A SYMMETRY-BREAKING DIAL FOR EQUIVARIANT QUANTUM MACHINE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Symmetry-preserving circuits encode conservation laws into quantum machine learning models, but a fully equivariant circuit cannot distinguish states related by that symmetry. This creates a structural blind spot on mirror-pair tasks, where the label identifies one of two symmetry-related ground-state branches. We introduce a continuous *symmetry-breaking dial* that rescales symmetry-breaking gate angles while keeping the parameter count fixed, interpolating between equivariant and unconstrained ansatze. We prove that the equivariant endpoint has exactly 50% accuracy on every mirror pair and that the class separation is odd in the dial, with a linear coefficient determined in closed form by the initialization and the data. With a preconditioned optimizer, the smallest dial value reaching 99.9% accuracy follows $\lambda^\star(T) \approx 0.3/T$ where the grid resolves it, and extends to a calibrated joint law in budget and task strength. On transverse-field mirror-pair tasks, this prescription keeps the dial below 10^{-2}, breaking only a fraction of a percent of the symmetry, while reaching full accuracy. The transport law persists across depth, register size, and a decade of task strength, with recalibrated constants. We also identify a zero-transport boundary where the circuit cannot propagate the mirror distinction to the readout. Beyond quantum models, whenever the label is defined by the symmetry, full equivariance is a provable ceiling rather than a safe default.

1 INTRODUCTION

Symmetry is the most reliable inductive bias in machine learning. When the problem obeys a conservation law, building that law into the model removes parameters it would otherwise have to learn, and it improves both optimization and generalization (Larocca et al., 2022b; Gard et al., 2020; Wecker et al., 2015; Bravyi et al., 2017). Quantum models follow the same recipe, and a circuit whose gates commute with the symmetry respects it by construction, so such designs are now standard across variational quantum eigensolvers and quantum machine learning (QML) classifiers (Cerezo et al., 2021; Kandala et al., 2017; Havlíček et al., 2019; Cerezo et al., 2022). Realized as equivariant gate-sets (Meyer et al., 2023; Nguyen et al., 2024), they come with trainability and generalization guarantees in structured settings (Schatzki et al., 2022; Skolik et al., 2023; Zheng et al., 2023; Sauvage et al., 2024), the quantum counterpart of geometric deep learning (Bronstein et al., 2021; Cohen & Welling, 2016), and they rest on one assumption that is rarely stated, that respecting the symmetry is free.

That assumption fails whenever the label is defined by the symmetry itself. A circuit and readout that both respect the symmetry cannot distinguish inputs related by the symmetry, whatever their capacity, depth or training time (Thanasilp et al., 2024). Symmetry-broken phases are everywhere in physics, and the canonical example is telling which ferromagnetic branch of a transverse-field chain a ground state belongs to, the two branches being mirror images of each other under the symmetry. For such *mirror-pair* tasks the equivariant model is exactly blind.

Between fully symmetric and blind on one side, and fully unconstrained and wasteful on the other, lies an entire spectrum, and the practitioner's question is quantitative. *How much symmetry must be broken before the model can see, and what does breaking cost?* Existing work establishes either the benefits of equivariance (Larocca et al., 2022b; Nguyen et al., 2024) or the expressivity cost of hard constraints (Schuld et al., 2021; Sim et al., 2019), and noise breaks equivariance in measurable ways (Tüysüz et al., 2024), but the intermediate regime has not been mapped. In current practice

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.

the amount of breaking is chosen by hand, by trying the equivariant ansatz, observing failure on mirror-defined distinctions, and then relaxing constraints by some untutored amount.

We map that spectrum with a single continuous dial. We split the circuit into a symmetry-respecting block and a symmetry-breaking block, and the dial scales only the angles of the second, leaving the parameter count unchanged. Because both ends of the dial are ansatze practitioners already use, every setting in between is a genuine interpolation between them. A two-qubit instance shows why the dial controls what the model can see. Let the mirror flip one Pauli operator and the readout be one the mirror leaves unchanged. A single breaking rotation then splits the readout into a term the mirror preserves and a term it flips, so the two classes separate by twice the second term, starting at zero and growing with the dial setting at a rate fixed by one expectation value on the data. Our contribution is a set of design laws for this dial, four results that together replace hand-tuning with a prescription.

- We prove that when both the circuit and the readout respect the symmetry, the two classes of a mirror-pair task receive identical outputs for every parameter setting, so the accuracy is pinned at exactly 50%, independent of capacity, depth, optimizer and budget. The blind spot is architectural rather than a failure of training, so more compute cannot remove it.
- We derive the rate at which turning the dial on separates the two classes. The separation reverses sign when the dial reverses, so it passes through zero and grows linearly at a rate fixed in closed form by the initialization and the data.
- We establish that under a preconditioned optimizer the smallest sufficient breaking falls in inverse proportion to the training budget in the resolved range, that the trade extends to a two-variable form in budget and task strength, and that unpreconditioned gradient descent stalls near full breaking. Both keep their functional form while their constants rescale, and the contrast singles out preconditioning as what makes a small dial usable.
- We show that the laws travel across a doubled depth, register sizes up to fourteen qubits and a decade of task strength, with only their constants changing, so retargeting costs a recalibration rather than a fresh sweep. The sample- and trainability-cost exemptions and the zero-transport boundary condense into one design-law summary. All empirical constants are established for simulated transverse-field mirror-pair tasks in the noiseless full-batch setting and need recalibration outside this family.

2 THE SYMMETRY-BREAKING DIAL

We introduce the setting through Figure 1. A mirror-pair task presents the model with two ground states that differ only by the symmetry, and the dial decides how much of that symmetry the circuit may break.

2.1 MIRROR-PAIR TASKS

Let $G = \langle g \rangle$ be the order-two group generated by g acting on n qubits, with X_i, Y_i, Z_i the Pauli operators on site i. We use the global parity flip $g = \prod_i Z_i$, which negates X and Y while fixing Z; this is the $\mathbb{Z}_2$ of the transverse-field Ising chain (Pfeuty, 1970), written with the coupling on X and the transverse field on Z, so that g exchanges the two ferromagnetic branches. A *mirror-pair task* consists of states

$$|+_\varepsilon\rangle = \mathrm{GS}\big[H(h, +\varepsilon)\big], \qquad |-_\varepsilon\rangle = g\,|+_\varepsilon\rangle, \qquad y = \pm 1, \tag{1}$$

where $\mathrm{GS}[\cdot]$ denotes the ground state of the bracketed Hamiltonian and $H(h,\varepsilon) = -\sum_i X_i X_{i+1} - h\sum_i Z_i - \varepsilon \sum_i X_i$ satisfies $g\,H(h,\varepsilon)\,g = H(h,-\varepsilon)$ exactly, since the longitudinal field ε is the only term that breaks the symmetry and flipping its sign is equivalent to conjugating by g. The two classes of the task are therefore related by the symmetry for every value of the transverse field h, and the magnitude ε sets the intrinsic difficulty of the distinction, the two classes being the states generated at $+\varepsilon$ and $-\varepsilon$, so a smaller ε hides the difference more deeply. The model reads out through a G-invariant observable,

$$R(\rho) \;=\; \text{mean over bonds of } \langle Z_a Z_{a+1}\rangle, \tag{2}$$

which is invariant under the mirror, so it is blind to which branch the state occupies, and is the kind of measurement a symmetry-respecting device performs naturally. Here ρ is the prepared state, and

