LEARNING WHEN AND HOW TO USE MEASUREMENT FEEDBACK IN DYNAMIC QUANTUM CIRCUITS

Anonymous authors
Paper under double-blind review

ABSTRACT

Mid-circuit measurements let a circuit steer later operations with information it has just created. Yet the feedforward rule that maps outcomes to gates is still designed by hand, one bespoke rule per protocol, while k measured qubits create 2^k outcome contingencies. External controllers move it outside the circuit, leaving open when conditioning helps. We make the feedforward rule a trainable part of the circuit. Exact branch decomposition enumerates the measurement outcomes inside the optimization loop, so the rule angles and every circuit parameter descend one balanced branch objective together, with no sampling estimator. It also yields analytic floors that bind any parameter-matched circuit ignoring the outcome, a spectral energy gap and a chance-level classification bound, which certify before training whether the outcome can change the attainable objective. Across energy and classification tasks the learned rule crosses those floors, clears every unconditional variant, and at matched budgets and restarts ties or exceeds the best baseline on 26 of 27 grid cells. Under a hardware-informed deployment model its net value is task dependent, repaying the feedback latency on classification while energy recovery crosses to a static preference. A mid-circuit measurement is thus a computational resource whose value can be certified and whose use can be learned.

1 INTRODUCTION

A mid-computation discrete event can both select later operations and report on the state of the system, so whether conditioning on its outcome pays, and how to design the conditioning rule, is a general question for hybrid learning systems. Dynamic quantum circuits are its sharpest instance, where a mid-circuit measurement yields a classical outcome and conditioned gates act on it before the final readout. That classical layer is where a processor acts on what it has just learned about itself, which is why the fault-tolerant stack is built out of it. Error correction spends its syndromes this way (Fowler et al., 2012; Terhal, 2015; Google Quantum AI, 2023), and so do teleportation-based gates, lattice surgery, and a growing family of measurement-driven variational algorithms (Deshpande et al., 2024; Lemelin et al., 2025), with hardware platforms executing such circuits routinely (Kim et al., 2023).

Existing practice handles the rule, the map from outcome to conditioned operations, in two ways. Design it by hand, as in the standard correction that flips the measured qubit, which works only for tasks with known structure, one bespoke design per task; or learn it with external machinery, reinforcement-learning (RL) policies over measurement records (Nie et al., 2026; Wang & You, 2025; Nautrup et al., 2019), neural decoders of outcome strings (Alam & Clark, 2024), or fixed heuristic searches (Lemelin et al., 2025). Both routes share a hidden premise, that the rule must be specified before training or optimized outside the circuit, and neither keeps pace with the hardware. Each protocol ships its own rule, and the rule space grows as 2^k with the measured register, so every larger register is a fresh design problem.

The premise now binds. Dynamic parameterized quantum circuits (DPQCs) are trainable and at least as expressive as their unitary counterparts (Deshpande et al., 2024), yet the feedforward rule there is designed by hand; external policies split the optimization into two systems and give up end-to-end differentiability; and whether a task benefits from conditioning is answered by trial and error. The obstacle is structural. The measurement is a discrete event that blocks gradients, and the standard

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.

escapes, score-function estimators and biased surrogates (Williams, 1992; Bengio et al., 2013), pay a per-step variance or bias that grows with the outcome register.

We take the direct route. Parameterize the rule by a handful of angles, put it inside the circuit, and descend the same loss as every other parameter. The enabler is *exact branch decomposition*. For the small registers relevant to variational feedforward, we simulate every measurement branch explicitly and train on the average of the exact branch losses, keeping the probability-gradient normalizers in the autograd graph. That average is our *balanced branch objective*. It pays every branch the same attention regardless of how likely that branch is, and the floors below bound it, whereas the outcome-weighted sum $\sum_m p_m \ell_m$ is the expectation of one sampled run. The same decomposition yields analytic floors, lower bounds that bind any circuit ignoring the outcome, the measurement-independent or unconditional case. A spectral gap of $\frac{1}{2}(E_1 - E_0)$ on energy tasks and a chance-level bound on classification turn the vague question into *does the learned rule cross a floor that no unconditional circuit can cross?* We contribute four things.

- We derive analytic floors for parameter-matched measurement-independent circuits that bound every circuit ignoring the outcome, turning whether the channel can pay into a quantity a designer can evaluate before any training run.
- We place the feedforward rule inside the circuit as four trainable angles and optimize them with every gate parameter by exact branch enumeration under a balanced branch objective, no sampling estimator in the loop; the rule crosses the floors wherever the hand prior stops covering the task.
- One four-angle rule family wins or ties on 26 of the 27 grid cells at matched-restart budgets; it clears every unconditional variant, wins strictly against a tuned hand family and on real data, and matches the hand rule where that rule is already optimal.
- We evaluate the learned rule under a hardware-informed deployment model that combines readout flips, feedback waiting, and conditional-pulse errors. The resulting task-dependent crossover separates settings where outcome-aware control repays its feedback cost from settings where an unconditional circuit is preferable.

2 Related work

Four lines of work surround our question, and none trains the rule itself, inside the circuit, on the task objective. Dynamic-circuit architectures fix the rule in advance, learned controllers and gradient estimators reach the measurement from outside the circuit, and landscape analyses study how measurement reshapes training.

Closest to us, Deshpande et al. (2024) introduced and analyzed dynamic quantum circuits with the feedforward rule fixed in advance, proving no barren plateaus, width-exponential gradient decay, and expressivity beyond unitary circuits of equal depth; our trainability analysis builds on those gradients, and our delta is to make the rule itself the trained object. Alam & Clark (2024) variationally optimize pre- and post-measurement gates with tensor-network sweeps but train a decoder only for extensive ancilla registers, and Lemelin et al. (2025) systematize heuristic mid-circuit rules. Feedback policies (Wang & You, 2025), RL controllers (Nie et al., 2026), outside-trained reservoir readouts (Gonon et al., 2025), and RL recovery of preparation circuits (Liao et al., 2025) all keep the rule or the readout outside the trained circuit, and hardware-level error mitigation is an orthogonal engineering layer (Santos & Uzdin, 2025; Tong et al., 2026). Our delta has two parts. Four rule angles live inside the circuit and train end-to-end on the task objective, with no decoder and no estimator, and they come with analytic floors, a certificate that prices the channel before training with no analogue in either a hand rule or a trained decoder.

On the gradient side the measurement is a discrete stochastic node, and the standard escapes are the score-function (REINFORCE) estimator (Williams, 1992; Schulman et al., 2015), straight-through biased surrogates (Bengio et al., 2013), Gumbel-softmax relaxations (Jang et al., 2017), and parameter-shift rules, unbiased for gate parameters but not for the branch choice (Sweke et al., 2020; Schuld et al., 2019). Small outcome spaces let the enumeration behind our exact branch decomposition remove the estimator entirely; for large registers such as quantum error-correcting syndrome strings, enumeration is exponential and external learned decoders (Alam & Clark, 2024; Nautrup et al., 2019) remain the right tool, our scope the small-register regime.

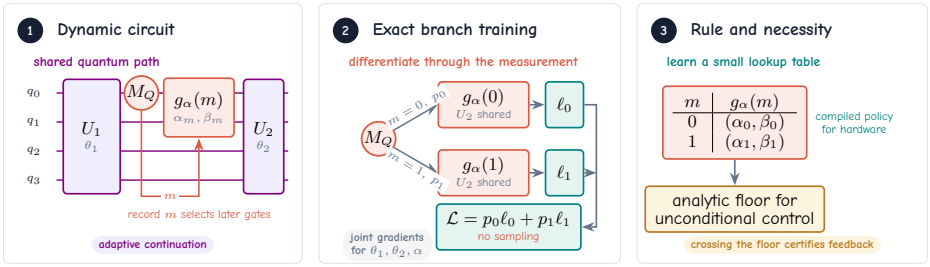


Figure 1: **Learning when and how to use measurement feedback.** *Panel 1 (dynamic circuit):* a preparation layer $U_1(\theta_1)$, one mid-circuit measurement of a register qubit Q yielding outcome m, conditioned gates executing the rule $g_\alpha(m) = (\alpha_m, \beta_m)$, and a shared processing layer $U_2(\theta_2)$ (ring / brick / star differ only in entangler connectivity). *Panel 2 (exact branch training):* the measurement is not sampled; both branches share one autograd graph, the loss the average of the exact branch losses. *Panel 3 (rule and necessity):* the learned rule compiles to a small lookup table over measurement outcomes, and analytic floors certify when the measurement record is genuinely used.

Intermediate measurements also reshape training landscapes through entanglement phase transitions (Wiersema et al., 2023) and measurement-induced landscape transitions (Gyawali et al., 2023), and systematic trainability failures persist inside DPQCs despite anti-concentration of the outcome statistics (Shirgure et al., 2026). Our parameterization keeps the rule structurally outcome-dependent, the failure mode flagged there, and our floors certify at task level that the conditional channel is *used*. Our tasks and circuits are standard variational primitives, from hardware-efficient variational quantum eigensolvers (VQE) (Peruzzo et al., 2014; Kandala et al., 2017) and QAOA (quantum approximate optimization algorithm) style classification (Farhi et al., 2014) to feature-space (Havlíček et al., 2019; Schuld & Killoran, 2019), convolutional (Cong et al., 2019) and generative models (Zoufal et al., 2019; Biamonte et al., 2017; Cerezo et al., 2021; Mitarai et al., 2018; Pérez-Salinas et al., 2020), with barren plateaus (McClean et al., 2018; Larocca et al., 2024; Wang et al., 2021) and the expressiveness–gradient link (Holmes et al., 2022) motivating the DPQC regime, mitigation running from layerwise training (Skolik et al., 2021) to symmetry-aware ansatz (circuit-template) design (Bravyi et al., 2020), while simultaneous-perturbation methods (Spall, 1992) sidestep the measurement event only by paying per-step estimation noise, and expressibility and data-power analyses (Sim et al., 2019; Huang et al., 2021; Caro et al., 2022) frame the classification tasks alongside device characterization (Gambetta et al., 2017; Tindall et al., 2024).

3 Method

3.1 Problem setup

An n-qubit DPQC in our setting makes one measurement before the readout, and its result decides the angles of a few later gates. The circuit splits at the measurement site into a preparation layer $U_1(\theta_1)$, a computational-basis (Z-basis) measurement of the register qubit Q whose outcome $m \in \{0, 1\}$ the rule consumes, and a processing layer $U_2(\theta_2)$ preceded by *conditioned gates* whose angles are functions of m. The conditional circuit family we study is

$$|\psi_m\rangle \;=\; U_2(\theta_2)\; R_Y\big(\beta_m\big)_r\; R_X\big(\alpha_m\big)_Q\; P_m\, U_1(\theta_1)\,|0^n\rangle, \tag{1}$$

where P_m projects the measured qubit onto $|m\rangle$, r is the readout qubit, the one measured at the end ($r = Q{+}1$ by default, the hub in the star architecture), and the rule

$$g_\alpha(m) = \big(\alpha_m, \beta_m\big), \qquad \alpha_m, \beta_m \in \mathbb{R}, \tag{2}$$

holds four parameters $\{\alpha_0, \alpha_1, \beta_0, \beta_1\}$. A *manual* rule fixes $g(m) = (\pi m, 0)$, the standard "flip the measured qubit" correction. An *unconditional* rule uses the same gate set with m-independent angles. The learned rule trains all four angles jointly with θ_1, θ_2.

