# Quantum Feature Accessibility as a Mechanism for Shortcut Learning under Distribution Shift

**Anonymous authors**
Paper under double-blind review

## Abstract

Which features do quantum machine learning (QML) models learn when stable and shortcut signals compete? We show that statistical predictiveness alone does not explain feature reliance: quantum processing makes some predictive signals more accessible than others. We develop a learning-theoretic account of quantum feature accessibility and prove, under explicit assumptions, that population-risk minimizers of local quantum models with affine heads rely only on a competing cue, even when joint features encode label information. Controlled interventions and encoding ablations reveal how the same processing bias can favor stable signals or shortcuts. Experiments on realistic shifts show task-dependent generalization, while MonarQ evaluations reveal hardware-dependent changes in stable-signal prediction. Together, these results link quantum processing to learned reliance and show why out-of-distribution (OOD) generalization depends on the stability of the signals used.

## 1 Introduction

Robustness to distribution shift is essential for the reliable deployment of machine learning systems (Koh et al., 2021). Deep learning has an extensive literature on out-of-distribution (OOD) generalization, including shortcut-learning analyses and systematic benchmarks (Geirhos et al., 2020; Gulrajani & Lopez-Paz, 2021; Lynch et al., 2023). A recurring challenge is reliance on shortcuts whose predictive value changes across environments.

In quantum machine learning (QML), encoding, circuit structure, and measurement constrain feature use. Information encoded in a quantum state need not reach the readout or influence training. Hardware noise can suppress measurable distinctions and training gradients (Wang et al., 2021), limiting usable predictive information. Signals visible in ideal circuits may therefore become difficult to resolve with finite measurement resources.

Prior work characterizes encoding-dependent expressivity (Schuld et al., 2021), quantum inductive and spectral biases (Kübler et al., 2021; Duffy & Jastrzebski, 2025), and finite-data generalization (Caro et al., 2022). Complementing OOD guarantees for quantum dynamics (Caro et al., 2023), data-dependent generalization analysis (Haug & Kim, 2024), and test-time adaptation to distribution shifts and circuit noise (Jian et al., 2024), we ask: *how do quantum processing restrictions shape reliance on competing stable and shortcut signals, and generalization when feature–label correlations in classical data change?*

This work develops a learning-theoretic and empirical account of **quantum feature accessibility**, defined as the extent to which a predictive signal can be used under specified quantum processing and learning resources. Our goal is to explain learned reliance and its consequences under distribution shift. Our theory distinguishes encoded information, measurement visibility, and learned reliance. Under explicit conditions, we prove that population-risk minimizers of local quantum models with affine heads are cue-only, and bound near-optimal deviations under shift. Supporting analyses address observable restrictions, early reliance growth, and finite-shot reliability.



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Our systematic experiments vary feature competition and quantum processing, using interventions to probe reliance that accuracy alone cannot reveal. Realistic distribution shifts examine the account's broader relevance. Our findings show why statistical predictiveness alone cannot explain reliance, how circuit structure changes access to stable signals, and when the same quantum bias favors stable-feature learning or shortcut reliance. Controlled hardware evaluations further illustrate why a smaller ID–OOD gap need not indicate improved OOD performance.

## 2 Problem Formulation and Preliminaries

**Learning setting.** Environment $e$ defines an input–label distribution $P_e(X, Y)$. Training pools $m$ environments (Arjovsky et al., 2019) as $P_{\text{tr}} = \sum_{e=1}^{m} \alpha_e P_e$, where $\alpha_e \geq 0$ and $\sum_e \alpha_e = 1$ specify training proportions. Environment labels need not be observed. Testing uses a potentially different distribution $P_{\text{te}}$. Fixed preprocessing is implicit in $f$ and shared across environments. For loss $\ell$, population risk is $R(f; P) = \mathbb{E}_{(X,Y)\sim P}[\ell(f(X), Y)]$ . Training uses empirical risk under $P_{\text{tr}}$; OOD performance is assessed through $R(f; P_{\text{te}})$. Stable features retain their predictive label relationship across the environments considered; shortcuts are predictive during training but this relationship changes under test shift (Geirhos et al., 2020). Each theoretical result states its distributional assumptions.

**Quantum predictor.** A circuit with parameters $\theta$ produces an $n$-qubit state with density matrix $\rho_\theta(x)$, encoding data once or repeatedly between trainable layers (Schuld et al., 2021). Its expected measurement outcomes and prediction are (Cerezo et al., 2021)

$$z_{\theta,k}(x) = \text{Tr}[O_k \rho_\theta(x)] \,, \qquad f_{\theta,w}(x) = h_w(\mathbf{z}_\theta(x)) \,.$$

Here $\mathbf{z}_\theta$ collects expectations of Hermitian observables $O_1, \ldots, O_K$, and $h_w$ is a classical prediction head with parameters $w$. Both observables and head are specified for each model.

**Measurements and feature reliance.** Physical noise affects states or readout; finite shots introduce sampling uncertainty (Cerezo et al., 2021). Finite-shot predictions use $\widehat{\mathbf{z}}_\theta(x)$; risk averages over measurement randomness. We specify circuit, training, and shot budgets and assess reliance through prediction or loss changes under controlled feature perturbations (Fisher et al., 2019), documenting fixed components and label treatment.

## 3 Quantum Feature Accessibility: A Theoretical Analysis

Quantum feature accessibility can favor stable signals or shortcuts; OOD effects depend on signal stability. We study readout restrictions, cue-only population optima in local circuits with affine heads, early learning, and measurement resources. Proofs appear in Appendix A.

### 3.1 From encoded information to accessible predictors

We first ask what the readout and head can recover. One upload followed by an input-independent channel $\mathcal{E}_\theta$ gives (Schuld et al., 2021; Gratsea & Huembeli, 2023)

$$z_{\theta,k}(x) = \text{Tr}[\widetilde{O}_{\theta,k} \rho_{\text{enc}}(x)], \qquad \widetilde{O}_{\theta,k} = \mathcal{E}_\theta^\dagger(O_k). \tag{1}$$

**Proposition 1** (Readout restrictions)**.** *For $\mathcal{V} = \text{span}_{\mathbb{R}}\{\widetilde{O}_{\theta,k} : \theta, k\}$ and $\Delta\rho = \rho_{\text{enc}}(x) - \rho_{\text{enc}}(x')$, the readouts agree for all $\theta$ iff $\text{Tr}(A\Delta\rho) = 0$ for all $A \in \mathcal{V}$. Any head using these expectations then predicts identically.* (Proof: App. A.2.)

**Proposition 2** (Distinguishability bounds, restated from Watrous, 2018)**.** *For states $\rho, \sigma$ and Hermitian $O$, $|\,\text{Tr}[O(\rho-\sigma)]| \leq 2\|O\|_\infty D(\rho,\sigma)$, where $D(\rho,\sigma) = \frac{1}{2}\|\rho-\sigma\|_1$. Applying the same quantum channel to states cannot increase $D$; unitaries preserve it.* (Proof: App. A.2.)

Distinguishability need not ensure recovery from a restricted readout. For fixed diagnostic distribution $Q$, affine recoverability $\mathcal{R}^2_\phi(\theta; Q)$ equals 1 minus minimum affine MSE for recovering $\phi(X)$ from $\mathbf{z}_\theta(X)$, normalized by the target's positive, finite variance under $Q$ (App. A.8). Recoverability fixes $\theta$; training selects it.

### 3.2 A population-risk mechanism for feature competition

We next connect representational restrictions to population feature reliance.

**Proposition 3** (Local-model additivity). *Product encoding, one coordinate and local readout per qubit, input-independent product channels, and an affine head give* $f(x) = b + \sum_j w_j z_{\theta,j}(x_j)$ (2). *Local layers and linear grouping preserve additivity.* (Proof: App. A.3.)

Write $X = (X_r, X_c)$ for relational inputs and a candidate cue. If $X_{r,j} \perp (Y, X_c)$ for every $j$, additive $a(X_r)$ has constant conditional mean. Yet $X_r = (U, YU)$ with fair $U \perp (Y, X_c)$ encodes $Y$ in their product. Removable relational weights and a free bias give closure.

**Theorem 4** (Cue-only population optima). *For* $Y = \pm 1$ *and* $\ell(t, y) = \log(1 + e^{-yt})$, *let* $\mathcal{F}$ *consist of logits* $f = a(X_r) + g(X_c)$ *with integrable* $f, a$ *under* $P_{\mathrm{tr}}$. *For every admissible decomposition assume constant* $\mu_a = \mathbb{E}_{P_{\mathrm{tr}}}[a \mid Y, X_c]$ *and* $\bar{f} = \mu_a + g \in \mathcal{F}$. *Then*

$$R(\bar{f}; P_{\mathrm{tr}}) \le R(f; P_{\mathrm{tr}}), \tag{3}$$

*with equality iff* $f = \bar{f}$ *a.s.; every attained finite minimizer is cue-only.* (Proof: App. A.4.)

**Proposition 5** (Near-optimal logits). *Under Theorem 4, let* $R_* = \inf_{u \in \mathcal{F}} R(u; P_{\mathrm{tr}})$. *If* $|f| \le B < \infty$ *a.s. and* $R(f; P_{\mathrm{tr}}) - R_* \le \varepsilon$, *then* $\mathbb{E}_{P_{\mathrm{tr}}}(f - \bar{f})^2 \le 2\varepsilon/\kappa_B$ (4), *where* $\kappa_B = e^B/(1 + e^B)^2$. (Proof: App. A.4.)

**Proposition 6** (OOD consequences). *If* $f = \bar{f}(X_c)$ $P^X_{\mathrm{tr}}$*-a.s. and* $P^X_{\mathrm{te}} \ll P^X_{\mathrm{tr}}$, *equality holds at test; a fixed joint cue–label law preserves risk. Under Proposition 5 with* $dP^X_{\mathrm{te}}/dP^X_{\mathrm{tr}} \le M < \infty$, $|R(f; P_{\mathrm{te}}) - R(\bar{f}; P_{\mathrm{te}})| \le \sqrt{2M\varepsilon/\kappa_B}$ (5). (Proof: App. A.5.)

Local additivity gives $\mathcal{R}^2_Y = 0$ with basis encoding and independent fair $U, Y, X_c$ under $Q$; CNOT target-$Z$ readout permits $\mathcal{R}^2_Y = 1$ (App. A.4). Proposition 6 transfers cue reliance and possible cue–label-shift failures; useful bounds require controlled $B, M, \varepsilon$.

### 3.3 Learning dynamics and feature reliance

For competing usable signals, fix before training comparably scaled pair laws $\nu_j$ changing only feature $j$, with predictive orientations $s_j(x, x') \in \{-1, +1\}$. For $\lambda = (\theta, w)$ set $q_j(\lambda) = \mathbb{E}_{\nu_j}[s_j(f_\lambda(x) - f_\lambda(x'))]$, $d_j = \nabla q_j(\lambda_0)$, and $g_0 = -\nabla_\lambda R(f_\lambda; P_{\mathrm{tr}})|_{\lambda_0}$.

**Proposition 7** (Early reliance growth; standard calculus). *For differentiable* $q_j$, *gradient flow gives* $\partial_t q_j(\lambda_t)|_{t=0} = d_j^\top g_0$ (6). *For* $\lambda_1 = \lambda_0 + \eta g_0$, $\eta > 0$, $q_j \in C^2$, *and* $\|\nabla^2 q_j\|_{\mathrm{op}} \le L_j$ *along the step,* $|q_j(\lambda_1) - q_j(\lambda_0) - \eta d_j^\top g_0| \le \frac{1}{2}\eta^2 L_j \|g_0\|_2^2$ (7). (Proof: App. A.6.)

Alignment predicts initial changes only; test with the actual optimizer direction.

### 3.4 Noise and measurement reliability

Finite shots can alter decisions after learning.

**Proposition 8** (Finite-shot reliability, via Hoeffding, 1963). *For a fixed affine logit, average* $N$ *independent shots per Pauli readout, allowing dependence among the* $K$ *readouts. For* $m = |f(x)| > 0$, $W = \|w\|_1 > 0$, $\Pr(\operatorname{sign} \widehat{f} \ne \operatorname{sign} f) \le \min\{1, 2Ke^{-Nm^2/(2W^2)}\}$ (8). $f$ *uses the measured circuit's expectations;* $W = 0$ *gives* $\widehat{f} = f$. (Proof: App. A.7.)

With circuit and head fixed, varying shots probes sampling through $m/W$; comparing ideal and noisy expectations isolates physical noise.

## 4 Controlled Analysis of Feature Competition

Accuracy alone does not reveal which features a model uses. We introduce a controlled suite with known stable and shortcut signals to examine how the accessibility mechanisms in Section 3 shape feature reliance and OOD behavior.

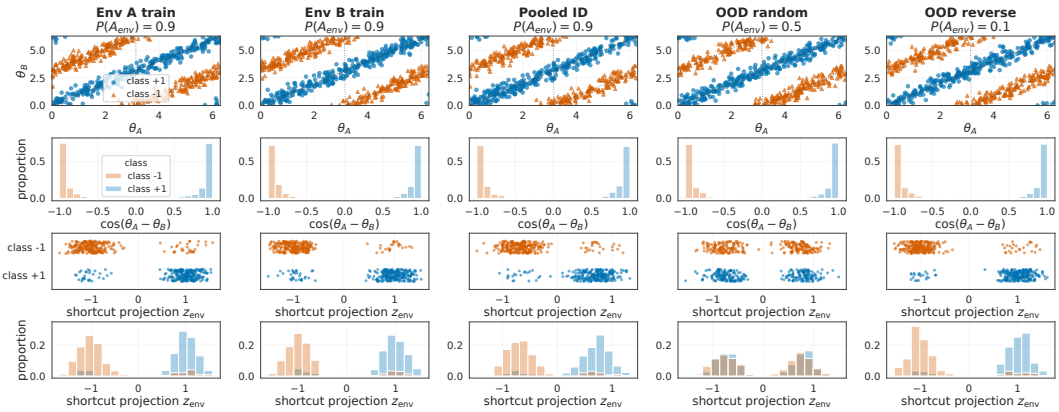


Figure 1: **Q-LINK across environments.** Rows: joint phases $(\theta_A, \theta_B)$; derived $\cos(\theta_A - \theta_B)$ distributions; local shortcut projections (class-wise samples and histograms).

The five tasks span geometric, spectral, and relational settings, with prescribed changes in shortcut–label associations. For geometric competition, **Q-SPHERE** pairs a nonlinear spherical decision rule with a directional shortcut. **Spirals-Signatures** adapts classical two-spiral geometry and established spiral-shortcut designs (Lang & Witbrock, 1988; Parascandolo et al., 2021) with a noisy directional cue and controlled shortcut–label agreement. Both examine competition between distributed geometry and simple directional cues. For spectral competition, our controlled **Spurious-Fourier** adaptation of WOODS (Gagnon-Audet et al., 2023) uses selected Fourier magnitudes and our shift protocol, preserving competition between stable high-frequency components and low-frequency shortcuts. Our **Q-FOS** construction combines stable, shortcut, and nuisance bands in full multipeak spectra. These settings probe how spectral representation and nuisance structure shape stable-signal accessibility. Finally, **Q-LINK** pairs a continuous joint phase signal with a local shortcut, complementing XOR-based benchmarks (Qiu et al., 2024). Classes concentrate near aligned versus opposite phases, while individual phase marginals carry no label information by construction. This targets the distinction between local and joint access in Theorem 4. Figure 1 illustrates competing signals across environments using Q-LINK. Definitions, construction details, and task visualizations appear in Appendices B.2, B.3, B.4, B.5, and B.6, respectively.

### 4.1 Feature reliance under controlled simulation

We compare logistic regression, an MLP, and 3 VQCs on 5 tasks with 3 seeds (Figure 2). All VQCs use six qubits, two layers, and affine heads; entangling models add CNOT rings, and reuploading models also repeat the encoding. Masks probe fixed predictors; shortcut-flip sensitivity measures mean absolute probability change. Architectures, rationale, preprocessing, and evaluation: Appendices C.2.1 and C.2.

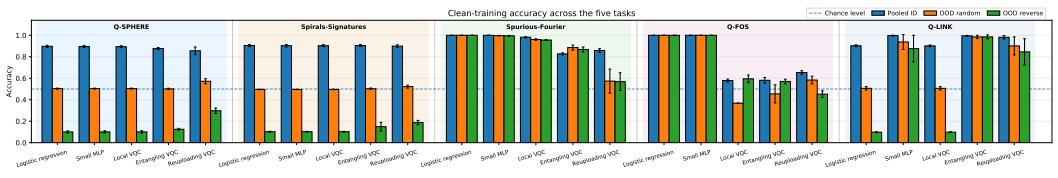


Figure 2: **Architecture rankings depend on the task.** Pooled ID overlaps training; Random and OOD reversed are held out. Appendix C.3 gives all-task values and uncertainty.

**Feature reliance across task families.** Q-SPHERE and Spirals-Signatures show high source but poor reversal accuracy. Q-SPHERE reuploading reaches $64.6 \pm 1.6\%$ on stable-only reversed inputs versus $29.7 \pm 2.6\%$ on full inputs, illustrating Section 3.1's distinction

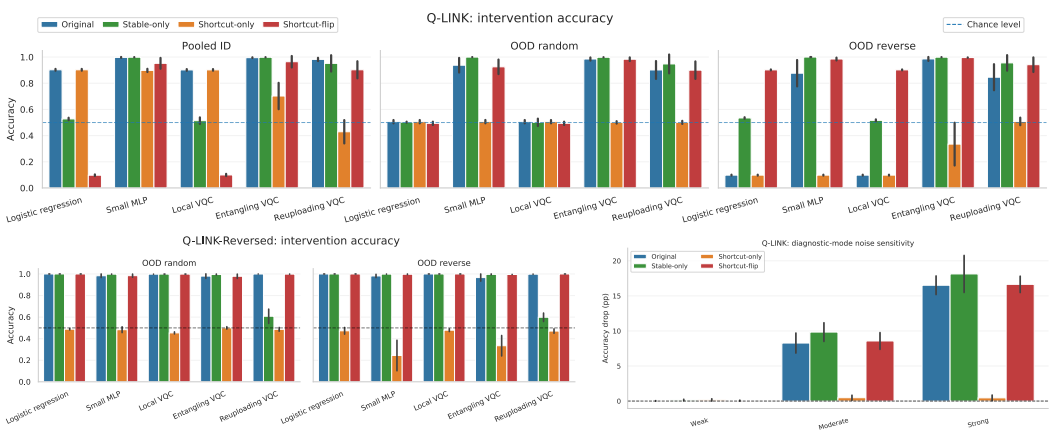


Figure 3: **Reliance, role reversal, and noise.** Top: Q-LINK interventions; bottom left: role reversal; bottom right: noise-induced accuracy drops pooled over VQCs/channels/splits. All-task details: Appendices C.4, C.8, and C.5, respectively.

between feature availability and learned reliance: useful stable-block performance can coexist with cue dependence. Under reported baseline pipelines, the local VQC leads quantum models on Spurious-Fourier, while all 3 trail classical baselines on Q-FOS, motivating joint analysis of architecture and encoding. All-task masks and shortcut perturbations appear in Appendix C.4; preprocessing and evaluation in Appendix C.2.

**Joint access and cue stability.** Local additivity (Proposition 3) and $X_{r,j} \perp (Y, X_c)$ give Q-LINK's relational terms constant conditional means. Theorem 4 makes every attained finite population optimum cue-only; under these assumptions, Proposition 5 bounds bounded predictors' deviations from cue-only logits by excess population risk. Local accuracies (stable-only source: $51.4 \pm 3.1\%$; reversal: $9.8 \pm 0.5\%$) are consistent with this mechanism; entangling counterparts ($99.8 \pm 0.2\%$, $98.6 \pm 1.9\%$) show usable joint information (Figure 3). The MLP also generalizes. Making the local cue stable in Q-LINK-Reversed yields 100% local OOD reversed accuracy (Appendix C.8). Thus the same access restriction can favor stable prediction or shortcut reliance, illustrating Proposition 6's cue-law perspective.

**Encoding changes the outcome.** For single-upload models, encoding changes can alter which state contrasts the readout detects (Equation 1; Proposition 1). Halving the $R_Y$ scale raises local and entangling Q-FOS reversal accuracy to 100%, but harms reuploading on Q-SPHERE and Spirals-Signatures; doubling improves reversal accuracy on both. Switching to $R_X$ improves Spurious-Fourier reuploading. Feature diagnostics connect performance changes to reliance: halving scale raises Q-LINK reuploading reversal accuracy from $84.6 \pm 12.2\%$ to $99.7 \pm 0.2\%$, with 100% stable-only accuracy, while shortcut-flip sensitivity falls from $0.149 \pm 0.073$ to $0.032 \pm 0.019$. These concordant diagnostics support reduced shortcut dependence with preserved stable-signal performance. Local Q-LINK remains near 10% reversal across tested encodings, consistent with Theorem 4's structural restriction. Appendix C.7 provides generator, scale, upload-order, and feature-intervention results.

**Inference noise and measurement.** Proposition 2 bounds observable contrasts by distinguishability, which common channels cannot increase. Noise can degrade successful Q-LINK and Spurious-Fourier predictors, while Q-FOS responses vary by architecture and channel (Table 1). Q-SPHERE and Spirals-Signatures can gain reversal accuracy toward chance while losing ID accuracy, suggesting weakened predictions rather than stable-signal recovery. These responses are consistent with reliance-dependent noise effects. For sampling, Proposition 8 relates decision reliability to shot budget and normalized logit margin. With 1024 shots, full-input seed means differ from analytic accuracy by at most about

2.1 percentage points across tasks, VQCs, and held-out splits, indicating limited aggregate sensitivity (Appendix C.5).

Table 1: **Inference noise after clean training.** Mean OOD reversed accuracy (%) under moderate → strong noise. Detailed results for Spirals-Signatures, Q-FOS, weak noise, and controls appear in Appendix C.5.

| VQC | Amplitude | Depolarizing | Phase |
|---|---|---|---|
| Q-SPHERE | | | |
| Local | $11.2_{\pm 2.2} \rightarrow 11.2_{\pm 2.2}$ | $11.2_{\pm 2.2} \rightarrow 11.1_{\pm 2.0}$ | $11.2_{\pm 2.2} \rightarrow 11.2_{\pm 2.2}$ |
| Entangling | $12.2_{\pm 1.7} \rightarrow 50.6_{\pm 2.2}$ | $21.6_{\pm 9.5} \rightarrow 40.0_{\pm 16.2}$ | $13.1_{\pm 2.1} \rightarrow 15.2_{\pm 2.8}$ |
| Reuploading | $33.3_{\pm 3.7} \rightarrow 51.2_{\pm 1.3}$ | $43.7_{\pm 2.7} \rightarrow 49.4_{\pm 0.7}$ | $44.7_{\pm 2.4} \rightarrow 47.8_{\pm 2.0}$ |
| Spurious-Fourier | | | |
| Local | $94.7_{\pm 1.8} \rightarrow 91.1_{\pm 2.2}$ | $95.6_{\pm 0.2} \rightarrow 95.4_{\pm 0.2}$ | $90.6_{\pm 2.8} \rightarrow 87.1_{\pm 2.1}$ |
| Entangling | $71.0_{\pm 8.7} \rightarrow 51.2_{\pm 5.9}$ | $84.6_{\pm 8.1} \rightarrow 51.7_{\pm 5.8}$ | $87.3_{\pm 2.2} \rightarrow 87.2_{\pm 2.2}$ |
| Reuploading | $55.7_{\pm 8.2} \rightarrow 51.3_{\pm 5.6}$ | $47.9_{\pm 8.8} \rightarrow 45.8_{\pm 9.2}$ | $49.1_{\pm 11.8} \rightarrow 48.7_{\pm 9.3}$ |
| Q-LINK | | | |
| Local | $7.9_{\pm 1.2} \rightarrow 8.0_{\pm 1.0}$ | $7.9_{\pm 1.2} \rightarrow 7.9_{\pm 1.2}$ | $8.0_{\pm 1.0} \rightarrow 8.7_{\pm 0.6}$ |
| Entangling | $93.9_{\pm 8.6} \rightarrow 72.2_{\pm 15.8}$ | $95.0_{\pm 1.8} \rightarrow 86.0_{\pm 4.8}$ | $83.3_{\pm 17.4} \rightarrow 69.8_{\pm 19.8}$ |
| Reuploading | $82.3_{\pm 6.5} \rightarrow 79.2_{\pm 6.7}$ | $65.4_{\pm 0.2} \rightarrow 52.1_{\pm 1.3}$ | $52.9_{\pm 2.2} \rightarrow 52.1_{\pm 1.3}$ |

**Training noise and learned reliance.** Common ideal evaluation isolates training effects. For each architecture and condition, Table 2 averages the 5 tasks equally within seed, then reports mean and SD over 3 seeds. Proposition 7 (Section 3.3) links initial reliance growth to gradient alignment. Here, stable-only accuracy and shortcut-flip sensitivity assess feature reliance after noise-aware training. Mixed training cycles amplitude, depolarizing, and phase channels, one per update, at moderate/strong settings. Strong phase-aware training lowers entangling stable-only accuracy (71.4% → 62.2%) and OOD accuracy while increasing shortcut-flip sensitivity (0.314 → 0.474), supporting changed reliance. Local task-averaged accuracy changes little; reuploading improves modestly under strong phase-aware training. Appendix C.6 gives per-task results, diagnostics, and trajectories; Appendices C.1 and C.9 map experiments to theory.

Table 2: **Noise-aware training under ideal evaluation.** Five-task mean OOD reversed accuracy (%); subscripts: ± seed SD. Per-task results: Appendix C.6.

| | Moderate | | | Strong | | |
|---|---|---|---|---|---|---|
| Training | Local | Entangling | Reuploading | Local | Entangling | Reuploading |
| Clean | $37.0_{\pm 1.1}$ | $53.7_{\pm 1.0}$ | $46.0_{\pm 3.6}$ | $37.0_{\pm 1.1}$ | $53.7_{\pm 1.0}$ | $46.0_{\pm 3.6}$ |
| Amplitude | $36.6_{\pm 1.0}$ | $54.7_{\pm 1.3}$ | $49.6_{\pm 2.9}$ | $36.3_{\pm 1.0}$ | $54.6_{\pm 1.7}$ | $48.2_{\pm 2.8}$ |
| Depolarizing | $36.9_{\pm 1.3}$ | $53.4_{\pm 0.3}$ | $43.8_{\pm 1.6}$ | $36.9_{\pm 1.2}$ | $56.0_{\pm 0.9}$ | $47.7_{\pm 3.0}$ |
| Phase | $37.0_{\pm 1.1}$ | $51.9_{\pm 1.2}$ | $47.8_{\pm 3.5}$ | $36.9_{\pm 1.1}$ | $42.0_{\pm 4.6}$ | $49.4_{\pm 2.0}$ |
| Mixed | $36.7_{\pm 1.1}$ | $54.6_{\pm 1.5}$ | $49.6_{\pm 2.3}$ | $36.8_{\pm 1.2}$ | $49.0_{\pm 7.8}$ | $47.5_{\pm 2.3}$ |

### 4.2 Feature reliance on MonarQ

We evaluate the transfer of simulation-trained checkpoints to MonarQ, which uses superconducting quantum computing technology (Calcul Québec, 2024). We repeat the corresponding VQC evaluations from Section 4.1, comparing hardware execution with analytic simulation (noiseless expectations) and finite-shot simulation (sampled noiseless measurements). This subsection highlights the key findings; detailed task-wise results appear in Appendix D, with the evaluation protocol and validation in Appendices D.1 and D.6.

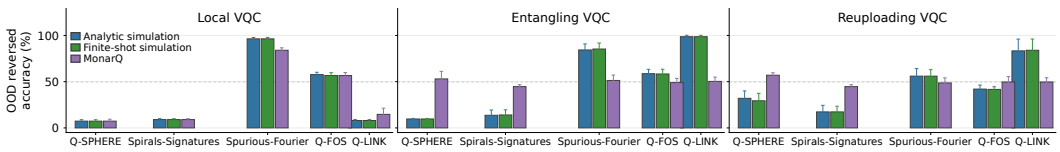


Figure 4: **Clean-trained checkpoint transfer.** OOD reversed accuracy (%), mean ± seed SD. Split/intervention details: Appendix D.3; prediction agreement: Appendix D.4.

Table 3: **Noise-aware VQCs on MonarQ.** OOD reversed accuracy (%) after noise-aware simulation training; means over three seeds, with SD in subscripts. Clean baselines are listed as Local / Entangling / Reuploading. Other tasks/controls: Appendix D.2; paired changes: Appendix D.5.

| | Moderate | | | Strong | | |
|---|---|---|---|---|---|---|
| Training | Local | Entangling | Reuploading | Local | Entangling | Reuploading |
| **Q-SPHERE** | Clean: $7.3_{\pm 2.1}$ / $53.0_{\pm 8.2}$ / $57.0_{\pm 2.6}$ | | | | | |
| Amplitude | $7.3_{\pm 1.5}$ | $53.0_{\pm 8.2}$ | $57.0_{\pm 2.6}$ | $8.3_{\pm 1.5}$ | $48.0_{\pm 6.1}$ | $57.0_{\pm 2.6}$ |
| Depolarizing | $9.3_{\pm 2.1}$ | $53.0_{\pm 8.2}$ | $47.0_{\pm 7.0}$ | $8.0_{\pm 1.7}$ | $53.0_{\pm 8.2}$ | $50.7_{\pm 8.4}$ |
| Phase | $7.3_{\pm 1.5}$ | $53.0_{\pm 8.2}$ | $57.0_{\pm 2.6}$ | $7.3_{\pm 1.5}$ | $53.0_{\pm 8.2}$ | $57.0_{\pm 2.6}$ |
| Mixed | $7.3_{\pm 1.5}$ | $53.0_{\pm 8.2}$ | $57.0_{\pm 2.6}$ | $7.3_{\pm 1.5}$ | $43.0_{\pm 2.6}$ | $57.7_{\pm 2.5}$ |
| **Spurious-Fourier** | Clean: $84.0_{\pm 2.6}$ / $51.3_{\pm 5.9}$ / $48.7_{\pm 5.5}$ | | | | | |
| Amplitude | $88.3_{\pm 5.5}$ | $51.3_{\pm 5.9}$ | $46.0_{\pm 3.6}$ | $92.0_{\pm 7.2}$ | $54.7_{\pm 7.1}$ | $51.3_{\pm 5.9}$ |
| Depolarizing | $90.0_{\pm 4.0}$ | $51.3_{\pm 5.9}$ | $48.3_{\pm 8.1}$ | $85.0_{\pm 2.6}$ | $51.0_{\pm 6.2}$ | $47.3_{\pm 5.5}$ |
| Phase | $90.3_{\pm 5.0}$ | $51.3_{\pm 5.9}$ | $48.0_{\pm 5.6}$ | $94.7_{\pm 2.1}$ | $51.3_{\pm 5.9}$ | $54.0_{\pm 5.6}$ |
| Mixed | $94.3_{\pm 2.5}$ | $51.3_{\pm 5.9}$ | $47.3_{\pm 6.8}$ | $87.7_{\pm 9.1}$ | $52.0_{\pm 6.9}$ | $52.3_{\pm 5.8}$ |
| **Q-LINK** | Clean: $14.7_{\pm 6.7}$ / $50.3_{\pm 4.6}$ / $49.7_{\pm 4.6}$ | | | | | |
| Amplitude | $8.3_{\pm 1.2}$ | $50.3_{\pm 4.6}$ | $48.0_{\pm 1.7}$ | $9.7_{\pm 0.6}$ | $50.3_{\pm 4.6}$ | $48.3_{\pm 4.2}$ |
| Depolarizing | $8.0_{\pm 1.0}$ | $51.0_{\pm 2.6}$ | $50.3_{\pm 4.2}$ | $10.3_{\pm 1.2}$ | $50.3_{\pm 4.6}$ | $46.3_{\pm 1.2}$ |
| Phase | $8.7_{\pm 0.6}$ | $53.7_{\pm 1.2}$ | $50.3_{\pm 4.6}$ | $8.0_{\pm 1.0}$ | $53.7_{\pm 1.2}$ | $54.0_{\pm 1.0}$ |
| Mixed | $11.3_{\pm 2.1}$ | $50.3_{\pm 4.6}$ | $48.7_{\pm 2.9}$ | $8.7_{\pm 0.6}$ | $51.0_{\pm 5.3}$ | $46.3_{\pm 1.2}$ |

We examine whether simulated feature reliance and OOD behavior persist on MonarQ. Figure 4 compares fixed clean-trained checkpoints across execution settings: analytic and finite-shot predictions largely agree (Appendix D.4), but hardware alters performance. Local geometric and Q-LINK failures persist; other VQCs' geometric reversal gains accompany ID losses, while Q-FOS reversal remains modest. Entangling stable-only ID accuracy falls from approximately 100% in simulation to 49.0% on Q-LINK and 49.3% on Spurious-Fourier; local Spurious-Fourier accuracy retains 100% (Appendix D.3). This illustrates the distinction between sampling reliability (Proposition 8) and implementation-dependent readouts (Equation 1, for single-upload models). Table 3 instead compares training conditions on the same MonarQ platform. Spurious-Fourier local OOD reversed accuracy rises from 84.0% (clean) to 94.7% (strong phase-aware) and 94.3% (moderate mixed), despite analytic accuracy remaining near 96%. Mixed training cycles amplitude, depolarizing, and phase channels, one per update (Appendix D.1). Other architectures remain near chance on this task, with no broad recovery across other tasks (Appendix D.2; paired changes: Appendix D.5). Hardware thus exposes stable-signal prediction failures and checkpoint-transfer differences concealed by noiseless accuracy. Detailed results for all 5 tasks appear in Appendix D.

## 5 Feature Reliance under Realistic Distribution Shifts

We examine whether the accessibility framework of Section 3 and controlled findings of Section 4 extend to realistic learning scenarios. Our 10 tasks span 5 families, combining real-data shortcut constructions and natural shifts across different signal structures and interaction requirements.

Table 4: **Performance on shared frozen representations at $d = 24$.** Balanced accuracy (%) over 3 seeds; subscripts show SD. Scores follow our preparation protocols and are not directly comparable to benchmark results. See Appendices E.2, E.3, and E.4 for protocols, additional results, and group metrics.

| Family | Task | Split | Logistic regression | MLP | Local VQC | Entangling VQC | Reuploading VQC |
|---|---|---|---|---|---|---|---|
| Appearance | Q-Colored-MNIST | Pooled ID | $87.5_{\pm 0.5}$ | $91.5_{\pm 0.9}$ | $75.6_{\pm 6.4}$ | $72.3_{\pm 7.6}$ | $67.0_{\pm 4.4}$ |
| | | OOD reversed | $36.6_{\pm 3.2}$ | $57.0_{\pm 10.8}$ | $43.3_{\pm 5.4}$ | $35.2_{\pm 12.3}$ | $40.3_{\pm 3.0}$ |
| | Q-Colored-Fashion-MNIST | Pooled ID | $94.0_{\pm 0.6}$ | $94.1_{\pm 0.3}$ | $87.1_{\pm 7.6}$ | $75.1_{\pm 11.8}$ | $77.6_{\pm 7.3}$ |
| | | OOD reversed | $78.0_{\pm 1.7}$ | $78.7_{\pm 1.8}$ | $81.8_{\pm 5.7}$ | $71.0_{\pm 14.0}$ | $73.2_{\pm 10.1}$ |
| | Waterbirds | Validation | $83.5_{\pm 1.7}$ | $85.6_{\pm 1.5}$ | $81.5_{\pm 2.2}$ | $78.4_{\pm 3.0}$ | $57.7_{\pm 11.4}$ |
| | | Test | $84.9_{\pm 1.0}$ | $86.2_{\pm 1.7}$ | $82.8_{\pm 1.5}$ | $77.8_{\pm 5.2}$ | $57.9_{\pm 11.1}$ |
| Frequency | MNIST-Fourier | Pooled ID | $88.8_{\pm 0.6}$ | $94.2_{\pm 1.1}$ | $84.1_{\pm 2.1}$ | $80.5_{\pm 5.0}$ | $63.3_{\pm 15.5}$ |
| | | OOD reversed | $48.3_{\pm 2.8}$ | $76.2_{\pm 3.6}$ | $40.0_{\pm 2.4}$ | $24.7_{\pm 5.5}$ | $41.0_{\pm 22.1}$ |
| | WOODS Spurious-Fourier | Pooled ID | $85.9_{\pm 0.7}$ | $85.8_{\pm 1.2}$ | $82.5_{\pm 3.8}$ | $83.3_{\pm 2.7}$ | $77.4_{\pm 10.4}$ |
| | | OOD test-out | $16.0_{\pm 0.8}$ | $16.4_{\pm 4.8}$ | $34.4_{\pm 22.1}$ | $27.4_{\pm 11.6}$ | $36.8_{\pm 21.8}$ |
| Temporal | TCMNIST-Source | Pooled ID | $85.1_{\pm 0.2}$ | $84.8_{\pm 0.4}$ | $71.8_{\pm 8.3}$ | $54.0_{\pm 6.5}$ | $60.0_{\pm 1.6}$ |
| | | OOD reversed | $9.9_{\pm 1.4}$ | $10.4_{\pm 1.4}$ | $24.5_{\pm 8.0}$ | $45.4_{\pm 8.3}$ | $37.4_{\pm 1.7}$ |
| | TCMNIST-Time | Pooled ID | $85.4_{\pm 0.4}$ | $84.4_{\pm 0.3}$ | $78.3_{\pm 4.4}$ | $61.1_{\pm 3.7}$ | $56.4_{\pm 6.2}$ |
| | | OOD reversed | $10.2_{\pm 0.6}$ | $11.6_{\pm 1.2}$ | $18.6_{\pm 6.2}$ | $41.1_{\pm 5.0}$ | $43.3_{\pm 9.9}$ |
| Domain | Camelyon17 | ID validation | $92.5_{\pm 0.4}$ | $92.5_{\pm 0.4}$ | $90.0_{\pm 2.4}$ | $82.8_{\pm 0.1}$ | $82.8_{\pm 2.2}$ |
| | | OOD test | $92.6_{\pm 1.0}$ | $90.5_{\pm 1.3}$ | $88.6_{\pm 3.1}$ | $79.6_{\pm 1.4}$ | $82.7_{\pm 0.7}$ |
| | Yearbook | Pooled ID | $98.3_{\pm 0.3}$ | $97.2_{\pm 0.7}$ | $95.1_{\pm 1.1}$ | $91.2_{\pm 5.5}$ | $71.5_{\pm 17.3}$ |
| | | OOD test | $85.5_{\pm 0.2}$ | $84.5_{\pm 0.2}$ | $87.0_{\pm 0.9}$ | $79.6_{\pm 5.6}$ | $61.5_{\pm 13.1}$ |
| Relational | RO-CLEVR | Pooled ID | $88.5_{\pm 0.6}$ | $87.6_{\pm 1.0}$ | $83.4_{\pm 0.8}$ | $78.8_{\pm 2.8}$ | $76.1_{\pm 1.7}$ |
| | | OOD reversed | $43.8_{\pm 4.1}$ | $46.1_{\pm 3.7}$ | $24.1_{\pm 4.9}$ | $15.7_{\pm 1.7}$ | $21.3_{\pm 1.5}$ |

**Appearance** family tasks (Q-Colored-MNIST, Q-Colored-Fashion-MNIST, Waterbirds) contrast object information with color/background cues. **Frequency** tasks (MNIST-Fourier, WOODS Spurious-Fourier) examine spectral competition; the latter uses a different construction from Section 4's controlled adaptation. **Temporal** tasks (TCMNIST-Source, TCMNIST-Time) contrast cross-observation relations with changing color cues. **Domain** tasks (Camelyon17, Yearbook) test hospital and historical-era shifts. Our **relational** task, RO-CLEVR, contrasts object relations with a local appearance shortcut. Definitions, adaptations, and sources appear in Appendix E.1.

Following QML dimensionality-reduction practice (Grant et al., 2018; Bowles et al., 2024), we fit preprocessing on training data only and freeze nested inputs across models ($d = 6$ baseline; $d \in \{12, 24\}$ scaling), with one feature per qubit. Post-preparation audits check data integrity and task-specific shifts, while held-out probes assess target and annotated shortcut recoverability (Appendix E.2.5). Logistic regression, an MLP, and local, entangling, and reuploading VQCs share inputs and splits. Here Pooled ID denotes held-out source validation. We report balanced accuracy across 3 seeds without OOD-based selection (Appendix E.2). These simulations extend the analysis across realistic shifts and representation sizes, complementing the controlled MonarQ evaluation.

### 5.1 Feature reliance and learning dynamics

We examine whether Section 4.1's task-dependent behavior persists on frozen realistic representations. Table 4 reports all tasks at the largest tested size, $d = 24$; Appendix E.3 provides $d \in \{6, 12\}$ and additional splits. VQCs use the fixed step-300 endpoint.

Proposition 1 distinguishes encoded information from usable predictions, motivating shared-input comparisons. VQC reversal accuracy is poor on Q-Colored-MNIST but higher on Q-Colored-Fashion-MNIST despite identical cue-agreement schedules. On MNIST-Fourier, the MLP reaches 76.2% while all VQCs remain below 50%, demonstrating that useful reversal prediction is achievable from their shared inputs. All models struggle on WOODS Spurious-Fourier. Temporal VQCs exceed classical reversal scores but have lower ID fit and below-chance OOD means. RO-CLEVR combines substantial ID fit with reversal failure even with entangling circuits, unlike Q-LINK: joint processing alone does not ensure stable relational learning. Waterbirds retains useful natural-shift accuracy except for reuploading;

Camelyon17 and Yearbook retain substantial OOD performance for several models. Yet Waterbirds' local VQC reaches 82.8% balanced test accuracy versus 55.3% worst-group accuracy; Yearbook also shows group disparities (Appendix E.4). Constructed reversal failures accord with Proposition 6's cue-law perspective; additional splits appear in Appendix E.3.

Table 5: **Learning beyond the early checkpoint.** At $d = 24$, $\Delta = \mathrm{BA}_{300} - \mathrm{BA}_{100}$, in percentage points: mean ± SD of 3 paired seed changes. Splits match Table 4. All-task trajectories at $d = 6, 12, 24$: Appendix E.6.

| Task | Local VQC | | Entangling VQC | | Reuploading VQC | |
|---|---|---|---|---|---|---|
| | ΔID | ΔOOD | ΔID | ΔOOD | ΔID | ΔOOD |
| Q-Colored-MNIST | $+4.5 \pm 13.0$ | $+1.2 \pm 11.3$ | $+8.3 \pm 4.1$ | $-6.6 \pm 11.5$ | $+11.7 \pm 13.4$ | $-8.9 \pm 4.7$ |
| Q-Colored-Fashion-MNIST | $+1.4 \pm 6.7$ | $+2.1 \pm 6.5$ | $-1.6 \pm 9.8$ | $-0.9 \pm 12.7$ | $+11.9 \pm 7.9$ | $+13.9 \pm 3.3$ |
| Waterbirds | $+1.9 \pm 6.0$ | $+2.9 \pm 5.2$ | $+4.2 \pm 2.0$ | $+4.6 \pm 1.7$ | $+6.3 \pm 10.2$ | $+6.2 \pm 9.9$ |
| MNIST-Fourier | $+4.4 \pm 8.8$ | $+5.3 \pm 16.8$ | $+6.9 \pm 10.1$ | $-0.5 \pm 8.3$ | $+13.0 \pm 16.5$ | $-13.8 \pm 25.0$ |
| WOODS Spurious-Fourier | $+0.9 \pm 3.4$ | $-1.2 \pm 16.1$ | $+13.5 \pm 10.5$ | $-14.4 \pm 9.3$ | $-3.7 \pm 9.3$ | $+1.0 \pm 16.7$ |
| TCMNIST-Source | $+5.7 \pm 6.7$ | $-7.4 \pm 8.9$ | $+3.3 \pm 3.3$ | $-2.7 \pm 4.4$ | $+5.8 \pm 3.9$ | $-8.9 \pm 3.7$ |
| TCMNIST-Time | $+14.2 \pm 6.4$ | $-16.0 \pm 7.7$ | $+11.2 \pm 3.4$ | $-8.8 \pm 5.1$ | $+4.4 \pm 7.0$ | $-3.3 \pm 12.7$ |
| Camelyon17 | $+1.7 \pm 2.0$ | $+1.2 \pm 4.5$ | $+20.2 \pm 10.5$ | $+14.3 \pm 12.5$ | $+7.7 \pm 1.8$ | $+4.6 \pm 6.0$ |
| Yearbook | $-0.1 \pm 1.2$ | $+2.8 \pm 1.4$ | $+24.4 \pm 18.5$ | $+17.9 \pm 15.7$ | $+23.1 \pm 17.7$ | $+15.5 \pm 9.2$ |
| RO-CLEVR | $+7.2 \pm 9.7$ | $-5.0 \pm 11.6$ | $+11.9 \pm 1.8$ | $-14.8 \pm 5.7$ | $+24.6 \pm 8.4$ | $-29.0 \pm 8.2$ |

From steps 100 to 300, RO-CLEVR reuploading gains 24.6 ID points but loses 29.0 OOD points; both temporal tasks show opposing mean changes for every VQC (Table 5). Q-Colored-Fashion-MNIST, Waterbirds, Camelyon17, and Yearbook generally improve on both measures. Continued fitting can deepen shortcut-reversal failure, though it need not hurt OOD performance. These checkpoints characterize finite-training accuracy changes; Proposition 7 describes initial feature-specific reliance growth. Full trajectories appear in Appendix E.6.

As $d$ rises from 6 to 24, entangling TCMNIST-Source OOD accuracy increases from 21.0% to 45.4% while ID falls from 75.1% to 54.0%: both approach chance. Entangling RO-CLEVR reversal accuracy instead falls from 28.3% to 15.7% (Appendix E.3). Since representation and VQC width change together, these comparisons do not isolate qubit-count effects. Appendix E.5 details the theory connection.

## 6 Accessibility Shapes Reliance and OOD Generalization

Quantum processing restrictions shape feature reliance, whose OOD consequences depend on whether the relied-on signal remains predictive under shift. Under explicit assumptions, the theory (Section 3) characterizes population-risk minimizers of local models with affine heads as cue-only. Controlled interventions, encoding ablations, and feature-role reversals show how accessibility can favor stable signals or shortcuts. Statistical predictiveness alone is therefore insufficient to explain the observed patterns of feature reliance. Realistic tasks show useful generalization under several natural shifts alongside severe shortcut-reversal failures. Training trajectories reveal opposing ID and OOD changes, while group metrics expose disparities hidden by aggregate accuracy. These findings motivate evaluating feature reliance alongside absolute ID and OOD performance.

## 7 Conclusion

Our theory and experiments support quantum feature accessibility as a mechanism linking quantum processing, learned reliance, and OOD generalization. Controlled interventions, realistic tasks, and MonarQ evaluations provide complementary evidence across the settings studied. Future work should test whether improving access to stable signals yields more reliable OOD generalization.

### AI assistance with writing

AI tools were used to assist with manuscript drafting and editing, including spelling, grammar, clarity, and readability. The authors take responsibility for the final text.

## References


Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*, 2019. doi: 10.48550/arXiv.1907.02893. URL https://arxiv.org/abs/1907.02893.

Joseph Bowles, Shahnawaz Ahmed, and Maria Schuld. Better than classical? the subtle art of benchmarking quantum machine learning models. *arXiv preprint arXiv:2403.07059*, 2024. doi: 10.48550/arXiv.2403.07059. URL https://arxiv.org/abs/2403.07059.

Calcul Québec. Official inauguration of MonarQ, a quantum computer dedicated to research and innovation, September 2024. URL https://www.calculquebec.ca/en/nouvelle/official-inauguration-of-monarq-a-quantum-computer-dedicated-to-research-and-innovation/. Published September 25, 2024.

Matthias C. Caro, Hsin-Yuan Huang, M. Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J. Coles. Generalization in quantum machine learning from few training data. *Nature Communications*, 13:4919, 2022. doi: 10.1038/s41467-022-32550-3. URL https://www.nature.com/articles/s41467-022-32550-3.

Matthias C. Caro, Hsin-Yuan Huang, Nicholas Ezzell, Joe Gibbs, Andrew T. Sornborger, Lukasz Cincio, Patrick J. Coles, and Zoë Holmes. Out-of-distribution generalization for learning quantum dynamics. *Nature Communications*, 14:3751, 2023. doi: 10.1038/s41467-023-39381-w. URL https://www.nature.com/articles/s41467-023-39381-w.

M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3:625–644, 2021. doi: 10.1038/s42254-021-00348-9. URL https://www.nature.com/articles/s42254-021-00348-9.

Callum Duffy and Marcin Jastrzebski. Spectral bias in variational quantum machine learning. *arXiv preprint arXiv:2506.22555*, 2025. doi: 10.48550/arXiv.2506.22555. URL https://arxiv.org/abs/2506.22555.

Aaron Fisher, Cynthia Rudin, and Francesca Dominici. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, 20(177):1–81, 2019. URL https://jmlr.org/papers/v20/18-760.html.

Jean-Christophe Gagnon-Audet, Kartik Ahuja, Mohammad-Javad Darvishi-Bayazi, Pooneh Mousavi, Guillaume Dumas, and Irina Rish. WOODS: Benchmarks for out-of-distribution generalization in time series. *Transactions on Machine Learning Research*, 2023. URL https://openreview.net/forum?id=mvftzofTYQ.

Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, Matthias Bethge, and Felix A. Wichmann. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2:665–673, 2020. doi: 10.1038/s42256-020-00257-z. URL https://www.nature.com/articles/s42256-020-00257-z.

Shiry Ginosar, Kate Rakelly, Sarah Sachs, Brian Yin, and Alexei A. Efros. A century of portraits: A visual historical record of american high school yearbooks. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, December 2015. URL https://openaccess.thecvf.com/content_iccv_2015_workshops/w18/html/Ginosar_A_Century_of_ICCV_2015_paper.html.

Edward Grant, Marcello Benedetti, Shuxiang Cao, Andrew Hallam, Joshua Lockhart, Vid Stojevic, Andrew G. Green, and Simone Severini. Hierarchical quantum classifiers. *npj Quantum Information*, 4:65, 2018. doi: 10.1038/s41534-018-0116-9. URL https://www.nature.com/articles/s41534-018-0116-9.

Aikaterini Gratsea and Patrick Huembeli. The effect of the processing and measurement operators on the expressive power of quantum models. *Quantum Machine Intelligence*, 5(2):32, 2023. doi: 10.1007/s42484-023-00109-0. URL https://link.springer.com/article/10.1007/s42484-023-00109-0.

Ishaan Gulrajani and David Lopez-Paz. In search of lost domain generalization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=lQdXeXDoWtI.

Tobias Haug and M. S. Kim. Generalization of quantum machine learning models using quantum fisher information metric. *Physical Review Letters*, 133:050603, 2024. doi: 10.1103/PhysRevLett.133.050603. URL https://arxiv.org/abs/2303.13462.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830. URL https://doi.org/10.1080/01621459.1963.10500830.

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://papers.nips.cc/paper_files/paper/2018/hash/5a4be1fa34e62bb8a6ec6b91d2462f5a-Abstract.html.

Damien Jian, Yu-Chao Huang, and Hsi-Sheng Goan. Test-time training with quantum auto-encoder: From distribution shift to noisy quantum circuits. *arXiv preprint arXiv:2411.06828*, 2024. doi: 10.48550/arXiv.2411.06828. URL https://arxiv.org/abs/2411.06828.

Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C. Lawrence Zitnick, and Ross Girshick. CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2901–2910, 2017. doi: 10.1109/CVPR.2017.215. URL https://openaccess.thecvf.com/content_cvpr_2017/html/Johnson_CLEVR_A_Diagnostic_CVPR_2017_paper.html.

Ian T. Jolliffe and Jorge Cadima. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016. doi: 10.1098/rsta.2015.0202. URL https://doi.org/10.1098/rsta.2015.0202.

Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, Tony Lee, Etienne David, Ian Stavness, Wei Guo, Berton Earnshaw, Imran Haque, Sara M Beery, Jure Leskovec, Anshul Kundaje, Emma Pierson, Sergey Levine, Chelsea Finn, and Percy Liang. WILDS: A benchmark of in-the-wild distribution shifts. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5637–5664. PMLR, 2021. URL https://proceedings.mlr.press/v139/koh21a.html.

Jonas M. Kübler, Simon Buchholz, and Bernhard Schölkopf. The inductive bias of quantum kernels. In *Advances in Neural Information Processing Systems*, volume 34, pp. 12661–12673, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/69adc1e107f7f7d035d7baf04342e1ca-Abstract.html.

Kevin J. Lang and Michael J. Witbrock. Learning to tell two spirals apart. In *Proceedings of the 1988 Connectionist Models Summer School*, pp. 52–59. Morgan Kaufmann, 1988. URL https://gwern.net/doc/ai/nn/fully-connected/1988-lang.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.

Aengus Lynch, Gbètondji J-S Dovonon, Jean Kaddour, and Ricardo Silva. Spawrious: A benchmark for fine control of spurious correlation biases. *arXiv preprint arXiv:2303.05470*, 2023. URL https://arxiv.org/abs/2303.05470.

Antonio Anna Mele, Armando Angrisani, Soumik Ghosh, Sumeet Khatri, Jens Eisert, Daniel Stilck França, and Yihui Quek. Noise-induced shallow circuits and absence of barren plateaus, 2024. URL https://arxiv.org/abs/2403.13927.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. doi: 10.48550/arXiv.2304.07193. URL https://arxiv.org/abs/2304.07193.

Giambattista Parascandolo, Alexander Neitz, Antonio Orvieto, Luigi Gresele, and Bernhard Schölkopf. Learning explanations that are hard to vary. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=hb1sDDSLbV.

Guanwen Qiu, Da Kuang, and Surbhi Goel. Complexity matters: Feature learning in the presence of spurious correlations. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 41658–41697. PMLR, 2024. URL https://proceedings.mlr.press/v235/qiu24e.html.

Shiori Sagawa, Pang Wei Koh, Tatsunori B. Hashimoto, and Percy Liang. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=ryxGuJrFvS.

Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103(3):032430, 2021. doi: 10.1103/PhysRevA.103.032430. URL https://arxiv.org/abs/2008.08605.

Samson Wang, Enrico Fontana, M. Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J. Coles. Noise-induced barren plateaus in variational quantum algorithms. *Nature Communications*, 12:6961, 2021. doi: 10.1038/s41467-021-27045-6. URL https://www.nature.com/articles/s41467-021-27045-6.

John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. doi: 10.1017/9781316848142. URL https://cs.uwaterloo.ca/~watrous/TQI/.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017. doi: 10.48550/arXiv.1708.07747. URL https://arxiv.org/abs/1708.07747.

Huaxiu Yao, Caroline Choi, Bochuan Cao, Yoonho Lee, Pang Wei Koh, and Chelsea Finn. Wild-Time: A benchmark of in-the-wild distribution shift over time. In *Advances in Neural Information Processing Systems*, volume 35, pp. 10309–10324. Curran Associates, Inc., 2022. doi: 10.52202/068431-0749. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/43119db5d59f07cc08fca7ba6820179a-Abstract-Datasets_and_Benchmarks.html.

# A Proofs and Supporting Results for Section 3

This appendix proves the results of Section 3 using observable duality and quantum-channel contraction (Watrous, 2018), encoding-dependent function classes (Schuld et al., 2021; Gratsea & Huembeli, 2023), conditional convexity, first-order learning dynamics (Jacot et al., 2018), and bounded-measurement concentration (Hoeffding, 1963). These standard tools connect specified quantum processing restrictions to feature competition and distribution shift; they are not claimed as new.

## A.1 Notation and scope

We retain $\rho_\theta(x)$, physical observables $O_k$, expectation vector $\mathbf{z}_\theta(x)$, and $f_{\theta,w}(x) = h_w(\mathbf{z}_\theta(x))$ from Section 2. The *effective* observable $\widetilde{O}_{\theta,k}$ acts on the encoded state and does not replace $O_k$. All systems are finite dimensional. The norms $\|\cdot\|_1$, $\|\cdot\|_\infty$, and $\|\cdot\|_2$ denote matrix trace, operator, and Euclidean vector norms; for a readout vector $w$, $\|w\|_1$ denotes its vector $\ell_1$ norm.

In binary results, $f$ is a real logit, labels are $\{-1, +1\}$, and classification thresholds at zero with a fixed tie rule. A sigmoid converts the logit to a probability and is outside the affine logit assumption. Unhatted expectations are exact, possibly for a physically noisy circuit; hats denote finite-shot estimates. Comparisons fix probability laws, preprocessing maps, and contrast orientations. Superscript $X$ denotes an input marginal; $P^X \ll Q^X$ means absolute continuity: every $Q^X$-null input set is $P^X$-null.

**Operational meaning of accessibility.** Accessibility claims must specify the feature $\phi$, diagnostic law $Q$, encoding, circuit, readout, head class, and resources. For a fixed circuit, affine head, and exact expectations, Equation (16) defines recovery diagnostic $\mathcal{R}^2_\phi$; finite-shot diagnostics also specify shot count and physical noise channel. Learning claims additionally specify initialization, optimizer, and update budget; Proposition 7 addresses only the initial change of a fixed signed contrast. These distinct quantities do not form a universal accessibility score: recoverability does not imply training selects a feature, nor selection that its label relation survives shift. Fix $Q$, feature definitions, readout and head conventions, and contrast scales before inspecting outcomes.

Recovery $\mathcal{R}^2_\phi(\theta; Q)$ assesses a fixed representation with an optimally refitted affine probe. Its supremum over a specified circuit parameter set measures capacity and need not be attained. At learned parameters $\widehat{\theta}$, it assesses representation content, not the trained head's reliance. Low recovery there does not establish a class-wide obstruction; high recovery does not ensure training selects the recovering head. Across-run associations with final reliance alone do not identify causality. Controlled encoding, circuit, or readout changes with a fixed diagnostic law and separate reliance measurement can test a proposed mechanism, subject to other changes the intervention induces.

## A.2 Observable restrictions and state distinguishability

*Proof of Proposition 1.* By the definition of the adjoint under the trace pairing,

$$\mathrm{Tr}[O_k \mathcal{E}_\theta(\rho)] = \mathrm{Tr}[\mathcal{E}^\dagger_\theta(O_k)\rho].$$

Subtracting this identity for the two encoded states expresses the expectation difference as $\mathrm{Tr}[\widetilde{O}_{\theta,k}\Delta\rho]$. Vanishing against each generator of $\mathcal{V}$ is equivalent, by linearity, to vanishing against its span. Identical vectors are identical inputs to any fixed head. For $\mathcal{E}_\theta(\rho) = U_\theta \rho U_\theta^\dagger$, cyclicity of trace gives Equation (1). □

The result concerns the specified expectation representation: equal local expectations need not imply equal joint outcome distributions available to another model, although single-Pauli expectations determine their binary marginals. Interleaved uploads generally prevent an input-independent processing channel after a fixed first upload; $\rho_\theta(x)$ remains well defined.

*Proof of Proposition 2.* These standard trace-norm facts are recalled from Watrous (2018); the proof is included for completeness. Write $A = \rho - \sigma = A_+ - A_-$ for its positive and negative parts. Since $-\|O\|_\infty I \preceq O \preceq \|O\|_\infty I$, applying these inequalities to the positive matrices $A_+$ and $A_-$ bounds the absolute trace pairing by $\|O\|_\infty(\mathrm{Tr}\, A_+ + \mathrm{Tr}\, A_-) =$

$\|O\|_\infty \|A\|_1$. This also yields the Hermitian dual formula $\|A\|_1 = \max_{-I \preceq H \preceq I} \operatorname{Tr}(HA)$, with equality at the sign of $A$. The adjoint of a trace-preserving positive map is unital and positive, so $-I \preceq H \preceq I$ implies $-I \preceq \mathcal{E}^\dagger(H) \preceq I$. Applying the dual formula to $\mathcal{E}(A)$ proves contraction. Unitary conjugation preserves all singular values and hence the trace norm. □

For an $L$-Lipschitz head in Euclidean norm, this further gives

$$|f(x) - f(x')| \le 2L\Big(\sum_k \|O_k\|_\infty^2\Big)^{1/2} D(\rho_\theta(x), \rho_\theta(x')).$$

An unrestricted head needs scale control for such a bound. Small nonzero state or expectation differences alone do not preclude noiseless classification or discrimination with arbitrarily many copies.

**How cross-qubit processing changes the readout.** For CNOT $U$ with the first qubit controlling the second, target-$Z$ measurement after $U$ on computational basis states returns $(-1)^{a\oplus b} = (-1)^a(-1)^b$. Both sides are diagonal, proving $U^\dagger(I \otimes Z)U = Z \otimes Z$: a local physical measurement can access a joint encoded-state observable. This gate identity neither requires every input to become entangled nor guarantees a prescribed circuit represents a task's stable predictor.

**A quantitative local/nonlocal decomposition.** For a bipartition $A|B$ of dimensions $d_A, d_B$, define $\Pi_B(H) = I_A \otimes \operatorname{Tr}_A(H)/d_A$. If $H = \mathcal{E}_\theta^\dagger(I_A \otimes O_B)$, then

$$|\operatorname{Tr}[H\Delta\rho]| \le \|O_B\|_\infty \|\operatorname{Tr}_A \Delta\rho\|_1 + \|H - \Pi_B(H)\|_\infty \|\Delta\rho\|_1. \tag{9}$$

To prove this, split $H = \Pi_B(H) + (H - \Pi_B(H))$ and apply the trace bound to each term. The local term equals $\operatorname{Tr}[(\operatorname{Tr}_A H/d_A)\operatorname{Tr}_A \Delta\rho]$. Both the adjoint channel and normalized partial trace are unital positive maps, so their operator norms on Hermitian inputs do not increase. This bounds $\|\operatorname{Tr}_A H/d_A\|_\infty$ by $\|O_B\|_\infty$ and proves the display. If $\mathcal{E}_\theta = \mathcal{E}_A \otimes \mathcal{E}_B$, its adjoint sends the physical observable to $I_A \otimes \mathcal{E}_B^\dagger(O_B)$, making the second term zero. A contrast with identical $B$ marginals is then invisible to this readout.

### A.3 Local circuits and additive logits

*Proof of Proposition 3.* For an observable $O_j$ on qubit $j$, the trace factors as

$$z_{\theta,j}(x) = \operatorname{Tr}[O_j \mathcal{E}_{\theta,j}(\rho_j(x_j))] \prod_{i \ne j} \operatorname{Tr}[\mathcal{E}_{\theta,i}(\rho_i(x_i))] = z_{\theta,j}(x_j),$$

because each channel preserves trace. Substitution into the affine head proves additivity. Composition of local layers leaves the product-channel structure unchanged. □

For the concrete single-upload $R_y(\xi_j(x_j))|0\rangle$ encoding, arbitrary single-qubit unitary processing, and Pauli-$Z$ measurement, write $U_j^\dagger Z U_j = u_{jx}X + u_{jy}Y + u_{jz}Z$ with $u_{jx}^2 + u_{jy}^2 + u_{jz}^2 = 1$. The encoded Bloch vector is $(\sin\xi_j, 0, \cos\xi_j)$. Consequently,

$$f(x) = b + \sum_j \big[\beta_j \sin\xi_j(x_j) + \gamma_j \cos\xi_j(x_j)\big], \tag{10}$$

where $\beta_j = w_j u_{jx}$ and $\gamma_j = w_j u_{jz}$. Fixed coordinatewise standardization, clipping, and encoding scale enter $\xi_j$. This restricted Fourier family accords with encoding-dependent expressivity (Schuld et al., 2021). Repeated local uploads of one coordinate can enlarge its univariate family but preserve additivity. Cross-coordinate preprocessing, different coordinates on one qubit, joint measurements, cross-qubit gates, and nonlinear heads require separate analysis. Parameter sharing preserves additivity but may affect the closure condition below.

### A.4 Population competition: exact and approximate results

**Precise class assumptions.** Let $\mathcal{F}$ be a class of measurable real logits $f(x_r, x_c) = a(x_r) + g(x_c)$, each integrable under $P_{\mathrm{tr}}$. For every admissible decomposition, assume $a(X_r)$ is

integrable and

$$\mathbb{E}_{P_{\mathrm{tr}}}[a(X_r) \mid Y, X_c] = \mu_a \quad \text{almost surely,} \tag{11}$$

where $\mu_a$ is a finite constant. Assume also $\bar{f}(x_c) = \mu_a + g(x_c) \in \mathcal{F}$. This closure is satisfied by local affine readouts with independently zeroable relational weights and a free bias: keep the cue processing and weights fixed, zero the relational weights, and absorb $\mu_a$ into the bias. The theorem concerns unregularized population logistic risk. Adding parameter penalties or coupling the removable weights to the cue can change this conclusion.

*Proof of Theorem 4.* For fixed $y \in \{-1, +1\}$, $\ell(t, y) = \log(1 + e^{-yt})$ has second derivative

$$\partial_1^2 \ell(t, y) = \frac{e^{yt}}{(1 + e^{yt})^2} > 0.$$

It is therefore strictly convex. Conditional on $(Y, X_c)$, the mean of $f(X)$ is $\mu_a + g(X_c) = \bar{f}(X_c)$. Conditional Jensen gives

$$\ell(\bar{f}(X_c), Y) \le \mathbb{E}[\ell(f(X), Y) \mid Y, X_c].$$

Integrating proves Equation (3). Under finite risk the nonnegative Jensen gap integrates to zero exactly when it is zero almost surely. Strict convexity makes this equivalent to $f(X) = \bar{f}(X_c)$ almost surely conditional on $(Y, X_c)$, and hence almost surely under the joint distribution. Since this equality depends only on $X$, it is also a $P_{\mathrm{tr}}^X$-almost-sure equality.

Let $\mathcal{F}_c$ be the subclass of cue-only functions in $\mathcal{F}$. Closure gives a member of $\mathcal{F}_c$ of no greater risk for each finite-risk member of $\mathcal{F}$, and $\mathcal{F}_c \subseteq \mathcal{F}$. Thus the two infima agree whenever a finite-risk member exists. An attained finite minimum must attain equality in Jensen and is cue-only almost surely. This does not require an attained minimum to exist, does not require parameters themselves to vanish, and does not assert equality away from the training support. □

The conditional mean condition is stronger than $\mathbb{E}[Y X_{r,j}] = 0$. For coordinate-additive $a(x_r) = \sum_{j \in r} a_j(x_j)$, fixed marginals $\mathcal{L}(X_{r,j} \mid Y, X_c)$ suffice; block-additive functions require fixed full block marginals. Relational coordinates need not be independent: their dependence can encode the label.

**An exact architecture-to-reliance construction.** Let $Y, U, C \in \{-1, +1\}$, where $Y$ is balanced, $U$ is balanced and independent of $(Y, C)$, and $\Pr(C = Y \mid Y) = p \in (1/2, 1)$ for both labels. Set $X_r = (X_1, X_2) = (U, YU)$ and $X_c = C$. The relation is fully predictive because $X_1 X_2 = Y$. Nevertheless, conditional on any $(Y, C)$, $U$ and $YU$ are each balanced. Hence both coordinates are individually independent of $(Y, C)$, and every integrable additive relational contribution satisfies Equation (11).

Encode each $x_j$ as the computational basis state $|(1 - x_j)/2\rangle$. Product processing, single-qubit measurements, and an affine head give Proposition 3; removable relational weights and a free bias give the theorem's closure. If identity processing is available for the cue qubit and it is measured in $Z$, the class represents $C \log(p/(1-p))$. The conditional logistic calculation below shows that this is the cue-only population optimum, so Theorem 4 implies that every attained finite population minimizer in this additive class is cue-only. The relation has not disappeared from the encoded input; the specified architecture and objective prevent it from improving the optimum.

A CNOT from the first relational qubit to the second removes this obstruction: its target-$Z$ expectation is $X_1 X_2 = Y$ by Appendix A.2, making the relation representable by an affine head. The computational basis inputs remain product states; the mechanism is access to a joint observable, not entanglement generation. Representability does not guarantee an optimizer finds this circuit or prefers the relation over an available cue.

The same construction gives a diagnostic prediction independent of the training correlation. Fix $Q$ by making $U, Y, C$ independent and balanced while retaining $X_1 = U$ and $X_2 = YU$. For any local product processing and single-qubit readout, every relational readout is a function of just $U$ or $YU$, each independent of $Y$ under $Q$; cue readouts depend only on the independent $C$. Thus $v_Y = 0$ and Equation (16) yields $\mathcal{R}_Y^2 = 0$. With the CNOT target readout, $Y$ is a readout coordinate, giving $\mathcal{R}_Y^2 = 1$ by zero reconstruction error. With an unchanged cue-$Z$ readout, $\mathcal{R}_C^2 = 1$ in both circuits. This example predicts a change

in relational recovery without imposing a stable–shortcut tradeoff. It also illustrates why a common diagnostic law matters: under the training law, $C$ alone predicts $Y$ and would contaminate a purported measure of relational recovery.

*Proof of Proposition 5.* On $[-B, B]$, the logistic second derivative is at least $\kappa_B = e^B/(1+e^B)^2$. Moreover, $\bar{f} = \mathbb{E}[f \mid Y, X_c]$ belongs to this interval almost surely. Strong convexity gives

$$\ell(f, Y) \geq \ell(\bar{f}, Y) + \partial_1 \ell(\bar{f}, Y)(f - \bar{f}) + \frac{\kappa_B}{2}(f - \bar{f})^2.$$

The linear term has zero conditional expectation given $(Y, X_c)$. Therefore

$$\frac{\kappa_B}{2}\mathbb{E}[(f - \bar{f})^2] \leq R(f; P_{\mathrm{tr}}) - R(\bar{f}; P_{\mathrm{tr}}) \leq R(f; P_{\mathrm{tr}}) - R_* \leq \varepsilon,$$

which proves the claim. For bounded Pauli expectations, any particular finite affine head satisfies $|f| \leq |b| + \|w\|_1$, although this need not give a small or uniform bound over trained models. □

### A.5 Distribution shift and the paired relational constructions

*Proof of Proposition 6.* The input set on which $f(X) \neq \bar{f}(X_c)$ has zero training probability and hence, by absolute continuity, zero test probability. Replacing $f$ by $\bar{f}$ in the test risk is therefore valid. Its expectation depends only on the joint law of $(X_c, Y)$, so equality of this law gives equality of risks.

For the quantitative statement, instead assume that the hypotheses of Proposition 5 hold and $r(x) = dP^X_{\mathrm{te}}/dP^X_{\mathrm{tr}}(x) \leq M$ almost surely. Then

$$\mathbb{E}_{P_{\mathrm{te}}}[(f - \bar{f})^2] = \mathbb{E}_{P_{\mathrm{tr}}}[r(X)(f - \bar{f})^2] \leq M\mathbb{E}_{P_{\mathrm{tr}}}[(f - \bar{f})^2] \leq \frac{2M\varepsilon}{\kappa_B}.$$

Since $|\partial_1 \ell(t, y)| = 1/(1 + e^{yt}) \leq 1$, logistic loss is 1-Lipschitz in its logit. The triangle inequality followed by Cauchy–Schwarz therefore gives

$$|R(f; P_{\mathrm{te}}) - R(\bar{f}; P_{\mathrm{te}})| \leq \mathbb{E}_{P_{\mathrm{te}}}|f - \bar{f}| \leq \sqrt{\mathbb{E}_{P_{\mathrm{te}}}[(f - \bar{f})^2]} \leq \sqrt{\frac{2M\varepsilon}{\kappa_B}}.$$

No assumption that the test conditional label law equals the training conditional law is needed for these two bounds: the discrepancy of the logits is a function of $X$, and the Lipschitz bound holds for either label. They compare the two predictors within the test environment, not test risk with training risk. The finite-logit hypothesis gives $|\bar{f}| \leq B$ training-almost surely and hence test-almost surely under this density-ratio condition, so both risks are finite. □

If the cue–label law shifts, exact cue-only risk can change in either direction; cue-only behavior determines neither the direction nor size of an OOD accuracy change. The quantitative bound transfers proximity to a cue-only reference within the test environment, not robustness: even zero discrepancy leaves its test risk unconstrained if the cue loses predictiveness.

**An illustrative exact reversal.** Suppose $Y$ is balanced, $C \in \{-1, +1\}$, and $\Pr(C = Y \mid Y) = p \in (1/2, 1)$ symmetrically for both labels. Bayes' rule gives $\Pr(Y = 1 \mid C = 1) = p$ and $\Pr(Y = 1 \mid C = -1) = 1 - p$. The unrestricted cue-only logistic optimum is

$$f^*_p(C) = C \log \frac{p}{1 - p}.$$

To verify this, conditional logistic risk at $C = c$ with posterior $\pi_c$ has derivative $\sigma(t) - \pi_c$, which vanishes uniquely at $t = \log(\pi_c/(1 - \pi_c))$. If the cue class represents this function, its classifier predicts $C$. Its accuracy in any environment with cue agreement $q$ is $q$, including $1/2$ under random agreement and $1 - p$ under symmetric reversal. This example is not an exact numerical accuracy formula for continuous noisy cues or differently pooled ID evaluations.

**Uniform-phase relational signal.** Consider the population construction

$$X_r = (\cos\Theta, \sin\Theta, \cos(\Theta+\Delta), \sin(\Theta+\Delta)), \tag{12}$$

where $\Theta \sim \mathrm{Unif}[0, 2\pi)$ is independent of $(Y, X_c, \Delta, e)$. The relative phase $\Delta$ can depend on the label and environment. Conditional on any $(Y, X_c, \Delta, e)$, each of the two angles is uniform modulo $2\pi$. Thus the marginal distribution of each two-coordinate circular block, and of each individual coordinate, is independent of $(Y, X_c, e)$. Taking expectations over $\Delta$ and the training environment preserves these marginals. Any coordinate-additive contribution, or any sum of separate functions of the two circular blocks, therefore satisfies Equation (11). Fixed coordinatewise preprocessing preserves this property because the pushforward of a common marginal remains common. Conditioning on a fitted preprocessing map, this argument applies to the population risk of the resulting fixed model.

The relationship between the blocks can still be predictive. For example, if $Y$ is balanced and $\Delta = \pi(1-Y)/2 + \zeta$ with independent $\zeta \sim \mathcal{N}(0, \sigma_\Delta^2)$, then

$$\cos\Delta = X_{r,1}X_{r,3} + X_{r,2}X_{r,4}, \qquad \mathbb{E}[Y\cos\Delta] = e^{-\sigma_\Delta^2/2} > 0.$$

The identity follows from the cosine subtraction formula. Also $\cos\Delta = Y\cos\zeta$, and the Gaussian characteristic function gives $\mathbb{E}\cos\zeta = e^{-\sigma_\Delta^2/2}$. Predictive information is consequently present in a joint relation even though every local marginal is label independent. Nonzero phase noise does not make that relation a perfectly deterministic label rule.

In the original orientation, the phase–label law is stable while a separate cue has the form $X_c = S_c d_e + \eta$, with $S_c \in \{-1, +1\}$ agreeing with $Y$ with environment-dependent probability and nondegenerate Gaussian $\eta$. The local model is restricted to the potentially shifting cue at a population optimum. In the role-reversed orientation, the local cue–label law is held fixed while a phase sign $S_r$ agrees with $Y$ with probability $p_e$. Taking $\Delta = \pi(1-S_r)/2 + \zeta$ gives $\mathbb{E}[Y\cos\Delta] = (2p_e - 1)e^{-\sigma_\Delta^2/2}$, while the same conditional-marginal argument holds. The local restriction now favors the stable cue. The result is neutral as to whether local or relational information is stable; those roles are specified by the environment family.

With nonzero wrapped-Gaussian phase variance and nondegenerate Gaussian cue noise, all these raw-input laws have positive densities on the common support $\mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{R}^2$, relative to the product of circular and Lebesgue measures. They are mutually absolutely continuous. Applying the same measurable preprocessing map preserves absolute continuity: a null set after preprocessing has a null preimage before it. This justifies the exact support-transfer statement for this population construction. It does not establish a uniformly bounded density ratio or quantify finite-sample optimization error.

**Classification transfer requires margin control.** The squared-logit bound in the proof of Proposition 6, and its logistic-risk bound, do not alone ensure small classification disagreement when $\bar{f}$ is near zero. For any $\tau > 0$, if $|\bar{f}| > \tau$ and $|f - \bar{f}| < \tau$, the two signs agree. A union bound and Markov's inequality therefore give

$$\Pr_{P_{\mathrm{te}}}(\operatorname{sign} f \neq \operatorname{sign} \bar{f}) \leq \Pr_{P_{\mathrm{te}}}(|\bar{f}| \leq \tau) + \frac{2M\varepsilon}{\kappa_B \tau^2}.$$

Classification-error differences are bounded by this disagreement probability because correctness indicators agree when predictions agree. These conditional bounds are not numerical claims for the experiments: common support supplies no $M$. Even finite $M$ can yield loose bounds with large density ratios, excess population risk, or logit bounds: $\kappa_B^{-1} = e^B + 2 + e^{-B}$. Observed training loss alone does not estimate $\varepsilon$ without population-risk and class-optimum information.

### A.6 Early loss-aligned reliance growth

Let $\lambda = (\theta, w)$ and abbreviate the scalar predictor by $f_\lambda$. Fix a probability law $\nu_j$ over ordered input pairs and a measurable orientation $s_j(x, x') \in \{-1, +1\}$. Positive signed contrast should have an explicitly chosen training-predictive meaning. These objects are not redefined after training or separately for each test environment. Specify which input factors are held fixed and, when constructing an intervention, whether the label is held fixed or recomputed. Define

$$q_j(\lambda) = \mathbb{E}_{\nu_j}[s_j(x, x')(f_\lambda(x) - f_\lambda(x'))], \quad \psi_0(x) = \nabla_\lambda f_\lambda(x)|_{\lambda=\lambda_0}.$$

For the chain rule, assume $q_j$ and population risk are finite and differentiable near $\lambda_0$, and the parameter path is differentiable at zero with derivative $g_0$. These assumptions on the resulting functions require no interchange of differentiation and expectation. The expanded formulas additionally require existence of the displayed derivative expectations and validity of differentiation under expectation, for example through an integrable local derivative bound. Then

$$g_0 = -\nabla_\lambda R(f_\lambda; P_{\mathrm{tr}})|_{\lambda=\lambda_0} = \mathbb{E}_{P_{\mathrm{tr}}}[-\partial_1 \ell(f_{\lambda_0}(X), Y)\, \psi_0(X)], \tag{13}$$

$$d_j = \nabla q_j(\lambda_0) = \mathbb{E}_{\nu_j}[s_j(x, x')(\psi_0(x) - \psi_0(x'))]. \tag{14}$$

For signed logistic loss, the scalar residual weight is $-\partial_1 \ell(f, Y) = Y/(1 + e^{Yf})$; for half squared error it is $Y - f$.

*Proof of Proposition 7.* The chain rule gives $\partial_t q_j(\lambda_t)|_{t=0} = \nabla q_j(\lambda_0)^\top \dot{\lambda}_0 = d_j^\top g_0$. For the discrete step $\lambda_1 = \lambda_0 + \eta g_0$, define

$$r_j = q_j(\lambda_1) - q_j(\lambda_0) - \eta d_j^\top g_0.$$

When $q_j$ is twice continuously differentiable on a neighborhood of this segment, Taylor's theorem with integral remainder gives

$$r_j = \eta^2 \int_0^1 (1-t) g_0^\top \nabla^2 q_j(\lambda_0 + t\eta g_0) g_0 \, dt.$$

Bounding the absolute quadratic form by $L_j \|g_0\|_2^2$ and integrating $1 - t$ gives $|r_j| \le \frac{1}{2}\eta^2 L_j \|g_0\|_2^2$, which is exactly Equation (7). □

For two feature contrasts $j, k$, a sufficient condition for a greater one-step *signed change* in $q_j$ than in $q_k$ is

$$(d_j - d_k)^\top g_0 > \tfrac{1}{2}\eta(L_j + L_k)\|g_0\|_2^2.$$

Indeed, writing $\Delta q_j = q_j(\lambda_1) - q_j(\lambda_0)$ and likewise for $k$, the two remainder identities give

$$\Delta q_j - \Delta q_k = \eta(d_j - d_k)^\top g_0 + r_j - r_k \ge \eta(d_j - d_k)^\top g_0 - \tfrac{1}{2}\eta^2(L_j + L_k)\|g_0\|_2^2,$$

where the inequality uses both absolute remainder bounds. Positivity proves the comparison. It compares changes, not final reliance levels, unless their initial levels also match. Both contrasts could decrease; positive growth of $q_j$ is ensured by the additional condition $d_j^\top g_0 > \frac{1}{2}\eta L_j \|g_0\|_2^2$. Contrast sizes and any normalizing constants must be fixed and comparable; making each contrast law a probability law alone does not make distinct perturbations equally strong. Weighted group comparisons follow by the same argument with nonnegative weights summing to one within each group.

**A prospective gradient test.** Before training, fix each feature's intervention law, held-fixed factors, label treatment, orientation, and normalization. Evaluate $d_j^\top g_0$ at a common initialization and compare its signed prediction with the same $q_j$'s change after a sufficiently small population-gradient step. Varying step size tests the second-order discrepancy scale; a numerical remainder guarantee additionally needs a valid $L_j$ bound. Feature comparisons use the same initialization and update and the preceding inequality; architecture comparisons preserve input intervention laws and conventions but recompute each update alignment. Empirical or adaptive updates use their realized direction $v_0$. This tests a prospective update-response prediction, not final accuracy or a post hoc contrast choice; it implies neither final reliance nor convergence.

For comparison to an unsigned gradient-variation diagnostic, define $G_j = \mathbb{E}_{\nu_j}\|\psi_0(x) - \psi_0(x')\|_2$. The triangle and Cauchy–Schwarz inequalities give $|d_j^\top g_0| \le G_j \|g_0\|_2$. A large $G_j$ supplies only an upper bound on possible alignment; it does not ensure positive growth. With the tangent kernel $K_0(x, u) = \psi_0(x)^\top \psi_0(u)$, the exact instantaneous predictor update is

$$\dot{f}_{\lambda_t}(x)|_{t=0} = \mathbb{E}_{P_{\mathrm{tr}}}[-\partial_1 \ell(f_{\lambda_0}(X), Y) K_0(x, X)].$$

Substitute $g_0$ into $\psi_0(x)^\top g_0$; no infinite-width limit or subsequent kernel constancy is assumed. For another update direction $v_0$, the chain rule gives $d_j^\top v_0$. A realized step $\lambda_1 = \lambda_0 + \eta v_0$ satisfies the same Taylor bound with $g_0$ replaced by $v_0$ if the Hessian bound

holds along its actual segment, by the same proof. Random-update claims condition on the realized direction, which need not equal the population gradient. Adam, minibatch randomness, estimated gradients, and long-time convergence require further analysis beyond the population-gradient claim.

### A.7 Physical noise and finite measurement budgets

**Physical noise.** Proposition 2 compares two states passed through the same input-independent channel. It does not imply that every fixed observable contrast decreases: a channel can rotate information into a previously insensitive measurement. An explicit attenuation law requires specifying both the channel and its location. For example, define the single-qubit depolarizing channel by $\mathcal{D}_p(\rho) = (1-p)\rho + pI/2$, with $0 \le p \le 1$. Its adjoint leaves $I$ fixed and multiplies each nonidentity Pauli by $1-p$. Therefore, for a Pauli string $P$ of weight $r$,

$$\mathrm{Tr}[P\mathcal{D}_p^{\otimes n}(\rho)] = (1-p)^r \,\mathrm{Tr}[P\rho]. \tag{15}$$

Tensor factorization of the adjoint action proves the formula when this product channel acts directly before measuring $P$. Final local-$Z$ readout has $r = 1$ even if $U_\theta^\dagger Z_j U_\theta$ is nonlocal. Interleaved noise requires propagation through the gate sequence; final effective-observable weight alone is insufficient. Comparisons must also match the implemented channel parameterization.

*Proof of Proposition 8.* For each coordinate and $t > 0$, Hoeffding's inequality gives

$$\Pr(|\widehat{z}_{\theta,k} - z_{\theta,k}| \ge t) \le 2\exp(-Nt^2/2),$$

since each outcome has range length two. Moreover,

$$|\widehat{f} - f| \le \sum_k |w_k|\,|\widehat{z}_{\theta,k} - z_{\theta,k}| \le W \max_k |\widehat{z}_{\theta,k} - z_{\theta,k}|.$$

A sign disagreement requires $|\widehat{f} - f| \ge m$. Apply the union bound at $t = m/W$, and cap the resulting probability by one. No independence across $k$ is needed. If $W = 0$, both logits equal the fixed bias. □

For disagreement probability at most $\delta \in (0,1)$, $N \ge 2W^2 m^{-2}\log(2K/\delta)$ shots per observable suffice; this is not a universal lower bound. Compatible joint measurements can share executions; separate estimation may require $KN$. The reference is the same possibly noisy circuit's expectation-valued predictor, not necessarily an ideal noiseless predictor. Within-observable independence and fixed means exclude unmodeled drift. Averaging the pointwise bound over inputs bounds disagreement probability and hence the absolute classification-error difference.

### A.8 Sensitivity, recovery, and readout ambiguity

These auxiliary diagnostics are not required for Theorem 4. Recovery specifies a feature, diagnostic law, readout, and head class. Its expectation-valued form assesses representation at fixed circuit parameters, without a training budget or assurance that the learned head attains its optimum.

**Linear feature recovery.** Fix a circuit and a diagnostic input law $Q$. Let $\phi(X)$ be a scalar feature with $0 < \mathrm{Var}_Q(\phi) < \infty$, and assume $\mathbf{z}_\theta$ has finite second moments. Center both under $Q$, and write $G = \mathrm{Cov}_Q(\mathbf{z}_\theta)$ and $v_\phi = \mathrm{Cov}_Q(\mathbf{z}_\theta, \phi)$. Writing $\mathcal{R}^2_\phi$ for $\mathcal{R}^2_\phi(\theta; Q)$, the fraction of feature variance recoverable by an affine head is

$$\mathcal{R}^2_\phi = 1 - \frac{\inf_{u,b}\mathbb{E}_Q[(\phi - b - u^\top \mathbf{z}_\theta)^2]}{\mathrm{Var}_Q(\phi)} = \frac{v_\phi^\top G^\dagger v_\phi}{\mathrm{Var}_Q(\phi)} \in [0,1], \tag{16}$$

where $G^\dagger$ is the Moore–Penrose inverse. To prove this, centering removes the optimal intercept and the remaining objective is $\mathrm{Var}(\phi) - 2u^\top v_\phi + u^\top G u$. A vector in the nullspace of $G$ has zero centered readout almost surely and hence zero pairing with $v_\phi$. Thus $v_\phi$ lies in the range of $G$, the minimum is attained at $u = G^\dagger v_\phi$ up to a nullspace vector, and the minimum value proves the formula. Nonnegativity of squared error and the zero predictor

give the interval bounds. This is standard least-squares projection, not a new information measure.

