ENTANGLEMENT AS A STRUCTURAL COMPLEXITY AXIS: A PAC-BAYESIAN VIEW OF GENERALIZATION IN QUANTUM POLICIES AND VALUE FUNCTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Parameterized quantum circuits (PQCs) are increasingly used as policies and value functions in quantum reinforcement learning (QRL), yet almost all evaluations report only average return, leaving open the question of *when* and *why* a quantum policy generalizes. We give a PAC-Bayesian answer. We derive a generalization bound for stochastic quantum policies whose complexity term is controlled not by the raw number of circuit parameters, but by the *log-determinant complexity of the Fisher geometry induced by the circuit*—a quantity that entangling connectivity inflates by expanding the readout's causal cone. Empirically, in a controlled setting that fixes the number and layout of trainable rotations and varies *only* the entangling connectivity, this acts as an independent axis of complexity: at a fixed parameter count the train-to-test generalization gap grows with the circuit's Fisher log-determinant complexity, whereas raw parameter count is the weakest predictor of the gap. The bound acts as a *ranking certificate*: at fixed parameter count it consistently distinguishes the non-entangled circuit from entangled alternatives, something a parameter-counting bound cannot do at all. We confirm the mechanism across settings—supervised classification (4–16 qubits, synthetic and real data), a reward-only quantum contextual bandit, and multi-step value-function generalization—where entangled circuits generalize worse than non-entangled ones of identical parameter count and the gap shrinks with sample size as predicted. Our strongest evidence is in these low-variance decision models; in genuine end-to-end multi-step *policy* learning the standard entangler's effect is statistically significant but return variance leaves the full ordering only partially resolved, so we frame the paper as a controlled study of the entanglement–generalization trade-off in quantum decision models rather than a solved account of end-to-end quantum RL. A partial-correlation analysis shows d_{eff} screens off the entangling pattern ($\Delta R^2 < 0.01$), consistent with readout causal-cone expansion as the operative mechanism and with the Meyer–Wallach entangling power as an empirical correlate rather than the driving variable; controls (matched training accuracy, alternative readout and optimizer) rule out optimization confounders, and the effect survives execution on an IBM Heron processor under real noise. Our results reframe quantum-policy design around an entanglement–generalization trade-off rather than expressivity alone.

1 INTRODUCTION

Quantum reinforcement learning (QRL) replaces the neural policy or value network of a classical agent with a parameterized quantum circuit (PQC). Recent work has made such hybrid agents practical—PPO-Q integrates a PQC policy into proximal policy optimization and validates it on superconducting hardware (Jin et al., 2025), and adaptive non-local observables show that the measurement layer itself is a key design axis (Lin et al., 2025). Yet the field is almost entirely empirical: papers report the average return of a chosen ansatz, and a reviewer is left to wonder whether a reported gain reflects a genuine inductive bias or merely the implicit regularization of a small model.

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.

We argue that the right question is not "does a quantum policy achieve higher return" but "*when does a quantum policy generalize, and what property of the circuit controls it*". This is the question PAC-Bayes theory is built to answer for stochastic predictors (McAllester, 1999; Maurer, 2004; Dziugaite & Roy, 2017), and it has recently been brought to reinforcement learning, yielding non-vacuous certificates for modern off-policy agents (Zitouni et al., 2025). In parallel, a PAC-Bayesian analysis of *supervised* quantum models has appeared (Rodriguez-Grasa et al., 2026), whose complexity term is built from learned parameter norms and sparsity. To our knowledge no PAC-Bayes analysis addresses quantum *policies*, and none isolates *entanglement* as the governing complexity.

This paper makes three contributions.

- **A PAC-Bayes bound for quantum policies** (Section 3) whose complexity term is the log-determinant complexity of the circuit-induced Fisher geometry rather than the parameter count. Under a data-dependent Gaussian posterior the KL complexity is governed by $d_{\text{eff}}(\theta)$, which for PQCs is inflated by entangling connectivity. We stress that d_{eff} is a Fisher log-volume, not a dimension count: it is the *rank* of the Fisher matrix that the parameter count bounds, while d_{eff} itself may exceed d.
- **An identification of entanglement as an independent complexity axis** (Section 4). Fixing the number of trainable rotations and varying only the entangling connectivity, the generalization gap increases with the Meyer–Wallach entangling power (Meyer & Wallach, 2002) *at fixed parameter count*—monotonically in the primary $L{=}2$ sweep, and strongly associated with it throughout—while parameter count itself is the weakest predictor of the gap. The axis is *readout-gated*: entanglement enlarges the readout's light-cone, so it is decisive under the local single-qubit readouts standard in QRL and is partly substituted by a global readout, which activates the same parameters directly (Section 7)—a control that confirms the light-cone account and shows d_{eff}, not the entangling label, is what tracks the gap.
- **Broad empirical validation**: supervised classification at 4–16 qubits (Sections 4–5), a reward-only quantum contextual bandit and multi-step value-function generalization (Section 6), optimization controls, real datasets, and a run on IBM Heron hardware (Section 7)—in all of which entangled circuits generalize worse than non-entangled ones of identical parameter count.

The structural pathway we test. The mechanism is *entangling connectivity* $\rightarrow$ *readout operator spreading* $\rightarrow$ *more active Fisher directions* $\rightarrow$ *larger* d_{eff} $\rightarrow$ *looser certificate*, with a larger gap *empirically associated* with that axis rather than implied by it, and it is *not* quantum-state entanglement as such: Q is an empirical correlate of the first arrow, and d_{eff} screens it off in a partial-correlation analysis (Section 4). We say "entanglement" where circuit-design vocabulary demands it, but every claim is about connectivity—a double-edged design axis, which enlarges the function space a quantum policy can represent while carrying a generalization cost that a parameter-counting analysis does not see and our bound makes explicit.

Scope. Our *strongest* evidence comes from low-variance instantiations of policies and value functions (single-observable classifiers and value heads, a one-step policy, Monte-Carlo evaluation of a multi-step value function); in end-to-end multi-step policy learning the standard entangler's effect is significant but return variance leaves the full ordering only partly resolved (Section 7). The paper is thus a controlled study of the entanglement–generalization trade-off in quantum decision models, not a solved account of end-to-end quantum RL.

2 Related Work

Quantum reinforcement learning. PQC agents have been trained with policy gradients, actor–critic and value-based methods (Jin et al., 2025; Luo et al., 2025; Wu et al., 2025), with recent attention to the measurement layer (Lin et al., 2025) and to data re-uploading and trainability (Coelho et al., 2024). These works optimize and report return; generalization is not analyzed.

Generalization of quantum models. Covering-number and Rademacher analyses give uniform bounds for quantum feature maps and PQCs (Banchi et al., 2021), and Caro et al. (2022) prove the

landmark result that a PQC with T parameterized gates generalizes from $\widetilde{O}(T)$ training points—a bound driven by *gate count*, not entanglement. Gil-Fuster et al. (2024) caution that such uniform-convergence bounds can be simultaneously satisfied and uninformative for the models one actually trains, motivating *model-dependent* capacity measures. The Fisher-information "effective dimension" is one such measure (Abbas et al., 2021). Our d_{eff} is this same log-determinant functional, but our contribution is orthogonal in two respects (Appendix C): they use it as a global, sample-size-dependent capacity for a model *class*, whereas we evaluate it at the *trained* parameters with a fixed γ and only to *rank* circuits of equal parameter count; and they do not connect it to circuit structure—our Proposition 1 attributes the inflation of the Fisher rank to connectivity-driven growth of the readout cone, which is the new mechanism. On the PAC-Bayes side, Rodriguez-Grasa et al. (2026) bound *supervised* quantum classifiers by a parameter-norm/sparsity term, with weak but positive correlations to the gap. We differ in treating *policies* (return-based generalization under Markov dependence) and in identifying *connectivity and gate structure*, rather than trainable parameter count, as the governing axis, isolating it at a fixed count and layout of trainable rotations.

PAC-Bayes for RL and control. PAC-Bayes control learns policies with certified generalization to novel environments (Majumdar et al., 2021), and recent work derives PAC-Bayes RL bounds that account for the mixing time of the induced Markov chain (Zitouni et al., 2025). We adopt this template and instantiate its complexity term for quantum policies.

3 A PAC-Bayes Bound for Quantum Policies

Setup. An agent interacts with an MDP M drawn from a distribution $\mathcal{D}$ over environments. A stochastic quantum policy $\pi_\theta(a \mid s)$ is defined by a PQC $U(s,\theta)$ acting on n qubits: the state s is encoded, the circuit is applied, and the action is read from a *single* readout observable $\langle Z_0\rangle_{s,\theta}$ (a binary policy $\pi_\theta(a{=}1 \mid s) = \sigma(\alpha\langle Z_0\rangle_{s,\theta})$; k-action policies use k observables and our analysis applies per observable). This single local readout is the default throughout the theory and the experiments; the two controls that deliberately vary it—an alternative local observable Z_1 and the global $\frac{1}{n}\sum_i\langle Z_i\rangle$—are reported in Section 7. We write $J(\pi_\theta) = \mathbb{E}_{M\sim\mathcal{D}}\mathbb{E}_{\tau\sim\pi_\theta}[R(\tau)]$ for the expected return with $R \in [0, R_{\max}]$, $\hat{J}(\pi_\theta)$ for its empirical estimate on m training environments, and consider a Gaussian posterior $Q = \mathcal{N}(\theta, \sigma^2 I_d)$ with prior $P = \mathcal{N}(\theta_0, \sigma^2 I_d)$.

Theorem 1 (PAC-Bayes bound for quantum policies, i.i.d. environments)**.** *Let the m training environments be drawn i.i.d. from $\mathcal{D}$ and write $R_M(\theta) = \mathbb{E}_{\tau\sim\pi_\theta,M}[R(\tau)] \in [0, R_{\max}]$ for the* expected *return of π_θ on environment M, so that the per-environment loss is a deterministic function of (θ, M) and carries no rollout randomness. For any $\delta \in (0,1)$, with probability at least $1-\delta$ over that draw,* simultaneously for every *posterior Q,*

$$J(\pi_Q) \;\geq\; \hat{J}(\pi_Q) \;-\; R_{\max}\sqrt{\frac{\mathrm{KL}(Q\|P) + \ln\frac{2\sqrt{m}}{\delta}}{2\,m}}, \qquad \mathrm{KL}(Q\|P) = \frac{\|\theta-\theta_0\|^2}{2\sigma^2}. \tag{1}$$

Theorem 1 is the i.i.d. statement and covers the experiments on which we evaluate the certificate directly (supervised classification and the contextual bandit); the remaining multi-step experiments are empirical tests of the structural mechanism rather than direct evaluations of the certificate, and extending the bound to Markov-dependent transitions needs a mixing-time correction (Zitouni et al., 2025) that we neither reprove nor rely on. Equation (1) is the standard McAllester bound; its only quantum-specific object is $\mathrm{KL}(Q\|P)$. Naively this depends only on the Euclidean displacement of the d parameters, suggesting that generalization is controlled by parameter count. Our central observation is that this is misleading once the posterior is shaped by the local geometry the circuit induces on the loss. In the settings carrying our evidence each "environment" is one i.i.d. *context* with a one-shot reward, so Theorem 1 applies directly with $m = N$.

From an isotropic to a Fisher-shaped posterior. The bound (1) is loose because the isotropic posterior ignores the geometry of the loss. The curvature of a quantum policy is the Fisher matrix $\hat{\mathcal{F}}(\theta) = \mathbb{E}_s[\nabla_\theta \ell_s\, \nabla_\theta \ell_s^\top]$ (eigenvalues $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$), which is highly anisotropic. For the single-observable Bernoulli readout $p_s = \sigma(\alpha\langle Z_0\rangle_{s,\theta})$ it has the closed form $\hat{\mathcal{F}}(\theta) = \mathbb{E}_s\big[\alpha^2\, p_s(1{-}p_s)\, \nabla_\theta f_s\, \nabla_\theta f_s^\top\big]$ with $f_s = \langle Z_0\rangle_{s,\theta}$: the loss Fisher of the logistic head *equals* the

output-gradient Fisher weighted by the Bernoulli variance. This is the object we diagonalize throughout (Appendix C), so the $\hat{\mathcal{F}}$ in the bound and the $\hat{\mathcal{F}}$ we measure are the same matrix, with rank governed by the readout cone (Proposition 1).