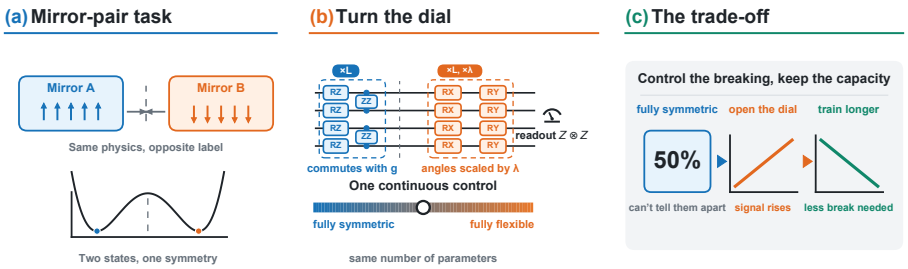

Figure 1: **The symmetry-breaking dial.** **(a)** Mirror-pair task, two ground states that are exact symmetry images and carry opposite labels, differing only in the magnetization component that the mirror negates. **(b)** The split circuit with gate families abbreviated. Blue gates (RZ and paired two-qubit generators) commute with the mirror and form the protected core, orange gates (RX, RY) anticommute with it and their angles are scaled by the dial λ; the readout measures the bond observable $Z \otimes Z$. Both sides repeat L times, symmetric layers first and the breaking block last. **(c)** The fully symmetric model cannot tell the pair apart and stays at 50% accuracy, opening the dial grows the class signal, and longer training reduces the breaking that is needed.

$R(|\phi\rangle)$ is the same functional on a pure state. Training minimizes a full-batch softplus loss on the signed margin $y\,R$, and test accuracy is threshold classification on held-out values of h, so train and test share one distribution. Because the label is generated by the symmetry rather than by a per-sample readout feature, the two classes are indistinguishable to any statistic of the readout alone at the equivariant end, where a two-sample comparison between them has no power (Gretton et al., 2012). Restoring that difference is what the dial has to do.

2.2 Equivariant block, breaking block, and the dial

We split the circuit by symmetry,

$$U(\boldsymbol{\theta};\lambda) \;=\; U_{\mathrm{brk}}(\lambda\boldsymbol{\theta}_{\mathrm{brk}})\, U_{\mathrm{sym}}(\boldsymbol{\theta}_{\mathrm{sym}}), \tag{3}$$

where $\boldsymbol{\theta}_{\mathrm{sym}}$ and $\boldsymbol{\theta}_{\mathrm{brk}}$ collect the angles of the two blocks, the generators of U_{sym} commute with g (here RZ_i on all sites and XX, YY, ZZ on alternating nearest-neighbour pairs) and those of U_{brk} anticommute with g (here RX_i, RY_i on all sites), with the symmetric block acting first; Figure 1b shows the layout. The split is read off the mirror directly, a single-qubit X- or Y-type gate flips sign once under g and so anticommutes, while a two-qubit Pauli-pair gate flips sign twice and so commutes, which is why the protected core is built from RZ and paired two-qubit gates and the breaking block from single-qubit X- and Y-rotations. The dial $\lambda \in [0,1]$ rescales only the breaking angles. Two properties make this a clean experimental instrument. At $\lambda = 0$ the breaking block is the identity, so the two endpoints are exactly the equivariant ansatz and the unconstrained hardware-efficient one, the standard baselines of the literature, and every setting between them is a genuine interpolation. The parameter count is identical for all λ, so a comparison across dial values compares symmetry budgets alone.

2.3 Training protocol and the operating point

We train each model with full-batch descent, and the budget T counts steps. The default optimizer is the preconditioned adaptive method Adam (Kingma & Ba, 2015), with plain gradient descent (GD) as the unpreconditioned contrast, a distinction the exchange law turns on. Because the two classes differ only by the symmetry, what matters is the budget at which the model crosses a fixed accuracy threshold, so the headline object is

$$\lambda^{\star}(T) \;=\; \min\{\lambda :\ \text{test accuracy} \geq 99.9\% \text{ within } T \text{ Adam steps}\}, \tag{4}$$

the smallest symmetry breaking that suffices at budget T. In the sweeps this minimum is estimated on the dial grid, so the reported $\lambda^{\star}$ is a grid-resolved sufficient upper bound unless stated otherwise. Defining it by a crossing rather than by a loss value keeps it comparable across optimizers and models, and it is the quantity the rest of the paper measures.

3 ANALYSIS OF THE DIAL

We prove three statements that characterize the dial, one for each end and one for the budget in between. Throughout, $\psi_- = g\psi_+$ for a reference state ψ_+, the readout R is G-invariant, and $U(\boldsymbol{\theta};\lambda)$ has the form of equation 3. We write $\Delta(\boldsymbol{\theta};\lambda) = R(U(\boldsymbol{\theta};\lambda)|\psi_+\rangle) - R(U(\boldsymbol{\theta};\lambda)|\psi_-\rangle)$ for the signed mirror margin at one initialization, the quantity that has to become nonzero before the task can be learned at all. We also write q for the mirror signal, the mean absolute readout margin between the two classes, $q = \mathbb{E}\,|\Delta|$; the group generator keeps the symbol g throughout. The three proofs are self-contained constructions, and we give them in full in Appendix A.

Theorem 1 (Structural ceiling)**.** *Whenever $U(\boldsymbol{\theta};\lambda)$ commutes with g, every mirror pair satisfies $R\big(U|\psi_+\rangle\big) = R\big(U|\psi_-\rangle\big)$. At the equivariant end, where the breaking block is the identity, this holds for every parameter setting, so on a test set organized as balanced mirror pairs, the accuracy of* any *threshold classifier with a fixed tie rule on the model is exactly 50%, independent of capacity, depth, optimizer, budget and sample size.*

The ceiling follows from the construction in one line. The two classes are symmetry images of each other, so they reach the readout with equal expectation, and more depth or more training moves both equally. A symmetry-constrained ansatz therefore fails on any mirror-defined distinction, which is why the dial has to be opened. Whether the budget is usable depends on how the descent treats a small gradient.

Theorem 2 (Odd transport)**.** *The mirror margin is exactly odd in the dial, $\Delta(\boldsymbol{\theta};-\lambda) = -\Delta(\boldsymbol{\theta};\lambda)$ for every initialization, hence $\Delta(\boldsymbol{\theta};\lambda) = s(\boldsymbol{\theta})\,\lambda + O(\lambda^3)$, where $s(\boldsymbol{\theta})$ is the transport slope at that initialization. The intercept is zero by symmetry and the transport is linear to leading order.*

Opening the dial therefore produces a signal with a definite sign and no threshold to be met, inherited from the initialization rather than tuned. The next statement gives $s(\boldsymbol{\theta})$ in closed form, so the transport slope, and through it the direction and magnitude of the initial signal, is computable before any training; the remaining ingredient, the budget trade that fixes the operating point $\lambda^\star$, is the subject of the last statement.

Corollary 1 (Susceptibility formula)**.** *Write the breaking gates as $U_{\mathrm{brk}} = \prod_k e^{-i\lambda\theta_k G_k/2}$ and let $\phi_+ = U_{\mathrm{sym}}|\psi_+\rangle$ be the state entering the breaking block. For every parameter setting,*

$$\Delta\big(\boldsymbol{\theta};\lambda\big) \;=\; -\,\lambda \sum_k \theta_k\, d_k(\boldsymbol{\theta}_{\mathrm{sym}}) \;+\; O(\lambda^3), \qquad d_k \;=\; \frac{d}{dt}\Big|_{t=0} R\Big(e^{itG_k}|\phi_+\rangle\Big), \tag{5}$$

where $R(\cdot)$ is the readout expectation in the bracketed state.