In the single-upload setting, for centered $\phi$ define $C_\phi = \mathbb{E}_Q[\phi(X)\rho_{\mathrm{enc}}(X)]$. Then $(v_\phi)_k = \mathrm{Tr}[\widetilde{O}_{\theta,k}C_\phi]$, by linearity and centering. This relates recovery to both the encoded feature operator and the available readout. A diagnostic law that independently varies stable and shortcut factors can prevent training correlations from making one factor appear recoverable merely through another. The chosen law must be stated; no such intervention is inferred from observational correlations alone.

For centered features $\Phi = (\phi_1, \ldots, \phi_d)^\top$ with $\mathbb{E}_Q[\Phi\Phi^\top] = I_d$, set $V = \mathbb{E}_Q[\mathbf{z}_\theta\Phi^\top]$ using the centered readout and $B = V^\top G^\dagger V$. Least-squares projection is an orthogonal contraction in $L^2(Q)$, so for any vector $a$, $a^\top Ba$ is the squared norm of the projection of $a^\top\Phi$ and lies between zero and $\|a\|_2^2$. Therefore

$$0 \preceq B \preceq I_d, \qquad \sum_{j=1}^{d} \mathcal{R}^2_{\phi_j} = \mathrm{Tr}\, B \le \mathrm{rank}(G) \le K.$$

Each eigenvalue is at most one and $\mathrm{rank}(B) \le \mathrm{rank}(G)$, proving the trace bound. Correlated features need not satisfy it without identity covariance. It gives no nontrivial two-feature bottleneck with six independent readouts and no universal stable–shortcut tradeoff.

**Weak signals and sampling.** If a centered unit-variance feature has noiseless readout $z = \epsilon\phi$, $\epsilon \neq 0$, an unrestricted affine head recovers it exactly by rescaling; small amplitude alone is not information loss. For a balanced binary feature $s$ with Pauli expectation $z = \epsilon s$, $0 < \epsilon \le 1$, an $N$-shot mean instead satisfies $\mathrm{Cov}(\widehat{z}, s) = \epsilon$ and $\mathrm{Var}(\widehat{z}) = \epsilon^2 + (1-\epsilon^2)/N$. The law of total variance and conditional unbiasedness prove these identities. Hence

$$\mathcal{R}^2_{s,N} = \frac{\epsilon^2}{\epsilon^2 + (1-\epsilon^2)/N}.$$

For any fixed target recovery in $(0, 1)$, the required shot count grows as $\epsilon^{-2}$ as $\epsilon \to 0$ in this example. More generally, conditionally unbiased noise $\widehat{\mathbf{z}} = \mathbf{z} + \eta$ adds $S = \mathbb{E}_Q\,\mathrm{Cov}(\eta \mid X)$ to $G$ and leaves $v_\phi$ unchanged, so Equation (16) uses $(G+S)^\dagger$. Physical noise need not be unbiased relative to the ideal circuit and is not covered by this replacement unless that assumption holds.

**Collisions are stronger than low linear recovery.** For a fixed expectation representation $Z = \mathbf{z}_\theta(X)$, the minimum binary classification error over all measurable heads is

$$\mathbb{E}_Z \min\{\Pr(Y = 1 \mid Z), \Pr(Y = -1 \mid Z)\}.$$

Conditional on $Z = z$, choosing the more probable class attains the smaller posterior error, proving the expression. For discrete outputs it equals $\sum_z \min\{\Pr(Z = z, Y = 1), \Pr(Z = z, Y = -1)\}$. For deterministic labels on a finite allowed input set, a head correct under every reweighting of that set exists exactly when every representation fiber contains only one label: necessity follows from identical inputs to the head, and sufficiency by assigning that common label to each fiber.

For example, let $s, c \in \{-1, +1\}$, $Y = s$ be balanced, and $\Pr(c = s \mid s) = p$ symmetrically. The scalar readout $z = s + c$ changes when $s$ changes at fixed $c$, yet both disagreement cases give $z = 0$ with opposite labels and equal mass. The Bayes error from this output is $(1-p)/2$. A shift changes the error by changing the mass in this collision. In contrast, $z = as + bc$ with $0 < a < b$ has four distinct outputs, permitting an arbitrary nonlinear head to decode $s$ exactly in noiseless arithmetic although no affine head can classify all four cases. In increasing readout order the labels are $-1, +1, -1, +1$, which cannot be separated by a single threshold. If $a > b > 0$, even the zero-threshold affine classifier succeeds. Thus neither sensitivity nor imperfect linear recovery is equivalent to unavoidable information loss for every head.

### A.9 Connection to empirical tests

The contribution is a conditional architecture-to-reliance mechanism. Encoding, channels, and observables constrain the quantum logit class; the population argument applies to any

class with the stated additive and closure properties, including classical additive models. The exact construction shows quantum processing can remove a representational obstruction. Neither it nor the bounds guarantees selection of the feature with greatest recovery, gradient magnitude, or measurement margin. When both signals can lower population risk, the theorem alone does not determine which training selects.

Applying the population result requires checking coordinate assignment, preprocessing, absence of cross-qubit interactions, observables, an affine logit with removable relational weights and free bias, and the generator's conditional marginals. Finite samples and adaptive optimization do not certify population optimality; empirical loss is not the excess population risk $\varepsilon$ of Proposition 5.

Table 6 maps claims to prospective tests; it does not assert every listed measurement was performed.

Table 6: What each result contributes and what would test its application.

| Result | Assumptions and useful test | Limit of the inference |
|---|---|---|
| Equation (16) | Fix the feature, $Q$, circuit, and affine probe class; estimate recovery on held-out diagnostic inputs. | Recovery at learned parameters differs from the class supremum and from reliance of the trained head; observational association alone is not causal. |
| Proposition 1 | Compute the effective readout or compare fixed input-pair expectations to identify contrasts it annihilates. | Blindness concerns the specified expectation representation; changing the measured information can remove it. |
| Proposition 2 | Check a shared channel and compare state distinguishability with the chosen observable contrast. | A standard upper bound on contrast is not a prediction of which feature is learned; a fixed contrast need not contract. |
| Proposition 3 and Theorem 4 | Verify coordinate assignment, product processing, affine head, conditional marginals, and closure; test the predicted cue-only optimum in the exact construction. | Low finite-sample loss does not certify a population optimum. The conclusion does not cover every encoding or nonlinear head. |
| Propositions 5 and 6 | Establish $B$, excess population risk, and support transfer; quantify $M$ and margins when using approximate OOD bounds. | Shared support alone gives no quantitative near-optimal transfer. Large constants can make the bounds uninformative. |
| Proposition 7 | Fix intervention scales and signs before the update; compare initial alignment with the subsequent signed response change. | This tests an initial or one-step prediction, not long-run reliance. The actual optimizer direction must be used. |
| Proposition 8 | Hold circuit and input fixed, estimate its expectation-valued margin, and repeat finite-shot predictions across shot counts. | Sampling disagreement is measured relative to the same circuit. Ideal-versus-noisy expectation changes require a separate comparison. |
| Equation (15) | Specify the physical noise channel and its circuit location; compare ideal and noisy expectations at fixed parameters. | The attenuation law is conditional on that noise model. Neither more shots nor a state-distance bound predicts noise-induced feature selection. |

Architecture comparisons test accessible function classes; fixed entanglers do not guarantee class containment, which must be checked. Post-training feature interventions test learned behavior under that intervention, not capacity or retraining on isolated features. Early-gradient tests require the specified update and comparable contrasts. Hardware inference and finite-shot tests assess noisy predictions and reliability, not the preceding optimization trajectory. ID/OOD comparisons must distinguish held-out ID environments from pooled training-distribution evaluation: different input laws need not satisfy the illustrative symmetric-reversal formula. These distinctions separate the theorem, diagnostics, and empirical evidence.

# B Controlled task definitions

Five geometric, spectral, and relational tasks specify competing stable and shortcut signals for Section 4.

## B.1 Common task design

We reserve $\theta$ for circuit parameters and $\vartheta, \varphi$ for data angles. Labels are $Y \in \{-1, +1\}$, stored as $(Y+1)/2$. *Stable* means an environment-invariant raw class-conditional mechanism, not error-free label recovery: Spirals-Signatures and Q-LINK sample $Y$ before drawing noisy, overlapping stable observations.

Primitive draws are independent unless specified. Source environments are $E_A, E_B$; held-out ID, Random, and OOD reversed complete the five settings:

$$E \in \{E_A, E_B, E_{\text{ID}}, E_{\text{rand}}, E_{\text{rev}}\}.$$

For each example, sample shortcut sign $S_E \in \{-1, +1\}$ as

$$S_E = \begin{cases} Y, & \text{with probability } p_E, \\ -Y, & \text{with probability } 1 - p_E, \end{cases} \qquad p_E = \begin{cases} 0.9, & E \in \{E_A, E_B, E_{\text{ID}}\}, \\ 0.5, & E = E_{\text{rand}}, \\ 0.1, & E = E_{\text{rev}}. \end{cases}$$

Figure headers report the population probability $P(A_{\text{env}}) = P(S_E = Y) = p_E$, not sample proportions. The shortcut is label-independent in Random and usually disagrees in OOD reversed.

Nonspectral environment cues have the form

$$\Phi_{\text{env}} = \mu S_E a_E + \sigma \xi, \qquad \xi \sim \mathcal{N}(0, I),$$

with task-specific $\mu$ and $\sigma$. The two-dimensional directions are

$$a_A^{(2)} = (1, 0), \qquad a_B^{(2)} = (0, 1), \qquad a_{\text{ID}}^{(2)} = a_{\text{rand}}^{(2)} = a_{\text{rev}}^{(2)} = \frac{(1, 1)}{\sqrt{2}},$$

and three-dimensional directions are

$$a_A^{(3)} = \frac{(1, 0.15, 0.05)}{\|(1, 0.15, 0.05)\|_2},$$
$$a_B^{(3)} = \frac{(0.10, 1, -0.10)}{\|(0.10, 1, -0.10)\|_2},$$
$$a_{\text{ID}}^{(3)} = a_{\text{rand}}^{(3)} = a_{\text{rev}}^{(3)} = \frac{(0.75, 0.75, 0)}{\|(0.75, 0.75, 0)\|_2}.$$

For nonspectral cues, $E_{\text{ID}}, E_{\text{rand}}, E_{\text{rev}}$ differ only in $p_E$; source environments also rotate the cue. Spectral frequency maps follow below.

**Pooled ID and the held-out ID generator.** Diagnostic columns show $E_A$, $E_B$, *Pooled ID*, $E_{\text{rand}}$, and $E_{\text{rev}}$. Equal source sizes give Pooled ID the mixture $\frac{1}{2}P_{E_A} + \frac{1}{2}P_{E_B}$ with agreement 0.9. This differs from the held-out $E_{\text{ID}}$ generator, which also has $p_E = 0.9$ but its own cue direction or frequency codebook.

**Saved inputs and normalization.** Stored inputs $X_{\text{gen}}$ concatenate stable and shortcut blocks for nonspectral tasks and contain the specified spectral coordinates otherwise. Before model preprocessing or encoding, dimensions are $6, 4, 10, 72, 6$ for Q-SPHERE, Spirals-Signatures, Spurious-Fourier, Q-FOS, and Q-LINK. Spectral splits use their own unlabeled coordinate means and standard deviations; source splits are standardized before pooling, without refitting. Stable mechanisms concern raw components, not identical finite-sample standardized marginals.

**Reading the task visualizations.** Blue denotes $Y = +1$ and orange $Y = -1$. Figures show stable geometry or raw spectra, class-conditional shortcut projections, and histograms. Descriptors are $d_{\text{env},i} = \Phi_{\text{env},i}$ for nonspectral tasks and specified below otherwise. Each

displayed batch uses

$$v = \frac{\overline{d}_+ - \overline{d}_-}{\|\overline{d}_+ - \overline{d}_-\|_2}, \qquad z_{\mathrm{env},i} = \langle d_{\mathrm{env},i}, v \rangle,$$

where $\overline{d}_\pm$ are class means. We use the first coordinate direction if a class is absent or the denominator is below $10^{-8}$. Histograms normalize bin probabilities within each class; scatter jitter aids visibility. Because $v$ is fitted to each batch's labels, its sign cannot establish reversal; finite-sample fluctuations can create separation in Random. Agreement is defined by $p_E$. Signed nonspectral comparisons instead use $\langle \Phi_{\mathrm{env}}, a_E \rangle$, whose generator direction is shared by $E_{\mathrm{ID}}, E_{\mathrm{rand}}, E_{\mathrm{rev}}$.

Raw waveforms, projected descriptors, and the Q-LINK relational statistic are visualization-only, not additional inputs. Figure angles $(\phi, \theta)$ map to spherical $(\varphi, \vartheta)$ below; Q-LINK $(\theta_A, \theta_B)$ map to $(\varphi_A, \varphi_B)$. All tasks share the shortcut-sign rule.

### B.2 Q-SPHERE

**Construction and purpose.** Q-SPHERE pairs a distributed nonlinear angular stable signal with a noisy directional cue. The sphere is classical input geometry, not quantum states or Bloch vectors (Figures 5 and 6).

**Mathematical definition.** Draw

$$u \sim \mathrm{Unif}[-1, 1], \qquad \varphi \sim \mathrm{Unif}[0, 2\pi), \qquad \vartheta = \arccos u,$$

together with $\epsilon_r \sim \mathcal{N}(0, 1)$, and set

$$r = 1 + 0.035\epsilon_r, \qquad \Phi_{\mathrm{stable}} = \begin{pmatrix} r \sin\vartheta \cos\varphi \\ r \sin\vartheta \sin\varphi \\ r \cos\vartheta \end{pmatrix}.$$

The untruncated $r$ has unbounded support concentrated near one. Set

$$h_{\mathrm{sph}} = \sin(3\varphi + 2.25\,\Phi_{\mathrm{stable},3}) + 0.25\sin(5\vartheta), \qquad Y = \begin{cases} +1, & h_{\mathrm{sph}} \geq 0, \\ -1, & h_{\mathrm{sph}} < 0. \end{cases}$$

Only Q-SPHERE derives $Y$ from a stable latent state; other tasks sample $Y$ first. The cue and saved input are

$$\Phi_{\mathrm{env}} = 1.05 S_E a_E^{(3)} + 0.28\xi, \quad \xi \sim \mathcal{N}(0, I_3), \qquad X_{\mathrm{gen}} = (\Phi_{\mathrm{stable}}, \Phi_{\mathrm{env}}) \in \mathbb{R}^6.$$

### B.3 Spirals-Signatures

**Origin and adaptation.** Spirals-Signatures adapts two-spiral geometry and spiral-shortcut designs (Lang & Witbrock, 1988; Parascandolo et al., 2021), pairing nonlinear stable coordinates with a noisy two-dimensional directional cue under the shared agreement protocol (Figure 7). Its accessibility contrast complements Q-SPHERE and depends on model and encoding.

**Mathematical definition.** First sample

$$Y \sim \mathrm{Unif}\{-1, +1\}, \qquad \tau \sim \mathrm{Unif}[0.5, 4\pi], \qquad \rho(\tau) = 0.105\tau,$$

and define $\delta_{+1} = 0$ and $\delta_{-1} = \pi$. The observed stable block is

$$\Phi_{\mathrm{stable}} = \begin{pmatrix} \rho(\tau)\cos(\tau + \delta_Y) \\ \rho(\tau)\sin(\tau + \delta_Y) \end{pmatrix} + 0.055\epsilon, \qquad \epsilon \sim \mathcal{N}(0, I_2).$$

Sampling $Y$ before perturbation yields overlapping class-conditional observations, not deterministic labels on the noisy plane. The cue and saved input are

$$\Phi_{\mathrm{env}} = S_E a_E^{(2)} + 0.20\xi, \quad \xi \sim \mathcal{N}(0, I_2), \qquad X_{\mathrm{gen}} = (\Phi_{\mathrm{stable}}, \Phi_{\mathrm{env}}) \in \mathbb{R}^4.$$

### B.4 Spurious-Fourier

**Origin and adaptation.** Adapted from WOODS Spurious-Fourier (Gagnon-Audet et al., 2023), this task retains stable high-frequency versus shortcut low-frequency competition,

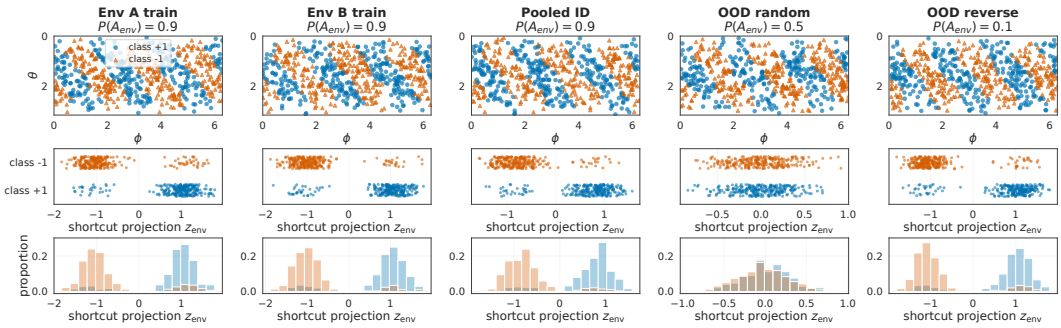


Figure 5: Q-SPHERE: labels in angular coordinates $(\varphi, \vartheta)$ (top); classwise cue projections $z_{\text{env}}$ and histograms (below). The stable mechanism is fixed; cue agreement varies. Projection convention: Appendix B.1.

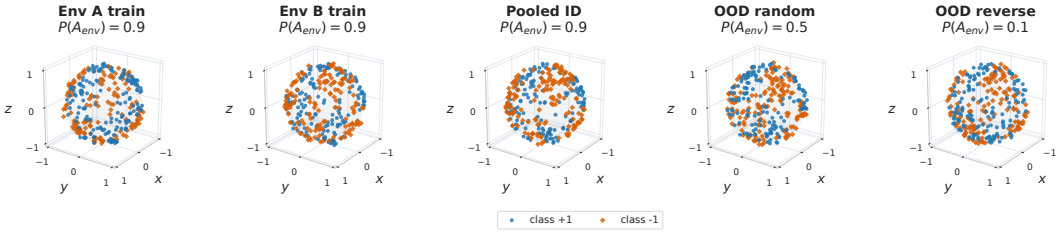


Figure 6: Q-SPHERE stable coordinates $\Phi_{\text{stable}} \in \mathbb{R}^3$ on the radially perturbed sphere, complementing Figure 5.

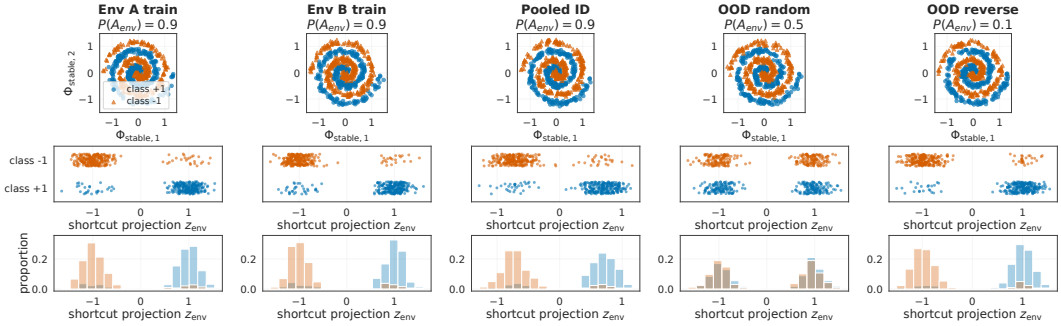


Figure 7: Spirals-Signatures: noisy stable observations (top); classwise signature projections $z_{\text{env}}$ and histograms (below). Signature agreement varies; vertical jitter is visualization-only.

with new codebooks, signal/noise parameters, the shared agreement protocol, and selected Fourier magnitudes replacing raw time-series inputs (Figure 8).

**Mathematical definition.** Sample $Y \sim \text{Unif}\{-1, +1\}$ and construct a length-$L = 128$ signal on $t_\ell = \ell/L$, $\ell = 0, \ldots, L-1$. The stable frequency is

$$k_{\text{st}}(Y) = \begin{cases} 19, & Y = +1, \\ 25, & Y = -1, \end{cases}$$

The shortcut-frequency map $k_{\text{env}}(E, S_E)$ is

| | $E_A$ | $E_B$ | $E_{\text{ID}}, E_{\text{rand}}, E_{\text{rev}}$ |
|---|---|---|---|
| $S_E = +1$ | 3 | 5 | 4 |
| $S_E = -1$ | 6 | 2 | 7 |

Draw independent phases

$$\alpha \sim \text{Unif}[-0.15, 0.15], \qquad \beta \sim \text{Unif}[-0.25, 0.25]$$

and independent $\epsilon_\ell \sim \mathcal{N}(0, 1)$. The raw waveform is

$$x_\ell = 0.75 \sin(2\pi k_{\text{st}} t_\ell + \alpha) + 1.05 \sin(2\pi k_{\text{env}} t_\ell + \beta) + 0.06 \sin(2\pi\, 11 t_\ell + 0.5) + 0.03\epsilon_\ell.$$

The fixed frequency-11 term is a nuisance. The unnormalized real Fourier magnitudes are

$$m_k = \left| \sum_{\ell=0}^{L-1} x_\ell \exp(-2\pi i k\ell/L) \right|, \qquad k = 0, \ldots, L/2.$$

With $B = (2, 3, 4, 5, 6, 7, 11, 19, 25, 31)$, the saved ten-coordinate feature is

$$X_{\text{gen},j} = \frac{m_{B_j} - \overline{m}_{B_j}}{\text{sd}(m_{B_j}) + 10^{-6}}, \qquad j = 1, \ldots, 10,$$

where splitwise statistics use the split size as the variance denominator.

**Visualization coordinates.** Figures retain raw waveforms and magnitude bins 0–39. The descriptor is

$$d_{\text{env}} = \big(k_{\text{env}}/10, \cos\beta, \sin\beta\big).$$

This generating-frequency/phase metadata is projected as in Appendix B.1; it neither projects nor augments the ten saved magnitudes.

**What the task isolates.** Held-out ID, Random, and OOD reversed share a shortcut codebook and differ only in $p_E$. Frequency refers to the source waveform; accessibility after magnitude extraction depends on encoding and model.

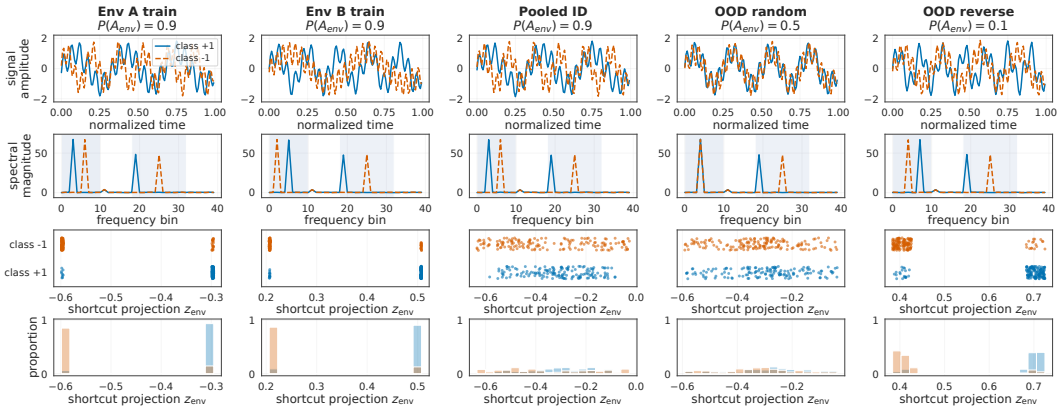


Figure 8: Spurious-Fourier: waveforms and magnitude spectra (above); classwise shortcut-descriptor projections and normalized histograms (below). Low-frequency codebooks vary across environments; the high-frequency mechanism is fixed. Descriptors are visualization-only.

### B.5 Q-FOS

**Construction and purpose.** Q-FOS (Quantum Fourier OOD Shortcuts) extends the spectral competition of Spurious-Fourier (Gagnon-Audet et al., 2023) to stable, shortcut, and nuisance peaks in 72 bins (Figure 9). It synthesizes spectra directly, without a time signal or Fourier transform.

**Mathematical definition.** Let the frequency bin be $b \in \{0, \ldots, 71\}$ and define

$$g(b; c, w) = \exp\left[ -\frac{1}{2} \left( \frac{b - c}{w} \right)^2 \right].$$

After sampling $Y \sim \text{Unif}\{-1, +1\}$ and $S_E$, define the stable centers and bridge center by

$$C_{\text{st}}(Y) = \begin{cases} \{47, 57\}, & Y = +1, \\ \{50, 61\}, & Y = -1, \end{cases} \qquad c_{\text{br}}(Y) = \begin{cases} 52, & Y = +1, \\ 55, & Y = -1. \end{cases}$$

The shortcut-center sets $C_{\text{env}}(E, S_E)$ are

| | $E_A$ | $E_B$ | $E_{\text{ID}}$ | $E_{\text{rand}}, E_{\text{rev}}$ |
|---|---|---|---|---|
| $S_E = +1$ | $\{4, 7\}$ | $\{8, 10\}$ | $\{5, 9\}$ | $\{6, 8\}$ |
| $S_E = -1$ | $\{5, 9\}$ | $\{3, 6\}$ | $\{4, 8\}$ | $\{5, 9\}$ |

For independent $J_1, J_2 \sim \mathrm{Unif}\{-1, 0, +1\}$, the nuisance centers are $C_{\mathrm{nuis}} = \{23+J_1, 28+J_2\}$. With independent $\epsilon_b \sim \mathcal{N}(0,1)$, the raw spectrum is

$$\widetilde{X}(b) = 0.03 + 1.05 \sum_{c \in C_{\mathrm{env}}} g(b; c, 1.05) + 0.42 \sum_{c \in C_{\mathrm{nuis}}} g(b; c, 1.45) + 0.85 \sum_{c \in C_{\mathrm{st}}} g(b; c, 1.20) + 0.18 g(b; c_{\mathrm{br}}, 2.5) + 0.015 \epsilon_b.$$

The saved 72-coordinate feature standardizes bins within each split (using split size as the variance denominator),

$$X_{\mathrm{gen}}(b) = \frac{\widetilde{X}(b) - \overline{\widetilde{X}}(b)}{\mathrm{sd}(\widetilde{X}(b)) + 10^{-6}}.$$

Shortcut, nuisance, and stable peaks have low-, middle-, and high-frequency centers, respectively, but Gaussian tails overlap. Noise is untruncated and raw spectra are not clipped at zero.

**Visualization coordinates.** For shortcut centers $c_1 < c_2$, the diagnostic descriptor is

$$d_{\mathrm{env}} = \big(c_1, c_2, S_E, (c_1 + c_2)/144\big).$$

Figure rows show raw spectra above and projections of this generating metadata below, following Appendix B.1. The descriptor does not project or augment $X_{\mathrm{gen}}$.

**What the task isolates.** Held-out ID and OOD differ in both $p_E$ and the shortcut-center codebook; Random and OOD reversed share a codebook and differ only in $p_E$. Overlapping peaks and label-independent nuisance structure complement selected Fourier magnitudes, testing access to stable bands under a different representation without holding all generating factors fixed.

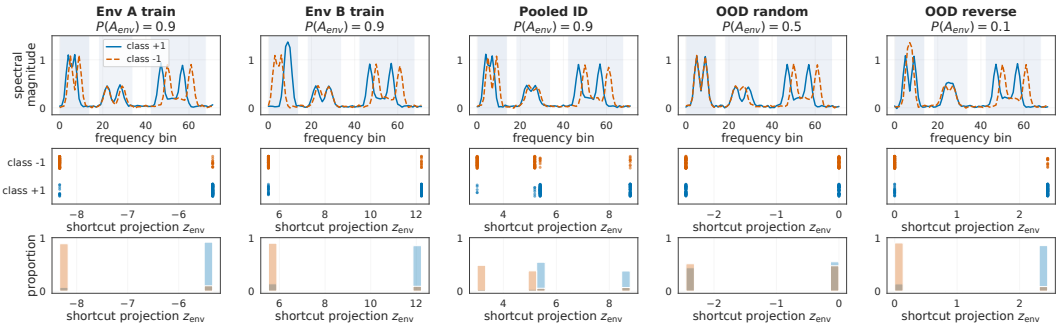


Figure 9: Q-FOS: raw spectra with low-frequency shortcut, mid-frequency nuisance, and high-frequency stable peaks (top); classwise descriptor projections and normalized histograms (below). Environments specify shortcut codebooks and agreement.

### B.6 Q-LINK

**Construction and purpose.** Q-LINK pairs a continuous joint phase signal with a local shortcut, relating to established XOR-based joint-information benchmarks (Qiu et al., 2024). Neither individual stable subsystem contains label information (Figures 10 and 11).

**Mathematical definition.** Sample

$$Y \sim \mathrm{Unif}\{-1, +1\}, \qquad \varphi_A \sim \mathrm{Unif}[0, 2\pi),$$

and then sample the relative displacement

$$\Delta \mid Y = +1 \sim \mathcal{N}(0, 0.38^2), \qquad \Delta \mid Y = -1 \sim \mathcal{N}(\pi, 0.38^2).$$

Set

$$\varphi_B = (\varphi_A + \Delta) \bmod 2\pi, \qquad x_A = (\cos\varphi_A, \sin\varphi_A), \qquad x_B = (\cos\varphi_B, \sin\varphi_B),$$

and let

$$\Phi_{\text{stable}} = (x_A, x_B) \in \mathbb{R}^4, \qquad r_{\text{link}} = \cos(\varphi_A - \varphi_B) = \cos\Delta.$$

Positive examples concentrate near aligned phases and negative examples near opposite phases. Because $Y$ determines the overlapping Gaussian laws of $\Delta$ independently of $\varphi_A$, $\text{sign}(r_{\text{link}})$ is not a deterministic label rule. For every class and environment,

$$\varphi_A \mid (Y, E) \sim \text{Unif}[0, 2\pi), \qquad \varphi_B \mid (Y, E) \sim \text{Unif}[0, 2\pi).$$

Within each environment, each subsystem is independent of $Y$ and the separately generated cue $\Phi_{\text{env}}$, while their relation depends on class. This population property does not imply identical sample histograms.

The local shortcut and saved input are

$$\Phi_{\text{env}} = S_E a_E^{(2)} + 0.22\xi, \quad \xi \sim \mathcal{N}(0, I_2), \qquad X_{\text{gen}} = (x_A, x_B, \Phi_{\text{env}}) \in \mathbb{R}^6.$$

The derived $r_{\text{link}}$ is not an input coordinate.

**What the task isolates.** Q-LINK separates marginal from relational information for Theorem 4. Cue-only behavior additionally requires that theorem's encoding, locality, readout, and optimization assumptions; it does not follow from the dataset alone.

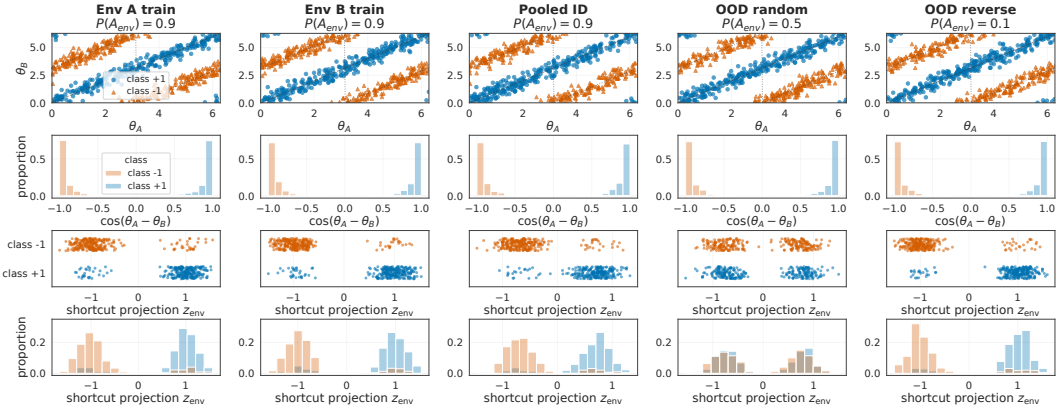


Figure 10: Q-LINK: joint phases $(\varphi_A, \varphi_B)$ and classwise $r_{\text{link}} = \cos(\varphi_A - \varphi_B)$ (first two rows); local cue projections and histograms (below). Batchwise projection signs do not establish reversal (Appendix B.1).

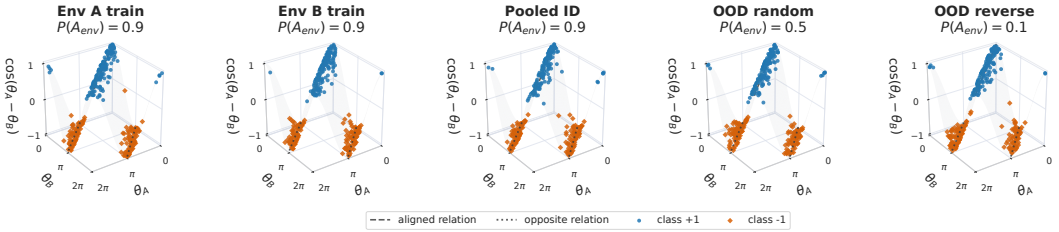


Figure 11: Q-LINK in $(\varphi_A, \varphi_B, r_{\text{link}})$ coordinates: positive examples concentrate near aligned phases and negative examples near opposite phases (Figure 10). The vertical coordinate is a derived statistic, not an input.

# C Controlled simulation: protocol and additional results

Protocols and complete results for the five primary tasks and Q-LINK-Reversed support Section 4.1; task definitions are in Appendix B. Figures retain information absent from numerical exports; MonarQ is reported separately.

Unless stated otherwise, entries are means ± sample SD over seeds $0, 1, 2$, with task, architecture, training and evaluation condition fixed. Accuracies and probability sensitivities are fractions; SD measures seed variation, not confidence intervals. A and B denote source-training subsets; Source ("Pooled ID") pools them and is not held-out ID. Random and Reverse denote held-out shifts. LR, Local, Ent., and Reup. denote logistic regression and the local, entangling, and reuploading VQCs. Gap denotes the stated ID/source accuracy minus OOD-reversed accuracy.

## C.1 Theory-to-evidence map

Table 7 maps the seven propositions and one theorem of Section 3 to behavioral evidence and its limits. Interventions do not measure the optimally refitted affine recoverability statistic of Section 3.1.

Table 7: Scope of the empirical connections to Section 3. Behavioral consistency does not establish an unmeasured inequality or population optimum.

| Formal result | Relevant evidence | Scope of the connection |
|---|---|---|
| Readout restrictions, Prop. 1 | Q-LINK interventions, Fig. 3; App. C.4. | Architecture-dependent stable-only behavior is consistent with access restrictions. No observable-span calculation establishes class-wide blindness. |
| Distinguishability, Prop. 2 | Fixed-checkpoint noise, App. C.5. | Prediction changes contextualize noise sensitivity. Trace distances and paired observable contrasts are not measured; accuracy need not decrease under a channel. |
| Local additivity, Prop. 3 | Circuit and preprocessing specification, App. C.2.1. | For Q-LINK, coordinatewise preprocessing, product encoding, local circuit operations, and an affine head yield additive logits. |
| Cue-only optima, Thm. 4 | Q-LINK interventions, Table 19. | Local cue reliance is consistent with the population mechanism. Finite training does not certify attainment of a population optimum. |
| Near-optimal logits, Prop. 5 | No direct empirical test. | Population excess risk, the uniform logit bound, and squared deviation from the conditional cue-only predictor are not estimated. |
| OOD consequences, Prop. 6 | Five-task comparison and Q-LINK role reversal, Figs. 2 and 3. | The stability of the favored signal affects OOD behavior. Accuracy contrasts do not estimate logistic-risk transfer constants. |
| Early reliance growth, Prop. 7 | Endpoint diagnostics and optimization context, App. C.6. | Initial signed feature contrasts, gradients, and realized update directions are absent; endpoint reliance does not test the initial-alignment identity. |
| Finite-shot reliability, Prop. 8 | Fixed-budget shot control, App. C.5. | One shot budget provides practical context. Margin-dependent disagreement probabilities and shot-count scaling are not measured. |

## C.2 Simulation protocol

We compare clean-trained logistic regression, an MLP, and three VQCs, then assess VQC inference noise, noise-aware training, and encoding changes. Q-LINK-Reversed separately tests feature-role reversal.

**Implementation and saved run records.** The same training implementation and task-specific preprocessing procedure are used across all three VQC architectures and seeds. The saved configurations for the seed-0 local-VQC runs on Spurious-Fourier and Q-FOS specify 6 qubits, 2 layers, 300 training updates, learning rate 0.04, and batch size 32. Their fitted preprocessors record standardization followed by PCA to six dimensions.

### C.2.1 Model architectures and input preprocessing

**Classical baselines.** Logistic regression uses an affine logit on standardized full inputs; the MLP has ReLU hidden layers $(64, 32)$ and a binary output. Neither baseline matches VQC parameter counts or feature dimensions. Settings are in Table 8.

**Inputs in the supplied code.** Classical models receive training-standardized full inputs. Quantum preprocessing uses training means and SDs to standardize coordinates (adding $10^{-8}$ to SD), projects inputs above six dimensions onto the standardized training matrix's first six principal components, and zero-pads smaller inputs to six. Coordinates are clipped to $[-\pi, \pi]$ for rotation angles; preprocessing is fixed at evaluation. In this implementation, Q-SPHERE and Q-LINK retain their six-coordinate assignment, while Spirals-Signatures gains two constant coordinates. The saved preprocessors for the seed-0 local-VQC runs confirm projections from 10 to 6 dimensions for Spurious-Fourier and from 72 to 6 for Q-FOS. The specified classical and quantum spectral pipelines therefore differ in input dimensionality.

**Quantum architectures.** All baseline VQCs use 6 qubits, 2 layers, and affine logit $f(x) = b + \sum_{j=0}^{5} w_j \langle Z_j \rangle_x$. Local applies one $R_Y$ upload and two layers of independent three-angle single-qubit rotations. Entangling adds the ordered CNOT ring $(0,1), (1,2), \ldots, (4,5), (5,0)$ after each rotation layer; reuploading also repeats $R_Y$ before each layer. Each has 36 circuit angles, 6 readout coefficients, and 1 bias (43 nominal parameters); a sigmoid maps the affine logit to probability. Fixed CNOT rings do not establish inclusion between parameterized model classes.

**Why compact models.** Shared qubit count, depth, readout, and nominal parameter count isolate connectivity and upload-schedule contrasts while permitting repeated evaluations. They imply neither equal expressivity nor optimal depth. Under their assumptions, Mele et al. (2024) show that noise can make deep random circuits effectively shallow for expectation estimation; this establishes neither depth optimality here nor the ineffectiveness of additional trained layers.

### C.2.2 Optimization and evaluation

Table 8 summarizes the full-run implementation settings. The saved configurations for the two spectral runs above agree with its VQC update budget, learning rate, and batch size. Seeds vary initialization and subset selection on fixed datasets, without regenerating them.

Table 8: Settings specified by the supplied full-run implementation.

| Component | Setting |
|---|---|
| Seeds and training samples | Seeds $0, 1, 2$; up to 1200 samples from each source environment, 2400 total. |
| Quantum objective and updates | Unregularized logistic loss; 300 Adam updates, learning rate 0.04, mini-batch size 32; analytic expectations and backpropagation. |
| Quantum initialization | Independent normal circuit angles and head coefficients with standard deviation 0.05; zero bias. Final-step checkpoint. |
| Logistic regression | At most 1000 solver iterations, otherwise library defaults. |
| MLP | Penalty parameter $10^{-4}$, at most 500 iterations, early stopping. |
| Clean and input diagnostics | Up to 1000 examples per source, pooled-source, randomized-shortcut, and reversed-shortcut split. |
| Noise evaluations | Up to 300 examples per held-out ID, randomized-shortcut, and reversed-shortcut split; comparisons use matched checkpoints and inputs. |

**Pooled-source and held-out ID are distinct.** Clean and intervention tables use Pooled ID: concatenated source samples from a reservoir overlapping training, measuring source fit, not independent generalization. Physical-noise and noise-aware evaluations use independently generated held-out ID with shortcut agreement 0.9, but potentially different cue direction or spectral codebook. These ID columns are not interchangeable. Random and OOD reversed are separately generated; shifts can change both shortcut agreement and task-specific environment properties (Appendix B).

### C.3 Clean training across the five primary tasks

Q-SPHERE and Spirals-Signatures show large source-to-reversal losses for all models. Reuploading improves on local-VQC reversal accuracy but remains below chance: partial behavioral change without robust stable-rule recovery.

On Spurious-Fourier, local-VQC reversal accuracy exceeds the interacting VQCs. On Q-FOS, both classical baselines achieve perfect accuracy; halving the quantum angle scale also yields perfect local and entangling reversal accuracy (Appendix C.7). These encoding-specific results establish neither intrinsic quantum disadvantage nor universal benefits from interactions. Q-FOS source heterogeneity is substantial: local-VQC accuracy is $0.383 \pm 0.033$ on A versus $0.762 \pm 0.053$ on B.

On Q-LINK, the local VQC follows logistic regression, while entangling and MLP models generalize; reuploading varies appreciably across seeds. This supports a representation-specific joint-relation contrast, not a universal model ranking. Small or negative source–reverse gaps alone do not establish task success.

Tables 9–13 give all splits and seed uncertainty underlying Figure 2.

Table 9: Q-SPHERE: clean-training accuracy and source–reverse gap. Conventions: Appendix C.

| Model | A | B | Source | Random | Reverse | Gap |
|---|---|---|---|---|---|---|
| LR | $0.886_{\pm 0.006}$ | $0.908_{\pm 0.005}$ | $0.898_{\pm 0.010}$ | $0.504_{\pm 0.005}$ | $0.100_{\pm 0.010}$ | $0.797_{\pm 0.016}$ |
| MLP | $0.884_{\pm 0.006}$ | $0.904_{\pm 0.006}$ | $0.896_{\pm 0.011}$ | $0.504_{\pm 0.004}$ | $0.101_{\pm 0.010}$ | $0.795_{\pm 0.018}$ |
| Local | $0.880_{\pm 0.002}$ | $0.908_{\pm 0.006}$ | $0.894_{\pm 0.009}$ | $0.504_{\pm 0.005}$ | $0.101_{\pm 0.010}$ | $0.793_{\pm 0.018}$ |
| Ent. | $0.855_{\pm 0.015}$ | $0.890_{\pm 0.003}$ | $0.877_{\pm 0.010}$ | $0.501_{\pm 0.006}$ | $0.125_{\pm 0.009}$ | $0.752_{\pm 0.018}$ |
| Reup. | $0.864_{\pm 0.026}$ | $0.845_{\pm 0.025}$ | $0.856_{\pm 0.036}$ | $0.572_{\pm 0.024}$ | $0.297_{\pm 0.026}$ | $0.558_{\pm 0.049}$ |

Table 10: Spirals-Signatures: clean-training accuracy and source–reverse gap. Conventions: Appendix C.

| Model | A | B | Source | Random | Reverse | Gap |
|---|---|---|---|---|---|---|
| LR | $0.901_{\pm 0.003}$ | $0.901_{\pm 0.001}$ | $0.905_{\pm 0.010}$ | $0.496_{\pm 0.002}$ | $0.102_{\pm 0.004}$ | $0.802_{\pm 0.009}$ |
| MLP | $0.900_{\pm 0.004}$ | $0.901_{\pm 0.003}$ | $0.903_{\pm 0.011}$ | $0.496_{\pm 0.002}$ | $0.102_{\pm 0.004}$ | $0.801_{\pm 0.010}$ |
| Local | $0.901_{\pm 0.003}$ | $0.899_{\pm 0.002}$ | $0.903_{\pm 0.010}$ | $0.496_{\pm 0.002}$ | $0.102_{\pm 0.004}$ | $0.801_{\pm 0.008}$ |
| Ent. | $0.902_{\pm 0.005}$ | $0.900_{\pm 0.002}$ | $0.905_{\pm 0.009}$ | $0.502_{\pm 0.007}$ | $0.149_{\pm 0.041}$ | $0.756_{\pm 0.048}$ |
| Reup. | $0.894_{\pm 0.006}$ | $0.899_{\pm 0.002}$ | $0.900_{\pm 0.012}$ | $0.522_{\pm 0.014}$ | $0.187_{\pm 0.021}$ | $0.713_{\pm 0.010}$ |

Table 11: Spurious-Fourier: clean-training accuracy and source–reverse gap. Conventions: Appendix C.

| Model | A | B | Source | Random | Reverse | Gap |
|---|---|---|---|---|---|---|
| LR | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.000_{\pm 0.000}$ |
| MLP | $1.000_{\pm 0.001}$ | $1.000_{\pm 0.001}$ | $1.000_{\pm 0.000}$ | $0.997_{\pm 0.002}$ | $0.996_{\pm 0.003}$ | $0.004_{\pm 0.003}$ |
| Local | $0.984_{\pm 0.004}$ | $0.986_{\pm 0.003}$ | $0.982_{\pm 0.006}$ | $0.961_{\pm 0.007}$ | $0.956_{\pm 0.003}$ | $0.026_{\pm 0.007}$ |
| Ent. | $0.833_{\pm 0.011}$ | $0.824_{\pm 0.007}$ | $0.827_{\pm 0.010}$ | $0.886_{\pm 0.024}$ | $0.868_{\pm 0.023}$ | $-0.041_{\pm 0.016}$ |
| Reup. | $0.866_{\pm 0.022}$ | $0.860_{\pm 0.014}$ | $0.858_{\pm 0.018}$ | $0.574_{\pm 0.112}$ | $0.569_{\pm 0.083}$ | $0.289_{\pm 0.083}$ |

Table 12: Q-FOS: clean-training accuracy and source–reverse gap. Conventions: Appendix C.

| Model | A | B | Source | Random | Reverse | Gap |
|---|---|---|---|---|---|---|
| LR | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.000_{\pm 0.000}$ |
| MLP | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.000_{\pm 0.000}$ |
| Local | $0.383_{\pm 0.033}$ | $0.762_{\pm 0.053}$ | $0.579_{\pm 0.014}$ | $0.368_{\pm 0.003}$ | $0.595_{\pm 0.036}$ | $-0.016_{\pm 0.027}$ |
| Ent. | $0.501_{\pm 0.072}$ | $0.652_{\pm 0.109}$ | $0.581_{\pm 0.027}$ | $0.455_{\pm 0.084}$ | $0.570_{\pm 0.021}$ | $0.011_{\pm 0.006}$ |
| Reup. | $0.552_{\pm 0.029}$ | $0.742_{\pm 0.053}$ | $0.653_{\pm 0.017}$ | $0.583_{\pm 0.036}$ | $0.452_{\pm 0.031}$ | $0.201_{\pm 0.022}$ |

### C.4 Post-training feature interventions

Fixed checkpoints are evaluated on original, stable-only, shortcut-only, and shortcut-perturbed inputs with unchanged preprocessing, labels, and threshold. Stable-only and shortcut-only retain designated coordinates and zero all others, including nuisance coordinates.

Table 13: Q-LINK: clean-training accuracy and source–reverse gap. Conventions: Appendix C.

| Model | A | B | Source | Random | Reverse | Gap |
|---|---|---|---|---|---|---|
| LR | $0.910_{\pm 0.003}$ | $0.882_{\pm 0.005}$ | $0.903_{\pm 0.009}$ | $0.506_{\pm 0.016}$ | $0.098_{\pm 0.005}$ | $0.805_{\pm 0.013}$ |
| MLP | $0.995_{\pm 0.006}$ | $0.998_{\pm 0.004}$ | $0.997_{\pm 0.005}$ | $0.938_{\pm 0.069}$ | $0.877_{\pm 0.124}$ | $0.121_{\pm 0.119}$ |
| Local | $0.909_{\pm 0.003}$ | $0.882_{\pm 0.005}$ | $0.902_{\pm 0.008}$ | $0.506_{\pm 0.016}$ | $0.098_{\pm 0.005}$ | $0.804_{\pm 0.012}$ |
| Ent. | $0.998_{\pm 0.001}$ | $0.990_{\pm 0.014}$ | $0.995_{\pm 0.004}$ | $0.985_{\pm 0.015}$ | $0.986_{\pm 0.019}$ | $0.010_{\pm 0.015}$ |
| Reup. | $0.985_{\pm 0.018}$ | $0.980_{\pm 0.016}$ | $0.982_{\pm 0.016}$ | $0.901_{\pm 0.084}$ | $0.846_{\pm 0.122}$ | $0.136_{\pm 0.111}$ |

Table 14 lists retained coordinates (zero-indexed, inclusive ranges). Masks precede fitted preprocessing. Spectral masks zero the generator-normalized representation, not physical components before normalization.

Table 14: Input-coordinate masks for the post-training diagnostics. Ranges include both endpoints.

| Task | Stable-only coordinates | Shortcut-only coordinates |
|---|---|---|
| Q-SPHERE | 0–2 | 3–5 |
| Spirals-Signatures | 0–1 | 2–3 |
| Spurious-Fourier | $7, 8$ (frequency bins $19, 25$) | 0–5 (frequency bins 2–7) |
| Q-FOS | 43–65 | 0–12 |
| Q-LINK | 0–3 | 4–5 |
| Q-LINK-Reversed (ablation) | 4–5 | 0–3 |

**Shortcut perturbations.** Q-SPHERE, Spirals-Signatures, and Q-LINK negate the shortcut block. Q-LINK-Reversed negates subsystem $B$ only, a $\pi$ phase shift reversing the relation while retaining the stable cue. Spurious-Fourier swaps columns $(1, 4)$ in source A, $(3, 0)$ in B, and $(2, 5)$ in held-out ID and both OOD splits; pooled-source evaluation instead negates the low-frequency block.

Q-FOS sequentially swaps columns $a + r, b + r$ for $r = -1, 0, 1$ around center pairs $(a, b)$: $(4, 5), (7, 9)$ in A; $(8, 3), (10, 6)$ in B; $(5, 4), (9, 8)$ in held-out ID; and $(6, 5), (8, 9)$ in both OOD splits. Pooled-source evaluation reverses columns 0–12. Overlapping window swaps are specified perturbations, not exact exchanges of isolated generative peaks.

**Summaries and interpretation.** For original, stable-only, and shortcut-only accuracies $A_0, A_s, A_c$, shortcut-removal and stable-removal changes are $A_0 - A_s$ and $A_0 - A_c$; balance is $C - S = A_c - A_s$. For shortcut transformation $T_c$, probability sensitivity $F = \mathbb{E}[|\widehat{p}(X) - \widehat{p}(T_c X)|]$ is estimated over the evaluation sample. It measures confidence change, not label disagreement. Table headers Original, Stable, Cue, and Flip denote the four input modes.

Interventions probe trained predictors: masks can leave input support and spectral perturbations can change the generator's conditional law. They describe transformation responses, not generative causal effects, isolated-signal retraining, or refitted affine recovery $\mathcal{R}^2_\phi$ (Section 3.1).

**What the interventions reveal.** Q-LINK source stable-only accuracies are $0.514 \pm 0.031$ (local), $0.998 \pm 0.002$ (entangling), and $0.952 \pm 0.077$ (reuploading). Local retains original accuracy on shortcut-only inputs; entangling responds less to cue perturbation. Subject to intervention caveats, this is consistent with local exclusion.

Spurious-Fourier reuploading has zero stable-only accuracy on every split. Masked anticorrelation can reflect learned orientation or calibration, not absent label information. Negative removal drops are retained; neither measures feature capacity or affine recovery (Section 3.1).

Tables 15–19 retain every model and split, probability sensitivity, and uncertainty in $C - S$.

### C.5 Fixed-checkpoint noise and measurement reliability

Clean-trained checkpoints remain fixed: ideal evaluation uses analytic expectations, physical noise uses analytic density-matrix simulation, and an otherwise ideal 1024-shot control isolates measurement sampling. Entries are accuracy percentages (mean ± sample SD over seeds $0, 1, 2$), holding task, architecture, split, and input mode fixed; no channels, architectures, or splits are pooled.

Table 15: Q-SPHERE: fixed-checkpoint interventions; $F$ is probability sensitivity. Definitions: Appendix C.4.

| Model | Split | Original | Stable | Cue | Flip | $C - S$ | $F$ |
|---|---|---|---|---|---|---|---|
| LR | A | $0.886_{\pm 0.006}$ | $0.506_{\pm 0.024}$ | $0.885_{\pm 0.006}$ | $0.114_{\pm 0.007}$ | $0.379_{\pm 0.029}$ | $0.728_{\pm 0.011}$ |
| | B | $0.908_{\pm 0.005}$ | $0.501_{\pm 0.013}$ | $0.908_{\pm 0.005}$ | $0.091_{\pm 0.006}$ | $0.407_{\pm 0.010}$ | $0.761_{\pm 0.005}$ |
| | Source | $0.898_{\pm 0.010}$ | $0.497_{\pm 0.021}$ | $0.897_{\pm 0.010}$ | $0.101_{\pm 0.011}$ | $0.400_{\pm 0.029}$ | $0.744_{\pm 0.003}$ |
| | Random | $0.504_{\pm 0.005}$ | $0.476_{\pm 0.025}$ | $0.504_{\pm 0.005}$ | $0.496_{\pm 0.005}$ | $0.028_{\pm 0.026}$ | $0.839_{\pm 0.002}$ |
| | Reverse | $0.100_{\pm 0.010}$ | $0.493_{\pm 0.023}$ | $0.100_{\pm 0.010}$ | $0.899_{\pm 0.010}$ | $-0.393_{\pm 0.022}$ | $0.843_{\pm 0.006}$ |
| MLP | A | $0.884_{\pm 0.006}$ | $0.525_{\pm 0.042}$ | $0.885_{\pm 0.004}$ | $0.116_{\pm 0.006}$ | $0.359_{\pm 0.038}$ | $0.602_{\pm 0.089}$ |
| | B | $0.904_{\pm 0.006}$ | $0.533_{\pm 0.054}$ | $0.904_{\pm 0.005}$ | $0.097_{\pm 0.004}$ | $0.371_{\pm 0.056}$ | $0.590_{\pm 0.118}$ |
| | Source | $0.896_{\pm 0.011}$ | $0.525_{\pm 0.051}$ | $0.897_{\pm 0.009}$ | $0.107_{\pm 0.010}$ | $0.371_{\pm 0.045}$ | $0.596_{\pm 0.102}$ |
| | Random | $0.504_{\pm 0.004}$ | $0.517_{\pm 0.047}$ | $0.504_{\pm 0.004}$ | $0.496_{\pm 0.004}$ | $-0.013_{\pm 0.047}$ | $0.679_{\pm 0.099}$ |
| | Reverse | $0.101_{\pm 0.010}$ | $0.553_{\pm 0.034}$ | $0.101_{\pm 0.010}$ | $0.899_{\pm 0.009}$ | $-0.452_{\pm 0.035}$ | $0.682_{\pm 0.098}$ |
| Local | A | $0.880_{\pm 0.002}$ | $0.529_{\pm 0.025}$ | $0.883_{\pm 0.006}$ | $0.119_{\pm 0.002}$ | $0.354_{\pm 0.025}$ | $0.721_{\pm 0.029}$ |
| | B | $0.908_{\pm 0.006}$ | $0.515_{\pm 0.009}$ | $0.909_{\pm 0.006}$ | $0.091_{\pm 0.006}$ | $0.394_{\pm 0.011}$ | $0.758_{\pm 0.022}$ |
| | Source | $0.894_{\pm 0.009}$ | $0.517_{\pm 0.039}$ | $0.896_{\pm 0.010}$ | $0.103_{\pm 0.010}$ | $0.379_{\pm 0.045}$ | $0.737_{\pm 0.020}$ |
| | Random | $0.504_{\pm 0.005}$ | $0.513_{\pm 0.040}$ | $0.504_{\pm 0.005}$ | $0.496_{\pm 0.004}$ | $-0.009_{\pm 0.036}$ | $0.881_{\pm 0.013}$ |
| | Reverse | $0.101_{\pm 0.010}$ | $0.517_{\pm 0.035}$ | $0.101_{\pm 0.010}$ | $0.899_{\pm 0.009}$ | $-0.416_{\pm 0.026}$ | $0.883_{\pm 0.013}$ |
| Ent. | A | $0.855_{\pm 0.015}$ | $0.496_{\pm 0.007}$ | $0.866_{\pm 0.016}$ | $0.152_{\pm 0.024}$ | $0.371_{\pm 0.023}$ | $0.609_{\pm 0.022}$ |
| | B | $0.890_{\pm 0.003}$ | $0.499_{\pm 0.026}$ | $0.898_{\pm 0.008}$ | $0.114_{\pm 0.005}$ | $0.398_{\pm 0.018}$ | $0.693_{\pm 0.028}$ |
| | Source | $0.877_{\pm 0.010}$ | $0.488_{\pm 0.021}$ | $0.883_{\pm 0.015}$ | $0.132_{\pm 0.021}$ | $0.395_{\pm 0.021}$ | $0.651_{\pm 0.013}$ |
| | Random | $0.501_{\pm 0.006}$ | $0.507_{\pm 0.036}$ | $0.505_{\pm 0.007}$ | $0.494_{\pm 0.009}$ | $-0.002_{\pm 0.030}$ | $0.671_{\pm 0.011}$ |
| | Reverse | $0.125_{\pm 0.009}$ | $0.490_{\pm 0.022}$ | $0.118_{\pm 0.013}$ | $0.874_{\pm 0.009}$ | $-0.372_{\pm 0.012}$ | $0.685_{\pm 0.015}$ |
| Reup. | A | $0.864_{\pm 0.026}$ | $0.641_{\pm 0.039}$ | $0.579_{\pm 0.182}$ | $0.285_{\pm 0.038}$ | $-0.062_{\pm 0.144}$ | $0.532_{\pm 0.072}$ |
| | B | $0.845_{\pm 0.025}$ | $0.625_{\pm 0.057}$ | $0.808_{\pm 0.058}$ | $0.303_{\pm 0.025}$ | $0.182_{\pm 0.083}$ | $0.470_{\pm 0.037}$ |
| | Source | $0.856_{\pm 0.036}$ | $0.632_{\pm 0.057}$ | $0.690_{\pm 0.095}$ | $0.300_{\pm 0.038}$ | $0.059_{\pm 0.042}$ | $0.495_{\pm 0.061}$ |
| | Random | $0.572_{\pm 0.024}$ | $0.647_{\pm 0.042}$ | $0.510_{\pm 0.010}$ | $0.567_{\pm 0.011}$ | $-0.137_{\pm 0.047}$ | $0.500_{\pm 0.038}$ |
| | Reverse | $0.297_{\pm 0.026}$ | $0.646_{\pm 0.016}$ | $0.265_{\pm 0.061}$ | $0.853_{\pm 0.017}$ | $-0.382_{\pm 0.073}$ | $0.510_{\pm 0.044}$ |

Table 16: Spirals-Signatures: fixed-checkpoint interventions; $F$ is probability sensitivity. Definitions: Appendix C.4.

| Model | Split | Original | Stable | Cue | Flip | $C - S$ | $F$ |
|---|---|---|---|---|---|---|---|
| LR | A | $0.901_{\pm 0.003}$ | $0.534_{\pm 0.024}$ | $0.901_{\pm 0.003}$ | $0.102_{\pm 0.003}$ | $0.367_{\pm 0.023}$ | $0.760_{\pm 0.004}$ |
| | B | $0.901_{\pm 0.001}$ | $0.564_{\pm 0.009}$ | $0.899_{\pm 0.001}$ | $0.105_{\pm 0.002}$ | $0.335_{\pm 0.009}$ | $0.756_{\pm 0.005}$ |
| | Source | $0.905_{\pm 0.010}$ | $0.556_{\pm 0.036}$ | $0.904_{\pm 0.009}$ | $0.100_{\pm 0.009}$ | $0.348_{\pm 0.037}$ | $0.759_{\pm 0.004}$ |
| | Random | $0.496_{\pm 0.002}$ | $0.512_{\pm 0.015}$ | $0.496_{\pm 0.002}$ | $0.504_{\pm 0.002}$ | $-0.016_{\pm 0.015}$ | $0.893_{\pm 0.003}$ |
| | Reverse | $0.102_{\pm 0.004}$ | $0.565_{\pm 0.011}$ | $0.102_{\pm 0.004}$ | $0.898_{\pm 0.004}$ | $-0.463_{\pm 0.008}$ | $0.888_{\pm 0.002}$ |
| MLP | A | $0.900_{\pm 0.004}$ | $0.537_{\pm 0.015}$ | $0.901_{\pm 0.004}$ | $0.109_{\pm 0.016}$ | $0.363_{\pm 0.012}$ | $0.447_{\pm 0.080}$ |
| | B | $0.901_{\pm 0.003}$ | $0.547_{\pm 0.041}$ | $0.898_{\pm 0.001}$ | $0.111_{\pm 0.004}$ | $0.351_{\pm 0.040}$ | $0.441_{\pm 0.121}$ |
| | Source | $0.903_{\pm 0.011}$ | $0.551_{\pm 0.039}$ | $0.904_{\pm 0.010}$ | $0.103_{\pm 0.016}$ | $0.353_{\pm 0.045}$ | $0.447_{\pm 0.096}$ |
| | Random | $0.496_{\pm 0.002}$ | $0.527_{\pm 0.020}$ | $0.496_{\pm 0.002}$ | $0.504_{\pm 0.002}$ | $-0.031_{\pm 0.021}$ | $0.560_{\pm 0.114}$ |
| | Reverse | $0.102_{\pm 0.004}$ | $0.566_{\pm 0.032}$ | $0.102_{\pm 0.004}$ | $0.898_{\pm 0.004}$ | $-0.463_{\pm 0.034}$ | $0.556_{\pm 0.114}$ |
| Local | A | $0.901_{\pm 0.003}$ | $0.532_{\pm 0.045}$ | $0.902_{\pm 0.003}$ | $0.100_{\pm 0.003}$ | $0.370_{\pm 0.042}$ | $0.771_{\pm 0.014}$ |
| | B | $0.899_{\pm 0.002}$ | $0.572_{\pm 0.041}$ | $0.898_{\pm 0.002}$ | $0.103_{\pm 0.001}$ | $0.326_{\pm 0.039}$ | $0.771_{\pm 0.004}$ |
| | Source | $0.903_{\pm 0.010}$ | $0.538_{\pm 0.056}$ | $0.904_{\pm 0.009}$ | $0.097_{\pm 0.010}$ | $0.365_{\pm 0.061}$ | $0.773_{\pm 0.006}$ |
| | Random | $0.496_{\pm 0.002}$ | $0.545_{\pm 0.029}$ | $0.496_{\pm 0.002}$ | $0.504_{\pm 0.002}$ | $-0.049_{\pm 0.031}$ | $0.946_{\pm 0.003}$ |
| | Reverse | $0.102_{\pm 0.004}$ | $0.552_{\pm 0.028}$ | $0.102_{\pm 0.004}$ | $0.898_{\pm 0.004}$ | $-0.450_{\pm 0.027}$ | $0.944_{\pm 0.003}$ |
| Ent. | A | $0.902_{\pm 0.005}$ | $0.483_{\pm 0.043}$ | $0.902_{\pm 0.003}$ | $0.112_{\pm 0.016}$ | $0.418_{\pm 0.043}$ | $0.741_{\pm 0.028}$ |
| | B | $0.900_{\pm 0.002}$ | $0.486_{\pm 0.036}$ | $0.898_{\pm 0.001}$ | $0.104_{\pm 0.003}$ | $0.412_{\pm 0.037}$ | $0.786_{\pm 0.007}$ |
| | Source | $0.905_{\pm 0.009}$ | $0.490_{\pm 0.038}$ | $0.904_{\pm 0.009}$ | $0.104_{\pm 0.016}$ | $0.414_{\pm 0.031}$ | $0.762_{\pm 0.017}$ |
| | Random | $0.502_{\pm 0.007}$ | $0.480_{\pm 0.046}$ | $0.496_{\pm 0.002}$ | $0.505_{\pm 0.005}$ | $0.016_{\pm 0.045}$ | $0.741_{\pm 0.088}$ |
| | Reverse | $0.149_{\pm 0.041}$ | $0.502_{\pm 0.030}$ | $0.103_{\pm 0.005}$ | $0.879_{\pm 0.021}$ | $-0.399_{\pm 0.028}$ | $0.737_{\pm 0.093}$ |
| Reup. | A | $0.894_{\pm 0.006}$ | $0.531_{\pm 0.015}$ | $0.901_{\pm 0.004}$ | $0.156_{\pm 0.037}$ | $0.370_{\pm 0.014}$ | $0.699_{\pm 0.032}$ |
| | B | $0.899_{\pm 0.002}$ | $0.545_{\pm 0.027}$ | $0.897_{\pm 0.003}$ | $0.113_{\pm 0.013}$ | $0.352_{\pm 0.027}$ | $0.787_{\pm 0.037}$ |
| | Source | $0.900_{\pm 0.012}$ | $0.534_{\pm 0.025}$ | $0.903_{\pm 0.008}$ | $0.136_{\pm 0.022}$ | $0.368_{\pm 0.018}$ | $0.741_{\pm 0.009}$ |
| | Random | $0.522_{\pm 0.014}$ | $0.534_{\pm 0.013}$ | $0.497_{\pm 0.004}$ | $0.515_{\pm 0.020}$ | $-0.037_{\pm 0.015}$ | $0.705_{\pm 0.022}$ |
| | Reverse | $0.187_{\pm 0.021}$ | $0.546_{\pm 0.020}$ | $0.110_{\pm 0.011}$ | $0.868_{\pm 0.010}$ | $-0.437_{\pm 0.011}$ | $0.698_{\pm 0.022}$ |

Table 17: Spurious-Fourier: fixed-checkpoint interventions; $F$ is probability sensitivity. Definitions: Appendix C.4.

| Model | Split | Original | Stable | Cue | Flip | $C - S$ | $F$ |
|---|---|---|---|---|---|---|---|
| LR | A | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.858_{\pm 0.014}$ | $1.000_{\pm 0.000}$ | $-0.142_{\pm 0.014}$ | $0.002_{\pm 0.000}$ |
| | B | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.812_{\pm 0.008}$ | $1.000_{\pm 0.000}$ | $-0.188_{\pm 0.008}$ | $0.002_{\pm 0.000}$ |
| | Source | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.834_{\pm 0.006}$ | $1.000_{\pm 0.000}$ | $-0.166_{\pm 0.006}$ | $0.002_{\pm 0.000}$ |
| | Random | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.503_{\pm 0.014}$ | $1.000_{\pm 0.000}$ | $-0.497_{\pm 0.014}$ | $0.000_{\pm 0.000}$ |
| | Reverse | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.490_{\pm 0.020}$ | $1.000_{\pm 0.000}$ | $-0.510_{\pm 0.020}$ | $0.000_{\pm 0.000}$ |
| MLP | A | $1.000_{\pm 0.001}$ | $1.000_{\pm 0.000}$ | $0.818_{\pm 0.046}$ | $0.990_{\pm 0.007}$ | $-0.182_{\pm 0.046}$ | $0.116_{\pm 0.020}$ |
| | B | $1.000_{\pm 0.001}$ | $1.000_{\pm 0.000}$ | $0.824_{\pm 0.010}$ | $0.991_{\pm 0.007}$ | $-0.176_{\pm 0.010}$ | $0.123_{\pm 0.039}$ |
| | Source | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.826_{\pm 0.016}$ | $0.988_{\pm 0.010}$ | $-0.174_{\pm 0.016}$ | $0.130_{\pm 0.031}$ |
| | Random | $0.997_{\pm 0.002}$ | $1.000_{\pm 0.000}$ | $0.504_{\pm 0.016}$ | $0.996_{\pm 0.005}$ | $-0.496_{\pm 0.016}$ | $0.024_{\pm 0.008}$ |
| | Reverse | $0.996_{\pm 0.003}$ | $1.000_{\pm 0.000}$ | $0.495_{\pm 0.016}$ | $0.996_{\pm 0.004}$ | $-0.505_{\pm 0.016}$ | $0.024_{\pm 0.008}$ |
| Local | A | $0.984_{\pm 0.004}$ | $1.000_{\pm 0.000}$ | $0.854_{\pm 0.022}$ | $0.872_{\pm 0.009}$ | $-0.146_{\pm 0.022}$ | $0.208_{\pm 0.008}$ |
| | B | $0.986_{\pm 0.003}$ | $1.000_{\pm 0.000}$ | $0.798_{\pm 0.009}$ | $0.873_{\pm 0.019}$ | $-0.202_{\pm 0.009}$ | $0.190_{\pm 0.019}$ |
| | Source | $0.982_{\pm 0.006}$ | $1.000_{\pm 0.000}$ | $0.827_{\pm 0.012}$ | $0.868_{\pm 0.002}$ | $-0.173_{\pm 0.012}$ | $0.192_{\pm 0.006}$ |
| | Random | $0.961_{\pm 0.007}$ | $1.000_{\pm 0.000}$ | $0.502_{\pm 0.012}$ | $0.961_{\pm 0.005}$ | $-0.498_{\pm 0.012}$ | $0.011_{\pm 0.003}$ |
| | Reverse | $0.956_{\pm 0.003}$ | $1.000_{\pm 0.000}$ | $0.483_{\pm 0.014}$ | $0.957_{\pm 0.004}$ | $-0.517_{\pm 0.014}$ | $0.011_{\pm 0.003}$ |
| Ent. | A | $0.833_{\pm 0.011}$ | $1.000_{\pm 0.000}$ | $0.828_{\pm 0.021}$ | $0.751_{\pm 0.014}$ | $-0.172_{\pm 0.021}$ | $0.207_{\pm 0.004}$ |
| | B | $0.824_{\pm 0.007}$ | $1.000_{\pm 0.000}$ | $0.800_{\pm 0.002}$ | $0.741_{\pm 0.016}$ | $-0.200_{\pm 0.002}$ | $0.199_{\pm 0.007}$ |
| | Source | $0.827_{\pm 0.010}$ | $1.000_{\pm 0.000}$ | $0.815_{\pm 0.019}$ | $0.740_{\pm 0.002}$ | $-0.185_{\pm 0.019}$ | $0.278_{\pm 0.007}$ |
| | Random | $0.886_{\pm 0.024}$ | $1.000_{\pm 0.000}$ | $0.509_{\pm 0.011}$ | $0.882_{\pm 0.018}$ | $-0.491_{\pm 0.011}$ | $0.130_{\pm 0.005}$ |
| | Reverse | $0.868_{\pm 0.023}$ | $1.000_{\pm 0.000}$ | $0.493_{\pm 0.010}$ | $0.880_{\pm 0.024}$ | $-0.507_{\pm 0.010}$ | $0.134_{\pm 0.004}$ |
| Reup. | A | $0.866_{\pm 0.022}$ | $0.000_{\pm 0.000}$ | $0.514_{\pm 0.055}$ | $0.617_{\pm 0.062}$ | $0.514_{\pm 0.055}$ | $0.286_{\pm 0.020}$ |
| | B | $0.860_{\pm 0.014}$ | $0.000_{\pm 0.000}$ | $0.513_{\pm 0.073}$ | $0.662_{\pm 0.055}$ | $0.513_{\pm 0.073}$ | $0.264_{\pm 0.008}$ |
| | Source | $0.858_{\pm 0.018}$ | $0.000_{\pm 0.000}$ | $0.505_{\pm 0.072}$ | $0.619_{\pm 0.076}$ | $0.505_{\pm 0.072}$ | $0.330_{\pm 0.027}$ |
| | Random | $0.574_{\pm 0.112}$ | $0.000_{\pm 0.000}$ | $0.508_{\pm 0.024}$ | $0.559_{\pm 0.093}$ | $0.508_{\pm 0.024}$ | $0.245_{\pm 0.066}$ |
| | Reverse | $0.569_{\pm 0.083}$ | $0.000_{\pm 0.000}$ | $0.499_{\pm 0.036}$ | $0.562_{\pm 0.104}$ | $0.499_{\pm 0.036}$ | $0.245_{\pm 0.073}$ |

Table 18: Q-FOS: fixed-checkpoint interventions; $F$ is probability sensitivity. Definitions: Appendix C.4.

| Model | Split | Original | Stable | Cue | Flip | $C - S$ | $F$ |
|---|---|---|---|---|---|---|---|
| LR | A | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.900_{\pm 0.009}$ | $1.000_{\pm 0.000}$ | $-0.100_{\pm 0.009}$ | $0.000_{\pm 0.000}$ |
| | B | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.893_{\pm 0.005}$ | $1.000_{\pm 0.000}$ | $-0.107_{\pm 0.005}$ | $0.000_{\pm 0.000}$ |
| | Source | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.895_{\pm 0.007}$ | $1.000_{\pm 0.000}$ | $-0.105_{\pm 0.007}$ | $0.000_{\pm 0.000}$ |
| | Random | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.507_{\pm 0.015}$ | $1.000_{\pm 0.000}$ | $-0.493_{\pm 0.015}$ | $0.000_{\pm 0.000}$ |
| | Reverse | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.097_{\pm 0.004}$ | $1.000_{\pm 0.000}$ | $-0.903_{\pm 0.004}$ | $0.000_{\pm 0.000}$ |
| MLP | A | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.818_{\pm 0.031}$ | $1.000_{\pm 0.000}$ | $-0.182_{\pm 0.031}$ | $0.046_{\pm 0.002}$ |
| | B | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.657_{\pm 0.143}$ | $1.000_{\pm 0.000}$ | $-0.343_{\pm 0.143}$ | $0.023_{\pm 0.010}$ |
| | Source | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.739_{\pm 0.079}$ | $1.000_{\pm 0.001}$ | $-0.261_{\pm 0.079}$ | $0.043_{\pm 0.025}$ |
| | Random | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.500_{\pm 0.008}$ | $1.000_{\pm 0.000}$ | $-0.500_{\pm 0.008}$ | $0.031_{\pm 0.012}$ |
| | Reverse | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.320_{\pm 0.142}$ | $1.000_{\pm 0.000}$ | $-0.680_{\pm 0.142}$ | $0.033_{\pm 0.017}$ |
| Local | A | $0.383_{\pm 0.033}$ | $0.337_{\pm 0.275}$ | $0.262_{\pm 0.021}$ | $0.235_{\pm 0.044}$ | $-0.075_{\pm 0.261}$ | $0.089_{\pm 0.017}$ |
| | B | $0.762_{\pm 0.053}$ | $0.328_{\pm 0.268}$ | $0.564_{\pm 0.053}$ | $0.127_{\pm 0.019}$ | $0.236_{\pm 0.256}$ | $0.171_{\pm 0.047}$ |
| | Source | $0.579_{\pm 0.014}$ | $0.340_{\pm 0.269}$ | $0.410_{\pm 0.033}$ | $0.530_{\pm 0.015}$ | $0.069_{\pm 0.245}$ | $0.157_{\pm 0.042}$ |
| | Random | $0.368_{\pm 0.003}$ | $0.337_{\pm 0.278}$ | $0.490_{\pm 0.015}$ | $0.438_{\pm 0.023}$ | $0.153_{\pm 0.288}$ | $0.054_{\pm 0.012}$ |
| | Reverse | $0.595_{\pm 0.036}$ | $0.341_{\pm 0.272}$ | $0.732_{\pm 0.038}$ | $0.780_{\pm 0.043}$ | $0.391_{\pm 0.309}$ | $0.061_{\pm 0.015}$ |
| Ent. | A | $0.501_{\pm 0.072}$ | $0.613_{\pm 0.205}$ | $0.558_{\pm 0.082}$ | $0.388_{\pm 0.124}$ | $-0.055_{\pm 0.253}$ | $0.084_{\pm 0.030}$ |
| | B | $0.652_{\pm 0.109}$ | $0.607_{\pm 0.193}$ | $0.435_{\pm 0.044}$ | $0.333_{\pm 0.164}$ | $-0.172_{\pm 0.227}$ | $0.120_{\pm 0.090}$ |
| | Source | $0.581_{\pm 0.027}$ | $0.606_{\pm 0.194}$ | $0.491_{\pm 0.076}$ | $0.563_{\pm 0.022}$ | $-0.116_{\pm 0.243}$ | $0.112_{\pm 0.084}$ |
| | Random | $0.455_{\pm 0.084}$ | $0.622_{\pm 0.187}$ | $0.522_{\pm 0.014}$ | $0.489_{\pm 0.057}$ | $-0.100_{\pm 0.175}$ | $0.042_{\pm 0.028}$ |
| | Reverse | $0.570_{\pm 0.021}$ | $0.610_{\pm 0.193}$ | $0.376_{\pm 0.106}$ | $0.674_{\pm 0.093}$ | $-0.234_{\pm 0.265}$ | $0.048_{\pm 0.034}$ |
| Reup. | A | $0.552_{\pm 0.029}$ | $0.657_{\pm 0.290}$ | $0.263_{\pm 0.202}$ | $0.691_{\pm 0.091}$ | $-0.394_{\pm 0.444}$ | $0.158_{\pm 0.010}$ |
| | B | $0.742_{\pm 0.053}$ | $0.665_{\pm 0.282}$ | $0.245_{\pm 0.209}$ | $0.275_{\pm 0.038}$ | $-0.420_{\pm 0.427}$ | $0.192_{\pm 0.036}$ |
| | Source | $0.653_{\pm 0.017}$ | $0.662_{\pm 0.286}$ | $0.256_{\pm 0.209}$ | $0.439_{\pm 0.032}$ | $-0.406_{\pm 0.441}$ | $0.163_{\pm 0.030}$ |
| | Random | $0.583_{\pm 0.036}$ | $0.671_{\pm 0.279}$ | $0.508_{\pm 0.011}$ | $0.472_{\pm 0.025}$ | $-0.163_{\pm 0.283}$ | $0.089_{\pm 0.013}$ |
| | Reverse | $0.452_{\pm 0.031}$ | $0.663_{\pm 0.282}$ | $0.305_{\pm 0.175}$ | $0.346_{\pm 0.084}$ | $-0.359_{\pm 0.434}$ | $0.086_{\pm 0.013}$ |

Table 19: Q-LINK: fixed-checkpoint interventions; $F$ is probability sensitivity. Definitions: Appendix C.4.

| Model | Split | Original | Stable | Cue | Flip | $C - S$ | $F$ |
|---|---|---|---|---|---|---|---|
| LR | A | $0.910_{\pm 0.003}$ | $0.522_{\pm 0.009}$ | $0.910_{\pm 0.003}$ | $0.089_{\pm 0.003}$ | $0.388_{\pm 0.012}$ | $0.772_{\pm 0.004}$ |
| | B | $0.882_{\pm 0.005}$ | $0.539_{\pm 0.014}$ | $0.883_{\pm 0.006}$ | $0.117_{\pm 0.006}$ | $0.343_{\pm 0.011}$ | $0.742_{\pm 0.003}$ |
| | Source | $0.903_{\pm 0.009}$ | $0.527_{\pm 0.012}$ | $0.903_{\pm 0.009}$ | $0.097_{\pm 0.008}$ | $0.375_{\pm 0.010}$ | $0.758_{\pm 0.003}$ |
| | Random | $0.506_{\pm 0.016}$ | $0.503_{\pm 0.005}$ | $0.506_{\pm 0.016}$ | $0.494_{\pm 0.016}$ | $0.004_{\pm 0.011}$ | $0.886_{\pm 0.001}$ |
| | Reverse | $0.098_{\pm 0.005}$ | $0.535_{\pm 0.007}$ | $0.098_{\pm 0.005}$ | $0.902_{\pm 0.005}$ | $-0.437_{\pm 0.012}$ | $0.889_{\pm 0.001}$ |
| MLP | A | $0.995_{\pm 0.006}$ | $0.999_{\pm 0.001}$ | $0.907_{\pm 0.008}$ | $0.939_{\pm 0.069}$ | $-0.093_{\pm 0.007}$ | $0.208_{\pm 0.093}$ |
| | B | $0.998_{\pm 0.004}$ | $1.000_{\pm 0.000}$ | $0.876_{\pm 0.011}$ | $0.974_{\pm 0.029}$ | $-0.124_{\pm 0.011}$ | $0.170_{\pm 0.076}$ |
| | Source | $0.997_{\pm 0.005}$ | $0.999_{\pm 0.002}$ | $0.898_{\pm 0.015}$ | $0.952_{\pm 0.051}$ | $-0.101_{\pm 0.014}$ | $0.190_{\pm 0.086}$ |
| | Random | $0.938_{\pm 0.069}$ | $1.000_{\pm 0.001}$ | $0.507_{\pm 0.016}$ | $0.925_{\pm 0.068}$ | $-0.493_{\pm 0.015}$ | $0.284_{\pm 0.113}$ |
| | Reverse | $0.877_{\pm 0.124}$ | $1.000_{\pm 0.000}$ | $0.098_{\pm 0.005}$ | $0.985_{\pm 0.014}$ | $-0.902_{\pm 0.005}$ | $0.286_{\pm 0.110}$ |
| Local | A | $0.909_{\pm 0.003}$ | $0.520_{\pm 0.031}$ | $0.910_{\pm 0.004}$ | $0.091_{\pm 0.005}$ | $0.390_{\pm 0.027}$ | $0.747_{\pm 0.025}$ |
| | B | $0.882_{\pm 0.005}$ | $0.509_{\pm 0.029}$ | $0.884_{\pm 0.004}$ | $0.118_{\pm 0.007}$ | $0.375_{\pm 0.025}$ | $0.741_{\pm 0.031}$ |
| | Source | $0.902_{\pm 0.008}$ | $0.514_{\pm 0.031}$ | $0.902_{\pm 0.008}$ | $0.099_{\pm 0.011}$ | $0.388_{\pm 0.023}$ | $0.743_{\pm 0.024}$ |
| | Random | $0.506_{\pm 0.016}$ | $0.501_{\pm 0.035}$ | $0.506_{\pm 0.016}$ | $0.494_{\pm 0.016}$ | $0.005_{\pm 0.028}$ | $0.930_{\pm 0.013}$ |
| | Reverse | $0.098_{\pm 0.005}$ | $0.516_{\pm 0.010}$ | $0.098_{\pm 0.005}$ | $0.902_{\pm 0.005}$ | $-0.418_{\pm 0.014}$ | $0.932_{\pm 0.014}$ |
| Ent. | A | $0.998_{\pm 0.001}$ | $0.999_{\pm 0.001}$ | $0.754_{\pm 0.151}$ | $0.995_{\pm 0.001}$ | $-0.245_{\pm 0.151}$ | $0.013_{\pm 0.012}$ |
| | B | $0.990_{\pm 0.014}$ | $0.999_{\pm 0.001}$ | $0.606_{\pm 0.189}$ | $0.939_{\pm 0.098}$ | $-0.393_{\pm 0.189}$ | $0.068_{\pm 0.089}$ |
| | Source | $0.995_{\pm 0.004}$ | $0.998_{\pm 0.002}$ | $0.702_{\pm 0.124}$ | $0.964_{\pm 0.054}$ | $-0.296_{\pm 0.123}$ | $0.041_{\pm 0.054}$ |
| | Random | $0.985_{\pm 0.015}$ | $0.998_{\pm 0.001}$ | $0.501_{\pm 0.012}$ | $0.983_{\pm 0.018}$ | $-0.497_{\pm 0.010}$ | $0.036_{\pm 0.046}$ |
| | Reverse | $0.986_{\pm 0.019}$ | $0.999_{\pm 0.001}$ | $0.335_{\pm 0.200}$ | $0.996_{\pm 0.002}$ | $-0.664_{\pm 0.200}$ | $0.033_{\pm 0.042}$ |
| Reup. | A | $0.985_{\pm 0.018}$ | $0.948_{\pm 0.084}$ | $0.311_{\pm 0.172}$ | $0.913_{\pm 0.083}$ | $-0.636_{\pm 0.182}$ | $0.124_{\pm 0.040}$ |
| | B | $0.980_{\pm 0.016}$ | $0.952_{\pm 0.077}$ | $0.554_{\pm 0.082}$ | $0.895_{\pm 0.082}$ | $-0.398_{\pm 0.160}$ | $0.120_{\pm 0.060}$ |
| | Source | $0.982_{\pm 0.016}$ | $0.952_{\pm 0.077}$ | $0.429_{\pm 0.109}$ | $0.903_{\pm 0.080}$ | $-0.522_{\pm 0.154}$ | $0.125_{\pm 0.050}$ |
| | Random | $0.901_{\pm 0.084}$ | $0.947_{\pm 0.089}$ | $0.501_{\pm 0.015}$ | $0.900_{\pm 0.082}$ | $-0.446_{\pm 0.096}$ | $0.151_{\pm 0.082}$ |
| | Reverse | $0.846_{\pm 0.122}$ | $0.955_{\pm 0.073}$ | $0.508_{\pm 0.036}$ | $0.942_{\pm 0.069}$ | $-0.447_{\pm 0.039}$ | $0.149_{\pm 0.073}$ |

Table 20 gives per-insertion channel parameters. Tiers do not represent equal physical severity across channels or calibrated device error rates.

Table 20: Per-insertion noise parameters. The same moderate and strong values define the corresponding noise-aware training groups.

| Channel | Weak | Moderate | Strong |
|---|---|---|---|
| Amplitude damping | 0.005 | 0.05 | 0.10 |
| Depolarizing | 0.005 | 0.08 | 0.15 |
| Phase damping | 0.005 | 0.50 | 0.60 |

Channels act on all qubits after each embedding, on each qubit after its trainable rotation, and on all qubits after each CNOT ring (or the corresponding local layer boundary). Two layers give five insertions per qubit for local/entangling and six for reuploading, unlike the theory's illustrative single global channel. Each held-out split uses up to 300 examples; ideal values may differ from the larger-sample clean overview (Figure 2).

**Architecture-specific effects across the suite.** Strong amplitude damping moves Q-SPHERE entangling reversal accuracy from $13.1 \pm 1.9\%$ to $50.6 \pm 2.2\%$ while ID falls from $86.4 \pm 1.0\%$ to $51.8 \pm 3.1\%$. Spirals-Signatures reuploading similarly approaches chance on both splits under strong depolarization. Reversal gains can accompany loss of the original rule without stable-feature recovery.

On Spurious-Fourier, strong depolarization preserves local reversal accuracy ($95.4 \pm 0.2\%$) but reduces entangling to $51.7 \pm 5.8\%$; strong phase damping instead preserves $87.2 \pm 2.2\%$. Q-FOS has weaker ideal reversal performance and no uniform preservation of a reliable stable predictor under strong noise.