Theorem 2 (Fisher-form bound)**.** *Let the prior be the data-independent isotropic Gaussian $P = \mathcal{N}(\theta_0, \tau^2 I_d)$ and the posterior the Gauss–Newton Gaussian $Q = \mathcal{N}(\theta, \Sigma)$ with $\Sigma = \tau^2(I_d + \gamma\hat{\mathcal{F}})^{-1}$, $\gamma > 0$. Then*

$$\mathrm{KL}(Q\|P) = \frac{\|\theta - \theta_0\|^2}{2\tau^2} + \tfrac{1}{2}\underbrace{\log\det\big(I_d + \gamma\hat{\mathcal{F}}\big)}_{=:\, d_{\mathrm{eff}}(\gamma)} - \tfrac{1}{2}\sum_{i=1}^{d}\frac{\gamma\lambda_i}{1+\gamma\lambda_i}, \tag{2}$$

and hence, under the assumptions of Theorem 1, with probability $\geq 1 - \delta$,

$$J(\pi_Q) \;\geq\; \hat{J}(\pi_Q) - R_{\max}\sqrt{\frac{\frac{\|\theta - \theta_0\|^2}{2\tau^2} + \frac{1}{2}\, d_{\mathrm{eff}}(\gamma) + \ln\frac{2\sqrt{m}}{\delta}}{2\,m}}. \tag{3}$$

Equation (3) is the bound that matters: its complexity is the distance term *plus* the *Fisher log-determinant complexity*

$$d_{\mathrm{eff}}(\gamma) = \log\det\big(I_d + \gamma\hat{\mathcal{F}}\big) = \sum_{i=1}^{d}\log(1+\gamma\lambda_i), \tag{4}$$

a quantity a parameter count cannot see. The final term of (2) is *subtracted* and lies in $[0, d/2]$; dropping it replaces KL by a larger upper surrogate, which conservatively *loosens* the gap bound while keeping it valid—we drop it only to expose d_{eff} as the operative complexity. What makes d_{eff} the right object is that it is driven by the *active* Fisher directions rather than the nominal ones: $d_{\mathrm{eff}}(\gamma) = \mathrm{rank}(\hat{\mathcal{F}})\log\gamma + O(1)$ for large γ, and a parameter outside the readout's causal cone contributes exactly zero (Proposition 1).

d_{eff} is a log-volume, not a dimension count. The name invites a false expectation, so we are explicit: it is $\mathrm{rank}(\hat{\mathcal{F}}) \leq d$ that the parameter count bounds, while $d_{\mathrm{eff}}(\gamma) = \sum_i \log(1+\gamma\lambda_i)$ is an *unnormalized Fisher log-volume* and is *not* bounded by d—it exceeds d in our own measurements (d_{eff}=21.8 at d=16, Table 11). This is not a pathology: the bound charges for how sharply the data curve each active direction, not merely for how many there are. We keep the symbol and the link to Abbas et al. (2021), but call the quantity the *Fisher log-determinant complexity* and never treat it as a dimension. Crucially the prior is data-*independent* (centred at the initialization θ_0), so (3) is a valid PAC-Bayes certificate even though the posterior *shape* Σ is data-dependent through $\hat{\mathcal{F}}$. Equation (2) is a direct Gaussian KL computation, proved in Appendix A.

Remark 1 (Choosing γ)**.** *Theorem 1 holds simultaneously for* every *posterior Q, so PAC-Bayes permits Q—including its covariance hyperparameters—to depend on the observed sample; it is the prior P that must be data-independent, and ours is, being centred at θ_0. No union bound over γ is needed. We fix $\gamma = 50$ across all experiments purely for comparability, and the ordering is stable across $\gamma \in \{10, \ldots, 200\}$ (Appendix C). A sample-size-dependent $\gamma \propto m$, as in Abbas et al. (2021), is* not *used here.*

Proposition 1 (Entangling connectivity enlarges the causal cone and activates Fisher directions)**.** *Fix the number and placement of the trainable single-qubit rotations and let the entangling connectivity vary. Let $\mathcal{C} \subseteq \{1, \ldots, d\}$ be the* potential *backward causal cone of the readout observable—a purely combinatorial set determined by the circuit's connectivity graph, independent of θ (Def. 1)—and let $\mathcal{L}(\theta) \subseteq \mathcal{C}$ be the parameters with a non-vanishing gradient. Then:* (i) *$\mathcal{C}$ is non-decreasing under the addition of entangling gates (Lemma 2), strictly increasing as soon as a gate couples the readout wire to a wire carrying rotations;* (ii) *the empirical Fisher formed from M inputs obeys $\mathrm{rank}(\hat{\mathcal{F}}) \leq \min\{|\mathcal{L}(\theta)|, M\} \leq |\mathcal{C}| \leq d$, with equality to $|\mathcal{L}(\theta)|$ generically when $M \geq |\mathcal{L}(\theta)|$ (Assumption 1); and* (iii) *with no entangling gates $\mathcal{C} = \{i : q(i) = 0\}$, so at most d/n parameters can be active—all at fixed parameter count d.*

Remark 2 (What is, and is not, claimed)**.** *Only one thing here is a theorem: monotonicity of the* structural *cone $\mathcal{C}$, which bounds $\mathrm{rank}(\hat{\mathcal{F}})$ from above. That the rank, and then $d_{\mathrm{eff}}(\gamma)$ at finite γ,*

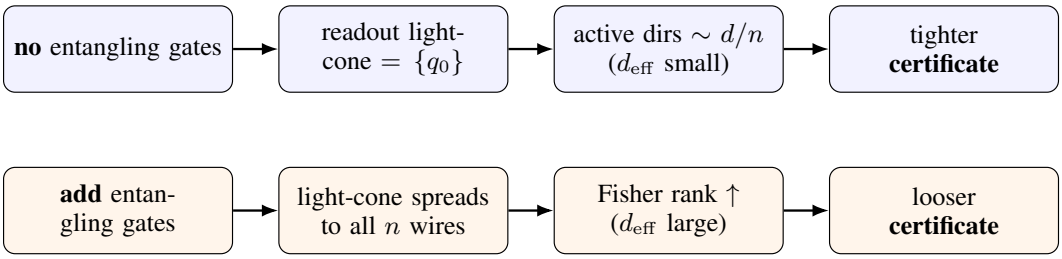


Figure 1: The mechanism, separating what is proved from what is measured. *Structurally:* at fixed parameter count, entangling gates enlarge the backward causal cone of the readout (Lemma 2), so more Fisher directions can be active (Proposition 1); where those directions carry non-negligible curvature this raises $d_{\text{eff}} = \log\det(I + \gamma\hat{\mathcal{F}})$ and hence loosens the certificate of (3). A looser upper bound does *not* by itself imply a larger true gap. *Empirically*, the observed gap does grow along the same axis—monotonically with d_{eff} at fixed d (Sections 4–7)—which is the regularity this paper measures, and which a parameter-counting analysis cannot see since d is unchanged.

actually move with it is generic rather than universal*—a newly admitted direction can stay inactive, and one with $\lambda_i \approx 0$ contributes ≈ 0 to d_{eff}—so both hold under generic non-commutation and away from barren plateaus, and we verify them empirically rather than deriving them. The inclusion $\mathcal{L}(\theta) \subseteq \mathcal{C}$ is also generally strict—a rotation inside the cone is inert if its generator commutes with the back-propagated observable, as the final R_Z on the readout wire always does. We therefore treat finite-γ d_{eff} monotonicity as an* empirical *regularity (Table 10 shows where it fails), not a theorem.*

Together these yield the falsifiable prediction we test at fixed parameter count: circuits whose connectivity gives a larger causal cone should—generically, off the plateau—have a larger d_{eff} and hence a looser bound (3) than circuits of *identical* parameter count with a smaller cone, and—empirically—a larger gap. No parameter-counting bound can express this. Proofs, scope conditions, a two-qubit worked example and the failure modes are in Appendix A; Figure 1 summarizes the mechanism.

4 Entanglement is an Independent Axis of Complexity

Design. To separate connectivity from parameter count we fix the layout of trainable rotations—L layers of R_Y, R_Z on each of n=4 qubits, so $d = 2nL$ parameters—and vary *only* the entangling connectivity after each rotation layer: `none` (no CNOTs), `linear` (a CNOT chain), and `full` (all-to-all CNOTs). All three share exactly the same d and rotation layout. These three do *not* fix the total gate count (0, $n-1$, $n(n-1)/2$ CNOTs), so gate count and cone reach co-vary across them. We separate the two with a matched-gate-count control (Appendix B, Table 5): three patterns with *exactly three* CNOTs but different readout-mediated connectivity. When the gates avoid the readout wire (`disjoint`=$\{(1,2),(2,3),(1,3)\}$, cone $\{0\}$) the circuit is *indistinguishable* from the unentangled one—identical gap, d_{eff} and live parameter count—while `linear` and `star`, with the same three gates but a cone covering all wires, give gaps of 0.146 and 0.372. Gate count alone therefore does not drive the effect; readout-mediated coupling and the resulting Fisher geometry do. Since `linear` and `star` share gate count *and* cone yet differ in d_{eff} and gap, cone membership alone is not sufficient either. The binary target $y = \text{sign}(x_1x_2 - x_3x_4 + \frac{1}{2}x_1)$ requires feature interactions, inputs are angle-encoded, the classifier is the single-observable readout $\hat{y} = \text{sign}(\langle Z_0\rangle)$, and we sweep the training-set size N and evaluate on 2000 held-out inputs, averaging 8 seeds per cell.

Result: the gap is genuine, and full-strength entanglement costs real accuracy. A key check is that the models actually generalize—otherwise a larger gap would only mean more memorization. Table 1 isolates depth L=2 (d=16) at two sample sizes. At the small N=16 the entangled circuit does memorize (train 0.92 vs 0.69), but test accuracy is near chance for all three—the ambiguous regime a reviewer should worry about. At N=64, however, *every* circuit generalizes above chance (test 0.61–0.66) and the gap is ordered none $<$ linear $<$ full at every N (Figure 2, left). Held-out *accuracy*, though, only drops visibly for the most entangled circuit: `full` reaches 0.608 against 0.656 for `none`, while `linear`'s 0.657 is statistically indistinguishable from `none`'s despite a gap

Table 1: Fixed parameter count (d=16, L=2), varying only the entangling connectivity, single-observable readout, 8 seeds. At N=64 all three circuits generalize well above chance and the gap is ordered none < linear < full; only `full`'s test accuracy is visibly lower, while `linear`'s is statistically indistinguishable from `none`'s—a larger gap need not mean worse held-out accuracy.

N_{train}	Entanglement	Q (M.–W.)	train acc	test acc	gap
16	`none`	0.00	0.69	0.59	0.101
16	`linear`	0.45	0.75	0.60	0.146
16	`full`	0.56	0.92	0.54	0.381
64	`none`	0.00	0.65	**0.656**	−0.01
64	`linear`	0.51	0.72	0.657	0.064
64	`full`	0.58	0.71	**0.608**	0.104

(0.064 vs. −0.01) that is already resolved beyond noise. We flag this because it recurs on real data (Table 9): the gap tracks d_{eff} tightly, but a larger gap does not by itself certify worse test accuracy—the two are correlated, not interchangeable, and we anchor the accuracy claim only where it is resolved. The effect is depth-gated (L=1 is inert for this single-qubit readout) and holds at d fixed. The slightly *negative* `none` gap at N=64 (−0.01) is within one seed standard error of zero, so we anchor quantitative contrasts at sample sizes where gaps are resolved and read near-chance rows only as tests of the *ordering*.

Result: d_{eff} is the best—and significantly best—predictor. Across a large grid of 300 configurations (five entangling patterns × five depths × three sample sizes × four seeds; Appendix C), the Spearman correlation with the gap is ρ=0.82 for the Fisher log-determinant complexity $d_{\text{eff}} = \log\det(I + \gamma\hat{\mathcal{F}})$, 0.73 for the learned displacement $\|\theta - \theta_0\|^2$, 0.72 for the entangling power Q—an *empirical correlate* of connectivity, not the causal variable—and 0.45 for parameter count. The margin of d_{eff} over the learned-norm term is *bootstrap-significant*: $\Delta\rho = 0.090$ with 95% CI $[0.041, 0.144]$. Thus the Fisher log-determinant complexity—the quantity that actually appears in bound (3)—is not merely nominally but significantly the best predictor of the gap, ahead of the learned-norm term used by Rodriguez-Grasa et al. (2026) and far ahead of raw parameter count; and, unlike the norm, entanglement is a structural property fixed before training that separates circuits of *equal* parameter count. A partial-correlation analysis on the same grid shows d_{eff} *screens off* the entangling pattern: adding the pattern on top of d_{eff} raises R^2 by < 0.01, while adding d_{eff} on top of the pattern raises it by 0.14 (Appendix C).

5 The Fisher Log-Determinant Complexity is the Right Capacity Measure

We now test Proposition 1 directly by computing d_{eff} at the trained parameters. At fixed depth, moving from `none` through `linear` to `full` inflates it monotonically, 3.3 → 9.3 → 21.8 at L=2, confirming that connectivity raises the complexity Eq. (3) charges for while d is fixed (per-configuration values and the Fisher spectrum in Appendix B, Table 11 and Figure 6). The gap shrinks with N (Figure 2, left), a basic PAC-Bayes consistency check, and increases with d_{eff} across circuits (Figure 2, right).