3.2 Learning the rule end-to-end

The obstacle is that a measurement is a discrete event, so gradient descent cannot flow through a sampled outcome. Training never samples the measurement (Figure 1, panel 2); both branches are

simulated exactly, and the loss averages their exact branch losses. Write the pre-measurement state as $|\psi\rangle = U_1(\theta_1)|0^n\rangle$ and decompose over the measured qubit,

$$|\psi\rangle = \sqrt{p_0}\,|0\rangle_Q|\phi_0\rangle + \sqrt{p_1}\,|1\rangle_Q|\phi_1\rangle, \tag{3}$$

with the Q-marginal of $|\phi_m\rangle$, the branch *payload* on the unmeasured sites, normalized. The post-measurement state on branch m is $|m\rangle_Q|\phi_m\rangle$ and the objective becomes the *balanced branch objective*

$$\mathcal{L}_{\mathrm{bal}}(\theta_1, \theta_2, \alpha) = \tfrac{1}{2} \sum_{m\in\{0,1\}} \ell\big(U_2\, g_\alpha(m)\, |m\rangle|\phi_m\rangle\big), \tag{4}$$

with each branch loss ℓ_m evaluated on its normalized payload. We train $\mathcal{L}_{\mathrm{bal}}$ because it pays the branches equally, rare branches the preparation produces included, and because the floors below bound it. The outcome-weighted average $\sum_m p_m \ell_m$ is a different quantity, the expectation a single sampled run returns and hence the cost a deployed circuit pays. The two coincide only when the branch probabilities are equal, which nothing in training enforces.

The decomposition is exact, and the normalized payloads carry θ_1-dependence through their normalizers, so gradients flow to θ_1 through *both* the branch payloads and the normalization channel, and to α through the branch-conditioned gates, with no estimator bias. The objective costs two state vectors per measurement, twice the memory of a unitary circuit at the same width, while REINFORCE adds sampling variance that can dominate at small sample budgets; for k measured qubits the factor is 2^k, hence the small-register scope.

Trainability carries over under one stated condition. The balanced branch objective adds no width-dependent suppression beyond the underlying branch circuits, provided the two branch gradients do not cancel each other. A gradient-variance scan up to $n{=}16$ confirms this, including the branch-collapse failure mode, one branch probability driven toward zero, that the outcome-weighted variant suppresses.

3.3 When is feedback necessary?

This subsection prices the channel before training. If the circuit ignores the outcome entirely, how well can it do? For broad task families the limit is analytic. Ignoring the outcome means an m-*independent* processing circuit acting on whatever payloads the preparation produces. The floors depend on the task, not on the training method, so a designer can price a measurement register before building it. We state them for our three task families.

Proposition 1 (Ground-state-energy floor; Ky Fan) *Let the two post-measurement branch states be orthonormal, as any unitary preparation produces, and let the Hamiltonian H have a non-degenerate ground state, with ground energy E_0 and first excited energy E_1. Then for any shared (m-independent) processing unitary V, the balanced energy $\frac{1}{2}(\ell_0 + \ell_1)$ obeys*

$$\tfrac{1}{2}\big(\ell_0(V) + \ell_1(V)\big) \;\geq\; E_0 + \tfrac{1}{2}\,(E_1 - E_0), \tag{5}$$

since two orthonormal inputs cannot both lie in the ground space, which is one-dimensional (Ky Fan's maximum principle on eigenvalue sums). A ground space of dimension $g \geq 2$ relaxes the statement to the trivial bound E_0.

For our ground-state-energy task, the transverse-field Ising model (TFIM) at ferromagnetic coupling $J{=}1$ and transverse field $h{=}1$ with $n{=}8$, the floor gap is $0.5\,(E_1 - E_0) = 0.1845$. A conditional rule evades the floor by steering each branch with its own angles.

Proposition 2 (Classification floors) *Two standard mechanisms force the cross-entropy (*CE*) to be at least* $\ln 2$ *for any m-independent circuit.* (i) Measure-and-reset: *if the measured qubit is reset after measurement and the two branches carry label-swapped payloads, the post-measurement states are identical across branches up to the label swap, so any shared readout distribution serves both labels; the optimal shared cross-entropy is exactly* $\ln 2$. (ii) XOR feature: *if the label is $y = x_0 \oplus x_1$ with x_0 the measured bit and the payload m-independent after reset, then conditioned on the payload the label is uniform, and no m-independent readout achieves* $\mathrm{CE} < \ln 2$ *in expectation over the data distribution.*

Both floors are tight in the same sense. The conditional family contains a rule with $\mathrm{CE} = 0$ (respond on the readout qubit with the branch-calibrated sign), so the floors measure the value of the classical channel itself.

The same argument covers a wider measurement register. With 2^k orthonormal branch states and a ground space of rank $g < 2^k$, any shared (m-independent) processing unitary V obeys

$$\tfrac{1}{2^k}\sum_{m=0}^{2^k-1} \ell_m(V) \;\geq\; E_0 + \tfrac{1}{2^k}\textstyle\sum_{j=1}^{2^k-1}(E_j - E_0),$$

so the floor rises with the payload spectrum, the excitation energies of the unmeasured sites. Our Hamiltonians have g=1, so the bound holds at each branch count we test.

The floors certify necessity of *conditioning*, not of learning. They bound every shared, m-independent processor and say nothing about whether the learned rule beats the best hand rule on a given task. What learning buys is coverage, one configuration spanning tasks that would otherwise need one hand design each, which the hand-family comparison below measures. Proofs of the floors, the variance analysis, the trainability scan, and the training loop are in Appendices B–D.

4 Experiments

The grid establishes the headline comparison against every baseline class; the scaling sweep shows the ordering is not a small-register artifact; the tuned hand family and the real-data benchmark are the two strict-win claims; the measurement-form sweep and the readout-noise study test the mechanism beyond one measurement and under imperfection; ablations close the section.

All experiments run at n=8 on standard DPQC primitives. *Ground-state energy* (VQE) prepares the TFIM ground state at J=h=1, $H = -\sum_i Z_i Z_{i+1} - \sum_i X_i$, in *both* branches, at the gap to the exact energy, floored by Proposition 1. *Phase classification* classifies six TFIM ground states by the sign of the ferromagnetic order parameter; the head resets the measured qubit and passes only the classical outcome to a conditioned block whose payload is the *opposite* sign's state, so the measured bit alone carries no label information (Proposition 2(i)). *XOR classification* is a 5-bit, 32-point dataset stored in the register's amplitudes, with label $y = x_0 \oplus x_1$, x_0 measured and reset (Proposition 2(ii)).

Architectures are ring (controlled-Z (CZ) chain), brick (alternating even and odd entangler layers), and star (hub at the readout qubit), each at depth $R \in \{2, 4, 6\}$, i.e. R blocks before and after the measurement; reset-based tasks exclude Q from the pre-measurement blocks so the floors stay exact. Training uses Adam (Kingma & Ba, 2015), 400 (energy) or 500 (classification) steps, lr 0.1/0.05, exact state-vector evaluation, 3 seeds per cell, and every reported energy respects the variational bound.

The baselines fall into three classes, all sharing the ansatz and gate budget. *Information controls* remove the channel (no conditioned gates) or freeze it into m-independent angles (uncond4, the parameter-matched control; uncond8 for depth, uncond2/16 for the budget sweep). *Prior controls* fix the rule by hand, the manual $\alpha_m = \pi m$ rule, the same rule on the readout, and the random rule. *Estimator controls* train through the measurement by sampling (REINFORCE (Williams, 1992; Schulman et al., 2015)) or by a straight-through surrogate (Bengio et al., 2013). Unconditional modes get the same or larger budgets on sites spread around Q and r. Per-cell tables, the ablation sweep, the rule inventories, and the simulator self-tests are in Appendices E–K.

4.1 Performance comparison

The breadth claim needs every baseline class in one comparison, so the grid sweeps three task families, three architectures, and three depths under one protocol (Figure 2; Table 1, the same 27 cells as per-mode medians). Three readings. (1) The floors are real and exact. In both classification families every unconditional variant converges to $\ln 2 = 0.6931$ in every cell, matching Proposition 2; on ground-state energy the unconditional medians sit at 0.55–0.58 gap, 1.2–4.2$\times$ above the Ky Fan floor 0.184. (2) The learned rule closes the conditional gap. On XOR it reaches 0.000 cross-entropy in every cell, pinning median and worst case at zero; on phase classification the median drops to

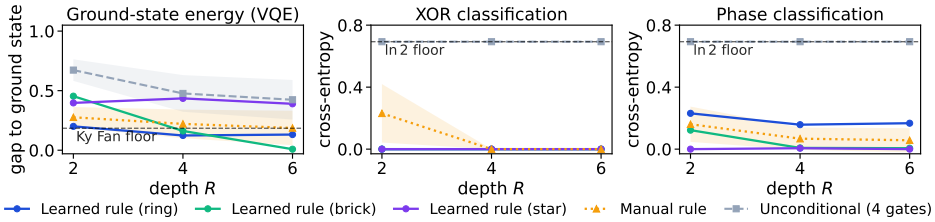


Figure 2: **Cost vs. depth across task families and architectures (learned rule).** Every panel carries the same five series: the learned rule, one solid curve per architecture (3 seeds); the manual πm rule (orange dotted); the parameter-matched unconditional control (grey dashed, band = spread); and the dashed black m-independent limit of each family.

Table 1: **Grid summary** (per family, median over its 9 cells, i.e. 3 architectures $\times$ 3 depths, each cell the mean over 3 seeds; cost in family units, depths $R \in \{2, 4, 6\}$). Lower is better; the m-independent floors are 0.184 on the ground state and $\ln 2 = 0.693$ on both classification tasks. Under the matched-restart protocol, ties declared inside 10^{-3}, the learned rule ties or beats the best baseline in $26/27$ per-cell comparisons and never trails an unconditional circuit; the only cell it does not is ring R2, where manual leads by 0.003 at the conditional optimum.

Rule variant	Ground-state (gap)	Phase (CE)	XOR (CE)
No conditioned gates	0.620	0.693	0.693
Manual rule	0.204	0.035	0.000
Manual rule (readout)	0.519	0.029	0.000
Random rule	0.577	0.233	0.236
Unconditional (4 gates)	0.552	0.693	0.693
Unconditional (8 gates)	0.579	0.693	0.693
Sampled rule (REINFORCE)	0.232	0.058	0.000
Learned rule (ours)	**0.200**	**0.008**	**0.000**

0.008 (range 0.00–0.23); on ground-state energy the median is 0.200 against 0.204 for the manual rule. It beats *every* unconditional variant in *every* cell and ties or beats the best baseline in $26/27$ cells. The count uses the matched-restart protocol with ties declared inside 10^{-3}, and the nine margins above tolerance are settled under one shared restart budget per cell (Appendix E). (3) Hand rules are priors, not solutions. The manual πm rule ties the learned rule on ground-state energy and fails on classification, where only the learned rule reaches 0.000.

Which architecture wins is family-dependent, star (hub on the readout) leading on phase classification (mean CE 0.002 against 0.185 for ring), ring most consistent on energy at shallow depth, XOR flat in all three, and every comparison above holds inside each architecture, so the choice carries none of the claim. A hand rule is a point of the learned family ($\alpha_m{=}\pi m$, $\beta_m{=}0$), so it cannot beat the learned rule at the global optimum; where it leads, the gap is optimization error rather than a limit of the family.