On Q-LINK, local accuracy remains below 9% under strong noise. Entangling falls from $98.6 \pm 2.0\%$ ideally to $72.2 \pm 15.8\%$, $86.0 \pm 4.8\%$, and $69.8 \pm 19.8\%$ under strong amplitude, depolarizing, and phase noise; reuploading falls from $84.0 \pm 12.3\%$ to $79.2 \pm 6.7\%$, $52.1 \pm 1.3\%$, and $52.1 \pm 1.3\%$. Low susceptibility need not imply useful OOD performance. The pooled mode visualization remains in Figure 3 (bottom right); tables separate architecture, split, and channel.

**Sampling and the scope of the physical-noise comparison.** A single shot budget does not test margin-conditioned disagreement or scaling in Proposition 8. Physical-noise accuracy does not measure trace distances (Proposition 2), entanglement, or optimal affine recoverability (Appendix A.8). Masking caveats remain (Appendix C.4).

**Reading the detailed tables.** Per-task ideal/shot controls and three channel tables retain all VQCs, held-out splits, and input modes: 1980 means from 5940 seed evaluations. Original is the full competing-feature task; Stable, Cue, and Flip denote stable-only, shortcut-only, and shortcut-flip. The three-task main table selects these evaluations; role reversal is separate (Appendix C.8).

| | | ID | | Random | | OOD reversed | |
|---|---|---|---|---|---|---|---|
| Input | VQC | Ideal | 1024 shots | Ideal | 1024 shots | Ideal | 1024 shots |
| Original | Local | $89.9_{\pm 0.2}$ | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $50.6_{\pm 1.6}$ | $11.2_{\pm 2.2}$ | $11.2_{\pm 2.2}$ |
| | Entangling | $86.4_{\pm 1.0}$ | $86.1_{\pm 1.3}$ | $50.2_{\pm 2.4}$ | $50.2_{\pm 2.2}$ | $13.1_{\pm 1.9}$ | $12.9_{\pm 2.0}$ |
| | Reuploading | $85.3_{\pm 0.9}$ | $83.8_{\pm 1.4}$ | $57.7_{\pm 1.5}$ | $57.3_{\pm 0.6}$ | $28.8_{\pm 3.3}$ | $28.6_{\pm 3.3}$ |
| Stable-only | Local | $52.3_{\pm 6.7}$ | $51.7_{\pm 6.2}$ | $49.4_{\pm 5.3}$ | $50.2_{\pm 6.0}$ | $52.8_{\pm 2.5}$ | $52.7_{\pm 3.1}$ |
| | Entangling | $49.8_{\pm 1.0}$ | $49.3_{\pm 0.9}$ | $52.1_{\pm 3.3}$ | $51.9_{\pm 3.1}$ | $47.7_{\pm 2.1}$ | $47.2_{\pm 2.0}$ |
| | Reuploading | $64.6_{\pm 4.7}$ | $63.9_{\pm 4.9}$ | $64.4_{\pm 3.1}$ | $63.4_{\pm 2.4}$ | $62.2_{\pm 3.1}$ | $61.7_{\pm 1.5}$ |
| Shortcut-only | Local | $89.9_{\pm 0.2}$ | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $50.6_{\pm 1.6}$ | $11.2_{\pm 2.2}$ | $11.2_{\pm 2.2}$ |
| | Entangling | $87.4_{\pm 1.6}$ | $87.7_{\pm 1.2}$ | $50.9_{\pm 2.0}$ | $50.7_{\pm 2.3}$ | $12.4_{\pm 2.2}$ | $12.3_{\pm 2.4}$ |
| | Reuploading | $72.2_{\pm 5.5}$ | $71.7_{\pm 5.6}$ | $50.0_{\pm 0.3}$ | $50.1_{\pm 0.8}$ | $26.2_{\pm 5.1}$ | $26.8_{\pm 6.3}$ |
| Shortcut-flip | Local | $10.1_{\pm 0.2}$ | $10.1_{\pm 0.2}$ | $49.3_{\pm 1.5}$ | $49.3_{\pm 1.5}$ | $88.9_{\pm 2.0}$ | $88.9_{\pm 2.0}$ |
| | Entangling | $14.0_{\pm 3.1}$ | $14.4_{\pm 3.4}$ | $48.8_{\pm 2.1}$ | $48.6_{\pm 2.3}$ | $86.7_{\pm 2.7}$ | $87.0_{\pm 2.8}$ |
| | Reuploading | $28.7_{\pm 0.6}$ | $28.3_{\pm 1.0}$ | $56.8_{\pm 4.3}$ | $56.6_{\pm 4.2}$ | $84.1_{\pm 2.3}$ | $84.0_{\pm 3.2}$ |

Table 21: Q-SPHERE: ideal analytic versus 1024-shot accuracy (%) for fixed clean-trained checkpoints. ID is held out; circuits are otherwise ideal.

| | | Weak $p = 0.005$ | | | Moderate $p = 0.05$ | | | Strong $p = 0.1$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | OOD | | | OOD | | | OOD |
| Input | VQC | ID | Random | reversed | ID | Random | reversed | ID | Random | reversed |
| Original | Local | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $11.2_{\pm 2.2}$ | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $11.2_{\pm 2.2}$ | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $11.2_{\pm 2.2}$ |
| | Entangling | $86.8_{\pm 0.7}$ | $50.3_{\pm 2.1}$ | $12.6_{\pm 2.0}$ | $88.1_{\pm 0.7}$ | $51.0_{\pm 1.2}$ | $12.2_{\pm 1.7}$ | $51.8_{\pm 3.1}$ | $52.4_{\pm 5.1}$ | $50.6_{\pm 2.2}$ |
| | Reuploading | $85.6_{\pm 1.2}$ | $57.7_{\pm 1.5}$ | $28.6_{\pm 3.2}$ | $75.1_{\pm 4.5}$ | $55.1_{\pm 1.6}$ | $33.3_{\pm 3.7}$ | $52.2_{\pm 2.7}$ | $54.9_{\pm 1.3}$ | $51.2_{\pm 1.3}$ |
| Stable-only | Local | $52.0_{\pm 6.5}$ | $50.0_{\pm 5.7}$ | $52.9_{\pm 2.7}$ | $50.8_{\pm 5.6}$ | $48.4_{\pm 5.0}$ | $50.2_{\pm 3.0}$ | $49.7_{\pm 4.6}$ | $47.6_{\pm 4.6}$ | $48.8_{\pm 1.2}$ |
| | Entangling | $50.8_{\pm 1.3}$ | $50.8_{\pm 2.7}$ | $49.0_{\pm 1.5}$ | $51.8_{\pm 3.1}$ | $52.4_{\pm 5.1}$ | $50.6_{\pm 2.2}$ | $51.8_{\pm 3.1}$ | $52.4_{\pm 5.1}$ | $50.6_{\pm 2.2}$ |
| | Reuploading | $64.2_{\pm 6.0}$ | $65.1_{\pm 3.8}$ | $61.7_{\pm 4.2}$ | $54.9_{\pm 4.2}$ | $59.1_{\pm 5.4}$ | $56.1_{\pm 4.6}$ | $51.8_{\pm 3.1}$ | $54.7_{\pm 1.5}$ | $51.4_{\pm 1.5}$ |
| Shortcut-only | Local | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $11.2_{\pm 2.2}$ | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $11.2_{\pm 2.2}$ | $89.9_{\pm 0.2}$ | $50.7_{\pm 1.5}$ | $11.2_{\pm 2.2}$ |
| | Entangling | $87.8_{\pm 1.0}$ | $50.8_{\pm 2.2}$ | $12.2_{\pm 2.4}$ | $89.4_{\pm 0.4}$ | $51.1_{\pm 1.2}$ | $11.6_{\pm 2.0}$ | $77.1_{\pm 3.1}$ | $52.6_{\pm 1.8}$ | $22.9_{\pm 1.3}$ |
| | Reuploading | $71.9_{\pm 5.7}$ | $50.3_{\pm 1.5}$ | $26.7_{\pm 4.7}$ | $58.0_{\pm 11.0}$ | $55.1_{\pm 0.4}$ | $42.1_{\pm 10.1}$ | $51.8_{\pm 3.1}$ | $54.7_{\pm 1.5}$ | $51.4_{\pm 1.5}$ |
| Shortcut-flip | Local | $10.1_{\pm 0.2}$ | $49.3_{\pm 1.5}$ | $88.9_{\pm 2.0}$ | $10.1_{\pm 0.2}$ | $49.3_{\pm 1.5}$ | $88.9_{\pm 2.0}$ | $10.1_{\pm 0.2}$ | $49.4_{\pm 1.6}$ | $88.9_{\pm 2.0}$ |
| | Entangling | $13.3_{\pm 2.4}$ | $48.2_{\pm 1.6}$ | $87.4_{\pm 2.5}$ | $13.2_{\pm 0.8}$ | $49.3_{\pm 0.3}$ | $87.2_{\pm 2.0}$ | $51.7_{\pm 3.2}$ | $52.4_{\pm 5.1}$ | $50.6_{\pm 2.2}$ |
| | Reuploading | $28.6_{\pm 1.7}$ | $56.7_{\pm 4.7}$ | $83.8_{\pm 2.5}$ | $36.4_{\pm 4.5}$ | $56.2_{\pm 3.7}$ | $74.4_{\pm 4.8}$ | $51.7_{\pm 3.2}$ | $54.9_{\pm 1.5}$ | $52.1_{\pm 1.9}$ |

Table 22: Q-SPHERE: fixed-checkpoint accuracy (%) under amplitude damping. Ideal/shot references: Table 21.

| | | Weak $p = 0.005$ | | | Moderate $p = 0.08$ | | | Strong $p = 0.15$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | OOD | | | OOD | | | OOD |
| Input | VQC | ID | Random | reversed | ID | Random | reversed | ID | Random | reversed |
| Original | Local | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $11.2_{\pm 2.2}$ | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $11.2_{\pm 2.2}$ | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $11.1_{\pm 2.0}$ |
| | Entangling | $86.6_{\pm 1.1}$ | $50.2_{\pm 2.4}$ | $13.0_{\pm 2.1}$ | $76.7_{\pm 8.9}$ | $49.3_{\pm 1.5}$ | $21.6_{\pm 9.5}$ | $60.3_{\pm 13.5}$ | $50.1_{\pm 5.4}$ | $40.0_{\pm 16.2}$ |
| | Reuploading | $84.9_{\pm 1.3}$ | $57.7_{\pm 1.5}$ | $28.4_{\pm 3.2}$ | $66.6_{\pm 10.5}$ | $54.6_{\pm 4.6}$ | $43.7_{\pm 2.7}$ | $52.3_{\pm 9.1}$ | $49.0_{\pm 6.8}$ | $49.4_{\pm 0.7}$ |
| Stable-only | Local | $52.3_{\pm 6.7}$ | $49.7_{\pm 5.5}$ | $53.0_{\pm 2.7}$ | $52.8_{\pm 7.5}$ | $49.9_{\pm 6.0}$ | $51.8_{\pm 4.6}$ | $49.3_{\pm 4.2}$ | $47.8_{\pm 5.0}$ | $51.0_{\pm 4.9}$ |
| | Entangling | $49.7_{\pm 0.7}$ | $52.0_{\pm 2.6}$ | $47.3_{\pm 2.4}$ | $52.1_{\pm 2.8}$ | $52.2_{\pm 5.0}$ | $50.0_{\pm 1.3}$ | $53.4_{\pm 3.0}$ | $51.3_{\pm 4.8}$ | $50.1_{\pm 1.5}$ |
| | Reuploading | $64.6_{\pm 4.8}$ | $64.8_{\pm 3.1}$ | $62.2_{\pm 3.1}$ | $51.1_{\pm 2.2}$ | $51.4_{\pm 9.2}$ | $53.6_{\pm 4.6}$ | $50.3_{\pm 5.8}$ | $50.0_{\pm 8.5}$ | $52.4_{\pm 5.4}$ |
| Shortcut-only | Local | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $11.2_{\pm 2.2}$ | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $11.1_{\pm 2.0}$ | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $11.1_{\pm 2.0}$ |
| | Entangling | $87.4_{\pm 1.6}$ | $50.9_{\pm 2.0}$ | $12.4_{\pm 2.2}$ | $87.3_{\pm 1.5}$ | $50.1_{\pm 2.5}$ | $11.8_{\pm 2.4}$ | $62.9_{\pm 17.9}$ | $50.0_{\pm 5.5}$ | $37.7_{\pm 20.2}$ |
| | Reuploading | $72.6_{\pm 5.6}$ | $49.8_{\pm 0.5}$ | $26.2_{\pm 5.7}$ | $56.6_{\pm 10.3}$ | $46.6_{\pm 3.4}$ | $41.4_{\pm 6.7}$ | $49.0_{\pm 3.9}$ | $48.4_{\pm 5.9}$ | $48.4_{\pm 1.7}$ |
| Shortcut-flip | Local | $10.1_{\pm 0.2}$ | $49.3_{\pm 1.5}$ | $88.9_{\pm 2.0}$ | $10.1_{\pm 0.2}$ | $49.3_{\pm 1.5}$ | $88.9_{\pm 2.0}$ | $10.1_{\pm 0.2}$ | $49.3_{\pm 1.5}$ | $88.8_{\pm 2.2}$ |
| | Entangling | $14.0_{\pm 3.1}$ | $48.8_{\pm 2.1}$ | $86.6_{\pm 2.8}$ | $24.3_{\pm 6.2}$ | $49.2_{\pm 0.7}$ | $77.3_{\pm 9.5}$ | $43.6_{\pm 16.0}$ | $51.0_{\pm 4.9}$ | $58.8_{\pm 16.4}$ |
| | Reuploading | $28.8_{\pm 1.2}$ | $56.9_{\pm 4.4}$ | $84.2_{\pm 2.2}$ | $42.2_{\pm 6.1}$ | $54.4_{\pm 8.1}$ | $68.1_{\pm 8.9}$ | $47.6_{\pm 2.7}$ | $50.2_{\pm 8.9}$ | $54.3_{\pm 8.7}$ |

Table 23: Q-SPHERE: fixed-checkpoint accuracy (%) under depolarizing. Ideal/shot references: Table 21.

### C.6 Noise-aware training and common evaluation conditions

M/S denote moderate/strong noise. Nine training groups comprise clean and moderate/strong amplitude, depolarizing, phase, and mixed noise, with Table 20 parameters and

| Input | VQC | Weak $p = 0.005$ | | | Moderate $p = 0.5$ | | | Strong $p = 0.6$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ID | Random | OOD reversed | ID | Random | OOD reversed | ID | Random | OOD reversed |
| Original | Local | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $11.2_{\pm 2.2}$ | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $11.2_{\pm 2.2}$ | $89.8_{\pm 0.4}$ | $50.6_{\pm 1.6}$ | $11.2_{\pm 2.2}$ |
| | Entangling | $86.4_{\pm 1.0}$ | $50.2_{\pm 2.4}$ | $13.1_{\pm 1.9}$ | $85.3_{\pm 0.3}$ | $51.0_{\pm 2.3}$ | $13.1_{\pm 2.1}$ | $83.6_{\pm 1.5}$ | $51.3_{\pm 0.7}$ | $15.2_{\pm 2.8}$ |
| | Reuploading | $85.3_{\pm 0.9}$ | $57.7_{\pm 1.5}$ | $28.4_{\pm 3.1}$ | $62.2_{\pm 13.4}$ | $50.7_{\pm 4.7}$ | $44.7_{\pm 2.4}$ | $56.9_{\pm 14.9}$ | $48.9_{\pm 6.4}$ | $47.8_{\pm 2.0}$ |
| Stable-only | Local | $52.2_{\pm 6.6}$ | $49.4_{\pm 5.3}$ | $52.9_{\pm 2.7}$ | $50.2_{\pm 5.5}$ | $47.3_{\pm 5.0}$ | $48.0_{\pm 2.6}$ | $48.8_{\pm 4.0}$ | $46.4_{\pm 3.7}$ | $48.0_{\pm 1.8}$ |
| | Entangling | $49.6_{\pm 1.1}$ | $52.0_{\pm 3.5}$ | $47.4_{\pm 1.9}$ | $51.8_{\pm 3.1}$ | $52.4_{\pm 5.1}$ | $50.6_{\pm 2.2}$ | $51.8_{\pm 3.1}$ | $52.4_{\pm 5.1}$ | $50.6_{\pm 2.2}$ |
| | Reuploading | $64.7_{\pm 4.8}$ | $64.6_{\pm 3.0}$ | $62.6_{\pm 3.1}$ | $55.3_{\pm 7.8}$ | $56.4_{\pm 12.0}$ | $58.9_{\pm 7.5}$ | $54.0_{\pm 9.3}$ | $54.2_{\pm 12.7}$ | $55.9_{\pm 9.1}$ |
| Shortcut-only | Local | $89.9_{\pm 0.2}$ | $50.6_{\pm 1.6}$ | $11.2_{\pm 2.2}$ | $89.8_{\pm 0.4}$ | $50.7_{\pm 1.5}$ | $11.2_{\pm 2.2}$ | $89.8_{\pm 0.4}$ | $50.7_{\pm 1.5}$ | $11.3_{\pm 2.3}$ |
| | Entangling | $87.6_{\pm 1.4}$ | $50.9_{\pm 2.0}$ | $12.4_{\pm 2.2}$ | $87.9_{\pm 1.2}$ | $51.3_{\pm 0.9}$ | $11.9_{\pm 2.2}$ | $87.6_{\pm 2.0}$ | $51.7_{\pm 0.9}$ | $13.1_{\pm 3.4}$ |
| | Reuploading | $72.0_{\pm 5.2}$ | $50.0_{\pm 0.3}$ | $26.3_{\pm 5.4}$ | $54.0_{\pm 8.0}$ | $48.8_{\pm 4.0}$ | $44.4_{\pm 5.9}$ | $52.3_{\pm 8.2}$ | $48.3_{\pm 4.3}$ | $46.4_{\pm 4.5}$ |
| Shortcut-flip | Local | $10.1_{\pm 0.2}$ | $49.3_{\pm 1.5}$ | $88.9_{\pm 2.0}$ | $10.1_{\pm 0.2}$ | $49.3_{\pm 1.5}$ | $88.9_{\pm 2.0}$ | $10.1_{\pm 0.2}$ | $49.3_{\pm 1.5}$ | $88.9_{\pm 2.0}$ |
| | Entangling | $14.0_{\pm 3.1}$ | $48.9_{\pm 1.9}$ | $86.7_{\pm 2.7}$ | $14.8_{\pm 0.8}$ | $48.7_{\pm 2.0}$ | $86.2_{\pm 2.7}$ | $17.4_{\pm 0.8}$ | $48.4_{\pm 2.5}$ | $84.0_{\pm 2.9}$ |
| | Reuploading | $28.9_{\pm 0.5}$ | $57.0_{\pm 4.2}$ | $84.2_{\pm 2.5}$ | $41.8_{\pm 7.0}$ | $51.1_{\pm 7.5}$ | $62.3_{\pm 13.9}$ | $43.3_{\pm 6.5}$ | $50.1_{\pm 8.0}$ | $57.3_{\pm 14.1}$ |

Table 24: Q-SPHERE: fixed-checkpoint accuracy (%) under phase damping. Ideal/shot references: Table 21.

| Input | VQC | ID | | Random | | OOD reversed | |
|---|---|---|---|---|---|---|---|
| | | Ideal | 1024 shots | Ideal | 1024 shots | Ideal | 1024 shots |
| Original | Local | $90.8_{\pm 0.7}$ | $90.8_{\pm 0.7}$ | $49.6_{\pm 2.5}$ | $49.6_{\pm 2.5}$ | $9.3_{\pm 0.9}$ | $9.3_{\pm 0.9}$ |
| | Entangling | $88.7_{\pm 2.2}$ | $87.8_{\pm 3.5}$ | $50.0_{\pm 2.7}$ | $50.2_{\pm 3.0}$ | $12.7_{\pm 3.5}$ | $13.2_{\pm 4.0}$ |
| | Reuploading | $86.7_{\pm 1.3}$ | $86.4_{\pm 1.9}$ | $52.0_{\pm 2.3}$ | $51.7_{\pm 2.4}$ | $18.0_{\pm 1.8}$ | $18.3_{\pm 1.8}$ |
| Stable-only | Local | $56.3_{\pm 5.2}$ | $55.0_{\pm 1.5}$ | $55.2_{\pm 2.5}$ | $55.8_{\pm 3.0}$ | $55.6_{\pm 4.8}$ | $54.8_{\pm 3.2}$ |
| | Entangling | $50.0_{\pm 3.4}$ | $51.0_{\pm 2.6}$ | $50.7_{\pm 4.0}$ | $51.8_{\pm 2.8}$ | $48.7_{\pm 4.0}$ | $50.3_{\pm 2.3}$ |
| | Reuploading | $52.9_{\pm 4.0}$ | $52.7_{\pm 3.7}$ | $54.0_{\pm 3.5}$ | $53.0_{\pm 3.3}$ | $55.1_{\pm 1.1}$ | $55.4_{\pm 0.7}$ |
| Shortcut-only | Local | $90.8_{\pm 0.7}$ | $90.8_{\pm 0.7}$ | $49.6_{\pm 2.5}$ | $49.6_{\pm 2.5}$ | $9.3_{\pm 0.9}$ | $9.3_{\pm 0.9}$ |
| | Entangling | $90.8_{\pm 0.7}$ | $90.8_{\pm 0.7}$ | $49.9_{\pm 2.7}$ | $49.9_{\pm 2.7}$ | $9.3_{\pm 0.9}$ | $9.3_{\pm 0.9}$ |
| | Reuploading | $90.2_{\pm 1.1}$ | $89.8_{\pm 1.6}$ | $49.7_{\pm 2.2}$ | $49.9_{\pm 2.2}$ | $9.7_{\pm 0.9}$ | $9.6_{\pm 0.8}$ |
| Shortcut-flip | Local | $9.2_{\pm 0.7}$ | $9.2_{\pm 0.7}$ | $50.4_{\pm 2.5}$ | $50.4_{\pm 2.5}$ | $90.7_{\pm 0.9}$ | $90.7_{\pm 0.9}$ |
| | Entangling | $13.3_{\pm 3.8}$ | $14.0_{\pm 4.7}$ | $52.0_{\pm 1.7}$ | $51.2_{\pm 2.0}$ | $89.3_{\pm 1.7}$ | $88.9_{\pm 2.2}$ |
| | Reuploading | $17.7_{\pm 3.0}$ | $17.3_{\pm 3.2}$ | $51.2_{\pm 4.1}$ | $51.3_{\pm 4.0}$ | $87.7_{\pm 2.5}$ | $87.7_{\pm 2.1}$ |

Table 25: Spirals-Signatures: ideal analytic versus 1024-shot accuracy (%) for fixed clean-trained checkpoints. ID is held out; circuits are otherwise ideal.

| Input | VQC | Weak $p = 0.005$ | | | Moderate $p = 0.05$ | | | Strong $p = 0.1$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ID | Random | OOD reversed | ID | Random | OOD reversed | ID | Random | OOD reversed |
| Original | Local | $90.8_{\pm 0.7}$ | $49.6_{\pm 2.5}$ | $9.3_{\pm 0.9}$ | $90.8_{\pm 0.7}$ | $49.6_{\pm 2.5}$ | $9.3_{\pm 0.9}$ | $90.7_{\pm 0.6}$ | $49.3_{\pm 2.5}$ | $9.6_{\pm 1.0}$ |
| | Entangling | $89.0_{\pm 1.9}$ | $50.3_{\pm 3.0}$ | $12.6_{\pm 3.4}$ | $90.0_{\pm 0.3}$ | $49.9_{\pm 3.3}$ | $10.2_{\pm 0.8}$ | $69.3_{\pm 17.4}$ | $53.9_{\pm 2.9}$ | $31.3_{\pm 18.9}$ |
| | Reuploading | $86.6_{\pm 1.7}$ | $52.0_{\pm 2.5}$ | $17.6_{\pm 1.4}$ | $86.0_{\pm 4.0}$ | $51.7_{\pm 2.3}$ | $17.0_{\pm 0.9}$ | $76.1_{\pm 8.6}$ | $50.6_{\pm 1.3}$ | $27.4_{\pm 10.3}$ |
| Stable-only | Local | $55.8_{\pm 2.4}$ | $57.1_{\pm 2.2}$ | $56.6_{\pm 0.5}$ | $52.8_{\pm 3.4}$ | $49.9_{\pm 2.5}$ | $52.2_{\pm 1.0}$ | $51.3_{\pm 0.9}$ | $48.1_{\pm 1.6}$ | $51.1_{\pm 1.0}$ |
| | Entangling | $51.3_{\pm 3.2}$ | $51.4_{\pm 3.4}$ | $51.3_{\pm 2.3}$ | $51.0_{\pm 3.2}$ | $52.1_{\pm 7.2}$ | $51.2_{\pm 5.0}$ | $49.3_{\pm 1.2}$ | $52.1_{\pm 1.9}$ | $49.8_{\pm 1.7}$ |
| | Reuploading | $52.7_{\pm 4.3}$ | $54.0_{\pm 3.2}$ | $55.0_{\pm 1.5}$ | $47.0_{\pm 5.2}$ | $48.9_{\pm 1.8}$ | $48.1_{\pm 4.8}$ | $48.0_{\pm 0.9}$ | $49.4_{\pm 2.0}$ | $44.6_{\pm 0.8}$ |
| Shortcut-only | Local | $90.8_{\pm 0.7}$ | $49.6_{\pm 2.5}$ | $9.3_{\pm 0.9}$ | $90.8_{\pm 0.7}$ | $49.6_{\pm 2.5}$ | $9.3_{\pm 0.9}$ | $90.7_{\pm 0.6}$ | $49.3_{\pm 2.7}$ | $9.6_{\pm 1.3}$ |
| | Entangling | $90.8_{\pm 0.7}$ | $49.9_{\pm 2.7}$ | $9.3_{\pm 0.9}$ | $89.2_{\pm 2.0}$ | $50.6_{\pm 4.4}$ | $12.1_{\pm 3.4}$ | $66.4_{\pm 20.3}$ | $51.8_{\pm 1.2}$ | $31.9_{\pm 19.1}$ |
| | Reuploading | $90.2_{\pm 1.1}$ | $49.7_{\pm 2.2}$ | $9.7_{\pm 0.9}$ | $90.0_{\pm 2.0}$ | $49.7_{\pm 2.5}$ | $9.2_{\pm 0.7}$ | $61.9_{\pm 23.5}$ | $52.0_{\pm 1.5}$ | $36.1_{\pm 22.3}$ |
| Shortcut-flip | Local | $9.2_{\pm 0.7}$ | $50.4_{\pm 2.5}$ | $90.7_{\pm 0.9}$ | $9.2_{\pm 0.7}$ | $50.4_{\pm 2.5}$ | $90.7_{\pm 0.9}$ | $9.8_{\pm 0.8}$ | $50.6_{\pm 2.3}$ | $90.4_{\pm 0.7}$ |
| | Entangling | $13.2_{\pm 3.7}$ | $51.3_{\pm 1.7}$ | $89.6_{\pm 1.6}$ | $11.1_{\pm 1.7}$ | $50.9_{\pm 2.5}$ | $89.1_{\pm 1.9}$ | $30.1_{\pm 17.2}$ | $53.4_{\pm 3.3}$ | $71.1_{\pm 17.1}$ |
| | Reuploading | $17.8_{\pm 3.0}$ | $51.1_{\pm 3.5}$ | $87.9_{\pm 2.2}$ | $18.4_{\pm 1.7}$ | $51.0_{\pm 4.0}$ | $88.2_{\pm 2.3}$ | $27.7_{\pm 5.6}$ | $51.2_{\pm 2.2}$ | $76.0_{\pm 8.9}$ |

Table 26: Spirals-Signatures: fixed-checkpoint accuracy (%) under amplitude damping. Ideal/shot references: Table 25.

| Input | VQC | Weak $p = 0.005$ | | | Moderate $p = 0.08$ | | | Strong $p = 0.15$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ID | Random | OOD reversed | ID | Random | OOD reversed | ID | Random | OOD reversed |
| Original | Local | $90.8_{\pm 0.7}$ | $49.6_{\pm 2.5}$ | $9.3_{\pm 0.9}$ | $90.8_{\pm 0.7}$ | $49.6_{\pm 2.5}$ | $9.3_{\pm 0.9}$ | $90.8_{\pm 0.7}$ | $49.6_{\pm 2.5}$ | $9.3_{\pm 0.9}$ |
| | Entangling | $88.7_{\pm 2.2}$ | $50.2_{\pm 2.9}$ | $12.7_{\pm 3.5}$ | $83.7_{\pm 2.4}$ | $49.1_{\pm 3.7}$ | $17.1_{\pm 1.1}$ | $62.8_{\pm 13.0}$ | $50.8_{\pm 3.2}$ | $38.1_{\pm 16.4}$ |
| | Reuploading | $86.2_{\pm 1.5}$ | $51.9_{\pm 2.4}$ | $18.1_{\pm 1.3}$ | $56.7_{\pm 11.3}$ | $51.2_{\pm 4.6}$ | $43.9_{\pm 11.3}$ | $50.9_{\pm 1.5}$ | $48.1_{\pm 1.6}$ | $50.2_{\pm 1.7}$ |
| Stable-only | Local | $56.4_{\pm 5.4}$ | $55.3_{\pm 2.6}$ | $55.7_{\pm 5.0}$ | $57.2_{\pm 0.5}$ | $56.1_{\pm 3.3}$ | $56.8_{\pm 0.8}$ | $56.6_{\pm 1.5}$ | $56.0_{\pm 4.0}$ | $57.0_{\pm 4.3}$ |
| | Entangling | $49.4_{\pm 3.2}$ | $50.1_{\pm 4.2}$ | $48.0_{\pm 3.7}$ | $48.8_{\pm 2.2}$ | $48.1_{\pm 1.6}$ | $48.4_{\pm 4.0}$ | $50.6_{\pm 1.1}$ | $48.3_{\pm 1.5}$ | $49.7_{\pm 2.0}$ |
| | Reuploading | $51.6_{\pm 3.0}$ | $53.3_{\pm 3.0}$ | $54.1_{\pm 1.0}$ | $51.0_{\pm 1.7}$ | $48.7_{\pm 1.3}$ | $50.2_{\pm 1.7}$ | $50.9_{\pm 1.5}$ | $48.1_{\pm 1.6}$ | $50.2_{\pm 1.7}$ |
| Shortcut-only | Local | $90.8_{\pm 0.7}$ | $49.6_{\pm 2.5}$ | $9.3_{\pm 0.9}$ | $90.8_{\pm 0.7}$ | $49.6_{\pm 2.5}$ | $9.3_{\pm 0.9}$ | $90.8_{\pm 0.7}$ | $49.6_{\pm 2.5}$ | $9.3_{\pm 0.9}$ |
| | Entangling | $90.8_{\pm 0.7}$ | $49.9_{\pm 2.7}$ | $9.3_{\pm 0.9}$ | $90.7_{\pm 0.9}$ | $49.7_{\pm 2.5}$ | $9.3_{\pm 0.9}$ | $68.4_{\pm 20.5}$ | $49.9_{\pm 2.2}$ | $33.4_{\pm 21.6}$ |
| | Reuploading | $90.0_{\pm 1.5}$ | $49.8_{\pm 2.2}$ | $9.8_{\pm 0.7}$ | $60.0_{\pm 17.1}$ | $49.8_{\pm 2.3}$ | $40.1_{\pm 17.8}$ | $50.9_{\pm 1.5}$ | $48.1_{\pm 1.6}$ | $50.2_{\pm 1.7}$ |
| Shortcut-flip | Local | $9.2_{\pm 0.7}$ | $50.4_{\pm 2.5}$ | $90.7_{\pm 0.9}$ | $9.2_{\pm 0.7}$ | $50.4_{\pm 2.5}$ | $90.7_{\pm 0.9}$ | $9.2_{\pm 0.7}$ | $50.4_{\pm 2.5}$ | $90.7_{\pm 0.9}$ |
| | Entangling | $13.4_{\pm 4.0}$ | $51.7_{\pm 2.0}$ | $89.1_{\pm 1.9}$ | $17.4_{\pm 1.4}$ | $50.6_{\pm 2.0}$ | $86.1_{\pm 3.3}$ | $37.7_{\pm 16.9}$ | $48.1_{\pm 3.6}$ | $63.6_{\pm 15.9}$ |
| | Reuploading | $18.1_{\pm 3.0}$ | $50.9_{\pm 3.7}$ | $87.9_{\pm 2.5}$ | $43.6_{\pm 11.5}$ | $47.7_{\pm 2.2}$ | $56.9_{\pm 11.5}$ | $50.9_{\pm 1.5}$ | $48.1_{\pm 1.6}$ | $50.2_{\pm 1.7}$ |

Table 27: Spirals-Signatures: fixed-checkpoint accuracy (%) under depolarizing. Ideal/shot references: Table 25.

| Input | VQC | Weak $p = 0.005$ | | | Moderate $p = 0.5$ | | | Strong $p = 0.6$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ID | Random | OOD reversed | ID | Random | OOD reversed | ID | Random | OOD reversed |
| Original | Local | $90.8_{\pm 0.7}$ | $49.6_{\pm 2.5}$ | $9.3_{\pm 0.9}$ | $90.6_{\pm 0.5}$ | $49.3_{\pm 2.5}$ | $9.6_{\pm 1.0}$ | $90.2_{\pm 0.7}$ | $48.9_{\pm 2.8}$ | $10.0_{\pm 1.2}$ |
| | Entangling | $88.7_{\pm 2.2}$ | $50.0_{\pm 2.7}$ | $12.7_{\pm 3.5}$ | $87.6_{\pm 0.8}$ | $50.8_{\pm 0.8}$ | $15.0_{\pm 0.9}$ | $85.6_{\pm 2.5}$ | $51.4_{\pm 1.2}$ | $18.1_{\pm 4.9}$ |
| | Reuploading | $86.6_{\pm 1.5}$ | $52.0_{\pm 2.3}$ | $18.1_{\pm 1.8}$ | $65.0_{\pm 12.9}$ | $52.0_{\pm 4.2}$ | $38.9_{\pm 14.0}$ | $61.6_{\pm 11.2}$ | $51.3_{\pm 3.5}$ | $40.6_{\pm 12.2}$ |
| Stable-only | Local | $56.4_{\pm 4.4}$ | $55.2_{\pm 2.2}$ | $55.8_{\pm 4.0}$ | $51.3_{\pm 0.9}$ | $48.1_{\pm 1.6}$ | $51.1_{\pm 1.0}$ | $51.3_{\pm 0.9}$ | $48.1_{\pm 1.6}$ | $51.1_{\pm 1.0}$ |
| | Entangling | $49.9_{\pm 3.5}$ | $50.2_{\pm 4.3}$ | $48.9_{\pm 3.7}$ | $51.2_{\pm 6.0}$ | $49.9_{\pm 3.5}$ | $53.0_{\pm 6.8}$ | $48.6_{\pm 4.2}$ | $48.2_{\pm 2.2}$ | $50.7_{\pm 2.9}$ |
| | Reuploading | $52.8_{\pm 4.8}$ | $54.1_{\pm 3.3}$ | $55.3_{\pm 1.7}$ | $51.8_{\pm 1.3}$ | $52.1_{\pm 7.2}$ | $54.7_{\pm 4.1}$ | $52.9_{\pm 1.6}$ | $53.2_{\pm 6.0}$ | $55.3_{\pm 3.1}$ |
| Shortcut-only | Local | $90.8_{\pm 0.7}$ | $49.6_{\pm 2.5}$ | $9.3_{\pm 0.9}$ | $90.7_{\pm 0.6}$ | $49.4_{\pm 2.7}$ | $9.6_{\pm 1.3}$ | $90.3_{\pm 0.7}$ | $49.0_{\pm 2.7}$ | $9.9_{\pm 1.5}$ |
| | Entangling | $90.8_{\pm 0.7}$ | $49.9_{\pm 2.7}$ | $9.3_{\pm 0.9}$ | $88.7_{\pm 3.5}$ | $49.9_{\pm 2.3}$ | $12.2_{\pm 4.2}$ | $85.0_{\pm 9.8}$ | $50.6_{\pm 1.5}$ | $15.9_{\pm 10.5}$ |
| | Reuploading | $90.2_{\pm 1.1}$ | $49.7_{\pm 2.2}$ | $9.7_{\pm 0.9}$ | $73.2_{\pm 19.8}$ | $49.7_{\pm 2.3}$ | $28.1_{\pm 21.1}$ | $70.7_{\pm 18.5}$ | $50.1_{\pm 2.5}$ | $31.2_{\pm 19.5}$ |
| Shortcut-flip | Local | $9.2_{\pm 0.7}$ | $50.4_{\pm 2.5}$ | $90.7_{\pm 0.9}$ | $9.7_{\pm 0.9}$ | $50.6_{\pm 2.3}$ | $90.4_{\pm 0.7}$ | $10.1_{\pm 1.1}$ | $50.9_{\pm 3.0}$ | $89.7_{\pm 0.7}$ |
| | Entangling | $13.3_{\pm 3.8}$ | $51.9_{\pm 1.9}$ | $89.3_{\pm 1.7}$ | $14.7_{\pm 2.4}$ | $51.2_{\pm 3.5}$ | $88.3_{\pm 0.7}$ | $19.1_{\pm 5.6}$ | $51.1_{\pm 3.4}$ | $85.8_{\pm 2.2}$ |
| | Reuploading | $17.7_{\pm 3.0}$ | $51.1_{\pm 4.1}$ | $87.8_{\pm 2.3}$ | $38.0_{\pm 14.2}$ | $49.0_{\pm 2.9}$ | $65.7_{\pm 13.2}$ | $40.6_{\pm 12.1}$ | $47.3_{\pm 2.3}$ | $63.1_{\pm 11.0}$ |

Table 28: Spirals-Signatures: fixed-checkpoint accuracy (%) under phase damping. Ideal/shot references: Table 25.

| Input | VQC | ID | | Random | | OOD reversed | |
|---|---|---|---|---|---|---|---|
| | | Ideal | 1024 shots | Ideal | 1024 shots | Ideal | 1024 shots |
| Original | Local | $94.3_{\pm 1.8}$ | $94.9_{\pm 1.6}$ | $95.1_{\pm 1.2}$ | $95.1_{\pm 1.2}$ | $95.6_{\pm 0.2}$ | $95.6_{\pm 0.2}$ |
| | Entangling | $88.2_{\pm 1.0}$ | $89.1_{\pm 0.8}$ | $87.6_{\pm 3.6}$ | $88.0_{\pm 2.9}$ | $87.1_{\pm 1.8}$ | $87.6_{\pm 1.8}$ |
| | Reuploading | $57.0_{\pm 10.3}$ | $55.8_{\pm 9.5}$ | $57.8_{\pm 8.9}$ | $55.7_{\pm 7.9}$ | $57.3_{\pm 9.8}$ | $58.9_{\pm 7.5}$ |
| Stable-only | Local | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| | Entangling | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| | Reuploading | $0.0_{\pm 0.0}$ | $0.9_{\pm 1.5}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ |
| Shortcut-only | Local | $48.9_{\pm 5.4}$ | $49.1_{\pm 5.6}$ | $48.6_{\pm 1.6}$ | $48.6_{\pm 1.3}$ | $48.8_{\pm 2.8}$ | $48.8_{\pm 2.5}$ |
| | Entangling | $49.7_{\pm 3.2}$ | $49.1_{\pm 3.7}$ | $49.7_{\pm 2.5}$ | $49.1_{\pm 1.9}$ | $51.0_{\pm 1.8}$ | $50.3_{\pm 1.0}$ |
| | Reuploading | $50.1_{\pm 1.6}$ | $50.7_{\pm 1.5}$ | $48.8_{\pm 1.6}$ | $51.6_{\pm 1.8}$ | $48.7_{\pm 2.0}$ | $50.2_{\pm 1.6}$ |
| Shortcut-flip | Local | $94.6_{\pm 2.0}$ | $94.6_{\pm 2.0}$ | $95.6_{\pm 1.6}$ | $95.2_{\pm 1.5}$ | $95.7_{\pm 0.7}$ | $95.9_{\pm 0.4}$ |
| | Entangling | $87.3_{\pm 2.0}$ | $86.9_{\pm 1.7}$ | $86.9_{\pm 4.0}$ | $86.4_{\pm 4.5}$ | $88.0_{\pm 3.5}$ | $87.4_{\pm 3.7}$ |
| | Reuploading | $59.3_{\pm 6.6}$ | $59.6_{\pm 6.0}$ | $56.2_{\pm 7.0}$ | $57.1_{\pm 6.8}$ | $57.8_{\pm 10.8}$ | $57.9_{\pm 8.5}$ |

Table 29: Spurious-Fourier: ideal analytic versus 1024-shot accuracy (%) for fixed clean-trained checkpoints. ID is held out; circuits are otherwise ideal.

| Input | VQC | Weak $p = 0.005$ | | | Moderate $p = 0.05$ | | | Strong $p = 0.1$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ID | Random | OOD reversed | ID | Random | OOD reversed | ID | Random | OOD reversed |
| Original | Local | $94.3_{\pm 1.8}$ | $95.1_{\pm 1.4}$ | $95.6_{\pm 0.2}$ | $94.8_{\pm 2.4}$ | $94.3_{\pm 1.8}$ | $94.7_{\pm 1.8}$ | $92.9_{\pm 4.4}$ | $90.4_{\pm 1.7}$ | $91.1_{\pm 2.2}$ |
| | Entangling | $88.7_{\pm 1.2}$ | $87.1_{\pm 3.4}$ | $87.2_{\pm 1.8}$ | $70.3_{\pm 2.9}$ | $67.7_{\pm 2.1}$ | $71.0_{\pm 8.7}$ | $51.8_{\pm 3.4}$ | $49.2_{\pm 0.4}$ | $51.2_{\pm 5.9}$ |
| | Reuploading | $57.1_{\pm 10.1}$ | $57.4_{\pm 8.3}$ | $57.4_{\pm 9.3}$ | $51.3_{\pm 4.6}$ | $53.9_{\pm 4.4}$ | $55.7_{\pm 8.2}$ | $46.4_{\pm 2.2}$ | $47.3_{\pm 2.3}$ | $51.3_{\pm 5.6}$ |
| Stable-only | Local | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| | Entangling | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $51.8_{\pm 3.5}$ | $49.1_{\pm 0.2}$ | $51.2_{\pm 5.9}$ |
| | Reuploading | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $0.1_{\pm 0.2}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $16.3_{\pm 28.3}$ | $16.3_{\pm 28.3}$ | $19.1_{\pm 33.1}$ |
| Shortcut-only | Local | $49.1_{\pm 5.4}$ | $48.7_{\pm 1.5}$ | $48.8_{\pm 3.3}$ | $51.4_{\pm 6.3}$ | $50.2_{\pm 2.0}$ | $48.9_{\pm 4.0}$ | $54.6_{\pm 6.0}$ | $49.2_{\pm 2.5}$ | $49.3_{\pm 4.1}$ |
| | Entangling | $48.9_{\pm 3.7}$ | $48.8_{\pm 1.6}$ | $50.3_{\pm 1.3}$ | $50.9_{\pm 3.4}$ | $50.0_{\pm 0.6}$ | $50.3_{\pm 4.4}$ | $51.8_{\pm 3.5}$ | $49.1_{\pm 0.2}$ | $51.2_{\pm 5.9}$ |
| | Reuploading | $50.0_{\pm 0.9}$ | $48.7_{\pm 1.5}$ | $48.7_{\pm 2.6}$ | $50.9_{\pm 1.8}$ | $48.0_{\pm 1.2}$ | $51.1_{\pm 2.7}$ | $50.1_{\pm 2.6}$ | $48.6_{\pm 1.1}$ | $53.4_{\pm 3.2}$ |
| Shortcut-flip | Local | $94.6_{\pm 2.0}$ | $95.4_{\pm 1.6}$ | $95.9_{\pm 0.7}$ | $94.1_{\pm 3.0}$ | $94.2_{\pm 1.7}$ | $95.0_{\pm 1.8}$ | $92.4_{\pm 4.4}$ | $91.0_{\pm 2.0}$ | $91.7_{\pm 2.8}$ |
| | Entangling | $87.6_{\pm 2.7}$ | $87.2_{\pm 3.9}$ | $87.7_{\pm 4.5}$ | $71.2_{\pm 2.7}$ | $69.4_{\pm 5.4}$ | $70.9_{\pm 9.5}$ | $51.9_{\pm 3.7}$ | $49.0_{\pm 0.0}$ | $51.1_{\pm 6.0}$ |
| | Reuploading | $59.4_{\pm 6.2}$ | $56.6_{\pm 6.4}$ | $57.8_{\pm 10.1}$ | $58.9_{\pm 2.5}$ | $55.8_{\pm 5.7}$ | $56.6_{\pm 8.1}$ | $50.9_{\pm 2.7}$ | $48.6_{\pm 2.5}$ | $52.6_{\pm 4.3}$ |

Table 30: Spurious-Fourier: fixed-checkpoint accuracy (%) under amplitude damping. Ideal/shot references: Table 29.

| Input | VQC | Weak $p = 0.005$ | | | Moderate $p = 0.08$ | | | Strong $p = 0.15$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ID | Random | OOD reversed | ID | Random | OOD reversed | ID | Random | OOD reversed |
| Original | Local | $94.3_{\pm 1.8}$ | $95.1_{\pm 1.2}$ | $95.6_{\pm 0.2}$ | $94.3_{\pm 1.8}$ | $95.1_{\pm 1.4}$ | $95.6_{\pm 0.2}$ | $94.7_{\pm 2.1}$ | $95.2_{\pm 1.3}$ | $95.4_{\pm 0.2}$ |
| | Entangling | $88.3_{\pm 0.9}$ | $87.4_{\pm 3.7}$ | $87.0_{\pm 2.0}$ | $84.2_{\pm 3.9}$ | $82.8_{\pm 6.5}$ | $84.6_{\pm 8.1}$ | $52.1_{\pm 3.4}$ | $49.3_{\pm 0.3}$ | $51.7_{\pm 5.8}$ |
| | Reuploading | $56.2_{\pm 9.6}$ | $57.0_{\pm 7.9}$ | $56.9_{\pm 9.3}$ | $49.0_{\pm 3.4}$ | $48.9_{\pm 6.2}$ | $47.9_{\pm 8.8}$ | $46.4_{\pm 2.4}$ | $42.8_{\pm 7.3}$ | $45.8_{\pm 9.2}$ |
| Stable-only | Local | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| | Entangling | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $68.2_{\pm 27.7}$ | $66.1_{\pm 29.3}$ | $67.7_{\pm 28.6}$ |
| | Reuploading | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $16.3_{\pm 28.3}$ | $16.3_{\pm 28.3}$ | $19.1_{\pm 33.1}$ |
| Shortcut-only | Local | $48.9_{\pm 5.4}$ | $48.4_{\pm 1.4}$ | $48.8_{\pm 2.8}$ | $49.1_{\pm 5.7}$ | $48.3_{\pm 1.5}$ | $48.4_{\pm 2.9}$ | $49.9_{\pm 6.6}$ | $48.7_{\pm 1.5}$ | $49.0_{\pm 2.6}$ |
| | Entangling | $49.4_{\pm 3.4}$ | $49.8_{\pm 2.7}$ | $51.0_{\pm 1.8}$ | $49.6_{\pm 6.3}$ | $49.4_{\pm 2.0}$ | $50.4_{\pm 1.7}$ | $51.6_{\pm 3.6}$ | $49.1_{\pm 0.2}$ | $51.2_{\pm 5.9}$ |
| | Reuploading | $50.7_{\pm 2.0}$ | $48.6_{\pm 1.2}$ | $48.4_{\pm 2.5}$ | $49.9_{\pm 3.2}$ | $49.3_{\pm 1.5}$ | $48.6_{\pm 1.0}$ | $49.9_{\pm 2.7}$ | $50.3_{\pm 1.2}$ | $52.6_{\pm 3.6}$ |
| Shortcut-flip | Local | $94.6_{\pm 2.0}$ | $95.6_{\pm 1.6}$ | $95.7_{\pm 0.7}$ | $94.6_{\pm 2.0}$ | $95.6_{\pm 1.6}$ | $95.8_{\pm 0.5}$ | $94.6_{\pm 2.0}$ | $95.7_{\pm 1.5}$ | $95.4_{\pm 0.4}$ |
| | Entangling | $87.3_{\pm 2.1}$ | $87.0_{\pm 3.8}$ | $87.9_{\pm 3.7}$ | $85.2_{\pm 3.1}$ | $82.1_{\pm 8.2}$ | $83.9_{\pm 9.7}$ | $51.8_{\pm 3.5}$ | $49.2_{\pm 0.2}$ | $51.2_{\pm 5.9}$ |
| | Reuploading | $59.0_{\pm 5.9}$ | $56.3_{\pm 6.9}$ | $57.1_{\pm 9.5}$ | $51.8_{\pm 1.7}$ | $49.9_{\pm 5.8}$ | $49.7_{\pm 8.3}$ | $49.9_{\pm 3.7}$ | $44.3_{\pm 7.3}$ | $48.2_{\pm 8.8}$ |

Table 31: Spurious-Fourier: fixed-checkpoint accuracy (%) under depolarizing. Ideal/shot references: Table 29.

| Input | VQC | Weak $p = 0.005$ | | | Moderate $p = 0.5$ | | | Strong $p = 0.6$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ID | Random | OOD reversed | ID | Random | OOD reversed | ID | Random | OOD reversed |
| Original | Local | $94.3_{\pm 1.8}$ | $95.1_{\pm 1.2}$ | $95.6_{\pm 0.2}$ | $91.9_{\pm 3.4}$ | $89.7_{\pm 1.9}$ | $90.6_{\pm 2.8}$ | $88.4_{\pm 5.3}$ | $85.7_{\pm 1.7}$ | $87.1_{\pm 2.1}$ |
| | Entangling | $88.2_{\pm 1.0}$ | $87.6_{\pm 3.6}$ | $87.0_{\pm 2.0}$ | $87.9_{\pm 1.5}$ | $87.2_{\pm 3.0}$ | $87.3_{\pm 2.2}$ | $87.9_{\pm 1.5}$ | $87.1_{\pm 3.7}$ | $87.2_{\pm 2.2}$ |
| | Reuploading | $57.0_{\pm 10.3}$ | $57.7_{\pm 8.7}$ | $57.3_{\pm 9.8}$ | $48.4_{\pm 3.4}$ | $47.9_{\pm 6.6}$ | $49.1_{\pm 11.8}$ | $49.2_{\pm 5.7}$ | $45.4_{\pm 6.5}$ | $48.7_{\pm 9.3}$ |
| Stable-only | Local | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| | Entangling | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| | Reuploading | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $16.3_{\pm 28.3}$ | $16.3_{\pm 28.3}$ | $19.1_{\pm 33.1}$ | $34.9_{\pm 30.4}$ | $32.8_{\pm 28.4}$ | $34.3_{\pm 30.3}$ |
| Shortcut-only | Local | $49.0_{\pm 5.2}$ | $48.6_{\pm 1.3}$ | $48.8_{\pm 2.8}$ | $54.1_{\pm 5.3}$ | $49.3_{\pm 3.7}$ | $49.8_{\pm 3.5}$ | $53.2_{\pm 3.7}$ | $49.2_{\pm 2.9}$ | $48.8_{\pm 5.0}$ |
| | Entangling | $49.7_{\pm 3.2}$ | $49.7_{\pm 2.5}$ | $51.0_{\pm 1.8}$ | $49.4_{\pm 3.7}$ | $49.2_{\pm 2.1}$ | $50.7_{\pm 1.2}$ | $49.3_{\pm 2.9}$ | $48.9_{\pm 2.7}$ | $50.6_{\pm 1.1}$ |
| | Reuploading | $50.0_{\pm 2.0}$ | $48.8_{\pm 1.6}$ | $48.8_{\pm 2.2}$ | $49.7_{\pm 0.9}$ | $51.2_{\pm 2.0}$ | $52.2_{\pm 3.6}$ | $51.4_{\pm 3.4}$ | $50.4_{\pm 2.2}$ | $51.6_{\pm 5.8}$ |
| Shortcut-flip | Local | $94.6_{\pm 2.0}$ | $95.6_{\pm 1.6}$ | $95.7_{\pm 0.7}$ | $91.6_{\pm 4.2}$ | $89.8_{\pm 1.3}$ | $91.1_{\pm 2.5}$ | $88.3_{\pm 5.2}$ | $86.1_{\pm 1.2}$ | $88.4_{\pm 3.0}$ |
| | Entangling | $87.3_{\pm 2.0}$ | $86.9_{\pm 4.0}$ | $88.0_{\pm 3.5}$ | $87.2_{\pm 2.1}$ | $86.6_{\pm 4.2}$ | $87.7_{\pm 3.9}$ | $87.3_{\pm 2.8}$ | $86.6_{\pm 3.9}$ | $87.6_{\pm 4.1}$ |
| | Reuploading | $59.2_{\pm 6.8}$ | $56.3_{\pm 7.2}$ | $57.4_{\pm 10.6}$ | $52.1_{\pm 2.7}$ | $49.9_{\pm 7.6}$ | $51.2_{\pm 9.0}$ | $50.8_{\pm 4.6}$ | $46.7_{\pm 5.5}$ | $49.1_{\pm 7.2}$ |

Table 32: Spurious-Fourier: fixed-checkpoint accuracy (%) under phase damping. Ideal/shot references: Table 29.

| Input | VQC | ID | | Random | | OOD reversed | |
|---|---|---|---|---|---|---|---|
| | | Ideal | 1024 shots | Ideal | 1024 shots | Ideal | 1024 shots |
| Original | Local | $88.7_{\pm 0.7}$ | $88.8_{\pm 0.5}$ | $36.4_{\pm 2.9}$ | $36.4_{\pm 2.8}$ | $61.2_{\pm 2.3}$ | $60.6_{\pm 1.0}$ |
| | Entangling | $71.6_{\pm 18.2}$ | $71.1_{\pm 19.0}$ | $44.3_{\pm 8.2}$ | $45.8_{\pm 9.6}$ | $57.2_{\pm 2.5}$ | $56.9_{\pm 2.7}$ |
| | Reuploading | $47.0_{\pm 2.3}$ | $46.4_{\pm 3.7}$ | $58.1_{\pm 5.9}$ | $58.2_{\pm 6.9}$ | $42.0_{\pm 2.7}$ | $41.6_{\pm 2.5}$ |
| Stable-only | Local | $34.8_{\pm 23.9}$ | $36.4_{\pm 21.0}$ | $35.0_{\pm 29.1}$ | $36.1_{\pm 27.2}$ | $35.8_{\pm 28.1}$ | $36.3_{\pm 27.2}$ |
| | Entangling | $57.2_{\pm 20.9}$ | $52.3_{\pm 12.5}$ | $63.2_{\pm 17.9}$ | $56.8_{\pm 7.1}$ | $60.6_{\pm 16.8}$ | $55.4_{\pm 8.0}$ |
| | Reuploading | $63.2_{\pm 26.7}$ | $57.4_{\pm 16.7}$ | $69.0_{\pm 26.0}$ | $65.2_{\pm 19.5}$ | $65.6_{\pm 29.3}$ | $61.0_{\pm 21.5}$ |
| Shortcut-only | Local | $89.7_{\pm 1.2}$ | $89.7_{\pm 1.2}$ | $50.4_{\pm 2.4}$ | $50.1_{\pm 2.4}$ | $75.2_{\pm 3.0}$ | $75.1_{\pm 2.8}$ |
| | Entangling | $25.0_{\pm 17.1}$ | $26.8_{\pm 16.7}$ | $52.8_{\pm 2.7}$ | $52.6_{\pm 2.5}$ | $37.0_{\pm 12.6}$ | $37.0_{\pm 12.2}$ |
| | Reuploading | $52.1_{\pm 20.8}$ | $51.4_{\pm 22.5}$ | $51.4_{\pm 2.8}$ | $51.7_{\pm 2.6}$ | $28.1_{\pm 16.0}$ | $29.7_{\pm 17.6}$ |
| Shortcut-flip | Local | $88.1_{\pm 1.3}$ | $88.2_{\pm 1.2}$ | $40.9_{\pm 2.4}$ | $41.4_{\pm 2.5}$ | $79.9_{\pm 4.2}$ | $79.7_{\pm 3.7}$ |
| | Entangling | $71.1_{\pm 17.2}$ | $72.2_{\pm 16.8}$ | $48.1_{\pm 6.4}$ | $47.7_{\pm 6.4}$ | $66.4_{\pm 8.5}$ | $65.9_{\pm 9.4}$ |
| | Reuploading | $39.6_{\pm 5.1}$ | $41.8_{\pm 6.0}$ | $48.9_{\pm 2.9}$ | $49.2_{\pm 4.1}$ | $31.4_{\pm 6.2}$ | $32.0_{\pm 6.2}$ |

Table 33: Q-FOS: ideal analytic versus 1024-shot accuracy (%) for fixed clean-trained checkpoints. ID is held out; circuits are otherwise ideal.

| Input | VQC | Weak $p = 0.005$ | | | Moderate $p = 0.05$ | | | Strong $p = 0.1$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ID | Random | OOD reversed | ID | Random | OOD reversed | ID | Random | OOD reversed |
| Original | Local | $88.8_{\pm 0.8}$ | $35.7_{\pm 3.2}$ | $60.6_{\pm 2.8}$ | $86.7_{\pm 5.6}$ | $38.3_{\pm 6.7}$ | $58.8_{\pm 3.5}$ | $76.4_{\pm 14.1}$ | $44.0_{\pm 7.7}$ | $55.2_{\pm 3.7}$ |
| | Entangling | $71.7_{\pm 18.0}$ | $44.6_{\pm 8.6}$ | $57.2_{\pm 3.1}$ | $72.8_{\pm 19.2}$ | $41.9_{\pm 8.6}$ | $58.1_{\pm 2.3}$ | $65.8_{\pm 20.4}$ | $44.3_{\pm 10.7}$ | $56.0_{\pm 2.3}$ |
| | Reuploading | $46.4_{\pm 3.3}$ | $58.6_{\pm 4.8}$ | $41.7_{\pm 1.8}$ | $48.0_{\pm 1.8}$ | $55.4_{\pm 2.2}$ | $45.8_{\pm 4.3}$ | $50.6_{\pm 4.9}$ | $52.6_{\pm 3.0}$ | $48.8_{\pm 2.0}$ |
| Stable-only | Local | $37.8_{\pm 18.7}$ | $35.8_{\pm 27.8}$ | $37.0_{\pm 26.0}$ | $51.0_{\pm 4.8}$ | $51.0_{\pm 4.2}$ | $51.0_{\pm 2.2}$ | $51.0_{\pm 4.8}$ | $51.0_{\pm 4.2}$ | $51.0_{\pm 2.2}$ |
| | Entangling | $57.9_{\pm 22.1}$ | $62.9_{\pm 17.3}$ | $60.9_{\pm 17.4}$ | $47.0_{\pm 3.4}$ | $51.2_{\pm 4.1}$ | $51.7_{\pm 1.5}$ | $50.9_{\pm 4.7}$ | $51.0_{\pm 4.2}$ | $51.0_{\pm 2.2}$ |
| | Reuploading | $63.3_{\pm 26.9}$ | $69.3_{\pm 26.6}$ | $65.6_{\pm 29.3}$ | $65.2_{\pm 30.2}$ | $69.3_{\pm 26.6}$ | $65.8_{\pm 29.7}$ | $50.6_{\pm 4.9}$ | $52.6_{\pm 3.0}$ | $48.8_{\pm 2.0}$ |
| Shortcut-only | Local | $89.7_{\pm 1.2}$ | $50.4_{\pm 1.6}$ | $74.9_{\pm 2.7}$ | $88.4_{\pm 0.7}$ | $48.1_{\pm 3.0}$ | $66.6_{\pm 5.2}$ | $72.4_{\pm 19.2}$ | $50.8_{\pm 3.6}$ | $55.1_{\pm 3.7}$ |
| | Entangling | $25.9_{\pm 16.9}$ | $52.7_{\pm 2.5}$ | $37.8_{\pm 11.7}$ | $26.8_{\pm 16.7}$ | $54.0_{\pm 2.2}$ | $40.3_{\pm 6.9}$ | $27.8_{\pm 24.6}$ | $53.9_{\pm 4.0}$ | $34.4_{\pm 18.2}$ |
| | Reuploading | $53.8_{\pm 18.3}$ | $53.3_{\pm 1.0}$ | $30.8_{\pm 20.9}$ | $54.3_{\pm 7.7}$ | $52.7_{\pm 2.6}$ | $37.3_{\pm 17.4}$ | $47.4_{\pm 1.6}$ | $52.8_{\pm 2.7}$ | $48.0_{\pm 2.3}$ |
| Shortcut-flip | Local | $88.1_{\pm 1.3}$ | $41.3_{\pm 2.0}$ | $79.3_{\pm 3.8}$ | $85.9_{\pm 5.7}$ | $41.1_{\pm 2.7}$ | $76.3_{\pm 6.7}$ | $75.1_{\pm 15.2}$ | $43.1_{\pm 4.6}$ | $65.9_{\pm 10.0}$ |
| | Entangling | $71.2_{\pm 17.5}$ | $47.1_{\pm 5.9}$ | $66.9_{\pm 8.4}$ | $71.9_{\pm 18.7}$ | $45.9_{\pm 5.5}$ | $69.2_{\pm 10.0}$ | $65.4_{\pm 20.1}$ | $48.1_{\pm 6.5}$ | $65.3_{\pm 14.4}$ |
| | Reuploading | $39.7_{\pm 5.6}$ | $49.0_{\pm 2.7}$ | $31.1_{\pm 6.5}$ | $44.4_{\pm 1.4}$ | $52.1_{\pm 3.9}$ | $40.9_{\pm 6.4}$ | $50.6_{\pm 4.9}$ | $52.6_{\pm 3.0}$ | $48.8_{\pm 2.0}$ |

Table 34: Q-FOS: fixed-checkpoint accuracy (%) under amplitude damping. Ideal/shot references: Table 33.

| Input | VQC | Weak $p = 0.005$ | | | Moderate $p = 0.08$ | | | Strong $p = 0.15$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ID | Random | OOD reversed | ID | Random | OOD reversed | ID | Random | OOD reversed |
| Original | Local | $88.7_{\pm 0.7}$ | $36.4_{\pm 2.9}$ | $60.9_{\pm 2.3}$ | $88.1_{\pm 0.5}$ | $38.0_{\pm 3.8}$ | $58.3_{\pm 3.5}$ | $84.2_{\pm 4.5}$ | $39.4_{\pm 6.8}$ | $55.6_{\pm 2.7}$ |
| | Entangling | $71.6_{\pm 18.3}$ | $44.3_{\pm 8.2}$ | $57.2_{\pm 2.8}$ | $63.7_{\pm 23.3}$ | $44.0_{\pm 10.7}$ | $53.3_{\pm 4.9}$ | $60.1_{\pm 23.4}$ | $46.6_{\pm 8.3}$ | $50.0_{\pm 3.0}$ |
| | Reuploading | $46.8_{\pm 2.9}$ | $58.0_{\pm 5.5}$ | $41.3_{\pm 2.3}$ | $44.6_{\pm 0.7}$ | $58.1_{\pm 2.5}$ | $42.4_{\pm 1.3}$ | $46.6_{\pm 2.7}$ | $52.8_{\pm 2.7}$ | $49.4_{\pm 2.4}$ |
| Stable-only | Local | $34.6_{\pm 24.2}$ | $34.8_{\pm 29.5}$ | $35.6_{\pm 28.5}$ | $31.4_{\pm 27.2}$ | $32.0_{\pm 28.8}$ | $32.8_{\pm 28.4}$ | $15.4_{\pm 26.8}$ | $18.6_{\pm 32.1}$ | $17.0_{\pm 29.2}$ |
| | Entangling | $57.2_{\pm 20.9}$ | $63.3_{\pm 18.1}$ | $61.0_{\pm 17.6}$ | $31.1_{\pm 27.1}$ | $34.2_{\pm 29.7}$ | $32.6_{\pm 28.3}$ | $49.0_{\pm 4.8}$ | $49.0_{\pm 4.2}$ | $49.0_{\pm 2.2}$ |
| | Reuploading | $63.3_{\pm 26.9}$ | $69.3_{\pm 26.6}$ | $65.6_{\pm 29.3}$ | $46.6_{\pm 2.7}$ | $52.8_{\pm 2.7}$ | $49.4_{\pm 2.4}$ | $46.6_{\pm 2.7}$ | $52.8_{\pm 2.7}$ | $49.4_{\pm 2.4}$ |
| Shortcut-only | Local | $89.7_{\pm 1.2}$ | $50.7_{\pm 2.2}$ | $75.2_{\pm 2.7}$ | $89.7_{\pm 1.2}$ | $49.7_{\pm 4.1}$ | $75.9_{\pm 3.5}$ | $89.6_{\pm 1.5}$ | $49.7_{\pm 3.0}$ | $73.7_{\pm 9.6}$ |
| | Entangling | $27.1_{\pm 16.7}$ | $52.2_{\pm 1.8}$ | $38.1_{\pm 11.2}$ | $34.7_{\pm 21.0}$ | $51.4_{\pm 1.0}$ | $39.8_{\pm 15.8}$ | $35.4_{\pm 19.6}$ | $50.1_{\pm 2.3}$ | $44.1_{\pm 8.5}$ |
| | Reuploading | $52.2_{\pm 20.3}$ | $52.2_{\pm 1.7}$ | $27.3_{\pm 15.0}$ | $51.8_{\pm 7.5}$ | $52.3_{\pm 3.4}$ | $50.6_{\pm 5.0}$ | $46.6_{\pm 2.7}$ | $52.8_{\pm 2.7}$ | $49.6_{\pm 2.5}$ |
| Shortcut-flip | Local | $88.1_{\pm 1.3}$ | $40.9_{\pm 2.4}$ | $79.9_{\pm 4.2}$ | $86.7_{\pm 2.1}$ | $41.3_{\pm 2.0}$ | $76.9_{\pm 6.3}$ | $82.8_{\pm 5.5}$ | $42.0_{\pm 1.3}$ | $72.9_{\pm 8.7}$ |
| | Entangling | $71.0_{\pm 17.5}$ | $47.2_{\pm 6.0}$ | $66.6_{\pm 8.9}$ | $63.0_{\pm 22.6}$ | $47.7_{\pm 5.5}$ | $61.1_{\pm 14.5}$ | $59.7_{\pm 22.7}$ | $46.3_{\pm 8.7}$ | $53.6_{\pm 8.5}$ |
| | Reuploading | $38.9_{\pm 5.1}$ | $48.4_{\pm 3.0}$ | $30.8_{\pm 6.4}$ | $37.0_{\pm 4.7}$ | $49.2_{\pm 3.0}$ | $32.7_{\pm 4.2}$ | $46.6_{\pm 2.7}$ | $52.8_{\pm 2.7}$ | $49.4_{\pm 2.4}$ |

Table 35: Q-FOS: fixed-checkpoint accuracy (%) under depolarizing. Ideal/shot references: Table 33.

| Input | VQC | Weak $p = 0.005$ | | | Moderate $p = 0.5$ | | | Strong $p = 0.6$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ID | Random | OOD reversed | ID | Random | OOD reversed | ID | Random | OOD reversed |
| Original | Local | $88.7_{\pm 0.7}$ | $36.3_{\pm 3.3}$ | $61.1_{\pm 2.5}$ | $85.9_{\pm 3.0}$ | $38.7_{\pm 5.6}$ | $56.6_{\pm 1.7}$ | $82.1_{\pm 6.7}$ | $40.0_{\pm 4.7}$ | $54.3_{\pm 0.9}$ |
| | Entangling | $71.6_{\pm 18.0}$ | $44.2_{\pm 8.4}$ | $56.9_{\pm 2.8}$ | $63.6_{\pm 23.1}$ | $45.4_{\pm 9.9}$ | $52.6_{\pm 4.4}$ | $60.0_{\pm 19.7}$ | $47.1_{\pm 10.1}$ | $51.4_{\pm 3.7}$ |
| | Reuploading | $46.9_{\pm 2.5}$ | $58.0_{\pm 5.8}$ | $41.8_{\pm 2.7}$ | $47.3_{\pm 1.7}$ | $52.8_{\pm 2.5}$ | $46.8_{\pm 3.7}$ | $46.6_{\pm 2.5}$ | $52.2_{\pm 0.8}$ | $47.0_{\pm 3.3}$ |
| Stable-only | Local | $35.0_{\pm 23.5}$ | $35.0_{\pm 29.1}$ | $35.9_{\pm 27.9}$ | $51.0_{\pm 4.8}$ | $51.0_{\pm 4.2}$ | $51.0_{\pm 2.2}$ | $51.0_{\pm 4.8}$ | $51.0_{\pm 4.2}$ | $51.0_{\pm 2.2}$ |
| | Entangling | $57.2_{\pm 20.9}$ | $63.1_{\pm 17.7}$ | $60.3_{\pm 16.5}$ | $34.2_{\pm 30.5}$ | $35.8_{\pm 31.2}$ | $34.9_{\pm 30.2}$ | $33.6_{\pm 29.4}$ | $34.1_{\pm 29.6}$ | $35.0_{\pm 30.3}$ |
| | Reuploading | $62.9_{\pm 26.1}$ | $69.0_{\pm 26.0}$ | $65.6_{\pm 29.3}$ | $46.6_{\pm 2.7}$ | $52.8_{\pm 2.7}$ | $49.4_{\pm 2.4}$ | $46.6_{\pm 2.7}$ | $52.8_{\pm 2.7}$ | $49.4_{\pm 2.4}$ |
| Shortcut-only | Local | $89.7_{\pm 1.2}$ | $50.3_{\pm 2.4}$ | $75.2_{\pm 3.0}$ | $88.3_{\pm 0.9}$ | $50.1_{\pm 3.3}$ | $58.9_{\pm 4.4}$ | $85.2_{\pm 6.0}$ | $50.4_{\pm 3.2}$ | $56.0_{\pm 2.5}$ |
| | Entangling | $25.2_{\pm 17.1}$ | $53.0_{\pm 2.7}$ | $37.7_{\pm 11.8}$ | $34.7_{\pm 21.0}$ | $52.8_{\pm 2.7}$ | $41.7_{\pm 12.6}$ | $34.7_{\pm 21.0}$ | $51.6_{\pm 1.1}$ | $40.8_{\pm 14.1}$ |
| | Reuploading | $51.8_{\pm 20.5}$ | $51.6_{\pm 2.6}$ | $28.1_{\pm 16.2}$ | $42.9_{\pm 12.5}$ | $52.7_{\pm 2.3}$ | $41.1_{\pm 14.4}$ | $43.6_{\pm 9.7}$ | $52.8_{\pm 2.8}$ | $47.1_{\pm 5.6}$ |
| Shortcut-flip | Local | $88.1_{\pm 1.3}$ | $40.9_{\pm 2.4}$ | $79.9_{\pm 4.2}$ | $84.7_{\pm 3.5}$ | $41.4_{\pm 0.7}$ | $74.7_{\pm 5.2}$ | $82.1_{\pm 4.6}$ | $41.4_{\pm 0.5}$ | $71.1_{\pm 3.9}$ |
| | Entangling | $70.9_{\pm 17.1}$ | $48.1_{\pm 6.4}$ | $66.4_{\pm 8.5}$ | $62.3_{\pm 21.6}$ | $47.8_{\pm 6.4}$ | $59.8_{\pm 12.9}$ | $59.4_{\pm 18.8}$ | $49.2_{\pm 8.0}$ | $56.1_{\pm 9.4}$ |
| | Reuploading | $39.7_{\pm 5.0}$ | $48.9_{\pm 2.9}$ | $31.3_{\pm 6.0}$ | $44.8_{\pm 1.3}$ | $49.3_{\pm 1.9}$ | $42.0_{\pm 3.6}$ | $45.4_{\pm 2.6}$ | $51.6_{\pm 1.7}$ | $45.8_{\pm 3.0}$ |

Table 36: Q-FOS: fixed-checkpoint accuracy (%) under phase damping. Ideal/shot references: Table 33.

| | | ID | | Random | | OOD reversed | |
|---|---|---|---|---|---|---|---|
| Input | VQC | Ideal | 1024 shots | Ideal | 1024 shots | Ideal | 1024 shots |
| Original | Local | $90.9_{\pm 1.0}$ | $90.9_{\pm 1.0}$ | $50.3_{\pm 1.0}$ | $50.4_{\pm 1.0}$ | $7.9_{\pm 1.2}$ | $7.9_{\pm 1.2}$ |
| | Entangling | $99.3_{\pm 0.3}$ | $99.2_{\pm 0.5}$ | $99.0_{\pm 1.0}$ | $98.7_{\pm 0.7}$ | $98.6_{\pm 2.0}$ | $98.3_{\pm 2.3}$ |
| | Reuploading | $94.4_{\pm 6.1}$ | $94.6_{\pm 6.4}$ | $89.2_{\pm 9.2}$ | $89.1_{\pm 8.8}$ | $84.0_{\pm 12.3}$ | $83.0_{\pm 12.3}$ |
| Stable-only | Local | $48.4_{\pm 2.8}$ | $48.4_{\pm 3.7}$ | $50.7_{\pm 3.0}$ | $50.2_{\pm 1.6}$ | $51.1_{\pm 1.4}$ | $51.9_{\pm 0.7}$ |
| | Entangling | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $99.9_{\pm 0.2}$ | $99.9_{\pm 0.2}$ | $99.9_{\pm 0.2}$ | $99.9_{\pm 0.2}$ |
| | Reuploading | $93.9_{\pm 10.3}$ | $92.3_{\pm 13.0}$ | $95.1_{\pm 8.2}$ | $93.7_{\pm 10.4}$ | $95.2_{\pm 7.7}$ | $92.2_{\pm 12.3}$ |
| Shortcut-only | Local | $90.9_{\pm 1.0}$ | $90.9_{\pm 1.0}$ | $50.3_{\pm 1.0}$ | $50.3_{\pm 1.0}$ | $7.9_{\pm 1.2}$ | $7.9_{\pm 1.2}$ |
| | Entangling | $69.8_{\pm 17.0}$ | $69.2_{\pm 14.3}$ | $50.8_{\pm 1.1}$ | $53.7_{\pm 1.9}$ | $33.4_{\pm 22.2}$ | $33.2_{\pm 18.5}$ |
| | Reuploading | $49.7_{\pm 3.8}$ | $49.2_{\pm 3.7}$ | $50.1_{\pm 0.5}$ | $51.6_{\pm 1.9}$ | $52.4_{\pm 2.2}$ | $54.2_{\pm 1.1}$ |
| Shortcut-flip | Local | $9.1_{\pm 1.0}$ | $9.1_{\pm 1.0}$ | $49.7_{\pm 1.0}$ | $49.7_{\pm 1.0}$ | $92.1_{\pm 1.2}$ | $92.1_{\pm 1.2}$ |
| | Entangling | $99.1_{\pm 1.0}$ | $98.7_{\pm 1.8}$ | $98.7_{\pm 1.8}$ | $98.8_{\pm 1.6}$ | $99.7_{\pm 0.3}$ | $99.4_{\pm 0.7}$ |
| | Reuploading | $84.7_{\pm 11.0}$ | $83.3_{\pm 11.0}$ | $89.4_{\pm 8.5}$ | $89.1_{\pm 8.7}$ | $94.0_{\pm 6.3}$ | $93.7_{\pm 6.1}$ |

Table 37: Q-LINK: ideal analytic versus 1024-shot accuracy (%) for fixed clean-trained checkpoints. ID is held out; circuits are otherwise ideal.

| Input | VQC | Weak $p = 0.005$ | | | Moderate $p = 0.05$ | | | Strong $p = 0.1$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ID | Random | OOD reversed | ID | Random | OOD reversed | ID | Random | OOD reversed |
| Original | Local | $90.9_{\pm 1.0}$ | $50.3_{\pm 1.0}$ | $7.9_{\pm 1.2}$ | $90.9_{\pm 1.0}$ | $50.3_{\pm 1.0}$ | $7.9_{\pm 1.2}$ | $90.8_{\pm 0.8}$ | $50.4_{\pm 1.0}$ | $8.0_{\pm 1.0}$ |
| | Entangling | $99.4_{\pm 0.5}$ | $99.1_{\pm 0.8}$ | $98.7_{\pm 1.5}$ | $98.3_{\pm 1.5}$ | $95.6_{\pm 3.4}$ | $93.9_{\pm 8.6}$ | $66.8_{\pm 5.4}$ | $66.2_{\pm 10.4}$ | $72.2_{\pm 15.8}$ |
| | Reuploading | $94.7_{\pm 5.8}$ | $89.3_{\pm 8.8}$ | $83.6_{\pm 12.3}$ | $93.4_{\pm 3.2}$ | $87.9_{\pm 6.6}$ | $82.3_{\pm 6.5}$ | $84.9_{\pm 8.4}$ | $83.0_{\pm 8.6}$ | $79.2_{\pm 6.7}$ |
| Stable-only | Local | $47.7_{\pm 1.5}$ | $50.0_{\pm 1.9}$ | $50.9_{\pm 1.7}$ | $46.8_{\pm 0.2}$ | $49.3_{\pm 0.9}$ | $50.1_{\pm 2.9}$ | $46.8_{\pm 0.2}$ | $49.3_{\pm 0.9}$ | $50.1_{\pm 2.9}$ |
| | Entangling | $100.0_{\pm 0.0}$ | $99.9_{\pm 0.2}$ | $99.9_{\pm 0.2}$ | $99.0_{\pm 1.5}$ | $98.8_{\pm 1.3}$ | $99.0_{\pm 1.5}$ | $71.2_{\pm 21.4}$ | $69.9_{\pm 17.7}$ | $74.1_{\pm 18.6}$ |
| | Reuploading | $93.4_{\pm 11.1}$ | $95.4_{\pm 7.6}$ | $93.2_{\pm 11.2}$ | $81.0_{\pm 29.2}$ | $85.8_{\pm 21.5}$ | $83.4_{\pm 26.1}$ | $74.3_{\pm 26.7}$ | $78.7_{\pm 20.9}$ | $77.9_{\pm 24.7}$ |
| Shortcut-only | Local | $90.9_{\pm 1.0}$ | $50.3_{\pm 1.0}$ | $7.9_{\pm 1.2}$ | $90.9_{\pm 1.0}$ | $50.3_{\pm 1.0}$ | $7.9_{\pm 1.2}$ | $90.7_{\pm 0.6}$ | $50.7_{\pm 0.6}$ | $7.9_{\pm 1.2}$ |
| | Entangling | $88.8_{\pm 1.5}$ | $50.9_{\pm 1.3}$ | $10.3_{\pm 1.9}$ | $46.8_{\pm 0.2}$ | $49.3_{\pm 0.9}$ | $50.1_{\pm 2.9}$ | $46.8_{\pm 0.2}$ | $49.3_{\pm 0.9}$ | $50.1_{\pm 2.9}$ |
| | Reuploading | $49.3_{\pm 3.8}$ | $50.4_{\pm 1.1}$ | $52.7_{\pm 1.9}$ | $48.7_{\pm 4.1}$ | $49.8_{\pm 0.2}$ | $53.3_{\pm 0.9}$ | $46.3_{\pm 7.8}$ | $49.0_{\pm 1.5}$ | $56.1_{\pm 5.4}$ |
| Shortcut-flip | Local | $9.1_{\pm 1.0}$ | $49.7_{\pm 1.0}$ | $92.1_{\pm 1.2}$ | $9.1_{\pm 1.0}$ | $49.7_{\pm 1.0}$ | $92.1_{\pm 1.2}$ | $9.2_{\pm 0.8}$ | $49.7_{\pm 1.0}$ | $91.9_{\pm 0.8}$ |
| | Entangling | $99.0_{\pm 1.2}$ | $98.7_{\pm 1.8}$ | $99.7_{\pm 0.3}$ | $94.6_{\pm 8.6}$ | $95.0_{\pm 5.5}$ | $98.1_{\pm 0.4}$ | $68.7_{\pm 18.4}$ | $67.4_{\pm 10.2}$ | $70.3_{\pm 4.9}$ |
| | Reuploading | $84.9_{\pm 11.4}$ | $89.3_{\pm 9.0}$ | $94.0_{\pm 6.0}$ | $82.4_{\pm 6.3}$ | $87.7_{\pm 5.5}$ | $92.8_{\pm 3.2}$ | $80.1_{\pm 6.0}$ | $83.4_{\pm 6.5}$ | $84.9_{\pm 8.6}$ |

Table 38: Q-LINK: fixed-checkpoint accuracy (%) under amplitude damping. Ideal/shot references: Table 37.

| Input | VQC | Weak $p = 0.005$ | | | Moderate $p = 0.08$ | | | Strong $p = 0.15$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ID | Random | OOD reversed | ID | Random | OOD reversed | ID | Random | OOD reversed |
| Original | Local | $90.9_{\pm 1.0}$ | $50.3_{\pm 1.0}$ | $7.9_{\pm 1.2}$ | $90.9_{\pm 1.0}$ | $50.4_{\pm 1.0}$ | $7.9_{\pm 1.2}$ | $90.9_{\pm 1.0}$ | $50.4_{\pm 1.0}$ | $7.9_{\pm 1.2}$ |
| | Entangling | $99.3_{\pm 0.3}$ | $99.0_{\pm 1.0}$ | $98.6_{\pm 2.0}$ | $97.2_{\pm 1.7}$ | $96.2_{\pm 0.4}$ | $95.0_{\pm 1.8}$ | $88.7_{\pm 3.7}$ | $86.1_{\pm 3.0}$ | $86.0_{\pm 4.8}$ |
| | Reuploading | $94.4_{\pm 6.1}$ | $89.1_{\pm 9.0}$ | $83.7_{\pm 12.5}$ | $70.7_{\pm 9.6}$ | $68.7_{\pm 8.2}$ | $65.4_{\pm 0.2}$ | $49.0_{\pm 3.8}$ | $49.3_{\pm 0.9}$ | $52.1_{\pm 1.3}$ |
| Stable-only | Local | $48.3_{\pm 2.6}$ | $50.7_{\pm 3.0}$ | $51.2_{\pm 1.3}$ | $48.6_{\pm 3.0}$ | $51.8_{\pm 4.9}$ | $52.8_{\pm 2.2}$ | $50.3_{\pm 6.1}$ | $51.8_{\pm 4.9}$ | $55.1_{\pm 6.0}$ |
| | Entangling | $100.0_{\pm 0.0}$ | $99.9_{\pm 0.2}$ | $99.9_{\pm 0.2}$ | $98.7_{\pm 1.2}$ | $98.8_{\pm 0.8}$ | $98.4_{\pm 0.5}$ | $89.1_{\pm 3.4}$ | $87.2_{\pm 3.5}$ | $87.6_{\pm 5.2}$ |
| | Reuploading | $93.8_{\pm 10.5}$ | $94.9_{\pm 8.6}$ | $95.4_{\pm 7.3}$ | $64.1_{\pm 14.9}$ | $66.3_{\pm 14.5}$ | $65.9_{\pm 11.5}$ | $49.0_{\pm 3.8}$ | $49.3_{\pm 0.9}$ | $52.1_{\pm 1.3}$ |
| Shortcut-only | Local | $90.9_{\pm 1.0}$ | $50.3_{\pm 1.0}$ | $7.9_{\pm 1.2}$ | $90.9_{\pm 1.0}$ | $50.3_{\pm 1.0}$ | $7.9_{\pm 1.2}$ | $90.9_{\pm 1.0}$ | $50.4_{\pm 1.0}$ | $7.9_{\pm 1.2}$ |
| | Entangling | $67.0_{\pm 18.0}$ | $50.9_{\pm 1.0}$ | $35.3_{\pm 23.2}$ | $64.8_{\pm 19.8}$ | $51.2_{\pm 1.1}$ | $37.4_{\pm 23.8}$ | $51.2_{\pm 3.7}$ | $50.4_{\pm 1.1}$ | $51.7_{\pm 2.0}$ |
| | Reuploading | $49.3_{\pm 3.8}$ | $50.4_{\pm 1.1}$ | $52.7_{\pm 1.9}$ | $49.0_{\pm 3.8}$ | $49.3_{\pm 0.9}$ | $52.1_{\pm 1.3}$ | $49.0_{\pm 3.8}$ | $49.3_{\pm 0.9}$ | $52.1_{\pm 1.3}$ |
| Shortcut-flip | Local | $9.1_{\pm 1.0}$ | $49.7_{\pm 1.0}$ | $92.1_{\pm 1.2}$ | $9.1_{\pm 1.0}$ | $49.7_{\pm 1.0}$ | $92.1_{\pm 1.2}$ | $9.1_{\pm 1.0}$ | $49.7_{\pm 1.0}$ | $92.1_{\pm 1.2}$ |
| | Entangling | $98.9_{\pm 1.4}$ | $98.4_{\pm 1.7}$ | $99.6_{\pm 0.5}$ | $95.2_{\pm 2.0}$ | $94.8_{\pm 1.3}$ | $97.2_{\pm 1.8}$ | $87.6_{\pm 2.2}$ | $86.1_{\pm 3.0}$ | $86.7_{\pm 5.7}$ |
| | Reuploading | $84.2_{\pm 12.0}$ | $89.2_{\pm 9.4}$ | $93.9_{\pm 6.2}$ | $64.7_{\pm 3.8}$ | $67.8_{\pm 6.0}$ | $71.3_{\pm 9.2}$ | $49.0_{\pm 3.8}$ | $49.3_{\pm 0.9}$ | $52.1_{\pm 1.3}$ |

Table 39: Q-LINK: fixed-checkpoint accuracy (%) under depolarizing. Ideal/shot references: Table 37.

| Input | VQC | Weak $p = 0.005$ | | | Moderate $p = 0.5$ | | | Strong $p = 0.6$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ID | Random | OOD reversed | ID | Random | OOD reversed | ID | Random | OOD reversed |
| Original | Local | $90.9_{\pm 1.0}$ | $50.3_{\pm 1.0}$ | $7.9_{\pm 1.2}$ | $90.7_{\pm 0.6}$ | $50.4_{\pm 1.0}$ | $8.0_{\pm 1.0}$ | $90.0_{\pm 0.6}$ | $50.9_{\pm 0.5}$ | $8.7_{\pm 0.6}$ |
| | Entangling | $99.3_{\pm 0.3}$ | $99.0_{\pm 1.0}$ | $98.6_{\pm 2.0}$ | $95.9_{\pm 1.6}$ | $89.8_{\pm 9.4}$ | $83.3_{\pm 17.4}$ | $86.1_{\pm 6.1}$ | $75.8_{\pm 10.8}$ | $69.8_{\pm 19.8}$ |
| | Reuploading | $94.4_{\pm 6.1}$ | $89.1_{\pm 9.0}$ | $83.8_{\pm 12.6}$ | $52.6_{\pm 0.7}$ | $51.1_{\pm 2.2}$ | $52.9_{\pm 2.2}$ | $49.9_{\pm 3.3}$ | $49.9_{\pm 1.5}$ | $52.1_{\pm 1.3}$ |
| Stable-only | Local | $48.4_{\pm 2.8}$ | $50.6_{\pm 2.8}$ | $50.9_{\pm 1.7}$ | $46.8_{\pm 0.2}$ | $49.3_{\pm 0.9}$ | $50.1_{\pm 2.9}$ | $46.8_{\pm 0.2}$ | $49.3_{\pm 0.9}$ | $50.1_{\pm 2.9}$ |
| | Entangling | $100.0_{\pm 0.0}$ | $99.9_{\pm 0.2}$ | $99.9_{\pm 0.2}$ | $94.4_{\pm 7.9}$ | $92.8_{\pm 10.5}$ | $94.7_{\pm 7.3}$ | $80.7_{\pm 23.5}$ | $80.1_{\pm 22.7}$ | $81.4_{\pm 20.6}$ |
| | Reuploading | $94.0_{\pm 10.1}$ | $95.1_{\pm 8.2}$ | $95.2_{\pm 7.7}$ | $55.3_{\pm 5.4}$ | $56.7_{\pm 8.1}$ | $58.4_{\pm 6.9}$ | $52.3_{\pm 5.2}$ | $53.3_{\pm 7.3}$ | $55.2_{\pm 6.0}$ |
| Shortcut-only | Local | $90.9_{\pm 1.0}$ | $50.3_{\pm 1.0}$ | $7.9_{\pm 1.2}$ | $90.7_{\pm 0.6}$ | $50.6_{\pm 0.8}$ | $7.9_{\pm 1.2}$ | $90.3_{\pm 0.0}$ | $51.0_{\pm 0.3}$ | $8.7_{\pm 0.6}$ |
| | Entangling | $69.6_{\pm 17.1}$ | $50.9_{\pm 1.3}$ | $33.6_{\pm 22.2}$ | $52.7_{\pm 1.5}$ | $50.4_{\pm 0.5}$ | $50.6_{\pm 2.1}$ | $51.8_{\pm 2.7}$ | $50.6_{\pm 1.0}$ | $51.2_{\pm 1.8}$ |
| | Reuploading | $49.7_{\pm 3.8}$ | $50.1_{\pm 0.5}$ | $52.4_{\pm 2.2}$ | $49.0_{\pm 3.8}$ | $49.3_{\pm 0.9}$ | $52.0_{\pm 1.2}$ | $49.0_{\pm 3.8}$ | $49.3_{\pm 0.9}$ | $52.0_{\pm 1.2}$ |
| Shortcut-flip | Local | $9.1_{\pm 1.0}$ | $49.7_{\pm 1.0}$ | $92.1_{\pm 1.2}$ | $9.3_{\pm 0.6}$ | $49.7_{\pm 1.0}$ | $91.7_{\pm 0.6}$ | $10.0_{\pm 0.6}$ | $49.3_{\pm 0.6}$ | $91.2_{\pm 0.7}$ |
| | Entangling | $99.1_{\pm 1.0}$ | $98.7_{\pm 1.8}$ | $99.7_{\pm 0.3}$ | $84.0_{\pm 19.4}$ | $88.4_{\pm 10.9}$ | $94.9_{\pm 1.5}$ | $70.9_{\pm 22.8}$ | $74.6_{\pm 13.0}$ | $84.9_{\pm 7.4}$ |
| | Reuploading | $84.4_{\pm 10.9}$ | $89.4_{\pm 8.5}$ | $94.0_{\pm 6.3}$ | $50.2_{\pm 3.2}$ | $51.4_{\pm 2.2}$ | $55.0_{\pm 1.9}$ | $49.1_{\pm 3.7}$ | $50.0_{\pm 1.7}$ | $52.7_{\pm 1.9}$ |

Table 40: Q-LINK: fixed-checkpoint accuracy (%) under phase damping. Ideal/shot references: Table 37.

the Appendix C.5 insertion schedule. Mixed training cycles channels, one per update, not a composite. Common ideal evaluation first separates learned-predictor from evaluation-channel changes.