The bound is primarily a ranking certificate. We do not claim the bound is numerically tight; its value is that at fixed parameter count it *consistently separates the non-entangled circuit from entangled alternatives*—a distinction no parameter-counting bound can make. Evaluating the right-hand side of Eq. (3) directly (τ^2=1, γ=50, δ=0.05, $R_{\max}$=1; Appendix B, Table 2) shows that (i) it is non-vacuous—below the trivial unit gap and within ≈ 2.3–$4.5\times$ of the deterministic gap where that gap is non-negligible; and (ii), the property we rely on, its complexity ranks the circuits none < linear < full at each N. Point (i) is a numerical comparison, not a certificate for θ: a bound on the gap of π_Q that happens to exceed the gap of θ certifies nothing about θ. The certificate proper is checked against π_Q next. *Which object is certified.* Eq. (3) is a statement about the stochastic predictor π_Q, whereas the "true gap" column reports the deterministic θ. Rather than bridge them by Lemma 1—a

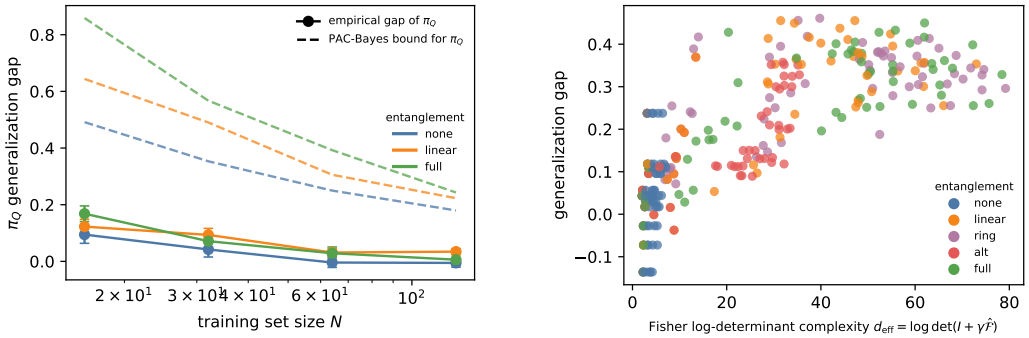


Figure 2: **Left:** the empirical generalization gap of the *stochastic* predictor π_Q (solid, 64 posterior draws per seed) and the PAC-Bayes bound of Eq. (3) for that same π_Q (dashed), versus training-set size N at identical parameter count. Plotting π_Q rather than the deterministic θ makes the panel a statement about the object the theorem certifies, with no derandomization step. The bound upper-bounds the gap at every N and decays with N at the same rate, and `none` is lowest throughout; the `linear`/`full` order is resolved only at N=16 (Table 3 and its discussion). **Right:** across 300 configurations the gap increases with $d_{\text{eff}} = \log\det(I + \gamma\hat{\mathcal{F}})$ (Spearman $\rho = 0.82$); colour encodes entangling connectivity, which drives d_{eff} up at fixed parameter count.

derandomization *argument* whose local expansion we do not rigorously control—we evaluate π_Q itself, sampling 64 draws from $Q = \mathcal{N}(\theta, \tau^2(I+\gamma\hat{\mathcal{F}})^{-1})$ per seed over $N \in \{16, 32, 64, 128\}$ (Appendix B, Table 3; Figure 2, left). The bound holds for π_Q everywhere and `none` has the smallest stochastic gap at every N, so the claim that matters transfers to the certified object. The *three-way* ordering does not transfer as cleanly: it holds at N=16 (0.095/0.123/0.168) but `linear` and `full` exchange places from N=32 on, because posterior averaging shrinks the gap most for the circuit with the largest d_{eff} (`full`: $0.381 \to 0.168$), whose eleven nearly flat Fisher directions absorb the full prior variance. We report this as a limitation of the certificate rather than of the mechanism.

Budgeted, not banished, entanglement. The design implication is not "use less entanglement"—entanglement is what buys a quantum model its expressivity. The trade-off is explicit (Appendix B, Figure 3): as d_{eff} grows with entanglement, training accuracy rises, but test accuracy does *not* follow—the extra training fit is spent on complexity the held-out data does not reward, so the gap opens. The right principle is to *budget* entanglement against its d_{eff} cost, exactly as the bound prices it, rather than to maximize expressivity.

The effect persists from 6 to 16 qubits. To check that the finding is not an artifact of the 4-qubit toy, we repeat the fixed-parameter-count comparison at $n \in \{6, 8, 10, 12, 16\}$ qubits, with the interaction target extended to $n/2$ pairwise products (Appendix B, Table 6). At every size the gap is monotone none $<$ linear $<$ full and the all-to-all circuit (d up to 64 trainable angles, a 2^{16}-dimensional state) exceeds the non-entangled one by 4.5–6.9 standard errors, with the effect growing at larger n (full-gap $0.16 \to 0.34$ from n=6 to n=16). At $n \geq 10$ the target is hard for these shallow circuits and held-out accuracy approaches chance, so those rows test the *ordering* only; at n=6, 8 accuracy stays well above chance. The causal-cone argument carries no dependence on n.

6 Quantum RL Policies: Bandit and Value-Function Generalization

We now test the prediction on quantum *policies* trained by reward maximization. We use a contextual bandit—the canonical one-step RL problem—which keeps genuine RL feedback (the agent observes only the reward of the action it takes, never the correct action) while removing the trajectory-length variance that makes full-MDP return estimates noisy. A context $x \sim U[-1, 1]^4$ is drawn, the optimal action follows the *interaction* rule $a^\star(x) = \mathbb{1}[x_1x_2 - x_3x_4 + \frac{1}{2}x_1 > 0]$, and the reward is 1 iff the

sampled action matches $a^\star$. A VQC policy $\pi_\theta(a{=}1 \mid x) = \sigma(\alpha\langle Z_0\rangle)$ is trained by REINFORCE (Williams, 1992) on N contexts and we measure the reward gap against 2000 held-out contexts, fixing the rotation layout ($L{=}2$, $d{=}16$) and varying only the connectivity; each cell averages 16 seeds.

The prediction holds cleanly (Appendix B, Table 7; reward-space view in Appendix B, Figure 4). Held-out reward is well above chance (0.60–0.66), so the policies genuinely generalize, and at fixed parameter count the entangled ones generalize *worse*. Both entanglers are significant—`full` exceeds `none` by 3.2–3.9 standard errors at every N and `linear` by 2.1–4.2—so the effect does not rest on one entangler. At smaller N the ordering is the monotone none $<$ linear $<$ full predicted by d_{eff} (only at $N{=}100$, where all gaps are small, does `linear` edge `full`). As in the supervised case the entangled policies fit the training contexts better (0.86 vs. 0.74 at $N{=}16$) and the gap shrinks with N as Eq. (3) anticipates—the same mechanism, now for a policy learned from reward alone.

Multi-step value-function generalization. To go beyond one step we test the generalization of a *value function* in a genuine finite-horizon MDP. States $x \in [-1, 1]^4$ evolve by a fixed deterministic map, the per-step reward is the interaction rule, and the value $V^\star(x) = \sum_{t=0}^{T-1} \gamma^t r(x_t)$ ($T{=}6$, $\gamma{=}0.9$) accumulates future rewards along the trajectory. We fit a single-observable VQC value head to $V^\star$ on N states by Monte-Carlo regression—a standard policy-evaluation primitive—and report the R^2 gap against 1000 held-out states (Appendix B, Figure 4, right). At $N{=}64$, where held-out R^2 is positive for all circuits, the gap is monotone none $<$ linear $<$ full ($0.089 < 0.179 < 0.299$); at $N{=}16$ the entangled value functions have *negative* held-out R^2, so we treat that row as a stress test only. Scope: Monte-Carlo value estimation is a supervised-flavoured RL primitive, which is both its strength (low variance) and its limit. Under *bootstrapped* fitted-Q iteration the effect washes out—consistent with the bound, since the bootstrap target is itself a strong regularizer—so we claim the penalty only for the Monte-Carlo form.

7 Controls, Scope, and a Real-Hardware Test

The effect is not explained by the tested optimization confounders. Varying the entangling connectivity holds the parameter count fixed but could, in principle, change the gap through optimization or trainability rather than through d_{eff}. We rule this out with three controls at $N{=}64$ (Appendix B, Figure 5, left): *matched training accuracy* (early-stopping each circuit at the same train accuracy) still leaves `full` with $\sim 3\times$ the gap of `none` (0.093 vs. 0.032) and lower test accuracy; a *different readout* (Z_1) preserves the ordering ($0.074/0.126/0.131$); and a *different optimizer* (SGD) preserves it ($0.034/0.065/0.089$).

Local vs. global readout: the effect is the readout light-cone, and d_{eff} is the invariant. The most adversarial test of the mechanism is also the most revealing. Because our single readout is the *local* observable $\langle Z_0\rangle$, a rotation on a wire that never reaches qubit 0 has identically zero gradient (Lemma 2), so the unentangled `none` circuit is effectively a *smaller* model—only $\sim d/n$ of its parameters are causally live. One might therefore object that "entanglement raises the gap at fixed d" is merely "entanglement recruits otherwise-dead parameters." That objection is correct, and it *is* the mechanism; we test it by replacing the local readout with the *global* one $\frac{1}{n}\sum_i \langle Z_i\rangle$, under which every wire couples to the output directly (Appendix B, Table 4). The robust finding is that the global readout sharply raises `none`'s live-parameter count, d_{eff} *and* gap at both sizes—at $n{=}4$ from (2 live, $d_{\text{eff}}{=}2.6$, gap$=-0.008$) to $(7, 14.0, 0.095)$, at $n{=}6$ from $(2, 2.9, 0.018)$ to $(11, 23.6, 0.173)$: the readout alone activates the parameters entanglement would have, and d_{eff} and the gap move together. The `full`$-$`none` *contrast* is then n-dependent, which we report honestly: at $n{=}4$ one global observable nearly saturates the cone even for `none` (7 of 16 live) and the contrast is largely absorbed ($+0.112$, 4.1 s.e. local $\rightarrow$ $+0.033$, 0.9 s.e. global), whereas at $n{=}6$, where 24 parameters cannot be saturated by one observable, it *persists* ($+0.114$, 3.8 s.e. vs. $+0.115$, 4.2 s.e.). Connectivity is thus not dispensable but *readout-gated*: what governs d_{eff} is how many parameters are causally coupled to the readout, a global readout raises the floor of that coupling, and in *every* cell it is d_{eff}, not the entangling label, that tracks the gap. This confirms Proposition 1 and sharpens the thesis: connectivity is the dominant structural driver of d_{eff} for the local readouts standard in quantum RL (Jin et al., 2025; Lin et al., 2025), partly substitutable by, never orthogonal to, the readout.

Beyond one target, and on real data. The effect is not an artifact of one synthetic rule. On a random-Fourier target and on labels from an entangled teacher PQC the gap is again ordered none $<$ linear $<$ full; 4-bit parity is the instructive exception—it is not representable without entanglement, so `none` and `linear` sit at chance while `full` is the only circuit that *learns* it (test 0.90), which is exactly why the right principle is to budget entanglement, not remove it (Appendix B, Table 8). On standard real datasets—Iris, Breast Cancer (Wisconsin), Wine, and Digits (3 vs. 8), reduced to four features and binarized—we run two comparisons. The first (Iris, Breast Cancer, Wine at N=40, 6 seeds) orders the gap none $<$ linear $<$ full, co-monotone with d_{eff} (Table 9). The second, more important for the practical claim, is a higher-powered test of *held-out accuracy* on the two datasets where accuracy resolves a difference (Wine and Digits, pooling $N \in \{16, 24, 32\}$, 22 seeds): test accuracy drops from 0.884 to 0.870 on Wine (2.5 s.e.) and 0.911 to 0.900 on Digits (3.1 s.e.), far above chance. Entanglement measurably *hurts* test accuracy on these two datasets—not merely widening the gap—which is what makes "budget entanglement" operational rather than rhetorical. This does not hold dataset-by-dataset, however: on Iris (Table 9) `full` has both the largest gap *and* the highest accuracy (0.76 vs. 0.70 for `none`), so we scope the accuracy claim to the two datasets large enough to resolve it rather than assert it as universal; the gap ordering itself, which is what d_{eff} predicts, does hold on all three.

Where the mechanism holds, and where it stops. Two boundaries delimit its reach and we test both (Appendix B). In a deliberately *learnable* regime (n=6, lower-order target, N=48) all three circuits reach held-out accuracy 0.72–0.77 and the ordering survives cleanly (gaps $0.042/0.043/0.165$, `full`$-$`none`$= +0.123$ at 5.8 s.e., d_{eff} co-monotone $4.2/7.8/21.7$)—the effect is sharpest exactly where the circuits genuinely generalize. Barren plateaus (McClean et al., 2018; Cerezo et al., 2021), however, erode the mechanism quantifiably (Table 10): at fixed depth d_{eff} *falls* from 13.2 at n=4 to 3.1 at n=12 even as d rises from 32 to 96, since the eigenvalues shrink faster than the rank grows—connectivity inflates d_{eff} only off the plateau. A complementary ablation (fixing the pattern, varying depth) shows the gap rising in lock-step with d_{eff}, so d_{eff} predicts the gap either way, whereas parameter count tracks it only when it co-varies with d_{eff}.

End-to-end multi-step policy learning. Finally, we train policies *end-to-end* by REINFORCE (with a value baseline, horizon 8, 20 seeds) on the multi-step return of an interaction-gated MDP—a genuine policy-learning setting, not a value regression. The return generalization gap of the standard hardware-efficient (`linear`) entangler is *significantly* larger than the non-entangled circuit's at the small N=8 (0.62 vs. 0.16, 2.8 s.e.), confirming the mechanism end-to-end. The all-to-all circuit and the larger-N regime are too high-variance to resolve. We therefore report the `linear`-vs-`none` effect as genuine multi-step evidence, but continue to treat the low-variance bandit and value-function settings as our primary probes.

The effect survives on real quantum hardware. We run the trained 4-qubit classifiers on an IBM Heron device (`ibm_aachen`, 156 qubits) at 4096 shots (Appendix B, Figure 5, right; calibration and shot statistics in Appendix C). Despite real gate and readout noise—with no readout error mitigation—the models run above chance and the ordering is preserved (0.14 for `none` versus 0.33/0.31 for `linear`/`full`): the cost our bound predicts is measurable on today's hardware, not a simulation artifact.

8 Discussion and Conclusion

We gave a PAC-Bayesian account of generalization for quantum policies in which the operative complexity is the log-determinant complexity of the circuit-induced Fisher geometry—inflated by entangling connectivity through expansion of the readout's causal cone, and invisible to parameter counting. Isolating connectivity at fixed parameter *and* gate count, we showed it is an independent predictor of the gap, stronger than parameter count and than the learned-norm term, across supervised classification at 4–16 qubits, real datasets, a reward-only contextual bandit, multi-step value-function generalization, and hardware. As an ansatz-selection rule: prefer the smaller d_{eff} whenever the larger buys no validation gain—a criterion that, unlike a parameter budget, distinguishes circuits of identical size. *Limitations.* The bound is loose and certifies π_Q, on which it separates `none` from the entangled circuits at every N but resolves `linear` from `full` only at N=16 (Table 3), and our end-to-end policy evidence holds only for `linear` at small N.

