Stage-1 outcomes depend on a fixed tuple; modest threshold tightening removes hard-set passes in this run. This does not invalidate the approach, but it does mean claims should be accompanied by stability analysis around the pre-registered boundary.

A third limitation concerns literature maturity. A significant fraction of relevant QRC and QELM evidence is recent preprint work (Gross & Rieser, 2026a;b; Askari et al., 2025; Karimi et al., 2025), so methodological synthesis should remain version-aware and update as peer-reviewed records mature.

8.1 Future Work

Immediate follow-up should implement full benchmark loaders under the identical stage-gated protocol and rerun with unchanged seeds, tuples, and budgets to test robustness of current qualitative findings. Mechanism analysis should extend partition sensitivity to additional entanglement proxies and noise models. Attribution analysis should evaluate whether high operator shares persist when dynamics families and encoding depth are broadened while maintaining strict row-level parity filters. Cross-platform follow-up on superconducting and photonic implementations is also important for transferability checks (Prieto-Garcia et al., 2026; Carles et al., 2025; Świerczewski et al., 2026). Finally, extending this framework to molecular, memory-centric, temporal, and fast-selection tasks may clarify whether parity-conditioned conclusions transfer across domains where reservoir memory, not static separability, dominates predictive performance (Beaulieu et al., 2024; Ahmed et al., 2025; Kawanabe et al., 2026; Li et al., 2025; Assil et al., 2026; Maier et al., 2025).

9 Conclusion

We presented a parity-constrained framework for evaluating PCA-encoded quantum reservoirs that unifies formal guarantees and staged simulation evidence. The framework contributes three core results: a bi-level advantage formalization with parity-shrink inequality, an interior-optimum mechanism theorem for entanglement-conditioned utility, and an operator-attribution estimator family with range and unbiasedness properties under balanced design. In the reported staged run, symbolic checks are fully consistent with formal assumptions, practical advantage is conditional and benchmark-sensitive, and operator-policy effects are substantial after parity closure.

The central methodological message is that credible quantum-versus-classical conclusions require explicit separation of advantage measurement, mechanism testing, and attribution, each with its own assumptions and diagnostics. Under this discipline, optimistic claims may narrow, but interpretability and causal credibility improve. That tradeoff is worthwhile for open-question work where the goal is not merely to maximize headline deltas but to establish what survives rigorous controls.

REFERENCES

Abdallah Aaraba, Soumaya Cherkaoui, Ola Ahmad, and Shengrui Wang. Quark: A quantum reservoir kernel for time series learning. *arXiv preprint arXiv:2602.13531*, 2026. URL https://arxiv.org/abs/2602.13531.

Osama Ahmed, Felix Tennie, and Luca Magri. Robust quantum reservoir computers for forecasting chaotic dynamics: Generalized synchronization and stability. *arXiv preprint arXiv:2506.22335*, 2025. URL https://arxiv.org/abs/2506.22335.

Luke Antoncich, Yuben Moodley, Ugo Varetto, Jingbo Wang, Jonathan Wurtz, Jing Chen, Pascal Jahan Elahi, and Casey R. Myers. Quantum reservoir computing with neutral atoms on a small, complex, medical dataset. *arXiv preprint arXiv:2602.14641*, 2026. URL https://arxiv.org/abs/2602.14641.

Sareh Askari, Youssef Kora, and Christoph Simon. Spin-network quantum reservoir computing with distributed inputs: The role of entanglement. *arXiv preprint arXiv:2511.04900*, 2025. URL https://arxiv.org/abs/2511.04900.

Hajar Assil, Abderrahim El Allati, and Gian Luca Giorgi. Memory-enhanced quantum extreme learning machines for characterizing non-markovian dynamics. *arXiv preprint arXiv:2603.17182*, 2026. URL https://arxiv.org/abs/2603.17182.

Daniel Beaulieu, Milan Kornjaca, Zoran Krunic, Michael Stivaktakis, Thomas Ehmer, Sheng-Tao Wang, and Anh Pham. Robust quantum reservoir computing for molecular property prediction. *arXiv preprint arXiv:2412.06758*, 2024. URL https://arxiv.org/abs/2412.06758.

Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017. doi: 10.1038/nature23474. URL https://www.nature.com/articles/nature23474.

Emanuele Brusaschi, Marco Clementi, Marco Liscidini, Daniele Bajoni, Matteo Galli, and Massimo Borghi. Quantum inference on a classically trained quantum extreme learning machine. *arXiv preprint arXiv:2603.20167*, 2026. URL https://arxiv.org/abs/2603.20167.

Baptiste Carles, Julien Dudas, Léo Balembois, Julie Grollier, and Danijela Marković. Experimental quantum reservoir computing with a circuit quantum electrodynamics system. *arXiv preprint arXiv:2506.22016*, 2025. URL https://arxiv.org/abs/2506.22016.