Tables 41 and 42 average five tasks within seed, then summarize three seed means per architecture. Tables 43–45 and 46–48 average architectures within seed per task. These marginal views do not identify joint task-by-architecture effects. Accuracy is in percent, Gap is held-out ID minus OOD reversed in percentage points, and flip sensitivity is a fraction; entries are mean ± seed SD ($n = 3$).

Under ideal evaluation, local reversal accuracy stays within 36.3–37.0%. Strong phase-aware training lowers entangling reversal and stable-only accuracy while increasing flip sensitivity; reuploading instead shows a modest mean reversal gain relative to seed variation. Training noise therefore has no uniform direction. Subject to masking caveats, these diagnostics suggest altered reliance but do not measure initial feature-gradient alignment (Proposition 7).

Noisy evaluation uses each checkpoint's designated channel/tier; mixed checkpoints average separate evaluations under all three channels of that tier. The archived clean-trained noisy baseline pools architectures and averages channels; it is not a matched architecture-specific baseline and is not subtracted to claim training benefits. A smaller ID–OOD gap may reflect ID loss. Comparisons are descriptive three-seed summaries without paired significance tests.

**Additional split and optimization information.** Figures 12–14 add Random accuracy absent from tables, averaging architectures within seed and summarizing seed uncertainty. Figures 15 and 16 add optimization trajectories. Loss changes identify neither the learned feature nor feature-gradient alignment.

### C.7 Encoding sensitivity and feature reliance

We vary the shared generator ($R_Y/R_X$), $R_Y$ angle multiplier $s$ $(0.5, 1, 2)$, and reuploading-stage order ($R_Y$–$R_Y$, $R_X$–$R_X$, $R_Y$–$R_X$, $R_X$–$R_Y$). Generator/scale comparisons cover all

Table 41: **Architecture-specific training results under ideal evaluation.** Tasks averaged within seed ($n = 3$); architectures separate. Metric conventions: Appendix C.6.

| Training | ID | OOD rev. | Gap | Stable-only | Shortcut-only | Flip sens. |
|---|---|---|---|---|---|---|
| **Local VQC** | | | | | | |
| Clean | $90.9_{\pm 0.2}$ | $37.0_{\pm 1.1}$ | $53.9_{\pm 1.0}$ | $58.4_{\pm 5.3}$ | $82.0_{\pm 1.0}$ | $0.556_{\pm 0.003}$ |
| Amplitude (M) | $90.7_{\pm 0.5}$ | $36.6_{\pm 1.0}$ | $54.1_{\pm 1.1}$ | $56.5_{\pm 5.0}$ | $82.0_{\pm 1.0}$ | $0.574_{\pm 0.001}$ |
| Depolarizing (M) | $91.3_{\pm 0.3}$ | $36.9_{\pm 1.3}$ | $54.4_{\pm 1.5}$ | $56.2_{\pm 6.9}$ | $82.2_{\pm 1.0}$ | $0.597_{\pm 0.001}$ |
| Phase (M) | $91.3_{\pm 0.3}$ | $37.0_{\pm 1.1}$ | $54.3_{\pm 1.3}$ | $54.7_{\pm 4.4}$ | $82.2_{\pm 0.8}$ | $0.587_{\pm 0.001}$ |
| Amplitude (S) | $89.5_{\pm 2.3}$ | $36.3_{\pm 1.0}$ | $53.2_{\pm 2.1}$ | $53.6_{\pm 4.6}$ | $81.7_{\pm 0.8}$ | $0.587_{\pm 0.000}$ |
| Depolarizing (S) | $91.3_{\pm 0.3}$ | $36.9_{\pm 1.2}$ | $54.4_{\pm 1.4}$ | $60.5_{\pm 2.7}$ | $82.3_{\pm 0.8}$ | $0.604_{\pm 0.000}$ |
| Phase (S) | $91.3_{\pm 0.2}$ | $36.9_{\pm 1.1}$ | $54.4_{\pm 1.3}$ | $54.8_{\pm 6.2}$ | $82.1_{\pm 0.9}$ | $0.593_{\pm 0.001}$ |
| Mixed (M) | $91.2_{\pm 0.3}$ | $36.7_{\pm 1.1}$ | $54.6_{\pm 1.3}$ | $53.7_{\pm 7.3}$ | $82.3_{\pm 0.8}$ | $0.587_{\pm 0.001}$ |
| Mixed (S) | $91.1_{\pm 0.5}$ | $36.8_{\pm 1.2}$ | $54.3_{\pm 1.6}$ | $58.8_{\pm 9.1}$ | $82.3_{\pm 0.8}$ | $0.597_{\pm 0.001}$ |
| **Entangling VQC** | | | | | | |
| Clean | $86.8_{\pm 3.5}$ | $53.7_{\pm 1.0}$ | $33.1_{\pm 2.8}$ | $71.4_{\pm 4.2}$ | $64.5_{\pm 5.4}$ | $0.314_{\pm 0.014}$ |
| Amplitude (M) | $80.6_{\pm 4.4}$ | $54.7_{\pm 1.3}$ | $26.0_{\pm 5.2}$ | $71.4_{\pm 5.2}$ | $63.3_{\pm 4.2}$ | $0.270_{\pm 0.025}$ |
| Depolarizing (M) | $86.4_{\pm 3.3}$ | $53.4_{\pm 0.3}$ | $33.0_{\pm 3.2}$ | $69.9_{\pm 9.1}$ | $59.3_{\pm 2.6}$ | $0.391_{\pm 0.021}$ |
| Phase (M) | $88.3_{\pm 2.4}$ | $51.9_{\pm 1.2}$ | $36.4_{\pm 3.6}$ | $64.3_{\pm 5.6}$ | $62.8_{\pm 4.8}$ | $0.387_{\pm 0.025}$ |
| Amplitude (S) | $75.3_{\pm 2.2}$ | $54.6_{\pm 1.7}$ | $20.7_{\pm 3.2}$ | $63.2_{\pm 5.9}$ | $59.0_{\pm 1.0}$ | $0.228_{\pm 0.017}$ |
| Depolarizing (S) | $87.6_{\pm 3.9}$ | $56.0_{\pm 0.9}$ | $31.7_{\pm 3.1}$ | $62.2_{\pm 5.7}$ | $63.7_{\pm 8.7}$ | $0.364_{\pm 0.008}$ |
| Phase (S) | $85.5_{\pm 5.1}$ | $42.0_{\pm 4.6}$ | $43.6_{\pm 5.4}$ | $62.2_{\pm 5.2}$ | $67.8_{\pm 3.8}$ | $0.474_{\pm 0.034}$ |
| Mixed (M) | $88.9_{\pm 0.5}$ | $54.6_{\pm 1.5}$ | $34.3_{\pm 2.0}$ | $63.1_{\pm 5.4}$ | $65.7_{\pm 0.6}$ | $0.355_{\pm 0.034}$ |
| Mixed (S) | $87.2_{\pm 1.1}$ | $49.0_{\pm 7.8}$ | $38.1_{\pm 6.8}$ | $58.6_{\pm 2.7}$ | $61.7_{\pm 4.4}$ | $0.371_{\pm 0.061}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $74.1_{\pm 2.9}$ | $46.0_{\pm 3.6}$ | $28.1_{\pm 0.9}$ | $54.9_{\pm 6.9}$ | $62.9_{\pm 2.6}$ | $0.323_{\pm 0.009}$ |
| Amplitude (M) | $75.4_{\pm 2.5}$ | $49.6_{\pm 2.9}$ | $25.8_{\pm 2.3}$ | $58.3_{\pm 7.8}$ | $62.8_{\pm 6.1}$ | $0.337_{\pm 0.019}$ |
| Depolarizing (M) | $69.2_{\pm 1.7}$ | $43.8_{\pm 1.6}$ | $25.4_{\pm 3.0}$ | $59.6_{\pm 3.5}$ | $61.7_{\pm 5.8}$ | $0.393_{\pm 0.026}$ |
| Phase (M) | $66.9_{\pm 0.1}$ | $47.8_{\pm 3.5}$ | $19.1_{\pm 3.4}$ | $72.3_{\pm 3.9}$ | $65.7_{\pm 2.0}$ | $0.315_{\pm 0.007}$ |
| Amplitude (S) | $67.9_{\pm 0.4}$ | $48.2_{\pm 2.8}$ | $19.7_{\pm 2.4}$ | $65.0_{\pm 5.1}$ | $58.7_{\pm 9.0}$ | $0.276_{\pm 0.004}$ |
| Depolarizing (S) | $63.6_{\pm 2.0}$ | $47.7_{\pm 3.0}$ | $15.9_{\pm 1.1}$ | $60.4_{\pm 2.1}$ | $62.7_{\pm 5.5}$ | $0.313_{\pm 0.005}$ |
| Phase (S) | $66.9_{\pm 2.8}$ | $49.4_{\pm 2.0}$ | $17.5_{\pm 1.1}$ | $69.0_{\pm 4.3}$ | $67.1_{\pm 6.2}$ | $0.311_{\pm 0.018}$ |
| Mixed (M) | $73.6_{\pm 1.6}$ | $49.6_{\pm 2.3}$ | $24.0_{\pm 1.7}$ | $68.9_{\pm 17.0}$ | $67.4_{\pm 4.5}$ | $0.337_{\pm 0.011}$ |
| Mixed (S) | $69.0_{\pm 2.7}$ | $47.5_{\pm 2.3}$ | $21.5_{\pm 2.5}$ | $77.7_{\pm 6.8}$ | $70.0_{\pm 8.6}$ | $0.305_{\pm 0.025}$ |

Table 42: **Architecture-specific training results under designated noisy evaluation.** Tasks averaged within seed ($n = 3$); architectures separate. Metric conventions: Appendix C.6. Mixed rows average three channels at the stated tier.

| Training | ID | OOD rev. | Gap | Stable-only | Shortcut-only | Flip sens. |
|---|---|---|---|---|---|---|
| **Local VQC** | | | | | | |
| Amplitude (M) | $91.1_{\pm 0.3}$ | $37.1_{\pm 1.2}$ | $54.0_{\pm 1.2}$ | $60.7_{\pm 2.9}$ | $81.9_{\pm 0.9}$ | $0.555_{\pm 0.002}$ |
| Depolarizing (M) | $91.3_{\pm 0.2}$ | $36.7_{\pm 0.9}$ | $54.6_{\pm 1.0}$ | $54.6_{\pm 6.6}$ | $82.2_{\pm 1.1}$ | $0.554_{\pm 0.002}$ |
| Phase (M) | $91.3_{\pm 0.3}$ | $36.9_{\pm 1.0}$ | $54.3_{\pm 1.2}$ | $57.4_{\pm 5.3}$ | $82.3_{\pm 0.8}$ | $0.554_{\pm 0.004}$ |
| Amplitude (S) | $91.0_{\pm 0.4}$ | $36.5_{\pm 1.3}$ | $54.5_{\pm 1.2}$ | $57.9_{\pm 5.7}$ | $79.3_{\pm 3.6}$ | $0.554_{\pm 0.002}$ |
| Depolarizing (S) | $91.3_{\pm 0.3}$ | $36.6_{\pm 0.8}$ | $54.7_{\pm 0.9}$ | $59.6_{\pm 2.3}$ | $82.4_{\pm 1.0}$ | $0.544_{\pm 0.000}$ |
| Phase (S) | $91.3_{\pm 0.3}$ | $36.9_{\pm 1.0}$ | $54.4_{\pm 1.2}$ | $57.5_{\pm 5.1}$ | $82.2_{\pm 0.9}$ | $0.554_{\pm 0.003}$ |
| Mixed (M) | $91.2_{\pm 0.3}$ | $36.7_{\pm 1.1}$ | $54.5_{\pm 1.2}$ | $55.3_{\pm 5.9}$ | $81.9_{\pm 0.5}$ | $0.550_{\pm 0.002}$ |
| Mixed (S) | $91.2_{\pm 0.4}$ | $36.6_{\pm 0.9}$ | $54.6_{\pm 1.2}$ | $55.9_{\pm 5.3}$ | $81.4_{\pm 0.5}$ | $0.533_{\pm 0.002}$ |
| **Entangling VQC** | | | | | | |
| Amplitude (M) | $87.2_{\pm 2.5}$ | $53.7_{\pm 0.9}$ | $33.5_{\pm 2.1}$ | $70.1_{\pm 0.4}$ | $62.2_{\pm 4.7}$ | $0.310_{\pm 0.013}$ |
| Depolarizing (M) | $85.4_{\pm 4.1}$ | $53.3_{\pm 1.1}$ | $32.2_{\pm 4.5}$ | $69.4_{\pm 2.1}$ | $59.4_{\pm 5.0}$ | $0.233_{\pm 0.011}$ |
| Phase (M) | $86.8_{\pm 4.4}$ | $47.2_{\pm 4.1}$ | $39.6_{\pm 8.4}$ | $64.0_{\pm 5.3}$ | $66.3_{\pm 2.7}$ | $0.327_{\pm 0.042}$ |
| Amplitude (S) | $90.6_{\pm 2.1}$ | $53.4_{\pm 0.6}$ | $37.2_{\pm 1.7}$ | $66.8_{\pm 6.0}$ | $61.1_{\pm 4.9}$ | $0.320_{\pm 0.010}$ |
| Depolarizing (S) | $87.7_{\pm 2.9}$ | $55.9_{\pm 1.0}$ | $31.8_{\pm 2.9}$ | $62.5_{\pm 5.4}$ | $59.8_{\pm 4.4}$ | $0.103_{\pm 0.035}$ |
| Phase (S) | $83.0_{\pm 4.2}$ | $38.5_{\pm 4.2}$ | $44.5_{\pm 8.0}$ | $59.0_{\pm 9.2}$ | $71.1_{\pm 2.8}$ | $0.356_{\pm 0.041}$ |
| Mixed (M) | $87.1_{\pm 0.6}$ | $51.7_{\pm 1.1}$ | $35.5_{\pm 0.8}$ | $65.6_{\pm 4.1}$ | $61.5_{\pm 1.6}$ | $0.244_{\pm 0.016}$ |
| Mixed (S) | $78.1_{\pm 2.6}$ | $50.0_{\pm 7.3}$ | $28.1_{\pm 4.7}$ | $59.0_{\pm 4.4}$ | $55.1_{\pm 3.0}$ | $0.184_{\pm 0.031}$ |
| **Reuploading VQC** | | | | | | |
| Amplitude (M) | $79.5_{\pm 2.6}$ | $46.9_{\pm 3.4}$ | $32.6_{\pm 4.1}$ | $61.8_{\pm 4.0}$ | $67.6_{\pm 9.2}$ | $0.333_{\pm 0.025}$ |
| Depolarizing (M) | $69.1_{\pm 2.1}$ | $43.9_{\pm 2.3}$ | $25.2_{\pm 3.7}$ | $58.6_{\pm 3.7}$ | $61.4_{\pm 7.8}$ | $0.170_{\pm 0.023}$ |
| Phase (M) | $69.5_{\pm 1.4}$ | $48.0_{\pm 1.3}$ | $21.5_{\pm 2.6}$ | $74.3_{\pm 3.3}$ | $65.3_{\pm 1.7}$ | $0.159_{\pm 0.012}$ |
| Amplitude (S) | $77.0_{\pm 2.0}$ | $47.0_{\pm 2.0}$ | $29.9_{\pm 0.2}$ | $78.2_{\pm 4.1}$ | $70.2_{\pm 3.1}$ | $0.326_{\pm 0.014}$ |
| Depolarizing (S) | $57.6_{\pm 1.8}$ | $46.7_{\pm 2.3}$ | $10.9_{\pm 3.3}$ | $54.8_{\pm 7.9}$ | $60.2_{\pm 2.8}$ | $0.054_{\pm 0.005}$ |
| Phase (S) | $69.0_{\pm 1.6}$ | $50.8_{\pm 3.3}$ | $18.3_{\pm 3.8}$ | $72.9_{\pm 3.2}$ | $63.8_{\pm 2.8}$ | $0.121_{\pm 0.023}$ |
| Mixed (M) | $71.4_{\pm 1.6}$ | $50.9_{\pm 1.8}$ | $20.5_{\pm 0.3}$ | $66.5_{\pm 8.7}$ | $65.2_{\pm 1.0}$ | $0.154_{\pm 0.009}$ |
| Mixed (S) | $64.0_{\pm 1.3}$ | $52.0_{\pm 2.7}$ | $12.0_{\pm 4.0}$ | $66.1_{\pm 3.7}$ | $59.1_{\pm 3.3}$ | $0.097_{\pm 0.013}$ |

Table 43: **Task-level training summaries under ideal evaluation (part 1 of 3).** Architectures averaged within seed ($n = 3$), not model-resolved task effects. Metric conventions: Appendix C.6.

| Training | ID | OOD rev. | Gap | Stable-only | Shortcut-only | Flip sens. |
|---|---|---|---|---|---|---|
| **Q-SPHERE** | | | | | | |
| Clean | $87.2_{\pm 0.6}$ | $17.7_{\pm 2.4}$ | $69.5_{\pm 2.9}$ | $55.6_{\pm 1.3}$ | $83.2_{\pm 2.3}$ | $0.682_{\pm 0.025}$ |
| Amplitude (M) | $82.3_{\pm 1.2}$ | $21.6_{\pm 1.4}$ | $60.7_{\pm 2.3}$ | $54.0_{\pm 3.4}$ | $86.0_{\pm 0.3}$ | $0.673_{\pm 0.015}$ |
| Depolarizing (M) | $84.6_{\pm 0.8}$ | $22.0_{\pm 5.5}$ | $62.6_{\pm 5.9}$ | $54.3_{\pm 2.0}$ | $80.6_{\pm 8.5}$ | $0.795_{\pm 0.047}$ |
| Phase (M) | $82.6_{\pm 1.3}$ | $24.5_{\pm 1.8}$ | $58.1_{\pm 2.5}$ | $53.9_{\pm 2.8}$ | $84.5_{\pm 1.8}$ | $0.703_{\pm 0.022}$ |
| Amplitude (S) | $75.8_{\pm 4.0}$ | $25.6_{\pm 4.4}$ | $50.2_{\pm 8.3}$ | $49.6_{\pm 3.9}$ | $82.9_{\pm 0.8}$ | $0.619_{\pm 0.038}$ |
| Depolarizing (S) | $80.3_{\pm 1.3}$ | $26.8_{\pm 1.8}$ | $53.5_{\pm 1.5}$ | $54.7_{\pm 0.4}$ | $77.7_{\pm 7.2}$ | $0.690_{\pm 0.025}$ |
| Phase (S) | $82.5_{\pm 0.6}$ | $25.6_{\pm 2.0}$ | $57.0_{\pm 2.5}$ | $53.9_{\pm 2.9}$ | $83.1_{\pm 3.4}$ | $0.698_{\pm 0.017}$ |
| Mixed (M) | $84.7_{\pm 1.7}$ | $22.7_{\pm 3.1}$ | $62.0_{\pm 4.1}$ | $53.2_{\pm 2.7}$ | $84.0_{\pm 2.8}$ | $0.699_{\pm 0.012}$ |
| Mixed (S) | $84.0_{\pm 1.8}$ | $20.8_{\pm 5.2}$ | $63.3_{\pm 6.9}$ | $54.5_{\pm 1.4}$ | $86.7_{\pm 0.6}$ | $0.677_{\pm 0.042}$ |
| **Spirals-Signatures** | | | | | | |
| Clean | $88.7_{\pm 1.3}$ | $13.3_{\pm 1.0}$ | $75.4_{\pm 2.2}$ | $53.1_{\pm 2.7}$ | $90.6_{\pm 0.8}$ | $0.790_{\pm 0.026}$ |
| Amplitude (M) | $83.3_{\pm 2.9}$ | $17.9_{\pm 3.1}$ | $65.4_{\pm 5.9}$ | $50.3_{\pm 0.8}$ | $83.4_{\pm 2.8}$ | $0.720_{\pm 0.049}$ |
| Depolarizing (M) | $87.0_{\pm 1.1}$ | $13.1_{\pm 2.4}$ | $73.9_{\pm 1.8}$ | $53.0_{\pm 1.2}$ | $88.7_{\pm 2.1}$ | $0.898_{\pm 0.021}$ |
| Phase (M) | $88.1_{\pm 1.9}$ | $15.3_{\pm 1.8}$ | $72.8_{\pm 3.6}$ | $53.4_{\pm 0.8}$ | $90.3_{\pm 0.8}$ | $0.851_{\pm 0.042}$ |
| Amplitude (S) | $78.2_{\pm 2.6}$ | $20.9_{\pm 2.7}$ | $57.3_{\pm 5.2}$ | $50.5_{\pm 1.2}$ | $80.1_{\pm 7.0}$ | $0.671_{\pm 0.031}$ |
| Depolarizing (S) | $84.4_{\pm 3.1}$ | $15.5_{\pm 1.0}$ | $69.0_{\pm 3.2}$ | $49.2_{\pm 1.7}$ | $90.7_{\pm 0.7}$ | $0.904_{\pm 0.031}$ |
| Phase (S) | $86.9_{\pm 2.1}$ | $18.0_{\pm 1.9}$ | $68.9_{\pm 1.8}$ | $55.9_{\pm 1.8}$ | $87.4_{\pm 4.8}$ | $0.798_{\pm 0.041}$ |
| Mixed (M) | $87.2_{\pm 2.7}$ | $14.7_{\pm 1.6}$ | $72.5_{\pm 4.2}$ | $52.0_{\pm 2.0}$ | $89.3_{\pm 2.6}$ | $0.831_{\pm 0.044}$ |
| Mixed (S) | $86.3_{\pm 2.0}$ | $14.5_{\pm 1.3}$ | $71.8_{\pm 3.1}$ | $51.9_{\pm 2.5}$ | $89.7_{\pm 0.4}$ | $0.812_{\pm 0.014}$ |

Table 44: **Task-level training summaries under ideal evaluation (part 2 of 3).** Architectures averaged within seed ($n = 3$), not model-resolved task effects. Metric conventions: Appendix C.6.

| Training | ID | OOD rev. | Gap | Stable-only | Shortcut-only | Flip sens. |
|---|---|---|---|---|---|---|
| **Spurious-Fourier** | | | | | | |
| Clean | $79.9_{\pm 3.4}$ | $80.0_{\pm 3.5}$ | $-0.1_{\pm 2.2}$ | $66.7_{\pm 0.0}$ | $49.6_{\pm 2.4}$ | $0.126_{\pm 0.025}$ |
| Amplitude (M) | $75.3_{\pm 3.0}$ | $75.9_{\pm 2.9}$ | $-0.6_{\pm 4.6}$ | $77.0_{\pm 9.0}$ | $51.9_{\pm 0.2}$ | $0.164_{\pm 0.009}$ |
| Depolarizing (M) | $75.0_{\pm 1.8}$ | $75.4_{\pm 2.4}$ | $-0.4_{\pm 2.0}$ | $66.7_{\pm 0.0}$ | $51.4_{\pm 3.6}$ | $0.147_{\pm 0.022}$ |
| Phase (M) | $78.4_{\pm 0.9}$ | $79.3_{\pm 2.2}$ | $-0.9_{\pm 2.1}$ | $100.0_{\pm 0.0}$ | $52.0_{\pm 1.5}$ | $0.144_{\pm 0.014}$ |
| Amplitude (S) | $70.5_{\pm 3.8}$ | $71.1_{\pm 2.3}$ | $-0.6_{\pm 5.3}$ | $88.2_{\pm 10.3}$ | $51.0_{\pm 1.5}$ | $0.115_{\pm 0.007}$ |
| Depolarizing (S) | $77.6_{\pm 2.6}$ | $78.9_{\pm 1.1}$ | $-1.3_{\pm 1.8}$ | $66.7_{\pm 0.0}$ | $51.9_{\pm 3.7}$ | $0.097_{\pm 0.028}$ |
| Phase (S) | $80.4_{\pm 1.7}$ | $80.7_{\pm 0.1}$ | $-0.3_{\pm 1.7}$ | $100.0_{\pm 0.0}$ | $52.1_{\pm 5.3}$ | $0.145_{\pm 0.010}$ |
| Mixed (M) | $79.6_{\pm 1.6}$ | $79.0_{\pm 1.1}$ | $0.6_{\pm 2.4}$ | $88.9_{\pm 19.2}$ | $52.1_{\pm 4.8}$ | $0.171_{\pm 0.016}$ |
| Mixed (S) | $75.3_{\pm 5.0}$ | $74.6_{\pm 4.2}$ | $0.7_{\pm 3.2}$ | $100.0_{\pm 0.0}$ | $52.6_{\pm 2.9}$ | $0.165_{\pm 0.017}$ |
| **Q-FOS** | | | | | | |
| Clean | $69.1_{\pm 6.8}$ | $53.5_{\pm 1.2}$ | $15.6_{\pm 6.4}$ | $51.7_{\pm 7.6}$ | $55.6_{\pm 11.0}$ | $0.018_{\pm 0.002}$ |
| Amplitude (M) | $76.3_{\pm 4.4}$ | $56.4_{\pm 1.4}$ | $19.9_{\pm 3.7}$ | $48.1_{\pm 15.0}$ | $60.2_{\pm 14.9}$ | $0.018_{\pm 0.003}$ |
| Depolarizing (M) | $69.3_{\pm 3.8}$ | $53.0_{\pm 2.3}$ | $16.3_{\pm 2.9}$ | $53.2_{\pm 28.5}$ | $51.3_{\pm 2.1}$ | $0.038_{\pm 0.005}$ |
| Phase (M) | $72.5_{\pm 6.8}$ | $55.0_{\pm 1.5}$ | $17.5_{\pm 5.8}$ | $32.6_{\pm 7.2}$ | $52.5_{\pm 2.2}$ | $0.017_{\pm 0.001}$ |
| Amplitude (S) | $71.2_{\pm 8.5}$ | $52.5_{\pm 2.9}$ | $18.7_{\pm 5.8}$ | $33.2_{\pm 14.4}$ | $52.6_{\pm 1.8}$ | $0.028_{\pm 0.004}$ |
| Depolarizing (S) | $69.0_{\pm 4.1}$ | $52.7_{\pm 0.5}$ | $16.3_{\pm 3.6}$ | $55.8_{\pm 9.6}$ | $61.6_{\pm 17.0}$ | $0.047_{\pm 0.008}$ |
| Phase (S) | $71.5_{\pm 10.9}$ | $56.1_{\pm 2.1}$ | $15.4_{\pm 8.8}$ | $32.3_{\pm 19.0}$ | $59.4_{\pm 7.0}$ | $0.014_{\pm 0.002}$ |
| Mixed (M) | $75.9_{\pm 6.5}$ | $56.3_{\pm 4.3}$ | $19.6_{\pm 2.3}$ | $35.2_{\pm 16.0}$ | $55.5_{\pm 3.7}$ | $0.025_{\pm 0.008}$ |
| Mixed (S) | $76.4_{\pm 3.7}$ | $56.1_{\pm 4.0}$ | $20.3_{\pm 0.7}$ | $39.8_{\pm 10.6}$ | $54.4_{\pm 7.9}$ | $0.023_{\pm 0.006}$ |

Table 45: **Task-level training summaries under ideal evaluation (part 3 of 3).** Architectures averaged within seed ($n = 3$), not model-resolved task effects. Metric conventions: Appendix C.6.

| Training | ID | OOD rev. | Gap | Stable-only | Shortcut-only | Flip sens. |
|---|---|---|---|---|---|---|
| **Q-LINK** | | | | | | |
| Clean | $94.9_{\pm 2.4}$ | $63.5_{\pm 4.3}$ | $31.4_{\pm 2.0}$ | $80.8_{\pm 4.0}$ | $70.1_{\pm 5.1}$ | $0.373_{\pm 0.034}$ |
| Amplitude (M) | $94.0_{\pm 1.8}$ | $63.1_{\pm 7.3}$ | $30.9_{\pm 5.8}$ | $80.9_{\pm 4.0}$ | $65.3_{\pm 1.5}$ | $0.395_{\pm 0.059}$ |
| Depolarizing (M) | $95.5_{\pm 0.4}$ | $59.9_{\pm 2.0}$ | $35.6_{\pm 2.3}$ | $82.2_{\pm 0.1}$ | $66.6_{\pm 2.7}$ | $0.423_{\pm 0.011}$ |
| Phase (M) | $89.1_{\pm 4.1}$ | $53.7_{\pm 8.2}$ | $35.3_{\pm 5.0}$ | $79.0_{\pm 3.0}$ | $72.0_{\pm 8.1}$ | $0.432_{\pm 0.029}$ |
| Amplitude (S) | $92.0_{\pm 1.9}$ | $61.6_{\pm 6.4}$ | $30.5_{\pm 8.0}$ | $81.6_{\pm 1.0}$ | $65.7_{\pm 12.3}$ | $0.386_{\pm 0.032}$ |
| Depolarizing (S) | $92.8_{\pm 1.6}$ | $60.3_{\pm 2.9}$ | $32.5_{\pm 2.3}$ | $78.7_{\pm 4.2}$ | $66.0_{\pm 10.8}$ | $0.396_{\pm 0.026}$ |
| Phase (S) | $85.0_{\pm 2.5}$ | $33.5_{\pm 6.0}$ | $51.5_{\pm 8.3}$ | $67.6_{\pm 2.0}$ | $79.9_{\pm 7.8}$ | $0.641_{\pm 0.072}$ |
| Mixed (M) | $95.6_{\pm 1.0}$ | $62.1_{\pm 2.2}$ | $33.5_{\pm 1.4}$ | $80.2_{\pm 2.3}$ | $78.3_{\pm 2.6}$ | $0.404_{\pm 0.030}$ |
| Mixed (S) | $90.1_{\pm 2.1}$ | $56.4_{\pm 8.8}$ | $33.7_{\pm 10.3}$ | $78.9_{\pm 1.4}$ | $73.3_{\pm 6.9}$ | $0.445_{\pm 0.072}$ |

Table 46: **Task-level training summaries under designated noisy evaluation (part 1 of 3).** Architectures averaged within seed ($n = 3$), not model-resolved task effects. Metric conventions: Appendix C.6.

| Training | ID | OOD rev. | Gap | Stable-only | Shortcut-only | Flip sens. |
|---|---|---|---|---|---|---|
| **Q-SPHERE** | | | | | | |
| Amplitude (M) | $88.3_{\pm 0.7}$ | $17.3_{\pm 1.8}$ | $71.0_{\pm 1.8}$ | $55.1_{\pm 2.4}$ | $87.6_{\pm 0.8}$ | $0.709_{\pm 0.008}$ |
| Depolarizing (M) | $83.9_{\pm 2.1}$ | $22.4_{\pm 6.7}$ | $61.5_{\pm 8.3}$ | $55.0_{\pm 1.6}$ | $81.0_{\pm 10.1}$ | $0.509_{\pm 0.048}$ |
| Phase (M) | $83.6_{\pm 0.8}$ | $23.7_{\pm 2.5}$ | $60.0_{\pm 3.1}$ | $54.5_{\pm 1.3}$ | $84.3_{\pm 1.4}$ | $0.548_{\pm 0.011}$ |
| Amplitude (S) | $89.1_{\pm 0.5}$ | $15.6_{\pm 1.6}$ | $73.5_{\pm 1.1}$ | $54.0_{\pm 4.3}$ | $89.3_{\pm 0.2}$ | $0.742_{\pm 0.018}$ |
| Depolarizing (S) | $78.3_{\pm 0.9}$ | $28.2_{\pm 2.4}$ | $50.1_{\pm 2.9}$ | $54.9_{\pm 0.5}$ | $75.0_{\pm 0.4}$ | $0.332_{\pm 0.001}$ |
| Phase (S) | $83.3_{\pm 0.9}$ | $24.0_{\pm 2.1}$ | $59.3_{\pm 3.1}$ | $55.0_{\pm 0.7}$ | $83.1_{\pm 2.7}$ | $0.518_{\pm 0.007}$ |
| Mixed (M) | $82.5_{\pm 1.1}$ | $24.2_{\pm 2.7}$ | $58.3_{\pm 3.1}$ | $53.7_{\pm 2.7}$ | $82.8_{\pm 0.3}$ | $0.506_{\pm 0.002}$ |
| Mixed (S) | $74.6_{\pm 1.0}$ | $27.5_{\pm 2.6}$ | $47.0_{\pm 3.6}$ | $52.3_{\pm 1.3}$ | $75.5_{\pm 2.1}$ | $0.420_{\pm 0.014}$ |
| **Spirals-Signatures** | | | | | | |
| Amplitude (M) | $89.3_{\pm 0.2}$ | $12.6_{\pm 0.5}$ | $76.7_{\pm 0.7}$ | $51.7_{\pm 2.7}$ | $89.7_{\pm 0.8}$ | $0.781_{\pm 0.017}$ |
| Depolarizing (M) | $88.1_{\pm 1.0}$ | $13.7_{\pm 2.9}$ | $74.4_{\pm 1.9}$ | $53.0_{\pm 1.5}$ | $89.7_{\pm 1.0}$ | $0.667_{\pm 0.008}$ |
| Phase (M) | $88.0_{\pm 1.3}$ | $15.5_{\pm 0.7}$ | $72.5_{\pm 1.5}$ | $52.3_{\pm 0.6}$ | $90.1_{\pm 0.9}$ | $0.661_{\pm 0.032}$ |
| Amplitude (S) | $90.6_{\pm 0.5}$ | $10.4_{\pm 0.9}$ | $80.2_{\pm 1.1}$ | $52.3_{\pm 3.8}$ | $90.7_{\pm 0.6}$ | $0.816_{\pm 0.012}$ |
| Depolarizing (S) | $81.6_{\pm 2.9}$ | $17.1_{\pm 3.3}$ | $64.5_{\pm 5.4}$ | $51.3_{\pm 0.9}$ | $90.8_{\pm 0.7}$ | $0.488_{\pm 0.057}$ |
| Phase (S) | $86.6_{\pm 2.1}$ | $16.9_{\pm 1.7}$ | $69.7_{\pm 0.4}$ | $53.7_{\pm 4.7}$ | $89.6_{\pm 0.5}$ | $0.631_{\pm 0.025}$ |
| Mixed (M) | $86.0_{\pm 2.0}$ | $14.7_{\pm 1.2}$ | $71.3_{\pm 3.1}$ | $52.8_{\pm 2.0}$ | $89.5_{\pm 1.7}$ | $0.622_{\pm 0.007}$ |
| Mixed (S) | $79.1_{\pm 1.3}$ | $21.2_{\pm 0.7}$ | $57.9_{\pm 1.9}$ | $51.4_{\pm 0.8}$ | $81.7_{\pm 0.9}$ | $0.536_{\pm 0.014}$ |

Table 47: **Task-level training summaries under designated noisy evaluation (part 2 of 3).** Architectures averaged within seed ($n = 3$), not model-resolved task effects. Metric conventions: Appendix C.6.

| Training | ID | OOD rev. | Gap | Stable-only | Shortcut-only | Flip sens. |
|---|---|---|---|---|---|---|
| **Spurious-Fourier** | | | | | | |
| Amplitude (M) | $81.2_{\pm 1.5}$ | $80.1_{\pm 1.1}$ | $1.1_{\pm 1.9}$ | $77.6_{\pm 9.5}$ | $49.9_{\pm 2.1}$ | $0.119_{\pm 0.009}$ |
| Depolarizing (M) | $72.9_{\pm 2.4}$ | $73.2_{\pm 2.1}$ | $-0.3_{\pm 2.1}$ | $66.7_{\pm 0.0}$ | $50.0_{\pm 1.9}$ | $0.071_{\pm 0.007}$ |
| Phase (M) | $78.3_{\pm 1.0}$ | $78.7_{\pm 1.7}$ | $-0.4_{\pm 0.9}$ | $100.0_{\pm 0.0}$ | $51.6_{\pm 3.0}$ | $0.074_{\pm 0.002}$ |
| Amplitude (S) | $84.0_{\pm 0.6}$ | $83.8_{\pm 1.7}$ | $0.1_{\pm 1.5}$ | $100.0_{\pm 0.0}$ | $52.0_{\pm 4.5}$ | $0.079_{\pm 0.016}$ |
| Depolarizing (S) | $76.0_{\pm 0.8}$ | $75.4_{\pm 1.7}$ | $0.7_{\pm 1.6}$ | $66.7_{\pm 0.0}$ | $51.4_{\pm 2.7}$ | $0.032_{\pm 0.006}$ |
| Phase (S) | $80.2_{\pm 0.4}$ | $80.3_{\pm 1.4}$ | $-0.1_{\pm 1.3}$ | $100.0_{\pm 0.0}$ | $49.3_{\pm 3.1}$ | $0.066_{\pm 0.002}$ |
| Mixed (M) | $77.0_{\pm 1.8}$ | $76.4_{\pm 0.5}$ | $0.7_{\pm 2.2}$ | $88.8_{\pm 11.1}$ | $51.9_{\pm 5.0}$ | $0.079_{\pm 0.009}$ |
| Mixed (S) | $71.2_{\pm 3.7}$ | $72.2_{\pm 2.4}$ | $-1.0_{\pm 3.6}$ | $86.5_{\pm 4.4}$ | $51.2_{\pm 2.9}$ | $0.053_{\pm 0.003}$ |
| **Q-FOS** | | | | | | |
| Amplitude (M) | $75.0_{\pm 2.5}$ | $55.6_{\pm 2.2}$ | $19.4_{\pm 0.3}$ | $56.2_{\pm 7.0}$ | $58.0_{\pm 16.3}$ | $0.012_{\pm 0.002}$ |
| Depolarizing (M) | $69.2_{\pm 5.0}$ | $52.7_{\pm 1.8}$ | $16.5_{\pm 3.7}$ | $47.7_{\pm 13.5}$ | $56.0_{\pm 6.3}$ | $0.013_{\pm 0.005}$ |
| Phase (M) | $71.0_{\pm 10.1}$ | $54.0_{\pm 2.7}$ | $17.0_{\pm 9.0}$ | $42.8_{\pm 15.3}$ | $58.8_{\pm 7.3}$ | $0.008_{\pm 0.001}$ |
| Amplitude (S) | $71.2_{\pm 1.1}$ | $52.0_{\pm 1.3}$ | $19.3_{\pm 0.3}$ | $48.9_{\pm 14.8}$ | $52.8_{\pm 2.6}$ | $0.015_{\pm 0.005}$ |
| Depolarizing (S) | $70.4_{\pm 1.6}$ | $52.1_{\pm 2.0}$ | $18.3_{\pm 1.6}$ | $48.2_{\pm 2.0}$ | $57.0_{\pm 7.1}$ | $0.011_{\pm 0.003}$ |
| Phase (S) | $67.6_{\pm 7.9}$ | $54.3_{\pm 2.0}$ | $13.3_{\pm 6.3}$ | $41.1_{\pm 10.6}$ | $62.8_{\pm 6.4}$ | $0.006_{\pm 0.001}$ |
| Mixed (M) | $75.7_{\pm 5.1}$ | $55.8_{\pm 3.8}$ | $19.9_{\pm 1.4}$ | $36.3_{\pm 7.7}$ | $53.3_{\pm 4.1}$ | $0.012_{\pm 0.003}$ |
| Mixed (S) | $76.3_{\pm 2.9}$ | $56.2_{\pm 2.6}$ | $20.1_{\pm 2.3}$ | $38.4_{\pm 4.8}$ | $49.9_{\pm 1.7}$ | $0.007_{\pm 0.002}$ |

Table 48: **Task-level training summaries under designated noisy evaluation (part 3 of 3).** Architectures averaged within seed ($n = 3$), not model-resolved task effects. Metric conventions: Appendix C.6.

| Training | ID | OOD rev. | Gap | Stable-only | Shortcut-only | Flip sens. |
|---|---|---|---|---|---|---|
| **Q-LINK** | | | | | | |
| Amplitude (M) | $95.9_{\pm 1.7}$ | $63.8_{\pm 7.5}$ | $32.1_{\pm 5.9}$ | $80.4_{\pm 3.3}$ | $67.8_{\pm 6.4}$ | $0.375_{\pm 0.052}$ |
| Depolarizing (M) | $95.6_{\pm 0.7}$ | $61.2_{\pm 0.8}$ | $34.4_{\pm 1.6}$ | $81.9_{\pm 0.3}$ | $61.5_{\pm 0.3}$ | $0.335_{\pm 0.005}$ |
| Phase (M) | $91.7_{\pm 3.2}$ | $48.4_{\pm 10.7}$ | $43.3_{\pm 8.3}$ | $76.4_{\pm 5.2}$ | $71.6_{\pm 9.0}$ | $0.443_{\pm 0.062}$ |
| Amplitude (S) | $96.0_{\pm 0.9}$ | $66.4_{\pm 5.0}$ | $29.6_{\pm 4.1}$ | $83.1_{\pm 1.8}$ | $66.1_{\pm 4.7}$ | $0.347_{\pm 0.029}$ |
| Depolarizing (S) | $87.9_{\pm 3.4}$ | $59.2_{\pm 1.4}$ | $28.7_{\pm 3.4}$ | $73.9_{\pm 2.7}$ | $63.3_{\pm 2.6}$ | $0.307_{\pm 0.002}$ |
| Phase (S) | $87.8_{\pm 1.8}$ | $34.8_{\pm 13.2}$ | $53.0_{\pm 11.9}$ | $66.0_{\pm 7.4}$ | $77.1_{\pm 2.0}$ | $0.497_{\pm 0.076}$ |
| Mixed (M) | $95.0_{\pm 0.3}$ | $61.0_{\pm 1.7}$ | $34.0_{\pm 1.6}$ | $81.0_{\pm 0.6}$ | $70.2_{\pm 0.9}$ | $0.361_{\pm 0.019}$ |
| Mixed (S) | $87.7_{\pm 0.3}$ | $53.8_{\pm 7.2}$ | $33.8_{\pm 7.5}$ | $73.0_{\pm 4.0}$ | $67.6_{\pm 4.8}$ | $0.337_{\pm 0.051}$ |

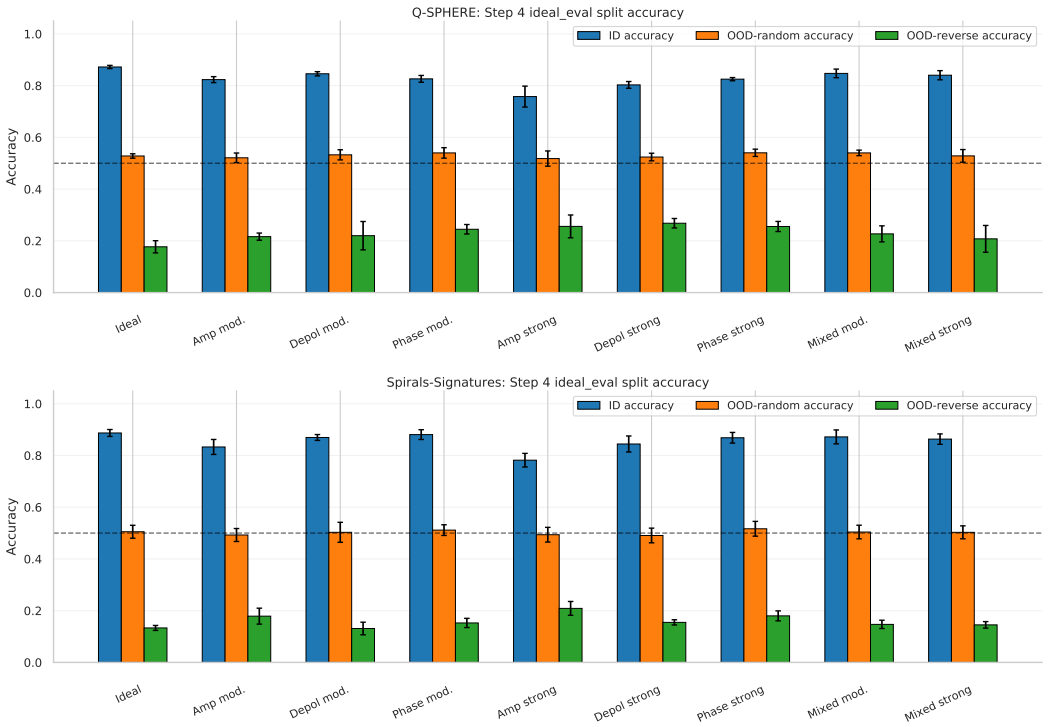


Figure 12: **Additional Random-split training information.** Geometric tasks: Q-SPHERE (top) and Spirals-Signatures (bottom). Ideal evaluation; architectures averaged within seed, with seed error bars.

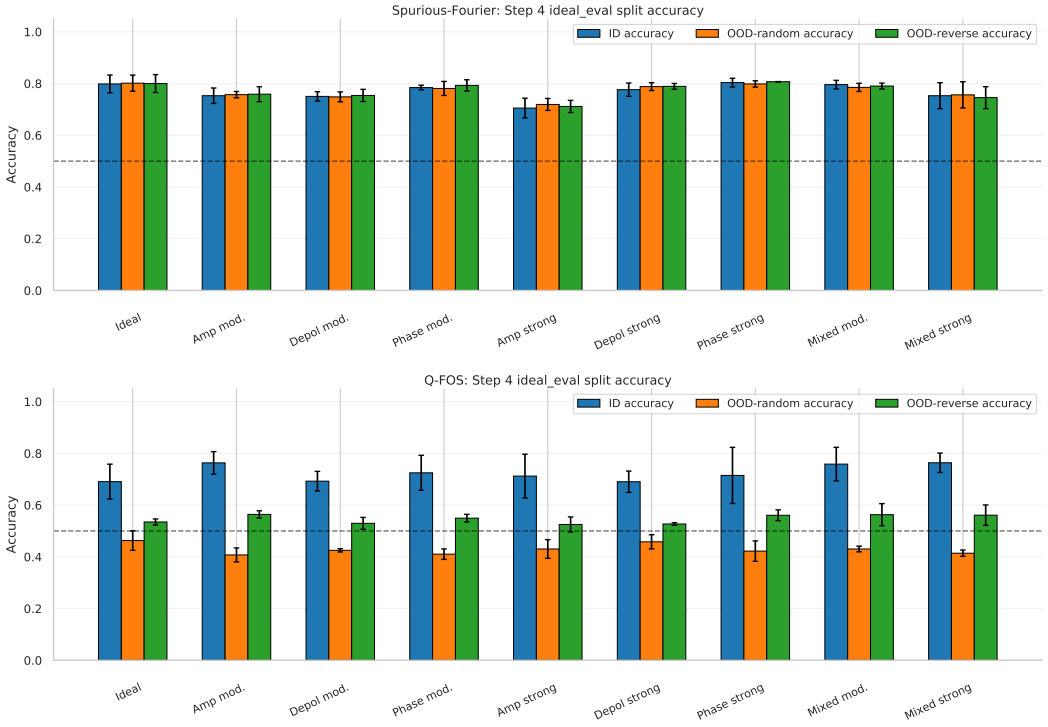


Figure 13: **Additional Random-split training information.** Spectral tasks: Spurious-Fourier (top) and Q-FOS (bottom). Ideal evaluation; architectures averaged within seed, with seed error bars.

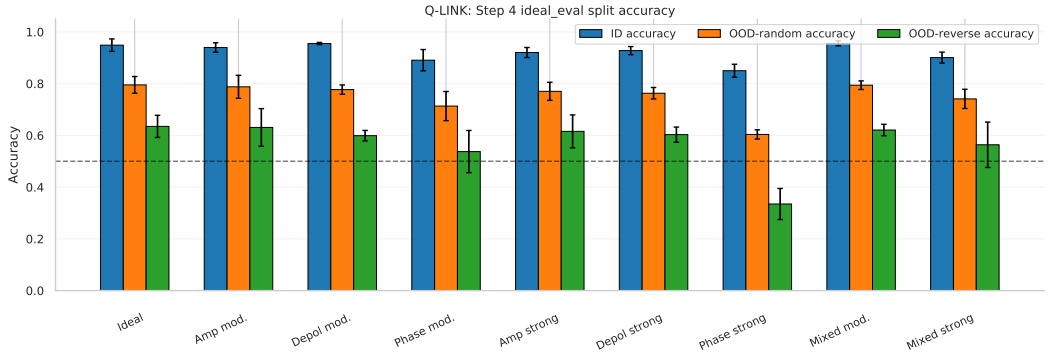


Figure 14: **Additional Random-split training information.** Relational task: Q-LINK. Ideal evaluation; architectures averaged within seed, with seed error bars.

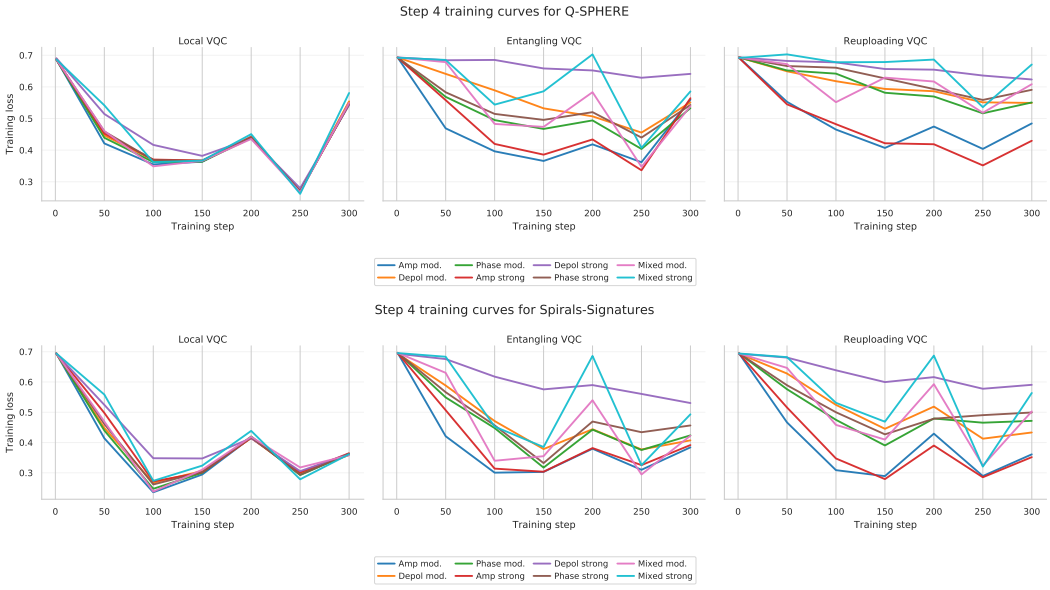


Figure 15: **Optimization trajectories under the nine training conditions.** Q-SPHERE (top) and Spirals-Signatures (bottom). Loss curves do not test feature-gradient alignment.

VQCs; stage order applies only to reuploading. Scale multiplies angles after clipping to $[-\pi, \pi]$; each stage precedes its rotations and CNOT ring. Unless specified, $s = 1$; shared $R_Y/R_X$ denote the corresponding identical-generator pairs in reuploading.

Coverage contains 285 configuration–task–model–seed records at 300 steps, six qubits, two layers, and seeds $0, 1, 2$. Repeated baselines are not independent replications. Tables 49–53 consolidate identical records from final training-time exports per task/model. Pooled ID reuses the source-training reservoir; Gap is Pooled ID minus OOD reversed, computed within seed before summarizing.

**Encoding changes depend on both task and architecture.** Halving $R_Y$ scale gives perfect local and entangling Q-FOS reversal accuracy and improves reuploading, but lowers Q-SPHERE reuploading performance. Scale effects depend on task and architecture.

Replacing $R_Y$ by $R_X$ improves Spurious-Fourier reuploading while local performance stays similar; Q-LINK local remains near 0.098 and entangling above 0.985. Changing reuploading from $R_Y$–$R_Y$ to $R_Y$–$R_X$ improves Q-LINK but harms Q-SPHERE. These change existing-stage generators, neither isolating representation from optimization nor testing reuploading's presence versus absence.

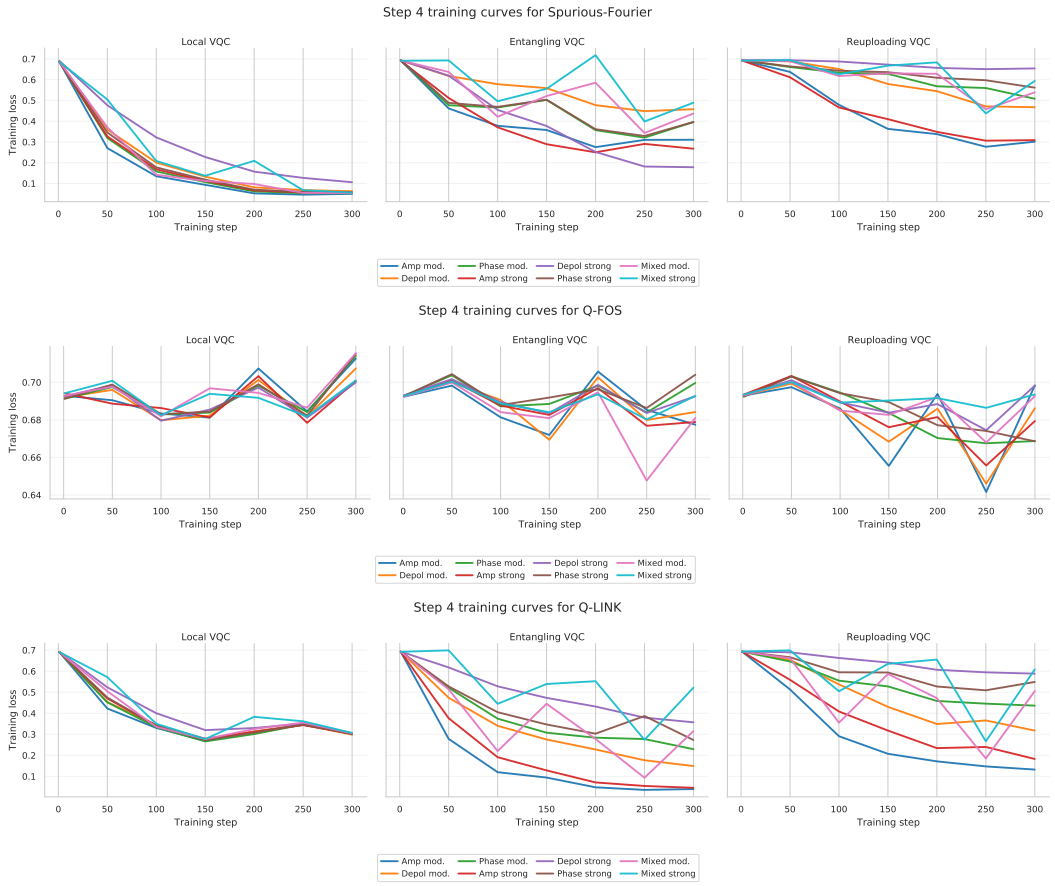


Figure 16: **Optimization trajectories under the nine training conditions.** Spurious-Fourier (top), Q-FOS (middle), and Q-LINK (bottom). Loss curves do not test feature-gradient alignment.

**Training-time feature diagnostics.** Training-time diagnostics for Q-SPHERE, Spirals-Signatures, and Q-LINK retain the trained encoding and preprocessing. Before preprocessing, stable-only zeros the shortcut block, shortcut-only zeros the stable block, and shortcut-flip negates the shortcut block. For positive-class probability $p(x)$,

$$S_{\text{flip}} = \frac{1}{n}\sum_{i=1}^{n}\left|p(x_i) - p(x_i^{\text{flip}})\right|.$$

This is probability change, not accuracy change. Results separate task, architecture, encoding, and split (mean $\pm$ sample SD over seeds $0, 1, 2$). These diagnostics are unavailable for the spectral tasks.

**Separate checkpoint-replay diagnostics.** Replay diagnostics are excluded: missing encoding metadata can trigger $R_Y$, scale-1 defaults, producing original-mode accuracies inconsistent with training-time exports. Reported diagnostics use the original predictor without reconstruction.

**Encoding-dependent reliance.** For Q-LINK reuploading, halving $R_Y$ scale improves stable-only reversal accuracy and lowers flip sensitivity alongside full-input gains (Table 53). Doubling Q-SPHERE entangling scale similarly improves both diagnostics. For Spirals-Signatures reuploading, doubling scale lowers both flip sensitivity and stable-only accuracy: reduced shortcut response need not recover stable prediction.

**Complete training-time feature diagnostics.** Tables 54–58, 59–63, and 64–68 report the training-time diagnostics defined above for all splits of Q-SPHERE, Spirals-Signatures,

Table 49: Q-SPHERE: encoding sensitivity by architecture. Definitions and paired-gap convention: Appendix C.7.

| Model | Encoding | Pooled ID | Random | OOD reversed | Gap |
|---|---|---|---|---|---|
| Local | $R_Y$ | $0.894_{\pm 0.009}$ | $0.504_{\pm 0.005}$ | $0.101_{\pm 0.010}$ | $0.793_{\pm 0.018}$ |
| | $R_X$ | $0.895_{\pm 0.010}$ | $0.504_{\pm 0.005}$ | $0.101_{\pm 0.010}$ | $0.794_{\pm 0.017}$ |
| | $R_Y$, $s = 0.5$ | $0.897_{\pm 0.010}$ | $0.504_{\pm 0.005}$ | $0.101_{\pm 0.009}$ | $0.797_{\pm 0.016}$ |
| | $R_Y$, $s = 2$ | $0.763_{\pm 0.010}$ | $0.512_{\pm 0.004}$ | $0.125_{\pm 0.019}$ | $0.638_{\pm 0.028}$ |
| Entangling | $R_Y$ | $0.877_{\pm 0.010}$ | $0.501_{\pm 0.006}$ | $0.125_{\pm 0.009}$ | $0.752_{\pm 0.018}$ |
| | $R_X$ | $0.870_{\pm 0.011}$ | $0.495_{\pm 0.001}$ | $0.123_{\pm 0.011}$ | $0.747_{\pm 0.021}$ |
| | $R_Y$, $s = 0.5$ | $0.897_{\pm 0.011}$ | $0.504_{\pm 0.005}$ | $0.101_{\pm 0.009}$ | $0.797_{\pm 0.015}$ |
| | $R_Y$, $s = 2$ | $0.793_{\pm 0.049}$ | $0.754_{\pm 0.066}$ | $0.723_{\pm 0.073}$ | $0.070_{\pm 0.059}$ |
| Reuploading | $R_Y$ | $0.856_{\pm 0.036}$ | $0.572_{\pm 0.024}$ | $0.297_{\pm 0.026}$ | $0.558_{\pm 0.049}$ |
| | $R_X$ | $0.872_{\pm 0.035}$ | $0.504_{\pm 0.012}$ | $0.151_{\pm 0.065}$ | $0.721_{\pm 0.100}$ |
| | $R_Y$, $s = 0.5$ | $0.894_{\pm 0.011}$ | $0.504_{\pm 0.006}$ | $0.109_{\pm 0.014}$ | $0.785_{\pm 0.022}$ |
| | $R_Y$, $s = 2$ | $0.660_{\pm 0.015}$ | $0.629_{\pm 0.030}$ | $0.620_{\pm 0.031}$ | $0.039_{\pm 0.040}$ |
| | $R_Y$–$R_X$ | $0.852_{\pm 0.034}$ | $0.512_{\pm 0.015}$ | $0.148_{\pm 0.065}$ | $0.704_{\pm 0.097}$ |
| | $R_X$–$R_Y$ | $0.842_{\pm 0.041}$ | $0.619_{\pm 0.037}$ | $0.393_{\pm 0.077}$ | $0.449_{\pm 0.115}$ |

Table 50: Spirals-Signatures: encoding sensitivity by architecture. Definitions and paired-gap convention: Appendix C.7.

| Model | Encoding | Pooled ID | Random | OOD reversed | Gap |
|---|---|---|---|---|---|
| Local | $R_Y$ | $0.903_{\pm 0.010}$ | $0.496_{\pm 0.002}$ | $0.102_{\pm 0.004}$ | $0.801_{\pm 0.008}$ |
| | $R_X$ | $0.903_{\pm 0.010}$ | $0.496_{\pm 0.002}$ | $0.102_{\pm 0.004}$ | $0.801_{\pm 0.008}$ |
| | $R_Y$, $s = 0.5$ | $0.904_{\pm 0.009}$ | $0.496_{\pm 0.002}$ | $0.102_{\pm 0.004}$ | $0.802_{\pm 0.008}$ |
| | $R_Y$, $s = 2$ | $0.652_{\pm 0.015}$ | $0.496_{\pm 0.002}$ | $0.104_{\pm 0.005}$ | $0.548_{\pm 0.016}$ |
| Entangling | $R_Y$ | $0.905_{\pm 0.009}$ | $0.502_{\pm 0.007}$ | $0.149_{\pm 0.041}$ | $0.756_{\pm 0.048}$ |
| | $R_X$ | $0.902_{\pm 0.007}$ | $0.504_{\pm 0.014}$ | $0.131_{\pm 0.049}$ | $0.771_{\pm 0.044}$ |
| | $R_Y$, $s = 0.5$ | $0.905_{\pm 0.010}$ | $0.496_{\pm 0.002}$ | $0.102_{\pm 0.004}$ | $0.803_{\pm 0.009}$ |
| | $R_Y$, $s = 2$ | $0.695_{\pm 0.014}$ | $0.507_{\pm 0.005}$ | $0.507_{\pm 0.071}$ | $0.188_{\pm 0.064}$ |
| Reuploading | $R_Y$ | $0.900_{\pm 0.012}$ | $0.522_{\pm 0.014}$ | $0.187_{\pm 0.021}$ | $0.713_{\pm 0.010}$ |
| | $R_X$ | $0.904_{\pm 0.007}$ | $0.493_{\pm 0.012}$ | $0.126_{\pm 0.007}$ | $0.777_{\pm 0.013}$ |
| | $R_Y$, $s = 0.5$ | $0.904_{\pm 0.010}$ | $0.496_{\pm 0.002}$ | $0.102_{\pm 0.004}$ | $0.801_{\pm 0.009}$ |
| | $R_Y$, $s = 2$ | $0.698_{\pm 0.026}$ | $0.562_{\pm 0.022}$ | $0.548_{\pm 0.036}$ | $0.150_{\pm 0.059}$ |
| | $R_Y$–$R_X$ | $0.902_{\pm 0.009}$ | $0.501_{\pm 0.009}$ | $0.132_{\pm 0.020}$ | $0.770_{\pm 0.026}$ |
| | $R_X$–$R_Y$ | $0.897_{\pm 0.012}$ | $0.518_{\pm 0.027}$ | $0.163_{\pm 0.020}$ | $0.735_{\pm 0.009}$ |

Table 51: Spurious-Fourier: encoding sensitivity by architecture. Definitions and paired-gap convention: Appendix C.7.

| Model | Encoding | Pooled ID | Random | OOD reversed | Gap |
|---|---|---|---|---|---|
| Local | $R_Y$ | $0.982_{\pm 0.006}$ | $0.961_{\pm 0.007}$ | $0.956_{\pm 0.003}$ | $0.026_{\pm 0.007}$ |
| | $R_X$ | $0.983_{\pm 0.005}$ | $0.961_{\pm 0.006}$ | $0.957_{\pm 0.003}$ | $0.026_{\pm 0.006}$ |
| | $R_Y$, $s = 0.5$ | $0.984_{\pm 0.005}$ | $0.965_{\pm 0.004}$ | $0.957_{\pm 0.006}$ | $0.027_{\pm 0.010}$ |
| | $R_Y$, $s = 2$ | $0.591_{\pm 0.006}$ | $0.291_{\pm 0.006}$ | $0.274_{\pm 0.010}$ | $0.317_{\pm 0.009}$ |
| Entangling | $R_Y$ | $0.827_{\pm 0.010}$ | $0.886_{\pm 0.024}$ | $0.868_{\pm 0.023}$ | $-0.041_{\pm 0.016}$ |
| | $R_X$ | $0.852_{\pm 0.034}$ | $0.882_{\pm 0.034}$ | $0.870_{\pm 0.039}$ | $-0.017_{\pm 0.024}$ |
| | $R_Y$, $s = 0.5$ | $0.983_{\pm 0.006}$ | $0.966_{\pm 0.003}$ | $0.959_{\pm 0.005}$ | $0.025_{\pm 0.010}$ |
| | $R_Y$, $s = 2$ | $0.755_{\pm 0.020}$ | $0.438_{\pm 0.016}$ | $0.438_{\pm 0.029}$ | $0.316_{\pm 0.046}$ |
| Reuploading | $R_Y$ | $0.858_{\pm 0.018}$ | $0.574_{\pm 0.112}$ | $0.569_{\pm 0.083}$ | $0.289_{\pm 0.083}$ |
| | $R_X$ | $0.980_{\pm 0.006}$ | $0.960_{\pm 0.006}$ | $0.954_{\pm 0.010}$ | $0.026_{\pm 0.009}$ |
| | $R_Y$, $s = 0.5$ | $0.982_{\pm 0.007}$ | $0.967_{\pm 0.005}$ | $0.954_{\pm 0.002}$ | $0.028_{\pm 0.006}$ |
| | $R_Y$, $s = 2$ | $0.736_{\pm 0.015}$ | $0.496_{\pm 0.031}$ | $0.488_{\pm 0.030}$ | $0.248_{\pm 0.017}$ |
| | $R_Y$–$R_X$ | $0.984_{\pm 0.005}$ | $0.966_{\pm 0.008}$ | $0.958_{\pm 0.004}$ | $0.027_{\pm 0.008}$ |
| | $R_X$–$R_Y$ | $0.878_{\pm 0.023}$ | $0.737_{\pm 0.106}$ | $0.737_{\pm 0.125}$ | $0.141_{\pm 0.104}$ |

Table 52: Q-FOS: encoding sensitivity by architecture. Definitions and paired-gap convention: Appendix C.7.

| Model | Encoding | Pooled ID | Random | OOD reversed | Gap |
|---|---|---|---|---|---|
| Local | $R_Y$ | $0.579_{\pm 0.014}$ | $0.368_{\pm 0.003}$ | $0.595_{\pm 0.036}$ | $-0.016_{\pm 0.027}$ |
| | $R_X$ | $0.574_{\pm 0.009}$ | $0.363_{\pm 0.028}$ | $0.598_{\pm 0.027}$ | $-0.023_{\pm 0.021}$ |
| | $R_Y$, $s = 0.5$ | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.000_{\pm 0.000}$ |
| | $R_Y$, $s = 2$ | $0.583_{\pm 0.018}$ | $0.608_{\pm 0.007}$ | $0.388_{\pm 0.013}$ | $0.195_{\pm 0.019}$ |
| Entangling | $R_Y$ | $0.581_{\pm 0.027}$ | $0.455_{\pm 0.084}$ | $0.570_{\pm 0.021}$ | $0.011_{\pm 0.006}$ |
| | $R_X$ | $0.602_{\pm 0.018}$ | $0.382_{\pm 0.011}$ | $0.570_{\pm 0.026}$ | $0.032_{\pm 0.044}$ |
| | $R_Y$, $s = 0.5$ | $0.957_{\pm 0.006}$ | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $-0.043_{\pm 0.006}$ |
| | $R_Y$, $s = 2$ | $0.601_{\pm 0.017}$ | $0.610_{\pm 0.010}$ | $0.434_{\pm 0.019}$ | $0.168_{\pm 0.019}$ |
| Reuploading | $R_Y$ | $0.653_{\pm 0.017}$ | $0.583_{\pm 0.036}$ | $0.452_{\pm 0.031}$ | $0.201_{\pm 0.022}$ |
| | $R_X$ | $0.615_{\pm 0.046}$ | $0.487_{\pm 0.048}$ | $0.443_{\pm 0.018}$ | $0.173_{\pm 0.057}$ |
| | $R_Y$, $s = 0.5$ | $1.000_{\pm 0.001}$ | $0.642_{\pm 0.074}$ | $0.624_{\pm 0.077}$ | $0.376_{\pm 0.077}$ |
| | $R_Y$, $s = 2$ | $0.624_{\pm 0.036}$ | $0.501_{\pm 0.032}$ | $0.448_{\pm 0.023}$ | $0.176_{\pm 0.058}$ |
| | $R_Y$–$R_X$ | $0.617_{\pm 0.048}$ | $0.521_{\pm 0.062}$ | $0.409_{\pm 0.018}$ | $0.208_{\pm 0.058}$ |
| | $R_X$–$R_Y$ | $0.636_{\pm 0.012}$ | $0.576_{\pm 0.038}$ | $0.482_{\pm 0.060}$ | $0.154_{\pm 0.051}$ |

Table 53: Q-LINK: encoding sensitivity by architecture. Definitions and paired-gap convention: Appendix C.7.

| Model | Encoding | Pooled ID | Random | OOD reversed | Gap |
|---|---|---|---|---|---|
| Local | $R_Y$ | $0.902_{\pm 0.008}$ | $0.506_{\pm 0.016}$ | $0.098_{\pm 0.005}$ | $0.804_{\pm 0.012}$ |
| | $R_X$ | $0.902_{\pm 0.008}$ | $0.506_{\pm 0.016}$ | $0.098_{\pm 0.005}$ | $0.804_{\pm 0.012}$ |
| | $R_Y$, $s = 0.5$ | $0.902_{\pm 0.008}$ | $0.506_{\pm 0.016}$ | $0.098_{\pm 0.005}$ | $0.804_{\pm 0.011}$ |
| | $R_Y$, $s = 2$ | $0.640_{\pm 0.024}$ | $0.507_{\pm 0.016}$ | $0.099_{\pm 0.006}$ | $0.542_{\pm 0.029}$ |
| Entangling | $R_Y$ | $0.995_{\pm 0.004}$ | $0.985_{\pm 0.015}$ | $0.986_{\pm 0.019}$ | $0.010_{\pm 0.015}$ |
| | $R_X$ | $0.991_{\pm 0.012}$ | $0.995_{\pm 0.006}$ | $0.998_{\pm 0.001}$ | $-0.007_{\pm 0.012}$ |
| | $R_Y$, $s = 0.5$ | $0.999_{\pm 0.001}$ | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.001}$ | $-0.000_{\pm 0.001}$ |
| | $R_Y$, $s = 2$ | $0.909_{\pm 0.026}$ | $0.744_{\pm 0.284}$ | $0.736_{\pm 0.291}$ | $0.173_{\pm 0.265}$ |
| Reuploading | $R_Y$ | $0.982_{\pm 0.016}$ | $0.901_{\pm 0.084}$ | $0.846_{\pm 0.122}$ | $0.136_{\pm 0.111}$ |
| | $R_X$ | $0.982_{\pm 0.030}$ | $0.914_{\pm 0.146}$ | $0.865_{\pm 0.234}$ | $0.117_{\pm 0.204}$ |
| | $R_Y$, $s = 0.5$ | $0.999_{\pm 0.001}$ | $0.998_{\pm 0.003}$ | $0.997_{\pm 0.002}$ | $0.002_{\pm 0.002}$ |
| | $R_Y$, $s = 2$ | $0.785_{\pm 0.070}$ | $0.575_{\pm 0.057}$ | $0.513_{\pm 0.037}$ | $0.272_{\pm 0.092}$ |
| | $R_Y$–$R_X$ | $0.966_{\pm 0.057}$ | $0.966_{\pm 0.056}$ | $0.968_{\pm 0.053}$ | $-0.002_{\pm 0.005}$ |
| | $R_X$–$R_Y$ | $0.950_{\pm 0.065}$ | $0.863_{\pm 0.169}$ | $0.748_{\pm 0.321}$ | $0.203_{\pm 0.256}$ |

and Q-LINK. Repeated baselines appear once. Stable-only/shortcut-only report accuracy; flip sensitivity follows the definition above. All metrics use $[0, 1]$; the encoding and Pooled ID conventions above apply.

Table 54: **Q-SPHERE: Env A train.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.529_{\pm 0.025}$ | $0.883_{\pm 0.006}$ | $0.721_{\pm 0.029}$ |
| | $R_X$, $s = 1$ | $0.515_{\pm 0.027}$ | $0.884_{\pm 0.007}$ | $0.723_{\pm 0.012}$ |
| | $R_Y$, $s = 0.5$ | $0.505_{\pm 0.037}$ | $0.884_{\pm 0.005}$ | $0.719_{\pm 0.037}$ |
| | $R_Y$, $s = 2$ | $0.627_{\pm 0.019}$ | $0.694_{\pm 0.010}$ | $0.452_{\pm 0.026}$ |
| Entangling | $R_Y$, $s = 1$ | $0.496_{\pm 0.007}$ | $0.866_{\pm 0.016}$ | $0.609_{\pm 0.022}$ |
| | $R_X$, $s = 1$ | $0.530_{\pm 0.021}$ | $0.868_{\pm 0.022}$ | $0.614_{\pm 0.042}$ |
| | $R_Y$, $s = 0.5$ | $0.495_{\pm 0.006}$ | $0.882_{\pm 0.009}$ | $0.708_{\pm 0.017}$ |
| | $R_Y$, $s = 2$ | $0.694_{\pm 0.163}$ | $0.439_{\pm 0.094}$ | $0.062_{\pm 0.042}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.641_{\pm 0.039}$ | $0.579_{\pm 0.182}$ | $0.532_{\pm 0.072}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.597_{\pm 0.015}$ | $0.869_{\pm 0.024}$ | $0.650_{\pm 0.065}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $0.619_{\pm 0.045}$ | $0.883_{\pm 0.008}$ | $0.701_{\pm 0.037}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.577_{\pm 0.129}$ | $0.484_{\pm 0.021}$ | $0.096_{\pm 0.017}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.599_{\pm 0.082}$ | $0.824_{\pm 0.049}$ | $0.583_{\pm 0.054}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.712_{\pm 0.025}$ | $0.830_{\pm 0.013}$ | $0.428_{\pm 0.055}$ |

### C.8 Role reversal: Q-LINK-Reversed

Q-LINK-Reversed retains six coordinates with reversed roles: an invariant local cue and environment-dependent relational shortcut. It is a Q-LINK ablation, not a sixth primary task.