4.2 Scaling to larger qubit registers

The grid fixes the register at 8 qubits; is the picture an artifact of that size? We train its ground-state-energy cell (ring, $R{=}2$) at $n \in \{8, \ldots, 20\}$ with every method at its grid budget, against per-n exact diagonalization (Figure 3a). (1) The conditional methods hold a width-flat gap. Learned moves from 0.200 at $n{=}8$ to 0.243 at $n{=}20$, and manual stays within 0.006 of it at every n, so the grid tie is not an $n{=}8$ coincidence. (2) The Ky Fan floor falls from 0.185 to 0.077 and no trained method follows it down, while the m-independent family sits at 0.81–1.11 gap, $4.1\times$ the learned gap at $n{=}8$ and $4.6\times$ at $n{=}20$. Size changes the absolute gaps but not the ordering, so CPU-sized results transfer to hardware-roadmap widths.

4.3 Comparison with a tuned family of hand rules

On the standard ground-state-energy task the learned rule ties the manual πm rule, as expected when the hand prior is already optimal. This block asks whether a tuned *family* of hand rules changes that

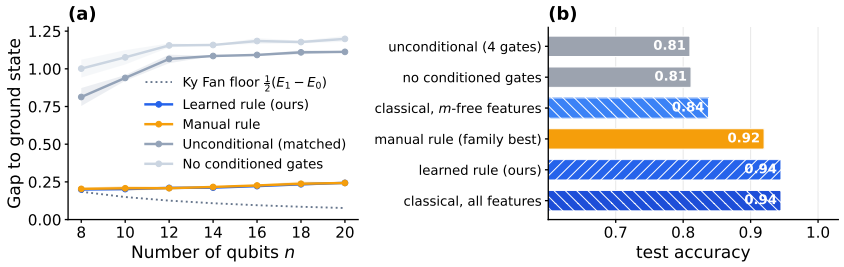


Figure 3: **Size scaling and real-data benchmark of the advantage. (a)** Gap to the exact ground state versus qubit number $n \in \{8, \ldots, 20\}$ (TFIM recovery, ring, R=2, 3 seeds, band = std). **(b)** Real-data machine-learning benchmark (MNIST 3v5 digits, principal-component (PCA) features, branch-threshold label, construction in the text). Test accuracy of the four quantum modes and two classical reference bars, on the m-free features (mean of logistic regression and an RBF-SVM, a radial-basis-function support vector machine, error bar spans the two) and on all features (0.945, full information).

Table 2: **Tuned hand-rule family vs. the learned rule** (gap to the exact ground state, mean over 3 seeds; lower is better). The family varies the branch angles over $\{0, \pi\}$ on Q and $\{0, \pm\pi/2, \pi\}$ on the readout and its neighbors, each of the 12 members trains θ_1, θ_2 for 800 steps with 3 seeds, and the best member is reported; the learned rule is one run of the same head at 400 steps and 3 seeds. Unconditional = six shared gates matched to the head; floor = the analytic bound $\frac{1}{2}(E_1(h) - E_0(h))$.

h	Learned rule	Best hand rule	Unconditional	Floor
0.5	**0.020**	0.023	1.318	0.003
1.0	**0.373**	0.428	1.212	0.184
2.0	**0.484**	0.834	1.970	1.095

picture, at transverse fields straddling the critical point h=1, against a 12-rule family and a matched unconditional control on the same head (Table 2).

The learned rule wins at every h. Three readings. (1) The manual family optimum tracks the textbook rule ($\alpha = (0, \pi)$) to within 0.005–0.05, so the learned rule beats the strongest hand prior available. (2) The learned advantage *grows* with the stretch, about 1.1× at h=0.5 and 1.72× at h=2.0, so the farther the task moves from the hand rule's design regime, the more it has to adapt. (3) At h=2.0 the learned gap falls below the unconditional floor while every unconditional variant stays above it, so the measurement channel buys energy that no shared circuit can reach. This is the paper's strict-win claim on energy. Benchmark against a tuned family, not a single rule, before declaring a hand prior sufficient.

4.4 Real-data classification through the measurement channel

The grid tasks are synthetic; this block asks whether the same separation appears for a real image statistic, where the baseline would be a classical model on the same features. The task is constructed on MNIST 3v5 with a branch-dependent threshold label rather than the digit identity, so it is a constructed task on real features, not a standard 3-vs-5 benchmark. Each dataset is reduced to eight principal components (PCA) fit on the train split and rescaled to $[0, 1]$; the feature most informative about the digit label, worth 0.175 bits of mutual information with it, is binarized at its train median and encoded on the measured qubit, so the measurement is deterministic and the reset leaves m as the feature's only surviving channel. The label is $y = [f_1 > \tau_m]$ with branch-dependent thresholds τ_0 = 70th percentile of f_1 given m=0 and τ_1 = 30th given m=1; the two differ, so no fixed rule is optimal in both branches, and the task inherits the Proposition 2(ii) floor.

Figure 3b shows the result on MNIST 3v5; the comparison axis is the measurement channel itself, channel on versus off at fixed architecture. Channel off, the two m-independent modes sit at 0.810–0.812 accuracy; channel on, the learned rule reaches 0.945 ± 0.000 and the best member of a 12-rule hand family reaches 0.920. Its trained branch angles converge to a flip rule on the readout qubit with small corrections the fixed family cannot express, worth the +2.5 points. The channel is therefore worth its budget exactly when the measured bit is informative about the label, a quantity one can

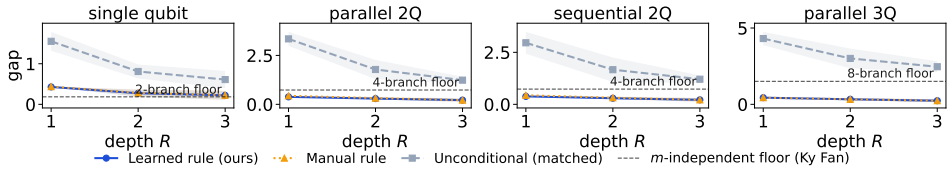


Figure 4: **The advantage generalizes across measurement forms.** Gap to the exact ground state vs. depth, every method at a matched budget (best of 6 restarts per seed, 3 seeds; band = spread across cells and architectures). Each panel is one form and the dashed line is the certified floor for its branch count; the tally is in the text.

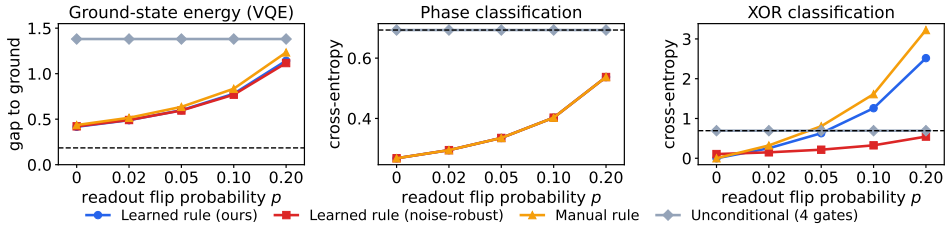


Figure 5: **Readout noise.** Cost vs. mid-measurement readout flip probability p (trained circuit frozen; 12 jitter repetitions, angle jitter of 0.1 rad moving every curve by at most 0.05 except the XOR policies). "Noise-robust" = the learned rule trained against the flipped-control mixture defined below; "unconditional" = uncond4.

estimate before training. A CIFAR-10 cat-vs-dog variant gives a bit worth only 0.006 bits and no gain from conditioning, and with matched information the classical models reach the same level once the measured feature is included (Appendix I).

4.5 Generalization to richer measurement forms

Real syndrome extraction measures registers, so this block reruns the experiment in four richer forms, moved sites (Q=3, Q=6), parallel measurement of two qubits, two sequential layers (the second rule layer reads both outcomes), and parallel measurement of three qubits (8 branches).

Every form tells the same story (Figure 4). The matched control stays at or above the certified floor of its branch count, and the learned rule stays below the control by a factor that grows with the register, from 2.8–3.6 at one measured qubit to 9.1–10.3 at three. More branches raise the floor that binds the alternative, and the conditional family passes under it. The learned gap (0.21–0.38 and 0.24–0.44) falls below the 4- and 8-branch floors (0.732 and 1.507), while at two branches the floor is only 0.184 and the gap stays above it, so there the advantage is empirical rather than certified. Under the matched budget the rule also beats the hand-designed πm rule outright in 17 of the 45 cells and ties it within 0.005 in 23. Repeating the sweep at h=2.0 makes the certified separation uniform, two-branch forms included (Appendix F).

4.6 Readout noise and deployment value

Real hardware misreads the measurement, so a rule tuned to the clean channel inherits the error; this block asks how the learned rules fare when the readout flips with probability p, and whether training repairs it.

The readout flip executes branch m with branch $1{-}m$'s angles; unconditional circuits are structurally immune but pay at p=0 already. We freeze each trained circuit and sweep $p \in [0, 0.20]$ (Figure 5). On ground-state energy the clean margin decides. Ring R1 (margin 0.96) survives the whole sweep (0.42 $\rightarrow$ 1.14, control flat at 1.38), while thin-margin ring R3 (0.61) is overtaken at p=0.02. Phase classification survives the entire sweep (0.27 $\rightarrow$ 0.54 at p=0.20, still below 0.693). On XOR the clean rule is calibrated to the noiseless channel, so flips invert the label-response correlation and the cost *exceeds* chance (2.52 at p=0.20). The noise-aware mixture

$$\mathcal{L}_p = \tfrac{1}{2} \sum_m \Big[(1{-}p)\, \ell_m\big(g_\alpha(m)\big) \;+\; p\, \ell_m\big(g_\alpha(1{-}m)\big) \Big], \tag{6}$$

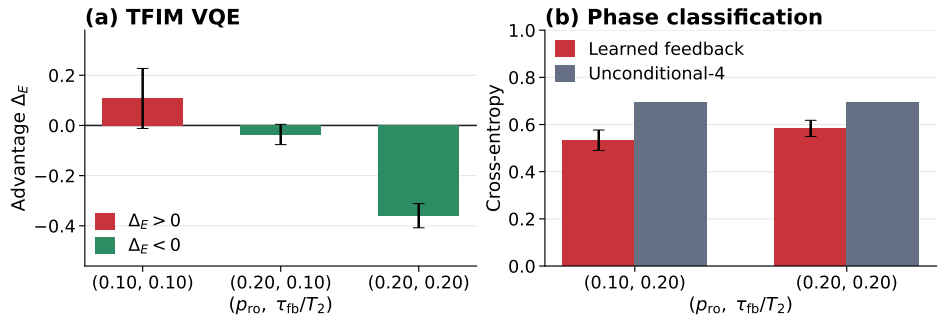


Figure 6: **Task-dependent deployment value of feedback.** Learned rule against Unconditional-4, 8-qubit ring, R=2, readout-flip training, 3 seeds $\times$ 32 trajectories, under readout error p_{ro}, feedback waiting $\exp(-\tau_{\mathrm{fb}}/T_2)$, and 0.005 depolarizing error per pulse. **(a)** TFIM VQE advantage $\Delta_E = E_{\mathrm{uncond4}} - E_{\mathrm{learned}}$, positive favoring learned feedback; **(b)** branch-averaged cross-entropy on phase classification. Error bars are seed standard deviations and the controls are defined in the text.

with ℓ_m the branch loss of the balanced branch objective, so the clean term reduces to $\mathcal{L}_{\mathrm{bal}}$ at p=0, is exactly differentiable in our branch decomposition (each branch executes the opposite branch's angles) and restores the advantage across the whole sweep (XOR settles at 0.545 at p=0.20, below floor and baselines), with every energy variant landing on the same curve to 10^{-3} and little cost elsewhere ($0.000 \to 0.105$ on XOR). Train against the mixture when the readout error rate is known, and trust a conditional rule in proportion to its clean margin.