The slope $s(\boldsymbol{\theta})$ is a closed-form property of the initialization and the data, one circuit evaluation per breaking gate and no fit. It predicts sign and magnitude initialization by initialization, and numerically it reproduces the measured slope to three or four significant digits across all 24 (seed, ε) pairs, with a worst-case relative deviation below 10^{-3}. This is the certificate the practitioner gets for free, the slope of a candidate circuit being readable before training starts.

Proposition 1 (Exchange under preconditioned descent)**.** *Model the preconditioned optimizer as moving each coordinate by at most η per step, so that after T steps $\sum_k |\theta_k^{(T)}| \le \sum_k |\theta_k^{(0)}| + M\eta T$, where M is the number of breaking gates. By equation 5 the mirror signal obeys $q \le \lambda\, d_{\max} \sum_k |\theta_k|$ with $d_{\max} = \max_k |d_k|$. If full separability requires the signal to reach a fixed threshold $q^\star$, then $\lambda \;\ge\; q^\star \big/ \big(d_{\max}\big(\sum_k |\theta_k^{(0)}| + M\eta T\big)\big)$, a necessary condition decreasing as $1/T$.*

The bound fixes the shape of the trade-off and leaves its constant to the geometry of the optimizer. A preconditioned update takes its step from the learning rate rather than from the gradient magnitude, so the breaking parameters keep moving at rate η where the gradient is small and the signal accumulates with the budget. Unpreconditioned descent behaves in the opposite way, since a breaking parameter enters only through the angle $\lambda\theta$, so its gradient carries a factor λ and a plain step moves the signal by $O(\lambda^2)$, leaving the dial to clear the floor set by its own initial signal.

4 EXPERIMENTS

To make every comparison that follows well defined, we fix the baselines and the default configuration here. "Full breaking" $\lambda = 1$ is the unconstrained hardware-efficient baseline, "$\lambda = 0$" the fully

Table 1: **Design-law summary** (main setting $n = 10$, $L = 3$, $40/40$ split, Adam; $g_0 = q(\lambda)$ at initialization is the untrained mirror gap), verified in the subsections of Sec. 4.

	Law	Form	Regime
1	Structural ceiling	acc $= 50\%$ at $\lambda = 0$, exactly	all budgets, depths, sizes
2	Transport	$g_0 = 0.3546\,\lambda$, intercept 0, s closed form (Cor. 1)	$\lambda \le 3 \times 10^{-3}$; plateau 0.024 to 0.026
3	Exchange (Adam)	$\lambda^\star(T) \approx 0.3/T$	resolved for $T \in [50, 200]$, upper bounds to 800
4	Two-variable law	$\lambda^\star \approx 1.8 \times 10^{-3}/(T\,\varepsilon^{1.5})$	calibrated fit, within $1.3\times$ at $T = 400$
5	Optimizer dependence	plain GD needs $\lambda \approx 1$; Adam, no floor observed	$T \le 400$
6	Sample cost	none in training states	4 training states suffice ($\lambda \ge 3 \times 10^{-3}$)
7	Trainability cost	none (variance in dial coordinates)	full range
8	Boundary	$s \equiv 0$ for half-register mirrors	diagonal conjugated mirror

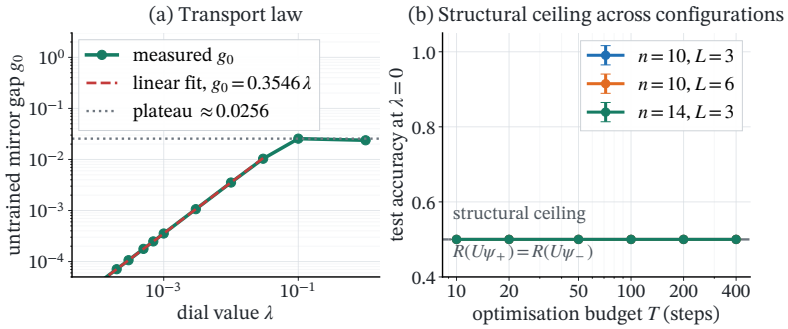

Figure 2: **Transport law and structural ceiling. (a)** Untrained mirror gap $g_0 = q(\lambda)$ at initialization against dial value λ (log–log). The dashed line is the forced-origin fit over $\lambda \in [10^{-4}, 3\times10^{-3}]$, extrapolated as a reference, and the dotted line marks the plateau. **(b)** Accuracy at $\lambda = 0$ against budget for depth $L = 3$, depth $L = 6$ and $n = 14$, all sitting at exactly 0.500 with zero variance across seeds, the ceiling of Thm. 1.

equivariant baseline, and plain gradient descent the unpreconditioned optimizer baseline, with the practitioner's default of grid-searching λ by hand completing the set of comparison points. Unless stated otherwise, we train the main configuration, $n = 10$, $L = 3$ (132 parameters, 72 symmetric and 60 breaking), on $40/40$ train/test states from interleaved h-grids at $\varepsilon = 0.03$, with five random initializations and shaded bands at ±1 standard deviation. All experiments run on a single workstation, and we list every configuration, seed and checkpoint in Appendix B so each sweep can be reproduced exactly.

We test the three statements of Sec. 3 on the mirror-pair task and measure what they leave open. Table 1 states the outcome up front as eight design laws, rows 1–3 the tested statements, rows 4–8 the regularities the sweeps add, and the subsections verify each row in turn. The symmetric end is exactly blind, transport away from it is linear with a slope computable at initialization, the budget trade is quantitative under a preconditioned optimizer, the kept symmetry costs nothing in samples or trainability, and the boundary class where transport vanishes is identified.

4.1 The ceiling is exact, and the transport is linear

We verify the two endpoint statements of Sec. 3 in two steps, the ceiling of Thm. 1 first, then the transport of Thm. 2 with the dial on. At $\lambda = 0$ the untrained mirror gap is $g_0 = 0$ at machine precision, because the negative class is the exact symmetry image of the positive one. The trained accuracy is 0.500 ± 0.000 across all seeds and every budget checkpoint of each sweep (Fig. 2b), and the ceiling does not move when the budget grows $40\times$, the depth doubles, or the register grows to $n = 14$. The equivariant model is blind.

Turning the dial on, the untrained gap follows $g_0 = s\lambda$ with $s = 0.3546$, where $g_0 = q(\lambda)$ is the mirror signal q at initialization. The ratio is constant to within 0.1% across $\lambda \in [10^{-4}, 3 \times 10^{-3}]$, then the branch bends over gradually and the gap settles on a plateau of 0.024 to 0.026 by $\lambda \approx 0.1$, so a larger dial buys no further untrained signal (Fig. 2a). The signed margin is exactly odd, $\Delta(-\lambda) =$

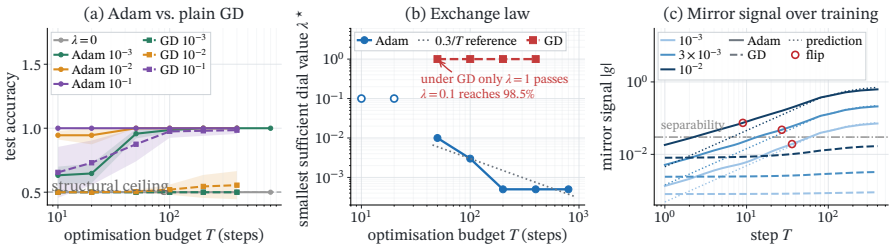

Figure 3: **The exchange law and the optimizer baseline.** **(a)** Test accuracy vs. budget for Adam (solid) and plain GD (dashed), mean ± 1 std over five seeds. **(b)** Smallest sufficient dial value $\lambda^\star(T)$, a grid-resolved upper bound at each budget. Adam tracks $\lambda^\star \approx 0.3/T$ where the grid resolves the crossing (open markers at the grid ceiling), while plain GD stalls, only $\lambda = 1$ reaching the threshold. **(c)** Mirror signal q over training; the three styles (solid Adam, dashed plain GD, dotted prediction derived below) each appear at $\lambda = 10^{-3}, 3 \times 10^{-3}, 10^{-2}$ (light to dark). The Adam signal climbs by more than an order of magnitude, and the circles mark the step at which the accuracy first reaches 99.9%.