Sample a balanced label $Y \in \{-1, +1\}$ and define

$$Z = Y d + \eta, \qquad d = \frac{(1, 0.35)}{\sqrt{1 + 0.35^2}}, \qquad \eta \sim \mathcal{N}(0, 0.22^2 I_2). \tag{17}$$

Table 55: **Q-SPHERE: Env B train.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.515_{\pm 0.009}$ | $0.909_{\pm 0.006}$ | $0.758_{\pm 0.022}$ |
| | $R_X$, $s = 1$ | $0.502_{\pm 0.034}$ | $0.909_{\pm 0.006}$ | $0.754_{\pm 0.023}$ |
| | $R_Y$, $s = 0.5$ | $0.500_{\pm 0.024}$ | $0.908_{\pm 0.006}$ | $0.768_{\pm 0.021}$ |
| | $R_Y$, $s = 2$ | $0.607_{\pm 0.005}$ | $0.713_{\pm 0.004}$ | $0.437_{\pm 0.022}$ |
| Entangling | $R_Y$, $s = 1$ | $0.499_{\pm 0.026}$ | $0.898_{\pm 0.008}$ | $0.693_{\pm 0.028}$ |
| | $R_X$, $s = 1$ | $0.523_{\pm 0.034}$ | $0.896_{\pm 0.007}$ | $0.681_{\pm 0.016}$ |
| | $R_Y$, $s = 0.5$ | $0.509_{\pm 0.010}$ | $0.908_{\pm 0.005}$ | $0.763_{\pm 0.007}$ |
| | $R_Y$, $s = 2$ | $0.711_{\pm 0.150}$ | $0.497_{\pm 0.073}$ | $0.058_{\pm 0.031}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.625_{\pm 0.057}$ | $0.808_{\pm 0.058}$ | $0.470_{\pm 0.037}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.576_{\pm 0.021}$ | $0.882_{\pm 0.045}$ | $0.672_{\pm 0.075}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $0.578_{\pm 0.043}$ | $0.904_{\pm 0.011}$ | $0.716_{\pm 0.016}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.562_{\pm 0.109}$ | $0.496_{\pm 0.025}$ | $0.113_{\pm 0.030}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.608_{\pm 0.049}$ | $0.869_{\pm 0.069}$ | $0.616_{\pm 0.103}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.701_{\pm 0.036}$ | $0.818_{\pm 0.056}$ | $0.468_{\pm 0.087}$ |

Table 56: **Q-SPHERE: Pooled ID.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.517_{\pm 0.039}$ | $0.896_{\pm 0.010}$ | $0.737_{\pm 0.020}$ |
| | $R_X$, $s = 1$ | $0.502_{\pm 0.043}$ | $0.897_{\pm 0.010}$ | $0.736_{\pm 0.018}$ |
| | $R_Y$, $s = 0.5$ | $0.513_{\pm 0.022}$ | $0.898_{\pm 0.010}$ | $0.743_{\pm 0.031}$ |
| | $R_Y$, $s = 2$ | $0.618_{\pm 0.007}$ | $0.705_{\pm 0.008}$ | $0.436_{\pm 0.024}$ |
| Entangling | $R_Y$, $s = 1$ | $0.488_{\pm 0.021}$ | $0.883_{\pm 0.015}$ | $0.651_{\pm 0.013}$ |
| | $R_X$, $s = 1$ | $0.515_{\pm 0.005}$ | $0.884_{\pm 0.015}$ | $0.649_{\pm 0.012}$ |
| | $R_Y$, $s = 0.5$ | $0.501_{\pm 0.025}$ | $0.897_{\pm 0.012}$ | $0.736_{\pm 0.010}$ |
| | $R_Y$, $s = 2$ | $0.702_{\pm 0.167}$ | $0.463_{\pm 0.097}$ | $0.059_{\pm 0.036}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.632_{\pm 0.057}$ | $0.690_{\pm 0.095}$ | $0.495_{\pm 0.061}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.582_{\pm 0.003}$ | $0.876_{\pm 0.037}$ | $0.659_{\pm 0.070}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $0.611_{\pm 0.031}$ | $0.895_{\pm 0.013}$ | $0.708_{\pm 0.026}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.587_{\pm 0.115}$ | $0.480_{\pm 0.021}$ | $0.103_{\pm 0.022}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.603_{\pm 0.049}$ | $0.850_{\pm 0.039}$ | $0.602_{\pm 0.037}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.714_{\pm 0.038}$ | $0.829_{\pm 0.031}$ | $0.445_{\pm 0.066}$ |

Table 57: **Q-SPHERE: Random.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.513_{\pm 0.040}$ | $0.504_{\pm 0.005}$ | $0.881_{\pm 0.013}$ |
| | $R_X$, $s = 1$ | $0.497_{\pm 0.049}$ | $0.504_{\pm 0.005}$ | $0.880_{\pm 0.012}$ |
| | $R_Y$, $s = 0.5$ | $0.510_{\pm 0.059}$ | $0.504_{\pm 0.005}$ | $0.849_{\pm 0.024}$ |
| | $R_Y$, $s = 2$ | $0.613_{\pm 0.013}$ | $0.500_{\pm 0.003}$ | $0.709_{\pm 0.036}$ |
| Entangling | $R_Y$, $s = 1$ | $0.507_{\pm 0.036}$ | $0.505_{\pm 0.007}$ | $0.671_{\pm 0.011}$ |
| | $R_X$, $s = 1$ | $0.522_{\pm 0.019}$ | $0.502_{\pm 0.007}$ | $0.658_{\pm 0.031}$ |
| | $R_Y$, $s = 0.5$ | $0.495_{\pm 0.043}$ | $0.504_{\pm 0.005}$ | $0.824_{\pm 0.005}$ |
| | $R_Y$, $s = 2$ | $0.700_{\pm 0.166}$ | $0.485_{\pm 0.014}$ | $0.078_{\pm 0.055}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.647_{\pm 0.042}$ | $0.510_{\pm 0.010}$ | $0.500_{\pm 0.038}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.583_{\pm 0.019}$ | $0.499_{\pm 0.002}$ | $0.647_{\pm 0.111}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $0.602_{\pm 0.035}$ | $0.504_{\pm 0.004}$ | $0.794_{\pm 0.022}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.583_{\pm 0.122}$ | $0.469_{\pm 0.013}$ | $0.108_{\pm 0.021}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.571_{\pm 0.086}$ | $0.505_{\pm 0.006}$ | $0.664_{\pm 0.124}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.704_{\pm 0.026}$ | $0.507_{\pm 0.005}$ | $0.437_{\pm 0.093}$ |

Table 58: **Q-SPHERE: OOD reversed.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.517_{\pm 0.035}$ | $0.101_{\pm 0.010}$ | $0.883_{\pm 0.013}$ |
| | $R_X$, $s = 1$ | $0.497_{\pm 0.034}$ | $0.100_{\pm 0.010}$ | $0.882_{\pm 0.011}$ |
| | $R_Y$, $s = 0.5$ | $0.506_{\pm 0.037}$ | $0.101_{\pm 0.009}$ | $0.852_{\pm 0.022}$ |
| | $R_Y$, $s = 2$ | $0.624_{\pm 0.022}$ | $0.120_{\pm 0.010}$ | $0.708_{\pm 0.036}$ |
| Entangling | $R_Y$, $s = 1$ | $0.490_{\pm 0.022}$ | $0.118_{\pm 0.013}$ | $0.685_{\pm 0.015}$ |
| | $R_X$, $s = 1$ | $0.522_{\pm 0.009}$ | $0.116_{\pm 0.012}$ | $0.668_{\pm 0.031}$ |
| | $R_Y$, $s = 0.5$ | $0.484_{\pm 0.014}$ | $0.101_{\pm 0.010}$ | $0.832_{\pm 0.004}$ |
| | $R_Y$, $s = 2$ | $0.696_{\pm 0.160}$ | $0.622_{\pm 0.204}$ | $0.078_{\pm 0.051}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.646_{\pm 0.016}$ | $0.265_{\pm 0.061}$ | $0.510_{\pm 0.044}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.582_{\pm 0.021}$ | $0.130_{\pm 0.057}$ | $0.646_{\pm 0.113}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $0.633_{\pm 0.059}$ | $0.103_{\pm 0.011}$ | $0.800_{\pm 0.024}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.580_{\pm 0.121}$ | $0.514_{\pm 0.060}$ | $0.109_{\pm 0.016}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.615_{\pm 0.087}$ | $0.124_{\pm 0.043}$ | $0.671_{\pm 0.128}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.715_{\pm 0.034}$ | $0.146_{\pm 0.039}$ | $0.438_{\pm 0.083}$ |

Table 59: **Spirals-Signatures: Env A train.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.532_{\pm 0.045}$ | $0.902_{\pm 0.003}$ | $0.771_{\pm 0.014}$ |
| | $R_X$, $s = 1$ | $0.547_{\pm 0.028}$ | $0.902_{\pm 0.003}$ | $0.772_{\pm 0.022}$ |
| | $R_Y$, $s = 0.5$ | $0.557_{\pm 0.022}$ | $0.901_{\pm 0.004}$ | $0.767_{\pm 0.007}$ |
| | $R_Y$, $s = 2$ | $0.525_{\pm 0.051}$ | $0.636_{\pm 0.028}$ | $0.304_{\pm 0.013}$ |
| Entangling | $R_Y$, $s = 1$ | $0.483_{\pm 0.043}$ | $0.902_{\pm 0.003}$ | $0.741_{\pm 0.028}$ |
| | $R_X$, $s = 1$ | $0.524_{\pm 0.019}$ | $0.901_{\pm 0.003}$ | $0.759_{\pm 0.027}$ |
| | $R_Y$, $s = 0.5$ | $0.549_{\pm 0.018}$ | $0.902_{\pm 0.003}$ | $0.761_{\pm 0.018}$ |
| | $R_Y$, $s = 2$ | $0.525_{\pm 0.005}$ | $0.694_{\pm 0.010}$ | $0.462_{\pm 0.013}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.531_{\pm 0.015}$ | $0.901_{\pm 0.004}$ | $0.699_{\pm 0.032}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.487_{\pm 0.020}$ | $0.901_{\pm 0.004}$ | $0.761_{\pm 0.020}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $0.559_{\pm 0.017}$ | $0.902_{\pm 0.003}$ | $0.783_{\pm 0.013}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.460_{\pm 0.021}$ | $0.614_{\pm 0.012}$ | $0.138_{\pm 0.023}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.535_{\pm 0.025}$ | $0.902_{\pm 0.003}$ | $0.780_{\pm 0.035}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.509_{\pm 0.032}$ | $0.901_{\pm 0.004}$ | $0.711_{\pm 0.004}$ |

Table 60: **Spirals-Signatures: Env B train.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.572_{\pm 0.041}$ | $0.898_{\pm 0.002}$ | $0.771_{\pm 0.004}$ |
| | $R_X$, $s = 1$ | $0.570_{\pm 0.012}$ | $0.898_{\pm 0.002}$ | $0.765_{\pm 0.005}$ |
| | $R_Y$, $s = 0.5$ | $0.561_{\pm 0.039}$ | $0.898_{\pm 0.001}$ | $0.749_{\pm 0.010}$ |
| | $R_Y$, $s = 2$ | $0.520_{\pm 0.032}$ | $0.661_{\pm 0.013}$ | $0.311_{\pm 0.003}$ |
| Entangling | $R_Y$, $s = 1$ | $0.486_{\pm 0.036}$ | $0.898_{\pm 0.001}$ | $0.786_{\pm 0.007}$ |
| | $R_X$, $s = 1$ | $0.533_{\pm 0.022}$ | $0.898_{\pm 0.002}$ | $0.790_{\pm 0.010}$ |
| | $R_Y$, $s = 0.5$ | $0.574_{\pm 0.037}$ | $0.899_{\pm 0.001}$ | $0.755_{\pm 0.012}$ |
| | $R_Y$, $s = 2$ | $0.514_{\pm 0.028}$ | $0.694_{\pm 0.018}$ | $0.303_{\pm 0.030}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.545_{\pm 0.027}$ | $0.897_{\pm 0.003}$ | $0.787_{\pm 0.037}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.505_{\pm 0.030}$ | $0.897_{\pm 0.003}$ | $0.793_{\pm 0.015}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $0.589_{\pm 0.015}$ | $0.898_{\pm 0.002}$ | $0.773_{\pm 0.012}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.469_{\pm 0.022}$ | $0.513_{\pm 0.032}$ | $0.121_{\pm 0.030}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.545_{\pm 0.019}$ | $0.898_{\pm 0.002}$ | $0.793_{\pm 0.018}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.538_{\pm 0.048}$ | $0.896_{\pm 0.001}$ | $0.778_{\pm 0.013}$ |

Table 61: **Spirals-Signatures: Pooled ID.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.538_{\pm 0.056}$ | $0.904_{\pm 0.009}$ | $0.773_{\pm 0.006}$ |
| | $R_X$, $s = 1$ | $0.551_{\pm 0.026}$ | $0.904_{\pm 0.009}$ | $0.770_{\pm 0.006}$ |
| | $R_Y$, $s = 0.5$ | $0.578_{\pm 0.025}$ | $0.904_{\pm 0.010}$ | $0.759_{\pm 0.003}$ |
| | $R_Y$, $s = 2$ | $0.508_{\pm 0.057}$ | $0.656_{\pm 0.025}$ | $0.308_{\pm 0.010}$ |
| Entangling | $R_Y$, $s = 1$ | $0.490_{\pm 0.038}$ | $0.904_{\pm 0.009}$ | $0.762_{\pm 0.017}$ |
| | $R_X$, $s = 1$ | $0.547_{\pm 0.017}$ | $0.903_{\pm 0.009}$ | $0.775_{\pm 0.009}$ |
| | $R_Y$, $s = 0.5$ | $0.568_{\pm 0.011}$ | $0.904_{\pm 0.009}$ | $0.759_{\pm 0.011}$ |
| | $R_Y$, $s = 2$ | $0.515_{\pm 0.019}$ | $0.694_{\pm 0.012}$ | $0.380_{\pm 0.011}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.534_{\pm 0.025}$ | $0.903_{\pm 0.008}$ | $0.741_{\pm 0.009}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.477_{\pm 0.011}$ | $0.903_{\pm 0.008}$ | $0.775_{\pm 0.013}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $0.577_{\pm 0.005}$ | $0.904_{\pm 0.009}$ | $0.779_{\pm 0.003}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.459_{\pm 0.016}$ | $0.554_{\pm 0.014}$ | $0.128_{\pm 0.029}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.524_{\pm 0.027}$ | $0.904_{\pm 0.009}$ | $0.785_{\pm 0.014}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.508_{\pm 0.033}$ | $0.901_{\pm 0.011}$ | $0.746_{\pm 0.007}$ |

Table 62: **Spirals-Signatures: Random.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.545_{\pm 0.029}$ | $0.496_{\pm 0.002}$ | $0.946_{\pm 0.003}$ |
| | $R_X$, $s = 1$ | $0.547_{\pm 0.023}$ | $0.496_{\pm 0.002}$ | $0.945_{\pm 0.003}$ |
| | $R_Y$, $s = 0.5$ | $0.545_{\pm 0.029}$ | $0.496_{\pm 0.002}$ | $0.904_{\pm 0.001}$ |
| | $R_Y$, $s = 2$ | $0.521_{\pm 0.032}$ | $0.496_{\pm 0.002}$ | $0.717_{\pm 0.005}$ |
| Entangling | $R_Y$, $s = 1$ | $0.480_{\pm 0.046}$ | $0.496_{\pm 0.002}$ | $0.741_{\pm 0.088}$ |
| | $R_X$, $s = 1$ | $0.507_{\pm 0.012}$ | $0.498_{\pm 0.005}$ | $0.752_{\pm 0.099}$ |
| | $R_Y$, $s = 0.5$ | $0.537_{\pm 0.044}$ | $0.496_{\pm 0.002}$ | $0.883_{\pm 0.005}$ |
| | $R_Y$, $s = 2$ | $0.489_{\pm 0.039}$ | $0.508_{\pm 0.008}$ | $0.295_{\pm 0.022}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.534_{\pm 0.013}$ | $0.497_{\pm 0.004}$ | $0.705_{\pm 0.022}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.485_{\pm 0.021}$ | $0.498_{\pm 0.004}$ | $0.715_{\pm 0.006}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $0.556_{\pm 0.031}$ | $0.496_{\pm 0.002}$ | $0.883_{\pm 0.009}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.487_{\pm 0.014}$ | $0.500_{\pm 0.013}$ | $0.176_{\pm 0.042}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.563_{\pm 0.036}$ | $0.499_{\pm 0.003}$ | $0.727_{\pm 0.017}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.512_{\pm 0.031}$ | $0.496_{\pm 0.002}$ | $0.710_{\pm 0.024}$ |

Table 63: **Spirals-Signatures: OOD reversed.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.552_{\pm 0.028}$ | $0.102_{\pm 0.004}$ | $0.944_{\pm 0.003}$ |
| | $R_X$, $s = 1$ | $0.555_{\pm 0.019}$ | $0.102_{\pm 0.004}$ | $0.943_{\pm 0.003}$ |
| | $R_Y$, $s = 0.5$ | $0.580_{\pm 0.036}$ | $0.102_{\pm 0.004}$ | $0.900_{\pm 0.002}$ |
| | $R_Y$, $s = 2$ | $0.526_{\pm 0.042}$ | $0.103_{\pm 0.005}$ | $0.721_{\pm 0.006}$ |
| Entangling | $R_Y$, $s = 1$ | $0.502_{\pm 0.030}$ | $0.103_{\pm 0.005}$ | $0.737_{\pm 0.093}$ |
| | $R_X$, $s = 1$ | $0.537_{\pm 0.037}$ | $0.108_{\pm 0.010}$ | $0.751_{\pm 0.096}$ |
| | $R_Y$, $s = 0.5$ | $0.574_{\pm 0.044}$ | $0.102_{\pm 0.004}$ | $0.877_{\pm 0.006}$ |
| | $R_Y$, $s = 2$ | $0.513_{\pm 0.032}$ | $0.480_{\pm 0.125}$ | $0.301_{\pm 0.020}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.546_{\pm 0.020}$ | $0.110_{\pm 0.011}$ | $0.698_{\pm 0.022}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.485_{\pm 0.012}$ | $0.104_{\pm 0.004}$ | $0.719_{\pm 0.016}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $0.584_{\pm 0.006}$ | $0.102_{\pm 0.004}$ | $0.879_{\pm 0.008}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.478_{\pm 0.015}$ | $0.475_{\pm 0.303}$ | $0.173_{\pm 0.041}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.548_{\pm 0.029}$ | $0.106_{\pm 0.006}$ | $0.728_{\pm 0.024}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.536_{\pm 0.016}$ | $0.105_{\pm 0.005}$ | $0.712_{\pm 0.025}$ |

Table 64: **Q-LINK: Env A train.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.520_{\pm 0.031}$ | $0.910_{\pm 0.004}$ | $0.747_{\pm 0.025}$ |
| | $R_X$, $s = 1$ | $0.505_{\pm 0.008}$ | $0.911_{\pm 0.003}$ | $0.750_{\pm 0.021}$ |
| | $R_Y$, $s = 0.5$ | $0.502_{\pm 0.013}$ | $0.910_{\pm 0.003}$ | $0.759_{\pm 0.036}$ |
| | $R_Y$, $s = 2$ | $0.484_{\pm 0.019}$ | $0.675_{\pm 0.022}$ | $0.309_{\pm 0.006}$ |
| Entangling | $R_Y$, $s = 1$ | $0.999_{\pm 0.001}$ | $0.754_{\pm 0.151}$ | $0.013_{\pm 0.012}$ |
| | $R_X$, $s = 1$ | $0.999_{\pm 0.000}$ | $0.527_{\pm 0.031}$ | $0.001_{\pm 0.001}$ |
| | $R_Y$, $s = 0.5$ | $1.000_{\pm 0.000}$ | $0.753_{\pm 0.087}$ | $0.032_{\pm 0.004}$ |
| | $R_Y$, $s = 2$ | $0.522_{\pm 0.422}$ | $0.492_{\pm 0.004}$ | $0.015_{\pm 0.006}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.948_{\pm 0.084}$ | $0.311_{\pm 0.172}$ | $0.124_{\pm 0.040}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.963_{\pm 0.064}$ | $0.636_{\pm 0.241}$ | $0.130_{\pm 0.225}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $0.999_{\pm 0.000}$ | $0.499_{\pm 0.009}$ | $0.026_{\pm 0.017}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.357_{\pm 0.219}$ | $0.590_{\pm 0.143}$ | $0.126_{\pm 0.042}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.960_{\pm 0.067}$ | $0.501_{\pm 0.009}$ | $0.007_{\pm 0.009}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.918_{\pm 0.134}$ | $0.639_{\pm 0.235}$ | $0.244_{\pm 0.299}$ |

Table 65: **Q-LINK: Env B train.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.509_{\pm 0.029}$ | $0.884_{\pm 0.004}$ | $0.741_{\pm 0.031}$ |
| | $R_X$, $s = 1$ | $0.496_{\pm 0.014}$ | $0.884_{\pm 0.004}$ | $0.742_{\pm 0.021}$ |
| | $R_Y$, $s = 0.5$ | $0.518_{\pm 0.030}$ | $0.882_{\pm 0.005}$ | $0.749_{\pm 0.016}$ |
| | $R_Y$, $s = 2$ | $0.504_{\pm 0.021}$ | $0.610_{\pm 0.011}$ | $0.296_{\pm 0.006}$ |
| Entangling | $R_Y$, $s = 1$ | $0.999_{\pm 0.001}$ | $0.606_{\pm 0.189}$ | $0.068_{\pm 0.089}$ |
| | $R_X$, $s = 1$ | $0.999_{\pm 0.001}$ | $0.540_{\pm 0.063}$ | $0.012_{\pm 0.020}$ |
| | $R_Y$, $s = 0.5$ | $1.000_{\pm 0.001}$ | $0.733_{\pm 0.042}$ | $0.030_{\pm 0.003}$ |
| | $R_Y$, $s = 2$ | $0.528_{\pm 0.436}$ | $0.496_{\pm 0.013}$ | $0.014_{\pm 0.006}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.952_{\pm 0.077}$ | $0.554_{\pm 0.082}$ | $0.120_{\pm 0.060}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.956_{\pm 0.076}$ | $0.633_{\pm 0.222}$ | $0.119_{\pm 0.205}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $1.000_{\pm 0.001}$ | $0.515_{\pm 0.041}$ | $0.023_{\pm 0.016}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.338_{\pm 0.216}$ | $0.521_{\pm 0.179}$ | $0.152_{\pm 0.021}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.964_{\pm 0.060}$ | $0.510_{\pm 0.009}$ | $0.002_{\pm 0.001}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.915_{\pm 0.142}$ | $0.581_{\pm 0.152}$ | $0.163_{\pm 0.129}$ |

Table 66: **Q-LINK: Pooled ID.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.514_{\pm 0.031}$ | $0.902_{\pm 0.008}$ | $0.743_{\pm 0.024}$ |
| | $R_X$, $s = 1$ | $0.495_{\pm 0.002}$ | $0.902_{\pm 0.008}$ | $0.746_{\pm 0.017}$ |
| | $R_Y$, $s = 0.5$ | $0.502_{\pm 0.023}$ | $0.902_{\pm 0.008}$ | $0.755_{\pm 0.021}$ |
| | $R_Y$, $s = 2$ | $0.497_{\pm 0.010}$ | $0.643_{\pm 0.023}$ | $0.299_{\pm 0.009}$ |
| Entangling | $R_Y$, $s = 1$ | $0.998_{\pm 0.002}$ | $0.702_{\pm 0.124}$ | $0.041_{\pm 0.054}$ |
| | $R_X$, $s = 1$ | $0.998_{\pm 0.001}$ | $0.528_{\pm 0.058}$ | $0.005_{\pm 0.008}$ |
| | $R_Y$, $s = 0.5$ | $0.999_{\pm 0.001}$ | $0.752_{\pm 0.061}$ | $0.030_{\pm 0.002}$ |
| | $R_Y$, $s = 2$ | $0.524_{\pm 0.426}$ | $0.506_{\pm 0.009}$ | $0.015_{\pm 0.007}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.952_{\pm 0.077}$ | $0.429_{\pm 0.109}$ | $0.125_{\pm 0.050}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.955_{\pm 0.077}$ | $0.640_{\pm 0.235}$ | $0.126_{\pm 0.217}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $0.999_{\pm 0.001}$ | $0.511_{\pm 0.017}$ | $0.024_{\pm 0.016}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.357_{\pm 0.234}$ | $0.563_{\pm 0.143}$ | $0.144_{\pm 0.014}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.964_{\pm 0.059}$ | $0.501_{\pm 0.006}$ | $0.005_{\pm 0.005}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.916_{\pm 0.137}$ | $0.613_{\pm 0.202}$ | $0.210_{\pm 0.214}$ |

Table 67: **Q-LINK: Random.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.501_{\pm 0.035}$ | $0.506_{\pm 0.016}$ | $0.930_{\pm 0.013}$ |
| | $R_X$, $s = 1$ | $0.484_{\pm 0.008}$ | $0.506_{\pm 0.016}$ | $0.931_{\pm 0.010}$ |
| | $R_Y$, $s = 0.5$ | $0.479_{\pm 0.010}$ | $0.506_{\pm 0.016}$ | $0.896_{\pm 0.017}$ |
| | $R_Y$, $s = 2$ | $0.481_{\pm 0.009}$ | $0.505_{\pm 0.015}$ | $0.655_{\pm 0.005}$ |
| Entangling | $R_Y$, $s = 1$ | $0.998_{\pm 0.001}$ | $0.501_{\pm 0.012}$ | $0.036_{\pm 0.046}$ |
| | $R_X$, $s = 1$ | $0.999_{\pm 0.001}$ | $0.492_{\pm 0.019}$ | $0.006_{\pm 0.007}$ |
| | $R_Y$, $s = 0.5$ | $0.999_{\pm 0.001}$ | $0.511_{\pm 0.016}$ | $0.038_{\pm 0.000}$ |
| | $R_Y$, $s = 2$ | $0.534_{\pm 0.419}$ | $0.497_{\pm 0.012}$ | $0.028_{\pm 0.017}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.947_{\pm 0.089}$ | $0.501_{\pm 0.015}$ | $0.151_{\pm 0.082}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.959_{\pm 0.071}$ | $0.518_{\pm 0.007}$ | $0.107_{\pm 0.179}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $1.000_{\pm 0.001}$ | $0.499_{\pm 0.011}$ | $0.030_{\pm 0.019}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.341_{\pm 0.201}$ | $0.499_{\pm 0.019}$ | $0.177_{\pm 0.036}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.967_{\pm 0.055}$ | $0.499_{\pm 0.021}$ | $0.009_{\pm 0.005}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.908_{\pm 0.149}$ | $0.496_{\pm 0.011}$ | $0.217_{\pm 0.198}$ |

Table 68: **Q-LINK: OOD reversed.** Encoding diagnostics; definitions in Appendix C.7.

| VQC | Encoding | Stable-only | Shortcut-only | Flip sensitivity |
|---|---|---|---|---|
| Local | $R_Y$, $s = 1$ | $0.516_{\pm 0.010}$ | $0.098_{\pm 0.005}$ | $0.932_{\pm 0.014}$ |
| | $R_X$, $s = 1$ | $0.500_{\pm 0.019}$ | $0.098_{\pm 0.005}$ | $0.933_{\pm 0.011}$ |
| | $R_Y$, $s = 0.5$ | $0.502_{\pm 0.020}$ | $0.098_{\pm 0.005}$ | $0.899_{\pm 0.016}$ |
| | $R_Y$, $s = 2$ | $0.492_{\pm 0.023}$ | $0.098_{\pm 0.006}$ | $0.652_{\pm 0.003}$ |
| Entangling | $R_Y$, $s = 1$ | $0.999_{\pm 0.001}$ | $0.335_{\pm 0.200}$ | $0.033_{\pm 0.042}$ |
| | $R_X$, $s = 1$ | $0.999_{\pm 0.001}$ | $0.496_{\pm 0.016}$ | $0.005_{\pm 0.006}$ |
| | $R_Y$, $s = 0.5$ | $1.000_{\pm 0.000}$ | $0.199_{\pm 0.074}$ | $0.038_{\pm 0.001}$ |
| | $R_Y$, $s = 2$ | $0.532_{\pm 0.420}$ | $0.514_{\pm 0.007}$ | $0.028_{\pm 0.017}$ |
| Reuploading | $R_Y$–$R_Y$, $s = 1$ | $0.955_{\pm 0.073}$ | $0.508_{\pm 0.036}$ | $0.149_{\pm 0.073}$ |
| | $R_X$–$R_X$, $s = 1$ | $0.959_{\pm 0.072}$ | $0.359_{\pm 0.226}$ | $0.098_{\pm 0.166}$ |
| | $R_Y$–$R_Y$, $s = 0.5$ | $1.000_{\pm 0.000}$ | $0.471_{\pm 0.074}$ | $0.032_{\pm 0.019}$ |
| | $R_Y$–$R_Y$, $s = 2$ | $0.340_{\pm 0.210}$ | $0.328_{\pm 0.282}$ | $0.172_{\pm 0.030}$ |
| | $R_Y$–$R_X$, $s = 1$ | $0.970_{\pm 0.052}$ | $0.509_{\pm 0.015}$ | $0.008_{\pm 0.005}$ |
| | $R_X$–$R_Y$, $s = 1$ | $0.917_{\pm 0.138}$ | $0.365_{\pm 0.224}$ | $0.223_{\pm 0.208}$ |

The law of $(Z, Y)$ is invariant. Draw $S$ with $\Pr(S = Y \mid Y, e) = p_e$, conditionally independent of local-cue noise given $Y$, and an independent uniform $\Theta_A \in [0, 2\pi)$. Set

$$\Delta = \frac{\pi(1 - S)}{2} + \zeta, \quad \zeta \sim \mathcal{N}(0, 0.38^2), \quad \Theta_B = (\Theta_A + \Delta) \bmod 2\pi. \tag{18}$$

The input is

$$X = (\cos \Theta_A, \sin \Theta_A, \cos \Theta_B, \sin \Theta_B, Z_1, Z_2), \tag{19}$$

with $p_e = 0.9$ in sources/held-out ID, 0.5 under randomization, and 0.1 under reversal. The descriptive statistic $\cos(\Theta_A - \Theta_B)$ is not an input coordinate. Neither noisy signal is a deterministic label rule.

Uniform base phase makes each relational coordinate marginally independent of $(Y, Z)$ despite source joint predictiveness. With supplied coordinatewise preprocessing and local affine architecture (Appendix C.2), Theorem 4 favors the invariant local cue at an attained finite population optimum, supporting OOD stability under Proposition 6. Finite runs assess consistency, not population optimality or elimination of optimization/architecture differences.

**Observed role reversal.** Figure 3 (lower left) shows high original/stable-only and lower shortcut-only accuracy for local VQC on Random and OOD reversed; LR, MLP, and entangling similarly favor the invariant cue. Reuploading retains high original accuracy but weaker stable-only accuracy and source sensitivity to both blocks. Local access can therefore favor stable prediction or shortcut reliance, depending on feature roles.

Table 69 retains all five models and splits; the derived-summary plot adds source-removal uncertainty absent from the export.

**Complete accessibility-reversal ablation.** The derived-summary figure preserves paired source-removal uncertainty unavailable from marginal means/SDs. Noise plots retain architecture/mode comparisons lacking model-resolved numerical exports. Archived

Table 69: Q-LINK-Reversed fixed-checkpoint interventions. $C - S$ is shortcut-only minus stable-only accuracy; $\Delta_{\text{flip}}$ is original minus shortcut-flip accuracy (both means only). Other conventions: Appendix C.4.

| Model | Split | Original | Stable | Cue | Flip | $C - S$ | $\Delta_{\text{flip}}$ |
|---|---|---|---|---|---|---|---|
| LR | A | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.542_{\pm 0.031}$ | $1.000_{\pm 0.000}$ | $-0.458$ | $0.000$ |
| | B | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.527_{\pm 0.050}$ | $1.000_{\pm 0.000}$ | $-0.473$ | $0.000$ |
| | Source | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.536_{\pm 0.039}$ | $1.000_{\pm 0.000}$ | $-0.464$ | $0.000$ |
| | Random | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.488_{\pm 0.007}$ | $1.000_{\pm 0.000}$ | $-0.512$ | $0.000$ |
| | Reverse | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.473_{\pm 0.031}$ | $1.000_{\pm 0.000}$ | $-0.527$ | $0.000$ |
| MLP | A | $0.997_{\pm 0.004}$ | $0.999_{\pm 0.001}$ | $0.756_{\pm 0.158}$ | $0.973_{\pm 0.020}$ | $-0.243$ | $0.024$ |
| | B | $0.997_{\pm 0.004}$ | $0.998_{\pm 0.002}$ | $0.754_{\pm 0.143}$ | $0.984_{\pm 0.017}$ | $-0.244$ | $0.013$ |
| | Source | $0.997_{\pm 0.004}$ | $0.999_{\pm 0.002}$ | $0.751_{\pm 0.157}$ | $0.978_{\pm 0.018}$ | $-0.248$ | $0.019$ |
| | Random | $0.984_{\pm 0.017}$ | $0.998_{\pm 0.002}$ | $0.483_{\pm 0.028}$ | $0.984_{\pm 0.015}$ | $-0.515$ | $-0.001$ |
| | Reverse | $0.982_{\pm 0.016}$ | $0.999_{\pm 0.001}$ | $0.244_{\pm 0.142}$ | $0.997_{\pm 0.003}$ | $-0.755$ | $-0.016$ |
| Local | A | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.510_{\pm 0.009}$ | $1.000_{\pm 0.000}$ | $-0.490$ | $0.000$ |
| | B | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.514_{\pm 0.011}$ | $1.000_{\pm 0.000}$ | $-0.486$ | $0.000$ |
| | Source | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.503_{\pm 0.011}$ | $1.000_{\pm 0.000}$ | $-0.497$ | $0.000$ |
| | Random | $0.998_{\pm 0.001}$ | $0.999_{\pm 0.000}$ | $0.454_{\pm 0.010}$ | $0.999_{\pm 0.000}$ | $-0.545$ | $-0.001$ |
| | Reverse | $1.000_{\pm 0.000}$ | $1.000_{\pm 0.000}$ | $0.479_{\pm 0.013}$ | $1.000_{\pm 0.000}$ | $-0.521$ | $0.000$ |
| Ent. | A | $0.995_{\pm 0.001}$ | $0.993_{\pm 0.008}$ | $0.677_{\pm 0.100}$ | $0.966_{\pm 0.033}$ | $-0.316$ | $0.029$ |
| | B | $0.996_{\pm 0.004}$ | $0.997_{\pm 0.004}$ | $0.675_{\pm 0.093}$ | $0.971_{\pm 0.033}$ | $-0.322$ | $0.025$ |
| | Source | $0.996_{\pm 0.002}$ | $0.994_{\pm 0.007}$ | $0.690_{\pm 0.085}$ | $0.966_{\pm 0.031}$ | $-0.304$ | $0.030$ |
| | Random | $0.980_{\pm 0.021}$ | $0.995_{\pm 0.005}$ | $0.500_{\pm 0.012}$ | $0.979_{\pm 0.021}$ | $-0.495$ | $0.002$ |
| | Reverse | $0.967_{\pm 0.034}$ | $0.996_{\pm 0.005}$ | $0.335_{\pm 0.095}$ | $0.995_{\pm 0.001}$ | $-0.662$ | $-0.029$ |
| Reup. | A | $0.999_{\pm 0.001}$ | $0.580_{\pm 0.027}$ | $0.513_{\pm 0.030}$ | $0.998_{\pm 0.001}$ | $-0.067$ | $0.001$ |
| | B | $0.999_{\pm 0.001}$ | $0.568_{\pm 0.022}$ | $0.527_{\pm 0.032}$ | $1.000_{\pm 0.001}$ | $-0.041$ | $-0.000$ |
| | Source | $0.999_{\pm 0.001}$ | $0.585_{\pm 0.030}$ | $0.516_{\pm 0.034}$ | $0.999_{\pm 0.000}$ | $-0.069$ | $-0.000$ |
| | Random | $1.000_{\pm 0.001}$ | $0.608_{\pm 0.067}$ | $0.487_{\pm 0.018}$ | $0.999_{\pm 0.001}$ | $-0.121$ | $0.001$ |
| | Reverse | $0.998_{\pm 0.001}$ | $0.599_{\pm 0.040}$ | $0.471_{\pm 0.021}$ | $1.000_{\pm 0.000}$ | $-0.128$ | $-0.002$ |

training tables pool architectures: SD includes between-model variation, not seed-only uncertainty. They provide descriptive context, not evidence of a common training effect.

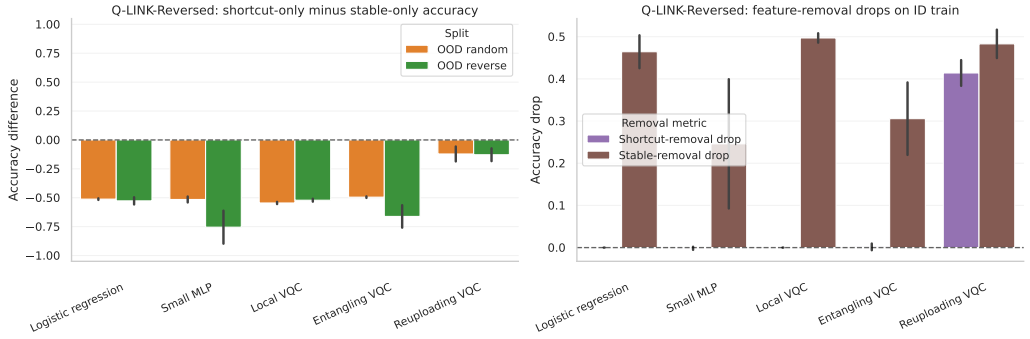


Figure 17: Q-LINK-Reversed intervention summaries: shortcut-only minus stable-only accuracy (left) and feature-removal drops (right), with paired-contrast uncertainty absent from the marginal table.

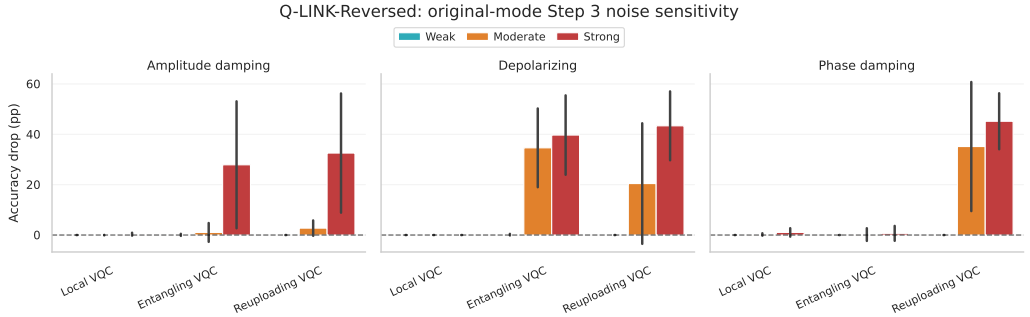


Figure 18: Q-LINK-Reversed original-mode drop from matched ideal analytic accuracy by architecture, channel, and tier. Bars/SD pool seeds and three held-out splits within model; SD is not seed-only.

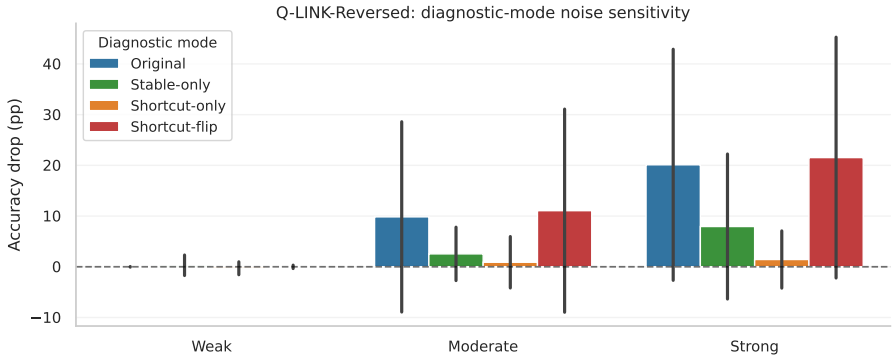


Figure 19: Q-LINK-Reversed held-out-ID drops by mode/tier. Bars/SD pool architectures, seeds, and channels, not seed-only uncertainty. Interpret small drops alongside ideal accuracy.

Table 70: Q-LINK-Reversed ideal-evaluation training summary. Mean ± SD pools model–seed entries: dispersion includes architectures and cannot establish model-specific effects.

| Training | ID | Reverse | Gap | Stable-only | Cue-only | $\Delta_{\text{flip}}$ |
|---|---|---|---|---|---|---|
| Clean | $0.999_{\pm 0.003}$ | $0.987_{\pm 0.028}$ | $0.012_{\pm 0.025}$ | $0.859_{\pm 0.211}$ | $0.551_{\pm 0.093}$ | $0.007_{\pm 0.019}$ |
| Amp mod. | $0.983_{\pm 0.019}$ | $0.937_{\pm 0.086}$ | $0.046_{\pm 0.072}$ | $0.933_{\pm 0.114}$ | $0.552_{\pm 0.124}$ | $0.044_{\pm 0.074}$ |
| Depol mod. | $0.936_{\pm 0.054}$ | $0.561_{\pm 0.371}$ | $0.375_{\pm 0.319}$ | $0.805_{\pm 0.258}$ | $0.746_{\pm 0.178}$ | $0.380_{\pm 0.318}$ |
| Phase mod. | $0.975_{\pm 0.037}$ | $0.812_{\pm 0.287}$ | $0.163_{\pm 0.251}$ | $0.896_{\pm 0.218}$ | $0.625_{\pm 0.170}$ | $0.173_{\pm 0.268}$ |
| Amp strong | $0.944_{\pm 0.061}$ | $0.905_{\pm 0.098}$ | $0.039_{\pm 0.069}$ | $0.909_{\pm 0.169}$ | $0.510_{\pm 0.065}$ | $0.026_{\pm 0.067}$ |
| Depol strong | $0.892_{\pm 0.088}$ | $0.469_{\pm 0.400}$ | $0.423_{\pm 0.318}$ | $0.673_{\pm 0.247}$ | $0.718_{\pm 0.138}$ | $0.411_{\pm 0.312}$ |
| Phase strong | $0.969_{\pm 0.046}$ | $0.794_{\pm 0.308}$ | $0.175_{\pm 0.262}$ | $0.920_{\pm 0.166}$ | $0.636_{\pm 0.171}$ | $0.175_{\pm 0.264}$ |
| Mixed mod. | $0.976_{\pm 0.034}$ | $0.881_{\pm 0.222}$ | $0.094_{\pm 0.189}$ | $0.916_{\pm 0.182}$ | $0.563_{\pm 0.143}$ | $0.103_{\pm 0.209}$ |
| Mixed strong | $0.932_{\pm 0.093}$ | $0.848_{\pm 0.269}$ | $0.084_{\pm 0.200}$ | $0.894_{\pm 0.195}$ | $0.550_{\pm 0.120}$ | $0.078_{\pm 0.190}$ |

Table 71: Q-LINK-Reversed noisy-evaluation training summary. Mean ± SD pools model–seed entries and each setting's available nonideal evaluation tags. Differing tag composition prevents matched training-effect comparisons; SD measures pooled heterogeneity.

| Training | ID | Reverse | Gap | Stable-only | Cue-only | $\Delta_{\text{flip}}$ |
|---|---|---|---|---|---|---|
| Clean | $0.849_{\pm 0.213}$ | $0.822_{\pm 0.224}$ | $0.027_{\pm 0.088}$ | $0.806_{\pm 0.233}$ | $0.540_{\pm 0.100}$ | $0.021_{\pm 0.078}$ |
| Amp mod. | $0.991_{\pm 0.016}$ | $0.970_{\pm 0.057}$ | $0.021_{\pm 0.045}$ | $0.870_{\pm 0.206}$ | $0.584_{\pm 0.129}$ | $0.024_{\pm 0.046}$ |
| Depol mod. | $0.932_{\pm 0.053}$ | $0.504_{\pm 0.386}$ | $0.428_{\pm 0.334}$ | $0.751_{\pm 0.259}$ | $0.748_{\pm 0.181}$ | $0.423_{\pm 0.330}$ |
| Phase mod. | $0.964_{\pm 0.053}$ | $0.814_{\pm 0.278}$ | $0.150_{\pm 0.225}$ | $0.834_{\pm 0.245}$ | $0.626_{\pm 0.181}$ | $0.158_{\pm 0.237}$ |
| Amp strong | $0.997_{\pm 0.004}$ | $0.988_{\pm 0.027}$ | $0.009_{\pm 0.024}$ | $0.891_{\pm 0.211}$ | $0.559_{\pm 0.104}$ | $0.007_{\pm 0.018}$ |
| Depol strong | $0.873_{\pm 0.099}$ | $0.440_{\pm 0.419}$ | $0.433_{\pm 0.323}$ | $0.681_{\pm 0.240}$ | $0.720_{\pm 0.128}$ | $0.410_{\pm 0.308}$ |
| Phase strong | $0.932_{\pm 0.108}$ | $0.801_{\pm 0.299}$ | $0.131_{\pm 0.214}$ | $0.860_{\pm 0.209}$ | $0.610_{\pm 0.163}$ | $0.114_{\pm 0.198}$ |
| Mixed mod. | $0.963_{\pm 0.047}$ | $0.851_{\pm 0.247}$ | $0.112_{\pm 0.214}$ | $0.871_{\pm 0.204}$ | $0.571_{\pm 0.131}$ | $0.114_{\pm 0.222}$ |
| Mixed strong | $0.852_{\pm 0.205}$ | $0.799_{\pm 0.257}$ | $0.053_{\pm 0.165}$ | $0.829_{\pm 0.215}$ | $0.535_{\pm 0.103}$ | $0.054_{\pm 0.149}$ |

In Tables 70 and 71, stable-only/cue-only are ID accuracies and $\Delta_{\text{flip}}$ is original minus shortcut-flip ID accuracy, not probability sensitivity. Architecture decomposition is unavailable; main-text interpretation uses model-resolved primary-task training results.

### C.9 What the simulations establish about the theory

Local Q-LINK models provide the strongest structural connection. Six-dimensional inputs avoid PCA; fixed coordinatewise standardization/clipping preserves each relational coordinate's independence from the label and local cue. Product encoding, local rotations, single-qubit readouts, independently removable relational head coefficients, and free bias satisfy Proposition 3 and Theorem 4. This establishes the structural connection to the population theorem; finite-sample Adam training does not certify attainment of a population optimum. Interventions support behavioral consistency without estimating excess risk or transfer constants (Propositions 5 and 6).

Fixed CNOT rings do not imply nested classes; successful joint processing does not prove entanglement necessary. Reuploading requires further analysis before applying the single-upload effective-observable representation. Geometric/spectral accuracy contrasts individ-

ually verify neither observable blindness nor the cue-only theorem's conditional-marginal assumptions.

Masks probe trained predictors, not optimally refitted affine recovery or class-wide obstructions. Endpoints and trajectories do not test Proposition 7, which requires an unmeasured fixed signed contrast and realized optimizer direction.

Physical-noise accuracy identifies neither feature-specific contraction nor coherence suppression. The 1024-shot control lacks per-input margins/disagreement and a shot-count sweep needed to assess Proposition 8. Noise-aware training changes the predictor rather than directly testing either noise bound. Accessibility can favor stable signals or shortcuts; OOD consequences depend on the selected signal and evaluation environment.

# D Hardware evaluation on MonarQ

We test clean-trained simulation-to-device transfer (Figure 4) and noise-aware simulation training (Table 3). Task-level diagnostics distinguish preserved predictions, lost discrimination, and changed reliance.

## D.1 Evaluation protocol and statistical reporting

Simulation-trained checkpoints are evaluated without parameter updates on MonarQ, Calcul Québec's superconducting quantum computer (Calcul Québec, 2024). Architectures follow Appendix C.2.1; tasks and interventions follow Appendices B.2–B.6. Here A denotes analytic simulation (exact noiseless expectations), F finite-shot simulation (sampled noiseless-circuit measurements), and H MonarQ (device measurements). F isolates sampling effects; H also includes hardware execution effects.

The design crosses 5 tasks, 3 VQC architectures, 9 training conditions, and seeds $\{0, 1, 2\}$, with 100 intended examples per split/input mode. Every run covers ID, Random, and OOD reversed splits and full competing-feature, stable-only, shortcut-only, and shortcut-flip inputs. Table 72 specifies nominal simulation-training channel parameters, not device-noise estimates. Mixed training cycles through amplitude, depolarizing, and phase channels, one per update, at its designated tier.

Results separate task, architecture, training condition, split, and intervention. Accuracy (%) is mean ± sample SD over three seeds $\{0, 1, 2\}$, with variance denominator $n - 1$. Paired changes subtract within corresponding seeds before aggregation, preserving covariance; their SD cannot be computed by subtracting SDs. Error bars describe seed variability, not confidence intervals over hardware sessions, and hardware differences may include session or calibration variation.

All reported conditions and paired comparisons use all three seeds. Coverage and raw-record reconciliation are described in Appendix D.6.

Table 72: **Simulation-training conditions.** Parameters apply only during training; hardware evaluation uses MonarQ.

| Training | Moderate | Strong |
|---|---|---|
| Clean | none | none |
| Amplitude | 0.05 | 0.10 |
| Depolarizing | 0.08 | 0.15 |
| Phase | 0.5 | 0.6 |
| Mixed | moderate channel cycle | strong channel cycle |

## D.2 Training conditions and transfer across tasks

Figures 20–24 show OOD reversed accuracy on full competing-feature inputs across all 9 training conditions, unlike the clean-only Figure 4. Rows separate VQC architectures; grouped bars compare A/F/H, whiskers show seed SD, and the dashed line marks 50%. Appendix D.5 reports paired changes from clean training.

On geometric tasks, local hardware predictors remain poor under reversal; entangling and reuploading predictors often approach chance. These increases can accompany lost ID discrimination and do not alone establish stable-feature recovery (Appendix D.3). On Spurious-Fourier, local predictors retain substantial reversed accuracy and benefit from several noise-aware checkpoints; entangling and reuploading predictors remain closer to chance. Q-FOS hardware performance remains modest: spectral task membership alone does not determine transfer; encoding, architecture, and checkpoint jointly matter.

Q-LINK shows the largest transfer loss for predictors exploiting simulated relational information; noise-aware training generally fails to preserve their reversed accuracy, while local predictors remain poor. Stable-only diagnostics distinguish simulated architectural access from its reliable device use. These checkpoint- and hardware-specific observations do not identify a unique physical noise mechanism.

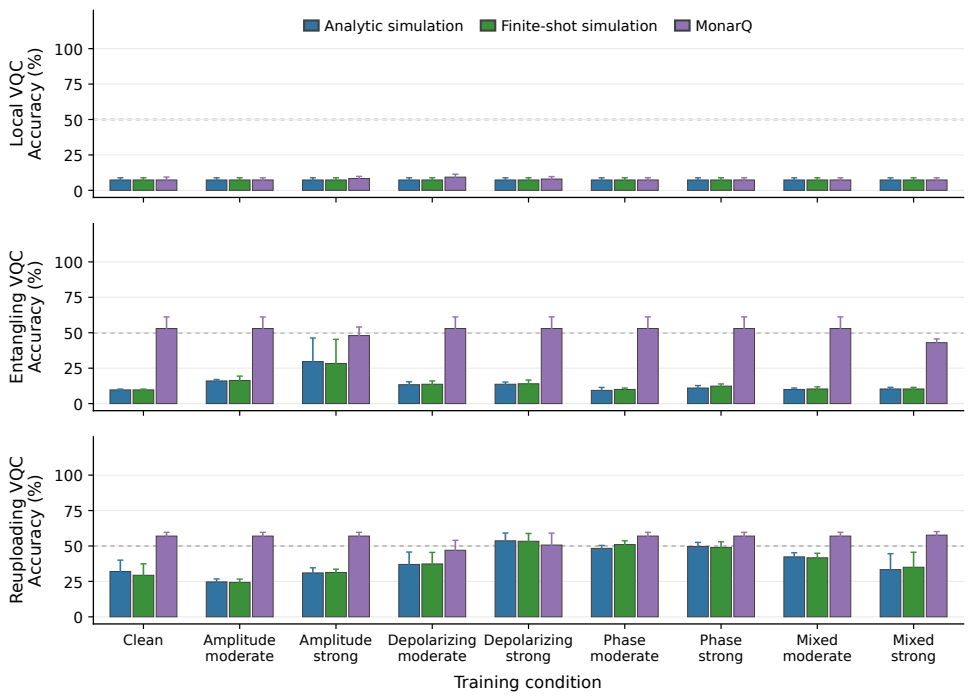


Figure 20: **Q-SPHERE: training and hardware transfer.** OOD reversed accuracy on full competing-feature inputs. Paired changes: Table 106; diagnostics: Appendix D.3.

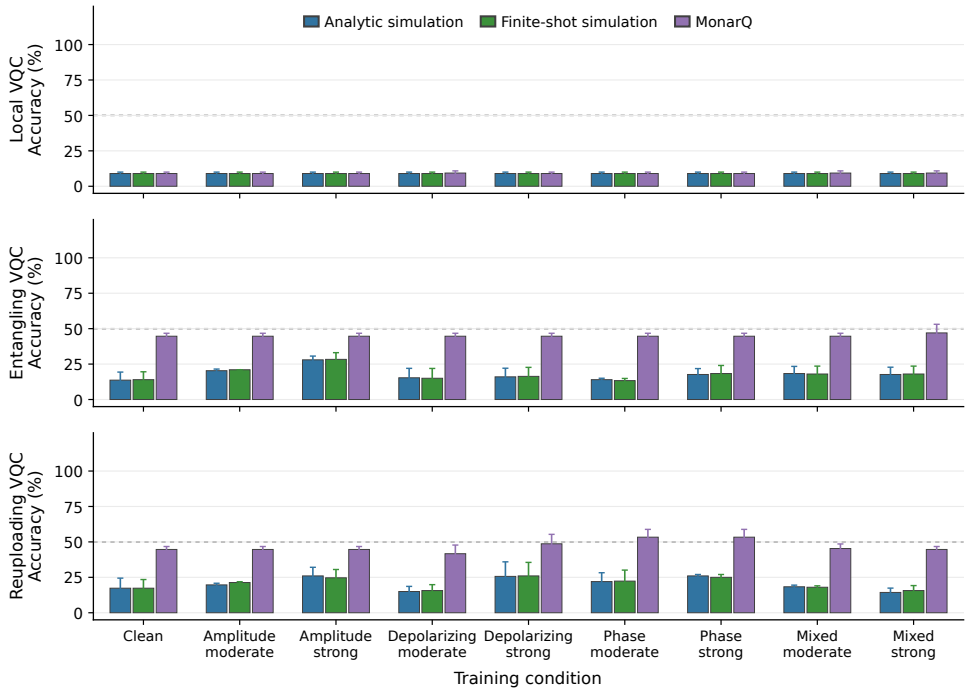


Figure 21: **Spirals-Signatures: training and hardware transfer.** OOD reversed accuracy on full competing-feature inputs. Paired changes: Table 107; diagnostics: Appendix D.3.

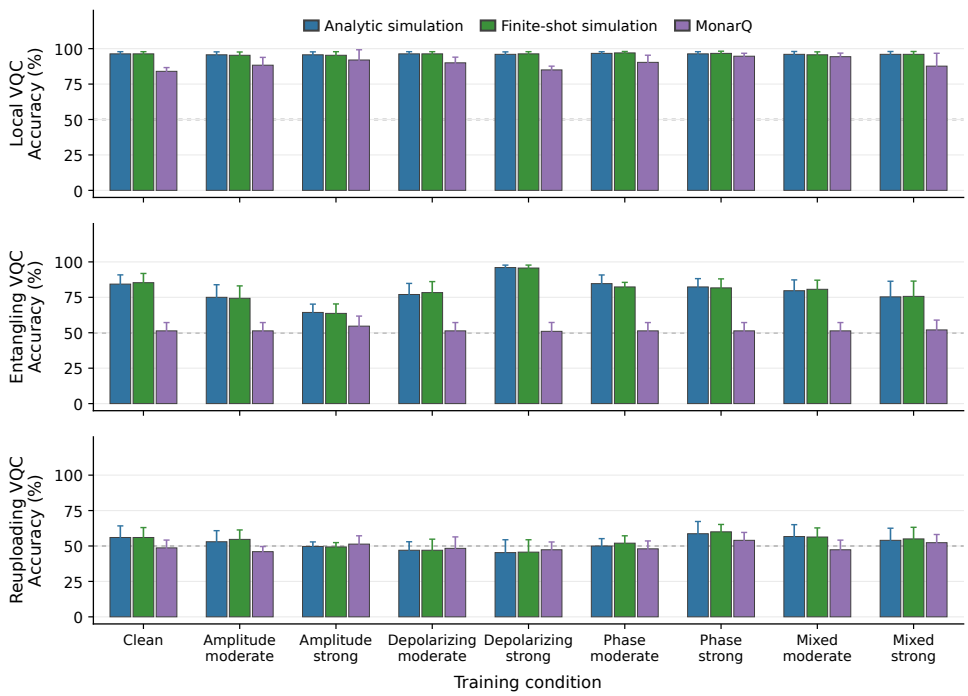


Figure 22: **Spurious-Fourier: training and hardware transfer.** OOD reversed accuracy on full competing-feature inputs. Paired changes: Table 108; diagnostics: Appendix D.3.

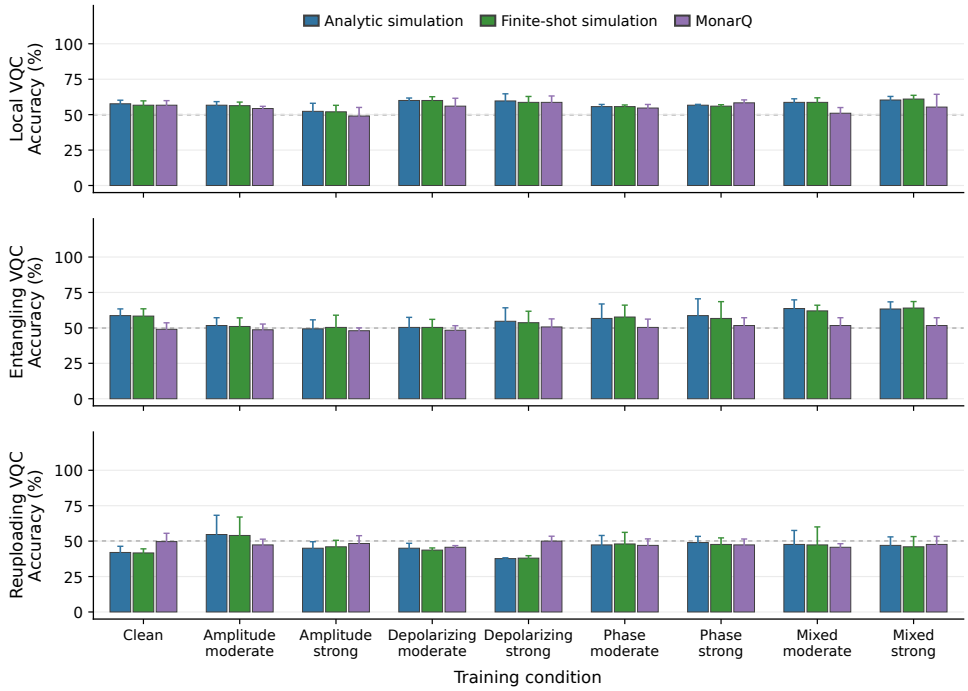


Figure 23: **Q-FOS: training and hardware transfer.** OOD reversed accuracy on full competing-feature inputs. Paired changes: Table 109; diagnostics: Appendix D.3.

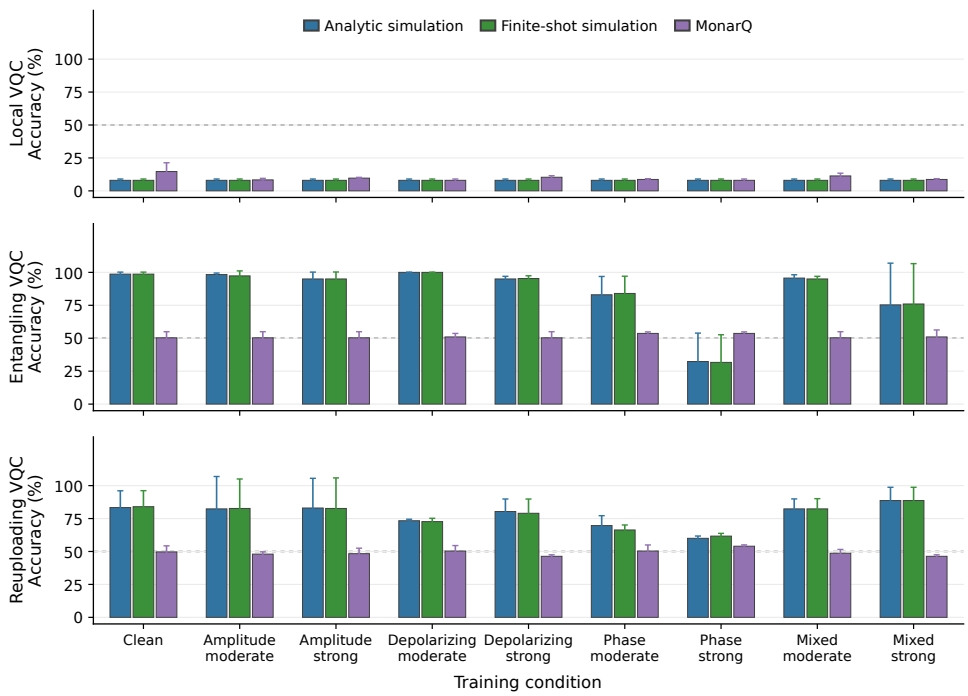


Figure 24: **Q-LINK: training and hardware transfer.** OOD reversed accuracy on full competing-feature inputs. Paired changes: Table 110; diagnostics: Appendix D.3.

## D.3 Split and intervention diagnostics

Tables cover every training condition and architecture for ID, Random, and the three OOD reversed interventions; full competing-feature OOD reversed results are in Appendix D.2.

Stable-only removes the shortcut block, shortcut-only removes the stable block, and shortcut-flip reverses the designated shortcut. These fitted-predictor interventions can leave training support and do not establish population optimality. Q-LINK's simulated-to-hardware stable-only loss shows that available relational prediction can fail during device execution; the local Spurious-Fourier model retains stronger stable-only performance. Similar full-task accuracies can therefore mask different intervention responses.

### D.3.1 Q-SPHERE

### D.3.2 Spirals-Signatures

### D.3.3 Spurious-Fourier

### D.3.4 Q-FOS

### D.3.5 Q-LINK

## D.4 Prediction agreement and logit changes

Tables 103–105 compare clean-trained predictions on matched full competing-feature examples. Agreement (%) is the fraction of matching labels; $d_{X,A}$ is the mean absolute logit difference between setting X and analytic simulation. Both are averaged within seed before seed aggregation. Similar accuracy need not imply matching predictions; differing model logit scales restrict discrepancy comparisons to a fixed predictor.

For entangling Q-LINK under reversal, F–A agreement is 99.3%, versus 50.3% for H–A, separating sampling variability from hardware transfer failure. Local Spirals-Signatures preserves its analytic decisions, including poor reversal accuracy: decision fidelity need not imply robustness. Consistent with Proposition 8, sampling fidelity concerns the measured circuit, whereas hardware execution can change its effective responses.

Table 73: **Q-SPHERE: ID diagnostics.** Full competing-feature task and stable-only accuracy (%), mean ± seed SD.

| Training | Full competing-feature task | | | Stable-only | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $49.7_{\pm 6.7}$ | $49.7_{\pm 6.7}$ | $48.7_{\pm 7.6}$ |
| Amplitude, moderate | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.0_{\pm 3.6}$ | $50.0_{\pm 7.9}$ | $47.3_{\pm 6.7}$ | $48.7_{\pm 4.7}$ |
| Amplitude, strong | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $49.3_{\pm 8.4}$ | $47.3_{\pm 7.6}$ | $53.7_{\pm 6.4}$ |
| Depolarizing, moderate | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $87.3_{\pm 2.9}$ | $49.3_{\pm 6.8}$ | $48.3_{\pm 5.9}$ | $51.3_{\pm 2.5}$ |
| Depolarizing, strong | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.0_{\pm 3.5}$ | $49.3_{\pm 6.8}$ | $47.7_{\pm 7.4}$ | $52.0_{\pm 4.4}$ |
| Phase, moderate | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $50.0_{\pm 7.2}$ | $49.7_{\pm 9.8}$ | $50.7_{\pm 3.1}$ |
| Phase, strong | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $50.0_{\pm 7.2}$ | $50.0_{\pm 5.6}$ | $47.0_{\pm 4.4}$ |
| Mixed, moderate | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $49.3_{\pm 7.6}$ | $50.0_{\pm 7.2}$ | $54.0_{\pm 7.2}$ |
| Mixed, strong | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.0_{\pm 3.5}$ | $50.0_{\pm 6.0}$ | $48.0_{\pm 6.1}$ | $51.0_{\pm 3.5}$ |
| **Entangling VQC** | | | | | | |
| Clean | $86.0_{\pm 5.3}$ | $85.0_{\pm 4.4}$ | $52.7_{\pm 4.2}$ | $47.0_{\pm 2.0}$ | $46.0_{\pm 0.0}$ | $52.7_{\pm 4.2}$ |
| Amplitude, moderate | $75.7_{\pm 3.5}$ | $76.3_{\pm 4.2}$ | $52.7_{\pm 4.2}$ | $47.3_{\pm 5.8}$ | $47.7_{\pm 4.0}$ | $52.7_{\pm 4.2}$ |
| Amplitude, strong | $69.0_{\pm 14.7}$ | $69.3_{\pm 13.8}$ | $48.3_{\pm 7.5}$ | $45.3_{\pm 6.5}$ | $44.0_{\pm 8.0}$ | $52.0_{\pm 4.0}$ |
| Depolarizing, moderate | $82.3_{\pm 4.5}$ | $82.3_{\pm 2.5}$ | $52.7_{\pm 4.2}$ | $46.7_{\pm 3.2}$ | $45.0_{\pm 3.0}$ | $50.7_{\pm 7.4}$ |
| Depolarizing, strong | $82.7_{\pm 2.5}$ | $81.7_{\pm 4.0}$ | $52.7_{\pm 3.5}$ | $51.0_{\pm 4.4}$ | $48.0_{\pm 6.1}$ | $52.3_{\pm 4.7}$ |
| Phase, moderate | $85.0_{\pm 4.0}$ | $85.0_{\pm 3.0}$ | $52.7_{\pm 4.2}$ | $48.0_{\pm 6.9}$ | $47.7_{\pm 7.0}$ | $52.7_{\pm 4.2}$ |
| Phase, strong | $84.7_{\pm 5.1}$ | $84.3_{\pm 6.0}$ | $52.7_{\pm 4.2}$ | $48.0_{\pm 4.4}$ | $46.7_{\pm 3.5}$ | $52.7_{\pm 4.2}$ |
| Mixed, moderate | $84.3_{\pm 4.0}$ | $85.0_{\pm 6.0}$ | $52.7_{\pm 4.2}$ | $47.7_{\pm 6.7}$ | $48.3_{\pm 6.7}$ | $52.7_{\pm 4.2}$ |
| Mixed, strong | $84.3_{\pm 5.7}$ | $84.0_{\pm 6.2}$ | $46.0_{\pm 2.0}$ | $56.0_{\pm 3.6}$ | $56.3_{\pm 2.5}$ | $46.0_{\pm 2.0}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $83.3_{\pm 6.0}$ | $82.0_{\pm 4.6}$ | $54.0_{\pm 2.0}$ | $63.3_{\pm 3.1}$ | $60.7_{\pm 2.5}$ | $54.0_{\pm 2.0}$ |
| Amplitude, moderate | $77.0_{\pm 3.5}$ | $77.7_{\pm 6.1}$ | $54.0_{\pm 2.0}$ | $58.3_{\pm 5.5}$ | $56.0_{\pm 2.0}$ | $54.0_{\pm 2.0}$ |
| Amplitude, strong | $69.7_{\pm 2.5}$ | $69.3_{\pm 0.6}$ | $54.0_{\pm 2.0}$ | $47.3_{\pm 1.2}$ | $47.3_{\pm 1.2}$ | $54.0_{\pm 2.0}$ |
| Depolarizing, moderate | $80.0_{\pm 3.6}$ | $78.7_{\pm 2.5}$ | $52.7_{\pm 4.9}$ | $61.3_{\pm 4.9}$ | $59.7_{\pm 6.4}$ | $50.0_{\pm 7.0}$ |
| Depolarizing, strong | $58.7_{\pm 5.5}$ | $61.7_{\pm 3.2}$ | $50.7_{\pm 5.8}$ | $54.7_{\pm 6.8}$ | $55.0_{\pm 6.2}$ | $55.0_{\pm 1.0}$ |
| Phase, moderate | $70.0_{\pm 1.0}$ | $68.0_{\pm 2.6}$ | $54.0_{\pm 2.0}$ | $59.7_{\pm 5.0}$ | $59.3_{\pm 7.5}$ | $54.0_{\pm 2.0}$ |
| Phase, strong | $69.0_{\pm 6.6}$ | $68.0_{\pm 6.9}$ | $54.0_{\pm 2.0}$ | $59.7_{\pm 5.9}$ | $60.0_{\pm 7.0}$ | $54.0_{\pm 2.0}$ |
| Mixed, moderate | $76.7_{\pm 5.8}$ | $77.3_{\pm 6.4}$ | $54.0_{\pm 2.0}$ | $60.3_{\pm 3.5}$ | $60.7_{\pm 5.8}$ | $54.0_{\pm 2.0}$ |
| Mixed, strong | $74.3_{\pm 4.7}$ | $74.7_{\pm 3.5}$ | $55.0_{\pm 1.0}$ | $61.0_{\pm 5.6}$ | $58.7_{\pm 5.5}$ | $54.3_{\pm 2.5}$ |

Table 74: **Q-SPHERE: ID diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ |
| Amplitude, moderate | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ |
| Amplitude, strong | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ |
| Depolarizing, moderate | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $87.3_{\pm 3.1}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $13.7_{\pm 3.2}$ |
| Depolarizing, strong | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ |
| Phase, moderate | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ |
| Phase, strong | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ |
| Mixed, moderate | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ |
| Mixed, strong | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $88.7_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ | $11.3_{\pm 4.2}$ |
| **Entangling VQC** | | | | | | |
| Clean | $86.3_{\pm 4.7}$ | $86.0_{\pm 4.4}$ | $52.7_{\pm 4.2}$ | $14.7_{\pm 1.2}$ | $15.3_{\pm 1.2}$ | $52.7_{\pm 4.2}$ |
| Amplitude, moderate | $85.0_{\pm 4.4}$ | $85.0_{\pm 4.4}$ | $52.7_{\pm 4.2}$ | $25.0_{\pm 1.0}$ | $24.0_{\pm 2.0}$ | $52.7_{\pm 4.2}$ |
| Amplitude, strong | $84.3_{\pm 5.0}$ | $84.3_{\pm 5.0}$ | $51.0_{\pm 4.4}$ | $30.0_{\pm 9.5}$ | $30.0_{\pm 8.5}$ | $49.7_{\pm 5.7}$ |
| Depolarizing, moderate | $85.0_{\pm 4.4}$ | $85.0_{\pm 4.4}$ | $47.3_{\pm 3.5}$ | $17.3_{\pm 2.5}$ | $17.0_{\pm 3.0}$ | $54.0_{\pm 8.5}$ |
| Depolarizing, strong | $85.0_{\pm 4.4}$ | $85.0_{\pm 4.4}$ | $52.7_{\pm 4.2}$ | $17.3_{\pm 1.5}$ | $16.7_{\pm 1.2}$ | $52.7_{\pm 4.2}$ |
| Phase, moderate | $86.3_{\pm 3.8}$ | $86.0_{\pm 3.6}$ | $52.7_{\pm 4.2}$ | $16.3_{\pm 2.9}$ | $16.7_{\pm 2.5}$ | $52.7_{\pm 4.2}$ |
| Phase, strong | $85.0_{\pm 5.3}$ | $85.3_{\pm 4.7}$ | $52.7_{\pm 4.2}$ | $17.0_{\pm 2.0}$ | $16.7_{\pm 2.5}$ | $52.7_{\pm 4.2}$ |
| Mixed, moderate | $86.3_{\pm 3.8}$ | $86.0_{\pm 3.6}$ | $52.7_{\pm 4.2}$ | $17.3_{\pm 2.3}$ | $18.0_{\pm 1.0}$ | $52.7_{\pm 4.2}$ |
| Mixed, strong | $87.7_{\pm 4.0}$ | $87.7_{\pm 4.0}$ | $46.0_{\pm 2.0}$ | $15.0_{\pm 5.3}$ | $13.7_{\pm 4.6}$ | $46.0_{\pm 2.0}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $71.7_{\pm 7.6}$ | $72.0_{\pm 6.1}$ | $54.0_{\pm 2.0}$ | $29.3_{\pm 2.5}$ | $29.7_{\pm 3.5}$ | $54.0_{\pm 2.0}$ |
| Amplitude, moderate | $82.7_{\pm 2.1}$ | $82.3_{\pm 3.1}$ | $54.0_{\pm 2.0}$ | $35.7_{\pm 4.9}$ | $36.7_{\pm 4.2}$ | $54.0_{\pm 2.0}$ |
| Amplitude, strong | $75.0_{\pm 3.0}$ | $73.7_{\pm 2.1}$ | $54.0_{\pm 2.0}$ | $37.7_{\pm 0.6}$ | $37.3_{\pm 0.6}$ | $54.0_{\pm 2.0}$ |
| Depolarizing, moderate | $65.0_{\pm 24.3}$ | $65.3_{\pm 22.8}$ | $53.3_{\pm 5.0}$ | $34.0_{\pm 6.0}$ | $35.0_{\pm 4.4}$ | $52.0_{\pm 5.3}$ |
| Depolarizing, strong | $54.3_{\pm 17.7}$ | $53.7_{\pm 16.2}$ | $51.3_{\pm 5.0}$ | $48.7_{\pm 10.2}$ | $51.0_{\pm 12.8}$ | $50.7_{\pm 6.1}$ |
| Phase, moderate | $76.3_{\pm 5.1}$ | $76.3_{\pm 6.7}$ | $54.0_{\pm 2.0}$ | $47.0_{\pm 6.2}$ | $45.7_{\pm 6.5}$ | $54.0_{\pm 2.0}$ |
| Phase, strong | $74.0_{\pm 6.6}$ | $76.0_{\pm 6.2}$ | $54.0_{\pm 2.0}$ | $45.3_{\pm 6.8}$ | $45.7_{\pm 6.4}$ | $54.0_{\pm 2.0}$ |
| Mixed, moderate | $75.3_{\pm 7.1}$ | $76.0_{\pm 5.3}$ | $54.0_{\pm 2.0}$ | $39.0_{\pm 2.0}$ | $40.0_{\pm 1.0}$ | $53.7_{\pm 2.5}$ |
| Mixed, strong | $81.0_{\pm 5.0}$ | $80.0_{\pm 2.6}$ | $56.0_{\pm 1.7}$ | $35.7_{\pm 8.1}$ | $36.3_{\pm 7.1}$ | $53.3_{\pm 4.0}$ |

Table 75: **Q-SPHERE: Random diagnostics.** Full competing-feature task and stable-only accuracy (%), mean ± seed SD.

| Training | Full competing-feature task | | | Stable-only | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $49.7_{\pm 9.7}$ | $48.3_{\pm 9.5}$ | $56.3_{\pm 4.7}$ |
| Amplitude, moderate | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $47.7_{\pm 7.6}$ | $48.3_{\pm 8.1}$ | $45.0_{\pm 7.2}$ |
| Amplitude, strong | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $47.3_{\pm 10.4}$ | $45.0_{\pm 7.9}$ | $55.3_{\pm 5.5}$ |
| Depolarizing, moderate | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $49.7_{\pm 9.7}$ | $48.7_{\pm 8.7}$ | $57.7_{\pm 2.9}$ |
| Depolarizing, strong | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $50.0_{\pm 1.0}$ | $51.3_{\pm 10.7}$ | $51.3_{\pm 10.8}$ | $59.7_{\pm 5.5}$ |
| Phase, moderate | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $50.0_{\pm 1.0}$ | $50.3_{\pm 9.9}$ | $49.7_{\pm 8.7}$ | $51.7_{\pm 5.1}$ |
| Phase, strong | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $50.3_{\pm 9.8}$ | $47.0_{\pm 7.5}$ | $44.0_{\pm 4.4}$ |
| Mixed, moderate | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $50.0_{\pm 1.0}$ | $50.0_{\pm 9.6}$ | $52.3_{\pm 11.6}$ | $54.7_{\pm 2.5}$ |
| Mixed, strong | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $50.0_{\pm 1.0}$ | $50.7_{\pm 10.2}$ | $49.7_{\pm 9.5}$ | $51.7_{\pm 12.3}$ |
| **Entangling VQC** | | | | | | |
| Clean | $47.0_{\pm 1.0}$ | $46.3_{\pm 1.5}$ | $53.7_{\pm 8.7}$ | $47.0_{\pm 9.5}$ | $46.0_{\pm 8.9}$ | $53.7_{\pm 8.7}$ |
| Amplitude, moderate | $46.3_{\pm 2.5}$ | $46.7_{\pm 2.5}$ | $53.7_{\pm 8.7}$ | $47.3_{\pm 9.7}$ | $47.3_{\pm 8.7}$ | $53.7_{\pm 8.7}$ |
| Amplitude, strong | $48.3_{\pm 5.5}$ | $47.3_{\pm 6.8}$ | $53.3_{\pm 8.6}$ | $46.7_{\pm 9.8}$ | $46.7_{\pm 9.9}$ | $52.3_{\pm 8.5}$ |
| Depolarizing, moderate | $47.0_{\pm 1.7}$ | $45.7_{\pm 0.6}$ | $53.7_{\pm 8.7}$ | $53.7_{\pm 3.5}$ | $52.7_{\pm 1.5}$ | $53.7_{\pm 8.7}$ |
| Depolarizing, strong | $47.7_{\pm 1.5}$ | $48.7_{\pm 2.5}$ | $53.7_{\pm 8.7}$ | $52.7_{\pm 3.2}$ | $53.3_{\pm 3.1}$ | $53.7_{\pm 8.7}$ |
| Phase, moderate | $47.7_{\pm 2.1}$ | $46.7_{\pm 3.1}$ | $53.7_{\pm 8.7}$ | $52.3_{\pm 11.2}$ | $51.3_{\pm 9.2}$ | $53.7_{\pm 8.7}$ |
| Phase, strong | $48.7_{\pm 1.5}$ | $48.3_{\pm 2.5}$ | $53.7_{\pm 8.7}$ | $49.3_{\pm 5.1}$ | $50.0_{\pm 6.2}$ | $53.7_{\pm 8.7}$ |
| Mixed, moderate | $47.7_{\pm 2.1}$ | $46.0_{\pm 2.6}$ | $53.7_{\pm 8.7}$ | $50.3_{\pm 9.5}$ | $48.7_{\pm 9.1}$ | $53.7_{\pm 8.7}$ |
| Mixed, strong | $51.0_{\pm 3.6}$ | $50.3_{\pm 4.2}$ | $42.3_{\pm 2.9}$ | $59.0_{\pm 2.0}$ | $59.7_{\pm 1.5}$ | $42.3_{\pm 2.9}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $56.0_{\pm 4.6}$ | $59.3_{\pm 3.1}$ | $57.7_{\pm 2.9}$ | $58.0_{\pm 3.6}$ | $55.7_{\pm 1.5}$ | $57.7_{\pm 2.9}$ |
| Amplitude, moderate | $53.3_{\pm 3.2}$ | $55.0_{\pm 3.6}$ | $57.7_{\pm 2.9}$ | $54.7_{\pm 3.2}$ | $54.3_{\pm 3.1}$ | $57.7_{\pm 2.9}$ |
| Amplitude, strong | $49.3_{\pm 3.2}$ | $49.3_{\pm 3.2}$ | $57.7_{\pm 2.9}$ | $43.0_{\pm 1.7}$ | $43.3_{\pm 2.1}$ | $57.7_{\pm 2.9}$ |
| Depolarizing, moderate | $55.3_{\pm 2.3}$ | $55.3_{\pm 2.3}$ | $50.7_{\pm 9.1}$ | $60.7_{\pm 1.5}$ | $58.7_{\pm 3.1}$ | $50.3_{\pm 9.3}$ |
| Depolarizing, strong | $59.0_{\pm 5.3}$ | $58.3_{\pm 6.7}$ | $54.7_{\pm 7.1}$ | $59.7_{\pm 6.7}$ | $59.0_{\pm 5.2}$ | $54.7_{\pm 7.1}$ |
| Phase, moderate | $60.0_{\pm 9.5}$ | $58.7_{\pm 7.8}$ | $57.7_{\pm 2.9}$ | $63.3_{\pm 1.5}$ | $64.7_{\pm 3.1}$ | $57.7_{\pm 2.9}$ |
| Phase, strong | $62.0_{\pm 3.5}$ | $60.3_{\pm 4.9}$ | $57.7_{\pm 2.9}$ | $61.0_{\pm 3.0}$ | $61.7_{\pm 3.8}$ | $57.7_{\pm 2.9}$ |
| Mixed, moderate | $54.7_{\pm 4.2}$ | $55.3_{\pm 5.0}$ | $57.7_{\pm 2.9}$ | $65.3_{\pm 3.2}$ | $63.0_{\pm 2.6}$ | $57.7_{\pm 2.9}$ |
| Mixed, strong | $53.0_{\pm 8.7}$ | $54.3_{\pm 7.6}$ | $55.7_{\pm 0.6}$ | $62.3_{\pm 6.1}$ | $63.7_{\pm 4.0}$ | $58.0_{\pm 2.0}$ |

Table 76: **Q-SPHERE: Random diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $50.0_{\pm 1.0}$ | $50.0_{\pm 1.0}$ | $50.0_{\pm 1.0}$ |
| Amplitude, moderate | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $50.0_{\pm 1.0}$ | $50.0_{\pm 1.0}$ | $50.0_{\pm 1.0}$ |
| Amplitude, strong | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $50.0_{\pm 1.0}$ | $50.0_{\pm 1.0}$ | $50.3_{\pm 1.5}$ |
| Depolarizing, moderate | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $50.3_{\pm 0.6}$ | $50.3_{\pm 1.5}$ | $50.3_{\pm 1.5}$ | $49.3_{\pm 0.6}$ |
| Depolarizing, strong | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $50.3_{\pm 1.5}$ | $50.3_{\pm 1.5}$ | $50.3_{\pm 1.5}$ |
| Phase, moderate | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $50.0_{\pm 1.0}$ | $50.0_{\pm 1.0}$ | $50.3_{\pm 1.5}$ | $50.3_{\pm 1.5}$ |
| Phase, strong | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $50.0_{\pm 1.0}$ | $50.0_{\pm 1.0}$ | $50.0_{\pm 1.0}$ |
| Mixed, moderate | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $50.0_{\pm 1.0}$ | $50.0_{\pm 1.0}$ | $50.0_{\pm 1.0}$ | $50.3_{\pm 1.5}$ |
| Mixed, strong | $49.7_{\pm 1.5}$ | $49.7_{\pm 1.5}$ | $50.0_{\pm 1.0}$ | $50.0_{\pm 1.0}$ | $50.3_{\pm 1.5}$ | $50.3_{\pm 1.5}$ |
| **Entangling VQC** | | | | | | |
| Clean | $47.7_{\pm 0.6}$ | $48.0_{\pm 1.0}$ | $53.7_{\pm 8.7}$ | $47.7_{\pm 3.2}$ | $47.0_{\pm 4.4}$ | $53.7_{\pm 8.7}$ |
| Amplitude, moderate | $47.0_{\pm 1.7}$ | $47.3_{\pm 1.5}$ | $53.7_{\pm 8.7}$ | $50.0_{\pm 3.5}$ | $50.0_{\pm 4.4}$ | $53.7_{\pm 8.7}$ |
| Amplitude, strong | $47.0_{\pm 1.7}$ | $47.0_{\pm 1.7}$ | $52.0_{\pm 8.5}$ | $48.7_{\pm 8.3}$ | $50.0_{\pm 8.5}$ | $49.3_{\pm 10.1}$ |
| Depolarizing, moderate | $47.7_{\pm 2.1}$ | $47.7_{\pm 2.1}$ | $53.7_{\pm 8.7}$ | $48.7_{\pm 2.1}$ | $50.0_{\pm 2.0}$ | $53.7_{\pm 8.7}$ |
| Depolarizing, strong | $48.0_{\pm 1.7}$ | $48.7_{\pm 2.1}$ | $53.7_{\pm 8.7}$ | $48.7_{\pm 1.5}$ | $49.0_{\pm 2.6}$ | $53.7_{\pm 8.7}$ |
| Phase, moderate | $47.3_{\pm 1.5}$ | $47.3_{\pm 1.5}$ | $53.7_{\pm 8.7}$ | $49.7_{\pm 2.5}$ | $49.7_{\pm 2.5}$ | $53.7_{\pm 8.7}$ |
| Phase, strong | $48.0_{\pm 1.7}$ | $48.0_{\pm 1.7}$ | $53.7_{\pm 8.7}$ | $50.7_{\pm 3.2}$ | $50.0_{\pm 2.6}$ | $53.7_{\pm 8.7}$ |
| Mixed, moderate | $47.3_{\pm 1.5}$ | $48.0_{\pm 1.7}$ | $53.7_{\pm 8.7}$ | $49.3_{\pm 3.5}$ | $49.7_{\pm 4.9}$ | $53.7_{\pm 8.7}$ |
| Mixed, strong | $50.7_{\pm 2.1}$ | $50.7_{\pm 2.1}$ | $42.3_{\pm 2.9}$ | $56.0_{\pm 1.0}$ | $54.0_{\pm 2.6}$ | $42.3_{\pm 2.9}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $51.3_{\pm 5.5}$ | $52.0_{\pm 5.3}$ | $57.7_{\pm 2.9}$ | $58.0_{\pm 2.6}$ | $56.7_{\pm 3.1}$ | $57.7_{\pm 2.9}$ |
| Amplitude, moderate | $46.3_{\pm 1.5}$ | $46.7_{\pm 2.1}$ | $57.7_{\pm 2.9}$ | $56.3_{\pm 3.2}$ | $55.3_{\pm 0.6}$ | $57.7_{\pm 2.9}$ |
| Amplitude, strong | $43.7_{\pm 5.1}$ | $43.0_{\pm 6.0}$ | $57.7_{\pm 2.9}$ | $50.7_{\pm 5.0}$ | $50.3_{\pm 5.1}$ | $57.7_{\pm 2.9}$ |
| Depolarizing, moderate | $51.0_{\pm 3.6}$ | $52.3_{\pm 2.5}$ | $51.0_{\pm 8.9}$ | $61.0_{\pm 3.6}$ | $62.7_{\pm 3.1}$ | $51.0_{\pm 9.5}$ |
| Depolarizing, strong | $48.3_{\pm 1.5}$ | $48.3_{\pm 2.3}$ | $54.7_{\pm 11.1}$ | $57.0_{\pm 3.0}$ | $59.7_{\pm 1.2}$ | $53.3_{\pm 9.3}$ |
| Phase, moderate | $47.7_{\pm 2.3}$ | $46.7_{\pm 2.1}$ | $57.7_{\pm 2.9}$ | $58.7_{\pm 5.7}$ | $60.3_{\pm 4.5}$ | $57.7_{\pm 2.9}$ |
| Phase, strong | $49.3_{\pm 4.5}$ | $50.3_{\pm 3.1}$ | $57.7_{\pm 2.9}$ | $59.7_{\pm 6.7}$ | $59.3_{\pm 5.5}$ | $57.7_{\pm 2.9}$ |
| Mixed, moderate | $48.7_{\pm 2.5}$ | $48.3_{\pm 3.1}$ | $57.7_{\pm 2.9}$ | $60.7_{\pm 2.1}$ | $61.7_{\pm 2.5}$ | $57.7_{\pm 2.9}$ |
| Mixed, strong | $48.3_{\pm 2.5}$ | $48.3_{\pm 2.1}$ | $56.7_{\pm 3.1}$ | $60.7_{\pm 8.7}$ | $59.7_{\pm 10.7}$ | $54.7_{\pm 1.5}$ |

Table 77: **Q-SPHERE: OOD reversed diagnostics.** Stable-only accuracy (%), mean ± seed SD.

| Training | Stable-only | | |
|---|---|---|---|
| | A | F | H |
| **Local VQC** | | | |
| Clean | $51.3_{\pm 6.5}$ | $49.7_{\pm 7.2}$ | $52.7_{\pm 7.5}$ |
| Amplitude, moderate | $51.7_{\pm 7.0}$ | $53.3_{\pm 7.6}$ | $57.3_{\pm 8.5}$ |
| Amplitude, strong | $50.3_{\pm 7.6}$ | $51.0_{\pm 7.9}$ | $52.3_{\pm 7.1}$ |
| Depolarizing, moderate | $50.0_{\pm 5.6}$ | $50.7_{\pm 7.4}$ | $58.0_{\pm 2.6}$ |
| Depolarizing, strong | $52.0_{\pm 8.2}$ | $51.0_{\pm 7.2}$ | $51.7_{\pm 5.0}$ |
| Phase, moderate | $53.0_{\pm 7.5}$ | $52.0_{\pm 6.2}$ | $52.3_{\pm 8.0}$ |
| Phase, strong | $53.7_{\pm 8.5}$ | $51.7_{\pm 7.0}$ | $49.3_{\pm 5.1}$ |
| Mixed, moderate | $52.0_{\pm 7.0}$ | $51.3_{\pm 5.7}$ | $55.0_{\pm 3.0}$ |
| Mixed, strong | $50.7_{\pm 4.9}$ | $50.0_{\pm 4.6}$ | $54.3_{\pm 6.7}$ |
| **Entangling VQC** | | | |
| Clean | $51.7_{\pm 8.4}$ | $50.0_{\pm 3.0}$ | $53.0_{\pm 8.2}$ |
| Amplitude, moderate | $48.7_{\pm 9.9}$ | $47.7_{\pm 9.9}$ | $53.0_{\pm 8.2}$ |
| Amplitude, strong | $48.0_{\pm 7.8}$ | $48.7_{\pm 7.2}$ | $47.0_{\pm 7.0}$ |
| Depolarizing, moderate | $51.0_{\pm 4.6}$ | $48.7_{\pm 6.7}$ | $53.0_{\pm 8.2}$ |
| Depolarizing, strong | $49.0_{\pm 10.4}$ | $46.7_{\pm 9.3}$ | $53.0_{\pm 8.2}$ |
| Phase, moderate | $49.7_{\pm 6.4}$ | $51.0_{\pm 5.3}$ | $53.0_{\pm 8.2}$ |
| Phase, strong | $49.0_{\pm 7.2}$ | $49.7_{\pm 5.5}$ | $53.0_{\pm 8.2}$ |
| Mixed, moderate | $48.0_{\pm 7.0}$ | $48.0_{\pm 7.0}$ | $53.0_{\pm 8.2}$ |
| Mixed, strong | $56.7_{\pm 0.6}$ | $57.7_{\pm 0.6}$ | $43.0_{\pm 2.6}$ |
| **Reuploading VQC** | | | |
| Clean | $68.0_{\pm 2.6}$ | $65.7_{\pm 3.1}$ | $57.0_{\pm 2.6}$ |
| Amplitude, moderate | $57.3_{\pm 7.2}$ | $56.3_{\pm 7.2}$ | $57.0_{\pm 2.6}$ |
| Amplitude, strong | $44.7_{\pm 2.5}$ | $45.3_{\pm 2.1}$ | $57.0_{\pm 2.6}$ |
| Depolarizing, moderate | $64.7_{\pm 1.5}$ | $64.3_{\pm 0.6}$ | $45.0_{\pm 8.7}$ |
| Depolarizing, strong | $60.0_{\pm 5.3}$ | $59.3_{\pm 3.8}$ | $50.0_{\pm 9.5}$ |
| Phase, moderate | $67.0_{\pm 4.4}$ | $66.0_{\pm 5.2}$ | $57.0_{\pm 2.6}$ |
| Phase, strong | $63.7_{\pm 3.8}$ | $63.0_{\pm 3.6}$ | $57.0_{\pm 2.6}$ |
| Mixed, moderate | $65.3_{\pm 4.0}$ | $64.0_{\pm 3.6}$ | $57.0_{\pm 2.6}$ |
| Mixed, strong | $65.3_{\pm 6.4}$ | $65.3_{\pm 6.4}$ | $57.7_{\pm 2.1}$ |