Deployment value. The outcome-aware rule also waits for readout, classical control, and the conditional pulse, a loop a measurement-independent circuit never runs. Adding that waiting channel, with coherence $\exp(-\tau_{\mathrm{fb}}/T_2)$, to the readout flip and a 0.005 pulse error, we evaluate the readout-aware rule against matched unconditional controls on an 8-qubit ring (details in Appendix H).

Learned feedback buys 0.108 energy over the parameter-matched Unconditional-4 at $(p_{\mathrm{ro}}, \tau_{\mathrm{fb}}/T_2) = (0.10, 0.10)$, but the ordering reverses at $(0.20, 0.10)$ (-0.036) and $(0.20, 0.20)$ (-0.360; Figure 6(a)). Against the execution-matched Unconditional-2 it stays ahead at $(0.20, 0.10)$, -8.984 ± 0.021 against -8.917 ± 0.014, so a larger static budget can outweigh the waiting cost even where the information still pays. The threshold is not universal. Phase classification keeps a feedback advantage at both high-cost anchors, 0.533 ± 0.043 and 0.583 ± 0.034 against 0.693 for Unconditional-4 (Figure 6(b)), because outcome-ignoring controls are pinned at the $\ln 2$ floor there.

5 Conclusion

We asked whether the rule that maps a mid-computation discrete event's outcome to later gates can itself be learned, dynamic quantum circuits being the concrete instance. *We derived* analytic floors that turn whether feedback pays into a diagnostic read off the pre-measurement state, priceable before any training run. *We proposed* the rule inside the circuit, both branches enumerated exactly and trained by one differentiable objective that weights the branches equally, with no sampling estimator, and the rule crosses those floors. *We validated* it on six experiment blocks. At matched restarts it ties or beats the best baseline in 26 of the 27 grid cells, no unconditional variant comes close, the tuned hand family loses at every field strength, and on real data it separates 0.945 from 0.920 and holds to 20 qubits, where each trained rule reduces to a short outcome table. The guarantees travel too. At a stretched field the learned gap passes under the floor of every branch count we test, and on the minimal two-angle head it reaches the best member of a twelve-rule family without paying that family's design cost. The certificate is the piece that outlives the trained rule. Under the deployment model that price is task dependent. It exceeds the feedback cost for phase classification at the tested high-cost anchors, while TFIM recovery crosses to an unconditional preference as readout error and delay increase. Future work pairs small-site enumeration with external decoders. Calibrating reset fidelity, crosstalk, and controller latency on one device would turn this crossover into a per-device map. Because the floor moves with the payload a measurement leaves behind, the same certificate can rank what to measure as well as how to respond.

Reproducibility statement

All experiments run on an Apple M4 Pro with a state-vector simulator, one thread per training run, independent cells shardable across cores, so the full grid, the noise, deployment and ablation sweeps, the n-scaling sweep to 20 qubits, and the matched-restart form sweeps, at $h{=}1$ and at the stretched field of Appendix F, all run on one workstation. The simulator passes exact-diagonalization and finite-difference self-tests, and every reported energy respects the variational bound. Code and per-seed results accompany the submission, and a claim-to-evidence map opens Appendix A, listing for each claim the subsection that establishes it and the appendix section that verifies it.

Ethics statement

This work is foundational simulator research on training algorithms for dynamic parameterized quantum circuits. It uses no human subjects and no personal data; the only external datasets are PCA projections of MNIST and CIFAR-10 used to demonstrate the measurement channel on real features. We see no direct dual-use or fairness concerns, and we have followed the ICLR Code of Ethics in preparing this submission.

AI use statement

We used an AI assistant for language polish, for drafting parts of the plotting and verification scripts, and for navigating the related literature. The authors proposed the original idea, designed the method and the proofs, and produced and verified every reported number; the AI assistant helped refine the idea through discussion and criticism, and assisted with the writing and the code. The scientific claims remain the authors'.

References

Faisal Alam and Bryan K. Clark. Learning dynamic quantum circuits for efficient state preparation. *arXiv preprint arXiv:2410.09030*, 2024.

Yoshua Bengio, Nicholas Léonard, and Aaron Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*, 2013.

Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549:195–202, 2017.

Sergey Bravyi, Alexander McKey, and Ashley Montanaro. Obstacles to variational quantum optimization from symmetry protection. *Physical Review Letters*, 125:060505, 2020.

Matthias C. Caro, Hsin-Yuan Huang, Marco Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J. Coles. Generalization bounds for quantum (classical) convolutional and fully connected neural networks. *Nature Communications*, 13:4919, 2022.

Marco 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.

Iris Cong, Soonwon Choi, and Mikhail D. Lukin. Quantum convolutional neural networks. *Nature Physics*, 15:1273–1278, 2019.

Abhinav Deshpande, Marcel Hinsche, Khadijeh Najafi, Kunal Sharma, Ryan Sweke, and Christa Zoufal. Dynamic parameterized quantum circuits: expressive and barren-plateau free. *arXiv preprint arXiv:2411.05760*, 2024.

Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014.

Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland. Surface codes: Towards practical large-scale quantum computation. *Physical Review A*, 86:032324, 2012.

Jay M. Gambetta, Jerry M. Chow, and Matthias Steffen. Building logical qubits in a superconducting quantum computing system. *npj Quantum Information*, 3:2, 2017.

Lukas Gonon, Rodrigo Martínez-Peña, and Juan-Pablo Ortega. Feedback-driven recurrent quantum neural network universality. *arXiv preprint arXiv:2506.16332*, 2025.

Google Quantum AI. Suppressing quantum errors by scaling a surface code logical qubit. *Nature*, 614:676–681, 2023.

Gaurav Gyawali, Sonny Rappaport, Tiago Sereno, and Michael J. Lawler. Measurement-induced landscape transitions and coding barren plateaus in hybrid variational quantum circuits. *arXiv preprint arXiv:2312.09135*, 2023.

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.

Zoe Holmes, Kunal Sharma, Marco 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. Power of data in quantum machine learning. *Nature Communications*, 12:2631, 2021.

Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. In *International Conference on Learning Representations*, 2017.

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.

Youngseok Kim, Andrew Eddins, Sajant Anand, Ken Xuan Wei, Ewout van den Berg, Sami Rosenblatt, Hasan Nayfeh, Yantao Wu, Michael Zaletel, Kristan Temme, and Abhinav Kandala. Evidence for the utility of quantum computing before fault tolerance. *Nature*, 618:500–505, 2023.

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

Martin Larocca, Ningshan Ju, Ethan Noel, et al. Review of barren plateaus in variational quantum computing. *arXiv preprint arXiv:2405.00781*, 2024.

Antoine Lemelin, Christophe Pere, Olivier Landon-Cardinal, and Camille Coti. Mid-circuit measurement as an algorithmic primitive. *arXiv preprint arXiv:2506.00118*, 2025.

Manwen Liao, Yan Zhu, Weitian Zhang, and Yuxiang Yang. Learning explicit circuit representations for quantum states from local measurements. In *International Conference on Learning Representations (ICLR)*, 2025.

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

Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. *Physical Review A*, 98:032309, 2018.

Hendrik Poulsen Nautrup, Nicolas Delfosse, Vedran Dunjko, Hans J. Briegel, and Nicolai Friis. Optimizing quantum error correction codes with reinforcement learning. *Quantum*, 3:215, 2019.

Xiaotian Nie, Tao Zhang, and Linghui Chen. Experiment-compatible measurement–feedback quantum state preparation with reinforcement learning. *arXiv preprint arXiv:2606.13005*, 2026.

Adrián Pérez-Salinas, Alba Cervera-Lierta, Elies Gil-Fuster, and José I. Latorre. Data re-uploading for a universal quantum classifier. *Quantum*, 4:226, 2020.

Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5:4213, 2014.

Jader P. Santos and Raam Uzdin. Drift-resilient mid-circuit measurement and state preparation error mitigation for dynamic circuits. *arXiv preprint arXiv:2506.11270*, 2025.

Maria Schuld and Nathan Killoran. Quantum machine learning in feature hilbert spaces. *Physical Review Letters*, 122:040504, 2019.

Maria Schuld, Ville Bergholm, Christian Gogolin, Måns Johansson, Christian Kokail, Martino Milazzo, Iacopo Ciaramella, and Jens Eisert. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99:032331, 2019.

John Schulman, Nicolas Heess, Theophane Weber, and Pieter Abbeel. Gradient estimation using stochastic computation graphs. In *Advances in Neural Information Processing Systems*, 2015.

Sumeet Shirgure, Efekan Kökcü, and Siyuan Niu. Challenges in barren plateau mitigation with dynamic parameterized quantum circuits. *arXiv preprint arXiv:2606.23751*, 2026.

Sukin Sim, Peter D. Chen, and Amber T. Lee. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Nature Communications*, 10: 2358, 2019.

Adrián Skolik, Jarrod R. McClean, Masoud Mohseni, Patrick van der Smagt, and Martin Leib. Layerwise learning for quantum neural networks. *Quantum Machine Intelligence*, 3:5, 2021.

James C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37(3):332–341, 1992.

Ryan Sweke, Frederik Wilde, Johannes Jakob Meyer, Maria Schuld, Paul K. Fährmann, Barthélémy Meynard-Piganeau, and Jens Eisert. Stochastic gradient descent for hybrid quantum-classical optimization. *Quantum*, 4:314, 2020.

Barbara M. Terhal. Quantum error correction for quantum memories. *Reviews of Modern Physics*, 87:307, 2015.

Joseph Tindall, Matthew Fishman, E. Miles Stoudenmire, and Dries Sels. Efficient tensor network simulation of ibm's eagle kicked ising experiment. *PRX Quantum*, 5:010308, 2024.

Christopher Tong, Liran Shirizly, Edward H. Chen, Derek S. Wang, and Bibek Pokharel. Learning error suppression strategies for dynamic quantum circuits. *arXiv preprint arXiv:2604.18734*, 2026.

Chuanxin Wang and Yi-Zhuang You. Machine learning for ground state preparation via measurement and feedback. *arXiv preprint arXiv:2502.06517*, 2025.

Samson Wang, Enrico Fontana, Marco Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J. Coles. Noise-induced barren plateaus in variational quantum algorithms. *Nature Communications*, 12:6961, 2021.

Roeland Wiersema, Cunlu Zhou, Juan Felipe Carrasquilla, and Yong Baek Kim. Measurement-induced entanglement phase transitions in variational quantum circuits. *SciPost Physics*, 14(6): 147, 2023.

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

Christa Zoufal, Aurélien Lucchi, and Stefan Woerner. Quantum generative adversarial networks. *npj Quantum Information*, 5:103, 2019.

A WHAT THE PAPER CLAIMS, AND WHERE TO CHECK IT

Each claim in the main text, with the place it is established and the evidence a reader can check directly. Numbers are means over 3 seeds unless noted, and appendix sections are named rather than numbered so that every pointer survives reordering.

Table 3: **Claim-to-evidence map for the main text.** Each row is a claim the paper makes, the subsection that establishes it, and the evidence behind it. Appendix locations are given by section name.