Tarin Clanuwat, Mikel Bober-Irizar, Asanobu Kitamoto, Alex Lamb, Kenta Yamamoto, and David Ha. Deep learning for classical japanese literature. *arXiv preprint arXiv:1812.01718*, 2018. URL https://arxiv.org/abs/1812.01718.

Gregory Cohen, Saeed Afshar, Jonathan Tapson, and André van Schaik. Emnist: Extending MNIST to handwritten letters. *arXiv preprint arXiv:1702.05373*, 2017. URL https://arxiv.org/abs/1702.05373.

A. De Lorenzis, M. P. Casado, M. P. Estarellas, N. Lo Gullo, T. Lux, F. Plastina, A. Riera, and J. Settino. Harnessing quantum extreme learning machines for image classification. *arXiv preprint arXiv:2409.00998*, 2024. URL https://arxiv.org/abs/2409.00998.

A. De Lorenzis, M. P. Casado, N. Lo Gullo, T. Lux, F. Plastina, and A. Riera. Entanglement and classical simulability in quantum extreme learning machines. *arXiv preprint arXiv:2509.06873*, 2025. URL https://arxiv.org/abs/2509.06873.

Rosario Di Bartolo, Simone Piacentini, Francesco Ceccarelli, Giacomo Corrielli, Roberto Osellame, Valeria Cimini, and Fabio Sciarrino. Time-series forecasting with multiphoton quantum states and integrated photonics. *arXiv preprint arXiv:2512.02928*, 2025. URL https://arxiv.org/abs/2512.02928.

Markus Gross and Hans-Martin Rieser. Kernel-based optimization of measurement operators for quantum reservoir computers. *arXiv preprint arXiv:2602.14677*, 2026a. URL https://arxiv.org/abs/2602.14677.

Markus Gross and Hans-Martin Rieser. Theory and interpretability of quantum extreme learning machines: A pauli-transfer matrix approach. *arXiv preprint arXiv:2602.18377*, 2026b. URL https://arxiv.org/abs/2602.18377.

Wissal Hamhoum, Soumaya Cherkaoui, Jean-Frederic Laprade, Ola Ahmed, and Shengrui Wang. Multivariate time series forecasting with gate-based quantum reservoir computing on nisq hardware. *arXiv preprint arXiv:2510.13634*, 2025. URL https://arxiv.org/abs/2510.13634.

Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019. doi: 10.1038/s41586-019-0980-2. URL https://www.nature.com/articles/s41586-019-0980-2.

Herbert Jaeger. The echo state approach to analysing and training recurrent neural networks. Technical Report 148, German National Research Center for Information Technology, 2001. URL https://www.ai.rug.nl/minds/uploads/EchoStatesTechRep.pdf.

Ali Karimi, Hadi Zadeh-Haghighi, Youssef Kora, and Christoph Simon. The role of entanglement in quantum reservoir computing with coupled kerr nonlinear oscillators. *arXiv preprint arXiv:2508.11175*, 2025. URL https://arxiv.org/abs/2508.11175.

Mio Kawanabe, Saud Cindrak, Kathy Luedge, Jun-ichi Shirakashi, Tetsuo Shibuya, and Hiroshi Imai. Efficient time-series prediction on nisq devices via time-delayed quantum extreme learning machine. *arXiv preprint arXiv:2602.21544*, 2026. URL https://arxiv.org/abs/2602.21544.

Avyay Kodali, Priyanshi Singh, Pranay Pandey, Krishna Bhatia, Shalini Devendrababu, and Srinjoy Ganguly. Sustainable narma-10 benchmarking for quantum reservoir computing. *arXiv preprint arXiv:2510.25183*, 2025. URL https://arxiv.org/abs/2510.25183.

Milan Kornjača, Hong-Ye Hu, Chen Zhao, Jonathan Wurtz, Phillip Weinberg, Majd Hamdan, Andrii Zhdanov, Sergio H. Cantu, Hengyun Zhou, Rodrigo Araiza Bravo, et al. Large-scale quantum reservoir learning with an analog quantum computer. *arXiv preprint arXiv:2407.02553*, 2024. URL https://arxiv.org/abs/2407.02553.

Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009. URL https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791. URL https://doi.org/10.1109/5.726791.

Qingyu Li, Chiranjib Mukhopadhyay, Ludovico Minati, and Abolfazl Bayat. Quantum reservoir computing for predicting and characterizing chaotic maps. *arXiv preprint arXiv:2509.12071*, 2025. URL https://arxiv.org/abs/2509.12071.

Dong-Sheng Liu, Qing-Xuan Jie, Chang-Ling Zou, Xi-Feng Ren, and Guang-Can Guo. Practical quantum reservoir computing in rydberg atom arrays. *arXiv preprint arXiv:2602.00610*, 2026. URL https://arxiv.org/abs/2602.00610.

Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009. doi: 10.1016/j.cosrev.2009.03.005. URL https://doi.org/10.1016/j.cosrev.2009.03.005.