Table 78: **Q-SPHERE: OOD reversed diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $7.3_{\pm 1.5}$ | $7.3_{\pm 1.5}$ | $7.3_{\pm 1.5}$ | $92.7_{\pm 1.5}$ | $92.7_{\pm 1.5}$ | $92.0_{\pm 1.7}$ |
| Amplitude, moderate | $7.3_{\pm 1.5}$ | $7.3_{\pm 1.5}$ | $7.3_{\pm 1.5}$ | $92.3_{\pm 1.2}$ | $92.7_{\pm 1.5}$ | $92.7_{\pm 1.5}$ |
| Amplitude, strong | $7.3_{\pm 1.5}$ | $7.3_{\pm 1.5}$ | $7.7_{\pm 1.5}$ | $92.3_{\pm 1.2}$ | $92.3_{\pm 1.2}$ | $92.7_{\pm 1.5}$ |
| Depolarizing, moderate | $7.3_{\pm 1.5}$ | $7.3_{\pm 1.5}$ | $8.0_{\pm 1.0}$ | $92.7_{\pm 1.5}$ | $92.7_{\pm 1.5}$ | $92.7_{\pm 1.5}$ |
| Depolarizing, strong | $7.3_{\pm 1.5}$ | $7.3_{\pm 1.5}$ | $7.3_{\pm 1.5}$ | $92.7_{\pm 1.5}$ | $92.7_{\pm 1.5}$ | $92.3_{\pm 2.1}$ |
| Phase, moderate | $7.3_{\pm 1.5}$ | $7.3_{\pm 1.5}$ | $8.0_{\pm 1.0}$ | $92.7_{\pm 1.5}$ | $92.7_{\pm 1.5}$ | $92.7_{\pm 1.5}$ |
| Phase, strong | $7.3_{\pm 1.5}$ | $7.3_{\pm 1.5}$ | $7.7_{\pm 1.5}$ | $92.7_{\pm 1.5}$ | $92.7_{\pm 1.5}$ | $92.7_{\pm 1.5}$ |
| Mixed, moderate | $7.3_{\pm 1.5}$ | $7.3_{\pm 1.5}$ | $7.3_{\pm 1.5}$ | $92.7_{\pm 1.5}$ | $92.7_{\pm 1.5}$ | $92.7_{\pm 1.5}$ |
| Mixed, strong | $7.3_{\pm 1.5}$ | $7.3_{\pm 1.5}$ | $7.3_{\pm 1.5}$ | $92.7_{\pm 1.5}$ | $92.7_{\pm 1.5}$ | $92.7_{\pm 1.5}$ |
| **Entangling VQC** | | | | | | |
| Clean | $9.0_{\pm 1.7}$ | $9.3_{\pm 0.6}$ | $53.0_{\pm 8.2}$ | $90.3_{\pm 3.2}$ | $90.0_{\pm 3.6}$ | $53.0_{\pm 8.2}$ |
| Amplitude, moderate | $9.7_{\pm 0.6}$ | $9.7_{\pm 0.6}$ | $53.0_{\pm 8.2}$ | $82.7_{\pm 6.4}$ | $83.3_{\pm 4.9}$ | $53.0_{\pm 8.2}$ |
| Amplitude, strong | $12.0_{\pm 4.4}$ | $11.3_{\pm 4.0}$ | $47.3_{\pm 6.7}$ | $72.7_{\pm 8.3}$ | $73.7_{\pm 7.2}$ | $44.0_{\pm 11.0}$ |
| Depolarizing, moderate | $10.3_{\pm 0.6}$ | $10.3_{\pm 0.6}$ | $53.0_{\pm 8.2}$ | $86.7_{\pm 2.5}$ | $88.0_{\pm 2.6}$ | $53.0_{\pm 8.2}$ |
| Depolarizing, strong | $10.7_{\pm 1.5}$ | $10.7_{\pm 1.5}$ | $53.0_{\pm 8.2}$ | $86.7_{\pm 1.5}$ | $87.0_{\pm 1.7}$ | $53.0_{\pm 8.2}$ |
| Phase, moderate | $8.3_{\pm 2.9}$ | $8.0_{\pm 2.6}$ | $53.0_{\pm 8.2}$ | $90.7_{\pm 3.8}$ | $90.3_{\pm 4.2}$ | $53.0_{\pm 8.2}$ |
| Phase, strong | $9.3_{\pm 1.2}$ | $9.0_{\pm 1.0}$ | $53.0_{\pm 8.2}$ | $90.0_{\pm 3.5}$ | $89.3_{\pm 4.2}$ | $53.0_{\pm 8.2}$ |
| Mixed, moderate | $8.3_{\pm 1.2}$ | $8.0_{\pm 1.7}$ | $53.0_{\pm 8.2}$ | $90.0_{\pm 4.4}$ | $88.0_{\pm 5.2}$ | $53.0_{\pm 8.2}$ |
| Mixed, strong | $8.0_{\pm 1.0}$ | $8.0_{\pm 1.0}$ | $43.0_{\pm 2.6}$ | $90.7_{\pm 2.1}$ | $90.3_{\pm 1.5}$ | $43.0_{\pm 2.6}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $24.3_{\pm 9.0}$ | $27.0_{\pm 11.1}$ | $57.0_{\pm 2.6}$ | $85.0_{\pm 2.6}$ | $84.7_{\pm 3.2}$ | $57.0_{\pm 2.6}$ |
| Amplitude, moderate | $13.3_{\pm 0.6}$ | $13.3_{\pm 1.2}$ | $57.0_{\pm 2.6}$ | $84.3_{\pm 1.2}$ | $83.7_{\pm 2.1}$ | $57.0_{\pm 2.6}$ |
| Amplitude, strong | $19.0_{\pm 4.4}$ | $19.3_{\pm 3.8}$ | $57.0_{\pm 2.6}$ | $74.7_{\pm 3.2}$ | $75.3_{\pm 2.9}$ | $57.0_{\pm 2.6}$ |
| Depolarizing, moderate | $37.7_{\pm 26.6}$ | $37.7_{\pm 27.6}$ | $48.7_{\pm 7.8}$ | $83.7_{\pm 2.9}$ | $81.3_{\pm 4.0}$ | $47.3_{\pm 7.5}$ |
| Depolarizing, strong | $42.0_{\pm 20.9}$ | $41.0_{\pm 20.2}$ | $50.3_{\pm 9.0}$ | $64.3_{\pm 3.2}$ | $64.0_{\pm 6.1}$ | $56.3_{\pm 1.5}$ |
| Phase, moderate | $19.3_{\pm 5.0}$ | $20.3_{\pm 8.5}$ | $57.0_{\pm 2.6}$ | $76.7_{\pm 2.1}$ | $77.0_{\pm 3.6}$ | $57.0_{\pm 2.6}$ |
| Phase, strong | $23.0_{\pm 14.0}$ | $23.3_{\pm 14.5}$ | $57.0_{\pm 2.6}$ | $75.0_{\pm 1.0}$ | $73.7_{\pm 1.5}$ | $57.0_{\pm 2.6}$ |
| Mixed, moderate | $17.7_{\pm 4.6}$ | $17.0_{\pm 2.6}$ | $57.0_{\pm 2.6}$ | $79.3_{\pm 4.0}$ | $81.0_{\pm 3.0}$ | $57.0_{\pm 2.6}$ |
| Mixed, strong | $16.0_{\pm 2.6}$ | $20.0_{\pm 7.9}$ | $55.7_{\pm 3.8}$ | $81.0_{\pm 3.6}$ | $81.7_{\pm 4.7}$ | $59.0_{\pm 2.6}$ |

Table 79: **Spirals-Signatures: ID diagnostics.** Full competing-feature task and stable-only accuracy (%), mean ± seed SD.

| Training | Full competing-feature task | | | Stable-only | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $53.3_{\pm 7.8}$ | $52.7_{\pm 3.8}$ | $57.0_{\pm 5.6}$ |
| Amplitude, moderate | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $49.7_{\pm 2.1}$ | $49.7_{\pm 2.1}$ | $51.7_{\pm 3.2}$ |
| Amplitude, strong | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $49.3_{\pm 1.5}$ | $49.7_{\pm 2.1}$ | $51.7_{\pm 1.5}$ |
| Depolarizing, moderate | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $90.7_{\pm 1.2}$ | $52.7_{\pm 2.9}$ | $51.7_{\pm 3.1}$ | $53.0_{\pm 3.6}$ |
| Depolarizing, strong | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $75.3_{\pm 4.0}$ | $50.3_{\pm 1.5}$ | $51.0_{\pm 2.0}$ | $51.0_{\pm 1.7}$ |
| Phase, moderate | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $51.7_{\pm 8.6}$ | $49.3_{\pm 4.9}$ | $49.3_{\pm 4.6}$ |
| Phase, strong | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $53.7_{\pm 4.6}$ | $49.7_{\pm 2.5}$ | $53.0_{\pm 1.0}$ |
| Mixed, moderate | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $49.7_{\pm 2.1}$ | $50.0_{\pm 1.7}$ | $49.3_{\pm 2.5}$ |
| Mixed, strong | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $49.0_{\pm 1.7}$ | $51.7_{\pm 4.0}$ | $52.0_{\pm 1.0}$ |
| **Entangling VQC** | | | | | | |
| Clean | $89.7_{\pm 4.0}$ | $88.7_{\pm 5.0}$ | $51.0_{\pm 1.7}$ | $45.7_{\pm 5.9}$ | $45.3_{\pm 4.6}$ | $51.0_{\pm 1.7}$ |
| Amplitude, moderate | $74.7_{\pm 3.1}$ | $74.7_{\pm 3.5}$ | $51.0_{\pm 1.7}$ | $49.0_{\pm 1.7}$ | $49.0_{\pm 1.7}$ | $51.0_{\pm 1.7}$ |
| Amplitude, strong | $66.7_{\pm 4.0}$ | $65.3_{\pm 5.7}$ | $51.0_{\pm 1.7}$ | $49.0_{\pm 1.7}$ | $49.0_{\pm 1.7}$ | $51.0_{\pm 1.7}$ |
| Depolarizing, moderate | $89.3_{\pm 2.5}$ | $89.0_{\pm 4.0}$ | $51.0_{\pm 1.7}$ | $50.0_{\pm 6.2}$ | $53.3_{\pm 6.8}$ | $52.3_{\pm 3.1}$ |
| Depolarizing, strong | $86.7_{\pm 4.5}$ | $87.0_{\pm 4.0}$ | $52.0_{\pm 2.6}$ | $45.3_{\pm 5.0}$ | $48.3_{\pm 4.0}$ | $51.0_{\pm 1.7}$ |
| Phase, moderate | $90.0_{\pm 4.6}$ | $89.3_{\pm 5.7}$ | $51.0_{\pm 1.7}$ | $44.7_{\pm 3.5}$ | $46.0_{\pm 2.6}$ | $51.0_{\pm 1.7}$ |
| Phase, strong | $91.7_{\pm 1.5}$ | $91.0_{\pm 1.0}$ | $51.0_{\pm 1.7}$ | $51.3_{\pm 4.6}$ | $51.7_{\pm 7.6}$ | $51.0_{\pm 1.7}$ |
| Mixed, moderate | $87.0_{\pm 4.4}$ | $86.3_{\pm 4.0}$ | $51.0_{\pm 1.7}$ | $49.3_{\pm 1.5}$ | $50.3_{\pm 2.1}$ | $51.0_{\pm 1.7}$ |
| Mixed, strong | $86.3_{\pm 4.7}$ | $85.7_{\pm 4.2}$ | $48.7_{\pm 3.5}$ | $50.0_{\pm 2.0}$ | $50.3_{\pm 3.2}$ | $52.0_{\pm 3.0}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $90.0_{\pm 1.7}$ | $89.0_{\pm 2.6}$ | $49.0_{\pm 3.0}$ | $56.7_{\pm 3.5}$ | $54.3_{\pm 5.7}$ | $50.3_{\pm 1.5}$ |
| Amplitude, moderate | $81.7_{\pm 6.7}$ | $79.3_{\pm 3.8}$ | $51.0_{\pm 1.7}$ | $49.0_{\pm 1.7}$ | $49.3_{\pm 2.3}$ | $51.0_{\pm 1.7}$ |
| Amplitude, strong | $79.7_{\pm 2.1}$ | $80.3_{\pm 5.0}$ | $51.0_{\pm 1.7}$ | $49.0_{\pm 1.7}$ | $49.0_{\pm 1.7}$ | $51.0_{\pm 1.7}$ |
| Depolarizing, moderate | $85.3_{\pm 3.5}$ | $85.3_{\pm 3.5}$ | $49.7_{\pm 5.5}$ | $52.7_{\pm 1.2}$ | $53.7_{\pm 1.2}$ | $50.0_{\pm 4.0}$ |
| Depolarizing, strong | $75.3_{\pm 5.5}$ | $74.7_{\pm 6.7}$ | $48.0_{\pm 1.7}$ | $48.0_{\pm 2.0}$ | $48.7_{\pm 0.6}$ | $51.3_{\pm 0.6}$ |
| Phase, moderate | $84.7_{\pm 3.5}$ | $85.3_{\pm 5.5}$ | $50.3_{\pm 2.1}$ | $53.3_{\pm 5.0}$ | $51.3_{\pm 5.1}$ | $50.3_{\pm 2.1}$ |
| Phase, strong | $76.7_{\pm 5.9}$ | $76.7_{\pm 5.8}$ | $50.3_{\pm 2.1}$ | $52.7_{\pm 2.1}$ | $53.0_{\pm 2.6}$ | $50.3_{\pm 2.1}$ |
| Mixed, moderate | $86.7_{\pm 2.3}$ | $86.7_{\pm 1.5}$ | $51.0_{\pm 2.0}$ | $58.7_{\pm 3.2}$ | $50.7_{\pm 3.2}$ | $50.7_{\pm 1.5}$ |
| Mixed, strong | $87.3_{\pm 2.3}$ | $86.3_{\pm 2.5}$ | $51.0_{\pm 1.7}$ | $50.3_{\pm 3.1}$ | $51.3_{\pm 5.9}$ | $51.0_{\pm 1.7}$ |

Table 80: **Spirals-Signatures: ID diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $8.3_{\pm 2.5}$ |
| Amplitude, moderate | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $8.3_{\pm 2.1}$ |
| Amplitude, strong | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ |
| Depolarizing, moderate | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $79.3_{\pm 9.5}$ | $8.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $14.7_{\pm 4.2}$ |
| Depolarizing, strong | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $8.3_{\pm 2.1}$ |
| Phase, moderate | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ |
| Phase, strong | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ |
| Mixed, moderate | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $91.7_{\pm 1.5}$ | $8.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $9.0_{\pm 2.6}$ |
| Mixed, strong | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ | $8.0_{\pm 2.0}$ |
| **Entangling VQC** | | | | | | |
| Clean | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $51.0_{\pm 1.7}$ | $11.0_{\pm 4.4}$ | $10.7_{\pm 4.2}$ | $51.0_{\pm 1.7}$ |
| Amplitude, moderate | $90.3_{\pm 1.5}$ | $90.0_{\pm 2.0}$ | $51.0_{\pm 1.7}$ | $23.3_{\pm 7.6}$ | $23.7_{\pm 7.4}$ | $51.0_{\pm 1.7}$ |
| Amplitude, strong | $91.0_{\pm 1.7}$ | $90.7_{\pm 2.3}$ | $51.0_{\pm 1.7}$ | $28.3_{\pm 6.1}$ | $28.3_{\pm 6.1}$ | $51.0_{\pm 1.7}$ |
| Depolarizing, moderate | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $47.3_{\pm 1.5}$ | $12.3_{\pm 3.8}$ | $12.7_{\pm 3.2}$ | $48.7_{\pm 9.5}$ |
| Depolarizing, strong | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $51.0_{\pm 1.7}$ | $17.7_{\pm 6.7}$ | $17.7_{\pm 5.9}$ | $51.0_{\pm 1.7}$ |
| Phase, moderate | $91.3_{\pm 3.1}$ | $91.3_{\pm 3.1}$ | $51.0_{\pm 1.7}$ | $14.7_{\pm 5.0}$ | $14.3_{\pm 5.1}$ | $51.0_{\pm 1.7}$ |
| Phase, strong | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $51.0_{\pm 1.7}$ | $17.0_{\pm 5.3}$ | $16.0_{\pm 3.6}$ | $51.0_{\pm 1.7}$ |
| Mixed, moderate | $91.3_{\pm 3.1}$ | $91.3_{\pm 3.1}$ | $51.0_{\pm 1.7}$ | $16.0_{\pm 2.6}$ | $16.3_{\pm 4.0}$ | $51.0_{\pm 1.7}$ |
| Mixed, strong | $92.0_{\pm 2.0}$ | $92.0_{\pm 2.0}$ | $47.7_{\pm 5.1}$ | $16.0_{\pm 2.0}$ | $16.0_{\pm 2.0}$ | $49.0_{\pm 3.0}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $91.0_{\pm 2.6}$ | $91.3_{\pm 2.1}$ | $50.3_{\pm 1.5}$ | $15.0_{\pm 2.6}$ | $15.7_{\pm 3.1}$ | $50.0_{\pm 1.7}$ |
| Amplitude, moderate | $70.3_{\pm 9.1}$ | $69.3_{\pm 7.6}$ | $51.0_{\pm 1.7}$ | $22.0_{\pm 5.3}$ | $22.0_{\pm 4.4}$ | $51.0_{\pm 1.7}$ |
| Amplitude, strong | $60.0_{\pm 19.9}$ | $59.7_{\pm 18.5}$ | $51.0_{\pm 1.7}$ | $25.7_{\pm 7.1}$ | $26.7_{\pm 5.9}$ | $51.0_{\pm 1.7}$ |
| Depolarizing, moderate | $87.0_{\pm 5.2}$ | $87.7_{\pm 4.0}$ | $49.3_{\pm 2.5}$ | $18.7_{\pm 1.2}$ | $18.7_{\pm 1.2}$ | $51.0_{\pm 6.6}$ |
| Depolarizing, strong | $91.7_{\pm 2.1}$ | $91.7_{\pm 2.1}$ | $52.3_{\pm 1.5}$ | $25.0_{\pm 4.4}$ | $24.7_{\pm 5.8}$ | $51.7_{\pm 0.6}$ |
| Phase, moderate | $90.3_{\pm 2.1}$ | $90.3_{\pm 2.1}$ | $50.0_{\pm 2.0}$ | $21.0_{\pm 5.3}$ | $20.3_{\pm 4.6}$ | $50.7_{\pm 2.3}$ |
| Phase, strong | $81.0_{\pm 15.6}$ | $80.7_{\pm 15.4}$ | $50.3_{\pm 2.1}$ | $30.3_{\pm 10.6}$ | $30.7_{\pm 9.5}$ | $50.7_{\pm 2.5}$ |
| Mixed, moderate | $87.7_{\pm 6.7}$ | $86.7_{\pm 8.5}$ | $51.3_{\pm 2.1}$ | $15.3_{\pm 5.1}$ | $16.0_{\pm 3.6}$ | $50.3_{\pm 1.2}$ |
| Mixed, strong | $89.7_{\pm 0.6}$ | $89.7_{\pm 1.5}$ | $51.0_{\pm 1.7}$ | $16.0_{\pm 1.0}$ | $15.7_{\pm 1.2}$ | $51.0_{\pm 1.7}$ |

Table 81: **Spirals-Signatures: Random diagnostics.** Full competing-feature task and stable-only accuracy (%), mean ± seed SD.

| Training | Full competing-feature task | | | Stable-only | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $47.3_{\pm 4.5}$ | $58.3_{\pm 10.3}$ | $54.7_{\pm 10.7}$ | $58.7_{\pm 4.0}$ |
| Amplitude, moderate | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $47.0_{\pm 8.5}$ | $47.0_{\pm 8.5}$ | $55.7_{\pm 5.0}$ |
| Amplitude, strong | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $48.7_{\pm 11.2}$ | $49.0_{\pm 11.8}$ | $52.7_{\pm 2.9}$ |
| Depolarizing, moderate | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $47.3_{\pm 4.5}$ | $56.7_{\pm 8.4}$ | $58.0_{\pm 9.2}$ | $54.3_{\pm 7.6}$ |
| Depolarizing, strong | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $47.0_{\pm 4.6}$ | $56.3_{\pm 7.4}$ | $55.0_{\pm 5.6}$ | $54.3_{\pm 7.6}$ |
| Phase, moderate | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $52.7_{\pm 7.0}$ | $55.0_{\pm 8.7}$ | $55.3_{\pm 6.7}$ |
| Phase, strong | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $54.0_{\pm 4.6}$ | $53.7_{\pm 5.5}$ | $54.0_{\pm 9.8}$ |
| Mixed, moderate | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $46.3_{\pm 4.2}$ | $47.0_{\pm 8.5}$ | $47.3_{\pm 8.1}$ | $53.3_{\pm 6.4}$ |
| Mixed, strong | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $46.3_{\pm 4.2}$ | $49.3_{\pm 10.5}$ | $50.0_{\pm 10.5}$ | $50.3_{\pm 5.8}$ |
| **Entangling VQC** | | | | | | |
| Clean | $46.7_{\pm 5.5}$ | $47.7_{\pm 6.4}$ | $54.3_{\pm 7.6}$ | $41.7_{\pm 7.4}$ | $44.0_{\pm 6.2}$ | $54.3_{\pm 7.6}$ |
| Amplitude, moderate | $46.7_{\pm 4.6}$ | $46.7_{\pm 4.7}$ | $54.3_{\pm 7.6}$ | $45.7_{\pm 7.6}$ | $45.7_{\pm 7.6}$ | $54.3_{\pm 7.6}$ |
| Amplitude, strong | $43.0_{\pm 4.4}$ | $43.7_{\pm 3.8}$ | $54.3_{\pm 7.6}$ | $45.7_{\pm 7.6}$ | $45.7_{\pm 7.6}$ | $54.3_{\pm 7.6}$ |
| Depolarizing, moderate | $48.3_{\pm 5.9}$ | $48.3_{\pm 5.1}$ | $54.3_{\pm 7.6}$ | $44.0_{\pm 4.4}$ | $45.0_{\pm 5.3}$ | $54.3_{\pm 7.6}$ |
| Depolarizing, strong | $49.0_{\pm 2.6}$ | $48.3_{\pm 2.5}$ | $54.3_{\pm 7.6}$ | $46.0_{\pm 2.0}$ | $46.3_{\pm 3.2}$ | $54.3_{\pm 7.6}$ |
| Phase, moderate | $48.3_{\pm 4.0}$ | $48.3_{\pm 4.2}$ | $54.3_{\pm 7.6}$ | $44.0_{\pm 4.4}$ | $45.0_{\pm 5.6}$ | $54.3_{\pm 7.6}$ |
| Phase, strong | $49.0_{\pm 3.5}$ | $50.0_{\pm 4.0}$ | $54.3_{\pm 7.6}$ | $49.0_{\pm 4.4}$ | $49.0_{\pm 4.6}$ | $54.3_{\pm 7.6}$ |
| Mixed, moderate | $48.3_{\pm 5.5}$ | $47.7_{\pm 5.5}$ | $54.3_{\pm 7.6}$ | $47.0_{\pm 7.5}$ | $46.3_{\pm 7.5}$ | $54.7_{\pm 8.1}$ |
| Mixed, strong | $46.7_{\pm 5.5}$ | $46.3_{\pm 4.0}$ | $52.7_{\pm 5.8}$ | $46.3_{\pm 8.1}$ | $46.3_{\pm 8.1}$ | $51.0_{\pm 5.0}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $49.0_{\pm 6.0}$ | $49.7_{\pm 6.1}$ | $52.7_{\pm 5.8}$ | $59.0_{\pm 4.6}$ | $58.7_{\pm 4.0}$ | $54.0_{\pm 7.2}$ |
| Amplitude, moderate | $43.3_{\pm 4.7}$ | $45.0_{\pm 4.0}$ | $54.3_{\pm 7.6}$ | $46.0_{\pm 8.2}$ | $46.0_{\pm 8.2}$ | $54.3_{\pm 7.6}$ |
| Amplitude, strong | $46.0_{\pm 1.0}$ | $46.0_{\pm 1.0}$ | $54.3_{\pm 7.6}$ | $45.7_{\pm 7.6}$ | $45.7_{\pm 7.6}$ | $54.3_{\pm 7.6}$ |
| Depolarizing, moderate | $47.3_{\pm 6.0}$ | $48.3_{\pm 4.5}$ | $57.0_{\pm 3.0}$ | $53.0_{\pm 5.2}$ | $55.3_{\pm 8.6}$ | $57.0_{\pm 3.6}$ |
| Depolarizing, strong | $46.3_{\pm 6.4}$ | $47.3_{\pm 5.7}$ | $57.0_{\pm 3.6}$ | $45.3_{\pm 5.5}$ | $44.7_{\pm 4.5}$ | $57.0_{\pm 3.6}$ |
| Phase, moderate | $49.3_{\pm 4.7}$ | $49.3_{\pm 4.7}$ | $52.7_{\pm 8.5}$ | $47.0_{\pm 11.5}$ | $47.3_{\pm 9.1}$ | $53.0_{\pm 8.5}$ |
| Phase, strong | $53.0_{\pm 4.6}$ | $52.3_{\pm 4.7}$ | $53.0_{\pm 8.5}$ | $55.3_{\pm 2.3}$ | $55.7_{\pm 5.8}$ | $52.3_{\pm 7.6}$ |
| Mixed, moderate | $47.0_{\pm 6.9}$ | $46.7_{\pm 7.2}$ | $54.0_{\pm 7.5}$ | $51.3_{\pm 6.5}$ | $50.0_{\pm 6.0}$ | $54.3_{\pm 7.6}$ |
| Mixed, strong | $47.0_{\pm 3.6}$ | $46.7_{\pm 2.1}$ | $54.3_{\pm 7.6}$ | $47.3_{\pm 7.6}$ | $45.0_{\pm 10.4}$ | $54.3_{\pm 7.6}$ |

Table 82: **Spirals-Signatures: Random diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ |
| Amplitude, moderate | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ |
| Amplitude, strong | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $47.0_{\pm 4.6}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ |
| Depolarizing, moderate | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $47.0_{\pm 4.6}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ |
| Depolarizing, strong | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $46.3_{\pm 4.2}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ |
| Phase, moderate | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ |
| Phase, strong | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ |
| Mixed, moderate | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $46.0_{\pm 3.6}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ | $53.7_{\pm 5.1}$ |
| Mixed, strong | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $46.3_{\pm 4.2}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ | $53.3_{\pm 4.7}$ |
| **Entangling VQC** | | | | | | |
| Clean | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $54.3_{\pm 7.6}$ | $52.0_{\pm 4.6}$ | $52.0_{\pm 5.0}$ | $54.3_{\pm 7.6}$ |
| Amplitude, moderate | $46.3_{\pm 5.5}$ | $45.7_{\pm 5.7}$ | $54.3_{\pm 7.6}$ | $51.3_{\pm 5.1}$ | $51.3_{\pm 6.7}$ | $54.3_{\pm 7.6}$ |
| Amplitude, strong | $47.0_{\pm 4.6}$ | $45.3_{\pm 5.9}$ | $54.3_{\pm 7.6}$ | $46.3_{\pm 6.1}$ | $44.7_{\pm 6.0}$ | $54.3_{\pm 7.6}$ |
| Depolarizing, moderate | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $54.3_{\pm 7.6}$ | $51.3_{\pm 3.5}$ | $51.0_{\pm 4.0}$ | $54.3_{\pm 7.6}$ |
| Depolarizing, strong | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $54.3_{\pm 7.6}$ | $52.7_{\pm 4.5}$ | $52.0_{\pm 4.6}$ | $54.3_{\pm 7.6}$ |
| Phase, moderate | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $54.3_{\pm 7.6}$ | $53.0_{\pm 3.0}$ | $53.3_{\pm 2.1}$ | $54.3_{\pm 7.6}$ |
| Phase, strong | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $54.3_{\pm 7.6}$ | $54.7_{\pm 3.2}$ | $54.3_{\pm 3.1}$ | $54.3_{\pm 7.6}$ |
| Mixed, moderate | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $54.3_{\pm 7.6}$ | $52.7_{\pm 6.0}$ | $53.0_{\pm 5.6}$ | $54.3_{\pm 7.6}$ |
| Mixed, strong | $46.7_{\pm 4.7}$ | $46.7_{\pm 4.7}$ | $53.0_{\pm 6.1}$ | $51.7_{\pm 5.0}$ | $51.7_{\pm 4.7}$ | $50.7_{\pm 5.0}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $46.7_{\pm 4.7}$ | $46.7_{\pm 5.7}$ | $54.7_{\pm 8.1}$ | $51.7_{\pm 3.8}$ | $52.7_{\pm 4.0}$ | $53.7_{\pm 6.8}$ |
| Amplitude, moderate | $42.7_{\pm 5.9}$ | $42.7_{\pm 5.9}$ | $54.3_{\pm 7.6}$ | $53.7_{\pm 3.5}$ | $52.7_{\pm 4.7}$ | $54.3_{\pm 7.6}$ |
| Amplitude, strong | $41.7_{\pm 2.5}$ | $42.0_{\pm 3.0}$ | $54.3_{\pm 7.6}$ | $51.3_{\pm 3.2}$ | $51.0_{\pm 2.6}$ | $54.3_{\pm 7.6}$ |
| Depolarizing, moderate | $46.0_{\pm 3.5}$ | $46.0_{\pm 3.5}$ | $54.0_{\pm 8.2}$ | $52.0_{\pm 2.6}$ | $53.0_{\pm 3.6}$ | $57.0_{\pm 3.6}$ |
| Depolarizing, strong | $47.3_{\pm 4.9}$ | $47.0_{\pm 5.3}$ | $57.0_{\pm 3.6}$ | $52.0_{\pm 2.6}$ | $51.7_{\pm 1.2}$ | $57.0_{\pm 3.6}$ |
| Phase, moderate | $47.3_{\pm 4.9}$ | $47.0_{\pm 4.4}$ | $52.7_{\pm 8.5}$ | $53.0_{\pm 5.6}$ | $53.3_{\pm 6.0}$ | $53.0_{\pm 8.5}$ |
| Phase, strong | $51.0_{\pm 3.6}$ | $50.3_{\pm 2.9}$ | $52.3_{\pm 7.6}$ | $56.3_{\pm 9.5}$ | $56.0_{\pm 10.5}$ | $52.7_{\pm 8.1}$ |
| Mixed, moderate | $45.0_{\pm 7.5}$ | $45.0_{\pm 8.5}$ | $54.3_{\pm 7.6}$ | $53.7_{\pm 4.0}$ | $52.7_{\pm 5.8}$ | $54.7_{\pm 7.8}$ |
| Mixed, strong | $46.3_{\pm 6.1}$ | $46.0_{\pm 5.0}$ | $54.3_{\pm 7.6}$ | $51.0_{\pm 4.6}$ | $50.7_{\pm 4.5}$ | $54.3_{\pm 7.6}$ |

Table 83: **Spirals-Signatures: OOD reversed diagnostics.** Stable-only accuracy (%), mean ± seed SD.

| Training | Stable-only | | |
|---|---|---|---|
| | A | F | H |
| **Local VQC** | | | |
| Clean | $50.3_{\pm 1.5}$ | $48.7_{\pm 1.5}$ | $51.3_{\pm 4.0}$ |
| Amplitude, moderate | $46.7_{\pm 5.5}$ | $46.7_{\pm 5.5}$ | $49.3_{\pm 6.0}$ |
| Amplitude, strong | $48.0_{\pm 4.6}$ | $47.7_{\pm 4.7}$ | $47.0_{\pm 2.0}$ |
| Depolarizing, moderate | $50.3_{\pm 3.5}$ | $52.0_{\pm 1.7}$ | $46.3_{\pm 1.2}$ |
| Depolarizing, strong | $48.3_{\pm 3.8}$ | $50.7_{\pm 3.5}$ | $44.7_{\pm 2.1}$ |
| Phase, moderate | $49.3_{\pm 5.1}$ | $53.0_{\pm 7.0}$ | $48.0_{\pm 1.0}$ |
| Phase, strong | $52.7_{\pm 7.2}$ | $57.7_{\pm 8.1}$ | $47.3_{\pm 2.5}$ |
| Mixed, moderate | $46.7_{\pm 5.5}$ | $46.7_{\pm 5.5}$ | $45.7_{\pm 3.1}$ |
| Mixed, strong | $55.3_{\pm 7.8}$ | $53.7_{\pm 7.0}$ | $48.3_{\pm 8.4}$ |
| **Entangling VQC** | | | |
| Clean | $49.7_{\pm 2.9}$ | $50.7_{\pm 2.1}$ | $44.7_{\pm 2.1}$ |
| Amplitude, moderate | $55.3_{\pm 2.1}$ | $55.3_{\pm 2.1}$ | $44.7_{\pm 2.1}$ |
| Amplitude, strong | $55.3_{\pm 2.1}$ | $55.3_{\pm 2.1}$ | $44.7_{\pm 2.1}$ |
| Depolarizing, moderate | $53.3_{\pm 3.5}$ | $53.3_{\pm 3.5}$ | $44.7_{\pm 2.1}$ |
| Depolarizing, strong | $42.3_{\pm 8.5}$ | $41.7_{\pm 9.0}$ | $44.7_{\pm 2.1}$ |
| Phase, moderate | $52.0_{\pm 1.7}$ | $52.0_{\pm 1.0}$ | $44.7_{\pm 2.1}$ |
| Phase, strong | $60.0_{\pm 5.3}$ | $57.7_{\pm 4.6}$ | $44.7_{\pm 2.1}$ |
| Mixed, moderate | $56.0_{\pm 3.0}$ | $55.3_{\pm 2.1}$ | $44.7_{\pm 2.1}$ |
| Mixed, strong | $55.3_{\pm 2.1}$ | $55.0_{\pm 2.0}$ | $44.7_{\pm 2.1}$ |
| **Reuploading VQC** | | | |
| Clean | $52.7_{\pm 7.2}$ | $52.3_{\pm 6.8}$ | $44.7_{\pm 2.1}$ |
| Amplitude, moderate | $56.0_{\pm 2.6}$ | $57.0_{\pm 4.0}$ | $44.7_{\pm 2.1}$ |
| Amplitude, strong | $55.3_{\pm 2.1}$ | $55.3_{\pm 2.1}$ | $44.7_{\pm 2.1}$ |
| Depolarizing, moderate | $48.0_{\pm 7.9}$ | $50.0_{\pm 7.2}$ | $49.0_{\pm 7.2}$ |
| Depolarizing, strong | $52.7_{\pm 10.1}$ | $51.3_{\pm 8.1}$ | $48.3_{\pm 7.1}$ |
| Phase, moderate | $56.7_{\pm 4.9}$ | $55.7_{\pm 2.5}$ | $53.3_{\pm 5.5}$ |
| Phase, strong | $52.0_{\pm 2.6}$ | $54.3_{\pm 3.5}$ | $53.3_{\pm 5.5}$ |
| Mixed, moderate | $56.3_{\pm 3.5}$ | $53.3_{\pm 1.5}$ | $44.3_{\pm 2.5}$ |
| Mixed, strong | $54.3_{\pm 1.2}$ | $54.0_{\pm 2.0}$ | $44.7_{\pm 2.1}$ |

Table 84: **Spirals-Signatures: OOD reversed diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ |
| Amplitude, moderate | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ |
| Amplitude, strong | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ |
| Depolarizing, moderate | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ |
| Depolarizing, strong | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ |
| Phase, moderate | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ |
| Phase, strong | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ |
| Mixed, moderate | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $9.7_{\pm 2.1}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ |
| Mixed, strong | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ | $90.7_{\pm 1.2}$ |
| **Entangling VQC** | | | | | | |
| Clean | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $44.7_{\pm 2.1}$ | $89.3_{\pm 2.1}$ | $88.7_{\pm 3.2}$ | $44.7_{\pm 2.1}$ |
| Amplitude, moderate | $10.7_{\pm 1.5}$ | $10.0_{\pm 1.0}$ | $44.7_{\pm 2.1}$ | $75.0_{\pm 2.6}$ | $76.3_{\pm 2.5}$ | $44.7_{\pm 2.1}$ |
| Amplitude, strong | $10.7_{\pm 1.5}$ | $11.0_{\pm 2.0}$ | $44.7_{\pm 2.1}$ | $68.3_{\pm 4.6}$ | $68.3_{\pm 5.5}$ | $44.7_{\pm 2.1}$ |
| Depolarizing, moderate | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $44.7_{\pm 2.1}$ | $87.7_{\pm 2.1}$ | $87.3_{\pm 1.5}$ | $44.7_{\pm 2.1}$ |
| Depolarizing, strong | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $44.7_{\pm 2.1}$ | $86.3_{\pm 4.0}$ | $86.3_{\pm 4.0}$ | $44.7_{\pm 2.1}$ |
| Phase, moderate | $9.0_{\pm 1.0}$ | $9.3_{\pm 0.6}$ | $44.7_{\pm 2.1}$ | $89.0_{\pm 1.7}$ | $88.7_{\pm 1.5}$ | $44.7_{\pm 2.1}$ |
| Phase, strong | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $44.7_{\pm 2.1}$ | $89.0_{\pm 1.0}$ | $88.7_{\pm 0.6}$ | $44.7_{\pm 2.1}$ |
| Mixed, moderate | $9.3_{\pm 0.6}$ | $9.3_{\pm 0.6}$ | $44.7_{\pm 2.1}$ | $87.7_{\pm 2.1}$ | $87.0_{\pm 1.0}$ | $44.7_{\pm 2.1}$ |
| Mixed, strong | $9.0_{\pm 1.0}$ | $9.0_{\pm 1.0}$ | $45.3_{\pm 3.2}$ | $85.0_{\pm 2.0}$ | $84.3_{\pm 2.5}$ | $46.3_{\pm 4.9}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $9.3_{\pm 0.6}$ | $9.3_{\pm 0.6}$ | $44.0_{\pm 1.0}$ | $91.3_{\pm 0.6}$ | $92.0_{\pm 1.0}$ | $45.3_{\pm 3.2}$ |
| Amplitude, moderate | $25.0_{\pm 9.5}$ | $24.3_{\pm 10.5}$ | $44.7_{\pm 2.1}$ | $84.0_{\pm 4.4}$ | $84.3_{\pm 6.7}$ | $44.7_{\pm 2.1}$ |
| Amplitude, strong | $40.3_{\pm 22.8}$ | $39.7_{\pm 22.2}$ | $44.7_{\pm 2.1}$ | $83.7_{\pm 4.7}$ | $83.7_{\pm 4.7}$ | $44.7_{\pm 2.1}$ |
| Depolarizing, moderate | $11.3_{\pm 4.2}$ | $10.7_{\pm 3.8}$ | $48.3_{\pm 7.1}$ | $86.0_{\pm 5.0}$ | $85.7_{\pm 5.1}$ | $47.3_{\pm 4.5}$ |
| Depolarizing, strong | $9.3_{\pm 1.5}$ | $9.3_{\pm 1.5}$ | $48.7_{\pm 6.7}$ | $76.3_{\pm 1.5}$ | $77.3_{\pm 1.5}$ | $48.3_{\pm 7.1}$ |
| Phase, moderate | $9.0_{\pm 1.0}$ | $9.3_{\pm 1.2}$ | $52.0_{\pm 5.0}$ | $85.7_{\pm 0.6}$ | $86.7_{\pm 1.5}$ | $53.7_{\pm 5.8}$ |
| Phase, strong | $18.0_{\pm 16.5}$ | $18.0_{\pm 16.5}$ | $53.3_{\pm 5.5}$ | $81.7_{\pm 5.9}$ | $81.0_{\pm 6.2}$ | $53.3_{\pm 5.5}$ |
| Mixed, moderate | $12.3_{\pm 5.8}$ | $12.3_{\pm 5.8}$ | $44.7_{\pm 2.1}$ | $85.3_{\pm 3.8}$ | $85.7_{\pm 3.2}$ | $44.3_{\pm 1.5}$ |
| Mixed, strong | $10.7_{\pm 1.2}$ | $10.7_{\pm 1.2}$ | $44.7_{\pm 2.1}$ | $87.3_{\pm 3.1}$ | $87.0_{\pm 3.6}$ | $44.7_{\pm 2.1}$ |

Table 85: **Spurious-Fourier: ID diagnostics.** Full competing-feature task and stable-only accuracy (%), mean ± seed SD.

| Training | Full competing-feature task | | | Stable-only | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $95.3_{\pm 2.1}$ | $95.0_{\pm 2.6}$ | $85.3_{\pm 7.8}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| Amplitude, moderate | $95.3_{\pm 2.1}$ | $95.3_{\pm 2.5}$ | $89.7_{\pm 3.5}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $99.3_{\pm 1.2}$ |
| Amplitude, strong | $94.7_{\pm 2.1}$ | $94.7_{\pm 2.1}$ | $92.0_{\pm 4.4}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| Depolarizing, moderate | $95.7_{\pm 1.5}$ | $95.0_{\pm 1.0}$ | $10.0_{\pm 7.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $3.3_{\pm 3.5}$ |
| Depolarizing, strong | $95.7_{\pm 0.6}$ | $96.0_{\pm 1.0}$ | $78.7_{\pm 8.4}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $93.0_{\pm 12.1}$ |
| Phase, moderate | $96.0_{\pm 1.0}$ | $96.0_{\pm 1.0}$ | $91.3_{\pm 3.5}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| Phase, strong | $95.3_{\pm 1.2}$ | $95.7_{\pm 1.5}$ | $94.3_{\pm 1.5}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| Mixed, moderate | $96.0_{\pm 1.7}$ | $97.0_{\pm 1.0}$ | $94.0_{\pm 2.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| Mixed, strong | $95.0_{\pm 1.0}$ | $95.7_{\pm 2.1}$ | $87.7_{\pm 5.5}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $86.7_{\pm 11.6}$ |
| **Entangling VQC** | | | | | | |
| Clean | $86.7_{\pm 1.5}$ | $88.7_{\pm 0.6}$ | $49.3_{\pm 5.8}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $49.3_{\pm 5.8}$ |
| Amplitude, moderate | $78.3_{\pm 3.1}$ | $79.0_{\pm 2.6}$ | $49.0_{\pm 6.9}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $49.0_{\pm 4.4}$ |
| Amplitude, strong | $72.0_{\pm 9.6}$ | $71.7_{\pm 9.5}$ | $50.0_{\pm 5.2}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $49.7_{\pm 4.7}$ |
| Depolarizing, moderate | $81.0_{\pm 5.0}$ | $81.0_{\pm 6.2}$ | $49.0_{\pm 5.2}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $46.3_{\pm 5.5}$ |
| Depolarizing, strong | $95.3_{\pm 0.6}$ | $95.0_{\pm 1.0}$ | $47.0_{\pm 1.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $48.7_{\pm 4.6}$ |
| Phase, moderate | $87.3_{\pm 2.5}$ | $87.3_{\pm 4.5}$ | $49.3_{\pm 5.8}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $49.3_{\pm 5.8}$ |
| Phase, strong | $87.0_{\pm 1.7}$ | $86.0_{\pm 1.7}$ | $49.3_{\pm 5.8}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $49.3_{\pm 5.8}$ |
| Mixed, moderate | $84.0_{\pm 3.6}$ | $83.7_{\pm 5.1}$ | $49.3_{\pm 5.8}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $49.3_{\pm 5.8}$ |
| Mixed, strong | $80.3_{\pm 6.5}$ | $79.3_{\pm 7.1}$ | $47.7_{\pm 4.6}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $48.3_{\pm 4.9}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $54.7_{\pm 9.7}$ | $56.3_{\pm 7.6}$ | $47.3_{\pm 6.1}$ | $0.0_{\pm 0.0}$ | $0.7_{\pm 1.2}$ | $48.0_{\pm 5.3}$ |
| Amplitude, moderate | $50.7_{\pm 11.0}$ | $51.7_{\pm 9.9}$ | $52.0_{\pm 5.3}$ | $30.0_{\pm 26.0}$ | $25.7_{\pm 17.6}$ | $52.0_{\pm 5.3}$ |
| Amplitude, strong | $53.0_{\pm 6.2}$ | $55.0_{\pm 4.6}$ | $49.3_{\pm 5.8}$ | $66.0_{\pm 29.9}$ | $66.3_{\pm 29.7}$ | $49.3_{\pm 5.8}$ |
| Depolarizing, moderate | $49.7_{\pm 7.5}$ | $50.3_{\pm 5.5}$ | $48.3_{\pm 4.0}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $51.3_{\pm 7.0}$ |
| Depolarizing, strong | $45.7_{\pm 6.8}$ | $43.7_{\pm 4.2}$ | $45.7_{\pm 1.2}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $47.0_{\pm 3.0}$ |
| Phase, moderate | $50.7_{\pm 5.5}$ | $51.7_{\pm 7.2}$ | $48.0_{\pm 5.3}$ | $100.0_{\pm 0.0}$ | $98.7_{\pm 1.5}$ | $46.7_{\pm 7.0}$ |
| Phase, strong | $58.7_{\pm 8.3}$ | $59.3_{\pm 9.3}$ | $46.0_{\pm 2.0}$ | $100.0_{\pm 0.0}$ | $99.7_{\pm 0.6}$ | $45.0_{\pm 1.0}$ |
| Mixed, moderate | $54.0_{\pm 5.0}$ | $55.7_{\pm 4.2}$ | $56.0_{\pm 1.7}$ | $66.7_{\pm 57.7}$ | $66.7_{\pm 57.7}$ | $53.3_{\pm 2.1}$ |
| Mixed, strong | $48.0_{\pm 3.6}$ | $49.7_{\pm 6.5}$ | $46.3_{\pm 0.6}$ | $100.0_{\pm 0.0}$ | $95.0_{\pm 5.0}$ | $45.3_{\pm 1.2}$ |

Table 86: **Spurious-Fourier: ID diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $51.7_{\pm 6.5}$ | $51.3_{\pm 6.5}$ | $49.3_{\pm 8.1}$ | $94.7_{\pm 2.3}$ | $94.7_{\pm 2.3}$ | $89.7_{\pm 6.5}$ |
| Amplitude, moderate | $50.7_{\pm 5.5}$ | $50.7_{\pm 4.5}$ | $50.0_{\pm 3.0}$ | $95.3_{\pm 2.1}$ | $95.7_{\pm 2.3}$ | $82.7_{\pm 4.7}$ |
| Amplitude, strong | $48.7_{\pm 3.2}$ | $48.7_{\pm 2.3}$ | $49.0_{\pm 6.2}$ | $94.7_{\pm 2.1}$ | $94.7_{\pm 2.1}$ | $93.0_{\pm 3.5}$ |
| Depolarizing, moderate | $52.3_{\pm 7.0}$ | $52.3_{\pm 4.9}$ | $49.3_{\pm 2.5}$ | $95.7_{\pm 1.5}$ | $95.3_{\pm 2.1}$ | $10.3_{\pm 7.5}$ |
| Depolarizing, strong | $53.3_{\pm 5.5}$ | $53.3_{\pm 5.5}$ | $49.3_{\pm 5.1}$ | $96.0_{\pm 1.0}$ | $96.0_{\pm 1.0}$ | $78.7_{\pm 9.5}$ |
| Phase, moderate | $52.7_{\pm 5.0}$ | $52.7_{\pm 5.5}$ | $50.0_{\pm 4.6}$ | $95.7_{\pm 0.6}$ | $95.7_{\pm 1.5}$ | $92.3_{\pm 5.0}$ |
| Phase, strong | $52.7_{\pm 6.0}$ | $52.7_{\pm 6.1}$ | $51.0_{\pm 4.4}$ | $95.0_{\pm 1.7}$ | $95.3_{\pm 1.2}$ | $92.7_{\pm 2.5}$ |
| Mixed, moderate | $53.7_{\pm 2.1}$ | $52.0_{\pm 2.0}$ | $50.0_{\pm 4.4}$ | $95.7_{\pm 0.6}$ | $95.7_{\pm 0.6}$ | $94.3_{\pm 2.1}$ |
| Mixed, strong | $52.7_{\pm 2.9}$ | $53.0_{\pm 3.6}$ | $50.3_{\pm 5.7}$ | $96.0_{\pm 2.6}$ | $96.0_{\pm 2.6}$ | $87.7_{\pm 3.8}$ |
| **Entangling VQC** | | | | | | |
| Clean | $52.0_{\pm 4.6}$ | $52.3_{\pm 5.1}$ | $49.3_{\pm 5.8}$ | $85.3_{\pm 2.5}$ | $85.7_{\pm 3.5}$ | $49.3_{\pm 5.8}$ |
| Amplitude, moderate | $53.0_{\pm 0.0}$ | $52.7_{\pm 1.5}$ | $49.3_{\pm 5.8}$ | $79.3_{\pm 7.4}$ | $78.7_{\pm 8.4}$ | $49.7_{\pm 6.4}$ |
| Amplitude, strong | $53.3_{\pm 0.6}$ | $53.7_{\pm 2.5}$ | $49.0_{\pm 6.2}$ | $66.7_{\pm 8.4}$ | $66.0_{\pm 8.7}$ | $49.0_{\pm 6.1}$ |
| Depolarizing, moderate | $52.7_{\pm 5.7}$ | $52.0_{\pm 4.4}$ | $50.7_{\pm 5.8}$ | $80.3_{\pm 5.0}$ | $79.7_{\pm 5.5}$ | $50.7_{\pm 4.5}$ |
| Depolarizing, strong | $52.3_{\pm 4.2}$ | $52.7_{\pm 5.0}$ | $48.7_{\pm 4.6}$ | $95.3_{\pm 0.6}$ | $95.3_{\pm 0.6}$ | $51.3_{\pm 9.2}$ |
| Phase, moderate | $54.0_{\pm 4.0}$ | $54.3_{\pm 3.5}$ | $49.3_{\pm 5.8}$ | $86.0_{\pm 2.6}$ | $86.3_{\pm 3.2}$ | $49.3_{\pm 5.8}$ |
| Phase, strong | $52.7_{\pm 4.0}$ | $51.3_{\pm 4.2}$ | $49.3_{\pm 5.8}$ | $85.0_{\pm 1.7}$ | $85.3_{\pm 3.5}$ | $49.3_{\pm 5.8}$ |
| Mixed, moderate | $53.7_{\pm 0.6}$ | $55.3_{\pm 2.1}$ | $49.3_{\pm 5.8}$ | $83.0_{\pm 5.2}$ | $83.3_{\pm 4.6}$ | $49.3_{\pm 5.8}$ |
| Mixed, strong | $54.0_{\pm 1.7}$ | $52.7_{\pm 1.2}$ | $51.3_{\pm 6.1}$ | $77.7_{\pm 10.1}$ | $76.0_{\pm 9.8}$ | $49.7_{\pm 3.5}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $51.7_{\pm 8.1}$ | $50.0_{\pm 7.9}$ | $49.0_{\pm 5.6}$ | $53.3_{\pm 13.0}$ | $52.7_{\pm 11.1}$ | $48.3_{\pm 5.9}$ |
| Amplitude, moderate | $59.0_{\pm 9.2}$ | $55.7_{\pm 10.5}$ | $52.0_{\pm 5.3}$ | $49.7_{\pm 5.0}$ | $49.7_{\pm 7.0}$ | $52.0_{\pm 5.3}$ |
| Amplitude, strong | $52.7_{\pm 8.1}$ | $53.0_{\pm 7.2}$ | $49.3_{\pm 5.8}$ | $52.7_{\pm 5.5}$ | $52.3_{\pm 6.5}$ | $49.3_{\pm 5.8}$ |
| Depolarizing, moderate | $57.0_{\pm 5.6}$ | $54.7_{\pm 9.0}$ | $50.7_{\pm 5.8}$ | $42.3_{\pm 5.7}$ | $40.0_{\pm 6.2}$ | $52.0_{\pm 8.5}$ |
| Depolarizing, strong | $57.7_{\pm 6.8}$ | $55.7_{\pm 6.7}$ | $48.7_{\pm 4.0}$ | $47.3_{\pm 12.1}$ | $45.3_{\pm 11.4}$ | $47.7_{\pm 6.4}$ |
| Phase, moderate | $60.7_{\pm 1.5}$ | $57.0_{\pm 2.6}$ | $46.7_{\pm 3.1}$ | $53.7_{\pm 0.6}$ | $51.7_{\pm 2.5}$ | $48.0_{\pm 5.3}$ |
| Phase, strong | $57.0_{\pm 13.0}$ | $53.3_{\pm 12.9}$ | $45.7_{\pm 1.5}$ | $62.7_{\pm 5.8}$ | $60.0_{\pm 6.1}$ | $48.7_{\pm 6.4}$ |
| Mixed, moderate | $55.7_{\pm 13.9}$ | $58.7_{\pm 12.7}$ | $52.0_{\pm 2.6}$ | $55.0_{\pm 4.0}$ | $52.7_{\pm 6.8}$ | $51.0_{\pm 6.1}$ |
| Mixed, strong | $54.0_{\pm 5.6}$ | $56.7_{\pm 5.5}$ | $45.7_{\pm 0.6}$ | $48.7_{\pm 8.4}$ | $49.3_{\pm 8.1}$ | $46.7_{\pm 1.2}$ |

Table 87: **Spurious-Fourier: Random diagnostics.** Full competing-feature task and stable-only accuracy (%), mean ± seed SD.

| Training | Full competing-feature task | | | Stable-only | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $96.7_{\pm 3.2}$ | $96.0_{\pm 2.6}$ | $86.7_{\pm 5.9}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| Amplitude, moderate | $96.0_{\pm 3.6}$ | $96.0_{\pm 3.6}$ | $78.3_{\pm 4.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $99.7_{\pm 0.6}$ |
| Amplitude, strong | $96.0_{\pm 3.0}$ | $96.0_{\pm 2.0}$ | $91.0_{\pm 1.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $99.7_{\pm 0.6}$ |
| Depolarizing, moderate | $97.0_{\pm 1.7}$ | $96.3_{\pm 2.1}$ | $75.7_{\pm 7.2}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| Depolarizing, strong | $97.7_{\pm 1.5}$ | $97.3_{\pm 1.5}$ | $82.3_{\pm 5.5}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $86.7_{\pm 19.6}$ |
| Phase, moderate | $97.3_{\pm 2.1}$ | $97.3_{\pm 2.1}$ | $92.3_{\pm 5.1}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $99.3_{\pm 1.2}$ |
| Phase, strong | $97.7_{\pm 2.5}$ | $96.7_{\pm 1.5}$ | $93.7_{\pm 3.1}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| Mixed, moderate | $97.3_{\pm 2.1}$ | $96.7_{\pm 2.1}$ | $93.7_{\pm 3.5}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| Mixed, strong | $96.3_{\pm 1.5}$ | $96.3_{\pm 1.5}$ | $86.0_{\pm 5.3}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $84.7_{\pm 25.7}$ |
| **Entangling VQC** | | | | | | |
| Clean | $86.7_{\pm 5.1}$ | $86.0_{\pm 3.0}$ | $51.7_{\pm 9.1}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $51.7_{\pm 9.1}$ |
| Amplitude, moderate | $77.3_{\pm 2.5}$ | $77.3_{\pm 1.5}$ | $51.7_{\pm 9.1}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $51.7_{\pm 9.1}$ |
| Amplitude, strong | $66.0_{\pm 1.7}$ | $66.0_{\pm 1.0}$ | $52.0_{\pm 9.2}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $52.0_{\pm 3.0}$ |
| Depolarizing, moderate | $80.3_{\pm 11.0}$ | $79.0_{\pm 8.0}$ | $51.7_{\pm 9.1}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $51.7_{\pm 9.1}$ |
| Depolarizing, strong | $98.3_{\pm 1.5}$ | $98.0_{\pm 2.0}$ | $51.7_{\pm 9.1}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $51.3_{\pm 9.6}$ |
| Phase, moderate | $84.3_{\pm 7.0}$ | $84.7_{\pm 5.5}$ | $51.7_{\pm 9.1}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $51.7_{\pm 9.1}$ |
| Phase, strong | $84.7_{\pm 5.1}$ | $85.0_{\pm 5.3}$ | $51.7_{\pm 9.1}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $51.7_{\pm 9.1}$ |
| Mixed, moderate | $82.0_{\pm 8.9}$ | $82.7_{\pm 8.1}$ | $51.7_{\pm 9.1}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $51.7_{\pm 9.1}$ |
| Mixed, strong | $79.0_{\pm 8.7}$ | $78.3_{\pm 8.4}$ | $52.7_{\pm 8.1}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $52.0_{\pm 9.8}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $64.0_{\pm 11.8}$ | $62.0_{\pm 12.3}$ | $51.0_{\pm 8.2}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $50.7_{\pm 8.1}$ |
| Amplitude, moderate | $60.3_{\pm 4.2}$ | $57.3_{\pm 4.0}$ | $45.0_{\pm 7.0}$ | $39.3_{\pm 34.1}$ | $28.3_{\pm 11.6}$ | $45.0_{\pm 7.0}$ |
| Amplitude, strong | $50.0_{\pm 3.6}$ | $51.0_{\pm 4.0}$ | $51.7_{\pm 9.1}$ | $66.0_{\pm 30.8}$ | $66.7_{\pm 30.0}$ | $51.7_{\pm 9.1}$ |
| Depolarizing, moderate | $51.7_{\pm 4.9}$ | $52.7_{\pm 5.1}$ | $53.7_{\pm 13.1}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $47.3_{\pm 9.3}$ |
| Depolarizing, strong | $49.3_{\pm 7.4}$ | $48.3_{\pm 5.5}$ | $53.3_{\pm 7.2}$ | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $50.7_{\pm 9.5}$ |
| Phase, moderate | $58.3_{\pm 6.8}$ | $54.0_{\pm 3.5}$ | $52.7_{\pm 5.5}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $49.7_{\pm 10.4}$ |
| Phase, strong | $66.7_{\pm 3.5}$ | $65.0_{\pm 2.6}$ | $55.3_{\pm 2.5}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $56.7_{\pm 3.2}$ |
| Mixed, moderate | $61.0_{\pm 5.3}$ | $61.7_{\pm 2.3}$ | $42.3_{\pm 5.9}$ | $66.7_{\pm 57.7}$ | $66.7_{\pm 57.7}$ | $44.7_{\pm 4.2}$ |
| Mixed, strong | $56.3_{\pm 11.5}$ | $56.0_{\pm 7.9}$ | $54.0_{\pm 5.6}$ | $100.0_{\pm 0.0}$ | $96.7_{\pm 5.8}$ | $54.3_{\pm 5.1}$ |

Table 88: **Spurious-Fourier: Random diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $54.0_{\pm 1.7}$ | $54.0_{\pm 1.0}$ | $56.7_{\pm 4.2}$ | $97.0_{\pm 2.6}$ | $97.0_{\pm 2.6}$ | $86.7_{\pm 7.1}$ |
| Amplitude, moderate | $50.7_{\pm 3.2}$ | $52.3_{\pm 3.1}$ | $56.0_{\pm 4.0}$ | $96.0_{\pm 3.6}$ | $95.7_{\pm 3.5}$ | $82.0_{\pm 2.6}$ |
| Amplitude, strong | $50.7_{\pm 5.8}$ | $50.7_{\pm 4.9}$ | $53.7_{\pm 4.2}$ | $95.7_{\pm 3.1}$ | $95.7_{\pm 3.1}$ | $88.3_{\pm 1.5}$ |
| Depolarizing, moderate | $53.7_{\pm 2.3}$ | $53.7_{\pm 1.5}$ | $52.0_{\pm 4.4}$ | $97.0_{\pm 2.6}$ | $97.3_{\pm 3.1}$ | $91.3_{\pm 4.9}$ |
| Depolarizing, strong | $54.0_{\pm 1.0}$ | $54.0_{\pm 2.0}$ | $54.7_{\pm 5.5}$ | $97.3_{\pm 1.2}$ | $97.3_{\pm 1.2}$ | $85.3_{\pm 8.0}$ |
| Phase, moderate | $53.7_{\pm 1.5}$ | $55.3_{\pm 1.5}$ | $55.0_{\pm 5.0}$ | $98.0_{\pm 2.0}$ | $97.3_{\pm 2.1}$ | $92.3_{\pm 4.9}$ |
| Phase, strong | $54.0_{\pm 1.7}$ | $54.7_{\pm 1.2}$ | $55.7_{\pm 5.0}$ | $97.3_{\pm 3.1}$ | $97.0_{\pm 2.6}$ | $93.7_{\pm 3.8}$ |
| Mixed, moderate | $51.3_{\pm 4.0}$ | $50.3_{\pm 3.5}$ | $56.0_{\pm 2.6}$ | $97.0_{\pm 2.0}$ | $97.3_{\pm 1.5}$ | $94.3_{\pm 2.5}$ |
| Mixed, strong | $50.0_{\pm 5.2}$ | $50.3_{\pm 4.7}$ | $53.3_{\pm 4.0}$ | $96.3_{\pm 2.3}$ | $96.3_{\pm 2.3}$ | $84.7_{\pm 6.0}$ |
| **Entangling VQC** | | | | | | |
| Clean | $53.0_{\pm 1.0}$ | $51.3_{\pm 3.1}$ | $51.7_{\pm 9.1}$ | $90.7_{\pm 2.3}$ | $90.3_{\pm 3.1}$ | $51.7_{\pm 9.1}$ |
| Amplitude, moderate | $53.0_{\pm 7.5}$ | $53.3_{\pm 7.5}$ | $51.7_{\pm 9.1}$ | $81.7_{\pm 6.7}$ | $81.3_{\pm 8.1}$ | $52.0_{\pm 8.5}$ |
| Amplitude, strong | $50.3_{\pm 6.1}$ | $50.3_{\pm 7.1}$ | $51.7_{\pm 5.7}$ | $69.0_{\pm 4.6}$ | $68.0_{\pm 4.6}$ | $50.0_{\pm 4.0}$ |
| Depolarizing, moderate | $54.3_{\pm 0.6}$ | $51.3_{\pm 2.5}$ | $51.7_{\pm 9.1}$ | $85.0_{\pm 6.0}$ | $82.3_{\pm 6.4}$ | $51.7_{\pm 9.1}$ |
| Depolarizing, strong | $54.0_{\pm 2.0}$ | $53.7_{\pm 1.5}$ | $53.3_{\pm 6.5}$ | $98.0_{\pm 1.7}$ | $97.7_{\pm 2.1}$ | $54.0_{\pm 5.6}$ |
| Phase, moderate | $50.3_{\pm 2.5}$ | $52.0_{\pm 1.0}$ | $51.7_{\pm 9.1}$ | $87.7_{\pm 4.9}$ | $88.0_{\pm 6.1}$ | $51.7_{\pm 9.1}$ |
| Phase, strong | $53.7_{\pm 2.5}$ | $52.7_{\pm 1.2}$ | $51.7_{\pm 9.1}$ | $87.7_{\pm 6.5}$ | $87.3_{\pm 5.7}$ | $51.7_{\pm 9.1}$ |
| Mixed, moderate | $49.3_{\pm 3.5}$ | $49.7_{\pm 3.2}$ | $51.7_{\pm 9.1}$ | $86.7_{\pm 5.5}$ | $86.7_{\pm 5.5}$ | $51.7_{\pm 9.1}$ |
| Mixed, strong | $52.0_{\pm 3.5}$ | $51.0_{\pm 4.4}$ | $51.3_{\pm 9.6}$ | $84.3_{\pm 8.0}$ | $82.7_{\pm 9.5}$ | $52.3_{\pm 9.3}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $52.0_{\pm 7.2}$ | $50.3_{\pm 5.5}$ | $50.3_{\pm 9.3}$ | $55.7_{\pm 9.5}$ | $56.3_{\pm 9.1}$ | $50.3_{\pm 9.3}$ |
| Amplitude, moderate | $52.3_{\pm 3.5}$ | $53.0_{\pm 5.2}$ | $45.0_{\pm 7.0}$ | $63.0_{\pm 5.3}$ | $61.7_{\pm 5.5}$ | $45.0_{\pm 7.0}$ |
| Amplitude, strong | $50.0_{\pm 7.2}$ | $49.7_{\pm 8.5}$ | $51.7_{\pm 9.1}$ | $49.3_{\pm 9.1}$ | $50.3_{\pm 8.6}$ | $51.7_{\pm 9.1}$ |
| Depolarizing, moderate | $47.7_{\pm 4.5}$ | $47.0_{\pm 2.0}$ | $49.0_{\pm 10.1}$ | $52.0_{\pm 3.5}$ | $50.3_{\pm 3.1}$ | $47.7_{\pm 9.3}$ |
| Depolarizing, strong | $54.7_{\pm 4.7}$ | $54.7_{\pm 4.0}$ | $48.0_{\pm 9.2}$ | $45.0_{\pm 1.7}$ | $47.0_{\pm 1.7}$ | $49.3_{\pm 9.6}$ |
| Phase, moderate | $53.3_{\pm 5.5}$ | $54.3_{\pm 5.9}$ | $50.3_{\pm 9.3}$ | $51.7_{\pm 6.8}$ | $51.7_{\pm 6.8}$ | $50.3_{\pm 9.3}$ |
| Phase, strong | $53.0_{\pm 2.6}$ | $54.3_{\pm 3.8}$ | $52.7_{\pm 5.5}$ | $55.3_{\pm 6.7}$ | $53.3_{\pm 4.9}$ | $56.3_{\pm 2.9}$ |
| Mixed, moderate | $50.0_{\pm 5.2}$ | $53.7_{\pm 4.0}$ | $43.3_{\pm 4.9}$ | $58.3_{\pm 0.6}$ | $57.0_{\pm 2.6}$ | $43.3_{\pm 3.1}$ |
| Mixed, strong | $52.3_{\pm 3.1}$ | $51.7_{\pm 3.8}$ | $55.0_{\pm 4.4}$ | $52.3_{\pm 8.6}$ | $51.7_{\pm 7.1}$ | $56.0_{\pm 3.6}$ |

Table 89: **Spurious-Fourier: OOD reversed diagnostics.** Stable-only accuracy (%), mean ± seed SD.

| Training | Stable-only | | |
|---|---|---|---|
| | A | F | H |
| **Local VQC** | | | |
| Clean | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $95.0_{\pm 8.7}$ |
| Amplitude, moderate | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| Amplitude, strong | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| Depolarizing, moderate | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $96.3_{\pm 6.4}$ |
| Depolarizing, strong | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $95.0_{\pm 8.7}$ |
| Phase, moderate | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $98.3_{\pm 2.9}$ |
| Phase, strong | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| Mixed, moderate | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ |
| Mixed, strong | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $92.0_{\pm 13.0}$ |
| **Entangling VQC** | | | |
| Clean | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $51.3_{\pm 5.9}$ |
| Amplitude, moderate | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $51.3_{\pm 5.9}$ |
| Amplitude, strong | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $53.0_{\pm 4.4}$ |
| Depolarizing, moderate | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $54.3_{\pm 5.5}$ |
| Depolarizing, strong | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $52.0_{\pm 5.2}$ |
| Phase, moderate | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $51.3_{\pm 5.9}$ |
| Phase, strong | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $51.3_{\pm 5.9}$ |
| Mixed, moderate | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $51.3_{\pm 5.9}$ |
| Mixed, strong | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $52.3_{\pm 4.9}$ |
| **Reuploading VQC** | | | |
| Clean | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $48.0_{\pm 5.6}$ |
| Amplitude, moderate | $26.7_{\pm 29.3}$ | $28.3_{\pm 15.0}$ | $46.0_{\pm 3.6}$ |
| Amplitude, strong | $65.0_{\pm 30.8}$ | $65.7_{\pm 30.1}$ | $51.3_{\pm 5.9}$ |
| Depolarizing, moderate | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $49.0_{\pm 5.3}$ |
| Depolarizing, strong | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $48.7_{\pm 4.5}$ |
| Phase, moderate | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $48.7_{\pm 4.5}$ |
| Phase, strong | $100.0_{\pm 0.0}$ | $99.7_{\pm 0.6}$ | $50.0_{\pm 2.6}$ |
| Mixed, moderate | $66.7_{\pm 57.7}$ | $68.0_{\pm 55.4}$ | $48.3_{\pm 5.5}$ |
| Mixed, strong | $100.0_{\pm 0.0}$ | $93.3_{\pm 11.5}$ | $51.0_{\pm 5.3}$ |

Table 90: **Spurious-Fourier: OOD reversed diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $49.3_{\pm 1.5}$ | $49.0_{\pm 1.7}$ | $53.3_{\pm 6.4}$ | $96.3_{\pm 1.5}$ | $96.3_{\pm 1.5}$ | $87.0_{\pm 4.0}$ |
| Amplitude, moderate | $51.0_{\pm 2.6}$ | $49.3_{\pm 0.6}$ | $54.3_{\pm 2.9}$ | $95.7_{\pm 2.1}$ | $95.7_{\pm 2.1}$ | $82.3_{\pm 4.7}$ |
| Amplitude, strong | $50.3_{\pm 2.1}$ | $50.7_{\pm 0.6}$ | $53.0_{\pm 2.6}$ | $95.7_{\pm 2.1}$ | $95.3_{\pm 2.5}$ | $89.7_{\pm 6.0}$ |
| Depolarizing, moderate | $50.0_{\pm 1.7}$ | $50.0_{\pm 2.6}$ | $53.7_{\pm 2.9}$ | $96.7_{\pm 1.2}$ | $96.7_{\pm 1.2}$ | $76.3_{\pm 4.0}$ |
| Depolarizing, strong | $48.3_{\pm 3.2}$ | $50.0_{\pm 2.6}$ | $53.7_{\pm 3.5}$ | $96.3_{\pm 1.5}$ | $96.3_{\pm 1.5}$ | $85.3_{\pm 7.2}$ |
| Phase, moderate | $48.3_{\pm 2.3}$ | $49.0_{\pm 3.5}$ | $52.0_{\pm 5.6}$ | $96.7_{\pm 1.2}$ | $96.0_{\pm 1.7}$ | $91.0_{\pm 5.6}$ |
| Phase, strong | $49.3_{\pm 2.5}$ | $49.3_{\pm 3.5}$ | $52.3_{\pm 4.5}$ | $96.7_{\pm 1.2}$ | $97.0_{\pm 1.0}$ | $95.3_{\pm 1.2}$ |
| Mixed, moderate | $50.0_{\pm 4.4}$ | $49.0_{\pm 5.6}$ | $53.3_{\pm 3.5}$ | $96.3_{\pm 1.5}$ | $96.7_{\pm 1.2}$ | $93.7_{\pm 2.5}$ |
| Mixed, strong | $49.3_{\pm 2.1}$ | $49.7_{\pm 2.1}$ | $51.3_{\pm 2.3}$ | $96.0_{\pm 2.0}$ | $96.0_{\pm 2.0}$ | $88.3_{\pm 8.1}$ |
| **Entangling VQC** | | | | | | |
| Clean | $48.0_{\pm 4.4}$ | $48.7_{\pm 5.0}$ | $51.3_{\pm 5.9}$ | $90.3_{\pm 1.5}$ | $89.7_{\pm 1.5}$ | $51.3_{\pm 5.9}$ |
| Amplitude, moderate | $48.3_{\pm 2.5}$ | $47.3_{\pm 3.1}$ | $51.7_{\pm 5.5}$ | $78.0_{\pm 3.6}$ | $79.0_{\pm 5.0}$ | $49.0_{\pm 2.0}$ |
| Amplitude, strong | $47.3_{\pm 2.5}$ | $46.7_{\pm 1.5}$ | $48.7_{\pm 3.8}$ | $67.7_{\pm 3.8}$ | $68.0_{\pm 4.4}$ | $49.0_{\pm 5.0}$ |
| Depolarizing, moderate | $49.0_{\pm 1.7}$ | $49.3_{\pm 3.2}$ | $50.0_{\pm 4.6}$ | $79.7_{\pm 6.4}$ | $78.3_{\pm 6.8}$ | $51.3_{\pm 5.9}$ |
| Depolarizing, strong | $49.7_{\pm 2.9}$ | $49.0_{\pm 2.0}$ | $55.7_{\pm 5.9}$ | $96.0_{\pm 1.7}$ | $96.3_{\pm 1.5}$ | $52.0_{\pm 5.2}$ |
| Phase, moderate | $49.7_{\pm 3.1}$ | $50.3_{\pm 4.7}$ | $51.3_{\pm 5.9}$ | $88.3_{\pm 3.5}$ | $87.7_{\pm 3.5}$ | $51.3_{\pm 5.9}$ |
| Phase, strong | $49.3_{\pm 2.5}$ | $47.0_{\pm 2.6}$ | $51.3_{\pm 5.9}$ | $87.7_{\pm 4.5}$ | $87.3_{\pm 3.5}$ | $51.3_{\pm 5.9}$ |
| Mixed, moderate | $51.0_{\pm 0.0}$ | $53.0_{\pm 1.7}$ | $51.3_{\pm 5.9}$ | $81.7_{\pm 7.4}$ | $81.7_{\pm 6.4}$ | $51.3_{\pm 5.9}$ |
| Mixed, strong | $49.7_{\pm 3.5}$ | $48.7_{\pm 4.9}$ | $51.0_{\pm 6.2}$ | $81.3_{\pm 10.1}$ | $81.0_{\pm 12.1}$ | $51.7_{\pm 5.5}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $49.3_{\pm 3.5}$ | $49.7_{\pm 5.5}$ | $48.0_{\pm 5.6}$ | $54.0_{\pm 8.9}$ | $55.0_{\pm 5.3}$ | $48.3_{\pm 6.0}$ |
| Amplitude, moderate | $48.3_{\pm 5.0}$ | $47.0_{\pm 3.6}$ | $46.0_{\pm 3.6}$ | $55.0_{\pm 7.0}$ | $55.0_{\pm 4.6}$ | $46.0_{\pm 3.6}$ |
| Amplitude, strong | $49.3_{\pm 3.2}$ | $48.7_{\pm 2.1}$ | $51.3_{\pm 5.9}$ | $52.7_{\pm 5.8}$ | $51.0_{\pm 6.2}$ | $51.3_{\pm 5.9}$ |
| Depolarizing, moderate | $49.3_{\pm 4.9}$ | $51.3_{\pm 5.1}$ | $50.3_{\pm 3.1}$ | $46.3_{\pm 6.5}$ | $47.3_{\pm 3.1}$ | $52.3_{\pm 2.1}$ |
| Depolarizing, strong | $47.3_{\pm 4.7}$ | $46.7_{\pm 4.0}$ | $48.7_{\pm 4.5}$ | $49.0_{\pm 10.1}$ | $48.0_{\pm 8.5}$ | $46.0_{\pm 6.2}$ |
| Phase, moderate | $44.3_{\pm 1.5}$ | $45.3_{\pm 2.5}$ | $48.3_{\pm 5.0}$ | $53.0_{\pm 6.9}$ | $53.3_{\pm 6.8}$ | $50.0_{\pm 2.6}$ |
| Phase, strong | $41.3_{\pm 7.6}$ | $45.3_{\pm 6.5}$ | $48.7_{\pm 4.5}$ | $56.7_{\pm 3.5}$ | $57.0_{\pm 5.0}$ | $48.3_{\pm 5.0}$ |
| Mixed, moderate | $42.7_{\pm 13.9}$ | $44.7_{\pm 15.3}$ | $45.7_{\pm 3.5}$ | $61.0_{\pm 7.2}$ | $60.7_{\pm 6.7}$ | $48.3_{\pm 5.5}$ |
| Mixed, strong | $47.0_{\pm 5.2}$ | $47.3_{\pm 6.7}$ | $54.3_{\pm 4.6}$ | $51.3_{\pm 9.7}$ | $51.0_{\pm 10.0}$ | $51.0_{\pm 5.3}$ |

Table 91: **Q-FOS: ID diagnostics.** Full competing-feature task and stable-only accuracy (%), mean ± seed SD.

| Training | Full competing-feature task | | | Stable-only | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $89.7_{\pm 1.2}$ | $89.3_{\pm 1.5}$ | $86.3_{\pm 5.5}$ | $38.3_{\pm 25.4}$ | $39.3_{\pm 23.7}$ | $38.3_{\pm 25.4}$ |
| Amplitude, moderate | $89.0_{\pm 1.0}$ | $89.7_{\pm 0.6}$ | $91.0_{\pm 1.7}$ | $35.0_{\pm 30.3}$ | $31.7_{\pm 27.8}$ | $24.0_{\pm 26.2}$ |
| Amplitude, strong | $81.7_{\pm 14.4}$ | $82.0_{\pm 13.9}$ | $81.0_{\pm 4.6}$ | $19.7_{\pm 28.2}$ | $24.3_{\pm 26.2}$ | $38.3_{\pm 18.7}$ |
| Depolarizing, moderate | $90.3_{\pm 1.5}$ | $90.3_{\pm 1.5}$ | $90.0_{\pm 1.7}$ | $30.0_{\pm 27.5}$ | $29.3_{\pm 27.3}$ | $34.3_{\pm 4.5}$ |
| Depolarizing, strong | $90.7_{\pm 2.1}$ | $90.7_{\pm 2.1}$ | $87.0_{\pm 5.3}$ | $50.3_{\pm 4.7}$ | $50.0_{\pm 5.3}$ | $48.7_{\pm 7.6}$ |
| Phase, moderate | $90.3_{\pm 1.5}$ | $90.3_{\pm 1.5}$ | $88.7_{\pm 4.2}$ | $17.0_{\pm 29.4}$ | $16.7_{\pm 28.9}$ | $21.3_{\pm 27.2}$ |
| Phase, strong | $90.3_{\pm 1.5}$ | $90.3_{\pm 1.5}$ | $88.3_{\pm 3.1}$ | $18.0_{\pm 31.2}$ | $18.7_{\pm 27.3}$ | $42.7_{\pm 17.9}$ |
| Mixed, moderate | $90.0_{\pm 1.7}$ | $90.3_{\pm 1.5}$ | $89.3_{\pm 1.5}$ | $25.7_{\pm 27.1}$ | $26.0_{\pm 27.1}$ | $20.3_{\pm 27.6}$ |
| Mixed, strong | $90.7_{\pm 2.1}$ | $90.7_{\pm 2.1}$ | $88.3_{\pm 3.5}$ | $33.3_{\pm 29.1}$ | $33.0_{\pm 28.9}$ | $39.3_{\pm 20.2}$ |
| **Entangling VQC** | | | | | | |
| Clean | $73.0_{\pm 15.5}$ | $72.7_{\pm 15.5}$ | $50.3_{\pm 4.7}$ | $62.3_{\pm 16.2}$ | $55.3_{\pm 5.1}$ | $48.3_{\pm 8.1}$ |
| Amplitude, moderate | $69.0_{\pm 22.9}$ | $69.3_{\pm 24.3}$ | $50.0_{\pm 5.3}$ | $61.0_{\pm 27.2}$ | $54.3_{\pm 26.5}$ | $53.7_{\pm 0.6}$ |
| Amplitude, strong | $74.3_{\pm 24.5}$ | $74.7_{\pm 24.0}$ | $49.7_{\pm 4.0}$ | $16.3_{\pm 28.3}$ | $22.0_{\pm 31.4}$ | $50.3_{\pm 4.0}$ |
| Depolarizing, moderate | $80.0_{\pm 14.7}$ | $77.3_{\pm 18.5}$ | $44.3_{\pm 4.7}$ | $50.3_{\pm 48.0}$ | $46.7_{\pm 31.1}$ | $47.7_{\pm 2.1}$ |
| Depolarizing, strong | $78.7_{\pm 17.0}$ | $79.0_{\pm 16.5}$ | $48.7_{\pm 5.0}$ | $16.0_{\pm 26.0}$ | $17.3_{\pm 25.0}$ | $48.7_{\pm 5.0}$ |
| Phase, moderate | $82.3_{\pm 15.0}$ | $83.3_{\pm 13.3}$ | $52.0_{\pm 3.5}$ | $29.0_{\pm 27.2}$ | $28.0_{\pm 27.1}$ | $52.0_{\pm 3.5}$ |
| Phase, strong | $77.0_{\pm 22.5}$ | $76.3_{\pm 22.9}$ | $52.0_{\pm 3.5}$ | $33.3_{\pm 29.1}$ | $33.0_{\pm 28.9}$ | $52.0_{\pm 3.5}$ |
| Mixed, moderate | $89.7_{\pm 1.5}$ | $90.3_{\pm 1.5}$ | $49.3_{\pm 4.2}$ | $15.0_{\pm 26.0}$ | $14.0_{\pm 22.5}$ | $49.0_{\pm 4.4}$ |
| Mixed, strong | $90.7_{\pm 3.1}$ | $90.7_{\pm 3.1}$ | $49.3_{\pm 4.2}$ | $0.0_{\pm 0.0}$ | $0.7_{\pm 1.2}$ | $49.7_{\pm 4.7}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $48.0_{\pm 2.0}$ | $44.7_{\pm 3.8}$ | $48.0_{\pm 3.5}$ | $64.7_{\pm 27.3}$ | $62.0_{\pm 22.7}$ | $48.0_{\pm 3.5}$ |
| Amplitude, moderate | $73.7_{\pm 18.6}$ | $74.7_{\pm 16.4}$ | $49.0_{\pm 4.4}$ | $50.7_{\pm 4.2}$ | $52.7_{\pm 6.1}$ | $52.0_{\pm 5.3}$ |
| Amplitude, strong | $52.0_{\pm 6.1}$ | $51.0_{\pm 4.4}$ | $50.7_{\pm 4.2}$ | $58.0_{\pm 19.1}$ | $58.0_{\pm 19.1}$ | $50.7_{\pm 4.2}$ |
| Depolarizing, moderate | $48.7_{\pm 3.1}$ | $48.7_{\pm 4.7}$ | $52.3_{\pm 5.1}$ | $84.7_{\pm 20.8}$ | $83.7_{\pm 19.6}$ | $48.0_{\pm 4.6}$ |
| Depolarizing, strong | $43.0_{\pm 3.6}$ | $42.7_{\pm 4.0}$ | $50.3_{\pm 5.5}$ | $100.0_{\pm 0.0}$ | $99.3_{\pm 1.2}$ | $53.0_{\pm 4.6}$ |
| Phase, moderate | $51.7_{\pm 2.9}$ | $50.3_{\pm 3.5}$ | $46.7_{\pm 1.2}$ | $56.7_{\pm 17.4}$ | $60.3_{\pm 22.1}$ | $46.7_{\pm 1.2}$ |
| Phase, strong | $49.3_{\pm 8.3}$ | $48.0_{\pm 8.7}$ | $46.7_{\pm 2.3}$ | $54.0_{\pm 1.0}$ | $53.0_{\pm 10.5}$ | $47.0_{\pm 1.7}$ |
| Mixed, moderate | $48.3_{\pm 17.6}$ | $49.3_{\pm 19.2}$ | $49.3_{\pm 4.2}$ | $65.7_{\pm 25.6}$ | $65.0_{\pm 19.9}$ | $49.3_{\pm 4.2}$ |
| Mixed, strong | $49.0_{\pm 10.8}$ | $48.7_{\pm 9.3}$ | $50.0_{\pm 5.3}$ | $84.7_{\pm 26.6}$ | $84.0_{\pm 26.0}$ | $45.0_{\pm 3.6}$ |

Table 92: **Q-FOS: ID diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $90.7_{\pm 2.1}$ | $90.7_{\pm 2.1}$ | $85.3_{\pm 3.2}$ | $89.0_{\pm 1.7}$ | $89.3_{\pm 2.3}$ | $85.3_{\pm 7.4}$ |
| Amplitude, moderate | $90.0_{\pm 2.6}$ | $90.0_{\pm 2.6}$ | $89.7_{\pm 3.1}$ | $89.0_{\pm 1.0}$ | $89.0_{\pm 1.0}$ | $88.7_{\pm 3.1}$ |
| Amplitude, strong | $89.0_{\pm 3.6}$ | $89.0_{\pm 3.6}$ | $85.0_{\pm 7.0}$ | $82.0_{\pm 12.2}$ | $82.3_{\pm 11.6}$ | $85.7_{\pm 2.5}$ |
| Depolarizing, moderate | $90.7_{\pm 2.1}$ | $90.7_{\pm 2.1}$ | $89.3_{\pm 2.5}$ | $90.0_{\pm 2.0}$ | $90.0_{\pm 2.0}$ | $87.3_{\pm 2.9}$ |
| Depolarizing, strong | $90.7_{\pm 2.1}$ | $90.7_{\pm 2.1}$ | $80.3_{\pm 7.4}$ | $90.0_{\pm 2.0}$ | $90.0_{\pm 2.0}$ | $85.0_{\pm 6.6}$ |
| Phase, moderate | $90.7_{\pm 2.1}$ | $90.7_{\pm 2.1}$ | $88.0_{\pm 2.0}$ | $90.0_{\pm 2.0}$ | $90.0_{\pm 2.0}$ | $87.3_{\pm 4.0}$ |
| Phase, strong | $90.7_{\pm 2.1}$ | $90.7_{\pm 2.1}$ | $87.3_{\pm 0.6}$ | $90.0_{\pm 2.0}$ | $90.0_{\pm 2.0}$ | $87.3_{\pm 5.7}$ |
| Mixed, moderate | $90.7_{\pm 2.1}$ | $90.7_{\pm 2.1}$ | $87.3_{\pm 4.0}$ | $90.0_{\pm 2.0}$ | $90.0_{\pm 2.0}$ | $86.3_{\pm 4.0}$ |
| Mixed, strong | $90.7_{\pm 2.1}$ | $90.7_{\pm 2.1}$ | $84.3_{\pm 6.8}$ | $90.0_{\pm 2.0}$ | $90.0_{\pm 2.0}$ | $88.0_{\pm 3.6}$ |
| **Entangling VQC** | | | | | | |
| Clean | $26.7_{\pm 23.8}$ | $30.0_{\pm 21.6}$ | $49.7_{\pm 5.9}$ | $73.3_{\pm 15.0}$ | $73.7_{\pm 13.6}$ | $49.3_{\pm 6.4}$ |
| Amplitude, moderate | $36.7_{\pm 23.4}$ | $35.0_{\pm 22.6}$ | $51.3_{\pm 3.1}$ | $69.3_{\pm 23.5}$ | $68.7_{\pm 20.5}$ | $50.7_{\pm 4.9}$ |
| Amplitude, strong | $13.3_{\pm 3.2}$ | $14.7_{\pm 5.5}$ | $52.3_{\pm 2.9}$ | $74.7_{\pm 24.0}$ | $75.3_{\pm 22.8}$ | $53.0_{\pm 1.7}$ |
| Depolarizing, moderate | $9.7_{\pm 1.5}$ | $15.7_{\pm 9.0}$ | $46.7_{\pm 1.2}$ | $76.3_{\pm 21.1}$ | $75.7_{\pm 22.2}$ | $46.7_{\pm 1.2}$ |
| Depolarizing, strong | $38.3_{\pm 47.3}$ | $38.3_{\pm 47.3}$ | $49.3_{\pm 4.2}$ | $78.3_{\pm 18.5}$ | $78.3_{\pm 18.5}$ | $49.0_{\pm 4.6}$ |
| Phase, moderate | $9.3_{\pm 2.1}$ | $9.3_{\pm 2.1}$ | $52.0_{\pm 3.5}$ | $82.0_{\pm 15.6}$ | $82.3_{\pm 15.0}$ | $52.0_{\pm 3.5}$ |
| Phase, strong | $24.0_{\pm 26.1}$ | $24.0_{\pm 26.1}$ | $52.0_{\pm 3.5}$ | $76.3_{\pm 21.1}$ | $76.0_{\pm 21.7}$ | $52.0_{\pm 3.5}$ |
| Mixed, moderate | $9.3_{\pm 2.1}$ | $9.7_{\pm 2.3}$ | $49.3_{\pm 4.2}$ | $89.3_{\pm 1.2}$ | $89.7_{\pm 1.5}$ | $49.3_{\pm 4.2}$ |
| Mixed, strong | $9.3_{\pm 2.1}$ | $9.3_{\pm 2.1}$ | $49.0_{\pm 3.6}$ | $90.7_{\pm 3.1}$ | $90.3_{\pm 2.5}$ | $49.3_{\pm 4.2}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $54.7_{\pm 20.5}$ | $58.0_{\pm 18.4}$ | $48.0_{\pm 3.5}$ | $39.3_{\pm 6.1}$ | $42.7_{\pm 6.7}$ | $48.0_{\pm 3.5}$ |
| Amplitude, moderate | $60.7_{\pm 42.9}$ | $60.3_{\pm 42.7}$ | $51.7_{\pm 4.9}$ | $70.3_{\pm 22.5}$ | $70.7_{\pm 22.7}$ | $50.7_{\pm 4.2}$ |
| Amplitude, strong | $52.0_{\pm 3.6}$ | $47.7_{\pm 3.5}$ | $50.7_{\pm 4.2}$ | $44.3_{\pm 7.0}$ | $45.0_{\pm 8.0}$ | $50.7_{\pm 4.2}$ |
| Depolarizing, moderate | $56.3_{\pm 13.3}$ | $54.7_{\pm 12.9}$ | $49.7_{\pm 2.5}$ | $39.3_{\pm 10.2}$ | $39.3_{\pm 11.0}$ | $47.0_{\pm 4.6}$ |
| Depolarizing, strong | $61.7_{\pm 2.1}$ | $59.7_{\pm 3.1}$ | $51.3_{\pm 3.1}$ | $31.0_{\pm 4.4}$ | $32.3_{\pm 3.8}$ | $51.3_{\pm 3.1}$ |
| Phase, moderate | $56.7_{\pm 14.2}$ | $57.0_{\pm 13.9}$ | $46.3_{\pm 1.5}$ | $45.7_{\pm 2.5}$ | $45.3_{\pm 3.8}$ | $46.7_{\pm 1.2}$ |
| Phase, strong | $65.3_{\pm 23.4}$ | $64.3_{\pm 20.3}$ | $46.7_{\pm 1.2}$ | $46.0_{\pm 4.4}$ | $46.7_{\pm 2.5}$ | $46.7_{\pm 1.2}$ |
| Mixed, moderate | $71.0_{\pm 16.7}$ | $71.3_{\pm 17.2}$ | $49.3_{\pm 4.2}$ | $41.3_{\pm 21.6}$ | $40.3_{\pm 22.2}$ | $49.7_{\pm 4.7}$ |
| Mixed, strong | $64.0_{\pm 22.9}$ | $63.3_{\pm 23.0}$ | $49.3_{\pm 4.2}$ | $41.7_{\pm 16.5}$ | $42.0_{\pm 15.5}$ | $47.0_{\pm 1.0}$ |