Reproducibility Statement

Experiments use PQCs of n=4 to 16 qubits simulated with PennyLane (`default.qubit` with backprop at $n \leq 8$; `lightning.qubit` with the adjoint method at $n \geq 10$), plus one run on the IBM Heron `ibm_aachen` processor; unless a scaling experiment states otherwise (Table 6), the default is n=4. Ansätze, target and environment distributions, optimizer settings, the exact Fisher estimator, hyperparameters (τ^2=1, γ=50) and seed counts are specified in Sections 4–7 and Appendices B and C; source code will be released with the camera-ready version.

References

Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature computational science*, 1(6):403–409, 2021.

Leonardo Banchi, Jason Pereira, and Stefano Pirandola. Generalization in quantum machine learning: A quantum information standpoint. *PRX Quantum*, 2(4):040321, 2021.

Matthias C Caro, Hsin-Yuan Huang, Marco Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J Coles. Generalization in quantum machine learning from few training data. *Nature Communications*, 13(1):4919, 2022.

Marco Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*, 12: 1791, 2021.

Rodrigo Coelho, André Sequeira, and Luís Paulo Santos. VQC-based reinforcement learning with data re-uploading: Performance and trainability. *Quantum Machine Intelligence*, 6(2):53, 2024.

Gintare Karolina Dziugaite and Daniel M Roy. Computing nonvacuous generalization bounds for deep (stochastic) neural networks with many more parameters than training data. In *Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence (UAI)*, 2017.

Elies Gil-Fuster, Jens Eisert, and Carlos Bravo-Prieto. Understanding quantum machine learning also requires rethinking generalization. *Nature Communications*, 15(1):2277, 2024.

Yu-Xin Jin, Zi-Wei Wang, Hong-Ze Xu, Wei-Feng Zhuang, Meng-Jun Hu, and Dong E Liu. PPO-Q: Proximal policy optimization with parametrized quantum policies or values. *arXiv preprint arXiv:2501.07085*, 2025.

Hsin-Yi Lin, Samuel Yen-Chi Chen, Huan-Hsin Tseng, and Shinjae Yoo. Quantum reinforcement learning by adaptive non-local observables. In *2025 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 2, pp. 241–246. IEEE, 2025.

Jie Luo, Jeremy Kulcsar, Xueyin Chen, Giulio Giaconi, and Georgios Korpas. Training hybrid deep quantum neural network for efficient reinforcement learning. *arXiv preprint arXiv:2503.09119*, 2025.

Anirudha Majumdar, Alec Farid, and Anoopkumar Sonar. PAC-Bayes control: Learning policies that provably generalize to novel environments. *The International Journal of Robotics Research*, 40(2-3):574–593, 2021.

Andreas Maurer. A note on the PAC Bayesian theorem. *arXiv preprint cs/0411099*, 2004.

David A McAllester. PAC-Bayesian model averaging. In *Proceedings of the twelfth annual conference on Computational learning theory*, pp. 164–170, 1999.

Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9:4812, 2018.

David A Meyer and Nolan R Wallach. Global entanglement in multiparticle systems. *Journal of Mathematical Physics*, 43(9):4273–4278, 2002.

Pablo Rodriguez-Grasa, Matthias C Caro, Jens Eisert, Elies Gil-Fuster, Franz J Schreiber, and Carlos Bravo-Prieto. A PAC-Bayesian approach to generalization for quantum models. *arXiv preprint arXiv:2603.22964*, 2026.

Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99(3):032331, 2019.

Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8:229–256, 1992.

Shaojun Wu, Shan Jin, Dingding Wen, Donghong Han, and Xiaoting Wang. Quantum reinforcement learning in continuous action space. *Quantum*, 9:1660, 2025.

Abdelkrim Zitouni, Mehdi Hennequin, Juba Agoun, Ryan Horache, Nadia Kabachi, and Omar Rivasplata. PAC-Bayesian reinforcement learning trains generalizable policies. *arXiv preprint arXiv:2510.10544*, 2025.

A PROOFS

A.1 PROOF OF THEOREM 1

We reduce the return bound to the classical McAllester PAC-Bayes theorem. Recall its statement for a loss taking values in $[0, 1]$.

Theorem 3 (McAllester, 1999; tightened by Maurer, 2004)**.** *Let $\ell(h, z) \in [0, 1]$, let $z_1, \dots, z_m$ be i.i.d. from a distribution $\mathcal{D}$, and let P be a prior over hypotheses chosen before seeing the data. For any $\delta \in (0, 1)$, with probability at least $1 - \delta$, simultaneously for all posteriors Q,*

$$\mathbb{E}_{h\sim Q}\,\mathbb{E}_{z\sim\mathcal{D}}\,\ell(h, z) \;\le\; \mathbb{E}_{h\sim Q}\,\frac{1}{m}\sum_{i=1}^{m}\ell(h, z_i) + \sqrt{\frac{\mathrm{KL}(Q\|P) + \ln\frac{2\sqrt{m}}{\delta}}{2m}}. \tag{5}$$

Take the hypothesis h to be a parameter vector θ (so Q, P are the Gaussians of Section 3), the sample $z_i = M_i$ to be the i-th training environment drawn i.i.d. from $\mathcal{D}$, and the loss to be the normalized expected regret $\ell(\theta, M) = 1 - R_M(\theta)/R_{\max} \in [0, 1]$ with $R_M(\theta) = \mathbb{E}_{\tau\sim\pi_\theta, M}[R(\tau)]$. Taking the *expected* return matters here: a single sampled rollout would carry policy and environment randomness beyond (θ, M), so ℓ would not be a function of the sample as McAllester's theorem requires. With the expectation, $\ell(\theta, M)$ is a deterministic $[0, 1]$-valued function of the drawn environment, which is exactly the setting of Theorem 3.

Which loss each experiment instantiates. The two families of experiments enter the bound differently, and we keep them separate. For supervised classification we instantiate Theorem 3 directly with the bounded $0/1$ loss of the *deterministic* decision rule $\hat{y} = \mathrm{sign}\langle Z_0\rangle$; the accuracy we report is that loss, not the expected reward of a stochastic policy. For the contextual bandit, Theorem 1 applies to the expected reward of the *stochastic* policy $\pi_\theta(a{=}1 \mid x) = \sigma(\alpha\langle Z_0\rangle)$, which we compute in closed form rather than by sampling actions. Both are bounded losses over i.i.d. contexts, so Theorem 3 applies in either case; what differs is only which predictor the loss refers to. Then $\mathbb{E}_Q\mathbb{E}_M\ell = 1 - J(\pi_Q)/R_{\max}$ and its empirical counterpart is $1 - \hat{J}(\pi_Q)/R_{\max}$. Substituting into Theorem 3 and multiplying by $R_{\max}$ gives

$$J(\pi_Q) \;\ge\; \hat{J}(\pi_Q) - R_{\max}\sqrt{\frac{\mathrm{KL}(Q\|P) + \ln\frac{2\sqrt{m}}{\delta}}{2m}}. \tag{6}$$

For a Gaussian posterior/prior with common covariance $\sigma^2 I_d$, $\mathrm{KL}(Q\|P) = \|\theta{-}\theta_0\|^2/(2\sigma^2)$, which is Eq. (1). Every step here is the i.i.d. McAllester argument applied to a bounded per-environment loss; nothing about mixing enters. If one instead pools the Markov-dependent transitions of a trajectory, the effective sample size is reduced by a mixing factor (Zitouni et al., 2025); we state this only to delimit Theorem 1 and make no use of it. □

A.2 Proof of Theorem 2

The KL divergence between two d-dimensional Gaussians $Q = \mathcal{N}(\theta, \Sigma)$ and $P = \mathcal{N}(\theta_0, \tau^2 I_d)$ is

$$\mathrm{KL}(Q\|P) = \tfrac{1}{2}\Big[\tfrac{1}{\tau^2}\operatorname{tr}\Sigma + \tfrac{1}{\tau^2}\|\theta-\theta_0\|^2 - d - \log\det\tfrac{\Sigma}{\tau^2}\Big]. \tag{7}$$

Substitute the Gauss–Newton covariance $\Sigma = \tau^2(I_d + \gamma\hat{\mathcal{F}})^{-1}$ and diagonalize $\hat{\mathcal{F}} = \sum_i \lambda_i u_i u_i^\top$. Then $\frac{1}{\tau^2}\operatorname{tr}\Sigma = \sum_i (1+\gamma\lambda_i)^{-1}$, $\log\det\frac{\Sigma}{\tau^2} = -\sum_i \log(1+\gamma\lambda_i)$, and using $d = \sum_i 1 = \sum_i \big[(1+\gamma\lambda_i)^{-1} + \frac{\gamma\lambda_i}{1+\gamma\lambda_i}\big]$ the trace and $-d$ terms combine to $-\sum_i \frac{\gamma\lambda_i}{1+\gamma\lambda_i}$. Collecting terms,

$$\mathrm{KL}(Q\|P) = \frac{\|\theta-\theta_0\|^2}{2\tau^2} + \tfrac{1}{2}\sum_i \log(1+\gamma\lambda_i) - \tfrac{1}{2}\sum_i \frac{\gamma\lambda_i}{1+\gamma\lambda_i}, \tag{8}$$

which is (2) since $\sum_i \log(1+\gamma\lambda_i) = \log\det(I_d + \gamma\hat{\mathcal{F}}) = d_{\mathrm{eff}}(\gamma)$. Each summand of the last term lies in $[0, \frac{1}{2})$, so $0 \le \frac{1}{2}\sum_i \frac{\gamma\lambda_i}{1+\gamma\lambda_i} \le d/2$ and dropping it can only enlarge the right-hand side of the bound. The prior P is fixed at initialization θ_0 and does not depend on the training environments, so Theorem 3 applies verbatim with this Q and P; substituting the resulting KL into the return bound of Theorem 1 gives (3). The asymptotics $d_{\mathrm{eff}}(\gamma) = \operatorname{rank}(\hat{\mathcal{F}})\log\gamma + O(1)$ as $\gamma \to \infty$ follow from $\log(1+\gamma\lambda_i) = \log\gamma + \log\lambda_i + o(1)$ for $\lambda_i > 0$, each zero eigenvalue contributing exactly 0. We stress that $d_{\mathrm{eff}}(\gamma)$ is *not* bounded by d: the elementary bound is $\log(1+\gamma\lambda_i) \le \gamma\lambda_i$, giving only $d_{\mathrm{eff}}(\gamma) \le \gamma \operatorname{tr}\hat{\mathcal{F}}$, and d_{eff} does exceed d in our measurements whenever the spectrum is large enough (Table 11). It is $\operatorname{rank}(\hat{\mathcal{F}}) \le d$, not d_{eff}, that the parameter count bounds; d_{eff} is an unnormalized log-volume of the Fisher geometry. □

A.3 Derandomization: from the stochastic π_Q to the deterministic θ

The bound (3) controls the gap of the posterior-averaged predictor π_Q, whereas the reported "true gap" is that of the deterministic trained θ. The following lemma bounds the difference and, crucially, shows the flat Fisher directions—which carry the full prior variance τ^2—do *not* spoil it, because the perturbation of a loss along a direction is weighted by that direction's curvature, which is the Fisher eigenvalue itself.

Lemma 1 (Flat-direction derandomization)**.** *Let $Q = \mathcal{N}(\theta, \Sigma)$ with $\Sigma = \tau^2(I_d + \gamma\hat{\mathcal{F}})^{-1}$, and let $L(\vartheta) = \mathbb{E}_s\,\ell(\vartheta, s)$ be a population loss whose Hessian at θ equals the Fisher $\hat{\mathcal{F}}(\theta)$ up to a residual term $\mathcal{R}$ that is $O(\|p-y\|)$ and vanishes in the Gauss–Newton (well-fit) regime. Then*

$$\big|\,\mathbb{E}_{\vartheta\sim Q}\,L(\vartheta) - L(\theta)\,\big| \;=\; \tfrac{1}{2}\operatorname{tr}\big(\Sigma\,\hat{\mathcal{F}}\big) + O(\|\Sigma\|^{3/2}) \;=\; \frac{\tau^2}{2}\sum_{i=1}^{d}\frac{\lambda_i}{1+\gamma\lambda_i} \;+\; O(\|\Sigma\|^{3/2}), \tag{9}$$

and the sum is bounded by $\dfrac{1}{\gamma}\cdot\dfrac{1}{2}\sum_i \dfrac{\gamma\lambda_i}{1+\gamma\lambda_i} \le \dfrac{d}{2\gamma}$. Consequently the train and test losses of π_Q and of θ each differ by at most $\frac{1}{2}\tau^2\sum_i \lambda_i/(1+\gamma\lambda_i) \le \tau^2 d/(2\gamma)$, so the two generalization gaps differ by at most $\tau^2\sum_i \lambda_i/(1+\gamma\lambda_i) \le \tau^2 d/\gamma$.