$-\Delta(\lambda)$, verified at machine precision for $n = 8, 10, 12, 14$ and $\lambda \in \{\pm 10^{-3}, \pm 10^{-2}, \pm 3 \times 10^{-2}\}$, matching Theorem 2.

4.2 The exchange law and its mechanism

To answer how much breaking a budget actually needs, we sweep the dial from 3×10^{-4} to 1 at fixed budgets and locate the smallest sufficient $\lambda^\star(T)$, a grid-resolved upper bound at each budget. With Adam it falls with budget (Fig. 3b). The five-seed sweep gives $\lambda^\star(50) = 10^{-2}$ and $\lambda^\star(100) = 3 \times 10^{-3}$, and a two-seed fine grid in the same panel gives $\lambda^\star(200) = \lambda^\star(400) = \lambda^\star(800) = 5 \times 10^{-4}$, each at the grid floor. In the budget range the grid resolves, $\lambda^\star(T)\, T$ lies between 0.1 and 0.5, consistent with

$$\lambda^\star(T) \approx 0.3/T, \tag{6}$$

the form Proposition 1 predicts. Beyond $T \approx 200$ the crossing sits at the grid floor, so $\lambda^\star$ is only bounded there and the $1/T$ reading is not resolved further.

To verify the mechanism behind the exchange law, we record the mirror signal q at every step. Under Adam the trajectories of q collapse onto one line per λ, $q(T) \approx \kappa\, \lambda \sum_{t \le T} \eta_t$, with a λ-independent amplification $\kappa \approx 9$–10; the dotted overlays in Fig. 3c show this prediction tracking the measured trajectories with no fit. Full separability then occurs at a nearly constant signal budget, $\lambda \sum_t \eta_t \in [0.0018, 0.0045]$ across the three resolved dial values, the exchange law seen from the inside, and the closed-form slope verifies seed by seed with no free parameter (Fig. 4c).

Every sweep so far uses Adam, so the exchange law could be a property of that optimizer rather than of symmetry breaking itself. To separate the two, we repeat the sweep under plain gradient descent, for which Prop. 1 predicts that no budget can purchase signal. The prediction holds. The signal freezes near its initial value (Fig. 3c, dashed), moving by 12% to 110% over the whole budget with increments scaling as λ^2, so the dial has to clear the floor set by its initial signal alone, and the budget needed to do so grows as $1/\lambda^2$. The same split appears at the level of the task (Fig. 3a): under Adam even $\lambda = 10^{-3}$ reaches full accuracy by $T = 200$, under plain GD $\lambda = 0.1$ stalls at 98.5% and only $\lambda = 1$ passes, $10^3\times$ more breaking at the matched budget, the contrast Fig. 3b measures. *Preconditioning is what makes the protection free.*

The grid-search alternative costs eight trainings to reach $\lambda = 10^{-3}$, and below $T = 50$ no value under 10^{-2} works, where equation 6 points. Every other route pays somewhere, in search, in trainings, or in symmetry; the exchange-law row is the only one that pays nothing at the calibrated task. The constants are themselves sweep estimates, so on a new task the prescription costs one small calibration where a full hand search costs a grid. Appendix B tabulates the search, the trainings and the kept symmetry of every route at the same 400-step budget.

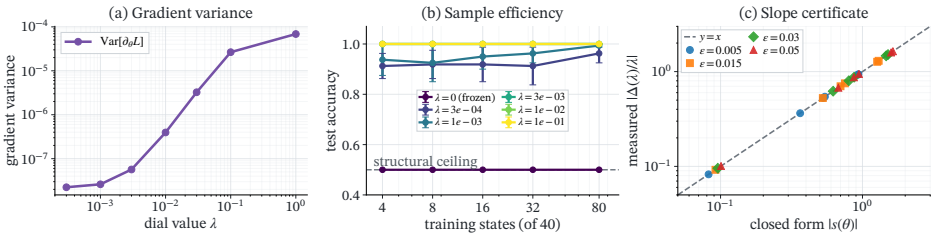

Figure 4: **The cost of protection and the certificate of the transport slope. (a)** Gradient variance at initialization in the dial coordinates grows with λ (from 2.2×10^{-8} at $\lambda = 0$ to 6.9×10^{-5} at $\lambda = 1$); part of the growth is the $\lambda\theta$ chain-rule factor, and no additional optimization barrier appears at small λ. **(b)** Test accuracy vs. number of training states, swept over $\{4, 8, 16, 32, 80\}$ with the 80-state test set fixed (200 steps). At four training states every tested $\lambda \geq 3 \times 10^{-3}$ is at full accuracy, so the three largest dial values coincide at 1.000; $\lambda = 10^{-3}$ sits at 0.938 and $\lambda = 0$ stays at exactly 0.500 regardless of data. **(c)** The susceptibility formula of equation 5 against the measured transport slope, seed by seed across ε and six initializations; the $y = x$ agreement with no free parameter certifies Cor. 1 directly.

4.3 The dial against its alternatives

We benchmark three alternative ways of opening a symmetry on the main task. (i) *Random-subset breaking* opens a random subset of k of the 60 breaking gates at full angle; the practitioner searches over the support size and the draw. (ii) *Penalty regularization* trains the unconstrained ansatz with a term $\mu\,\mathbb{E}|\boldsymbol{\theta}_{\mathrm{brk}}|$ pulling the breaking angles toward the symmetric point; the practitioner searches over μ. (iii) *A symmetry-matched compact model* is a translation-invariant (TI) equivariant ansatz with one shared angle per gate type per layer, 18 trainable parameters instead of 132, equipped with the dial.

All three routes reach full separability, and they differ in what must be searched first. The dial avoids the per-candidate grid search of the alternatives, its operating point coming from a formula, its endpoints from Thm. 1 and Cor. 1, its budget trade from the measured law (Prop. 1); the constants are calibrated on the task family, so what the dial removes is the one-training-per-candidate search, not the calibration itself. The subset and penalty routes search at one training per candidate, the cost that grows out of reach on hardware or at scale. Equipping the compact model with the dial reaches full accuracy with 18 trainable parameters, a $7.3\times$ reduction, at $\lambda = 10^{-2}$; within this class the fully broken setting is no longer the best operating point, with $\lambda = 1$ averaging 0.975 over three seeds at $T = 400$ against 1.000 for the interior dial, the constant rescaling as the signal-budget mechanism predicts.

4.4 The cost of protection in samples and in trainability

To answer whether the kept symmetry costs anything in samples or in trainability, we shrink the training set from 80 states to 4 and track the gradient variance at initialization (Fig. 4). The gradient variance *grows* with λ (Fig. 4a), so the equivariant model is the least trainable point on the axis, and switching the dial on only adds usable directions. One caveat sharpens what this measures. The variance is taken in the dial coordinates, and since a breaking gate enters through the angle $\lambda\theta$, part of the growth is a chain-rule factor of the parameterization. What we can support is that in the coordinates actually optimized, small dial values add gradient signal rather than suppressing it, and training matches this, every $\lambda > 0$ converging at the same learning rate without instability. Shrinking the training set to 4 states leaves every $\lambda \geq 3 \times 10^{-3}$ at full accuracy (Fig. 4b), while very small dials ($\lambda = 3 \times 10^{-4}$ stays between 0.91 and 0.96 at every training-set size) barely improve with more data, so optimization budget is the binding resource.