Table 93: **Q-FOS: Random diagnostics.** Full competing-feature task and stable-only accuracy (%), mean ± seed SD.

| Training | Full competing-feature task | | | Stable-only | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $40.0_{\pm 11.5}$ | $41.3_{\pm 10.3}$ | $41.7_{\pm 11.0}$ | $31.0_{\pm 26.2}$ | $32.0_{\pm 24.4}$ | $35.0_{\pm 19.3}$ |
| Amplitude, moderate | $41.0_{\pm 12.1}$ | $40.7_{\pm 12.7}$ | $43.3_{\pm 13.3}$ | $30.7_{\pm 26.7}$ | $27.0_{\pm 24.9}$ | $17.0_{\pm 27.7}$ |
| Amplitude, strong | $44.3_{\pm 11.6}$ | $45.3_{\pm 10.8}$ | $44.7_{\pm 10.4}$ | $25.7_{\pm 24.6}$ | $26.3_{\pm 24.7}$ | $28.3_{\pm 20.2}$ |
| Depolarizing, moderate | $39.0_{\pm 9.5}$ | $39.0_{\pm 9.5}$ | $44.3_{\pm 5.1}$ | $23.0_{\pm 21.7}$ | $22.7_{\pm 21.6}$ | $49.7_{\pm 6.5}$ |
| Depolarizing, strong | $38.7_{\pm 9.3}$ | $39.0_{\pm 10.8}$ | $38.0_{\pm 8.7}$ | $48.7_{\pm 5.5}$ | $45.0_{\pm 3.5}$ | $47.7_{\pm 4.2}$ |
| Phase, moderate | $41.7_{\pm 11.8}$ | $41.7_{\pm 11.8}$ | $42.3_{\pm 14.2}$ | $16.3_{\pm 28.3}$ | $16.0_{\pm 27.7}$ | $16.3_{\pm 28.3}$ |
| Phase, strong | $42.0_{\pm 12.3}$ | $42.0_{\pm 12.1}$ | $42.7_{\pm 8.4}$ | $17.0_{\pm 22.9}$ | $15.7_{\pm 17.8}$ | $45.3_{\pm 3.5}$ |
| Mixed, moderate | $39.3_{\pm 8.5}$ | $37.7_{\pm 8.6}$ | $43.0_{\pm 10.6}$ | $17.3_{\pm 22.7}$ | $17.3_{\pm 22.7}$ | $19.0_{\pm 26.0}$ |
| Mixed, strong | $38.7_{\pm 7.4}$ | $38.0_{\pm 7.0}$ | $39.0_{\pm 8.2}$ | $32.7_{\pm 28.9}$ | $32.0_{\pm 28.2}$ | $30.3_{\pm 25.7}$ |
| **Entangling VQC** | | | | | | |
| Clean | $49.7_{\pm 5.0}$ | $50.3_{\pm 5.5}$ | $49.7_{\pm 6.0}$ | $58.7_{\pm 21.2}$ | $52.7_{\pm 11.0}$ | $50.3_{\pm 7.1}$ |
| Amplitude, moderate | $46.7_{\pm 11.1}$ | $46.7_{\pm 8.5}$ | $49.3_{\pm 6.8}$ | $54.0_{\pm 37.0}$ | $50.7_{\pm 29.6}$ | $44.7_{\pm 1.2}$ |
| Amplitude, strong | $39.7_{\pm 15.3}$ | $40.0_{\pm 14.4}$ | $51.7_{\pm 7.1}$ | $20.0_{\pm 34.6}$ | $26.3_{\pm 35.0}$ | $45.7_{\pm 9.6}$ |
| Depolarizing, moderate | $38.3_{\pm 14.8}$ | $38.0_{\pm 13.2}$ | $51.3_{\pm 1.5}$ | $48.0_{\pm 43.1}$ | $44.7_{\pm 32.7}$ | $54.7_{\pm 3.2}$ |
| Depolarizing, strong | $42.7_{\pm 13.3}$ | $43.0_{\pm 13.7}$ | $48.3_{\pm 4.6}$ | $19.0_{\pm 32.9}$ | $20.7_{\pm 31.6}$ | $50.7_{\pm 7.5}$ |
| Phase, moderate | $42.0_{\pm 4.0}$ | $41.3_{\pm 4.2}$ | $46.0_{\pm 4.4}$ | $29.0_{\pm 25.1}$ | $28.7_{\pm 24.8}$ | $46.0_{\pm 4.4}$ |
| Phase, strong | $40.3_{\pm 6.4}$ | $40.0_{\pm 7.8}$ | $46.0_{\pm 4.4}$ | $33.7_{\pm 29.9}$ | $33.3_{\pm 29.5}$ | $46.0_{\pm 4.4}$ |
| Mixed, moderate | $37.7_{\pm 9.7}$ | $38.7_{\pm 10.7}$ | $49.7_{\pm 6.5}$ | $19.0_{\pm 32.9}$ | $17.3_{\pm 29.2}$ | $50.0_{\pm 6.6}$ |
| Mixed, strong | $36.0_{\pm 7.5}$ | $36.7_{\pm 8.1}$ | $50.0_{\pm 6.6}$ | $0.7_{\pm 1.2}$ | $2.0_{\pm 3.5}$ | $50.0_{\pm 6.6}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $58.0_{\pm 7.9}$ | $57.0_{\pm 6.1}$ | $54.0_{\pm 4.4}$ | $68.7_{\pm 27.4}$ | $64.0_{\pm 19.5}$ | $54.0_{\pm 4.4}$ |
| Amplitude, moderate | $38.3_{\pm 9.3}$ | $38.0_{\pm 8.7}$ | $51.0_{\pm 6.6}$ | $51.3_{\pm 7.2}$ | $53.0_{\pm 8.7}$ | $50.0_{\pm 6.6}$ |
| Amplitude, strong | $56.0_{\pm 9.5}$ | $55.7_{\pm 9.0}$ | $50.0_{\pm 6.6}$ | $65.7_{\pm 21.2}$ | $63.3_{\pm 16.4}$ | $50.0_{\pm 6.6}$ |
| Depolarizing, moderate | $56.3_{\pm 8.1}$ | $57.0_{\pm 9.6}$ | $48.7_{\pm 4.0}$ | $77.3_{\pm 20.5}$ | $78.0_{\pm 20.0}$ | $48.3_{\pm 5.1}$ |
| Depolarizing, strong | $60.3_{\pm 6.7}$ | $59.0_{\pm 6.0}$ | $49.3_{\pm 5.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $53.0_{\pm 9.2}$ |
| Phase, moderate | $54.7_{\pm 2.5}$ | $51.0_{\pm 1.0}$ | $55.7_{\pm 4.5}$ | $60.0_{\pm 23.1}$ | $55.0_{\pm 24.2}$ | $54.7_{\pm 3.2}$ |
| Phase, strong | $57.7_{\pm 2.5}$ | $56.3_{\pm 3.5}$ | $54.7_{\pm 3.2}$ | $51.0_{\pm 15.7}$ | $52.3_{\pm 15.3}$ | $54.3_{\pm 3.8}$ |
| Mixed, moderate | $54.3_{\pm 11.2}$ | $54.0_{\pm 10.4}$ | $50.7_{\pm 6.5}$ | $64.7_{\pm 30.9}$ | $65.7_{\pm 29.5}$ | $50.7_{\pm 6.5}$ |
| Mixed, strong | $56.3_{\pm 11.9}$ | $58.0_{\pm 10.4}$ | $52.7_{\pm 3.8}$ | $81.0_{\pm 32.9}$ | $78.0_{\pm 30.6}$ | $54.3_{\pm 3.1}$ |

Table 94: **Q-FOS: Random diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $49.3_{\pm 7.6}$ | $48.3_{\pm 7.6}$ | $45.7_{\pm 4.0}$ | $46.3_{\pm 5.7}$ | $46.3_{\pm 5.7}$ | $47.0_{\pm 6.1}$ |
| Amplitude, moderate | $48.3_{\pm 7.6}$ | $48.7_{\pm 7.2}$ | $47.0_{\pm 4.6}$ | $46.7_{\pm 5.9}$ | $46.3_{\pm 5.7}$ | $45.7_{\pm 8.7}$ |
| Amplitude, strong | $47.3_{\pm 9.2}$ | $47.7_{\pm 8.1}$ | $44.7_{\pm 5.5}$ | $47.0_{\pm 9.5}$ | $46.0_{\pm 7.9}$ | $48.7_{\pm 9.5}$ |
| Depolarizing, moderate | $51.0_{\pm 9.5}$ | $51.0_{\pm 8.7}$ | $49.0_{\pm 6.0}$ | $47.3_{\pm 5.5}$ | $46.7_{\pm 4.9}$ | $43.0_{\pm 4.4}$ |
| Depolarizing, strong | $51.7_{\pm 7.4}$ | $51.7_{\pm 8.1}$ | $48.0_{\pm 5.3}$ | $45.3_{\pm 5.5}$ | $45.3_{\pm 4.7}$ | $47.0_{\pm 6.2}$ |
| Phase, moderate | $47.7_{\pm 7.0}$ | $47.3_{\pm 6.0}$ | $48.3_{\pm 5.0}$ | $46.0_{\pm 5.3}$ | $45.3_{\pm 5.0}$ | $46.3_{\pm 4.7}$ |
| Phase, strong | $50.3_{\pm 8.7}$ | $49.0_{\pm 8.9}$ | $48.7_{\pm 5.0}$ | $45.0_{\pm 4.6}$ | $45.0_{\pm 5.6}$ | $45.0_{\pm 4.6}$ |
| Mixed, moderate | $51.0_{\pm 8.7}$ | $50.7_{\pm 9.7}$ | $50.3_{\pm 8.1}$ | $47.0_{\pm 5.2}$ | $47.0_{\pm 6.1}$ | $44.3_{\pm 8.1}$ |
| Mixed, strong | $54.7_{\pm 6.4}$ | $55.3_{\pm 5.8}$ | $55.0_{\pm 3.5}$ | $45.3_{\pm 6.4}$ | $46.3_{\pm 6.4}$ | $44.0_{\pm 6.2}$ |
| **Entangling VQC** | | | | | | |
| Clean | $48.7_{\pm 6.4}$ | $48.3_{\pm 5.9}$ | $47.7_{\pm 3.2}$ | $52.0_{\pm 5.6}$ | $49.3_{\pm 6.0}$ | $47.3_{\pm 2.9}$ |
| Amplitude, moderate | $49.7_{\pm 6.7}$ | $50.7_{\pm 6.4}$ | $45.7_{\pm 2.1}$ | $47.7_{\pm 8.1}$ | $49.3_{\pm 5.7}$ | $53.0_{\pm 10.1}$ |
| Amplitude, strong | $52.3_{\pm 9.0}$ | $52.0_{\pm 9.5}$ | $50.3_{\pm 7.1}$ | $46.7_{\pm 7.6}$ | $45.7_{\pm 7.1}$ | $52.0_{\pm 7.5}$ |
| Depolarizing, moderate | $48.3_{\pm 8.1}$ | $50.3_{\pm 9.1}$ | $54.0_{\pm 2.6}$ | $47.7_{\pm 7.8}$ | $46.7_{\pm 6.8}$ | $54.7_{\pm 3.2}$ |
| Depolarizing, strong | $49.0_{\pm 7.5}$ | $48.7_{\pm 8.1}$ | $49.0_{\pm 5.3}$ | $46.3_{\pm 6.4}$ | $46.7_{\pm 7.8}$ | $50.3_{\pm 7.0}$ |
| Phase, moderate | $43.7_{\pm 4.9}$ | $44.0_{\pm 4.6}$ | $46.0_{\pm 4.4}$ | $49.0_{\pm 1.7}$ | $50.0_{\pm 2.0}$ | $46.0_{\pm 4.4}$ |
| Phase, strong | $47.0_{\pm 7.9}$ | $47.0_{\pm 7.9}$ | $46.0_{\pm 4.4}$ | $48.0_{\pm 3.5}$ | $47.3_{\pm 2.5}$ | $49.7_{\pm 6.1}$ |
| Mixed, moderate | $50.3_{\pm 10.6}$ | $50.0_{\pm 10.5}$ | $50.0_{\pm 6.6}$ | $47.0_{\pm 5.3}$ | $47.3_{\pm 5.5}$ | $50.0_{\pm 6.6}$ |
| Mixed, strong | $48.7_{\pm 9.1}$ | $49.0_{\pm 8.9}$ | $50.0_{\pm 6.6}$ | $46.0_{\pm 5.6}$ | $46.3_{\pm 6.0}$ | $50.0_{\pm 6.6}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $50.7_{\pm 4.9}$ | $53.7_{\pm 0.6}$ | $54.0_{\pm 4.4}$ | $48.7_{\pm 2.1}$ | $49.7_{\pm 3.1}$ | $54.0_{\pm 4.4}$ |
| Amplitude, moderate | $48.7_{\pm 8.1}$ | $48.0_{\pm 8.9}$ | $52.7_{\pm 7.5}$ | $45.3_{\pm 5.5}$ | $46.0_{\pm 4.6}$ | $50.7_{\pm 6.5}$ |
| Amplitude, strong | $53.0_{\pm 2.0}$ | $52.3_{\pm 5.1}$ | $50.0_{\pm 6.6}$ | $49.0_{\pm 4.4}$ | $49.7_{\pm 3.5}$ | $50.0_{\pm 6.6}$ |
| Depolarizing, moderate | $47.0_{\pm 8.9}$ | $49.0_{\pm 10.5}$ | $46.0_{\pm 4.4}$ | $46.7_{\pm 2.5}$ | $47.7_{\pm 0.6}$ | $49.3_{\pm 6.7}$ |
| Depolarizing, strong | $50.7_{\pm 11.8}$ | $51.7_{\pm 11.9}$ | $49.0_{\pm 5.6}$ | $43.7_{\pm 3.5}$ | $43.3_{\pm 2.9}$ | $47.3_{\pm 2.9}$ |
| Phase, moderate | $51.7_{\pm 6.0}$ | $53.0_{\pm 3.5}$ | $54.7_{\pm 3.2}$ | $46.0_{\pm 2.6}$ | $46.3_{\pm 2.1}$ | $54.7_{\pm 3.2}$ |
| Phase, strong | $51.3_{\pm 1.5}$ | $53.0_{\pm 2.6}$ | $54.0_{\pm 4.4}$ | $51.0_{\pm 5.2}$ | $51.3_{\pm 8.1}$ | $56.7_{\pm 4.9}$ |
| Mixed, moderate | $48.7_{\pm 5.0}$ | $49.3_{\pm 6.1}$ | $50.3_{\pm 7.0}$ | $44.0_{\pm 2.0}$ | $44.3_{\pm 2.5}$ | $50.7_{\pm 6.5}$ |
| Mixed, strong | $49.7_{\pm 6.4}$ | $49.3_{\pm 5.9}$ | $55.0_{\pm 3.5}$ | $46.3_{\pm 5.9}$ | $46.0_{\pm 6.2}$ | $54.7_{\pm 3.2}$ |

Table 95: **Q-FOS: OOD reversed diagnostics.** Stable-only accuracy (%), mean ± seed SD.

| Training | Stable-only | | |
|---|---|---|---|
| | A | F | H |
| **Local VQC** | | | |
| Clean | $37.3_{\pm 31.5}$ | $39.0_{\pm 28.6}$ | $37.3_{\pm 31.5}$ |
| Amplitude, moderate | $36.0_{\pm 31.2}$ | $34.7_{\pm 29.3}$ | $19.7_{\pm 29.8}$ |
| Amplitude, strong | $23.3_{\pm 27.7}$ | $24.7_{\pm 27.3}$ | $44.3_{\pm 9.1}$ |
| Depolarizing, moderate | $27.3_{\pm 28.6}$ | $28.7_{\pm 28.5}$ | $34.7_{\pm 9.2}$ |
| Depolarizing, strong | $54.3_{\pm 2.5}$ | $52.0_{\pm 6.2}$ | $50.7_{\pm 4.2}$ |
| Phase, moderate | $18.0_{\pm 31.2}$ | $17.7_{\pm 30.6}$ | $24.7_{\pm 26.1}$ |
| Phase, strong | $19.0_{\pm 31.2}$ | $20.3_{\pm 22.2}$ | $50.7_{\pm 8.5}$ |
| Mixed, moderate | $22.0_{\pm 30.6}$ | $24.7_{\pm 29.3}$ | $28.0_{\pm 23.3}$ |
| Mixed, strong | $36.3_{\pm 31.6}$ | $35.7_{\pm 31.1}$ | $39.0_{\pm 28.6}$ |
| **Entangling VQC** | | | |
| Clean | $61.3_{\pm 18.1}$ | $55.7_{\pm 8.6}$ | $47.7_{\pm 6.5}$ |
| Amplitude, moderate | $55.3_{\pm 32.6}$ | $53.7_{\pm 26.0}$ | $50.0_{\pm 2.6}$ |
| Amplitude, strong | $15.0_{\pm 25.1}$ | $20.7_{\pm 27.6}$ | $48.0_{\pm 2.0}$ |
| Depolarizing, moderate | $46.7_{\pm 44.5}$ | $46.3_{\pm 39.0}$ | $46.3_{\pm 5.5}$ |
| Depolarizing, strong | $14.3_{\pm 24.8}$ | $15.0_{\pm 24.3}$ | $51.7_{\pm 5.5}$ |
| Phase, moderate | $26.0_{\pm 24.2}$ | $24.3_{\pm 24.0}$ | $50.3_{\pm 5.9}$ |
| Phase, strong | $30.3_{\pm 26.4}$ | $30.0_{\pm 26.2}$ | $52.0_{\pm 5.6}$ |
| Mixed, moderate | $14.3_{\pm 24.8}$ | $13.7_{\pm 22.8}$ | $51.7_{\pm 5.5}$ |
| Mixed, strong | $0.3_{\pm 0.6}$ | $1.0_{\pm 1.7}$ | $51.7_{\pm 5.5}$ |
| **Reuploading VQC** | | | |
| Clean | $65.3_{\pm 29.7}$ | $61.3_{\pm 22.9}$ | $49.7_{\pm 5.9}$ |
| Amplitude, moderate | $52.0_{\pm 5.0}$ | $54.0_{\pm 2.6}$ | $49.0_{\pm 6.6}$ |
| Amplitude, strong | $62.7_{\pm 23.9}$ | $56.3_{\pm 12.3}$ | $48.3_{\pm 5.5}$ |
| Depolarizing, moderate | $79.7_{\pm 24.1}$ | $79.7_{\pm 22.8}$ | $47.3_{\pm 4.5}$ |
| Depolarizing, strong | $100.0_{\pm 0.0}$ | $98.7_{\pm 2.3}$ | $43.0_{\pm 6.0}$ |
| Phase, moderate | $56.3_{\pm 18.5}$ | $52.0_{\pm 17.8}$ | $47.3_{\pm 4.2}$ |
| Phase, strong | $48.7_{\pm 4.0}$ | $50.3_{\pm 10.4}$ | $47.7_{\pm 10.6}$ |
| Mixed, moderate | $67.7_{\pm 28.5}$ | $65.7_{\pm 22.3}$ | $45.7_{\pm 2.5}$ |
| Mixed, strong | $85.7_{\pm 24.8}$ | $83.3_{\pm 23.1}$ | $48.0_{\pm 6.2}$ |

Table 96: **Q-FOS: OOD reversed diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $74.0_{\pm 2.6}$ | $74.3_{\pm 3.2}$ | $67.0_{\pm 9.8}$ | $78.3_{\pm 5.7}$ | $78.3_{\pm 5.7}$ | $73.0_{\pm 11.3}$ |
| Amplitude, moderate | $73.0_{\pm 2.0}$ | $72.7_{\pm 1.5}$ | $69.3_{\pm 8.1}$ | $77.3_{\pm 3.2}$ | $78.3_{\pm 2.5}$ | $72.3_{\pm 4.0}$ |
| Amplitude, strong | $72.7_{\pm 1.5}$ | $72.3_{\pm 1.2}$ | $70.7_{\pm 11.0}$ | $72.0_{\pm 8.9}$ | $73.0_{\pm 8.9}$ | $60.7_{\pm 8.7}$ |
| Depolarizing, moderate | $75.7_{\pm 2.5}$ | $75.0_{\pm 2.6}$ | $70.7_{\pm 1.2}$ | $82.0_{\pm 1.0}$ | $81.3_{\pm 0.6}$ | $58.3_{\pm 3.5}$ |
| Depolarizing, strong | $75.0_{\pm 3.5}$ | $74.3_{\pm 3.1}$ | $57.7_{\pm 2.5}$ | $80.7_{\pm 6.5}$ | $81.0_{\pm 6.6}$ | $64.7_{\pm 5.5}$ |
| Phase, moderate | $75.7_{\pm 4.0}$ | $76.0_{\pm 3.6}$ | $68.7_{\pm 9.3}$ | $79.3_{\pm 3.1}$ | $79.0_{\pm 3.6}$ | $72.3_{\pm 10.3}$ |
| Phase, strong | $75.3_{\pm 3.1}$ | $75.7_{\pm 3.5}$ | $70.0_{\pm 3.0}$ | $79.7_{\pm 2.5}$ | $79.0_{\pm 2.6}$ | $72.7_{\pm 8.5}$ |
| Mixed, moderate | $72.3_{\pm 3.5}$ | $73.0_{\pm 2.6}$ | $64.3_{\pm 6.0}$ | $81.3_{\pm 3.2}$ | $81.3_{\pm 4.2}$ | $76.7_{\pm 4.0}$ |
| Mixed, strong | $72.7_{\pm 5.5}$ | $73.7_{\pm 5.5}$ | $68.0_{\pm 4.0}$ | $81.7_{\pm 4.6}$ | $82.0_{\pm 5.3}$ | $78.0_{\pm 7.0}$ |
| **Entangling VQC** | | | | | | |
| Clean | $33.0_{\pm 13.7}$ | $35.7_{\pm 10.7}$ | $48.0_{\pm 6.0}$ | $67.3_{\pm 8.1}$ | $69.0_{\pm 7.2}$ | $50.0_{\pm 3.5}$ |
| Amplitude, moderate | $42.0_{\pm 5.6}$ | $40.3_{\pm 3.2}$ | $48.3_{\pm 4.5}$ | $60.3_{\pm 15.5}$ | $60.0_{\pm 13.5}$ | $50.3_{\pm 2.1}$ |
| Amplitude, strong | $34.3_{\pm 11.0}$ | $34.0_{\pm 8.9}$ | $47.0_{\pm 1.7}$ | $68.7_{\pm 18.8}$ | $67.7_{\pm 20.6}$ | $49.0_{\pm 2.6}$ |
| Depolarizing, moderate | $26.3_{\pm 9.1}$ | $30.3_{\pm 15.5}$ | $47.3_{\pm 4.2}$ | $66.3_{\pm 17.8}$ | $65.0_{\pm 17.4}$ | $50.7_{\pm 4.0}$ |
| Depolarizing, strong | $45.7_{\pm 36.7}$ | $45.3_{\pm 37.8}$ | $53.0_{\pm 5.3}$ | $70.3_{\pm 18.6}$ | $71.0_{\pm 19.1}$ | $53.3_{\pm 4.0}$ |
| Phase, moderate | $25.0_{\pm 10.6}$ | $26.3_{\pm 11.2}$ | $50.3_{\pm 5.9}$ | $77.3_{\pm 12.4}$ | $79.3_{\pm 9.9}$ | $50.3_{\pm 5.9}$ |
| Phase, strong | $31.0_{\pm 14.8}$ | $30.7_{\pm 15.1}$ | $51.7_{\pm 5.5}$ | $70.3_{\pm 20.2}$ | $71.3_{\pm 21.1}$ | $51.7_{\pm 5.5}$ |
| Mixed, moderate | $22.0_{\pm 5.6}$ | $23.0_{\pm 3.6}$ | $51.7_{\pm 5.5}$ | $82.3_{\pm 1.5}$ | $82.3_{\pm 1.2}$ | $52.0_{\pm 5.0}$ |
| Mixed, strong | $25.3_{\pm 2.3}$ | $23.7_{\pm 2.1}$ | $51.7_{\pm 5.5}$ | $82.7_{\pm 0.6}$ | $83.0_{\pm 1.0}$ | $51.7_{\pm 5.5}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $29.3_{\pm 18.1}$ | $29.3_{\pm 15.3}$ | $49.7_{\pm 5.9}$ | $31.3_{\pm 7.1}$ | $32.0_{\pm 8.9}$ | $49.7_{\pm 5.9}$ |
| Amplitude, moderate | $49.0_{\pm 20.0}$ | $49.7_{\pm 21.1}$ | $48.7_{\pm 6.0}$ | $65.0_{\pm 24.6}$ | $66.0_{\pm 22.6}$ | $46.7_{\pm 3.2}$ |
| Amplitude, strong | $35.7_{\pm 15.4}$ | $37.0_{\pm 13.1}$ | $48.3_{\pm 5.5}$ | $39.0_{\pm 6.0}$ | $38.7_{\pm 8.0}$ | $48.3_{\pm 5.5}$ |
| Depolarizing, moderate | $36.0_{\pm 15.4}$ | $37.0_{\pm 16.8}$ | $45.3_{\pm 2.1}$ | $30.0_{\pm 10.5}$ | $28.0_{\pm 10.0}$ | $52.7_{\pm 4.2}$ |
| Depolarizing, strong | $36.3_{\pm 18.0}$ | $36.0_{\pm 19.1}$ | $53.0_{\pm 4.6}$ | $22.0_{\pm 7.9}$ | $21.7_{\pm 9.6}$ | $51.3_{\pm 3.1}$ |
| Phase, moderate | $35.7_{\pm 10.0}$ | $39.0_{\pm 9.6}$ | $47.0_{\pm 4.6}$ | $41.0_{\pm 9.2}$ | $38.0_{\pm 7.2}$ | $46.7_{\pm 5.0}$ |
| Phase, strong | $50.0_{\pm 7.8}$ | $51.3_{\pm 8.4}$ | $48.3_{\pm 2.5}$ | $43.7_{\pm 7.6}$ | $46.7_{\pm 5.0}$ | $46.7_{\pm 4.0}$ |
| Mixed, moderate | $33.7_{\pm 15.0}$ | $35.3_{\pm 13.6}$ | $45.7_{\pm 2.5}$ | $34.0_{\pm 27.2}$ | $35.7_{\pm 27.4}$ | $45.7_{\pm 2.5}$ |
| Mixed, strong | $44.7_{\pm 16.4}$ | $44.0_{\pm 15.9}$ | $47.7_{\pm 5.7}$ | $36.7_{\pm 15.5}$ | $38.0_{\pm 18.4}$ | $47.3_{\pm 5.1}$ |

Table 97: **Q-LINK: ID diagnostics.** Full competing-feature task and stable-only accuracy (%), mean ± seed SD.

| Training | Full competing-feature task | | | Stable-only | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $49.0_{\pm 3.6}$ | $50.7_{\pm 5.5}$ | $51.3_{\pm 2.9}$ |
| Amplitude, moderate | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $89.7_{\pm 1.5}$ | $48.7_{\pm 1.2}$ | $49.0_{\pm 2.6}$ | $51.7_{\pm 3.2}$ |
| Amplitude, strong | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $87.7_{\pm 1.5}$ | $51.0_{\pm 2.6}$ | $50.3_{\pm 2.1}$ | $51.7_{\pm 3.2}$ |
| Depolarizing, moderate | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $49.0_{\pm 3.6}$ | $49.0_{\pm 3.6}$ | $46.3_{\pm 8.6}$ |
| Depolarizing, strong | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $88.3_{\pm 2.5}$ | $49.0_{\pm 3.6}$ | $50.7_{\pm 6.4}$ | $48.7_{\pm 4.0}$ |
| Phase, moderate | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $89.7_{\pm 1.5}$ | $50.7_{\pm 2.5}$ | $50.0_{\pm 3.5}$ | $51.7_{\pm 3.2}$ |
| Phase, strong | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $88.0_{\pm 2.0}$ | $49.0_{\pm 3.6}$ | $48.7_{\pm 5.0}$ | $51.7_{\pm 3.2}$ |
| Mixed, moderate | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $89.0_{\pm 2.0}$ | $48.3_{\pm 12.5}$ | $49.7_{\pm 13.6}$ | $50.7_{\pm 2.5}$ |
| Mixed, strong | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $55.3_{\pm 15.4}$ | $48.7_{\pm 5.0}$ | $51.7_{\pm 3.2}$ |
| **Entangling VQC** | | | | | | |
| Clean | $99.3_{\pm 0.6}$ | $99.3_{\pm 0.6}$ | $49.0_{\pm 3.6}$ | $100.0_{\pm 0.0}$ | $99.7_{\pm 0.6}$ | $49.0_{\pm 3.6}$ |
| Amplitude, moderate | $98.7_{\pm 1.5}$ | $99.0_{\pm 1.7}$ | $49.0_{\pm 3.6}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $49.0_{\pm 3.6}$ |
| Amplitude, strong | $96.0_{\pm 1.0}$ | $96.0_{\pm 1.0}$ | $49.0_{\pm 3.6}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $49.0_{\pm 3.6}$ |
| Depolarizing, moderate | $99.7_{\pm 0.6}$ | $99.7_{\pm 0.6}$ | $50.7_{\pm 6.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $48.0_{\pm 6.1}$ |
| Depolarizing, strong | $96.0_{\pm 2.6}$ | $94.7_{\pm 2.1}$ | $49.0_{\pm 3.6}$ | $96.3_{\pm 2.9}$ | $96.3_{\pm 2.9}$ | $49.0_{\pm 3.6}$ |
| Phase, moderate | $98.0_{\pm 2.0}$ | $97.7_{\pm 1.5}$ | $47.0_{\pm 1.0}$ | $97.0_{\pm 3.0}$ | $96.7_{\pm 4.2}$ | $47.0_{\pm 1.0}$ |
| Phase, strong | $92.0_{\pm 2.6}$ | $91.7_{\pm 2.1}$ | $47.0_{\pm 1.0}$ | $76.7_{\pm 20.8}$ | $76.0_{\pm 20.4}$ | $47.0_{\pm 1.0}$ |
| Mixed, moderate | $99.0_{\pm 0.0}$ | $98.7_{\pm 0.6}$ | $49.0_{\pm 3.6}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $49.0_{\pm 3.6}$ |
| Mixed, strong | $93.7_{\pm 1.2}$ | $92.0_{\pm 1.0}$ | $48.7_{\pm 4.0}$ | $87.7_{\pm 15.7}$ | $88.0_{\pm 15.9}$ | $50.0_{\pm 2.6}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $96.0_{\pm 3.5}$ | $95.0_{\pm 3.6}$ | $51.0_{\pm 3.6}$ | $96.3_{\pm 6.4}$ | $93.3_{\pm 11.5}$ | $51.0_{\pm 3.6}$ |
| Amplitude, moderate | $93.7_{\pm 5.1}$ | $92.3_{\pm 5.7}$ | $53.7_{\pm 1.5}$ | $92.0_{\pm 13.9}$ | $92.0_{\pm 13.9}$ | $50.7_{\pm 4.2}$ |
| Amplitude, strong | $91.0_{\pm 3.5}$ | $92.0_{\pm 2.6}$ | $50.3_{\pm 3.8}$ | $98.0_{\pm 1.0}$ | $97.3_{\pm 1.2}$ | $50.3_{\pm 3.8}$ |
| Depolarizing, moderate | $97.0_{\pm 1.7}$ | $97.0_{\pm 1.0}$ | $50.3_{\pm 7.2}$ | $99.7_{\pm 0.6}$ | $100.0_{\pm 0.0}$ | $47.3_{\pm 6.5}$ |
| Depolarizing, strong | $94.0_{\pm 2.6}$ | $93.0_{\pm 4.0}$ | $53.0_{\pm 1.0}$ | $93.0_{\pm 9.5}$ | $93.7_{\pm 8.4}$ | $53.0_{\pm 1.0}$ |
| Phase, moderate | $78.3_{\pm 12.7}$ | $75.3_{\pm 13.6}$ | $50.0_{\pm 4.6}$ | $89.7_{\pm 11.0}$ | $90.0_{\pm 10.4}$ | $50.7_{\pm 5.7}$ |
| Phase, strong | $77.3_{\pm 8.1}$ | $75.7_{\pm 7.5}$ | $47.0_{\pm 1.0}$ | $83.3_{\pm 10.4}$ | $82.0_{\pm 10.4}$ | $47.0_{\pm 1.0}$ |
| Mixed, moderate | $97.0_{\pm 3.6}$ | $96.3_{\pm 4.0}$ | $52.0_{\pm 2.0}$ | $94.7_{\pm 9.2}$ | $95.3_{\pm 8.1}$ | $53.7_{\pm 2.9}$ |
| Mixed, strong | $83.0_{\pm 7.8}$ | $83.0_{\pm 6.1}$ | $53.0_{\pm 1.0}$ | $93.7_{\pm 9.2}$ | $92.7_{\pm 11.0}$ | $53.0_{\pm 1.0}$ |

Table 98: **Q-LINK: ID diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $89.3_{\pm 1.5}$ | $10.0_{\pm 1.7}$ | $10.0_{\pm 1.7}$ | $11.3_{\pm 2.1}$ |
| Amplitude, moderate | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $86.7_{\pm 4.0}$ | $10.0_{\pm 1.7}$ | $10.0_{\pm 1.7}$ | $14.3_{\pm 4.0}$ |
| Amplitude, strong | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $89.7_{\pm 1.5}$ | $10.0_{\pm 1.7}$ | $10.0_{\pm 1.7}$ | $10.7_{\pm 2.1}$ |
| Depolarizing, moderate | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $10.0_{\pm 1.7}$ | $10.0_{\pm 1.7}$ | $10.3_{\pm 1.5}$ |
| Depolarizing, strong | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $89.7_{\pm 1.5}$ | $10.0_{\pm 1.7}$ | $10.0_{\pm 1.7}$ | $11.7_{\pm 2.5}$ |
| Phase, moderate | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $10.0_{\pm 1.7}$ | $10.0_{\pm 1.7}$ | $10.7_{\pm 1.5}$ |
| Phase, strong | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $78.7_{\pm 6.7}$ | $10.0_{\pm 1.7}$ | $10.0_{\pm 1.7}$ | $27.0_{\pm 5.3}$ |
| Mixed, moderate | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $89.3_{\pm 1.5}$ | $10.0_{\pm 1.7}$ | $10.0_{\pm 1.7}$ | $11.7_{\pm 2.3}$ |
| Mixed, strong | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $90.0_{\pm 1.7}$ | $10.0_{\pm 1.7}$ | $10.0_{\pm 1.7}$ | $10.3_{\pm 1.5}$ |
| **Entangling VQC** | | | | | | |
| Clean | $67.0_{\pm 19.2}$ | $65.3_{\pm 20.8}$ | $49.0_{\pm 3.6}$ | $98.7_{\pm 1.5}$ | $98.0_{\pm 2.6}$ | $49.0_{\pm 3.6}$ |
| Amplitude, moderate | $51.0_{\pm 3.6}$ | $53.3_{\pm 5.0}$ | $49.0_{\pm 3.6}$ | $98.7_{\pm 1.5}$ | $98.3_{\pm 1.5}$ | $49.0_{\pm 3.6}$ |
| Amplitude, strong | $51.0_{\pm 3.6}$ | $51.0_{\pm 3.6}$ | $49.0_{\pm 3.6}$ | $96.7_{\pm 0.6}$ | $97.0_{\pm 1.0}$ | $49.0_{\pm 3.6}$ |
| Depolarizing, moderate | $59.0_{\pm 14.2}$ | $49.3_{\pm 2.9}$ | $52.0_{\pm 4.4}$ | $99.7_{\pm 0.6}$ | $99.7_{\pm 0.6}$ | $46.0_{\pm 10.6}$ |
| Depolarizing, strong | $52.7_{\pm 4.2}$ | $56.7_{\pm 3.8}$ | $49.0_{\pm 3.6}$ | $96.0_{\pm 2.6}$ | $96.0_{\pm 3.5}$ | $49.0_{\pm 3.6}$ |
| Phase, moderate | $78.3_{\pm 21.9}$ | $79.7_{\pm 19.6}$ | $47.0_{\pm 1.0}$ | $84.0_{\pm 13.9}$ | $85.3_{\pm 12.7}$ | $47.0_{\pm 1.0}$ |
| Phase, strong | $89.7_{\pm 1.5}$ | $89.7_{\pm 1.5}$ | $47.0_{\pm 1.0}$ | $34.0_{\pm 21.7}$ | $34.0_{\pm 21.7}$ | $47.0_{\pm 1.0}$ |
| Mixed, moderate | $89.3_{\pm 1.2}$ | $83.7_{\pm 4.9}$ | $49.0_{\pm 3.6}$ | $96.3_{\pm 3.1}$ | $96.3_{\pm 3.1}$ | $49.0_{\pm 3.6}$ |
| Mixed, strong | $64.7_{\pm 20.2}$ | $66.0_{\pm 19.5}$ | $52.7_{\pm 4.5}$ | $77.0_{\pm 33.2}$ | $77.0_{\pm 34.8}$ | $48.0_{\pm 5.0}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $50.7_{\pm 3.1}$ | $49.3_{\pm 4.5}$ | $51.0_{\pm 3.6}$ | $88.0_{\pm 9.0}$ | $88.7_{\pm 7.0}$ | $51.0_{\pm 3.6}$ |
| Amplitude, moderate | $52.3_{\pm 2.9}$ | $54.0_{\pm 3.0}$ | $54.3_{\pm 2.5}$ | $82.3_{\pm 22.8}$ | $82.3_{\pm 24.5}$ | $53.7_{\pm 1.5}$ |
| Amplitude, strong | $50.7_{\pm 35.6}$ | $55.3_{\pm 31.1}$ | $50.3_{\pm 3.8}$ | $86.7_{\pm 11.8}$ | $85.7_{\pm 11.4}$ | $50.3_{\pm 3.8}$ |
| Depolarizing, moderate | $54.3_{\pm 2.5}$ | $54.3_{\pm 1.5}$ | $50.7_{\pm 4.2}$ | $75.0_{\pm 5.0}$ | $77.0_{\pm 3.0}$ | $52.3_{\pm 2.9}$ |
| Depolarizing, strong | $52.0_{\pm 30.2}$ | $49.0_{\pm 10.4}$ | $53.0_{\pm 1.0}$ | $81.0_{\pm 9.0}$ | $80.7_{\pm 8.6}$ | $53.0_{\pm 1.0}$ |
| Phase, moderate | $47.3_{\pm 2.3}$ | $49.0_{\pm 3.6}$ | $47.0_{\pm 1.7}$ | $66.3_{\pm 8.5}$ | $67.0_{\pm 8.9}$ | $49.3_{\pm 4.2}$ |
| Phase, strong | $56.3_{\pm 17.9}$ | $45.0_{\pm 8.7}$ | $47.0_{\pm 1.0}$ | $60.0_{\pm 5.3}$ | $62.0_{\pm 5.6}$ | $46.0_{\pm 1.0}$ |
| Mixed, moderate | $54.3_{\pm 2.3}$ | $58.0_{\pm 6.2}$ | $52.7_{\pm 1.2}$ | $82.0_{\pm 10.0}$ | $81.7_{\pm 10.1}$ | $53.3_{\pm 1.2}$ |
| Mixed, strong | $64.0_{\pm 19.9}$ | $52.3_{\pm 3.1}$ | $53.0_{\pm 1.0}$ | $87.7_{\pm 10.1}$ | $86.3_{\pm 12.2}$ | $53.0_{\pm 1.0}$ |

Table 99: **Q-LINK: Random diagnostics.** Full competing-feature task and stable-only accuracy (%), mean ± seed SD.

| Training | Full competing-feature task | | | Stable-only | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $48.7_{\pm 4.0}$ | $46.7_{\pm 4.6}$ | $47.7_{\pm 3.8}$ | $49.0_{\pm 4.4}$ |
| Amplitude, moderate | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $49.0_{\pm 1.0}$ | $47.0_{\pm 5.6}$ | $48.3_{\pm 3.5}$ | $51.7_{\pm 7.5}$ |
| Amplitude, strong | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $48.0_{\pm 5.0}$ | $51.3_{\pm 7.0}$ | $50.0_{\pm 6.2}$ | $44.3_{\pm 8.6}$ |
| Depolarizing, moderate | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $45.7_{\pm 5.7}$ | $45.7_{\pm 5.7}$ | $45.3_{\pm 5.9}$ |
| Depolarizing, strong | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $50.0_{\pm 2.6}$ | $45.7_{\pm 5.7}$ | $46.0_{\pm 5.6}$ | $47.7_{\pm 4.0}$ |
| Phase, moderate | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $46.3_{\pm 4.9}$ | $48.3_{\pm 4.7}$ | $51.7_{\pm 7.5}$ |
| Phase, strong | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $50.0_{\pm 4.6}$ | $45.7_{\pm 5.7}$ | $48.0_{\pm 6.1}$ | $51.7_{\pm 7.5}$ |
| Mixed, moderate | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $48.7_{\pm 5.0}$ | $47.0_{\pm 6.2}$ | $46.7_{\pm 4.7}$ | $51.7_{\pm 7.5}$ |
| Mixed, strong | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $51.3_{\pm 8.0}$ | $47.3_{\pm 7.2}$ | $51.7_{\pm 7.5}$ |
| **Entangling VQC** | | | | | | |
| Clean | $98.0_{\pm 1.0}$ | $97.3_{\pm 2.1}$ | $45.7_{\pm 5.7}$ | $99.7_{\pm 0.6}$ | $99.7_{\pm 0.6}$ | $45.7_{\pm 5.7}$ |
| Amplitude, moderate | $98.3_{\pm 1.5}$ | $98.3_{\pm 1.5}$ | $45.7_{\pm 5.7}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $45.7_{\pm 5.7}$ |
| Amplitude, strong | $96.7_{\pm 2.1}$ | $96.0_{\pm 2.0}$ | $45.7_{\pm 5.7}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $45.7_{\pm 5.7}$ |
| Depolarizing, moderate | $97.3_{\pm 1.5}$ | $97.3_{\pm 2.1}$ | $45.7_{\pm 5.7}$ | $99.3_{\pm 1.2}$ | $99.3_{\pm 1.2}$ | $45.7_{\pm 5.7}$ |
| Depolarizing, strong | $94.0_{\pm 3.0}$ | $93.0_{\pm 2.6}$ | $45.7_{\pm 5.7}$ | $95.7_{\pm 3.2}$ | $94.3_{\pm 4.0}$ | $45.7_{\pm 5.7}$ |
| Phase, moderate | $89.0_{\pm 7.8}$ | $89.0_{\pm 7.2}$ | $49.7_{\pm 7.8}$ | $96.3_{\pm 3.2}$ | $96.3_{\pm 3.5}$ | $49.7_{\pm 7.8}$ |
| Phase, strong | $64.7_{\pm 19.7}$ | $64.0_{\pm 19.3}$ | $49.7_{\pm 7.8}$ | $76.3_{\pm 20.1}$ | $76.3_{\pm 20.1}$ | $49.7_{\pm 7.8}$ |
| Mixed, moderate | $96.3_{\pm 1.5}$ | $95.7_{\pm 2.1}$ | $45.7_{\pm 5.7}$ | $99.3_{\pm 1.2}$ | $99.3_{\pm 1.2}$ | $45.7_{\pm 5.7}$ |
| Mixed, strong | $83.3_{\pm 15.1}$ | $84.0_{\pm 14.0}$ | $50.0_{\pm 5.3}$ | $85.3_{\pm 14.6}$ | $84.7_{\pm 15.0}$ | $49.0_{\pm 4.4}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $87.7_{\pm 13.3}$ | $89.7_{\pm 12.1}$ | $54.3_{\pm 5.7}$ | $94.0_{\pm 7.9}$ | $93.0_{\pm 9.6}$ | $54.0_{\pm 6.2}$ |
| Amplitude, moderate | $86.7_{\pm 16.2}$ | $86.7_{\pm 17.9}$ | $50.0_{\pm 8.2}$ | $95.0_{\pm 7.8}$ | $94.3_{\pm 9.8}$ | $50.3_{\pm 7.8}$ |
| Amplitude, strong | $87.7_{\pm 4.5}$ | $87.3_{\pm 5.5}$ | $44.3_{\pm 3.5}$ | $96.3_{\pm 3.5}$ | $95.7_{\pm 3.8}$ | $44.3_{\pm 3.5}$ |
| Depolarizing, moderate | $86.3_{\pm 8.5}$ | $87.7_{\pm 9.3}$ | $51.7_{\pm 7.0}$ | $99.0_{\pm 1.0}$ | $97.7_{\pm 1.2}$ | $54.0_{\pm 5.6}$ |
| Depolarizing, strong | $89.7_{\pm 5.5}$ | $88.7_{\pm 4.7}$ | $50.3_{\pm 7.8}$ | $96.7_{\pm 5.8}$ | $96.3_{\pm 3.8}$ | $50.3_{\pm 7.8}$ |
| Phase, moderate | $77.3_{\pm 13.1}$ | $76.3_{\pm 13.7}$ | $46.0_{\pm 5.6}$ | $92.0_{\pm 9.2}$ | $92.7_{\pm 8.1}$ | $46.3_{\pm 5.5}$ |
| Phase, strong | $67.0_{\pm 6.2}$ | $65.7_{\pm 3.8}$ | $49.3_{\pm 7.6}$ | $84.7_{\pm 8.1}$ | $84.3_{\pm 8.5}$ | $49.7_{\pm 7.8}$ |
| Mixed, moderate | $93.3_{\pm 6.0}$ | $93.0_{\pm 6.0}$ | $51.0_{\pm 7.0}$ | $97.3_{\pm 4.6}$ | $97.3_{\pm 3.8}$ | $49.7_{\pm 8.6}$ |
| Mixed, strong | $88.3_{\pm 6.5}$ | $87.0_{\pm 7.0}$ | $50.3_{\pm 7.8}$ | $97.0_{\pm 5.2}$ | $95.3_{\pm 8.1}$ | $50.3_{\pm 7.8}$ |

Table 100: **Q-LINK: Random diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $52.0_{\pm 6.0}$ | $52.0_{\pm 6.0}$ | $51.7_{\pm 8.5}$ |
| Amplitude, moderate | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $49.3_{\pm 4.2}$ | $52.0_{\pm 6.0}$ | $52.0_{\pm 6.0}$ | $51.7_{\pm 7.5}$ |
| Amplitude, strong | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $49.3_{\pm 4.5}$ | $52.0_{\pm 6.0}$ | $52.0_{\pm 6.0}$ | $51.7_{\pm 7.0}$ |
| Depolarizing, moderate | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $48.3_{\pm 6.0}$ | $52.0_{\pm 6.0}$ | $52.0_{\pm 6.0}$ | $51.7_{\pm 6.0}$ |
| Depolarizing, strong | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $48.7_{\pm 4.0}$ | $52.0_{\pm 6.0}$ | $52.0_{\pm 6.0}$ | $52.0_{\pm 7.0}$ |
| Phase, moderate | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $49.0_{\pm 4.6}$ | $52.0_{\pm 6.0}$ | $52.0_{\pm 6.0}$ | $52.0_{\pm 7.0}$ |
| Phase, strong | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $51.3_{\pm 2.3}$ | $52.0_{\pm 6.0}$ | $52.0_{\pm 6.0}$ | $52.0_{\pm 6.0}$ |
| Mixed, moderate | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $48.7_{\pm 5.0}$ | $52.0_{\pm 6.0}$ | $52.0_{\pm 6.0}$ | $51.7_{\pm 6.5}$ |
| Mixed, strong | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $48.3_{\pm 5.5}$ | $52.0_{\pm 6.0}$ | $52.0_{\pm 6.0}$ | $51.7_{\pm 6.5}$ |
| **Entangling VQC** | | | | | | |
| Clean | $51.0_{\pm 8.0}$ | $47.0_{\pm 5.2}$ | $45.7_{\pm 5.7}$ | $98.0_{\pm 1.0}$ | $98.0_{\pm 1.0}$ | $45.7_{\pm 5.7}$ |
| Amplitude, moderate | $54.3_{\pm 5.7}$ | $55.3_{\pm 4.0}$ | $45.7_{\pm 5.7}$ | $98.3_{\pm 1.5}$ | $98.3_{\pm 1.5}$ | $45.7_{\pm 5.7}$ |
| Amplitude, strong | $54.3_{\pm 5.7}$ | $54.3_{\pm 5.7}$ | $45.7_{\pm 5.7}$ | $96.3_{\pm 2.5}$ | $97.3_{\pm 1.5}$ | $45.7_{\pm 5.7}$ |
| Depolarizing, moderate | $54.0_{\pm 4.4}$ | $56.7_{\pm 3.8}$ | $45.7_{\pm 5.7}$ | $97.7_{\pm 1.5}$ | $97.3_{\pm 1.5}$ | $46.0_{\pm 5.6}$ |
| Depolarizing, strong | $56.7_{\pm 4.0}$ | $55.3_{\pm 2.9}$ | $45.7_{\pm 5.7}$ | $94.0_{\pm 3.6}$ | $92.0_{\pm 2.6}$ | $45.7_{\pm 5.7}$ |
| Phase, moderate | $44.7_{\pm 3.1}$ | $48.0_{\pm 5.0}$ | $48.3_{\pm 6.4}$ | $91.0_{\pm 6.2}$ | $91.0_{\pm 5.3}$ | $49.7_{\pm 7.8}$ |
| Phase, strong | $48.0_{\pm 6.0}$ | $48.0_{\pm 6.0}$ | $49.7_{\pm 7.8}$ | $64.7_{\pm 8.3}$ | $65.0_{\pm 9.6}$ | $49.7_{\pm 7.8}$ |
| Mixed, moderate | $48.3_{\pm 5.5}$ | $49.3_{\pm 4.5}$ | $45.7_{\pm 5.7}$ | $96.0_{\pm 2.6}$ | $95.3_{\pm 2.1}$ | $45.7_{\pm 5.7}$ |
| Mixed, strong | $53.7_{\pm 5.5}$ | $53.3_{\pm 3.1}$ | $51.3_{\pm 6.4}$ | $82.7_{\pm 18.1}$ | $82.0_{\pm 17.4}$ | $46.0_{\pm 5.3}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $52.0_{\pm 7.5}$ | $48.0_{\pm 9.8}$ | $54.3_{\pm 5.7}$ | $89.0_{\pm 10.1}$ | $87.0_{\pm 12.1}$ | $54.3_{\pm 5.7}$ |
| Amplitude, moderate | $52.3_{\pm 6.5}$ | $52.7_{\pm 6.5}$ | $51.0_{\pm 7.0}$ | $87.3_{\pm 14.2}$ | $87.3_{\pm 14.2}$ | $53.7_{\pm 5.5}$ |
| Amplitude, strong | $52.0_{\pm 3.6}$ | $50.0_{\pm 4.6}$ | $44.3_{\pm 3.5}$ | $86.7_{\pm 5.5}$ | $85.3_{\pm 5.0}$ | $44.3_{\pm 3.5}$ |
| Depolarizing, moderate | $43.0_{\pm 2.0}$ | $47.3_{\pm 4.0}$ | $54.3_{\pm 5.7}$ | $88.3_{\pm 2.1}$ | $87.7_{\pm 2.5}$ | $54.3_{\pm 5.7}$ |
| Depolarizing, strong | $49.7_{\pm 5.9}$ | $48.3_{\pm 6.4}$ | $50.3_{\pm 7.8}$ | $87.3_{\pm 4.5}$ | $88.0_{\pm 7.5}$ | $50.3_{\pm 7.8}$ |
| Phase, moderate | $49.0_{\pm 7.8}$ | $49.7_{\pm 5.5}$ | $45.7_{\pm 5.7}$ | $73.7_{\pm 9.6}$ | $74.0_{\pm 8.2}$ | $48.7_{\pm 6.7}$ |
| Phase, strong | $48.0_{\pm 5.6}$ | $49.3_{\pm 4.0}$ | $49.3_{\pm 7.6}$ | $65.7_{\pm 6.0}$ | $64.7_{\pm 7.6}$ | $50.0_{\pm 7.9}$ |
| Mixed, moderate | $51.0_{\pm 8.9}$ | $52.3_{\pm 4.9}$ | $50.7_{\pm 7.4}$ | $92.7_{\pm 4.2}$ | $94.0_{\pm 5.2}$ | $51.3_{\pm 6.7}$ |
| Mixed, strong | $47.3_{\pm 3.5}$ | $49.3_{\pm 5.7}$ | $50.3_{\pm 7.8}$ | $89.3_{\pm 9.0}$ | $87.7_{\pm 8.0}$ | $50.3_{\pm 7.8}$ |

Table 101: **Q-LINK: OOD reversed diagnostics.** Stable-only accuracy (%), mean ± seed SD.

| Training | Stable-only | | |
|---|---|---|---|
| | A | F | H |
| **Local VQC** | | | |
| Clean | $51.7_{\pm 6.1}$ | $52.7_{\pm 6.5}$ | $49.0_{\pm 4.0}$ |
| Amplitude, moderate | $52.3_{\pm 9.0}$ | $52.3_{\pm 3.1}$ | $48.7_{\pm 4.0}$ |
| Amplitude, strong | $48.3_{\pm 4.2}$ | $49.3_{\pm 4.0}$ | $48.3_{\pm 4.2}$ |
| Depolarizing, moderate | $50.3_{\pm 4.6}$ | $50.3_{\pm 4.6}$ | $50.3_{\pm 4.6}$ |
| Depolarizing, strong | $50.3_{\pm 4.6}$ | $51.7_{\pm 2.3}$ | $51.0_{\pm 5.3}$ |
| Phase, moderate | $53.7_{\pm 9.0}$ | $52.3_{\pm 7.0}$ | $48.3_{\pm 4.2}$ |
| Phase, strong | $50.3_{\pm 4.6}$ | $52.0_{\pm 5.6}$ | $48.3_{\pm 4.2}$ |
| Mixed, moderate | $46.0_{\pm 7.5}$ | $48.3_{\pm 5.0}$ | $48.3_{\pm 4.2}$ |
| Mixed, strong | $57.3_{\pm 5.9}$ | $53.3_{\pm 4.5}$ | $48.3_{\pm 4.2}$ |
| **Entangling VQC** | | | |
| Clean | $100.0_{\pm 0.0}$ | $99.7_{\pm 0.6}$ | $50.3_{\pm 4.6}$ |
| Amplitude, moderate | $99.7_{\pm 0.6}$ | $100.0_{\pm 0.0}$ | $50.3_{\pm 4.6}$ |
| Amplitude, strong | $98.7_{\pm 1.5}$ | $99.0_{\pm 1.0}$ | $50.3_{\pm 4.6}$ |
| Depolarizing, moderate | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $47.7_{\pm 4.6}$ |
| Depolarizing, strong | $95.7_{\pm 2.1}$ | $96.3_{\pm 1.5}$ | $50.7_{\pm 2.1}$ |
| Phase, moderate | $98.0_{\pm 2.0}$ | $97.7_{\pm 2.5}$ | $53.7_{\pm 1.2}$ |
| Phase, strong | $70.7_{\pm 20.8}$ | $71.7_{\pm 23.5}$ | $53.7_{\pm 1.2}$ |
| Mixed, moderate | $99.3_{\pm 0.6}$ | $99.7_{\pm 0.6}$ | $50.3_{\pm 4.6}$ |
| Mixed, strong | $90.3_{\pm 9.0}$ | $89.3_{\pm 9.1}$ | $47.0_{\pm 2.6}$ |
| **Reuploading VQC** | | | |
| Clean | $95.0_{\pm 7.8}$ | $92.7_{\pm 11.8}$ | $49.7_{\pm 4.6}$ |
| Amplitude, moderate | $94.3_{\pm 9.0}$ | $94.7_{\pm 9.2}$ | $47.3_{\pm 0.6}$ |
| Amplitude, strong | $97.0_{\pm 3.6}$ | $97.0_{\pm 3.5}$ | $48.3_{\pm 4.2}$ |
| Depolarizing, moderate | $99.3_{\pm 0.6}$ | $99.3_{\pm 0.6}$ | $49.3_{\pm 4.0}$ |
| Depolarizing, strong | $95.0_{\pm 7.8}$ | $92.7_{\pm 11.8}$ | $46.3_{\pm 1.2}$ |
| Phase, moderate | $90.0_{\pm 9.8}$ | $89.7_{\pm 9.1}$ | $50.3_{\pm 3.1}$ |
| Phase, strong | $82.0_{\pm 13.9}$ | $82.3_{\pm 13.6}$ | $53.0_{\pm 2.0}$ |
| Mixed, moderate | $98.0_{\pm 2.6}$ | $98.3_{\pm 2.9}$ | $46.3_{\pm 1.2}$ |
| Mixed, strong | $96.3_{\pm 4.7}$ | $96.7_{\pm 4.2}$ | $46.3_{\pm 1.2}$ |

Table 102: **Q-LINK: OOD reversed diagnostics.** Shortcut-only and shortcut-flip accuracy (%), mean ± seed SD.

| Training | Shortcut-only | | | Shortcut-flip | | |
|---|---|---|---|---|---|---|
| | A | F | H | A | F | H |
| **Local VQC** | | | | | | |
| Clean | $8.0_{\pm 1.0}$ | $8.0_{\pm 1.0}$ | $11.3_{\pm 3.2}$ | $92.0_{\pm 1.0}$ | $92.0_{\pm 1.0}$ | $87.7_{\pm 4.2}$ |
| Amplitude, moderate | $8.0_{\pm 1.0}$ | $8.0_{\pm 1.0}$ | $8.3_{\pm 1.2}$ | $92.0_{\pm 1.0}$ | $92.0_{\pm 1.0}$ | $91.3_{\pm 0.6}$ |
| Amplitude, strong | $8.0_{\pm 1.0}$ | $8.0_{\pm 1.0}$ | $8.3_{\pm 1.2}$ | $92.0_{\pm 1.0}$ | $92.0_{\pm 1.0}$ | $90.7_{\pm 0.6}$ |
| Depolarizing, moderate | $8.0_{\pm 1.0}$ | $8.0_{\pm 1.0}$ | $8.0_{\pm 1.0}$ | $92.0_{\pm 1.0}$ | $92.0_{\pm 1.0}$ | $91.7_{\pm 0.6}$ |
| Depolarizing, strong | $8.0_{\pm 1.0}$ | $8.0_{\pm 1.0}$ | $11.3_{\pm 5.9}$ | $92.0_{\pm 1.0}$ | $92.0_{\pm 1.0}$ | $90.0_{\pm 1.0}$ |
| Phase, moderate | $8.0_{\pm 1.0}$ | $8.0_{\pm 1.0}$ | $9.3_{\pm 1.5}$ | $92.0_{\pm 1.0}$ | $92.0_{\pm 1.0}$ | $90.7_{\pm 1.5}$ |
| Phase, strong | $8.0_{\pm 1.0}$ | $8.0_{\pm 1.0}$ | $8.3_{\pm 0.6}$ | $92.0_{\pm 1.0}$ | $92.0_{\pm 1.0}$ | $91.0_{\pm 0.0}$ |
| Mixed, moderate | $8.0_{\pm 1.0}$ | $8.0_{\pm 1.0}$ | $9.7_{\pm 0.6}$ | $92.0_{\pm 1.0}$ | $92.0_{\pm 1.0}$ | $91.0_{\pm 1.0}$ |
| Mixed, strong | $8.0_{\pm 1.0}$ | $8.0_{\pm 1.0}$ | $8.0_{\pm 1.0}$ | $92.0_{\pm 1.0}$ | $92.0_{\pm 1.0}$ | $91.7_{\pm 0.6}$ |
| **Entangling VQC** | | | | | | |
| Clean | $30.7_{\pm 20.3}$ | $30.0_{\pm 14.8}$ | $50.3_{\pm 4.6}$ | $99.7_{\pm 0.6}$ | $99.7_{\pm 0.6}$ | $50.3_{\pm 4.6}$ |
| Amplitude, moderate | $49.7_{\pm 4.6}$ | $50.0_{\pm 4.6}$ | $50.3_{\pm 4.6}$ | $98.3_{\pm 1.2}$ | $98.7_{\pm 0.6}$ | $50.3_{\pm 4.6}$ |
| Amplitude, strong | $49.7_{\pm 4.6}$ | $49.7_{\pm 4.6}$ | $50.3_{\pm 4.6}$ | $95.3_{\pm 4.6}$ | $94.7_{\pm 4.9}$ | $50.3_{\pm 4.6}$ |
| Depolarizing, moderate | $41.7_{\pm 14.7}$ | $48.0_{\pm 8.5}$ | $51.3_{\pm 2.9}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $53.7_{\pm 4.0}$ |
| Depolarizing, strong | $49.7_{\pm 3.1}$ | $44.3_{\pm 2.5}$ | $50.3_{\pm 4.6}$ | $95.0_{\pm 2.0}$ | $94.3_{\pm 2.5}$ | $50.3_{\pm 4.6}$ |
| Phase, moderate | $20.3_{\pm 21.4}$ | $24.7_{\pm 28.9}$ | $53.7_{\pm 1.2}$ | $99.0_{\pm 0.0}$ | $99.0_{\pm 0.0}$ | $53.7_{\pm 1.2}$ |
| Phase, strong | $8.0_{\pm 1.0}$ | $8.0_{\pm 1.0}$ | $53.7_{\pm 1.2}$ | $94.0_{\pm 1.7}$ | $92.3_{\pm 3.1}$ | $53.3_{\pm 0.6}$ |
| Mixed, moderate | $9.7_{\pm 2.1}$ | $13.3_{\pm 3.8}$ | $50.3_{\pm 4.6}$ | $98.3_{\pm 1.2}$ | $98.7_{\pm 0.6}$ | $50.3_{\pm 4.6}$ |
| Mixed, strong | $34.0_{\pm 22.5}$ | $36.3_{\pm 24.5}$ | $48.7_{\pm 4.0}$ | $94.3_{\pm 3.5}$ | $93.7_{\pm 2.5}$ | $45.7_{\pm 2.1}$ |
| **Reuploading VQC** | | | | | | |
| Clean | $47.0_{\pm 1.0}$ | $50.0_{\pm 4.4}$ | $49.7_{\pm 4.6}$ | $94.0_{\pm 7.2}$ | $93.3_{\pm 7.4}$ | $49.7_{\pm 4.6}$ |
| Amplitude, moderate | $50.7_{\pm 3.2}$ | $48.0_{\pm 3.6}$ | $44.7_{\pm 4.0}$ | $93.0_{\pm 6.6}$ | $92.3_{\pm 7.4}$ | $48.0_{\pm 1.7}$ |
| Amplitude, strong | $48.7_{\pm 36.9}$ | $45.0_{\pm 31.7}$ | $48.3_{\pm 4.2}$ | $90.7_{\pm 4.0}$ | $90.0_{\pm 1.7}$ | $48.3_{\pm 4.2}$ |
| Depolarizing, moderate | $43.0_{\pm 5.3}$ | $47.7_{\pm 7.4}$ | $49.7_{\pm 4.6}$ | $96.3_{\pm 1.2}$ | $95.7_{\pm 1.5}$ | $49.7_{\pm 4.6}$ |
| Depolarizing, strong | $44.0_{\pm 29.5}$ | $46.7_{\pm 11.0}$ | $46.7_{\pm 0.6}$ | $92.0_{\pm 3.0}$ | $90.7_{\pm 3.5}$ | $46.7_{\pm 0.6}$ |
| Phase, moderate | $54.0_{\pm 1.7}$ | $55.7_{\pm 2.9}$ | $51.7_{\pm 2.3}$ | $78.3_{\pm 9.9}$ | $78.3_{\pm 9.9}$ | $52.0_{\pm 1.7}$ |
| Phase, strong | $38.7_{\pm 17.9}$ | $52.7_{\pm 4.5}$ | $54.0_{\pm 1.0}$ | $75.0_{\pm 7.8}$ | $75.7_{\pm 9.8}$ | $54.3_{\pm 1.2}$ |
| Mixed, moderate | $41.0_{\pm 3.6}$ | $37.3_{\pm 2.9}$ | $46.7_{\pm 0.6}$ | $98.0_{\pm 1.0}$ | $97.0_{\pm 1.0}$ | $47.7_{\pm 0.6}$ |
| Mixed, strong | $36.0_{\pm 21.9}$ | $48.0_{\pm 9.2}$ | $46.3_{\pm 1.2}$ | $83.7_{\pm 9.0}$ | $83.0_{\pm 9.0}$ | $46.3_{\pm 1.2}$ |

Table 103: **Clean-trained prediction agreement: ID.** Full competing-feature inputs; agreement (%) and logit discrepancies are mean ± seed SD.

| VQC | F–A | H–A | H–F | $d_{F,A}$ | $d_{H,A}$ |
|---|---|---|---|---|---|
| **Q-SPHERE** | | | | | |
| Local VQC | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $0.04_{\pm 0.00}$ | $1.13_{\pm 0.02}$ |
| Entangling VQC | $98.3_{\pm 0.6}$ | $50.0_{\pm 4.0}$ | $49.7_{\pm 3.5}$ | $0.21_{\pm 0.02}$ | $3.10_{\pm 0.39}$ |
| Reuploading VQC | $97.3_{\pm 0.6}$ | $55.3_{\pm 10.8}$ | $56.0_{\pm 8.7}$ | $0.23_{\pm 0.00}$ | $2.57_{\pm 0.49}$ |
| **Spirals-Signatures** | | | | | |
| Local VQC | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $0.05_{\pm 0.00}$ | $1.81_{\pm 0.06}$ |
| Entangling VQC | $99.0_{\pm 1.0}$ | $48.0_{\pm 3.6}$ | $47.0_{\pm 4.6}$ | $0.12_{\pm 0.03}$ | $2.58_{\pm 0.12}$ |
| Reuploading VQC | $98.3_{\pm 1.2}$ | $47.0_{\pm 5.6}$ | $45.3_{\pm 5.9}$ | $0.21_{\pm 0.04}$ | $2.67_{\pm 0.03}$ |
| **Spurious-Fourier** | | | | | |
| Local VQC | $99.7_{\pm 0.6}$ | $83.3_{\pm 4.9}$ | $83.0_{\pm 4.4}$ | $0.07_{\pm 0.00}$ | $2.74_{\pm 0.32}$ |
| Entangling VQC | $96.7_{\pm 1.5}$ | $50.0_{\pm 7.8}$ | $49.3_{\pm 5.9}$ | $0.13_{\pm 0.01}$ | $1.93_{\pm 0.35}$ |
| Reuploading VQC | $94.3_{\pm 0.6}$ | $51.3_{\pm 2.9}$ | $53.0_{\pm 3.0}$ | $0.29_{\pm 0.02}$ | $1.65_{\pm 0.20}$ |
| **Q-FOS** | | | | | |
| Local VQC | $99.7_{\pm 0.6}$ | $96.7_{\pm 4.9}$ | $97.0_{\pm 4.4}$ | $0.01_{\pm 0.00}$ | $0.30_{\pm 0.09}$ |
| Entangling VQC | $98.3_{\pm 1.5}$ | $66.0_{\pm 21.7}$ | $66.3_{\pm 21.5}$ | $0.05_{\pm 0.02}$ | $0.50_{\pm 0.32}$ |
| Reuploading VQC | $91.3_{\pm 1.5}$ | $66.0_{\pm 5.3}$ | $62.7_{\pm 8.1}$ | $0.12_{\pm 0.02}$ | $0.88_{\pm 0.43}$ |
| **Q-LINK** | | | | | |
| Local VQC | $100.0_{\pm 0.0}$ | $99.3_{\pm 1.2}$ | $99.3_{\pm 1.2}$ | $0.05_{\pm 0.01}$ | $1.62_{\pm 0.08}$ |
| Entangling VQC | $99.3_{\pm 1.2}$ | $48.3_{\pm 4.2}$ | $49.0_{\pm 4.0}$ | $0.30_{\pm 0.01}$ | $5.89_{\pm 0.50}$ |
| Reuploading VQC | $99.0_{\pm 1.0}$ | $49.7_{\pm 4.5}$ | $49.3_{\pm 5.1}$ | $0.34_{\pm 0.02}$ | $4.24_{\pm 1.30}$ |

Table 104: **Clean-trained prediction agreement: Random.** Full competing-feature inputs; agreement (%) and logit discrepancies are mean ± seed SD.

| VQC | F–A | H–A | H–F | $d_{F,A}$ | $d_{H,A}$ |
|---|---|---|---|---|---|
| **Q-SPHERE** | | | | | |
| Local VQC | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $0.05_{\pm 0.00}$ | $1.21_{\pm 0.05}$ |
| Entangling VQC | $98.7_{\pm 1.5}$ | $42.7_{\pm 6.4}$ | $42.0_{\pm 6.1}$ | $0.20_{\pm 0.03}$ | $3.32_{\pm 0.43}$ |
| Reuploading VQC | $95.3_{\pm 2.1}$ | $53.7_{\pm 5.9}$ | $54.3_{\pm 5.7}$ | $0.25_{\pm 0.03}$ | $2.44_{\pm 0.32}$ |
| **Spirals-Signatures** | | | | | |
| Local VQC | $100.0_{\pm 0.0}$ | $99.3_{\pm 1.2}$ | $99.3_{\pm 1.2}$ | $0.05_{\pm 0.01}$ | $1.78_{\pm 0.03}$ |
| Entangling VQC | $98.3_{\pm 2.1}$ | $45.7_{\pm 6.4}$ | $44.7_{\pm 7.2}$ | $0.13_{\pm 0.03}$ | $2.67_{\pm 0.05}$ |
| Reuploading VQC | $99.3_{\pm 1.2}$ | $46.3_{\pm 3.1}$ | $45.7_{\pm 2.3}$ | $0.20_{\pm 0.03}$ | $2.71_{\pm 0.26}$ |
| **Spurious-Fourier** | | | | | |
| Local VQC | $99.3_{\pm 0.6}$ | $85.3_{\pm 7.0}$ | $85.3_{\pm 6.0}$ | $0.07_{\pm 0.00}$ | $2.62_{\pm 0.19}$ |
| Entangling VQC | $96.0_{\pm 2.0}$ | $51.0_{\pm 4.0}$ | $51.7_{\pm 2.3}$ | $0.12_{\pm 0.01}$ | $1.85_{\pm 0.05}$ |
| Reuploading VQC | $93.3_{\pm 0.6}$ | $52.3_{\pm 12.0}$ | $52.3_{\pm 12.6}$ | $0.28_{\pm 0.00}$ | $1.64_{\pm 0.09}$ |
| **Q-FOS** | | | | | |
| Local VQC | $98.7_{\pm 1.5}$ | $89.0_{\pm 9.5}$ | $89.7_{\pm 9.2}$ | $0.02_{\pm 0.00}$ | $0.17_{\pm 0.04}$ |
| Entangling VQC | $96.0_{\pm 3.0}$ | $62.0_{\pm 20.0}$ | $62.7_{\pm 17.6}$ | $0.05_{\pm 0.01}$ | $0.41_{\pm 0.22}$ |
| Reuploading VQC | $94.3_{\pm 3.1}$ | $50.0_{\pm 7.8}$ | $47.7_{\pm 7.2}$ | $0.11_{\pm 0.01}$ | $1.11_{\pm 0.42}$ |
| **Q-LINK** | | | | | |
| Local VQC | $100.0_{\pm 0.0}$ | $98.0_{\pm 2.6}$ | $98.0_{\pm 2.6}$ | $0.05_{\pm 0.00}$ | $1.74_{\pm 0.23}$ |
| Entangling VQC | $99.3_{\pm 1.2}$ | $45.7_{\pm 4.7}$ | $45.0_{\pm 3.6}$ | $0.32_{\pm 0.03}$ | $6.16_{\pm 0.49}$ |
| Reuploading VQC | $96.7_{\pm 2.9}$ | $46.7_{\pm 13.5}$ | $48.7_{\pm 13.3}$ | $0.32_{\pm 0.04}$ | $4.45_{\pm 0.92}$ |

Table 105: **Clean-trained prediction agreement: OOD reversed.** Full competing-feature inputs; agreement (%) and logit discrepancies are mean ± seed SD.

| VQC | F–A | H–A | H–F | $d_{F,A}$ | $d_{H,A}$ |
|---|---|---|---|---|---|
| **Q-SPHERE** | | | | | |
| Local VQC | $100.0_{\pm 0.0}$ | $99.3_{\pm 0.6}$ | $99.3_{\pm 0.6}$ | $0.04_{\pm 0.00}$ | $1.25_{\pm 0.14}$ |
| Entangling VQC | $100.0_{\pm 0.0}$ | $42.7_{\pm 7.8}$ | $42.7_{\pm 7.8}$ | $0.21_{\pm 0.01}$ | $3.17_{\pm 0.68}$ |
| Reuploading VQC | $92.0_{\pm 2.0}$ | $46.3_{\pm 8.5}$ | $45.0_{\pm 8.2}$ | $0.26_{\pm 0.04}$ | $2.80_{\pm 0.39}$ |
| **Spirals-Signatures** | | | | | |
| Local VQC | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $100.0_{\pm 0.0}$ | $0.05_{\pm 0.00}$ | $1.71_{\pm 0.10}$ |
| Entangling VQC | $99.7_{\pm 0.6}$ | $49.0_{\pm 10.6}$ | $49.3_{\pm 10.4}$ | $0.12_{\pm 0.02}$ | $2.56_{\pm 0.13}$ |
| Reuploading VQC | $98.7_{\pm 0.6}$ | $47.3_{\pm 3.2}$ | $47.3_{\pm 4.0}$ | $0.21_{\pm 0.03}$ | $2.62_{\pm 0.19}$ |
| **Spurious-Fourier** | | | | | |
| Local VQC | $100.0_{\pm 0.0}$ | $81.7_{\pm 2.1}$ | $81.7_{\pm 2.1}$ | $0.07_{\pm 0.00}$ | $2.66_{\pm 0.01}$ |
| Entangling VQC | $97.0_{\pm 2.0}$ | $51.7_{\pm 3.1}$ | $52.0_{\pm 3.5}$ | $0.13_{\pm 0.01}$ | $1.81_{\pm 0.21}$ |
| Reuploading VQC | $93.3_{\pm 3.2}$ | $46.7_{\pm 9.6}$ | $46.0_{\pm 8.9}$ | $0.29_{\pm 0.02}$ | $1.86_{\pm 0.05}$ |
| **Q-FOS** | | | | | |
| Local VQC | $97.7_{\pm 2.1}$ | $84.3_{\pm 3.2}$ | $84.0_{\pm 3.6}$ | $0.02_{\pm 0.00}$ | $0.16_{\pm 0.02}$ |
| Entangling VQC | $95.7_{\pm 2.1}$ | $65.7_{\pm 15.3}$ | $66.0_{\pm 17.3}$ | $0.05_{\pm 0.02}$ | $0.39_{\pm 0.23}$ |
| Reuploading VQC | $93.0_{\pm 5.0}$ | $47.7_{\pm 9.1}$ | $47.3_{\pm 11.9}$ | $0.11_{\pm 0.02}$ | $1.06_{\pm 0.46}$ |
| **Q-LINK** | | | | | |
| Local VQC | $100.0_{\pm 0.0}$ | $90.7_{\pm 9.5}$ | $90.7_{\pm 9.5}$ | $0.05_{\pm 0.00}$ | $1.92_{\pm 0.17}$ |
| Entangling VQC | $99.3_{\pm 1.2}$ | $50.3_{\pm 4.7}$ | $49.7_{\pm 4.0}$ | $0.31_{\pm 0.01}$ | $5.76_{\pm 0.16}$ |
| Reuploading VQC | $98.0_{\pm 1.0}$ | $37.7_{\pm 10.0}$ | $37.0_{\pm 8.7}$ | $0.38_{\pm 0.04}$ | $4.30_{\pm 0.98}$ |

### D.5 Paired changes from clean training

Tables report noise-aware minus clean-trained OOD reversed accuracy on full competing-feature inputs, separately by task, architecture, and execution setting. Entries are mean ± SD of within-seed differences in percentage points; positive values denote gains. Each comparison includes three matched seeds.

Local Spurious-Fourier gains $10.7 \pm 4.2$ MonarQ points with strong phase-aware training, versus $0.0 \pm 0.0$ analytically; mixed moderate training gains $10.3 \pm 3.5$ hardware points with little analytic change. Analytic accuracy alone misses transfer differences, but gains vary across architectures and tasks. These descriptive checkpoint-group comparisons do not separate training from hardware-session effects.

Table 106: **Q-SPHERE: change from clean training.** OOD reversed, full competing-feature inputs: paired mean ± SD over three seeds, in percentage points.

| Training | A | F | H |
|---|---|---|---|
| **Local VQC** | | | |
| Amplitude, moderate | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 1.0}$ |
| Amplitude, strong | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+1.0_{\pm 2.0}$ |
| Depolarizing, moderate | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+2.0_{\pm 0.0}$ |
| Depolarizing, strong | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+0.7_{\pm 1.5}$ |
| Phase, moderate | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 1.0}$ |
| Phase, strong | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 1.0}$ |
| Mixed, moderate | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 1.0}$ |
| Mixed, strong | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 1.0}$ |
| **Entangling VQC** | | | |
| Amplitude, moderate | $+6.3_{\pm 1.5}$ | $+6.7_{\pm 3.5}$ | $+0.0_{\pm 0.0}$ |
| Amplitude, strong | $+20.0_{\pm 17.1}$ | $+18.7_{\pm 17.5}$ | $-5.0_{\pm 8.7}$ |
| Depolarizing, moderate | $+3.7_{\pm 2.5}$ | $+4.0_{\pm 2.6}$ | $+0.0_{\pm 0.0}$ |
| Depolarizing, strong | $+4.0_{\pm 1.7}$ | $+4.3_{\pm 2.9}$ | $+0.0_{\pm 0.0}$ |
| Phase, moderate | $-0.3_{\pm 1.5}$ | $+0.3_{\pm 0.6}$ | $+0.0_{\pm 0.0}$ |
| Phase, strong | $+1.3_{\pm 1.5}$ | $+2.7_{\pm 1.5}$ | $+0.0_{\pm 0.0}$ |
| Mixed, moderate | $+0.3_{\pm 1.2}$ | $+0.7_{\pm 1.2}$ | $+0.0_{\pm 0.0}$ |
| Mixed, strong | $+0.7_{\pm 1.5}$ | $+0.7_{\pm 1.5}$ | $-10.0_{\pm 10.0}$ |
| **Reuploading VQC** | | | |
| Amplitude, moderate | $-7.3_{\pm 8.7}$ | $-5.0_{\pm 7.2}$ | $+0.0_{\pm 0.0}$ |
| Amplitude, strong | $-1.0_{\pm 10.8}$ | $+2.0_{\pm 7.2}$ | $+0.0_{\pm 0.0}$ |
| Depolarizing, moderate | $+5.0_{\pm 11.1}$ | $+8.0_{\pm 13.5}$ | $-10.0_{\pm 9.2}$ |
| Depolarizing, strong | $+21.7_{\pm 6.8}$ | $+24.0_{\pm 4.6}$ | $-6.3_{\pm 11.0}$ |
| Phase, moderate | $+16.3_{\pm 6.0}$ | $+21.7_{\pm 6.7}$ | $+0.0_{\pm 0.0}$ |
| Phase, strong | $+17.7_{\pm 5.7}$ | $+19.7_{\pm 9.0}$ | $+0.0_{\pm 0.0}$ |
| Mixed, moderate | $+10.3_{\pm 8.5}$ | $+12.3_{\pm 10.4}$ | $+0.0_{\pm 0.0}$ |
| Mixed, strong | $+1.3_{\pm 19.1}$ | $+5.7_{\pm 17.0}$ | $+0.7_{\pm 1.2}$ |

Table 107: **Spirals-Signatures: change from clean training.** OOD reversed, full competing-feature inputs: paired mean ± SD over three seeds, in percentage points.

| Training | A | F | H |
|---|---|---|---|
| **Local VQC** | | | |
| Amplitude, moderate | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ |
| Amplitude, strong | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ |
| Depolarizing, moderate | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+0.3_{\pm 0.6}$ |
| Depolarizing, strong | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ |
| Phase, moderate | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ |
| Phase, strong | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ |
| Mixed, moderate | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+0.3_{\pm 0.6}$ |
| Mixed, strong | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+0.3_{\pm 0.6}$ |
| **Entangling VQC** | | | |
| Amplitude, moderate | $+6.7_{\pm 5.5}$ | $+7.0_{\pm 5.6}$ | $+0.0_{\pm 0.0}$ |
| Amplitude, strong | $+14.3_{\pm 6.1}$ | $+14.3_{\pm 7.5}$ | $+0.0_{\pm 0.0}$ |
| Depolarizing, moderate | $+1.7_{\pm 2.3}$ | $+1.0_{\pm 2.6}$ | $+0.0_{\pm 0.0}$ |
| Depolarizing, strong | $+2.3_{\pm 7.6}$ | $+2.3_{\pm 7.8}$ | $+0.0_{\pm 0.0}$ |
| Phase, moderate | $+0.3_{\pm 5.0}$ | $-0.7_{\pm 4.0}$ | $+0.0_{\pm 0.0}$ |
| Phase, strong | $+4.0_{\pm 5.2}$ | $+4.3_{\pm 6.1}$ | $+0.0_{\pm 0.0}$ |
| Mixed, moderate | $+4.7_{\pm 8.1}$ | $+4.0_{\pm 8.7}$ | $+0.0_{\pm 0.0}$ |
| Mixed, strong | $+4.0_{\pm 5.3}$ | $+4.0_{\pm 5.6}$ | $+2.3_{\pm 4.0}$ |
| **Reuploading VQC** | | | |
| Amplitude, moderate | $+2.3_{\pm 7.4}$ | $+4.0_{\pm 6.6}$ | $+0.0_{\pm 0.0}$ |
| Amplitude, strong | $+8.7_{\pm 12.9}$ | $+7.3_{\pm 11.6}$ | $+0.0_{\pm 0.0}$ |
| Depolarizing, moderate | $-2.3_{\pm 5.9}$ | $-1.7_{\pm 5.0}$ | $-3.0_{\pm 5.2}$ |
| Depolarizing, strong | $+8.3_{\pm 14.8}$ | $+8.7_{\pm 14.0}$ | $+4.0_{\pm 6.9}$ |
| Phase, moderate | $+4.7_{\pm 4.0}$ | $+5.0_{\pm 4.4}$ | $+8.7_{\pm 7.6}$ |
| Phase, strong | $+8.7_{\pm 7.5}$ | $+7.7_{\pm 5.0}$ | $+8.7_{\pm 7.6}$ |
| Mixed, moderate | $+1.0_{\pm 7.0}$ | $+0.7_{\pm 5.1}$ | $+0.7_{\pm 1.2}$ |
| Mixed, strong | $-3.0_{\pm 6.1}$ | $-1.7_{\pm 5.7}$ | $+0.0_{\pm 0.0}$ |

Table 108: **Spurious-Fourier: change from clean training.** OOD reversed, full competing-feature inputs: paired mean ± SD over three seeds, in percentage points.

| Training | A | F | H |
|---|---|---|---|
| **Local VQC** | | | |
| Amplitude, moderate | $-0.7_{\pm 1.2}$ | $-1.0_{\pm 1.0}$ | $+4.3_{\pm 4.0}$ |
| Amplitude, strong | $-0.7_{\pm 1.2}$ | $-1.0_{\pm 1.7}$ | $+8.0_{\pm 6.6}$ |
| Depolarizing, moderate | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $+6.0_{\pm 2.6}$ |
| Depolarizing, strong | $-0.3_{\pm 0.6}$ | $+0.0_{\pm 1.0}$ | $+1.0_{\pm 2.6}$ |
| Phase, moderate | $+0.3_{\pm 0.6}$ | $+0.7_{\pm 0.6}$ | $+6.3_{\pm 7.4}$ |
| Phase, strong | $+0.0_{\pm 0.0}$ | $+0.3_{\pm 0.6}$ | $+10.7_{\pm 4.2}$ |
| Mixed, moderate | $-0.3_{\pm 1.5}$ | $-0.7_{\pm 1.2}$ | $+10.3_{\pm 3.5}$ |
| Mixed, strong | $-0.3_{\pm 1.5}$ | $-0.3_{\pm 1.5}$ | $+3.7_{\pm 10.6}$ |
| **Entangling VQC** | | | |
| Amplitude, moderate | $-9.3_{\pm 3.1}$ | $-11.0_{\pm 2.6}$ | $+0.0_{\pm 0.0}$ |
| Amplitude, strong | $-20.0_{\pm 2.0}$ | $-21.7_{\pm 1.2}$ | $+3.3_{\pm 5.5}$ |
| Depolarizing, moderate | $-7.3_{\pm 4.0}$ | $-7.0_{\pm 5.3}$ | $+0.0_{\pm 0.0}$ |
| Depolarizing, strong | $+11.7_{\pm 5.0}$ | $+10.3_{\pm 4.5}$ | $-0.3_{\pm 0.6}$ |
| Phase, moderate | $+0.3_{\pm 1.5}$ | $-3.0_{\pm 3.6}$ | $+0.0_{\pm 0.0}$ |
| Phase, strong | $-2.0_{\pm 2.0}$ | $-3.7_{\pm 3.1}$ | $+0.0_{\pm 0.0}$ |
| Mixed, moderate | $-4.7_{\pm 1.5}$ | $-4.7_{\pm 2.1}$ | $+0.0_{\pm 0.0}$ |
| Mixed, strong | $-9.0_{\pm 6.6}$ | $-9.7_{\pm 6.7}$ | $+0.7_{\pm 5.5}$ |
| **Reuploading VQC** | | | |
| Amplitude, moderate | $-3.0_{\pm 9.6}$ | $-1.3_{\pm 7.2}$ | $-2.7_{\pm 3.8}$ |
| Amplitude, strong | $-6.3_{\pm 11.4}$ | $-6.7_{\pm 9.8}$ | $+2.7_{\pm 11.2}$ |
| Depolarizing, moderate | $-9.0_{\pm 2.6}$ | $-9.0_{\pm 1.0}$ | $-0.3_{\pm 4.0}$ |
| Depolarizing, strong | $-10.7_{\pm 9.7}$ | $-10.3_{\pm 7.0}$ | $-1.3_{\pm 0.6}$ |
| Phase, moderate | $-6.0_{\pm 3.6}$ | $-4.0_{\pm 2.0}$ | $-0.7_{\pm 0.6}$ |
| Phase, strong | $+2.7_{\pm 14.4}$ | $+4.0_{\pm 11.0}$ | $+5.3_{\pm 10.1}$ |
| Mixed, moderate | $+0.7_{\pm 10.8}$ | $+0.3_{\pm 6.7}$ | $-1.3_{\pm 7.5}$ |
| Mixed, strong | $-2.0_{\pm 2.6}$ | $-1.0_{\pm 4.6}$ | $+3.7_{\pm 11.0}$ |

Table 109: **Q-FOS: change from clean training.** OOD reversed, full competing-feature inputs: paired mean ± SD over three seeds, in percentage points.

| Training | A | F | H |
|---|---|---|---|
| **Local VQC** | | | |
| Amplitude, moderate | $-1.0_{\pm 0.0}$ | $-0.3_{\pm 0.6}$ | $-2.3_{\pm 3.1}$ |
| Amplitude, strong | $-5.3_{\pm 5.9}$ | $-4.7_{\pm 4.0}$ | $-7.7_{\pm 5.0}$ |
| Depolarizing, moderate | $+2.3_{\pm 1.2}$ | $+3.3_{\pm 4.0}$ | $-0.7_{\pm 7.2}$ |
| Depolarizing, strong | $+2.0_{\pm 4.4}$ | $+2.0_{\pm 3.5}$ | $+2.0_{\pm 6.1}$ |
| Phase, moderate | $-2.0_{\pm 4.0}$ | $-1.0_{\pm 4.0}$ | $-2.0_{\pm 4.0}$ |
| Phase, strong | $-1.0_{\pm 2.6}$ | $-0.7_{\pm 3.5}$ | $+1.7_{\pm 5.0}$ |
| Mixed, moderate | $+1.0_{\pm 2.0}$ | $+2.0_{\pm 2.6}$ | $-5.7_{\pm 1.5}$ |
| Mixed, strong | $+2.7_{\pm 4.0}$ | $+4.3_{\pm 4.7}$ | $-1.3_{\pm 6.7}$ |
| **Entangling VQC** | | | |
| Amplitude, moderate | $-7.0_{\pm 1.7}$ | $-7.3_{\pm 2.3}$ | $-0.3_{\pm 0.6}$ |
| Amplitude, strong | $-9.3_{\pm 6.7}$ | $-8.0_{\pm 10.6}$ | $-1.0_{\pm 6.6}$ |
| Depolarizing, moderate | $-8.3_{\pm 4.0}$ | $-8.0_{\pm 5.6}$ | $-0.7_{\pm 6.4}$ |
| Depolarizing, strong | $-4.0_{\pm 6.9}$ | $-4.7_{\pm 6.7}$ | $+1.7_{\pm 10.0}$ |
| Phase, moderate | $-2.0_{\pm 7.2}$ | $-0.7_{\pm 5.5}$ | $+1.3_{\pm 10.1}$ |
| Phase, strong | $+0.0_{\pm 9.5}$ | $-1.7_{\pm 9.0}$ | $+2.7_{\pm 10.1}$ |
| Mixed, moderate | $+5.0_{\pm 6.1}$ | $+3.7_{\pm 5.5}$ | $+2.7_{\pm 10.1}$ |
| Mixed, strong | $+4.7_{\pm 5.7}$ | $+5.7_{\pm 5.0}$ | $+2.7_{\pm 10.1}$ |
| **Reuploading VQC** | | | |
| Amplitude, moderate | $+12.7_{\pm 15.1}$ | $+12.3_{\pm 13.3}$ | $-2.3_{\pm 2.1}$ |
| Amplitude, strong | $+3.0_{\pm 1.0}$ | $+4.3_{\pm 2.1}$ | $-1.3_{\pm 2.3}$ |
| Depolarizing, moderate | $+3.0_{\pm 6.6}$ | $+2.0_{\pm 3.0}$ | $-4.0_{\pm 5.6}$ |
| Depolarizing, strong | $-4.3_{\pm 4.0}$ | $-3.7_{\pm 2.5}$ | $+0.3_{\pm 4.5}$ |
| Phase, moderate | $+5.3_{\pm 10.8}$ | $+6.3_{\pm 10.5}$ | $-2.7_{\pm 4.6}$ |
| Phase, strong | $+7.0_{\pm 0.0}$ | $+6.0_{\pm 1.7}$ | $-2.3_{\pm 4.9}$ |
| Mixed, moderate | $+5.7_{\pm 13.2}$ | $+5.7_{\pm 14.4}$ | $-4.0_{\pm 4.0}$ |
| Mixed, strong | $+5.0_{\pm 7.8}$ | $+4.3_{\pm 7.1}$ | $-2.0_{\pm 5.3}$ |

Table 110: **Q-LINK: change from clean training.** OOD reversed, full competing-feature inputs: paired mean ± SD over three seeds, in percentage points.

| Training | A | F | H |
|---|---|---|---|
| **Local VQC** | | | |
| Amplitude, moderate | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $-6.3_{\pm 7.8}$ |
| Amplitude, strong | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $-5.0_{\pm 7.2}$ |
| Depolarizing, moderate | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $-6.7_{\pm 7.6}$ |
| Depolarizing, strong | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $-4.3_{\pm 5.9}$ |
| Phase, moderate | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $-6.0_{\pm 6.6}$ |
| Phase, strong | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $-6.7_{\pm 7.6}$ |
| Mixed, moderate | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $-3.3_{\pm 4.9}$ |
| Mixed, strong | $+0.0_{\pm 0.0}$ | $+0.0_{\pm 0.0}$ | $-6.0_{\pm 7.2}$ |
| **Entangling VQC** | | | |
| Amplitude, moderate | $-0.3_{\pm 2.1}$ | $-1.3_{\pm 4.5}$ | $+0.0_{\pm 0.0}$ |
| Amplitude, strong | $-3.7_{\pm 5.7}$ | $-3.7_{\pm 6.0}$ | $+0.0_{\pm 0.0}$ |
| Depolarizing, moderate | $+1.3_{\pm 1.5}$ | $+1.3_{\pm 1.5}$ | $+0.7_{\pm 3.5}$ |
| Depolarizing, strong | $-3.7_{\pm 3.2}$ | $-3.3_{\pm 3.1}$ | $+0.0_{\pm 0.0}$ |
| Phase, moderate | $-15.7_{\pm 12.9}$ | $-14.7_{\pm 12.1}$ | $+3.3_{\pm 5.8}$ |
| Phase, strong | $-66.3_{\pm 20.3}$ | $-67.0_{\pm 19.7}$ | $+3.3_{\pm 5.8}$ |
| Mixed, moderate | $-3.0_{\pm 4.0}$ | $-3.7_{\pm 3.5}$ | $+0.0_{\pm 0.0}$ |
| Mixed, strong | $-23.3_{\pm 32.9}$ | $-22.7_{\pm 31.9}$ | $+0.7_{\pm 1.2}$ |
| **Reuploading VQC** | | | |
| Amplitude, moderate | $-1.0_{\pm 21.5}$ | $-1.3_{\pm 19.0}$ | $-1.7_{\pm 2.9}$ |
| Amplitude, strong | $-0.3_{\pm 35.3}$ | $-1.3_{\pm 35.3}$ | $-1.3_{\pm 8.1}$ |
| Depolarizing, moderate | $-10.0_{\pm 12.3}$ | $-11.3_{\pm 10.1}$ | $+0.7_{\pm 1.2}$ |
| Depolarizing, strong | $-3.0_{\pm 15.5}$ | $-5.0_{\pm 17.5}$ | $-3.3_{\pm 5.8}$ |
| Phase, moderate | $-13.7_{\pm 20.2}$ | $-17.7_{\pm 15.9}$ | $+0.7_{\pm 9.2}$ |
| Phase, strong | $-23.3_{\pm 13.6}$ | $-22.3_{\pm 13.6}$ | $+4.3_{\pm 3.8}$ |
| Mixed, moderate | $-1.0_{\pm 5.6}$ | $-1.7_{\pm 5.5}$ | $-1.0_{\pm 1.7}$ |
| Mixed, strong | $+5.3_{\pm 22.5}$ | $+4.7_{\pm 21.9}$ | $-3.3_{\pm 5.8}$ |

### D.6 Coverage and raw-record validation

All 405 runs contribute 4860 split/intervention conditions. Every reported statistic uses all three seeds $\{0, 1, 2\}$, with 100 unique examples per seed and condition. Coverage is checked across task, architecture, training condition, split, and input mode.