Proof. Since $\mathbb{E}_{\vartheta\sim Q}[\vartheta - \theta] = 0$, a second-order Taylor expansion of L about θ gives $\mathbb{E}_Q L(\vartheta) - L(\theta) = \frac{1}{2}\,\mathbb{E}_Q[(\vartheta-\theta)^\top\nabla^2 L\,(\vartheta-\theta)] + O(\|\Sigma\|^{3/2}) = \frac{1}{2}\mathrm{tr}(\Sigma\nabla^2 L) + O(\|\Sigma\|^{3/2})$, the remainder controlled by the third derivative of the smooth PQC expectation and the Gaussian third moment. For the logistic head, $\nabla^2 L = \mathbb{E}_s[\alpha^2 p_s(1-p_s)\nabla f_s \nabla f_s^\top] + \mathbb{E}_s[\alpha(p_s - y_s)\nabla^2 f_s] = \hat{\mathcal{F}} + \mathcal{R}$, and $\mathcal{R}$ vanishes as the fit improves ($p_s \to y_s$). Substituting $\Sigma = \tau^2(I_d + \gamma\hat{\mathcal{F}})^{-1}$ and diagonalizing $\hat{\mathcal{F}}$ gives $\operatorname{tr}(\Sigma\hat{\mathcal{F}}) = \tau^2\sum_i \lambda_i/(1+\gamma\lambda_i)$, which is (9). Writing $\lambda_i/(1+\gamma\lambda_i) = \gamma^{-1}\cdot\gamma\lambda_i/(1+\gamma\lambda_i)$ and $\gamma\lambda_i/(1+\gamma\lambda_i) \in [0,1)$ gives the $d/2\gamma$ envelope. A flat direction ($\lambda_i \to 0$) contributes $\lambda_i/(1+\gamma\lambda_i) \to \lambda_i \to 0$ despite its posterior variance $\Sigma_{ii} \to \tau^2$: the variance is large but the curvature that converts it into a loss change is exactly λ_i, so flat directions are harmless. The gap statement follows by applying the bound to the train and test losses separately and using the triangle inequality. □

Status of Lemma 1. The expansion above is local: the remainder is $O(\|\Sigma\|^{3/2})$ with a constant set by the third derivative of the PQC expectation, and the identification $\nabla^2 L = \hat{\mathcal{F}} + \mathcal{R}$ discards a residual that vanishes only in the well-fit limit. Neither is quantified for our circuits. We therefore use the lemma as a *qualitative* account of why flat Fisher directions do not spoil derandomization—the posterior variance is largest exactly where the output is least sensitive—and not as a rigorous transfer of the certificate from π_Q to θ. A Monte-Carlo evaluation of π_Q needs no such bridge, and we report it in Table 3: the measured discrepancy between the two gaps stays inside the envelope this lemma predicts, so the argument is empirically consistent even though we do not claim it as a theorem.

Note that the derandomization cost $\frac{1}{2}\tau^2 \sum_i \lambda_i/(1+\gamma\lambda_i)$ is exactly $1/\gamma$ times the KL curvature term $\frac{1}{2}\sum_i \gamma\lambda_i/(1+\gamma\lambda_i)$ that we dropped in Eq. (2): the same quantity we discarded to expose d_{eff} reappears, scaled by $1/\gamma$, as the price of derandomization, and $\gamma{=}50$ makes it small ($\leq d/100$ for $\tau^2{=}1$). This ties the two approximations together and replaces the informal "low-temperature" argument in the main text.

A.4 Proof of Proposition 1

Write the readout expectation in the Heisenberg picture, $f(s,\theta) = \langle\psi(s)|\, U(\theta)^\dagger Z_0\, U(\theta)\,|\psi(s)\rangle$, and let gate i be a Pauli rotation $e^{-\mathrm{i}\theta_i G_i/2}$ with G_i a single-qubit Pauli on wire $q(i)$. Splitting the circuit at gate i as $U = U_{>i}\, e^{-\mathrm{i}\theta_i G_i/2}\, U_{\leq i}$ and defining the back-propagated observable $O_i(\theta) = U_{>i}^\dagger Z_0\, U_{>i}$, the parameter-shift identity (Schuld et al., 2019) gives

$$\partial_{\theta_i} f(s,\theta) = -\tfrac{\mathrm{i}}{2}\,\langle\phi_i(s)|\,[\,G_i,\, O_i(\theta)\,]\,|\phi_i(s)\rangle, \qquad |\phi_i(s)\rangle = U_{\leq i}|\psi(s)\rangle. \tag{10}$$

The (output-)Fisher matrix is $\hat{\mathcal{F}}_{ij} = \mathbb{E}_s[\,c(s)\,\partial_{\theta_i} f\,\partial_{\theta_j} f\,]$ with $c(s) = \mathrm{sc}^2 p(1{-}p) > 0$, so $\hat{\mathcal{F}} = \mathbb{E}_s[c(s)\, g_s g_s^\top]$ where $g_s = \nabla_\theta f(s,\theta)$; in particular $\operatorname{rank}(\hat{\mathcal{F}}) \leq |\{i :\ \partial_{\theta_i} f \not\equiv 0\}|$, with equality generically.

We separate two objects that the informal phrase "light-cone" conflates: a *structural* set fixed by the circuit's wiring, which is what behaves monotonically, and the *active* set of parameters that actually carry gradient, which is a subset of it and depends on commutation relations and on θ.

Definition 1 (Potential causal cone and active set)**.** *For each gate index i, let $\operatorname{cone}(i) \subseteq \{0,\ldots,n{-}1\}$ be the set of wires reachable backwards from the readout wire 0 through the two-qubit gates lying in $U_{>i}$, defined purely on the circuit's connectivity graph (no dependence on θ). The* potential causal cone *is $\mathcal{C} = \{i :\ q(i) \in \operatorname{cone}(i)\}$, and the* active set *is $\mathcal{L}(\theta) = \{i :\ [G_i, O_i(\theta)] \neq 0\}$, the parameters with a non-vanishing gradient in (10).*

Lemma 2 (Structural monotonicity, and strictness of the inclusion)**.** *Fix the placement of all rotation gates.* (i) *$\mathcal{L}(\theta) \subseteq \mathcal{C}$ for every θ.* (ii) *Adding two-qubit entangling gates (keeping all rotations) can only enlarge $\mathcal{C}$: $\mathcal{C}_{\text{without}} \subseteq \mathcal{C}_{\text{with}}$.* (iii) *With no entangling gates, $\mathcal{C} = \{i : q(i) = 0\}$, so at most d/n parameters can be active.* (iv) *The inclusion in (i) is in general* strict*: a rotation inside the cone whose generator commutes with its back-propagated observable is inert—in particular the last R_Z on the readout wire, for which $O_i = Z_0$ and $[Z_0, Z_0] = 0$.*

Proof. (i) $O_i(\theta) = U_{>i}^\dagger Z_0 U_{>i}$ is Z_0 conjugated by the gates after i. Single-qubit rotations act within their own wire and two-qubit gates can spread support only along the edges they realize, so the support of $O_i(\theta)$ is contained in $\operatorname{cone}(i)$ for every θ. If $q(i) \notin \operatorname{cone}(i)$ then O_i acts as the identity on wire $q(i)$ and $[G_i, O_i] = 0$, so $\partial_{\theta_i} f \equiv 0$ by (10) and $i \notin \mathcal{L}(\theta)$. (ii) $\operatorname{cone}(i)$ is defined by reachability in the connectivity graph, and adding edges cannot reduce a reachable set. (iii) With no two-qubit gates the only reachable wire is 0 itself. (iv) is the displayed counterexample: containment of the support in the cone is necessary for a non-zero gradient, not sufficient, because the commutator may still vanish. □

Note what (i)–(ii) do *not* say. It is not true in general that the Pauli support of $O_i(\theta)$ grows monotonically when gates are added—conjugation can cancel terms, so the realized support may shrink at particular θ. What is monotone is the structural set $\mathcal{C}$, which upper-bounds the support for *all* θ. This is why we phrase Proposition 1 in terms of $\mathcal{C}$ and treat the passage from $\mathcal{C}$ to d_{eff} as generic rather than universal.

Assumption 1 (Genericity and sufficient sampling)**.** *For generic parameters the nonzero gradients* $\{g_x\}$ *span* $\mathrm{span}\{e_i : i \in \mathcal{L}(\theta)\}$ *and the induced Fisher eigenvalues on this subspace are of comparable magnitude (no exact degeneracies or cancellations). For the* empirical *Fisher we additionally require* $M \geq |\mathcal{L}(\theta)|$*, so that the* M *sampled gradients can span the active subspace.*

Proof of Proposition 1. (ii) Both the population Fisher $\mathcal{F} = \mathbb{E}_x[c(x) g_x g_x^\top]$ and its empirical counterpart $\hat{\mathcal{F}} = \frac{1}{M}\sum_{x=1}^{M} c(x) g_x g_x^\top$ are supported on $\mathrm{span}\{e_i : i \in \mathcal{L}(\theta)\}$, since parameters outside the active set have $\partial_{\theta_i} f \equiv 0$ by (10). For the population matrix this gives $\mathrm{rank}(\mathcal{F}) \leq |\mathcal{L}(\theta)|$, with equality generically. The empirical matrix is a sum of M rank-one terms, so it obeys the stronger constraint $\mathrm{rank}(\hat{\mathcal{F}}) \leq \min\{|\mathcal{L}(\theta)|, M\}$: with M inputs one cannot realize more than M nonzero eigenvalues however many parameters are active. Equality holds generically once $M \geq |\mathcal{L}(\theta)|$ and the sampled gradients span the active subspace (Assumption 1). Finally $|\mathcal{L}(\theta)| \leq |\mathcal{C}| \leq d$ by Lemma 2(i). (i) and (iii) are Lemma 2(ii) and (iii). Strictness in (i) holds as soon as an added gate puts a rotation wire into $\mathrm{cone}(i)$, since that index enters $\mathcal{C}$ while d is unchanged. □

Population versus empirical rank. The distinction matters for the numbers we report. The generic rank of the population Fisher is $|\mathcal{L}(\theta)|$; that of the empirical Fisher is $\min\{M, |\mathcal{L}(\theta)|\}$. We use M=40 held-out inputs throughout (Appendix C), so the sampling constraint is inactive in the primary n=4 setting, where d=16 and the largest observed $|\mathcal{L}|$ is 12 (Table 4). It would bind at the top of our scaling study—at n=16, d=64 exceeds M—so the d_{eff} values in Table 6 should be read as computed from a rank-≤ 40 estimate of the Fisher rather than from the full spectrum. This affects the absolute d_{eff} scale at large n, not the within-n comparisons between connectivities, which all share the same M.

From the cone to d_{eff}: what does *not* follow. Proposition 1 bounds and orders *ranks*, and this is where the rigour stops. Two further steps are generic rather than universal, and we flag both. First, an index entering $\mathcal{C}$ need not enter $\mathcal{L}(\theta)$: the commutator in (10) may vanish (Lemma 2(iv)) or the parameters may be fine-tuned to a measure-zero cancellation. Second, and more importantly for the finite-γ quantity we actually measure, a rank increase raises $d_{\mathrm{eff}}(\gamma) = \sum_i \log(1 + \gamma\lambda_i)$ only if the newly activated eigenvalues are not vanishingly small: a mode with $\lambda_i \approx 0$ contributes ≈ 0 to d_{eff} however large the rank. Unconditional monotonicity of d_{eff} therefore holds only in the limit $\gamma \to \infty$, where $d_{\mathrm{eff}} \to \mathrm{rank}(\hat{\mathcal{F}}) \log \gamma$. At the fixed γ=50 used throughout, the statement "more entangling connectivity ⇒ larger d_{eff}" is an *empirical observation*, which we report as such and which Table 10 shows failing in exactly the predicted regime (barren plateaus, where all λ_i decay). The empirical counterpart is Table 11: at fixed depth (fixed d) the measured $d_{\mathrm{eff}} = \log\det(I + \gamma\hat{\mathcal{F}})$ rises monotonically from its no-entanglement value (none → linear → full), tracking the rank of the Fisher matrix. The strictness of $\mathcal{L} \subsetneq \mathcal{C}$ is also visible: at n=4, L=2 the unentangled circuit has $|\mathcal{C}| = 4$ (the four rotations on wire 0) but only 2 numerically live Fisher directions (Table 4). The final R_Z on the readout wire is inert by Lemma 2(iv). But the remaining gap between $|\mathcal{C}|$=4 and the 2 numerically live directions is *not* a numerical-threshold artifact: it is exact. With a single real encoding angle x_0 entering only through $R_Y(x_0)$, any sequence of single-qubit rotations confined to wire 0 leaves the encoded state in the XZ-plane of the Bloch sphere, so $\langle Y_0 \rangle \equiv 0$ and the readout is exactly $\langle Z_0 \rangle = a(\theta)\cos x_0 + b(\theta)\sin x_0$ for some a, b depending on all of wire 0's parameters. The gradient with respect to *any* such parameter is therefore a linear combination of $\cos x_0$ and $\sin x_0$ alone, so $\mathrm{rank}(\hat{\mathcal{F}}) \leq 2$ *exactly*, however many rotation layers L sit on wire 0. This is a case where Assumption 1's spanning condition fails, and it fails *structurally*—for every θ in this family, not on a measure-zero set—which is why we state $\mathrm{rank}(\hat{\mathcal{F}}) \leq |\mathcal{L}(\theta)| \leq |\mathcal{C}|$ as an inequality chain and never invoke the genericity equality for the `none` connectivity. We record this explicitly as failure mode (iv) below.