Claim	Where	Evidence
The floors bound any circuit that ignores the outcome	§3.3, proofs	Propositions 1 and 2, proved in the section that follows; the classification floors $\ln 2 = 0.693$ are attained by every unconditional variant in every classification cell, and the energy floor 0.1845 bounds them in every ground-state-energy cell
The balanced branch objective adds no width-dependent suppression	§3.2, variance and width scans	gradient-variance decomposition under the stated non-cancellation condition; 200 random initializations per point up to n=16
The advantage is broad, not one lucky cell	§4.1	matched-restart protocol, ties inside 10^{-3}, 26 of 27 cells at or above the best baseline; every unconditional variant beaten in every cell
The ordering is not a small-n artifact	§4.2	$n \in \{8, \ldots, 20\}$ against per-n exact diagonalization
Hand rules lose against a tuned family and on real data	§4.3, §4.4, real-data section	$h \in \{0.5, 1, 2\}$ with the best of 12 hand rules per field; constructed MNIST 3v5 task at 0.945 against 0.920
The advantage survives richer measurement	§4.5, form tables	four measurement forms, every method at best-of-6 restarts, matched controls, a floor per branch count; at h=2.0 the learned gap passes under the floor in every form and architecture
Robustness to readout error	§4.6	readout flip probability to p=0.20, noise-aware mixture training
Feedback value under deployment cost	§4.6, deployment section	readout error to p_{ro}=0.20, feedback waiting to τ_{fb}=0.20 T_2, conditional-pulse error 0.005; crossover against the parameter-matched control on the energy task, advantage retained on phase classification at the $\ln 2$ floor
The gains are not shared budget or placement	ablation section	shared budget $2 \to 16$ gates; on-Q against off-Q placement
Trained rules are implementable	learned-rules section	each trained rule reduces to a table of (α_m, β_m)
The reported numbers are trustworthy	self-tests	exact-diagonalization and finite-difference checks on the simulator

Scope. Three boundaries, each set by where the method or the experiment stops rather than by a failure. Exact enumeration costs 2^k state vectors per measurement, so the study stops at three measured qubits and makes no claim about extensive registers, where dedicated decoders remain the right tool. Every number comes from a state-vector simulator with a readout-flip and feedback-latency channel model, not from hardware. Within those boundaries the certificate is what transfers, since the floors read a property of the task rather than of the simulator.

B WHY THE FLOORS HOLD

Ground-state-energy floor (Proposition 1). Write the post-measurement branch states as $|m\rangle_Q|\phi_m\rangle$, with $|\phi_m\rangle$ the normalized payload on the unmeasured sites. The measured qubit is left in $|m\rangle_Q$ on branch m, so the two branch states are orthonormal. Any shared unitary V maps them to $V|m\rangle_Q|\phi_m\rangle$, which stay orthonormal. Write $\ell_m(V) = \langle\phi_m|V^\dagger H V|\phi_m\rangle$ and let Π_0 be the projector onto the ground space. We assume H has a non-degenerate ground state, so Π_0 has rank one. Then at most one of the two orthonormal states can lie fully in Π_0; hence at least one branch has overlap $\le 1-\gamma$ with the ground space for some $\gamma > 0$, and the Ritz variational bound, any normalized trial state has energy at least E_0, gives $\ell \ge E_0 + \gamma(E_1' - E_0)$ where E_1' is the bottom of the orthogonal complement. Since $E_1' \ge E_1$, the balanced average satisfies $\frac{1}{2}(\ell_0 + \ell_1) \ge E_0 + \frac{1}{2}(E_1 - E_0)$. A ground space of rank $g \ge 2$ admits g orthonormal branches at once, and the bound weakens to E_0; the Hamiltonians in this paper have g=1. The conditional family evades the bound because it applies *different* unitaries V_m to the two branches, each of which can map its branch into the ground space.

Algorithm 1 Exact branch-decomposition training of the feedforward rule

input: task loss ℓ, ansatz blocks, rule parameterization $g_\alpha(m) = (\alpha_m, \beta_m)$, mode $\in$ {learned, ff-reinf, ff-straight, uncond4, uncond8, manual, none}
initialize $\theta_1, \theta_2, \alpha$ (small random); Adam optimizer, learning rate η
for $t = 1, \ldots, T$ **do**
  $|\psi\rangle \leftarrow$ apply preparation blocks to $|0^n\rangle$
  decompose $|\psi\rangle = \sum_m \sqrt{p_m}\,|m\rangle_Q|\phi_m\rangle$; normalize each $|\phi_m\rangle$
  if mode is exact-branch (learned / uncond / manual / none) **then**
    for $m = 0, 1$ **do**
      apply $g_\alpha(m)$ (or shared angles for uncond/manual/none)
      apply processing blocks; record branch loss ℓ_m
    end for
    $\mathcal{L} \leftarrow \frac{1}{2}\sum_m \ell_m$
  else if mode = ff-reinf **then**
    sample $\hat{m} \sim \mathrm{Bernoulli}(p_1)$; execute branch $\hat{m}$'s rule; $\mathcal{L} \leftarrow \ell_{\hat{m}} - (\ell_{\hat{m}} - b)\,\nabla_\alpha \log p_{\hat{m}}$ (exponential-moving-average baseline b)
  else if mode = ff-straight **then**
    hard branch choice with soft p_m-weighted surrogate gradient
  end if
  backpropagate through the (exact or surrogate) graph; Adam step
end for

Proof of the multi-branch floor. The measurement-form sweep trains circuits with 2^k branches ($k = 2, 3$), and its unconditional controls are bounded by the k-branch version of the two-branch argument. The shared unitary V preserves orthonormality, so $\{V|\phi_m\rangle\}_{m=0}^{2^k-1}$ is an orthonormal set, and the Ky Fan maximum principle on eigenvalue sums (or, equivalently, 2^k copies of the Ritz bound averaged with the Courant-Fischer minimax characterization of eigenvalues) gives $\frac{1}{2^k}\sum_m \langle\phi_m|V^\dagger H V|\phi_m\rangle \geq \frac{1}{2^k}\sum_{j=0}^{2^k-1} E_j$ for any orthonormal 2^k-tuple of trial states. The step needs a ground space of rank $g < 2^k$; a rank-g ground space can host g of the branches at once and leaves only the bound E_0. Every Hamiltonian here has g=1, so the bound holds at each branch count we test. Subtracting E_0 and dividing by 2^k puts the bound in the form stated in the main text. It is the two-branch bound of Proposition 1 with E_1 replaced by the mean of the first $2^k - 1$ excited levels; the conditional family evades it as before, since each branch has its own unitary. □

For the three-qubit parallel form, the payload spectrum of the TFIM task gives $\frac{1}{8}\sum_{j=1}^{7}(E_j - E_0) = 1.507$, the value quoted in Figure 4 and in the multi-branch table below; the two-qubit forms use the analogous 4-branch value 0.732.

Classification floor (Proposition 2). (i) Measure-and-reset with label swap: the post-reset state is the same $|\phi\rangle$ on both branches (the reset erases the only carrier of m), while the labels are $y = 0$ on branch 0 and $y = 1$ on branch 1. For any shared circuit V, the readout distribution $q(\cdot)$ is the same on both branches, so $\mathrm{CE} = \frac{1}{2}[H(y{=}0) + H(y{=}1)] = \ln 2$ at the optimum of the shared family, achieved by $q = (1/2, 1/2)$. (ii) XOR feature: conditioned on the payload (which excludes x_0), the label is $x_0 \oplus x_1$ with x_0 uniform and independent, so $P(y = 1 \,|\, \text{payload}) = 1/2$; any readout distribution produces expected cross-entropy $\geq H(y \,|\, \text{payload}) = \ln 2$ by the information inequality, with equality iff the readout ignores the payload. Both bounds are tight for the conditional family. The rule can drive the readout to the correct label on each branch, achieving $\mathrm{CE} = 0$.

C WHETHER THE EXACT OBJECTIVE ADDS A PLATEAU

The exact branch objective of Eq. equation 4 replaces the sampling estimator of measurement-based DPQC training with an analytic sum. This section records why this replacement does not reintroduce a barren-plateau mechanism, and isolates the one scaling hazard that remains. Throughout, $|\psi(\theta_1)\rangle = U_1(\theta_1)|0^n\rangle$ decomposes as $\sqrt{p_0}\,|0\rangle_Q|\phi_0\rangle + \sqrt{p_1}\,|1\rangle_Q|\phi_1\rangle$ with normalized Q-marginals, $V_m = g_\alpha(m)\,U_2(\theta_2)$ denotes the branch-m processing circuit, and gradients are over a bounded parametrization $e^{-i\theta_j P_j}$ with generators P_j of unit norm.

Decomposition of the parameter gradient. The objective is the balanced form $\mathcal{L}_{\mathrm{bal}} = \frac{1}{2}\sum_m \ell_m(\theta_2, \alpha)$ with $\ell_m = \langle\phi_m|V_m^\dagger H V_m|\phi_m\rangle$, and $\mathcal{L}$ below denotes $\mathcal{L}_{\mathrm{bal}}$. The chain rule sep-

arates the two parameter groups,

$$\frac{\partial \mathcal{L}}{\partial \theta_j} = \frac{1}{2}\sum_m \Big(\frac{\partial \ell_m}{\partial \theta_j} - (\text{normalization terms})\Big), \qquad \frac{\partial \mathcal{L}}{\partial \alpha} = \frac{1}{2}\sum_m \frac{\partial \ell_m}{\partial \alpha}, \tag{7}$$

the left identity being the payload channel (gate parameters $\theta_j \in (\theta_1, \theta_2)$, with the normalization terms carrying the θ_1-dependence of the branch normalizers) and the right identity the rule channel. Three structural facts follow. (i) Every term is an exact expectation over norm-one states; no finite-difference or likelihood estimator enters, so the estimator variance of sampled training is removed entirely, not reduced. (ii) The rule channel treats the two branches symmetrically, each branch angle carrying half the branch loss; the outcome-weighted variant multiplies the branch-m terms by p_m, an attenuation of rare branches that the trainability scan below uses as a diagnostic. (iii) Each $\partial \ell_m / \partial \theta_j$ is a gradient of an ordinary unitary circuit on the $n-1$ unmeasured sites, so its variance obeys the standard Lie-algebraic scaling (Larocca et al., 2024), $O(1)$ for local costs with local generators, and damped only by the usual $1/\mathrm{poly}(n)$ factors of the effective (post-measurement) subsystem, not by the full n-qubit second moment.

Proposition 3 (No new plateau from the measurement channel) *Let the processing ansatz $\{U_2(\theta_2)\}$ act on the $n-1$ unmeasured sites with generators from a Lie algebra $\mathfrak{g}$ of dimension $\dim \mathfrak{g}$, and let the cost be local. Write $g_m = \partial \ell_m / \partial \theta_j$, and assume the branch cross term is non-negative over the initialization, $E_\theta[g_0 g_1] \geq 0$; this excludes a cancellation regime in which the two branches anti-align through the shared parameters. Then the second moment of the exact objective gradient over uniform initializations satisfies*

$$E_\theta\left[\Big(\frac{\partial \mathcal{L}}{\partial \theta_j}\Big)^2\right] \;\geq\; \frac{1}{4}\,\max_m E_\theta\big[g_m^2\big] \;\geq\; \frac{1}{4}\,\max_m \mathrm{Var}\left[\frac{\partial \ell_m}{\partial \theta_j}\right], \tag{8}$$

and each $\mathrm{Var}[\partial \ell_m / \partial \theta_j]$ inherits the barren-plateau scaling of the $(n-1)$-site processing circuit alone. In particular, if the branch circuits are free of barren plateaus, the exact objective gradient stays bounded away from zero in second moment, with no global 2^{-n} factor beyond that circuit's own scaling; removing the sampling estimator (versus REINFORCE-style training, whose variance grows as the shot budget shrinks) strictly reduces gradient variance.