4.5 Task strength enters as a two-variable law

How does the required breaking depend on the difficulty of the distinction itself? Figure 5 sweeps the intrinsic task perturbation ε over a decade. The $\lambda^\star(T)$ curves remain parallel on a log–log plot,

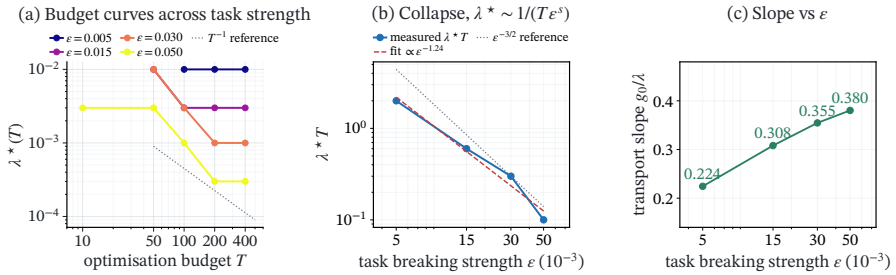

Figure 5: **Task strength enters as a shift, not a slope. (a)** $\lambda^\star(T)$ for $\varepsilon \in \{0.005, 0.015, 0.03, 0.05\}$, parallel curves on a log–log plot. **(b)** The collapse $\lambda^\star T$ vs. ε with an $\varepsilon^{-1.5}$ reference; the measured exponent on the grid-limited medians is -1.24. **(c)** The transport slope $s = g_0/\lambda$ increases with ε (0.224 at $\varepsilon = 0.005$ to 0.380 at $\varepsilon = 0.05$), since a stronger task perturbation makes the same dial value more readable at initialization.

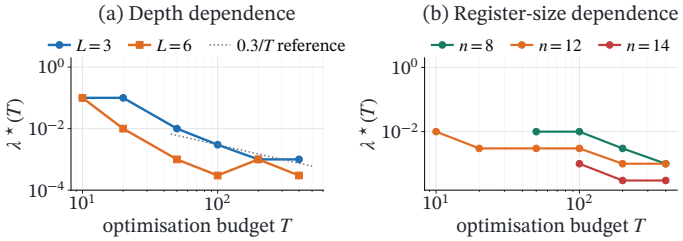

Figure 6: **The laws are robust. (a)** Depth $L = 6$ vs. $L = 3$, both following $\lambda^\star \propto 1/T$ with architecture-dependent constants, and both sitting at the exact ceiling at $\lambda = 0$. **(b)** Register size. At matched budget larger registers need *less* breaking ($\lambda^\star(400) = 3 \times 10^{-4}$ at $n = 14$ vs. $\sim 10^{-3}$ at $n = 8$–12).

the $1/T$ shape unchanged, while their level shifts monotonically (Fig. 5a). $\lambda^\star(400)$ falls from 10^{-2} at $\varepsilon = 0.005$ to 3×10^{-4} at $\varepsilon = 0.05$. A calibrated two-parameter summary of the probed range is

$$\lambda^\star(T, \varepsilon) \;\approx\; \frac{1.8 \times 10^{-3}}{T\,\varepsilon^{1.5}}, \tag{7}$$

which holds to within $1.3\times$ at the calibration budget $T = 400$ (Fig. 5b). The ε exponent is calibrated by the reference collapse of Fig. 5b, not independently verified; a direct fit over the $T \geq 100$ checkpoints on the grid-limited medians gives -1.24, and the coarse λ grid at small ε censors the crossing points. The transport slope $s = g_0/\lambda$ grows with ε (Fig. 5c, 0.224 to 0.380), so a stronger perturbation makes the same dial value more readable at initialization, and the ceiling stays at exactly 0.500 for every ε, as Theorem 1 requires. The fit reduces retargeting to recalibrating one constant, the budget trade of Sec. 4.2 supplying the rest.

4.6 Robustness across depth, size and task strength

To see whether the laws survive a deeper circuit and a wider register, we rerun the sweep along the two axes a practitioner would vary first (Fig. 6). Both statements of Sec. 3 survive in every configuration, the exact ceiling and the $1/T$ shape included. Larger registers need *less* breaking at matched budget, $\lambda^\star(400) = 3 \times 10^{-4}$ at $n = 14$, three times less than at $n = 8$–12 (Fig. 6b), consistent with the growing parameter count amplifying a weak breaking signal. The transport slope s is architecture-dependent (Fig. 6a, $s = 0.0599$ at $L = 6$ vs. 0.3546 at $L = 3$) while its linearity and zero intercept are not (Thm. 2).

4.7 A boundary where the dial transports nothing

To probe the premise of transport itself, we ask whether there exists a mirror class for which the dial transports nothing, and construct one whose conjugate stays diagonal. By equation 5, transport

requires $d_k \neq 0$ for some breaking gate, a generator that fails to commute with the readout on the state the symmetric block prepares. For a half-register mirror, $g = \prod_{i\,\mathrm{even}} Z_i$, with single-qubit breaking gates on the same support, this condition fails structurally, the conjugated mirror $UgU^\dagger$ remains a diagonal unitary and the gap is exactly zero for every λ and every invariant readout we tried. Appendix C documents the diagnostic.

5 Related work

Encoding conservation laws into ansätze is established practice, from symmetry-adapted variational schedules (Wecker et al., 2015) and symmetry-preserving state preparation (Gard et al., 2020) to qubit tapering by symmetries (Bravyi et al., 2017). Larocca et al. (2022b) give the general group-invariant framework that our equivariant block instantiates, and equivariant architectures now carry a full design toolkit, from general equivariant quantum neural network (EQNN) constructions (Nguyen et al., 2024) to gate symmetrization (Meyer et al., 2023) and applications on graphs, lattices and spin systems (Skolik et al., 2023; Zheng et al., 2023; Schatzki et al., 2022). What this toolkit does not quantify is the amount of breaking a mirror-type task needs.

Expressivity of parameterized circuits depends on the gate set and data encoding (Schuld et al., 2021; Sim et al., 2019), and its absence couples to trainability. Barren plateaus (McClean et al., 2018) arise from high expressibility (Holmes et al., 2022) and from controllability (Larocca et al., 2022a), their variance characterized exactly by the adjoint representation of the dynamical Lie group (Fontana et al., 2024) and unified by Ragone et al. (2024) (reviewed in Larocca et al., 2025). Concentration phenomena bound what QML models can learn from data (Thanasilp et al., 2024), generalization bounds quantify the sample cost (Caro et al., 2022), and information-theoretic arguments delimit where quantum models can beat classical ones (Huang et al., 2021; Liu et al., 2021). Theorem 1 places the mirror-task blindness in this landscape, and the transport and exchange laws extend it into the interior of the spectrum, where budget and optimizer, rather than capacity, set what a model can see.

Adaptive and Hamiltonian ansätze choose *which* generators to include. ADAPT-VQE grows circuits by gradient-greedy operator selection (Grimsley et al., 2019), and Hamiltonian variational ansätze reuse the problem's generators (Wiersema et al., 2020). Both are complementary to the dial, which instead fixes *which* gates may break the symmetry and by how much.