Scope, a worked example, and failure modes. We state precisely what Proposition 1 does and does not give. *Scope.* It applies to hardware-efficient ansätze in which (a) each trainable gate is a Pauli rotation, (b) the readout is a fixed Pauli observable, and (c) encoding gates precede the trainable block. The three claims must be kept apart, in decreasing order of strength. *(a) Structural causal-cone monotonicity* ($\mathcal{C}_{\mathrm{without}} \subseteq \mathcal{C}_{\mathrm{with}}$) is rigorous: it is Lemma 2(ii), a statement about reachability in the connectivity graph, with no genericity and no dependence on θ. *(b) The corresponding increase in the number, and hence the rank, of* active *Fisher directions* is *generic rather than universal*.

We do *not* claim $\mathrm{rank}(\hat{\mathcal{F}}_{\mathrm{with}}) \geq \mathrm{rank}(\hat{\mathcal{F}}_{\mathrm{without}})$ as a theorem: enlarging $\mathcal{C}$ enlarges the set of parameters that *may* be active, but nothing in the argument prevents a previously active direction from becoming inactive, or the back-propagated observable's realized support from shrinking at particular θ through cancellation. What (a) gives is monotonicity of the *upper envelope* $\mathrm{rank}(\hat{\mathcal{F}}) \leq |\mathcal{L}(\theta)| \leq |\mathcal{C}|$; the rank itself moving with it is generic, and we verify it empirically in our setting (Table 11, Figure 6). *(c) Effective-dimension monotonicity* $d_{\mathrm{eff}}(\gamma)_{\mathrm{with}} \geq d_{\mathrm{eff}}(\gamma)_{\mathrm{without}}$ is weaker still, holding unconditionally only as $\gamma \to \infty$; at finite γ a rank increase raises d_{eff} only if the new eigenvalues are not vanishingly small, since a mode with $\lambda \approx 0$ contributes ≈ 0. Only (a) is a theorem; (b) and (c) are empirical regularities of the regime we study.

Worked example. Take n=2 qubits, one layer, readout Z_0, trainable $R_Y(\theta_0)$ on qubit 0 and $R_Y(\theta_1)$ on qubit 1. Without a CNOT, $\partial_{\theta_1}\langle Z_0\rangle \equiv 0$ (the back-propagated observable is $Z_0 \otimes I$, which commutes with Y_1), so $\hat{\mathcal{F}} = \mathrm{diag}(c\langle\dots\rangle, 0)$ has rank 1. Inserting $\mathrm{CNOT}_{0\to1}$ before the readout maps $Z_0 \mapsto Z_0$ but $\mathrm{CNOT}_{1\to0}$ maps $Z_0 \mapsto Z_0Z_1$, which no longer commutes with Y_1; then $\partial_{\theta_1}\langle Z_0\rangle = -\frac{\mathrm{i}}{2}\langle[Y_1, Z_0Z_1]\rangle \neq 0$ generically and $\mathrm{rank}(\hat{\mathcal{F}}) = 2$. This is the mechanism in closed form.

Failure modes. The rank does *not* increase when the added entanglement is neutralized by structure: (i) a symmetry or commuting-gate configuration for which $[G_i, O_i(\theta)] = 0$ despite non-trivial support (e.g. O_i proportional to $Z_{q(i)}$ and $G_i = Z_{q(i)}$); (ii) fine-tuned parameters at which the new gradients vanish (a measure-zero set excluded by genericity); (iii) the barren-plateau regime, where all $\lambda_i \to 0$ exponentially, so although the *rank* may rise the finite-γ d_{eff} stays near zero—entanglement then inflates capacity only once the circuit is trained away from the plateau; and (iv) a *single active wire*: with one real encoding angle and no entangling gates, the Bloch-sphere argument above caps $\mathrm{rank}(\hat{\mathcal{F}}) \leq 2$ for the `none` connectivity regardless of depth. Unlike (ii), (iv) is not a measure-zero exception—it is the generic behavior of `none` throughout this paper—so we do not avoid it; we rely instead on Lemma 2(ii)'s structural statement about $\mathcal{C}$, which holds independently of how tight the rank bound is for any one connectivity, and on the empirical rank and d_{eff} values we report rather than on an assumed equality. (ii) and (iii) delimit where trained parameters can defeat the mechanism; (iv) is why we never claim more than $\mathrm{rank}(\hat{\mathcal{F}}) \leq |\mathcal{C}|$ for `none` specifically.

B Extended results

This appendix collects the supporting experiments referenced in the main text: the numerical evaluation of the certificate and its Monte-Carlo evaluation on the stochastic predictor, the accuracy–complexity trade-off, the qubit-count scaling study, the bandit and value-function figures, alternative targets and real datasets, the barren-plateau boundary, the depth ablation, the optimization controls and the hardware run, and the per-configuration Fisher table and spectrum.

The certificate, evaluated. Plugging the measured $\|\theta - \theta_0\|^2$ and $d_{\mathrm{eff}}(\gamma)$ into the complexity $C = \|\theta - \theta_0\|^2/(2\tau^2) + \frac{1}{2}d_{\mathrm{eff}}(\gamma)$ (with τ^2=1, γ=50, δ=0.05, $R_{\max}$=1) gives the certified bound on the gap, $\sqrt{(C + \ln\frac{2\sqrt{m}}{\delta})/2m}$, reported in Table 2 next to the deterministic gap. The bound is non-vacuous, loose by $\approx$ 2–4.5$\times$ where the gap is non-negligible, and correctly ordered at fixed parameter count—the ranking property the main text relies on. A parameter-counting complexity would be *identical* across the three rows of each block and could not order them at all. The deterministic column is included for reference only; the bound is a statement about π_Q, and Table 3 evaluates it against π_Q.

Evaluating the certified object directly: Monte-Carlo π_Q. The bound certifies the stochastic predictor π_Q, so we evaluate π_Q rather than relying on Lemma 1 to transfer the statement to θ. For each trained circuit we form the Gauss–Newton posterior $Q = \mathcal{N}(\theta, \Sigma)$, $\Sigma = \tau^2(I + \gamma\hat{\mathcal{F}})^{-1}$ with the same τ^2=1, γ=50 used everywhere, draw K=64 parameter vectors $\vartheta \sim Q$ per seed, and average the train and held-out accuracy over the draws; the *stochastic* gap is the difference of these posterior-averaged accuracies, which is exactly the object Eq. (3) bounds. Table 3 reports both gaps and the certificate over 8 seeds; the deterministic column reproduces Table 1 to within rounding, confirming the two runs share the training pipeline.

Table 2: The PAC-Bayes bound of Eq. (3) evaluated per circuit (L=2, τ^2=1, γ=50, δ=0.05), alongside the gap of the *deterministic* trained θ. The bound is non-vacuous and loose by only $\sim$2–4.5$\times$ where the gap is non-negligible, and—driven by d_{eff}—is correctly ordered at fixed parameter count. The last column is a numerical reference point, not a certified quantity: the bound certifies π_Q, and Table 3 checks it against π_Q directly.

N	connectivity	complexity C	certified gap bound	true gap
16	`none`	2.7	0.49	0.101
16	`linear`	8.6	0.65	0.146
16	`full`	19.2	0.87	0.381
64	`none`	2.2	0.25	−0.01
64	`linear`	6.2	0.31	0.064
64	`full`	14.2	0.40	0.104

Three things follow, and the third is a limitation. *(i)* The certificate is valid for the object it certifies: the bound exceeds the stochastic gap in every cell. *(ii)* `none` has the smallest stochastic gap at every N, so the claim that matters—an entangling circuit generalizes worse than the non-entangled circuit of identical size—holds for π_Q and not only for θ. *(iii)* The *three-way* ordering, however, transfers to π_Q only at N=16 ($0.095 < 0.123 < 0.168$). From N=32 on, `linear` and `full` exchange places (0.093 vs. 0.071 at N=32; 0.031 vs. 0.029 at N=64; 0.034 vs. 0.006 at N=128), the first two within one standard error and the last outside it. The reason is visible in the same table: posterior averaging shrinks the gap most where d_{eff} is largest (`full`: 0.381 $\to$ 0.168 at N=16), because a circuit whose Fisher rank is 5 out of d=16 carries eleven nearly flat directions that receive the full prior variance τ^2 and perturb it away from its sharper training fit. The bound's complexity term does not model this shrinkage, so it orders π_Q correctly only where the shrinkage is small. The largest gap discrepancy, 0.21, does lie inside the envelope $\tau^2 d/\gamma = 0.32$ that Lemma 1 predicts, so the derandomization argument is empirically consistent even though we do not claim it as a theorem.

Table 3: Monte-Carlo evaluation of the certified object (L=2, d=16, τ^2=1, γ=50, δ=0.05, K=64 posterior draws, 8 seeds, mean $\pm$ sem). The stochastic gap is the train−test accuracy difference of the posterior-averaged predictor π_Q—the quantity Eq. (3) actually bounds—and needs no derandomization step. The bound is valid for π_Q in every cell, and `none` has the smallest stochastic gap at every N. The full three-way ordering, however, is resolved only at N=16: from N=32 on, `linear` and `full` swap, mostly within one standard error. We report this rather than restrict the sweep.

N	connectivity	d_{eff}	stochastic gap (π_Q)	deterministic gap (θ)	bound
16	`none`	3.3	$\mathbf{0.095 \pm 0.031}$	0.101	0.49
16	`linear`	9.3	0.123 ± 0.026	0.146	0.64
16	`full`	21.8	0.168 ± 0.028	0.381	0.86
32	`none`	2.8	$\mathbf{0.042 \pm 0.027}$	0.048	0.35
32	`linear`	9.1	0.093 ± 0.023	0.169	0.49
32	`full`	19.3	0.071 ± 0.023	0.225	0.57
64	`none`	2.6	$\mathbf{-0.004 \pm 0.017}$	−0.008	0.25
64	`linear`	8.0	0.031 ± 0.014	0.064	0.31
64	`full`	12.7	0.029 ± 0.022	0.104	0.39
128	`none`	2.7	$\mathbf{-0.005 \pm 0.015}$	−0.016	0.18
128	`linear`	7.7	0.034 ± 0.010	0.062	0.22
128	`full`	9.9	0.006 ± 0.019	0.052	0.24

Local vs. global readout. Table 4 gives the per-cell numbers behind the readout control of Section 7: live-parameter counts, d_{eff} and gap under the local $\langle Z_0 \rangle$ and the global $\frac{1}{n}\sum_i \langle Z_i \rangle$ readout, at n=4 and n=6.

Table 4: Local vs. global readout (L=2, N=64, 8 seeds). A local readout leaves most of `none`'s parameters outside the causal cone, so recruiting them makes `full`−`none` significant; a global readout activates them directly, absorbing the contrast fully at n=4 but only partly at n=6. In every row d_{eff} tracks the gap, the entangling label does not.

n	readout	connectivity	live params	d_{eff}	gap
4	local $\langle Z_0 \rangle$	`none`	2/16	2.6	−0.008
4	local $\langle Z_0 \rangle$	`full`	5/16	12.9	0.104
4	global $\frac{1}{n}\sum_i Z_i$	`none`	7/16	14.0	0.095
4	global $\frac{1}{n}\sum_i Z_i$	`full`	12/16	30.7	0.128
6	local $\langle Z_0 \rangle$	`none`	2/24	2.9	0.018
6	local $\langle Z_0 \rangle$	`full`	7/24	21.1	0.132
6	global $\frac{1}{n}\sum_i Z_i$	`none`	11/24	23.6	0.173
6	global $\frac{1}{n}\sum_i Z_i$	`full`	17/24	35.4	0.289

Separating gate count from causal-cone reach. The three ansätze of Section 4 differ in both two-qubit gate count and cone reach, so on their own they cannot say which drives the gap. Table 5 fixes the gate count at exactly three CNOTs and varies only the cone. `disjoint`=$\{(1,2),(2,3),(1,3)\}$ places all three gates away from the readout wire, so by Lemma 2(iii) its cone is $\{0\}$; `star`=$\{(0,1),(0,2),(0,3)\}$ and `linear` both reach every wire.

The result is unambiguous: `disjoint` is *numerically identical* to `none` in gap, test accuracy, d_{eff} and live-parameter count at both sample sizes—its three entangling gates are invisible to $\langle Z_0 \rangle$—whereas `linear` and `star`, with the same three gates, reach gaps of 0.146 and 0.372 at N=16. This identity is not a surprise but a consequence: if qubit 0 participates in no two-qubit gate, the circuit factorizes and $\langle Z_0 \rangle$ depends on the wire-0 parameters alone, so the optimization trajectory coincides exactly. That is what makes it a clean control—it exhibits three CNOTs that provably cannot matter, removing gate count from the explanation.

A second reading is more informative. `star` and `linear` have the same gate count *and* the same cone $\{0,1,2,3\}$, yet differ sharply in d_{eff} (21.8 vs. 9.3) and gap (0.372 vs. 0.146). Taken together the two comparisons rule out both cruder summaries of the circuit: gate count does not determine complexity (`disjoint` vs. the other two), and cone membership does not determine it either (`linear` vs. `star`)—what remains is the Fisher geometry the connectivity induces, which is what d_{eff} measures and which orders all four patterns correctly (3.3, 3.3, 9.3, 21.8 against gaps 0.101, 0.101, 0.146, 0.372). This is evidence for the choice of complexity measure, not only for the mechanism.