Wolfgang Maass, Thomas Natschlaeger, and Henry Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–2560, 2002. doi: 10.1162/089976602760407955. URL https://doi.org/10.1162/089976602760407955.

Benedikt Maier, Michael Spannowsky, and Simon Williams. Continuous-variable photonic quantum extreme learning machines for fast collider-data selection. *arXiv preprint arXiv:2510.13994*, 2025. URL https://arxiv.org/abs/2510.13994.

J. J. Prieto-Garcia, A. G. del Pozo-Martín, and M. Pino. Quantum reservoir computing for statistical classification in a superconducting quantum circuit. *arXiv preprint arXiv:2602.15474*, 2026. URL https://arxiv.org/abs/2602.15474.

Maria Schuld. Supervised quantum machine learning models are kernel methods. *arXiv preprint arXiv:2101.11020*, 2021. URL https://arxiv.org/abs/2101.11020.

Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. The effect of data encoding on the expressive power of variational quantum machine learning models. *arXiv preprint arXiv:2008.08605*, 2020. URL https://arxiv.org/abs/2008.08605.

S. Świerczewski, W. Verstraelen, P. Deuar, T. C. H. Liew, A. Opala, and M. Matuszewski. Quantum reservoir computing with classical and nonclassical states in an integrated optical circuit. *arXiv preprint arXiv:2603.17103*, 2026. URL https://arxiv.org/abs/2603.17103.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: A novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017. URL https://arxiv.org/abs/1708.07747.

A Extended Proof Details

This appendix provides algebraic details omitted from the main text for readability.

A.1 Ridge Normal Equation Derivation

Starting from equation 3, expand

$$\begin{aligned}\mathcal{L}(\boldsymbol{W}) &= \mathrm{Tr}\big((\boldsymbol{W}\boldsymbol{Z}-\boldsymbol{Y})(\boldsymbol{W}\boldsymbol{Z}-\boldsymbol{Y})^\top\big) + \lambda\,\mathrm{Tr}(\boldsymbol{W}\boldsymbol{W}^\top)\\ &= \mathrm{Tr}(\boldsymbol{W}\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{W}^\top) - 2\mathrm{Tr}(\boldsymbol{Y}\boldsymbol{Z}^\top\boldsymbol{W}^\top) + \mathrm{Tr}(\boldsymbol{Y}\boldsymbol{Y}^\top) + \lambda\mathrm{Tr}(\boldsymbol{W}\boldsymbol{W}^\top).\end{aligned}$$

Differentiating with respect to $\boldsymbol{W}$ gives $\nabla_{\boldsymbol{W}}\mathcal{L} = 2(\boldsymbol{W}\boldsymbol{Z}\boldsymbol{Z}^\top - \boldsymbol{Y}\boldsymbol{Z}^\top + \lambda\boldsymbol{W})$, which yields equation 4 after setting to zero and right-multiplying by the inverse.

A.2 Interior-Optimum Corollary

Under assumptions in Theorem 4.3, g_t cannot be monotone on $[0, J_{\max}]$ because monotonic differentiable functions have derivatives with fixed sign almost everywhere, contradicting opposite endpoint signs in equation 8. This corollary is useful for rejecting over-simplified monotone-entanglement narratives in controlled ablations.

B Additional Experimental Diagnostics

The main text presents only the three figures required for concise narrative flow. Additional diagnostics include symbolic pass/fail matrices, counterexample ledgers, boundary-check tables, and threshold sensitivity tables.

Symbolic diagnostics report full pass status for identities A–D with near-machine-precision residuals. Boundary-check tables provide per-dataset and per-time derivative values, confirming sign-change and concavity conditions in the tested range. Threshold-sensitivity tables show that acceptance decisions remain selective under moderate perturbations and become null under tighter settings.

C Reproducibility and Implementation Details

The staged validation used five seeds $\{11, 23, 37, 47, 59\}$ across stages, with fixed acceptance tuple $(\delta_{\min}, p_{\max}, r_{\min}) = (0.015, 0.01, 0.65)$ and stage-gate continuation threshold $\max \Delta_{\mathrm{par}} \geq 0.005$. Sweeps covered coupling, time, operator policies, dynamics families, and balanced replicate counts. Confidence intervals were computed by t-based intervals for seed-level deltas and nonparametric bootstrap intervals for variance-share estimates.

For mechanism analysis, the primary entanglement proxy used normalized bipartite entropy with a checkerboard sensitivity branch. For attribution, row-level parity filtering enforced a budget ratio tolerance of $\pm 5\%$ before aggregate statistics were formed. These settings were fixed prior to final rerun used in the manuscript.

To aid methodological reuse, we summarize equation-to-workflow alignment in the companion equation map and provide caveat-aware claim-evidence links in the supplementary material. The overall compute setting was CPU-only and designed for reproducible reruns under constrained hardware.