For the 10 runs containing overlapping executions, raw-record checks verify example identity, execution settings, labels, completion status, numerical outputs, shot counts, and agreement with the sample manifest. Repeated successful evaluations refer to the same example, settings, labels, and noiseless output. Each example contributes its earliest fully validated successful evaluation by recorded timestamp, with CSV row order breaking timestamp ties. Selection does not depend on prediction accuracy. The same rule is applied throughout; repeated attempts do not increase sample size.

Condition summaries for these 10 runs are recomputed from the selected records; the remaining runs retain their previously validated summaries. Matching conditions and seeds are used for analytic, finite-shot, and hardware comparisons and for paired changes from clean training. No values are imputed, and no seed or condition is excluded from the reported results.

# E Realistic-task definitions and preparation

The 10 tasks in Section 5 combine constructed feature competition and native domain shifts on identical frozen inputs for all learners. The target is $Y \in \{0,1\}$, the constructed or annotated shortcut $C \in \{0,1\}$, and environment $E$. Constructed cues have agreement $p_e = \Pr(C = Y \mid E = e)$; hospital and year metadata identify native domains without assumed binary shortcut labels.

## E.1 Task definitions and design rationale

### E.1.1 Appearance: semantic targets and appearance cues

**Q-Colored-MNIST.** Following the colored-image design (Arjovsky et al., 2019), MNIST (LeCun et al., 1998) digit $D \in \{0,\dots,9\}$ defines $Y = \mathbf{1}\{D \geq 5\}$ without added label noise. Cue $C$ agrees with $Y$ with probability $p_e$: grayscale intensity occupies the red channel for $C = 0$ or green for $C = 1$, with other channels zero. Original training and test images are pooled and randomly repartitioned into source A, source B, ID validation, Random, and OOD reversed with fractions $(0.35, 0.35, 0.10, 0.10, 0.10)$ and agreements $(0.90, 0.80, 0.85, 0.50, 0.10)$, respectively. These project-specific partitions replace the original MNIST split.

**Q-Colored-Fashion-MNIST.** Fashion-MNIST (Xiao et al., 2017) uses the same construction, red/green cue, split fractions, and agreements, with clothing-class indices 0–4 mapped to $Y = 0$ and 5–9 to $Y = 1$.

**Waterbirds.** Waterbirds (Sagawa et al., 2020) supplies bird-type targets $Y$ and background cues $C$. Native training, validation, and test assignments are retained as `env_A_train`, `pooled_id_val`, and `ood_natural`. No Random split or artificial reversal is introduced; target–background combinations define four evaluation groups.

### E.1.2 Frequency: spatial and spectral representations

**MNIST-Fourier.** MNIST-Fourier uses the same binary digit target, image repartitioning, and shortcut agreements as Q-Colored-MNIST, but replaces color with an additive sinusoidal pattern. For normalized grayscale image $g(u,v) \in [0,1]$, where $u$ and $v$ are horizontal and vertical pixel coordinates, respectively, the observed image is

$$\begin{aligned} \widetilde{g}(u,v) &= \mathrm{clip}_{[0,1]}\left[g(u,v) + 0.35\, h_C(u,v)\right], \\ h_0(u,v) &= \sin(2\pi u/4), \qquad h_1(u,v) = \sin(2\pi v/7). \end{aligned} \tag{20}$$

The cue changes orientation and spatial frequency. The fixed descriptor concatenates a flattened $32\times32$ grayscale image and its flattened $\log(1+|\,\mathrm{rFFT}_2\,|)$, giving $1024+544 = 1568$ coordinates before shared training-only reduction.

**WOODS Spurious-Fourier.** We retain the WOODS signal construction and environment parameters (Gagnon-Audet et al., 2023), adding a fixed spectral descriptor. Each environment has 4000 signals, balanced in latent stable label $S \in \{0,1\}$. The target is $Y = S \oplus N$ with $N \sim \mathrm{Bernoulli}(0.25)$; $C$ agrees with $Y$ with probability $p_e \in \{0.10, 0.80, 0.90\}$. Stable frequencies are $f_S = 9\,\mathrm{Hz}$ for $S = 0$ and $7\,\mathrm{Hz}$ for $S = 1$; cue frequencies are $f_C = 2\,\mathrm{Hz}$ for $C = 0$ and $4\,\mathrm{Hz}$ for $C = 1$. Their pre-normalization Fourier amplitudes are 1 and 0.5, respectively. Each $(S,C)$ template is generated by an inverse real FFT of length 10000 at 100 Hz, normalized by its maximum, and partitioned into 50-sample windows at offsets $0, 2, \dots, 48$. Randomly permuted template-window pools supply examples.

A 20% holdout gives 3200 in-split and 800 out-split examples per environment. The $p_e = 0.80, 0.90$ in-splits provide 6400 training examples; their out-splits form 1600 pooled ID validation examples. The entire $p_e = 0.10$ environment is excluded from training: its `ood_test_in` (3200) and `ood_test_out` (800) partitions are both OOD, and neither is Random. Here *in/out* denotes within-environment partitioning.

The descriptor takes $\log(1 + |\,\mathrm{rFFT}_{50}\,|)$ at the first 24 positive-frequency bins $(2, 4, \dots, 48\,\mathrm{Hz})$. Coordinates are standardized on the 6400 training examples, clipped to $[-3, 3]$, and mapped to $[-\pi, \pi]$. Nested prefixes of length 6, 12, and 24 cover 2–12, 2–24, and 2–48 Hz, respectively; no PCA or label-based frequency selection is used.

#### E.1.3 Temporal: ordered relations and changing color cues

**Temporal Colored-MNIST: common construction.** Adapting WOODS temporal-colored-MNIST (Gagnon-Audet et al., 2023), we group consecutive MNIST images into 4-digit sequences, then permute complete sequences. At prediction times $t \in \{1, 2, 3\}$, the stable relation and target are

$$S_t = (D_{t-1} + D_t) \bmod 2, \qquad Y_t = S_t \oplus N_t, \qquad N_t \sim \text{Bernoulli}(0.25). \tag{21}$$

Current-frame red/green color $C_t$ agrees with $Y_t$ with probability $p_e$; the first frame is grayscale. Ordered previous/current inputs preserve the stable parity relation and current color cue. Target and shortcut-assignment noise are distinct.

**TCMNIST-Source.** Complete sequences form three approximately equal domains with $p_e \in \{0.90, 0.80, 0.10\}$. The first two provide training and the third OOD reversed evaluation. Each training domain holds out 20% of complete sequences for pooled ID validation. All three prediction times are used with constant within-domain agreement. Sequence-level splitting keeps training, ID, and OOD sequences disjoint.

**TCMNIST-Time.** Agreement varies over time: $p_1 = 0.90$, $p_2 = 0.80$, $p_3 = 0.10$. Complete sequences are split 80/20 into training/held-out sets. Training uses $t = 1, 2$; held-out sequences supply ID validation at $t = 1, 2$ and OOD reversed at $t = 3$. Training and evaluation sequences are disjoint; ID and OOD evaluation share held-out sequences at different times. Neither temporal variant has a Random split.

#### E.1.4 Domain: native hospital and historical shifts

**Camelyon17.** Camelyon17-WILDS (Koh et al., 2021) retains its binary tumor target and hospital metadata. Official training and ID validation remain `train` and `id_val`; official validation/test become `ood_val`/`ood_test`, representing held-out hospitals. No Random or reversed-cue condition is introduced. Hospital identifiers specify acquisition domains, not known causal or spurious image features.

**Yearbook.** We retain Yearbook's released binary portrait target (Ginosar et al., 2015) and Wild-Time Eval-Fix protocol (Yao et al., 2022). Years 1930–1970 form the ID period and 1971–2013 the OOD period. Within each ID year, released partitions 0 and 1 provide training and held-out ID examples. Only partition 1 of future years supplies OOD evaluation; future partition 0 is never used for training. Partition 2 contains the full per-year collection and serves only source-integrity checks. All released temporal coverage is retained without subsampling to a nominal size. Year is domain metadata; the shortcut-field compatibility value $-1$ means *unavailable* and is excluded from shortcut-recovery tests.

#### E.1.5 Relational: object relations and a local color cue

**Relational-OOD CLEVR (RO-CLEVR).** From pooled CLEVR training/validation scenes (Johnson et al., 2017), we retain scenes with a cube and sphere. Let $A$ and $B$ be the leftmost cube and sphere by horizontal pixel coordinates $x_A, x_B$; $A$ must be red or blue. Our adapted target and cue are

$$Y = \mathbf{1}\{x_A < x_B\}, \qquad C = \mathbf{1}\{\text{color}(A) = \text{blue}\}. \tag{22}$$

The stable signal is an ordered object relation and the cue is cube color. Scenes are selected without recoloring. Source A, source B, ID validation, Random, and OOD reversed receive fractions $(0.30, 0.30, 0.10, 0.15, 0.15)$ and agreements $(0.90, 0.80, 0.85, 0.50, 0.10)$, respectively.

Joint allocation balances both target classes within every split. Each 400-scene block allocates $(120, 120, 40, 60, 60)$ examples, consuming 137 cue-agreeing and 63 cue-disagreeing examples per target class. Available $(Y, C)$ counts limit the number of blocks; each cell is shuffled once and allocated by disjoint slices, without scene reuse or cell substitution. Object coordinates also define the color descriptor in Appendix E.2.4. Scene-level relations alone do not establish a quantum readout restriction after compression.

## E.2 Preparation and evaluation protocol

### E.2.1 Partitions and sample identity

Table 111 lists partitions: ID is held-out source validation, Random/reversal apply only where constructed, and native tests retain their meaning. Temporal splits are sequence-level; preprocessing fitting excludes all evaluation rows.

Table 111: Split protocols. A/B: constructed source environments. Evaluation partitions are excluded from preprocessing fitting and downstream training.

| Tasks | Training | ID validation | OOD evaluation |
|---|---|---|---|
| Q-Colored-MNIST, Q-Colored-Fashion-MNIST, MNIST-Fourier, RO-CLEVR | A/B | Pooled source holdout | Random; reversed |
| Waterbirds | Official train | Official validation | Official test |
| TCMNIST-Source | Source A/B | Held-out sequences | Reversed |
| TCMNIST-Time | Times $t = 1, 2$ | Held-out sequences, $t = 1, 2$ | Same held-out sequences, $t = 3$ |
| WOODS Spurious-Fourier | $p_e = 0.8, 0.9$ in-splits | Source out-splits | Both partitions of $p_e = 0.1$ |
| Camelyon17 | Official train | Official ID validation | Official OOD validation; test |
| Yearbook | ID-year partition 0 | ID-year partition 1 | Future-year partition 1 |

Manifests record source identifiers, targets, environments, splits, and available group/relation metadata; frozen arrays preserve row order and labels. Manifest, configuration, and array fingerprints record lineage. Learners share arrays and splits per task/dimension; diagnostic labels are not inputs.

### E.2.2 Feature extraction

Frozen DINOv2 ViT-S/14 (Oquab et al., 2023) runs in evaluation mode without gradient updates. RGB inputs are resized bicubically to $224 \times 224$ and normalized by means $(0.485, 0.456, 0.406)$ and standard deviations $(0.229, 0.224, 0.225)$. Colored-MNIST, Colored-Fashion-MNIST, Waterbirds, and Camelyon17 use the global token. RO-CLEVR and Yearbook concatenate it with an ordered $4 \times 4$ grid of adaptively averaged patch tokens, retaining position before compression. Temporal inputs use separate global descriptors per frame. MNIST-Fourier and WOODS Spurious-Fourier use the spatial-plus-spectral and spectral-only descriptors above, respectively.

### E.2.3 Training-only scaling and nested dimensions

Following related QML benchmarks (Grant et al., 2018; Bowles et al., 2024), dimension reduction enables one-feature-per-qubit experiments. Our chosen $d \in \{6, 12, 24\}$ compare frozen representations, not unrestricted learning on raw inputs; these dimensions are not prescribed by those references.

For single-branch PCA, a source-training-fitted StandardScaler transforms fixed descriptor $h(x)$ into $\widetilde{h}(x)$. Unwhitened PCA (Jolliffe & Cadima, 2016) retains 24 leading components, each restandardized using source-training rows only. For resulting component $z_j(x)$, the encoded coordinate is

$$q_j(x) = \frac{\pi}{3} \operatorname{clip}_{[-3,3]} z_j(x), \qquad x^{(d)} = (q_1(x), \ldots, q_d(x)). \tag{23}$$

PCA uses full singular-value decomposition; neither PCA nor scaling uses target/shortcut labels. Dimensions 6 and 12 are prefixes of the same 24-dimensional representation, with identical shared coordinates in $[-\pi, \pi]$.

**Temporal representation.** Frozen DINOv2 global features are extracted separately from previous/current $28 \times 28$ frame crops resized to $224 \times 224$. Each branch uses source-training rows only for independent standardization, 12-component PCA, and component standardization, followed by clipping to $[-3, 3]$ and scaling by $\pi/3$. Interleaving $(z_1^{\text{prev}}, z_1^{\text{curr}}, z_2^{\text{prev}}, z_2^{\text{curr}}, \ldots)$ provides nested $d \in \{6, 12, 24\}$ prefixes with $d/2$ coordinates per frame. Neither target nor shortcut labels fit the transformations.

**Spectral exception.** WOODS Spurious-Fourier applies training-fitted scaling and Equation 23's clipping map directly to its 24 ordered log-magnitude bins; prefixes select increasing frequency ranges. MNIST-Fourier applies shared PCA to its 1568 spatial-plus-spectral coordinates.

#### E.2.4 Fixed local color coordinate for RO-CLEVR

RO-CLEVR prepends a fixed local color statistic to spatial PCA features. A radius-24 crop centered on the cube's rounded CLEVR pixel coordinates is $49 \times 49$ except at image boundaries; an inscribed circular mask selects pixels. Scene metadata supplies only crop geometry; image pixels, not color-name or class-label fields, supply color.

For RGB pixel $u$ with channels in $[0, 1]$, define saturation $s_u$ and normalized weight $w_u$ by

$$\begin{aligned} s_u &= \frac{\max(R_u, G_u, B_u) - \min(R_u, G_u, B_u)}{\max\{\max(R_u, G_u, B_u), 10^{-6}\}}, \\ w_u &= \frac{s_u^2 + 10^{-4}}{\sum_v (s_v^2 + 10^{-4})}, \qquad c_{\text{raw}} = \sum_u w_u (B_u - R_u). \end{aligned} \tag{24}$$

Source-training standardization and the same clipping rule map the statistic to $[-\pi, \pi]$. Denoting it by $c(x)$ and joint spatial PCA coordinates by $q_j(x)$, the final input is

$$x^{(d)} = \big(c(x), q_1(x), \ldots, q_{d-1}(x)\big). \tag{25}$$

Dimensions 6, 12, and 24 contain one color and 5, 11, and 23 spatial PCA coordinates. Inputs are frozen across learners and exactly nested; additional diagnostic color statistics are excluded.

#### E.2.5 Preparation checks and frozen comparison

Checks verify source identifiers, labels, domains, splits, row counts, finite values, dimensions, nested prefixes, and provenance fingerprints. Constructed-cue tasks check class support, stable–shortcut cells, and specified agreements; native tasks check released partitions.

Separate from downstream model selection, class-balanced logistic probes are fitted on source-training rows for available stable, shortcut, target, or relation labels. Held-out recovery requires balanced chance plus 0.05 by default and at least 20 examples. Declared agreements allow discrepancy $\max\{0.05, 3\sqrt{p_e(1-p_e)/n_e}\}$ for environment size $n_e$. Unavailable annotations are excluded.

**Additional WOODS validation.** Stable and shortcut probes balance the four stable–shortcut cells, fit on source-training rows, and evaluate separately on source holdouts and the complete excluded environment: 3 dimensions $\times$ 2 targets $\times$ 2 evaluation domains give 12 gates. Matched quartets contain all four combinations at a common waveform-offset index; a 70/30 split of quartet identifiers gives 6 further gates using the frozen spectral scaler. The audit uses 2000 quartets and 20–2000 examples per selected joint cell.

Each gate uses 5 seeds, 500 bootstrap resamples, 499 permutations, and at least 100 evaluation examples. Acceptance requires all seeds to complete, mean balanced accuracy $\geq 0.55$, every seed's 95% bootstrap lower bound $> 0.5$, maximum permutation $p \leq 0.01$, and mean performance at least 0.05 above a shuffled-training-label control. Source-integrity, representation, and exact split audits are also required.

**Additional Yearbook validation.** Yearbook has no shortcut-recovery gate. Pooled ID and exact-year-conditioned probes test target retention at all 3 dimensions, fitting on ID training and assessing readiness on ID validation. Exact-year probes require both labels, at least 2 examples per year–label cell, 30 eligible years, and 300 test examples. Sampling caps each cell at 50 without forcing label balance. Training weights years equally; evaluation uses macro-over-year balanced accuracy. Bootstraps resample years; permutations and shuffled controls act within years. The audit uses 5 seeds, 500 bootstraps, 499 permutations, and the same 0.05 effect/control margins, above-chance confidence bounds, and $p \leq 0.01$ criteria. Future-year target metrics and year prediction are separate diagnostics, not readiness gates.

These audits concern input recoverability under probes, which does not imply accessibility to every downstream architecture.

**Downstream interface.** Logistic regression, a small MLP, and local, entangling, and reuploading VQCs share arrays and partitions. Input dimension equals qubit count; VQCs share a 2-layer family and angle-encoding interface, use seeds $\{0, 1, 2\}$, and train for 300 steps. Evaluation checkpoints are $\{1, 50, 100, 150, 200, 250, 300\}$ for $d \in \{6, 12\}$ and $\{1, 100, 200, 300\}$ for $d = 24$. Classical models are fitted once per task–dimension–seed. Hyperparameters/model choices use training or ID validation only; checkpoint schedules exclude OOD-based selection.

The primary metric is binary balanced accuracy,

$$\mathrm{BA} = \frac{1}{2} \sum_{y \in \{0,1\}} \frac{\sum_i \mathbf{1}\{Y_i = y, \widehat{Y}_i = y\}}{\sum_i \mathbf{1}\{Y_i = y\}}, \tag{26}$$

with ordinary accuracy and annotated group metrics also recorded. Across-seed summaries separate tasks, dimensions, and architectures. The baseline size is $d = 6$; $d \in \{12, 24\}$ tests scaling of nested inputs. The main tables display the largest tested size, $d = 24$; results at $d \in \{6, 12\}$ appear in Appendix E.3. Device access and budget limit this suite to simulation; transfer of all observed behavior to MonarQ hardware is not assumed.

### E.3 Detailed simulation results

Tables supplement Table 4 with $d \in \{6, 12\}$ and additional $d = 24$ held-out conditions: balanced accuracy (%), mean ± sample SD over 3 seeds, separately by architecture. VQCs use 300 updates without OOD-based checkpoint selection. Dimension changes both input coordinates and VQC width, so comparisons concern pipelines rather than isolated qubit-count effects.

#### E.3.1 Appearance family

At both lower dimensions (Table 112), Q-Colored-MNIST has a large reversal penalty, Q-Colored-Fashion-MNIST retains stronger reversal accuracy, and Waterbirds has similar aggregate ID/test performance. More coordinates particularly help the Q-Colored-MNIST MLP; VQC changes are smaller and architecture-dependent.

Table 112: Appearance tasks at $d \in \{6, 12\}$: held-out balanced accuracy (%), mean ± seed SD. For $d = 24$, see Table 4.

| Task | $d$ | Split | Logistic regression | MLP | Local VQC | Entangling VQC | Reuploading VQC |
|---|---|---|---|---|---|---|---|
| Q-Colored-MNIST | 6 | Pooled ID | $84.6_{\pm 0.9}$ | $85.9_{\pm 0.3}$ | $83.1_{\pm 1.4}$ | $81.2_{\pm 1.2}$ | $79.9_{\pm 0.9}$ |
| | | Random | $50.8_{\pm 1.0}$ | $59.7_{\pm 2.5}$ | $55.7_{\pm 2.9}$ | $59.6_{\pm 1.8}$ | $58.9_{\pm 0.9}$ |
| | | OOD reversed | $12.9_{\pm 2.0}$ | $29.0_{\pm 5.1}$ | $25.7_{\pm 3.1}$ | $33.1_{\pm 2.6}$ | $36.9_{\pm 4.0}$ |
| | 12 | Pooled ID | $84.7_{\pm 1.1}$ | $91.5_{\pm 1.0}$ | $82.9_{\pm 1.6}$ | $82.3_{\pm 1.0}$ | $76.3_{\pm 0.6}$ |
| | | Random | $52.7_{\pm 0.8}$ | $75.5_{\pm 5.1}$ | $55.9_{\pm 0.7}$ | $59.4_{\pm 0.7}$ | $57.7_{\pm 1.2}$ |
| | | OOD reversed | $17.4_{\pm 1.7}$ | $55.9_{\pm 6.6}$ | $26.4_{\pm 2.1}$ | $34.9_{\pm 0.5}$ | $38.1_{\pm 1.4}$ |
| Q-Colored-Fashion-MNIST | 6 | Pooled ID | $90.4_{\pm 0.8}$ | $91.6_{\pm 0.9}$ | $90.0_{\pm 1.2}$ | $90.0_{\pm 0.3}$ | $90.0_{\pm 0.5}$ |
| | | Random | $84.7_{\pm 1.4}$ | $86.4_{\pm 0.4}$ | $86.7_{\pm 1.9}$ | $83.6_{\pm 1.0}$ | $85.0_{\pm 1.8}$ |
| | | OOD reversed | $77.3_{\pm 2.6}$ | $79.7_{\pm 3.0}$ | $80.5_{\pm 2.9}$ | $76.6_{\pm 2.9}$ | $79.2_{\pm 1.0}$ |
| | 12 | Pooled ID | $93.5_{\pm 0.5}$ | $94.1_{\pm 0.3}$ | $91.7_{\pm 0.6}$ | $88.3_{\pm 1.5}$ | $87.1_{\pm 0.6}$ |
| | | Random | $85.8_{\pm 0.8}$ | $87.0_{\pm 1.8}$ | $84.3_{\pm 3.5}$ | $83.3_{\pm 1.2}$ | $82.3_{\pm 1.6}$ |
| | | OOD reversed | $76.2_{\pm 2.4}$ | $77.4_{\pm 3.3}$ | $76.3_{\pm 6.0}$ | $75.7_{\pm 2.4}$ | $77.5_{\pm 2.2}$ |
| Waterbirds | 6 | Pooled ID | $83.0_{\pm 0.4}$ | $81.5_{\pm 1.5}$ | $83.3_{\pm 1.5}$ | $80.5_{\pm 3.5}$ | $77.5_{\pm 2.4}$ |
| | | OOD test | $84.0_{\pm 0.9}$ | $83.1_{\pm 1.1}$ | $84.3_{\pm 1.3}$ | $80.9_{\pm 3.1}$ | $77.3_{\pm 1.9}$ |
| | 12 | Pooled ID | $84.1_{\pm 0.5}$ | $83.5_{\pm 0.5}$ | $85.8_{\pm 0.6}$ | $81.8_{\pm 0.7}$ | $75.2_{\pm 2.9}$ |
| | | OOD test | $84.3_{\pm 0.5}$ | $84.0_{\pm 0.9}$ | $85.4_{\pm 2.3}$ | $82.6_{\pm 1.7}$ | $74.6_{\pm 2.5}$ |

#### E.3.2 Frequency family

Increasing $d$ from 6 to 12 strongly improves the MNIST-Fourier MLP, with architecture-dependent VQC changes (Table 113). WOODS Spurious-Fourier retains high ID accuracy but performs poorly on both partitions of the excluded environment; neither partition is ID.

#### E.3.3 Temporal family

Both tasks retain large ID/reversal contrasts (Table 114). Entangling/reuploading reversal accuracy can rise with dimension while ID accuracy falls. Movement toward chance alone does not establish use of stable temporal information.

Table 113: Frequency tasks at $d \in \{6, 12\}$: held-out balanced accuracy (%), mean ± seed SD. For $d = 24$, see Table 4.

| Task | $d$ | Split | Logistic regression | MLP | Local VQC | Entangling VQC | Reuploading VQC |
|---|---|---|---|---|---|---|---|
| MNIST-Fourier | 6 | Pooled ID | $86.4_{\pm 0.8}$ | $88.2_{\pm 1.5}$ | $87.2_{\pm 1.5}$ | $86.0_{\pm 2.1}$ | $84.4_{\pm 1.2}$ |
| | | Random | $55.1_{\pm 1.6}$ | $63.2_{\pm 2.9}$ | $61.3_{\pm 2.6}$ | $61.6_{\pm 3.9}$ | $55.8_{\pm 3.1}$ |
| | | OOD reversed | $20.9_{\pm 3.8}$ | $35.8_{\pm 8.7}$ | $31.5_{\pm 6.5}$ | $35.6_{\pm 9.3}$ | $25.2_{\pm 8.7}$ |
| | 12 | Pooled ID | $86.6_{\pm 1.2}$ | $92.1_{\pm 0.7}$ | $87.1_{\pm 1.0}$ | $85.5_{\pm 0.6}$ | $84.4_{\pm 0.9}$ |
| | | Random | $59.8_{\pm 1.3}$ | $77.4_{\pm 1.5}$ | $63.8_{\pm 1.1}$ | $58.3_{\pm 1.9}$ | $55.6_{\pm 2.0}$ |
| | | OOD reversed | $28.8_{\pm 5.1}$ | $64.9_{\pm 0.4}$ | $37.0_{\pm 4.0}$ | $28.5_{\pm 3.0}$ | $21.8_{\pm 4.2}$ |
| WOODS Spurious-Fourier | 6 | Pooled ID | $85.6_{\pm 0.9}$ | $85.5_{\pm 0.3}$ | $85.1_{\pm 1.4}$ | $84.4_{\pm 1.2}$ | $83.1_{\pm 1.7}$ |
| | | Excluded env.: in-partition | $18.2_{\pm 1.8}$ | $17.2_{\pm 3.0}$ | $21.6_{\pm 2.7}$ | $25.4_{\pm 5.5}$ | $30.1_{\pm 7.6}$ |
| | | Excluded env.: out-partition | $16.8_{\pm 0.6}$ | $15.9_{\pm 3.3}$ | $20.0_{\pm 2.9}$ | $24.1_{\pm 6.3}$ | $26.7_{\pm 9.0}$ |
| | 12 | Pooled ID | $85.7_{\pm 0.8}$ | $84.7_{\pm 0.5}$ | $86.1_{\pm 1.2}$ | $85.2_{\pm 2.1}$ | $85.7_{\pm 1.2}$ |
| | | Excluded env.: in-partition | $18.9_{\pm 1.8}$ | $24.2_{\pm 3.7}$ | $15.0_{\pm 4.8}$ | $20.6_{\pm 9.0}$ | $16.8_{\pm 5.2}$ |
| | | Excluded env.: out-partition | $17.0_{\pm 0.7}$ | $22.0_{\pm 5.0}$ | $14.0_{\pm 3.8}$ | $19.5_{\pm 8.1}$ | $15.4_{\pm 4.9}$ |

Table 114: Temporal tasks at $d \in \{6, 12\}$: held-out balanced accuracy (%), mean ± seed SD. For $d = 24$, see Table 4.

| Task | $d$ | Split | Logistic regression | MLP | Local VQC | Entangling VQC | Reuploading VQC |
|---|---|---|---|---|---|---|---|
| TCMNIST-Source | 6 | Pooled ID | $83.3_{\pm 0.5}$ | $83.0_{\pm 0.6}$ | $80.4_{\pm 1.0}$ | $75.1_{\pm 1.1}$ | $73.8_{\pm 2.3}$ |
| | | OOD reversed | $12.0_{\pm 1.5}$ | $12.2_{\pm 1.5}$ | $15.3_{\pm 2.4}$ | $21.0_{\pm 2.9}$ | $24.7_{\pm 2.7}$ |
| | 12 | Pooled ID | $85.0_{\pm 0.1}$ | $84.5_{\pm 0.2}$ | $82.2_{\pm 1.2}$ | $62.3_{\pm 5.4}$ | $64.9_{\pm 3.4}$ |
| | | OOD reversed | $10.0_{\pm 1.2}$ | $10.6_{\pm 1.2}$ | $12.8_{\pm 1.9}$ | $35.5_{\pm 7.8}$ | $32.1_{\pm 3.6}$ |
| TCMNIST-Time | 6 | Pooled ID | $83.4_{\pm 0.3}$ | $82.7_{\pm 0.8}$ | $81.7_{\pm 0.5}$ | $76.7_{\pm 2.7}$ | $71.5_{\pm 2.0}$ |
| | | OOD reversed | $11.9_{\pm 0.9}$ | $12.6_{\pm 1.1}$ | $13.3_{\pm 1.5}$ | $18.4_{\pm 1.4}$ | $24.4_{\pm 1.8}$ |
| | 12 | Pooled ID | $85.4_{\pm 0.4}$ | $84.8_{\pm 0.5}$ | $83.6_{\pm 0.7}$ | $69.2_{\pm 2.1}$ | $63.5_{\pm 2.4}$ |
| | | OOD reversed | $10.2_{\pm 0.6}$ | $10.9_{\pm 0.5}$ | $11.8_{\pm 1.2}$ | $26.7_{\pm 2.8}$ | $34.0_{\pm 3.6}$ |

### E.3.4 Domain family

Camelyon17's OOD validation and test differ substantially in difficulty; test accuracy sometimes exceeds ID accuracy (Table 115). These native conditions are neither interchangeable nor shortcut reversals. Yearbook consistently loses accuracy on future years, with dimension effects varying by architecture.

Table 115: Domain tasks at $d \in \{6, 12\}$: held-out balanced accuracy (%), mean ± seed SD. For $d = 24$, see Table 4.

| Task | $d$ | Split | Logistic regression | MLP | Local VQC | Entangling VQC | Reuploading VQC |
|---|---|---|---|---|---|---|---|
| Camelyon17 | 6 | ID validation | $89.5_{\pm 1.0}$ | $90.7_{\pm 1.5}$ | $90.2_{\pm 1.3}$ | $86.3_{\pm 1.9}$ | $84.4_{\pm 0.7}$ |
| | | OOD validation | $80.6_{\pm 0.9}$ | $81.1_{\pm 0.2}$ | $79.7_{\pm 1.4}$ | $73.9_{\pm 1.4}$ | $73.8_{\pm 1.1}$ |
| | | OOD test | $92.2_{\pm 1.0}$ | $91.3_{\pm 0.2}$ | $90.3_{\pm 0.6}$ | $85.3_{\pm 3.6}$ | $85.2_{\pm 1.1}$ |
| | 12 | ID validation | $90.9_{\pm 1.2}$ | $91.6_{\pm 1.4}$ | $89.5_{\pm 3.0}$ | $89.5_{\pm 2.3}$ | $82.3_{\pm 1.8}$ |
| | | OOD validation | $75.5_{\pm 0.7}$ | $78.1_{\pm 1.0}$ | $76.4_{\pm 1.2}$ | $75.2_{\pm 2.8}$ | $74.2_{\pm 2.1}$ |
| | | OOD test | $91.0_{\pm 0.7}$ | $91.3_{\pm 0.3}$ | $88.5_{\pm 3.0}$ | $86.8_{\pm 0.9}$ | $81.0_{\pm 1.7}$ |
| Yearbook | 6 | Pooled ID | $97.3_{\pm 0.2}$ | $96.1_{\pm 1.2}$ | $96.9_{\pm 0.6}$ | $92.2_{\pm 1.9}$ | $91.0_{\pm 0.2}$ |
| | | Future-year test | $84.4_{\pm 0.5}$ | $81.7_{\pm 2.5}$ | $84.6_{\pm 2.0}$ | $80.1_{\pm 1.8}$ | $79.0_{\pm 1.3}$ |
| | 12 | Pooled ID | $97.8_{\pm 0.1}$ | $96.5_{\pm 0.8}$ | $96.5_{\pm 0.4}$ | $92.7_{\pm 2.6}$ | $87.7_{\pm 2.1}$ |
| | | Future-year test | $83.8_{\pm 0.6}$ | $80.5_{\pm 1.9}$ | $83.2_{\pm 3.3}$ | $82.8_{\pm 2.8}$ | $77.4_{\pm 2.2}$ |

### E.3.5 Relational family

RO-CLEVR combines high ID, poor reversal, and near-chance VQC Random accuracy (Table 116). Entangling operations alone therefore do not ensure stable relational learning under these inputs and training budgets.

### E.3.6 Additional held-out conditions

Table 117 adds Random splits, the second WOODS excluded-environment partition, and Camelyon17 OOD validation. Random and reversed conditions remain distinct, as do native validation/test shifts. Training data never substitute for an undefined Random split.

Table 116: Relational tasks at $d \in \{6, 12\}$: held-out balanced accuracy (%), mean ± seed SD. For $d = 24$, see Table 4.

| Task | $d$ | Split | Logistic regression | MLP | Local VQC | Entangling VQC | Reuploading VQC |
|---|---|---|---|---|---|---|---|
| RO-CLEVR | 6 | Pooled ID | $84.2_{\pm 0.4}$ | $84.4_{\pm 0.4}$ | $83.9_{\pm 0.6}$ | $72.6_{\pm 1.8}$ | $83.1_{\pm 0.2}$ |
| | | Random | $50.5_{\pm 0.5}$ | $51.3_{\pm 0.6}$ | $50.1_{\pm 0.4}$ | $52.4_{\pm 1.6}$ | $50.1_{\pm 0.4}$ |
| | | OOD reversed | $12.5_{\pm 1.2}$ | $13.2_{\pm 1.0}$ | $11.5_{\pm 1.8}$ | $28.3_{\pm 2.5}$ | $13.4_{\pm 3.8}$ |
| | 12 | Pooled ID | $84.5_{\pm 0.6}$ | $83.6_{\pm 1.2}$ | $84.0_{\pm 1.0}$ | $79.3_{\pm 2.9}$ | $83.2_{\pm 0.8}$ |
| | | Random | $52.8_{\pm 0.5}$ | $54.8_{\pm 2.0}$ | $51.0_{\pm 0.7}$ | $48.7_{\pm 0.5}$ | $49.8_{\pm 0.8}$ |
| | | OOD reversed | $16.3_{\pm 0.4}$ | $21.7_{\pm 6.1}$ | $13.1_{\pm 2.4}$ | $16.8_{\pm 3.0}$ | $12.7_{\pm 1.4}$ |

Table 117: Additional $d = 24$ conditions complementing Table 4: balanced accuracy (%), mean ± seed SD. Both WOODS partitions are OOD.

| Task | $d$ | Split | Logistic regression | MLP | Local VQC | Entangling VQC | Reuploading VQC |
|---|---|---|---|---|---|---|---|
| Q-Colored-MNIST | 24 | Random | $64.4_{\pm 0.7}$ | $76.5_{\pm 5.2}$ | $61.6_{\pm 1.4}$ | $54.9_{\pm 9.1}$ | $53.8_{\pm 3.4}$ |
| Q-Colored-Fashion-MNIST | 24 | Random | $87.3_{\pm 1.2}$ | $88.1_{\pm 0.7}$ | $83.5_{\pm 6.8}$ | $73.3_{\pm 12.8}$ | $75.3_{\pm 8.1}$ |
| MNIST-Fourier | 24 | Random | $68.8_{\pm 2.2}$ | $85.0_{\pm 4.0}$ | $63.2_{\pm 1.9}$ | $52.7_{\pm 2.0}$ | $54.6_{\pm 6.1}$ |
| WOODS Spurious-Fourier | 24 | Excluded env.: in-partition | $18.8_{\pm 1.4}$ | $18.0_{\pm 4.7}$ | $35.7_{\pm 21.9}$ | $29.0_{\pm 12.1}$ | $37.6_{\pm 21.0}$ |
| Camelyon17 | 24 | OOD validation | $74.8_{\pm 0.8}$ | $74.6_{\pm 3.0}$ | $76.6_{\pm 1.2}$ | $72.9_{\pm 4.1}$ | $73.3_{\pm 2.6}$ |
| RO-CLEVR | 24 | Random | $67.1_{\pm 1.8}$ | $68.7_{\pm 2.3}$ | $55.2_{\pm 2.9}$ | $50.1_{\pm 1.6}$ | $50.2_{\pm 0.7}$ |

Table 118: Appearance: worst-group accuracy (%), mean ± seed SD. Multi-group conditions and task-specific annotations follow Appendix E.4.

| Task | $d$ | Split | Logistic regression | MLP | Local VQC | Entangling VQC | Reuploading VQC |
|---|---|---|---|---|---|---|---|
| Q-Colored-MNIST | 6 | Pooled ID | $1.4_{\pm 1.4}$ | $8.9_{\pm 4.7}$ | $13.9_{\pm 4.9}$ | $12.1_{\pm 3.3}$ | $23.7_{\pm 11.4}$ |
| | | Random | $1.4_{\pm 0.9}$ | $12.5_{\pm 4.2}$ | $13.0_{\pm 6.5}$ | $17.4_{\pm 4.4}$ | $21.5_{\pm 3.6}$ |
| | | OOD reversed | $0.9_{\pm 0.5}$ | $10.3_{\pm 5.2}$ | $16.5_{\pm 4.1}$ | $14.9_{\pm 1.8}$ | $23.4_{\pm 6.3}$ |
| | 12 | Pooled ID | $3.7_{\pm 2.1}$ | $47.7_{\pm 15.8}$ | $17.6_{\pm 3.4}$ | $13.5_{\pm 2.4}$ | $14.8_{\pm 4.0}$ |
| | | Random | $3.5_{\pm 2.0}$ | $48.0_{\pm 10.8}$ | $9.1_{\pm 3.4}$ | $13.4_{\pm 5.2}$ | $14.9_{\pm 5.4}$ |
| | | OOD reversed | $5.0_{\pm 1.7}$ | $46.4_{\pm 7.4}$ | $11.1_{\pm 3.5}$ | $12.8_{\pm 5.6}$ | $15.3_{\pm 5.9}$ |
| | 24 | Pooled ID | $27.7_{\pm 5.6}$ | $54.3_{\pm 10.6}$ | $24.5_{\pm 3.7}$ | $17.8_{\pm 10.9}$ | $20.1_{\pm 15.2}$ |
| | | Random | $27.6_{\pm 5.1}$ | $52.5_{\pm 11.9}$ | $20.6_{\pm 7.9}$ | $17.8_{\pm 10.2}$ | $17.6_{\pm 13.2}$ |
| | | OOD reversed | $24.9_{\pm 3.8}$ | $50.2_{\pm 13.9}$ | $18.9_{\pm 5.5}$ | $19.5_{\pm 11.8}$ | $17.2_{\pm 12.4}$ |
| Q-Colored-Fashion-MNIST | 6 | Pooled ID | $68.6_{\pm 6.3}$ | $72.9_{\pm 2.7}$ | $75.9_{\pm 3.0}$ | $74.6_{\pm 5.3}$ | $75.0_{\pm 1.6}$ |
| | | Random | $74.0_{\pm 0.6}$ | $75.6_{\pm 4.8}$ | $79.4_{\pm 3.4}$ | $71.8_{\pm 6.1}$ | $76.8_{\pm 4.5}$ |
| | | OOD reversed | $72.6_{\pm 3.5}$ | $74.7_{\pm 4.6}$ | $75.8_{\pm 1.3}$ | $72.2_{\pm 5.1}$ | $76.2_{\pm 3.2}$ |
| | 12 | Pooled ID | $65.9_{\pm 7.6}$ | $65.4_{\pm 8.2}$ | $63.7_{\pm 14.7}$ | $71.5_{\pm 3.9}$ | $72.8_{\pm 1.5}$ |
| | | Random | $70.7_{\pm 0.7}$ | $71.0_{\pm 4.6}$ | $69.9_{\pm 13.7}$ | $72.2_{\pm 3.5}$ | $74.4_{\pm 3.6}$ |
| | | OOD reversed | $68.1_{\pm 3.4}$ | $69.1_{\pm 4.1}$ | $70.6_{\pm 10.9}$ | $69.6_{\pm 4.9}$ | $71.7_{\pm 4.2}$ |
| | 24 | Pooled ID | $68.1_{\pm 6.8}$ | $69.0_{\pm 1.8}$ | $68.1_{\pm 8.8}$ | $56.2_{\pm 14.4}$ | $55.5_{\pm 18.9}$ |
| | | Random | $74.5_{\pm 1.3}$ | $76.2_{\pm 1.6}$ | $69.7_{\pm 15.0}$ | $58.1_{\pm 19.7}$ | $59.6_{\pm 20.9}$ |
| | | OOD reversed | $71.3_{\pm 2.3}$ | $71.9_{\pm 2.7}$ | $70.7_{\pm 10.9}$ | $56.8_{\pm 18.1}$ | $61.6_{\pm 20.4}$ |
| Waterbirds | 6 | Pooled ID | $61.8_{\pm 2.0}$ | $56.3_{\pm 5.0}$ | $62.3_{\pm 3.0}$ | $51.3_{\pm 9.2}$ | $46.1_{\pm 5.7}$ |
| | | OOD test | $61.8_{\pm 3.7}$ | $56.9_{\pm 4.1}$ | $64.4_{\pm 2.4}$ | $53.1_{\pm 8.5}$ | $43.2_{\pm 6.7}$ |
| | 12 | Pooled ID | $64.0_{\pm 2.9}$ | $60.2_{\pm 2.6}$ | $69.0_{\pm 5.5}$ | $54.2_{\pm 4.5}$ | $43.9_{\pm 8.2}$ |
| | | OOD test | $63.5_{\pm 1.8}$ | $58.7_{\pm 1.7}$ | $66.2_{\pm 10.5}$ | $60.2_{\pm 4.1}$ | $42.7_{\pm 7.2}$ |
| | 24 | Pooled ID | $63.2_{\pm 3.5}$ | $65.2_{\pm 6.6}$ | $50.3_{\pm 10.2}$ | $50.4_{\pm 9.8}$ | $17.2_{\pm 10.9}$ |
| | | OOD test | $63.8_{\pm 2.7}$ | $62.9_{\pm 4.0}$ | $55.3_{\pm 5.9}$ | $52.6_{\pm 12.2}$ | $17.8_{\pm 12.1}$ |

### E.4 Group-level evaluation

Worst-group accuracy takes each seed's minimum annotated-group accuracy and reports mean ± sample SD over 3 seeds. Appearance, frequency, temporal, and relational evaluations each have 4 groups. Yearbook has 41 pooled-ID and 43 future-year groups; Camelyon17 has 3 ID-validation groups and 1 per OOD split. Single-group conditions are omitted. Group metrics can expose failures hidden by class-balanced accuracy; task-specific partitions preclude meaningful cross-task comparisons.

**Appearance and frequency.** Tables 118 and 119 show that strong aggregate Waterbirds accuracy can conceal poor worst-group performance; MNIST-Fourier VQCs can fit ID while failing on an annotated group.

Table 119: Frequency: worst-group accuracy (%), mean ± seed SD. Multi-group conditions and task-specific annotations follow Appendix E.4.

| Task | $d$ | Split | Logistic regression | MLP | Local VQC | Entangling VQC | Reuploading VQC |
|---|---|---|---|---|---|---|---|
| MNIST-Fourier | 6 | Pooled ID | $6.1_{\pm 3.4}$ | $19.2_{\pm 13.9}$ | $20.6_{\pm 4.1}$ | $14.4_{\pm 6.2}$ | $12.6_{\pm 11.9}$ |
| | | Random | $5.6_{\pm 2.4}$ | $22.7_{\pm 10.1}$ | $22.4_{\pm 3.4}$ | $17.9_{\pm 10.8}$ | $11.7_{\pm 9.2}$ |
| | | OOD reversed | $6.0_{\pm 2.6}$ | $22.6_{\pm 13.4}$ | $23.7_{\pm 5.2}$ | $17.2_{\pm 10.8}$ | $15.0_{\pm 11.9}$ |
| | 12 | Pooled ID | $9.3_{\pm 4.6}$ | $50.9_{\pm 3.6}$ | $19.4_{\pm 9.5}$ | $0.9_{\pm 1.6}$ | $0.9_{\pm 1.6}$ |
| | | Random | $14.8_{\pm 1.3}$ | $55.1_{\pm 3.3}$ | $19.7_{\pm 9.3}$ | $2.4_{\pm 2.5}$ | $0.4_{\pm 0.4}$ |
| | | OOD reversed | $12.0_{\pm 3.8}$ | $60.6_{\pm 0.7}$ | $20.5_{\pm 10.2}$ | $1.5_{\pm 1.4}$ | $0.2_{\pm 0.4}$ |
| | 24 | Pooled ID | $34.3_{\pm 3.7}$ | $67.7_{\pm 7.7}$ | $17.0_{\pm 11.7}$ | $2.8_{\pm 3.6}$ | $5.9_{\pm 8.1}$ |
| | | Random | $38.3_{\pm 2.2}$ | $71.3_{\pm 7.8}$ | $17.6_{\pm 3.0}$ | $3.1_{\pm 4.3}$ | $9.3_{\pm 11.3}$ |
| | | OOD reversed | $37.5_{\pm 2.5}$ | $72.6_{\pm 3.8}$ | $17.7_{\pm 5.2}$ | $3.8_{\pm 6.1}$ | $8.8_{\pm 11.6}$ |
| WOODS Spurious-Fourier | 6 | Pooled ID | $0.0_{\pm 0.0}$ | $3.6_{\pm 6.3}$ | $3.6_{\pm 6.3}$ | $5.7_{\pm 9.9}$ | $8.7_{\pm 15.1}$ |
| | | Excluded env.: in-partition | $0.0_{\pm 0.0}$ | $3.5_{\pm 6.1}$ | $1.9_{\pm 3.4}$ | $4.9_{\pm 8.5}$ | $9.4_{\pm 16.2}$ |
| | | Excluded env.: out-partition | $0.0_{\pm 0.0}$ | $3.3_{\pm 5.6}$ | $2.3_{\pm 4.1}$ | $5.3_{\pm 9.2}$ | $7.6_{\pm 13.1}$ |
| | 12 | Pooled ID | $0.0_{\pm 0.0}$ | $4.8_{\pm 1.9}$ | $0.0_{\pm 0.0}$ | $2.5_{\pm 4.3}$ | $0.0_{\pm 0.0}$ |
| | | Excluded env.: in-partition | $0.0_{\pm 0.0}$ | $3.2_{\pm 2.3}$ | $0.0_{\pm 0.0}$ | $3.4_{\pm 5.8}$ | $0.0_{\pm 0.0}$ |
| | | Excluded env.: out-partition | $0.0_{\pm 0.0}$ | $3.7_{\pm 1.5}$ | $0.0_{\pm 0.0}$ | $3.6_{\pm 6.3}$ | $0.0_{\pm 0.0}$ |
| | 24 | Pooled ID | $0.0_{\pm 0.0}$ | $2.3_{\pm 3.9}$ | $13.5_{\pm 21.0}$ | $11.1_{\pm 6.7}$ | $17.4_{\pm 30.2}$ |
| | | Excluded env.: in-partition | $0.0_{\pm 0.0}$ | $2.0_{\pm 3.4}$ | $14.8_{\pm 23.4}$ | $9.2_{\pm 9.5}$ | $16.0_{\pm 27.8}$ |
| | | Excluded env.: out-partition | $0.0_{\pm 0.0}$ | $1.8_{\pm 3.1}$ | $14.0_{\pm 21.9}$ | $7.8_{\pm 7.0}$ | $14.1_{\pm 24.5}$ |

**Temporal and relational.** Tables 120 and 121 expose severe group errors beyond the aggregate ID/reversal contrast, without directly identifying internal feature use.

Table 120: Temporal: worst-group accuracy (%), mean ± seed SD. Multi-group conditions and task-specific annotations follow Appendix E.4.

| Task | $d$ | Split | Logistic regression | MLP | Local VQC | Entangling VQC | Reuploading VQC |
|---|---|---|---|---|---|---|---|
| TCMNIST-Source | 6 | Pooled ID | $1.9_{\pm 1.0}$ | $1.9_{\pm 1.0}$ | $3.0_{\pm 3.1}$ | $13.8_{\pm 2.4}$ | $16.1_{\pm 3.6}$ |
| | | OOD reversed | $2.3_{\pm 0.4}$ | $2.4_{\pm 0.8}$ | $2.5_{\pm 1.3}$ | $10.6_{\pm 1.4}$ | $17.1_{\pm 2.3}$ |
| | 12 | Pooled ID | $0.0_{\pm 0.0}$ | $0.5_{\pm 0.9}$ | $0.5_{\pm 0.9}$ | $11.4_{\pm 9.6}$ | $24.5_{\pm 3.1}$ |
| | | OOD reversed | $0.0_{\pm 0.0}$ | $0.4_{\pm 0.3}$ | $1.7_{\pm 1.0}$ | $9.4_{\pm 4.4}$ | $23.2_{\pm 1.5}$ |
| | 24 | Pooled ID | $0.0_{\pm 0.0}$ | $0.0_{\pm 0.0}$ | $2.2_{\pm 1.5}$ | $8.7_{\pm 9.9}$ | $20.0_{\pm 12.8}$ |
| | | OOD reversed | $0.0_{\pm 0.0}$ | $0.2_{\pm 0.2}$ | $2.4_{\pm 0.5}$ | $10.2_{\pm 11.1}$ | $16.5_{\pm 13.2}$ |
| TCMNIST-Time | 6 | Pooled ID | $3.0_{\pm 2.7}$ | $2.9_{\pm 1.8}$ | $4.3_{\pm 2.0}$ | $13.8_{\pm 2.0}$ | $21.3_{\pm 2.7}$ |
| | | OOD reversed | $1.9_{\pm 0.4}$ | $2.1_{\pm 0.6}$ | $3.4_{\pm 0.6}$ | $9.1_{\pm 1.3}$ | $15.0_{\pm 4.8}$ |
| | 12 | Pooled ID | $0.0_{\pm 0.0}$ | $0.4_{\pm 0.7}$ | $0.9_{\pm 0.8}$ | $18.3_{\pm 6.0}$ | $26.8_{\pm 4.7}$ |
| | | OOD reversed | $0.0_{\pm 0.0}$ | $0.3_{\pm 0.5}$ | $0.5_{\pm 0.2}$ | $15.6_{\pm 2.9}$ | $20.1_{\pm 1.0}$ |
| | 24 | Pooled ID | $0.0_{\pm 0.0}$ | $1.0_{\pm 0.9}$ | $2.2_{\pm 3.7}$ | $16.9_{\pm 19.4}$ | $7.1_{\pm 6.5}$ |
| | | OOD reversed | $0.0_{\pm 0.0}$ | $1.2_{\pm 0.8}$ | $1.3_{\pm 1.7}$ | $12.8_{\pm 18.6}$ | $9.0_{\pm 13.2}$ |

Table 121: Relational: worst-group accuracy (%), mean ± seed SD. Multi-group conditions and task-specific annotations follow Appendix E.4.

| Task | $d$ | Split | Logistic regression | MLP | Local VQC | Entangling VQC | Reuploading VQC |
|---|---|---|---|---|---|---|---|
| RO-CLEVR | 6 | Pooled ID | $3.0_{\pm 0.6}$ | $3.1_{\pm 0.7}$ | $2.2_{\pm 0.8}$ | $19.0_{\pm 2.4}$ | $3.9_{\pm 5.7}$ |
| | | Random | $2.8_{\pm 0.4}$ | $3.8_{\pm 0.8}$ | $1.7_{\pm 0.9}$ | $18.2_{\pm 3.2}$ | $4.0_{\pm 2.8}$ |
| | | OOD reversed | $2.6_{\pm 0.6}$ | $3.1_{\pm 0.3}$ | $1.6_{\pm 0.6}$ | $20.9_{\pm 1.4}$ | $3.1_{\pm 2.9}$ |
| | 12 | Pooled ID | $7.4_{\pm 1.0}$ | $11.1_{\pm 2.3}$ | $2.6_{\pm 2.3}$ | $6.1_{\pm 3.4}$ | $2.6_{\pm 0.1}$ |
| | | Random | $6.7_{\pm 0.4}$ | $11.3_{\pm 4.0}$ | $3.0_{\pm 2.2}$ | $5.9_{\pm 3.4}$ | $2.3_{\pm 0.6}$ |
| | | OOD reversed | $6.1_{\pm 0.4}$ | $11.3_{\pm 6.4}$ | $2.4_{\pm 1.4}$ | $7.1_{\pm 4.4}$ | $1.9_{\pm 0.7}$ |
| | 24 | Pooled ID | $42.1_{\pm 6.2}$ | $42.1_{\pm 8.6}$ | $10.5_{\pm 10.8}$ | $7.0_{\pm 2.0}$ | $10.1_{\pm 7.1}$ |
| | | Random | $38.1_{\pm 3.2}$ | $41.5_{\pm 6.2}$ | $9.2_{\pm 11.1}$ | $5.7_{\pm 3.0}$ | $5.9_{\pm 2.3}$ |
| | | OOD reversed | $34.8_{\pm 3.4}$ | $39.1_{\pm 3.5}$ | $8.5_{\pm 9.7}$ | $4.1_{\pm 0.8}$ | $5.6_{\pm 2.9}$ |

**Native domain shifts.** Yearbook's worst-group test accuracy is much lower than its aggregate score and varies across seeds (Table 122). Camelyon17 contributes ID validation only; single-group OOD results appear above.

Table 122: Domain: worst-group accuracy (%), mean ± seed SD. Multi-group conditions and task-specific annotations follow Appendix E.4.

| Task | $d$ | Split | Logistic regression | MLP | Local VQC | Entangling VQC | Reuploading VQC |
|---|---|---|---|---|---|---|---|
| Camelyon17 | 6 | ID validation | $87.7_{\pm 1.3}$ | $88.9_{\pm 2.2}$ | $89.0_{\pm 1.9}$ | $85.0_{\pm 2.1}$ | $82.2_{\pm 1.3}$ |
| | 12 | ID validation | $88.7_{\pm 2.5}$ | $90.7_{\pm 1.2}$ | $88.1_{\pm 3.8}$ | $88.8_{\pm 1.8}$ | $79.9_{\pm 4.2}$ |
| | 24 | ID validation | $89.8_{\pm 2.0}$ | $90.0_{\pm 0.4}$ | $88.2_{\pm 3.2}$ | $80.1_{\pm 2.9}$ | $81.0_{\pm 4.3}$ |
| Yearbook | 6 | Pooled ID | $85.7_{\pm 2.3}$ | $82.9_{\pm 5.2}$ | $85.7_{\pm 2.3}$ | $68.3_{\pm 12.7}$ | $76.2_{\pm 1.4}$ |
| | | Future-year test | $41.8_{\pm 7.5}$ | $40.7_{\pm 6.4}$ | $43.9_{\pm 10.3}$ | $43.9_{\pm 10.3}$ | $40.7_{\pm 6.4}$ |
| | 12 | Pooled ID | $86.7_{\pm 2.9}$ | $84.6_{\pm 3.8}$ | $82.7_{\pm 3.9}$ | $71.9_{\pm 11.7}$ | $73.7_{\pm 2.3}$ |
| | | Future-year test | $40.7_{\pm 6.4}$ | $41.1_{\pm 8.4}$ | $43.4_{\pm 9.6}$ | $42.1_{\pm 8.4}$ | $36.5_{\pm 16.0}$ |
| | 24 | Pooled ID | $88.7_{\pm 5.3}$ | $80.3_{\pm 6.8}$ | $78.8_{\pm 8.0}$ | $69.0_{\pm 16.5}$ | $51.4_{\pm 14.6}$ |
| | | Future-year test | $46.3_{\pm 11.6}$ | $44.1_{\pm 10.6}$ | $45.7_{\pm 10.9}$ | $41.8_{\pm 7.5}$ | $21.1_{\pm 18.4}$ |

### E.5 Relation to the theoretical analysis

Proposition 1 distinguishes encoded from readout-recoverable information; Proposition 6 relates cue stability to OOD behavior. They organize these task/architecture/split comparisons. Theorem 4's coordinate-independence assumptions require separate verification after feature extraction and PCA; these results support the broader mechanism, not a new instance of that population theorem. Likewise, Proposition 7 concerns initial feature-specific reliance, whereas checkpoints measure finite-training changes in balanced accuracy.

### E.6 Checkpoint trajectories

Table 5 aggregates within-seed changes from updates 100 to 300. Curves use the checkpoint schedules in Appendix E.2; connecting lines only guide the eye. All plots share rows $d = 6, 12, 24$ and columns local, entangling, and reuploading VQCs. Curves show every task-specific held-out condition as mean balanced accuracy ± sample SD over 3 seeds; dashed lines mark 50%. Endpoints are fixed without OOD selection.

Appearance trajectories are in Figures 25, 26, and 27; frequency trajectories in Figures 28 and 29. Figures 30 and 31 give temporal ID/reversal curves; Figures 32 and 33 retain native-domain conditions. Figure 34 gives relational trajectories. Reading ID and shifted curves jointly distinguishes source fitting from improvements retained under shift.

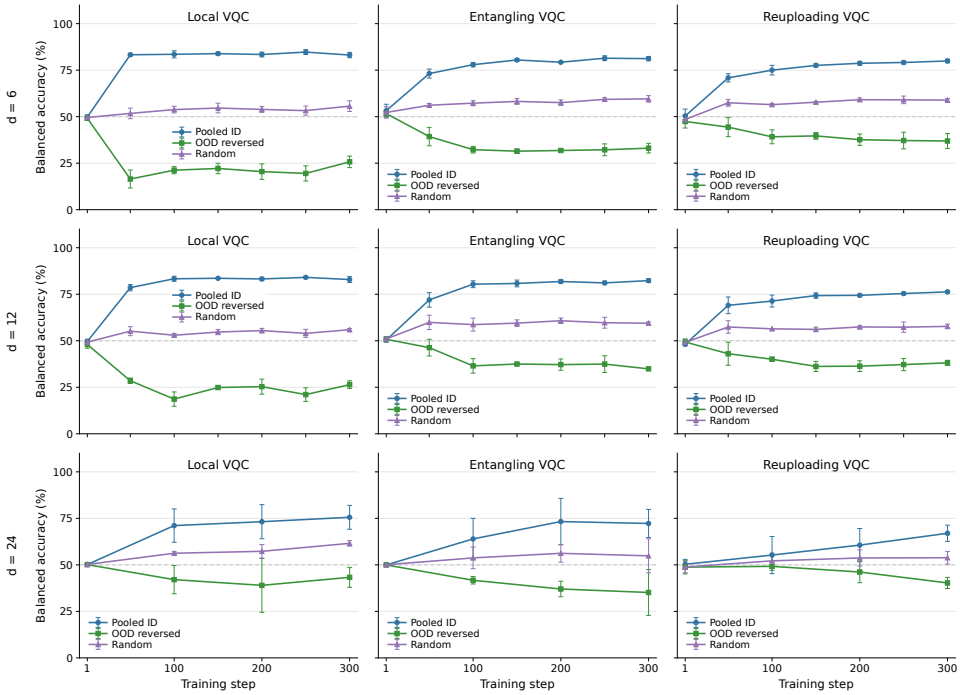


Figure 25: **Q-Colored-MNIST learning trajectories.** Layout and uncertainty: Appendix E.6.

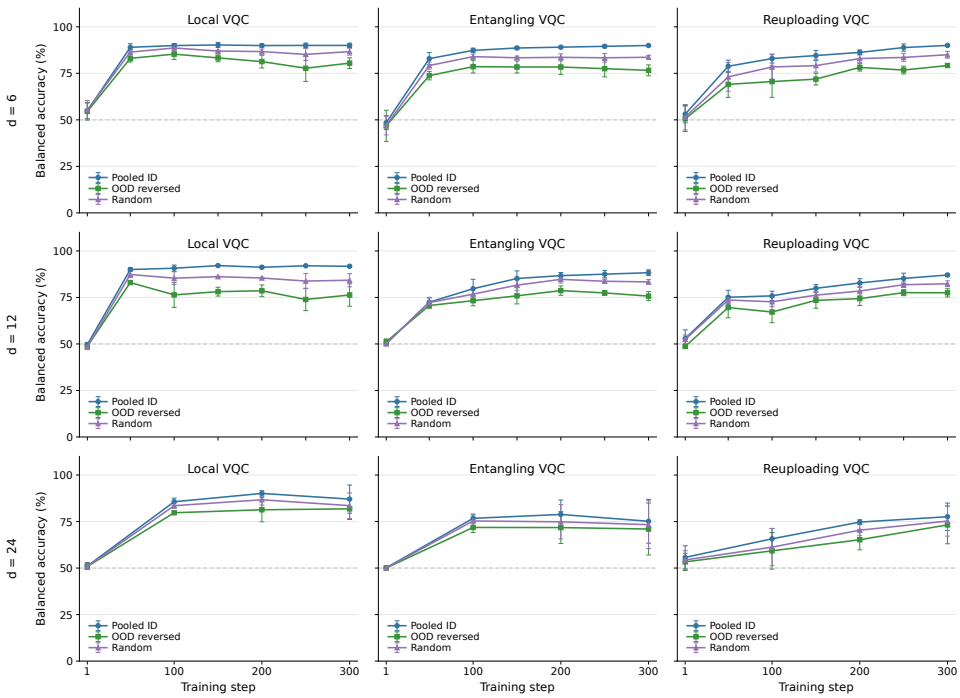


Figure 26: **Q-Colored-Fashion-MNIST learning trajectories.** Layout and uncertainty: Appendix E.6.

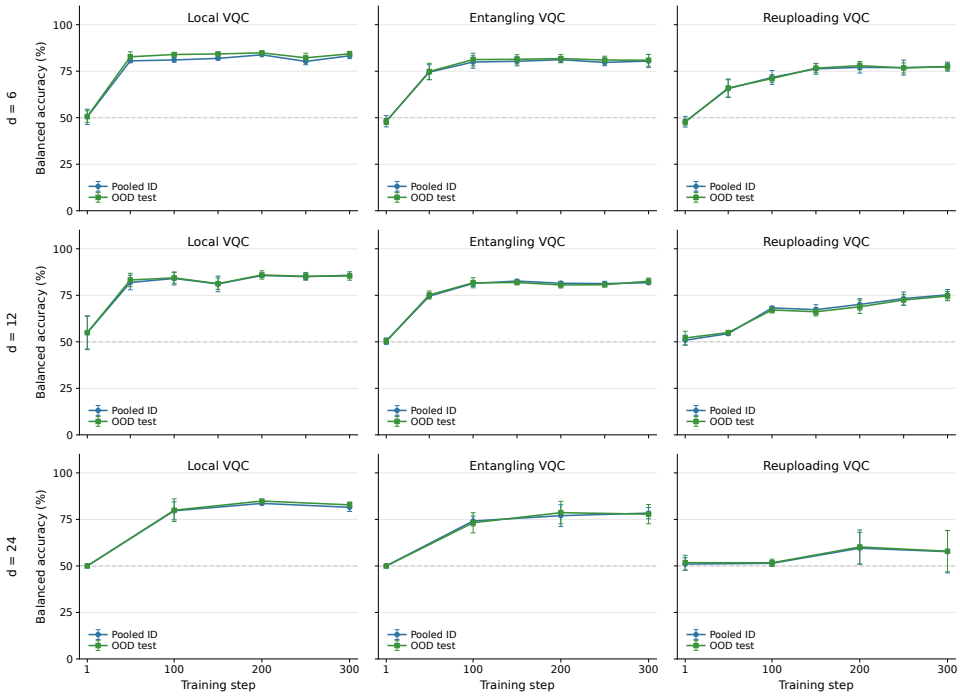


Figure 27: **Waterbirds learning trajectories.** Layout and uncertainty: Appendix E.6.

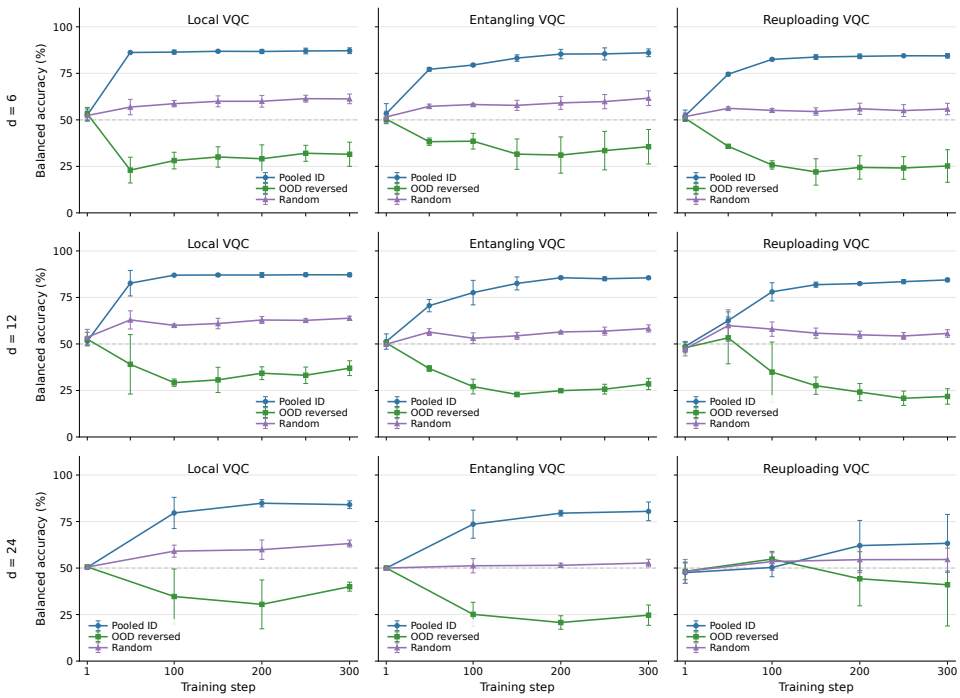


Figure 28: **MNIST-Fourier learning trajectories.** Layout and uncertainty: Appendix E.6.

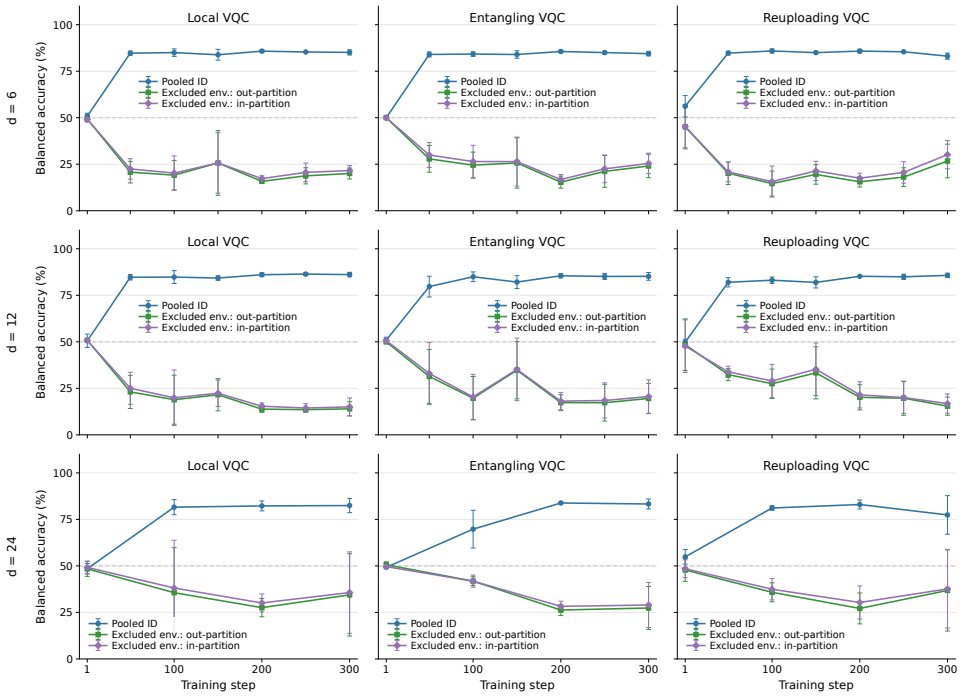


Figure 29: **WOODS Spurious-Fourier learning trajectories.** Layout and uncertainty: Appendix E.6.

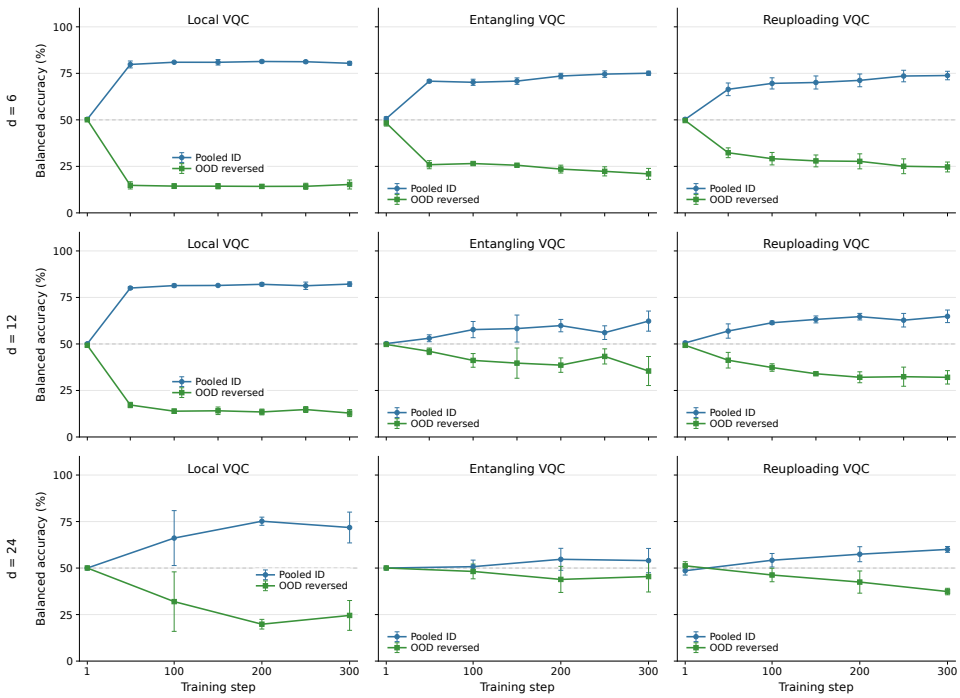


Figure 30: **TCMNIST-Source learning trajectories.** Layout and uncertainty: Appendix E.6.

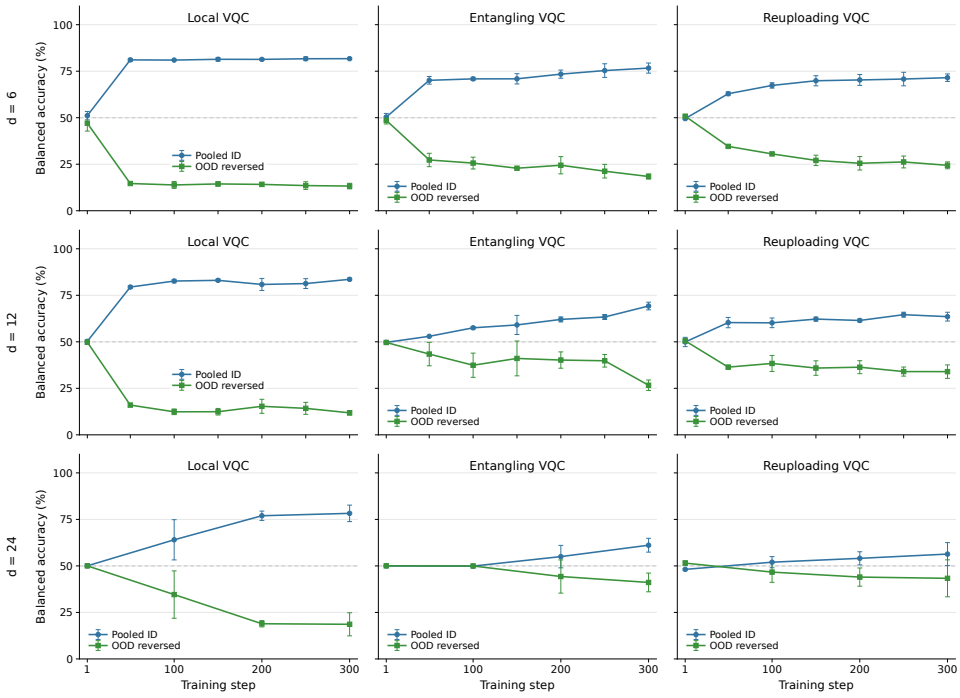


Figure 31: **TCMNIST-Time learning trajectories.** Layout and uncertainty: Appendix E.6.

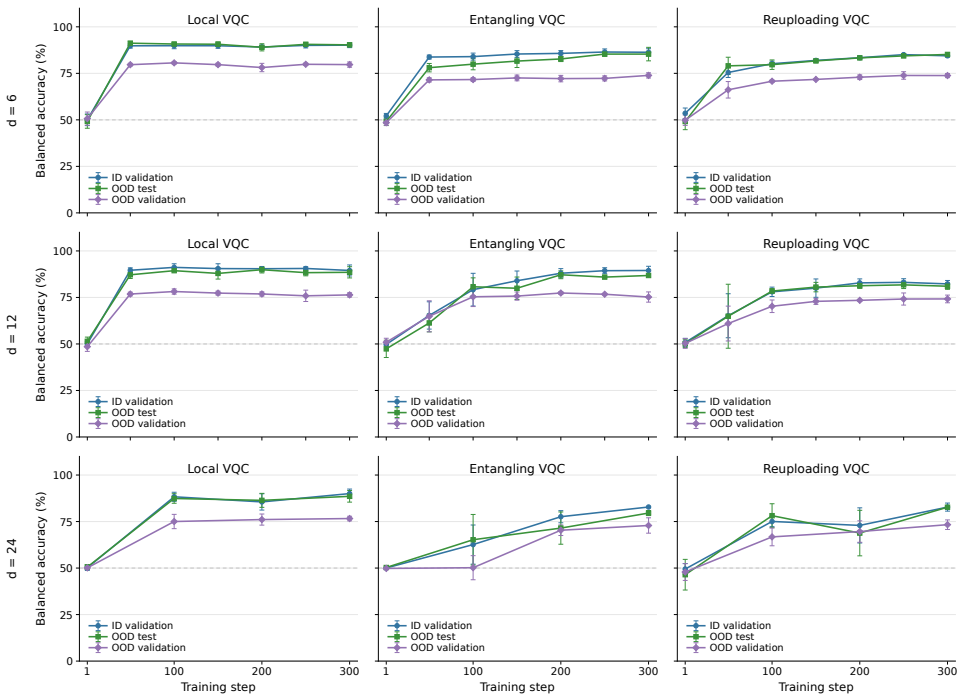


Figure 32: **Camelyon17 learning trajectories.** Layout and uncertainty: Appendix E.6.

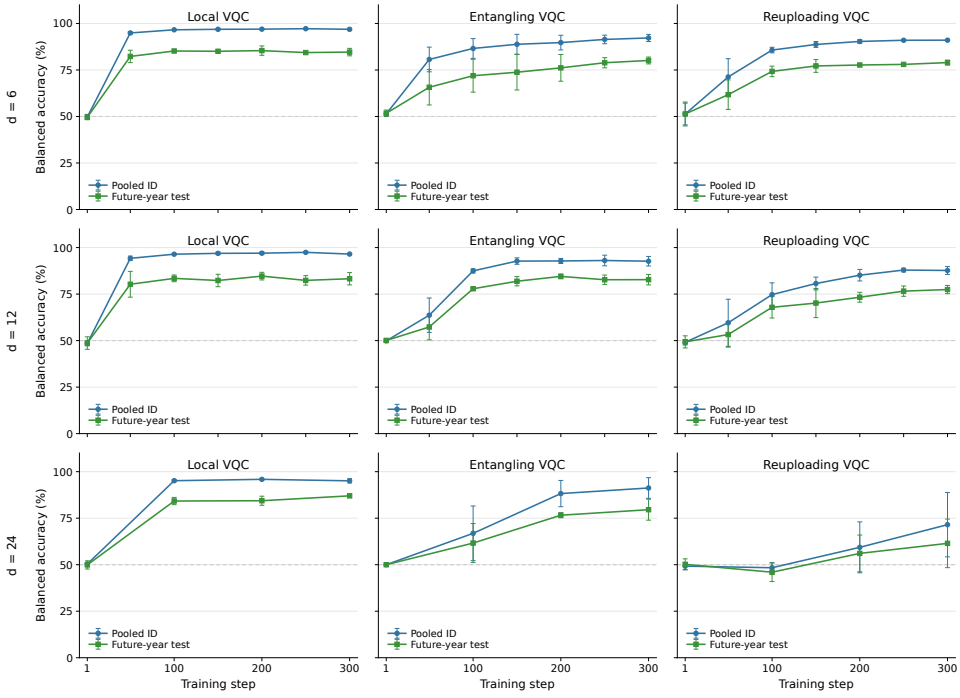


Figure 33: **Yearbook learning trajectories.** Layout and uncertainty: Appendix E.6.

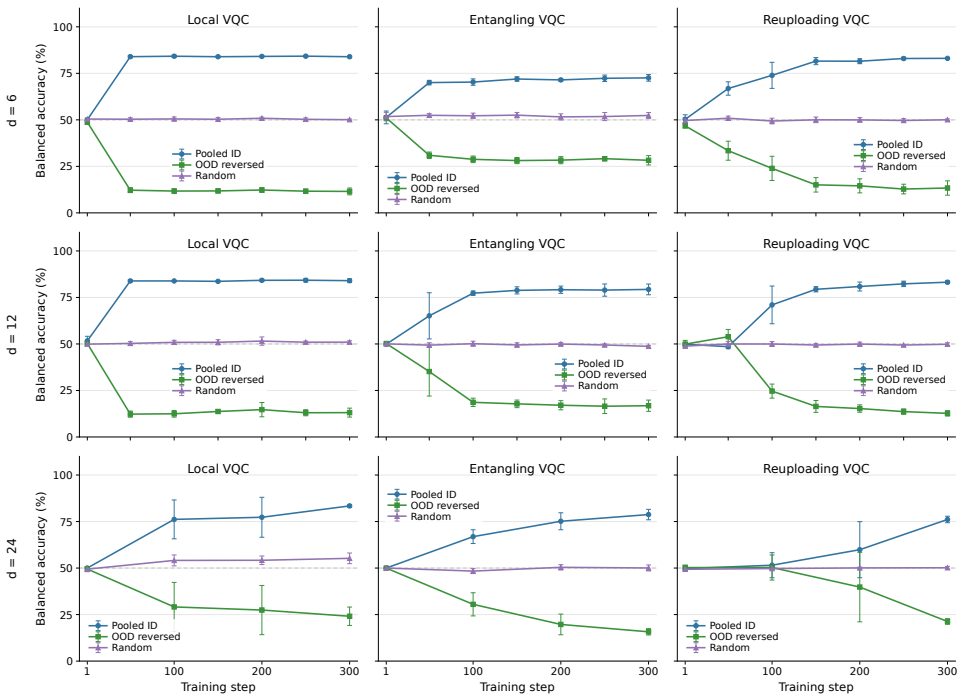


Figure 34: **RO-CLEVR learning trajectories.** Layout and uncertainty: Appendix E.6.