Table 5: Matched two-qubit-gate-count control (n=4, L=2, d=16, 8 seeds, mean ± sem). All three entangling patterns use *exactly three* CNOTs but differ in readout-mediated connectivity: `disjoint` has cone $\{0\}$, whereas `linear` and `star` both reach all wires but induce different Fisher geometries. Two comparisons follow. `disjoint` vs. `linear`/`star`: gate count does not determine complexity. `linear` vs. `star`: cone membership does not determine complexity either. d_{eff} orders all four rows.

pattern	CNOTs	cone	N	gap	test acc	d_{eff}	live
`none`	0	$\{0\}$	16	0.101 ± 0.040	0.587	3.3	1.9
`disjoint`	3	$\{0\}$	16	0.101 ± 0.040	0.587	3.3	1.9
`linear`	3	$\{0,1,2,3\}$	16	0.146 ± 0.039	0.604	9.3	3.5
`star`	3	$\{0,1,2,3\}$	16	$\mathbf{0.372 \pm 0.014}$	0.542	21.8	5.2
`none`	0	$\{0\}$	64	-0.008 ± 0.022	0.656	2.6	2.0
`disjoint`	3	$\{0\}$	64	-0.008 ± 0.022	0.656	2.6	2.0
`linear`	3	$\{0,1,2,3\}$	64	0.064 ± 0.025	0.657	8.0	4.0
`star`	3	$\{0,1,2,3\}$	64	$\mathbf{0.104 \pm 0.016}$	0.608	12.7	5.0

The accuracy–complexity trade-off. Figure 3 makes the "budget, don't banish" principle of Section 5 explicit: as d_{eff} grows with entanglement, *train* accuracy rises but *test* accuracy does not, so the gap—the vertical distance between the curves—widens.

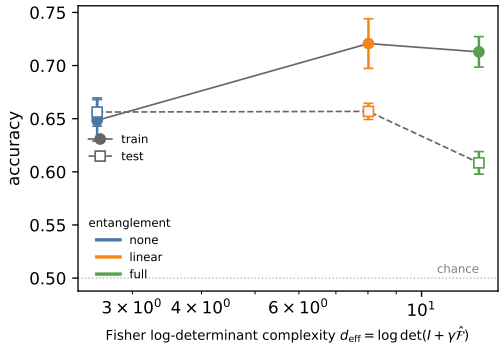


Figure 3: The accuracy–complexity trade-off (N=64, single-observable readout). As d_{eff} grows with connectivity (colour), *train* accuracy rises but *test* accuracy does not, so the gap widens: entanglement buys training fit that the test distribution does not reward. Points are means over 8 seeds; error bars are one standard error.

Scaling to 16 qubits. Table 6 reports the fixed-parameter-count comparison at $n \in \{6, 8, 10, 12, 16\}$ qubits with the interaction target extended to $n/2$ pairwise products. The gap is monotone none $<$ linear $<$ full at every size, and the effect strengthens with n.

Table 6: The entanglement–gap effect from n=6 to n=16 qubits (single-observable readout, fixed d=$2nL$, L=2; 8 seeds at n=6, 8, 4 at n=10, 12, 16). The gap is monotone none $<$ linear $<$ full at every size and `full` beats `none` by many standard errors, with the effect strengthening with n. (Held-out accuracy is above chance at n=6, 8 and approaches chance at $n \geq 10$, where the target is harder—see text.)

n (d)	N	gap `none`	gap `linear`	gap `full`	`full`−`none`
6 (24)	64	0.042	0.098	0.159	+0.117 (6.9σ)
6 (24)	128	0.035	0.061	0.104	+0.070 (4.5σ)
8 (32)	64	0.053	0.084	0.183	+0.130 (6.2σ)
8 (32)	128	0.033	0.071	0.101	+0.068 (4.6σ)
10† (40)	64	0.062	0.124	0.274	+0.212 (5.6σ)
12† (48)	64	0.026	0.095	0.295	+0.269 (6.4σ)
16† (64)	64	0.075	0.121	0.340	+0.265 (6.3σ)

† Held-out accuracy near chance at these sizes (the $n/2$-product target is hard for shallow circuits); these rows test the gap *ordering* only, not high-accuracy *learning*. Rows above the rule (n=6, 8) learn well above chance.

Table 7: Contextual-bandit (one-step RL) generalization: reward gap at fixed parameter count (d=16), mean ± sem over 16 seeds. Held-out reward is 0.60–0.66 vs. 0.5 chance, so this is genuine generalization; the entangled policies generalize worse, monotonically at the smaller N, and each beats `none` by ≥ 3 standard errors.

N_{train}	`none`	`linear`	`full`	`full`−`none`
16	0.136 ± 0.019	0.208 ± 0.018	$\mathbf{0.262 \pm 0.026}$	+0.125 (3.9σ)
40	0.092 ± 0.018	0.134 ± 0.009	$\mathbf{0.173 \pm 0.018}$	+0.081 (3.2σ)
100	0.026 ± 0.008	0.083 ± 0.011	0.069 ± 0.008	+0.043 (3.7σ)

Bandit and value-function figures. Figure 4 gives the reward-space view of the contextual-bandit result of Table 7: the entangled policies earn a higher *training* reward but not a higher *test* reward. its right panel shows the multi-step value-function gap, monotone none $<$ linear $<$ full at every training-set size.

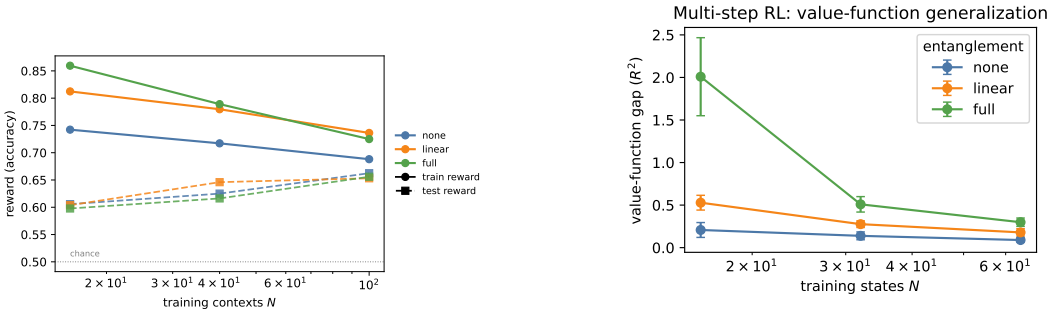


Figure 4: **Left:** contextual-bandit (one-step RL) *train* (solid) and *test* (dashed) reward versus the number of training contexts N, at fixed parameter count. The entangled policies earn a higher training reward but not a higher test reward—the vertical train–test gap, which Table 7 quantifies, is the entanglement cost. Averages over 16 seeds. **Right:** multi-step RL—the value-function generalization gap (R^2, Monte-Carlo policy evaluation, horizon 6) is monotone none $<$ linear $<$ full at every training-set size, at fixed parameter count. 8 seeds; error bars ± 1 s.e.

Beyond a single target. Table 8 repeats the fixed-parameter-count comparison on three further targets. On the Fourier target the gap is again ordered none $<$ linear $<$ full. The entangled-teacher target is hard for *all* student circuits at this sample size (test accuracy near chance for every connectivity); we include it only as a stress test of gap *behaviour*, not as evidence of successful learning. Parity is the instructive exception: it is not representable without entanglement, so `none` and `linear` sit at chance while `full` is the only circuit that learns it.

Table 8: The effect across target distributions (n=4, N=64, fixed d=16). On targets all circuits can fit (Fourier, teacher-PQC) entanglement widens the gap; on parity, which requires entanglement, only `full` learns the task (test 0.90 vs. chance).

target	`none` gap (test)	`linear` gap (test)	`full` gap (test)
random Fourier	0.038 (0.71)	0.045 (0.69)	0.103 (0.67)
entangled teacher	0.074 (0.50)	0.123 (0.50)	0.148 (0.51)
4-bit parity	0.067 (0.50)	0.116 (0.50)	0.050 (**0.90**)

Real (non-synthetic) benchmarks. Table 9 gives the per-dataset gaps and d_{eff} values behind the test-accuracy result quoted in Section 7.

Table 9: Real datasets (binary, 4 features, N=40, fixed d=16, 6 seeds). The gap is ordered none $<$ linear $<$ full and co-monotone with d_{eff} on all three; held-out accuracy itself is *not* monotone across datasets (`full` is highest on Iris, essentially tied on Breast Cancer, lowest on Wine), so we do not read this table as evidence that entanglement lowers accuracy—only the pooled Wine/Digits comparison in the main text (Section 7, 22 seeds) resolves that claim. Digits (3 vs. 8) is not listed here: it enters only the higher-powered pooled test-accuracy comparison of Section 7 ($N \in \{16, 24, 32\}$, 22 seeds), which is a different protocol from this single-N, 6-seed table.

dataset	`none` gap (d_{eff}, test)	`linear` gap (d_{eff}, test)	`full` gap (d_{eff}, test)
Iris	0.00 (3.0, 0.70)	0.061 (5.2, 0.65)	0.082 (18.2, 0.76)
Breast Cancer	0.025 (4.5, 0.90)	0.033 (10.3, 0.91)	0.069 (13.8, 0.90)
Wine	0.029 (5.1, 0.90)	0.059 (9.1, 0.88)	0.125 (13.2, 0.85)

Barren plateaus bound the mechanism. Our mechanism runs on the Fisher spectrum, which barren plateaus (McClean et al., 2018; Cerezo et al., 2021) flatten exponentially in qubit number. We confirm the erosion directly (Table 10): at a fixed moderate depth (L=4, all-to-all, random parameters) the log-determinant complexity *falls* monotonically with n even as the nominal parameter count rises—the Fisher eigenvalues (the $\langle Z_0 \rangle$ gradient variance) decay with n faster than the rank grows, so d_{eff} erodes toward its floor. This makes the boundary flagged in our failure-mode analysis (Appendix A) quantitative: entanglement inflates d_{eff} only while the circuit is off the plateau.

Table 10: Barren plateaus erode the mechanism's substrate. At fixed depth (L=4, all-to-all, random parameters, 6–12 seeds), the Fisher log-determinant complexity d_{eff} *decreases* with qubit number even as the nominal parameter count d grows. This is the regime our trainable-circuit experiments deliberately avoid.

n	4	6	8	10	12
nominal params d	32	48	64	80	96
d_{eff} (L=4, random)	13.2	8.4	4.5	3.2	3.1

Depth ablation: d_{eff}, not parameter count, is the stable predictor. The complementary ablation—*fixing* the entangling pattern and *varying* depth, so the parameter count changes—shows the gap rising with depth for both patterns (`linear`: L=1→4 gap 0.03→0.33; `full`: 0.03→0.31), and in lock-step with d_{eff} (which rises 2→50 and 2→61 respectively). At matched depth (matched parameter count) the more entangled pattern still has the larger d_{eff} and gap. Thus d_{eff} predicts the gap whether we vary connectivity at fixed depth or depth at fixed connectivity, whereas raw parameter count only tracks the gap when it happens to co-vary with d_{eff}.

Controls and the hardware run. Figure 5 shows the three optimization controls of Section 7 (left) and the hardware result (right).

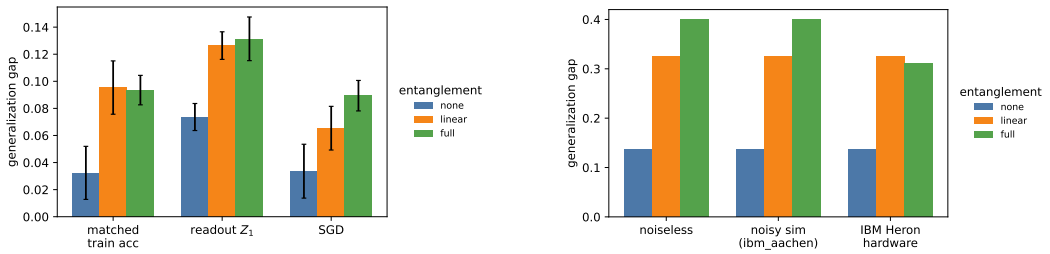


Figure 5: **Left:** the ordering (entangled circuits above `none`) survives three controls at N=64—matched training accuracy, a different readout observable, and a different optimizer—so the gap is not a mere optimization or trainability artifact. **Right:** the ordering is reproduced by a noise-model simulator of `ibm_aachen` and by execution on the `ibm_aachen` Heron processor itself under real noise; calibration and shot-noise details are in Appendix C.

Fisher log-determinant complexity and gap, per configuration. Table 11 gives $d_{\text{eff}} = \log\det(I + \gamma\hat{\mathcal{F}})$ (Eq. (4), fixed $\gamma = 50$) and the gap for every connectivity at depths L=2, 3 and N_{train}=16. At fixed depth (fixed d), d_{eff} increases monotonically with entangling connectivity in both blocks. The gap follows it monotonically at L=2 ($0.101 < 0.146 < 0.381$) but *not* at L=3, where `linear` and `full` are effectively tied and slightly inverted (0.440 vs. 0.431) while d_{eff} still separates them (29.5 vs. 36.5). We report this rather than smooth it over: our claim is that entangled circuits generalize worse than non-entangled ones of equal size and that d_{eff} is the best available predictor of the gap (ρ=0.82 over 300 configurations), not that the three-way ordering is exact in every cell. A single entangling layer (L=1) leaves the gap unchanged—the depth-gating referred to in the main text. The mechanism is visible directly in the Fisher eigenvalue spectrum (Figure 6).