The relaxation of equivariance itself has recently become a design object, in classical deep learning and in QML alike. Lawrence et al. (2025) break symmetries probabilistically by randomizing over symmetry-related outputs; approximately equivariant networks treat the deviation as a penalty (Wang et al., 2022; Cohen et al., 2019); subgroup selection asks which symmetry to keep (Baumann et al., 2026); statistical tests decide when finite measurements justify releasing a symmetry constraint (Chen, 2026); and hardware noise breaks equivariance in measurable, partly compensable ways (Tüysüz et al., 2024), as do Hamming-weight-preserving subspaces (Monbroussou et al., 2025). The dial operates on the remaining axis, scaling the breaking generators at constant parameter count.

6 Discussion and conclusion

We introduced a one-parameter symmetry-breaking dial that interpolates between equivariant and unconstrained ansatze at constant parameter count. Full symmetry leaves the model blind on mirror-pair tasks, an exact ceiling that no capacity, depth or training lifts. Opening the dial produces a class separation linear to leading order, with a closed-form slope computable before any training, and the required breaking under a preconditioned optimizer follows $\lambda^\star \approx 0.3/T$ in the resolved budget range, extending to a calibrated joint law in budget and task strength. The prescription sets the dial from the training budget while leaving almost all of the symmetry intact, and the laws travel across depth, register size and task strength with recalibrated constants.

Future work can extend the dial to other mirrors and hardware settings, where noise-induced breaking plays the role of the dial, and can turn the necessary condition of Corollary 1 into a design-time test for learnability by characterizing the diagonal-conjugate structure of Appendix C.

REPRODUCIBILITY STATEMENT

The supplementary material provides anonymized simulation and plotting code together with the configurations used for every reported sweep. The main text and Appendix B specify the Hamiltonian, circuit, loss, optimizer schedules, seed counts, dial grids, and checkpoints $T \in \{10, 20, 50, 100, 200, 400, 800\}$. We used full-batch state-vector simulation and independently verified every reported theorem and numerical claim from these artifacts.

ETHICS STATEMENT

This work is theoretical and computational. It uses synthetic quantum states generated in simulation and contains no human subjects, personal data, or external datasets. All authors have read and adhere to the ICLR Code of Ethics.

AI USE STATEMENT

Generative AI tools were used to implement simulation and plotting code and to draft and edit parts of the manuscript. We did not use generative AI tools to formulate the theoretical claims, supply proof ingredients, propose hypotheses, design the experiments, or interpret the results. The authors reviewed and verified all AI-assisted code and prose and take responsibility for the final text, claims, and artifacts. This disclosure will be matched in the OpenReview submission form.

REFERENCES

Markus Baumann, Claudia Linnhoff-Popien, and Jonas Stein. Symmetry alone is not an ansatz: task-aligned interactions in equivariant quantum circuits. *arXiv preprint arXiv:2606.20316*, 2026.

Sergey Bravyi, Jay M. Gambetta, Antonio Mezzacapo, and Kristan Sheldon. Tapering off qubits to simulate fermionic Hamiltonians. *arXiv preprint arXiv:1701.08213*, 2017.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

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

M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3:625–644, 2021.

M. Cerezo, Guillaume Verdon, Hsin-Yuan Huang, Lukasz Cincio, and Patrick J. Coles. Challenges and opportunities in quantum machine learning. *Nature Computational Science*, 2:567–576, 2022.

Zeyu Chen. Statistical symmetry release for equivariant quantum learning. *arXiv preprint arXiv:2609.11470*, 2026.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, 2016.

Taco S. Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant CNNs on homogeneous spaces. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Enrico Fontana, Dylan Herman, Shouvanik Chakrabarti, Niraj Kumar, Romina Yalovetzky, Jamie Heredge, Shree Hari Sureshbabu, and Marco Pistoia. The adjoint is all you need: characterizing barren plateaus in quantum ansätze. *Nature Communications*, 15:7171, 2024.

Bryan T. Gard, Linghua Zhu, George S. Barron, Nicholas J. Mayhall, Sophia E. Economou, and Edwin Barnes. Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm. *npj Quantum Information*, 6:10, 2020.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773, 2012.

Harper R. Grimsley, Sophia E. Economou, Edwin Barnes, and Nicholas J. Mayhall. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nature Communications*, 10:3007, 2019.

Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567:209–212, 2019.

Zoë Holmes, Kunal Sharma, M. Cerezo, and Patrick J. Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum*, 3:010313, 2022.

Hsin-Yuan Huang, Richard Kueng, and John Preskill. Information-theoretic bounds on quantum advantage in machine learning. *Physical Review Letters*, 126:190505, 2021.

Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549:242–246, 2017.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

Martín Larocca, Piotr Czarnik, Kunal Sharma, Gopikrishnan Muraleedharan, Patrick J. Coles, and M. Cerezo. Diagnosing barren plateaus with tools from quantum optimal control. *Quantum*, 6: 824, 2022a.

Martín Larocca, Frédéric Sauvage, Faris M. Sbahi, Guillaume Verdon, Patrick J. Coles, and M. Cerezo. Group-invariant quantum machine learning. *PRX Quantum*, 3:030341, 2022b.

Martín Larocca, Supanut Thanasilp, Samson Wang, Kunal Sharma, Jacob Biamonte, Patrick J. Coles, Lukasz Cincio, Jarrod R. McClean, Zoë Holmes, and M. Cerezo. Barren plateaus in variational quantum computing. *Nature Reviews Physics*, 7:174, 2025.

Hannah Lawrence, Vasco Portilheiro, Yan Zhang, and Sékou-Oumar Kaba. Improving equivariant networks with probabilistic symmetry breaking. In *International Conference on Learning Representations (ICLR)*, 2025.

Yunchao Liu, Srinivasan Arunachalam, and Kristan Temme. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics*, 17:1013–1017, 2021.

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.

Johannes Jakob Meyer, Marian Mularski, Elies Gil-Fuster, Antonio Anna Mele, Francesco Arzani, Alissa Wilms, and Jens Eisert. Exploiting symmetry in variational quantum machine learning. *PRX Quantum*, 4:010328, 2023.

Léo Monbroussou, Jonas Landman, Alex B. Grilo, Romain Kukla, and Elham Kashefi. Trainability and expressivity of Hamming-weight preserving quantum circuits for machine learning. *Quantum*, 9:1745, 2025.

Quynh T. Nguyen, Louis Schatzki, Paolo Braccia, Michael Ragone, Patrick J. Coles, Frédéric Sauvage, Martín Larocca, and M. Cerezo. Theory for equivariant quantum neural networks. *PRX Quantum*, 5:020328, 2024.

Pierre Pfeuty. The one-dimensional Ising model with a transverse field. *Annals of Physics*, 57: 79–90, 1970.

Michael Ragone, Bojko N. Bakalov, Frédéric Sauvage, Alexander F. Kemper, Carlos Ortiz Marrero, Martín Larocca, and M. Cerezo. A Lie algebraic theory of barren plateaus for deep parameterized quantum circuits. *Nature Communications*, 15:7172, 2024.

Frédéric Sauvage, Martín Larocca, Patrick J. Coles, and M. Cerezo. Building spatial symmetries into parameterized quantum circuits for faster training. *Quantum Science and Technology*, 9: 015029, 2024.

Louis Schatzki, Martín Larocca, Quynh T. Nguyen, Frédéric Sauvage, and M. Cerezo. Theoretical guarantees for permutation-equivariant quantum neural networks. *arXiv preprint arXiv:2210.09974*, 2022.

Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. Effect of data encoding on the expressive power of variational quantum machine learning models. *Physical Review A*, 103:032430, 2021.

Sukin Sim, Peter D. Johnson, and Alán Aspuru-Guzik. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Advanced Quantum Technologies*, 2:1900070, 2019.