Proof. For a processing-circuit parameter θ_j the exact gradient is the branch average $\partial \mathcal{L}/\partial \theta_j = \frac{1}{2}(g_0 + g_1)$ (the measured qubit is reset to $|m\rangle_Q$, so no cross-branch interference term survives at the processing gates; the two branches differ only in their input payload and conditioned angles). Expanding the second moment over the initialization gives $E_\theta[\frac{1}{4}(g_0+g_1)^2] = \frac{1}{4}\big(E_\theta[g_0^2] + E_\theta[g_1^2]\big) + \frac{1}{2}E_\theta[g_0 g_1]$. The diagonal terms satisfy $E_\theta[g_m^2] \geq 0$ and their sum is at least $2\max_m E_\theta[g_m^2]$; with the non-negative cross term assumed in the statement, $E_\theta[(\partial \mathcal{L}/\partial \theta_j)^2] \geq \frac{1}{4}\max_m E_\theta[g_m^2]$, and $E_\theta[g_m^2] = \mathrm{Var}[g_m] + (E_\theta g_m)^2 \geq \mathrm{Var}[g_m]$ completes the displayed chain. The single-branch variance is the standard variance of a parameterized circuit with local cost and generators in $\mathfrak{g}$ (Larocca et al., 2024; McClean et al., 2018), which is bounded below by $\Omega(1/\dim \mathfrak{g})$ and, for local generators with a shallow light cone, the region a local perturbation can reach within the circuit depth, stays $\Omega(1)$ in n; no $\min_m p_m$ factor enters, a constant-factor strengthening over the outcome-weighted variant, whose bound carries $(\min_m p_m)^2$. The claim about the sampling estimator is the standard shot-noise floor. REINFORCE over measurement outcomes has variance scaling as $1/N_{\mathrm{shots}}$ on top of the exact variance above, and is bounded away from zero at finite shots. □ The proposition isolates the processing-circuit scaling from the measurement channel. Two hazards remain outside its scope, both visible in Eq. equation 7. First, if a task design drives $\min_m p_m$ below $\mathrm{poly}^{-1}(n)$, the outcome-weighted variant soft-suppresses the rule channel and training the rule becomes slow; the equal-weight form is immune by construction, at the price of training collapsed branches (next paragraph). Second, each $\partial \ell_m / \partial \theta_j$ acts on the normalized branch state, whose normalizer is an $O(1)$ observable of the full n-qubit preparation circuit, so its variance is subject to the same light-cone analysis as any local observable of U_1; our preparation layers are shallow (depth $R \leq 6$ in the main grid), keeping this channel $O(1)$ at the widths simulated. This matches the empirical picture of the trainability scan below and of the grid. The exact objective trains wherever the underlying unitary ansatz trains, and the DPQC architecture (which removes barren plateaus by mid-circuit measurement and reset (Deshpande et al., 2024)) retains that advantage under exact branch training. The non-cancellation condition enters Proposition 3 as an assumption, and it is necessary for the

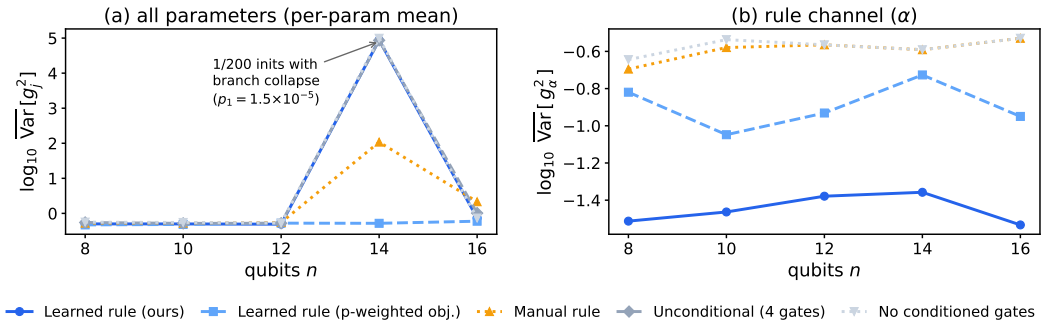


Figure 7: **Gradient-variance scan on the ground-state-energy task** (ring, R=2, 200 random initializations per point, the standard barren-plateau convention $\theta \sim U(0, \pi)$). (a) Textbook barren-plateau statistic. The per-parameter gradient variance over inits, averaged over all circuit parameters. The learned rule under the equal-weight objective is flat at ≈ -0.3 in log units from n=8 to n=16, so there is no barren-plateau decay. The curves spike at n=14 because a single initialization in 200 undergoes *branch collapse* ($p_1 = 1.5 \times 10^{-5}$). The per-branch normalization gradient scales as $1/\sqrt{p_1}$, so one collapsed branch dominates the raw variance; the manual rule, the unconditional control, and no conditioned gates inherit the same sensitivity, while the outcome-weighted variant (light blue, diagnostic) cancels the factor and stays flat. (b) Rule channel only ($\partial\mathcal{L}/\partial\alpha$, the feedforward parameters, per-parameter variance averaged over the two α's). The learned rule's variance is 0.8–1.0 orders of magnitude *below* the manual rule and the no-rule baseline at every width.

bound as stated. When the two payloads are related by a single-site flip that conjugates a generator's energy gradient to its negative, the two branch gradients anti-align exactly, the exact gradient in that parameter vanishes identically while each single-branch variance stays positive, and the displayed inequality fails.

D WHETHER TRAINABILITY SURVIVES LARGER WIDTHS

Proposition 3 predicts no n-scaling penalty from the measurement channel itself. We verify this on the ground-state-energy task with the ring architecture at R=2, measuring the standard barren-plateau diagnostic, the per-parameter variance of the analytic gradient over 200 random initializations ($\theta \sim U(0, \pi)$ per gate, the convention of McClean et al. (2018)), averaged over all parameters of the circuit, evaluated at initialization, for $n \in \{8, 10, 12, 14, 16\}$. The same scan runs the outcome-weighted variant as a diagnostic, isolating what the branch weights do to the normalization channel.

The learned rule under the equal-weight objective holds $\log_{10} \overline{\mathrm{Var}}[g_j^2] \approx -0.3$ flat across the whole sweep (Figure 7a). The gradient variance does not decay with n, matching the prediction that the decomposition adds no factor beyond the $(n-1)$-site processing circuit, whose shallow local cost and local generators keep single-branch variances $O(1)$ here. The other modes sit at the same level (between -0.25 and -0.33) at $n \in \{8, 10, 12\}$ and return to it at n=16 (between -0.13 and -0.22; the learned rule's slight elevation is the seed-spread of the same collapsed-initialization tail that dominates n=14, present but weak at n=16).

At n=14 the equal-weight statistic spikes by five orders of magnitude. This is not a plateau but the opposite. In exactly one initialization of 200 the preparation circuit produces a near-deterministic measurement outcome ($p_1 = 1.5 \times 10^{-5}$), and the per-branch normalization in the exact decomposition contributes gradients scaling as $1/\sqrt{p_1} \approx 260$. The equal-weight objective $\frac{1}{2}(\ell_0 + \ell_1)$ trains the collapsed branch even though it carries negligible sampling probability, so its gradient contains the full $1/\sqrt{p_1}$ factor. The outcome-weighted variant $\sum_m p_m \ell_m$, the expectation a sampled DPQC run estimates, cancels exactly this factor. The branch weight multiplies the normalized state, removing the spike (Figure 7a, light blue flat line). Manual, unconditional, and no-rule modes inherit the same sensitivity at collapsed branches (their gradients also act on the normalized branch states), so the spike is a property of the task family at branch-collapse inits, not of the learned rule. One collapsed initialization moves a single parameter's gradient by $\sim 10^3$, which under the averaging statistic lifts

Table 4: Matched-restart comparison for eight of the nine cells whose three-seed margin exceeds the tolerance, the ninth being ring R2 (one shared budget per cell, best-of-restarts reported, gap / CE in family units). R4*: 10 restarts instead of 6, 2× steps throughout.

Cell	learned	best other (mode)	outcome
vqe/brick/R2	0.228	0.228 (manual)	tie ($< 10^{-5}$)
vqe/brick/R4	0.014	0.013 (manual)	manual $+0.0005$
vqe/ring/R6	0.105	0.105 (manual)	learned $+0.0004$
vqe/star/R4*	0.323	0.346 (random-rule)	**learned wins**
phase/brick/R6	0.000	0.000 (ff-reinf)	tie at 0
phase/ring/R4	0.014	0.014 (manual-ro)	manual $+0.0004$
phase/ring/R6	0.011	0.011 (manual)	tie ($< 10^{-5}$)
phase/star/R4	0.000	0.000 (manual)	tie at 0

the affected modes to $10^{4.6}$–10^5 while the outcome-weighted variant, whose branch weight cancels the normalization factor, is untouched; the per-initialization $\max_j g_j^2$ statistic gives the same separation at a six-order magnitude (2.8×10^6 vs. 1.5), so the conclusion does not depend on how the per-parameter variances are aggregated. The failure is self-limiting in training, the $1/\sqrt{p_1}$ gradient pushes the preparation circuit away from the collapsed point, and no spike appears in any trained run of Sec. 4.1.

Isolating the feedforward parameters (Figure 7b), the learned rule's gradient variance sits 0.8–1.0 orders of magnitude below the manual rule and the no-rule baseline at every width, and below the outcome-weighted diagnostic by a similar margin. The separation is modest in magnitude because the rule channel is two parameters and the manual and no-rule channels share the same normalized branch states; the learned advantage is won during training, not at the initialization point.

For all widths the variational bound, no branch energy below the exact ground energy $E^\star$, holds on every branch, and the measurement-channel costs stay at the factor-2^k memory overhead discussed in Sec. 4.5.

E WHICH GRID CELLS THE LEARNED RULE WINS

The per-cell table below lists every cell of the $R \in \{1, 2, 3\}$ sweep (means over 3 seeds), and the main text quotes the $R \in \{2, 4, 6\}$ protocol. Ties are declared inside a tolerance of 10^{-3}, applied to every cell in both protocols. In this sweep the learned rule is strictly best in 3 cells and co-optimal (within 10^{-3}, sharing the better rank) in 18; in the 6 others a hand or sampled rule is ahead (four ground-state-energy cells, where manual leads by 0.01–0.23 gap or REINFORCE by 0.002, and two classification cells, where manual-ro or REINFORCE leads by $\leq$ 0.024). The learned rule beats both unconditional controls in every cell, and the variational bound (branch energies $\geq E^\star$, the exact ground energy) holds in every run.

The $R \in \{2, 4, 6\}$ protocol has 27 cells. In 18 of them the learned rule holds rank one, or a margin inside the tolerance. The other nine are settled under a matched restart budget, every method in a cell receiving the same budget and its best restart reported, 6 restarts with 2× steps throughout and 10 restarts for ring R2 and star R4. The table below collects eight of the nine; all eight land inside the tolerance, the widest margin 5×10^{-4} (brick R4, manual ahead), and star R4 becomes a strict learned win. The ninth, ring R2, is the one cell that stays with manual, ahead by 0.003 gap with both methods at the conditional optimum. Together with the 18, these eight give the $26/27$ reported for this protocol.