Table 11: Fisher log-determinant complexity $d_{\text{eff}} = \log\det(I + \gamma\hat{\mathcal{F}})$ and generalization gap. At each fixed (L, N_{train}) block the parameter count $d = 8L$ is identical across the three rows; only the entangling connectivity—and hence d_{eff} and the gap—changes. d_{eff} rises monotonically none → linear → full in both blocks; the gap does so at L=2 but ties at L=3 (0.440 vs. 0.431), which we report as-is.

depth	connectivity	N_{train}	Fisher d_{eff} (log-det)	gap
L=2	`none`	16	3.3	0.101
L=2	`linear`	16	9.3	0.146
L=2	`full`	16	21.8	0.381
L=3	`none`	16	4.0	0.101
L=3	`linear`	16	29.5	0.440
L=3	`full`	16	36.5	0.431

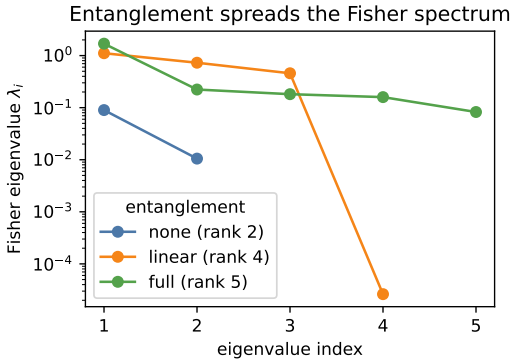


Figure 6: Fisher eigenvalue spectrum of the trained circuits at L=2 (single-observable readout, log scale). The non-entangled circuit has effectively two non-zero eigenvalues, `linear` four and `full` five (the ranks shown in the legend): connectivity raises the rank of the Fisher matrix and hence $d_{\text{eff}} = \log\det(I + \gamma\hat{\mathcal{F}})$, exactly the quantity charged by the bound. This is the spectral content of Proposition 1.

C Additional experimental detail

Setup. The primary setup, used unless a section states otherwise, is n=4 qubits simulated with PennyLane's `default.qubit` and the *single* local readout $\langle Z_0\rangle$. The exceptions are stated where they occur: the scaling study runs $n \in \{6, \ldots, 16\}$ (with `lightning.qubit` and the adjoint method at $n \geq 10$), and the readout controls of Section 7 replace $\langle Z_0\rangle$ by Z_1 or by the global $\frac{1}{n}\sum_i \langle Z_i\rangle$. Inputs are angle-encoded with R_Y; the ansatz is L layers of R_Y, R_Z per qubit (d=$2nL$ trainable angles) followed by a fixed CNOT pattern (`none`/`linear`/`full`). Figure 7 shows the three ansätze: they share the identical rotation layout (hence d) and differ only in the entangling block.

Supervised models (Sections 4–5) are binary classifiers $\hat{y} = \text{sign}(\langle Z_0\rangle)$ trained by Adam (600 steps, full batch, lr 0.08) with the cross-entropy loss on the interaction target $y = \text{sign}(x_1x_2 - x_3x_4 + \frac{1}{2}x_1)$, $x \sim U[-1, 1]^4$, and evaluated on 2000 held-out inputs. *RL policies* (Section 6) are contextual-bandit policies $\pi(a{=}1 \mid x) = \sigma(\alpha\langle Z_0\rangle)$ over contexts $x \sim U[-1, 1]^4$, trained by REINFORCE (600 updates, lr 0.08) on bandit feedback (reward 1 iff the sampled action equals $a^\star(x) = \mathbb{1}[x_1x_2 - x_3x_4 + \frac{1}{2}x_1 > 0]$, the correct label is never revealed); the gap is training reward minus reward on 2000 held-out contexts. Each cell averages 8 (supervised) or 16 (bandit) seeds. The Meyer–Wallach Q and the Fisher matrix are evaluated at the trained parameters over held-out inputs.

Fisher log-determinant complexity: exact computation. To remove ambiguity we specify the estimator precisely. (i) We use the *empirical* (Gauss–Newton) Fisher of the policy's Bernoulli

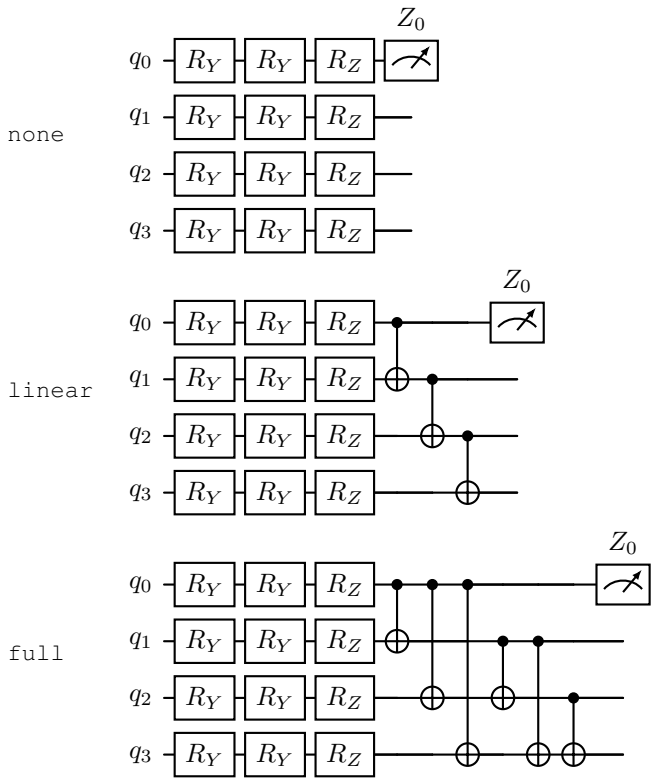


Figure 7: The three ansätze at n=4 (one layer of the repeated block shown; the block repeats L times, here between the encoding $R_Y(x_i)$ and the $\langle Z_0\rangle$ readout). All three use the *same* R_Y, R_Z rotations—hence the same parameter count d=$2nL$—and differ only in the entangling block: none (no CNOTs), a linear CNOT chain, or all-to-all CNOTs. This is the "fixed parameter count, vary only entanglement" comparison.

readout head—for the logistic head this *is* the loss Fisher of Section 3, written in the output-gradient form; it is the empirical, not the exact/population, matrix, and it is not a parameter-space metric of the state: $\hat{\mathcal{F}}(\theta) = \frac{1}{M}\sum_x c(x)\, g_x g_x^\top$ with $g_x = \nabla_\theta \langle Z_0\rangle_{x,\theta}$ and Bernoulli weight $c(x) = \alpha^2\, p(x)(1-p(x))$, $p = \sigma(\alpha\langle Z_0\rangle)$. (ii) It is evaluated at the *trained* parameters θ over M=40 *held-out* inputs drawn from the input distribution (not the training set); being an average of M rank-one terms, its rank is at most $\min\{M, |\mathcal{L}(\theta)|\}$ (Proposition 1), which is inactive at n=4 but binds at the largest n in Table 6. (iii) The per-input gradients g_x are computed by autograd (exact parameter-shift). (iv) Eigenvalues λ_i of $\hat{\mathcal{F}}$ are used *unnormalized* (no trace rescaling), which is why d_{eff} is a log-volume and can exceed d. (v) The complexity is $d_{\text{eff}} = \log\det(I + \gamma\hat{\mathcal{F}}) = \sum_i \log(1+\gamma\lambda_i)$ with the constant $\gamma = 50$ used identically everywhere (Remark 1). (vi) For a reported cell we compute d_{eff} per seed and then *average* d_{eff} across seeds (we do not pool the Fisher matrices). The participation ratio $(\sum_i \lambda_i)^2 / \sum_i \lambda_i^2$, when reported, is computed from the same spectra. *Why γ=50, and does it matter.* The value 50 places the observed spectra (eigenvalues $\lambda \in [10^{-2}, 2]$) in the informative band where $\gamma\lambda$ spans $\approx$ 0.5–100, so d_{eff} is neither saturated nor collapsed to zero. Crucially the conclusions we draw from d_{eff} are insensitive to this choice, since $d_{\text{eff}}(\gamma) = \sum_i \log(1+\gamma\lambda_i)$ is monotone in the Fisher spectrum. We recomputed d_{eff} for all 18 configurations of the depth×connectivity sweep at $\gamma \in \{10, 25, 50, 100, 200\}$: the Spearman correlation $\rho(d_{\text{eff}}, \text{gap})$ is $0.74, 0.70, 0.67, 0.67, 0.64$ respectively—stable in sign and magnitude—and the ordering none $<$ linear $<$ full holds at every γ (e.g. at L=2, d_{eff} goes 2.5→4.8→11.5 at γ=10 and 5.8→12.2→25.6 at γ=200). We fix a single $\gamma = 50$ only so that d_{eff} is numerically comparable across rows.

d_{eff} dominates the learned norm—a powered, bootstrapped test. A comparison between two correlations of 0.76 and 0.70 over 18 points would be underpowered, so we built a much larger and

more varied grid: 5 entangling patterns (`none`/`linear`/`ring`/`alt`/`full`) $\times$ depths $L \in \{1, \ldots, 5\}$ $\times$ $N \in \{16, 32, 64\}$ $\times$ 4 seeds $= 300$ configurations, which dissociates d_{eff} from the learned norm and from parameter count. Across these 300 points the Spearman correlations with the gap are $\rho(d_{\text{eff}}) = 0.82$, $\rho(\|\theta - \theta_0\|^2) = 0.73$, $\rho(Q) = 0.72$, $\rho(\text{params}) = 0.45$. A 5000-resample bootstrap over configurations gives $\Delta\rho = \rho(d_{\text{eff}}) - \rho(\text{norm}) = 0.090$ with 95% CI $[0.041, 0.144]$ (bootstrap $P(\Delta\rho{>}0){=}1.00$), and $\rho(d_{\text{eff}}) - \rho(Q) = 0.100$, CI $[0.046, 0.156]$: the Fisher log-determinant complexity is a *significantly* better predictor of the gap than the learned norm or the entangling power. A multiple regression on the same grid loads almost entirely on d_{eff} (standardized coefficient $+0.16$ vs. ≤ 0.06 in magnitude for the others), and adding the entangling pattern on top of d_{eff} raises R^2 by < 0.01 while adding d_{eff} on top of the pattern raises it by 0.14. Entanglement therefore acts on the gap *through* d_{eff}, which screens it off—the precise, and now powered, sense in which d_{eff} is the governing quantity.

Multi-step value-function generalization and controls. The value-function experiment (Section 6) fits a single-observable VQC to $V^\star(x) = \sum_{t=0}^{5} 0.9^t r(x_t)$ under the deterministic transition $x' = \tanh(1.3\,\text{roll}(x) + 0.2x)$ with r the interaction rule, by Adam regression (800 steps) on N states, reporting held-out R^2 over 1000 states. The controls of Section 7 use $N{=}64$: matched-accuracy early-stops at train accuracy 0.72; the readout control measures Z_1; the optimizer control uses SGD (lr 0.5). The hardware run trains the $N{=}16$ classifiers in simulation and evaluates $\langle Z_0 \rangle$ for all 56 inputs on `ibm_aachen` via `EstimatorV2` at 4096 shots, optimization level 1.

Hardware credibility: calibration, noise, and shot statistics. The device is IBM Heron `ibm_aachen` (156 qubits). At the time of execution its median two-qubit gate error was 1.8×10^{-3} (typical range 7×10^{-4} to a few $\times 10^{-3}$, excluding one inoperable pair) and median $T_1 \approx 218\,\mu$s. Readout error mitigation was *not* applied—the effect we report survives raw device noise. Three cross-checks support the hardware numbers (Figure 5, right). (i) A *noise-model simulator* built from the `ibm_aachen` calibration (`AerSimulator` with `NoiseModel.from_backend`) reproduces the ordering (gap $0.14/0.33/0.40$ for none/linear/full), i.e. the modelled noise does not remove the effect. (ii) *Shot noise* is small: at 4096 shots each $\langle Z_0 \rangle$ estimate has standard error $\approx 1/\sqrt{4096} \approx 0.016$, which flips a sign only for inputs within that margin of the decision boundary, so the accuracy-based gap is stable. (iii) The raw hardware gaps ($0.14/0.33/0.31$) agree with both the noiseless and noise-model values to within ≈ 0.1, and preserve the none $<$ {linear, full} ordering. We report the hardware result as corroboration of the simulation evidence, not as a separate quantitative claim.

Compute and reproducibility. Classical simulations use PennyLane with the PyTorch interface and analytic gradients: `default.qubit` with backpropagation in the primary $n \leq 8$ setting, and `lightning.qubit` with the adjoint method for the $n \geq 10$ scaling runs. Each supervised cell averages 8 seeds, each bandit/robustness/real-data cell 16/8/6 seeds, and the value-function and end-to-end policy-gradient cells 8/20 seeds; the pooled real-data accuracy comparison of Section 7 uses 22 seeds per cell. Error bars are one standard error. Targets: interaction rule $\text{sign}(x_1x_2 - x_3x_4 + \frac{1}{2}x_1)$ (main); 4-bit parity $\text{sign}(\prod_i \text{sign}(x_i))$; random Fourier $\text{sign}(\sum_{k=1}^{6} \cos(w_k \cdot x + b_k))$ with fixed w_k, b_k; entangled-teacher labels $\text{sign}(\langle Z_0 \rangle)$ of a fixed 2-layer `StronglyEntangling` teacher. Real datasets are the scikit-learn Iris (versicolor vs. virginica), Breast Cancer Wisconsin, Wine (classes 0/1) and Digits (3 vs. 8) sets, standardized, reduced to four features by PCA when needed, and min–max scaled to $[-1, 1]$. Digits is used only for the pooled test-accuracy comparison, not for Table 9.

AI USE STATEMENT

Parts of this paper's text, code and analysis were produced with the assistance of generative AI. The authors take full responsibility for the final content of this paper, including any text produced with the assistance of generative AI.