Andrea Skolik, Michele Cattelan, Sheir Yarkoni, Thomas Bäck, and Vedran Dunjko. Equivariant quantum circuits for learning on weighted graphs. *npj Quantum Information*, 9:47, 2023.

Supanut Thanasilp, Samson Wang, M. Cerezo, and Zoë Holmes. Exponential concentration in quantum kernel methods. *Nature Communications*, 15:5200, 2024.

Cenk Tüysüz, Su Yeon Chang, Maria Demidik, Karl Jansen, Sofia Vallecorsa, and Michele Grossi. Symmetry breaking in geometric quantum machine learning in the presence of noise. *PRX Quantum*, 5:030314, 2024.

Rui Wang, Robin Walters, and Rose Yu. Approximately equivariant networks for imperfectly symmetric dynamics. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, 2022.

Dave Wecker, Matthew B. Hastings, and Matthias Troyer. Progress towards practical quantum variational algorithms. *Physical Review A*, 92:042303, 2015.

Roeland Wiersema, Cunlu Zhou, Yvette de Sereville, Juan Felipe Carrasquilla, Yong Baek Kim, and Henry Yuen. Exploring entanglement and optimization within the Hamiltonian variational ansatz. *PRX Quantum*, 1:020319, 2020.

Han Zheng, Zimu Li, Junyu Liu, Sergii Strelchuk, and Risi Kondor. Speeding up learning quantum states through group equivariant convolutional quantum ansätze. *PRX Quantum*, 4:020327, 2023.

A Proofs

A.1 Proof of Theorem 1

Since $[U_{\mathrm{sym}}, g] = 0$ and $[R, g] = 0$ by construction, and U_{brk} is the identity at $\lambda = 0$, the equivariant end satisfies $Ug = gU$ and therefore

$$R\big(U|\psi_-\rangle\big) = R\big(Ug|\psi_+\rangle\big) = R\big(g\,U|\psi_+\rangle\big) = R\big(U|\psi_+\rangle\big),$$

where the last step uses the g-invariance of R. The two classes receive identical outputs for every $\boldsymbol{\theta}$, so a threshold classifier with a fixed tie rule assigns them the same predicted label and is correct on exactly half of every mirror pair, hence on exactly 50% of any balanced evaluation set. ■

A.2 Proof of Theorem 2

For the breaking generators, $g\,X_i\,g = -X_i$ and $g\,Y_i\,g = -Y_i$, whence $g\,RX_i(\vartheta)\,g = RX_i(-\vartheta) = RX_i(\vartheta)^\dagger$ and similarly for RY_i. Therefore $g\,U_{\mathrm{brk}}(\lambda\boldsymbol{\theta})\,g = U_{\mathrm{brk}}(-\lambda\boldsymbol{\theta})$, that is $U_{\mathrm{brk}}(-\lambda\boldsymbol{\theta})\,g = g\,U_{\mathrm{brk}}(\lambda\boldsymbol{\theta})$, and since $[U_{\mathrm{sym}}, g] = 0$,

$$U(\boldsymbol{\theta}; -\lambda)\,g \;=\; U_{\mathrm{brk}}(-\lambda\boldsymbol{\theta})\,U_{\mathrm{sym}}\,g \;=\; U_{\mathrm{brk}}(-\lambda\boldsymbol{\theta})\,g\,U_{\mathrm{sym}} \;=\; g\,U_{\mathrm{brk}}(\lambda\boldsymbol{\theta})\,U_{\mathrm{sym}} \;=\; g\,U(\boldsymbol{\theta}; \lambda).$$

Applying this to ψ_+ and using the g-invariance of R gives $R\big(U(-\lambda)\psi_+\big) = R\big(U(\lambda)\psi_-\big)$. Swapping the roles of the two classes (both are real ground states, so both constructions hold) yields $\Delta(\boldsymbol{\theta}; -\lambda) = -\Delta(\boldsymbol{\theta}; \lambda)$ at every initialization. An exactly odd analytic function has an odd power series, so $\Delta(\boldsymbol{\theta}; \lambda) = s(\boldsymbol{\theta})\lambda + O(\lambda^3)$ pointwise, and the same expansion holds for the expectation over any initialization. Numerically, $\Delta(-\lambda) = -\Delta(\lambda)$ holds to machine precision for $n = 8, 10, 12, 14$ and $\lambda \in \{\pm 10^{-3}, \pm 10^{-2}, \pm 3 \times 10^{-2}\}$. ■

A.3 Proof of Corollary 1

The circuit applies the symmetric block first and the breaking block second, so $U(\boldsymbol{\theta};\lambda) = U_{\mathrm{brk}}(\lambda\boldsymbol{\theta})\,U_{\mathrm{sym}}$ and $\phi_+ = U_{\mathrm{sym}}|\psi_+\rangle$ is the state entering the breaking block. To first order, $U_{\mathrm{brk}}^{\dagger}\,R\,U_{\mathrm{brk}} = R + \frac{i\lambda}{2}\sum_k \theta_k[G_k, R] + O(\lambda^2)$. Reflecting with g flips the sign of the commutator, $g\,[G_k, R]\,g = -[G_k, R]$, because breaking generators anticommute with g while the readout commutes; the two classes therefore carry opposite first-order terms, and their difference doubles them,

$$\Delta(\boldsymbol{\theta};\lambda) \;=\; i\lambda \sum_k \theta_k \left\langle \phi_+ \middle| [G_k, R] \middle| \phi_+ \right\rangle + O(\lambda^3).$$

Differentiating the bracketed expectation gives $\frac{d}{dt}\big|_0 R\big(e^{itG_k}|\phi_+\rangle\big) = i\,\langle\phi_+|[R, G_k]|\phi_+\rangle = -i\,\langle\phi_+|[G_k, R]|\phi_+\rangle$, which is equation 5. The remainder is the next odd order, of size $O(\lambda^2)$ relative to the linear term. Numerically the formula reproduces the measured signed slope $\Delta(\boldsymbol{\theta};\lambda)/\lambda$ for all 24 (seed, ε) pairs to three or four significant digits, with a typical relative deviation of a few parts in 10^4 and a worst case below 10^{-3}. ■

A.4 Proof of Proposition 1

The unit-step model is an idealized envelope on the optimizer, not a step-by-step description of Adam. We assume a preconditioned update whose per-coordinate displacement is bounded by η (Adam with bounded learning rate is well approximated by this, since its per-coordinate update is the learning rate times a bounded ratio, though the ratio is not exactly 1 at every step), so after T steps $\sum_k |\theta_k^{(T)}| \le \sum_k |\theta_k^{(0)}| + M\eta T$. By the susceptibility formula the mirror signal obeys $q \le \lambda\, d_{\max} \sum_k |\theta_k|$; if full separability requires $q \ge q^\star$, then $\lambda \ge q^\star \big/ \big(d_{\max}(\sum_k |\theta_k^{(0)}| + M\eta T)\big)$, decaying as $1/T$. The resolved flip points give $\lambda^\star \sum_t \eta_t \in [0.0018, 0.0045]$ (Sec. 4.2), the constant-signal-budget form of this bound. The $1/T$ reading holds while the schedule still accumulates, since the halving schedule drives $\sum_t \eta_t$ to 8 and the bound then stops improving with T. ■

Two features of the descent decide whether the budget can be used at all. A preconditioned update takes its step from the learning rate rather than from the gradient magnitude, so a breaking parameter keeps moving at rate η where the gradient is small, and the signal accumulates with the budget. Unpreconditioned descent behaves in the opposite way. A breaking parameter enters the circuit only through the angle $\lambda\theta$, so the loss gradient with respect to it carries a factor λ and vanishes linearly with the dial, and since the signal responds to the angle rather than to the parameter, a plain step moves the signal by $O(\lambda^2)$. Over 400 steps the signal moves by $+12\%$, $+35\%$ and $+110\%$ at $\lambda = 10^{-3}, 3\times10^{-3}, 10^{-2}$, with increments scaling as λ^2 (factors 9 and 94 per $3\times$ and $10\times$ λ). No budget in our range lifts a sub-threshold dial over the floor, which is why only $\lambda = 1$ separates under GD.