F WHETHER THE ADVANTAGE SURVIVES RICHER MEASUREMENT

The measurement-form sweep uses its own depth grid ($R \in \{1, 2, 3\}$, chosen to match the depth at which multi-branch floors are still trainable on CPU) and therefore keeps R=1, 3 rows in the tables below; the main grid's $R \in \{2, 4, 6\}$ protocol applies to the grid, the noise sweep, and the real-data benchmark.

Both tables report every method under one budget, best of 6 restarts per seed at the same 400 steps (mean over 3 seeds, ground-state-energy gap), so no comparison here depends on a single

Table 5: Single measured qubit at two sites (ground-state-energy gap, mean over 3 seeds; every method at best of 6 restarts). uncond = the unconditional control (uncond4). Floor 0.184 (binds the unconditional control only). Bold marks the better of the learned and manual rules in that cell; exact ties are bolded on the learned rule.

	Q=3 (learned / manual / unconditional)				Q=6 (learned / manual / unconditional)		
Cell	learn.	man.	unc.	Cell	learn.	man.	unc.
ring R1	**0.421**	0.435	1.545	**0.435**	0.435	1.239	
ring R2	0.198	**0.195**	0.772	**0.183**	0.195	0.777	
ring R3	**0.138**	0.144	0.632	**0.134**	0.140	0.584	
brick R1	**0.421**	0.435	1.929	**0.435**	0.435	1.413	
brick R2	**0.204**	0.228	0.719	0.277	**0.277**	0.723	
brick R3	0.170	**0.151**	0.298	0.142	**0.132**	0.502	
star R1	**0.421**	0.435	1.765	0.435	**0.435**	1.425	
star R2	**0.402**	0.406	0.679	**0.377**	0.381	1.184	
star R3	**0.377**	0.402	0.635	**0.367**	0.377	1.011	

Table 6: Parallel and sequential two-qubit forms and parallel three-qubit form (ground-state-energy gap, mean over 3 seeds; every method at best of 6 restarts). uncond = the unconditional control (uncond6/uncond8). Floors: 4 branches 0.732, 8 branches 1.507 (bind the unconditional control only). Bold marks the better of the learned and manual rules in that cell; exact ties are bolded on the learned rule.

	parallel-2Q			sequential-2Q			parallel-3Q		
Cell	learn.	man.	unc.	learn.	man.	unc.	learn.	man.	unc.
ring R1	**0.382**	0.435	3.010	**0.382**	0.435	2.320	0.435	**0.435**	4.318
ring R2	0.231	**0.223**	1.545	0.231	**0.223**	1.470	**0.231**	0.231	2.749
ring R3	0.137	**0.136**	1.232	0.138	**0.136**	1.162	0.159	**0.159**	2.337
brick R1	**0.382**	0.435	3.188	**0.382**	0.435	2.952	**0.435**	0.435	3.773
brick R2	**0.307**	0.383	1.338	**0.307**	0.383	1.066	0.383	**0.383**	2.381
brick R3	0.168	**0.166**	1.030	0.169	**0.166**	1.012	0.179	**0.174**	2.098
star R1	**0.382**	0.435	3.838	**0.382**	0.435	3.602	0.435	**0.435**	4.838
star R2	**0.341**	0.342	2.459	**0.341**	0.342	2.459	**0.377**	0.380	3.893
star R3	0.341	**0.341**	1.448	**0.340**	0.341	1.448	0.377	**0.377**	2.953

initialization. The first table lists the single-qubit measurement-form cells and the second the two- and three-qubit forms. The matched control is above both conditional methods in every cell. The manual πm rule is the per-bit conditional optimum on ground-state energy; under the matched budget the learned rule beats it outright in 17 of the 45 cells and ties it within 0.005 in 23, and the five losses are all within 0.019 gap.

Certified separation at a stretched field

The main-text sweep runs at h=1. Repeating all 45 cells at h=2.0 under the same budget (best of 6 restarts per seed, 400 steps, mean over 3 seeds) moves the floor statement (the figure below). The Ky Fan floors of the branch counts rise to 1.095, 2.057, and 3.385; the matched control stays above its floor in every cell, and the learned rule passes below it in 43 of the 45 cells. The two exceptions are two-branch cells in which half of the learned seeds fail to converge within the 400-step budget; taking the best depth per form and architecture, the learned gap is below the floor in all 15 groups. The two-branch caveat of the main text, where the learned gap stays above the h=1 floor of 0.184, is therefore removed at h=2.0, and the control-to-learned ratio runs from 3.6 to 103 in the multi-qubit forms.

A tuned family matched to the two-angle head

The form head carries two conditional angles, (α_0, α_1) on the measured qubit and (β_0, β_1) on the readout. Enumerating every angle-pair combination the head offers gives a 12-member family of tuned conditional hand rules, πm among them; each member fixes its angles and trains only θ_1, θ_2, exactly like the manual rule. The table below reports the comparison on two representative cells at both fields, under the same matched budget.

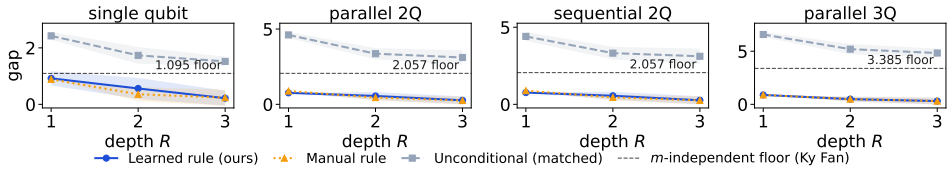


Figure 8: **The certified separation becomes uniform at a stretched field.** The form sweep of Figure 4 repeated at h=2.0 under the same matched budget. Each panel is one form, and the dashed line is the Ky Fan floor of its branch count at h=2.0 (1.095, 2.057, 3.385). The matched control sits above the floor throughout; the learned rule passes below it in every multi-qubit form and in all but two two-branch cells.

Table 7: **Tuned conditional family on the two-angle head** (ground-state-energy gap, best of 6 restarts per seed, mean over 3 seeds; lower is better). Every member fixes the conditional angles (α_0, α_1) on Q and (β_0, β_1) on the readout and trains only θ_1, θ_2, exactly like the manual rule. In every column the best member reproduces the learned rule's converged solution to four decimals. On this head the family's angle grid already covers the conditional optimum, so the parity is structural. In the last column the learned mean carries one unconverged seed; its converged seeds give 0.391, the best member's value.

Member	single-Q_3/ring, R=1		seq-2Q/brick, R=2	
$(\alpha_0, \alpha_1;\ \beta_0, \beta_1)$	h=1	h=2	h=1	h=2
$0, 0;\ 0, 0$	2.272	2.885	1.610	3.220
$0, 0;\ \pi/2, -\pi/2$	2.214	2.762	2.824	5.527
$0, 0;\ \pi/2, \pi/2$	2.513	3.885	1.360	3.125
$0, \pi;\ 0, 0\ (\pi m)$	0.436	0.885	0.383	0.516
$0, \pi;\ \pi/2, -\pi/2$	0.436	0.885	1.209	2.167
$0, \pi;\ \pi/2, \pi/2$	**0.421**	**0.762**	**0.307**	**0.391**
$\pi, 0;\ 0, 0$	0.436	0.885	0.383	0.516
$\pi, 0;\ \pi/2, -\pi/2$	0.436	0.885	1.209	2.167
$\pi, 0;\ \pi/2, \pi/2$	**0.421**	**0.762**	**0.307**	**0.391**
$\pi, \pi;\ 0, 0$	2.272	2.885	1.610	3.221
$\pi, \pi;\ \pi/2, -\pi/2$	2.214	2.762	2.825	5.528
$\pi, \pi;\ \pi/2, \pi/2$	2.513	3.885	1.360	3.125
learned	**0.421**	**0.762**	**0.307**	0.833

The comparison is bounded by the head. On the two-angle head the twelve members enumerate its conditional angle grid, and the best member equals the learned rule's converged solution to four decimals in every column, so family parity here says the learned rule finds the best member of this head, not that hand design is unnecessary in general. On the richer eight-parameter head of Sec. 4.3 fixed angle pairs cannot cover the head, and the learned rule wins by 1.1–1.7× (Table 2).

G WHETHER THE GAINS COME FROM BUDGET OR PLACEMENT

The figure below sweeps the number of sampled mid-circuit outcomes per step for the two sampled estimators against the exact objective (400 steps, 3 seeds, one representative cell per family). Exact evaluation and the straight-through surrogate give identical final costs in all three families at every budget. Even a single sample per step suffices for straight-through, because the gradient passed through the hard branch selection carries the same information as the exact probability weighting here. REINFORCE is the only estimator that pays for small budgets. On the ground-state-energy task it lands systematically above the exact cost (-9.640 vs. -9.648 energy, about 0.008) at $S \in \{1, 4, 16\}$ samples per step and closes the gap only at S=64; on both classification families it matches the exact cost at every budget. The exact objective therefore wins on variance and shot requirements, not on the optimum, and for small outcome registers there is no reason to accept estimator variance.

Placement matters (Figure 9b). Conditional gates belong next to the measured qubit for ground-state energy. On-Q placement reaches 0.20 gap where off-Q reaches 0.75 and far placement 0.79 (vqe/ring/R2), a 0.55–0.59 gap penalty. Classification is insensitive between on-Q and off-Q (0.231 both) but degrades with variance at far placement (0.60 ± 0.18 on phase/ring/R2, where the con-

Table 8: Per-cell costs of the $R \in \{1, 2, 3\}$ sweep (mean over 3 seeds). VQE (ground-state energy): gap to the exact ground state; classification: cross-entropy. For the $R \in \{2, 4, 6\}$ protocol of the main text, see Sec. 4.1 and Table 4.