B Experimental details

The data are ground states of $H(h,\varepsilon) = -\sum_i X_i X_{i+1} - h\sum_i Z_i - \varepsilon\sum_i X_i$ ($n = 10$ unless stated, periodic boundaries), computed by sparse Lanczos in `float32`; the negative class is the *exact* symmetry image $g\psi_+$ (a diagonal phase $(-1)^{\mathrm{popcount}(x)}$ per basis state), which removes any float-level asymmetry between the classes. The h-grid $\{0.70, \ldots, 1.30\}$ (40 values) is split alternately into 20 train and 20 test values, so both splits share the same marginal distribution; each h contributes one $+$ and one $-$ state, giving $40/40$ states at $\varepsilon = 0.03$.

The model is L layers of [RZ on all sites + XX, YY, ZZ on alternating nearest-neighbour pairs], followed by L blocks of [RX, RY on all sites], with the angles of the latter multiplied by λ. Each layer carries n RZ angles, 3 angles per brick-wall pair, and $2n$ RX/RY angles. The brick-wall alternates between the $\lceil (n-1)/2 \rceil$ pairs $(0,1), (2,3), \ldots$ and the $\lfloor (n-1)/2 \rfloor$ pairs $(1,2), (3,4), \ldots$, giving $P = \lceil L/2 \rceil \lceil (n-1)/2 \rceil + \lfloor L/2 \rfloor \lfloor (n-1)/2 \rfloor$ two-qubit pairs in total, so

$$n_{\mathrm{sym}} = Ln + 3P, \qquad n_{\mathrm{brk}} = 2Ln, \qquad n_{\mathrm{par}} = 3Ln + 3P.$$

For $n = 10$, $L = 3$ this gives $n_{\mathrm{par}} = 132$ with 72 symmetric and 60 breaking. The readout is the mean nearest-neighbour $\langle Z_a Z_{a+1} \rangle$.

Table 2: **Routes to full separability on the mirror-pair task** (main setting $n = 10$, $L = 3$, standard ansatz 132 parameters, $40/40$ split, three to five seeds per row). All routes, including every search training, run the same 400-step Adam budget. "Search" is what a practitioner must tune before the first usable model; "trainings" counts the runs of our sweep including the search. "Breaking fraction" is the dial coordinate, the fraction of the breaking-angle scale left open, $\lambda = 0$ being fully equivariant and $\lambda = 1$ unconstrained.

Route	Search	Trainings	Accuracy	Breaking fraction
Equivariant ($\lambda = 0$)	none	1	50% (ceiling)	0%
Unconstrained ($\lambda = 1$)	none	1	100%	100%
Dial, grid search	λ grid	8	100%	whatever the grid finds
Dial with plain GD	optimizer	1	$<99.9\%$ below $\lambda=1$	100% in practice
Dial at the exchange law	**none** ($\lambda = 0.3/T$)	**1**	**100%**	$0.3/T \approx$ **0.1%**
Random subset, k of 60 gates at full angle	support (k, draw)	15	100% for $k \geq 2$	$k/60$
Penalty $\mu\,\mathbb{E}\lvert\boldsymbol{\theta}_{\mathrm{brk}}\rvert$	strength μ	15	100% (all μ tried)	uncontrolled
TI-equivariant + dial (18 params)	none	1	100%	$\lambda \approx 1\%$

Optimization uses a full-batch softplus loss on the margin. Adam runs at learning rate 0.05, halved every 80 steps; the GD baseline at learning rate 0.25. Test accuracy is evaluated at every step and recorded at checkpoints $T \in \{10, 20, 50, 100, 200, 400\}$ (and 800 where stated); $\lambda^\star(T)$ uses the mean over seeds against a 99.9% threshold. Seeds are five for the headline sweeps and two to three for the fine grids as stated per figure.

The two-qubit gate was validated against the exact Pauli exponential (agreement 3×10^{-8}), and the mirror by $|\langle g\psi_+|\psi_-\rangle| = 1$ and by $gH(+\varepsilon)g = H(-\varepsilon)$ elementwise.

The signal-trajectory runs of Fig. 3c use $\lambda \in \{10^{-3}, 3\times10^{-3}, 10^{-2}\}$, one seed, full-batch, $\varepsilon = 0.03$, recording every step the test accuracy, the signed mean mirror margin and its absolute value on the training set, and the breaking angles. The amplification κ is the median of $q/(\lambda \sum_{t\leq T} \eta_t)$ over $T \leq 200$. The susceptibility check of Fig. 4c evaluates equation 5 by central finite differences ($t = 10^{-3}$) on the state entering the breaking block, for six seeds ($\boldsymbol{\theta} \sim \mathrm{Unif}[-\pi, \pi]$) and $\varepsilon \in \{0.005, 0.015, 0.03, 0.05\}$, against the measured $\Delta(\lambda)/\lambda$ at $\lambda = 10^{-3}$.

The alternative-route benchmarks of Sec. 4.3 use the following configurations. Random subset opens $k \in \{2, 4, 8, 16, 32\}$ of the 60 gates at full angle, three random draws each, Adam, 400 steps. Penalty adds $\mu\,\mathbb{E}|\boldsymbol{\theta}_{\mathrm{brk}}|$ to the unconstrained loss with $\mu \in \{10^{-4}, \ldots, 1\}$, three seeds. The compact model shares one angle per gate type per layer ($4L$ symmetric and $2L$ breaking parameters, 18 at $L = 3$) and sweeps $\lambda \in \{0, 3 \times 10^{-4}, \ldots, 1\}$ with three seeds. The depth probe runs the standard ansatz at $L \in \{6, 9, 12\}$ and $\lambda \in \{10^{-3}, 1\}$ for three seeds and 400 steps. At every depth the unconstrained ansatz reaches full accuracy at the first checkpoint ($T = 10$), and the dial at $\lambda = 10^{-3}$ reaches it by $T \leq 50$, so across the probed depths the measured dividends of the kept symmetry are the parameter reduction and the interior operating point of Sec. 4.3.

C THE ZERO-TRANSPORT BOUNDARY

For a half-register mirror $g = \prod_{i\,\mathrm{even}} Z_i$ with single-qubit breaking gates RX_i, RY_i (i even) we measure $\Delta(\lambda) \equiv 0$ exactly, robustly across the readout choice (nearest-neighbour ZZ, same-parity ZZ, same-parity XX), across gate subsets (only RX, only RY, both), and across a two-site mirror Z_0Z_1. The diagnostic behind Sec. 4.7 is direct. The conjugated mirror $P = U(\lambda)\, g\, U(\lambda)^\dagger$ is exactly diagonal (off-diagonal norm at float32 noise level), whereas for the global mirror of the main text P is dense, with off-diagonal norm rising from 3.6 to 16 over $\lambda \in [0.03, 1]$ at the diagnostic initializations. A diagonal P means $Ug\psi_+ = PU\psi_+$ differs from $U\psi_+$ by phases only, so every computational-basis outcome distribution, and hence every readout built from them, is class-independent. Characterizing precisely which (group, architecture, readout) triples leave P diagonal delimits where symmetry-breaking dials can work at all.