Task	Cell	no cond. gates	manual	manual (readout)	random	uncond4	uncond8	REINFORCE	learned
VQE	ring R1	2.272	0.435	2.020	1.652	1.381	1.387	0.895	**0.417**
	ring R2	1.001	0.204	0.829	0.761	0.771	0.771	**0.198**	0.200
	ring R3	0.674	**0.146**	0.622	0.518	0.645	0.659	1.008	0.301
	brick R1	2.272	**0.435**	2.434	1.723	1.726	1.655	0.493	0.435
	brick R2	0.923	**0.228**	1.026	0.866	0.552	0.579	0.232	0.454
	brick R3	0.640	**0.174**	0.366	0.508	0.347	0.307	0.201	0.186
	star R1	2.272	0.435	0.955	1.449	1.619	1.980	0.762	**0.416**
	star R2	0.697	0.397	0.541	0.902	0.695	0.688	0.977	**0.397**
	star R3	0.662	0.393	0.496	0.555	0.648	0.644	0.397	**0.390**
Phase	ring R1	0.693	**0.268**	**0.268**	0.413	0.693	0.693	0.274	**0.268**
	ring R2	0.693	0.231	**0.231**	0.393	0.693	0.693	0.234	**0.231**
	ring R3	0.693	0.230	0.230	0.334	0.693	0.693	**0.206**	0.230
	brick R1	0.693	**0.268**	**0.268**	0.374	0.693	0.693	**0.268**	**0.268**
	brick R2	0.693	0.251	**0.122**	0.315	0.693	0.693	0.125	**0.122**
	brick R3	0.693	0.129	**0.127**	0.275	0.693	0.693	0.136	0.128
	star R1	0.693	0.268	0.268	0.325	0.693	0.693	0.000	**0.000**
	star R2	0.693	**0.000**	**0.000**	0.181	0.693	0.693	0.000	0.000
	star R3	0.693	0.063	0.052	0.227	0.693	0.693	0.000	**0.000**
XOR	ring R1	0.701	0.107	0.213	0.229	0.693	0.693	**0.000**	0.000
	ring R2	0.693	0.462	**0.000**	0.236	0.693	0.693	**0.000**	**0.000**
	ring R3	0.693	**0.000**	0.000	0.238	0.693	0.693	0.000	0.000
	brick R1	0.701	0.107	0.213	0.165	0.693	0.693	**0.003**	**0.003**
	brick R2	0.693	0.231	0.000	0.187	0.693	0.693	**0.000**	**0.000**
	brick R3	0.693	**0.000**	**0.000**	0.165	0.693	0.693	**0.000**	**0.000**
	star R1	0.701	0.107	0.213	0.230	0.693	0.693	**0.000**	0.000
	star R2	0.693	**0.000**	**0.000**	0.236	0.693	0.693	**0.000**	**0.000**
	star R3	0.693	0.000	0.000	0.238	0.693	0.693	**0.000**	**0.000**

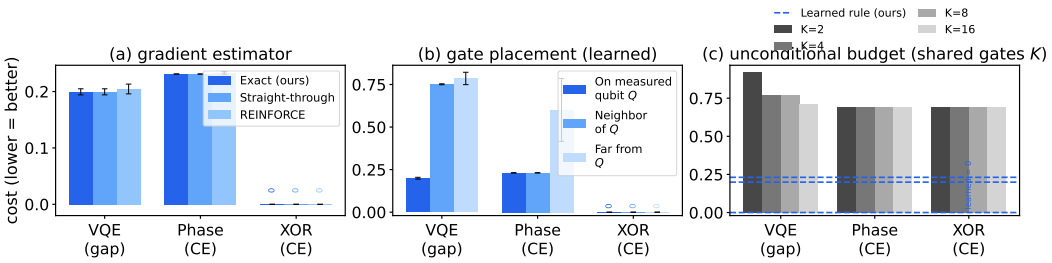


Figure 9: **Ablations** on one representative cell per family (raw costs: gap to ground state for the VQE task, cross-entropy for classification; lower is better). (a) Gradient estimator: exact backprop, straight-through, and REINFORCE land on the same cost in all three families, so the exact objective wins on variance and shot requirements, not on the converged value. (b) Gate placement: putting the rule on the measured qubit or its neighbor is what carries the advantage; moving it far away degrades the ground-state-energy cost by 0.55 gap and phase by 0.37 CE, while XOR is placement-insensitive (the reset itself carries the signal). (c) Unconditional budget: raising the shared-gate count K from 2 to 16 leaves every unconditional variant on its floor (grey), far above the learned cost (blue dashed, $= 0$ on both classification families). "Exact (ours)" denotes our differentiable branch objective; "straight-through" and "REINFORCE" denote the sampled surrogate estimators of Bengio et al. (2013) and Schulman et al. (2015).

ditional gates lose reliable access to the readout qubit). The floor perspective explains the split. Ground-state energy needs the correction to act on the branch orthogonality carrier (Q itself), classification needs any channel into the readout.

The shared budget does not close the gap (Figure 9c). Raising the unconditional budget from 2 to 16 shared parameters moves the ground-state-energy cost from -8.92 to -9.13 energy (gap $0.92 \rightarrow 0.71$), still far above the learned rule's -9.64 (gap 0.20); on both classification families the floor $\ln 2$ binds at every K. The unconditional ceiling is structural, matching Propositions 1 and 2; parameter count does not buy information.

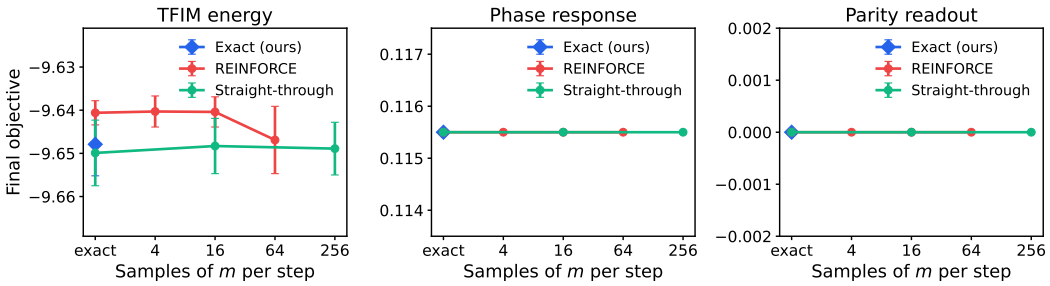


Figure 10: **Gradient-estimator budget sweep** on one representative cell per family (vqe/phase/xor, ring, R=2; final objective after 400 steps, mean $\pm$ std over 3 seeds; lower is better on the energy panel, higher is better elsewhere but all methods coincide). "exact" marks the analytic branch objective of this paper (no sampling); "REINFORCE" and "straight-through" are the sampled surrogate estimators of Schulman et al. (2015) and Bengio et al. (2013) at S sampled outcomes per step. Straight-through matches exact at every budget; REINFORCE pays about 0.008 energy at small budgets on the ground-state-energy task and catches up at S=64.

Table 9: **Deployment cost points.** Mean $\pm$ standard deviation over 3 seeds, 32 trajectories each, lower is better in both families. $\Delta = \mathcal{L}_{\text{uncond4}} - \mathcal{L}_{\text{learned}}$ is positive where learned feedback wins. Learned and Unconditional-2 execute two post-measurement pulses, Unconditional-4 four. The classification controls sit at the $\ln 2 = 0.69315$ floor with a seed spread of at most 3×10^{-4}, so their deviations are omitted.

Task	$(p_{\text{ro}}, \tau_{\text{fb}}/T_2)$	Learned	Uncond-2	Uncond-4	Δ
TFIM VQE (energy)	$(0.10, 0.10)$	-9.141 ± 0.116	-8.916 ± 0.041	-9.034 ± 0.027	$+0.108$
	$(0.20, 0.10)$	-8.984 ± 0.021	-8.917 ± 0.014	-9.020 ± 0.035	-0.036
	$(0.20, 0.20)$	-8.662 ± 0.032	-8.909 ± 0.046	-9.022 ± 0.036	-0.360
Phase (CE)	$(0.10, 0.20)$	0.533 ± 0.043	0.69315	0.69315	$+0.160$
	$(0.20, 0.20)$	0.583 ± 0.034	0.69315	0.69297	$+0.110$

The unconditional control needs no enumeration, since it is a unitary circuit. The noise-robustness pipeline trains it that way, and it matches its grid costs on the classification families exactly and lands on the same floor; on the ground-state-energy task the two routes agree to within the seed spread once the same equal-weight branch objective is used.

H THE DEPLOYMENT COST MODEL, IN FULL

The deployment study of §4.6 adds to the readout-flip mixture the two costs that only an outcome-aware rule pays. The register is the 8-qubit ring at R=2; the learned rule and both controls are trained under the flip mixture at the same cost point, 400 steps on the energy task and 500 on classification, and evaluation draws 32 stochastic trajectories for each of 3 seeds. Three channels act.

- *Readout error.* The mid-circuit outcome is misread with probability p_{ro}, and the circuit then executes the other branch's angles, exactly as in the mixture of Eq. equation 6.
- *Feedback waiting.* Between the measurement and the conditional pulse, every unmeasured qubit carries an independent Z dephasing channel of coherence $\exp(-\tau_{\text{fb}}/T_2)$. Only the learned rule waits, because its pulse depends on a classical outcome, whereas a measurement-independent circuit runs straight through. This term is what makes the comparison a test of the feedback loop rather than of noise in general.
- *Conditional-pulse error.* Each post-measurement single-qubit pulse carries depolarizing probability 0.005, charged to every method in proportion to the pulses it executes, two for the learned rule and Unconditional-2, four for Unconditional-4.

Only the waiting term is charged to the outcome-aware rule alone, so a positive Δ in Table 9 means the outcome information repaid the closed-loop latency at that cost point, against the stated control. The two controls answer different questions. Unconditional-4 is the parameter-matched control used

Table 10: Real-data benchmark (MNIST 3v5, branch-threshold label) per-mode test accuracy (mean $\pm$ std over 3 seeds; CE = branch-averaged cross-entropy on test). The manual bar reports the *best* member of a 12-rule family (below); the learned rule beats every member. Classical baselines use the same PCA-8 features; "(7f)" $= f_{1..7}$ (the m-free information set), "(8f)" = all features (full information). The m-independent modes plateau at 0.810–0.812, the accuracy attainable without the measured bit; the majority-class rate 0.565 marks the no-information level.

Rule variant	acc	CE
No conditioned gates	0.812 ± 0.002	0.360
Unconditional (4 gates)	0.810 ± 0.000	0.360
Manual rule (family best, flip)	0.920	0.294
Learned rule (ours)	$\mathbf{0.945 \pm 0.000}$	0.318
Logistic regression (7f)	0.815	n/a
RBF-SVM (7f)	0.860	n/a
Logistic regression (8f)	0.945	n/a
RBF-SVM (8f)	0.945	n/a
Majority class	0.565	n/a

throughout the paper; Unconditional-2 matches the number of post-measurement pulses the learned rule actually executes. At $(0.20, 0.10)$ the learned rule beats the second and loses to the first, which is why the main text reports both.

One caveat belongs here. This is a channel-level account of a dynamic circuit on a state-vector simulator. It prices feedback against readout error, waiting, and pulse error, and it does not model measurement-induced state degradation, reset fidelity, crosstalk, or the latency budget of the classical controller itself. The separation between the two tasks is therefore a statement about the information the outcome carries and about the floor that the task imposes on outcome-ignoring controls, not a calibration of any device.

I When the channel pays on real data

The comparison the paper makes is between information sets, channel on against channel off, and it holds on both sides. Logistic regression and an RBF-SVM on the m-free features reach 0.815–0.860, and both reach 0.945 once the measured feature is included, the level the learned rule reaches.

Labels are built as follows. PCA-8 is fit on the train split (600 points; MNIST digits $\{3, 5\}$, 200 test points) and min-max scaled by the train range. The measured feature is the column maximizing binarized mutual information $I(m; y)$ on the natural label ($I = 0.175$ bits, test 0.204); it is binarized at the train median and encoded $RY(\pi b)$ on the measured qubit, so the mid-circuit measurement is deterministic and the reset erases the feature. The remaining seven features are angle-encoded $RY(\pi f_i)$ on qubits 1..7; the pre-measurement blocks (RY + CZ path on qubits 1..7) never touch the measured qubit, and the post-measurement blocks re-entangle it (brick CZ; for the reset qubit in $|0\rangle$ the control acts as identity, so m-independent modes are unaffected). The label is $y = [f_1 > \tau_m]$ with $\tau_0 =$ the 70th percentile of f_1 over train points with m=0 and $\tau_1 =$ the 30th percentile over m=1 ($\tau_0 = 0.633$, $\tau_1 = 0.375$). The two thresholds are far enough apart that the branch-0 and branch-1 decision surfaces disagree on 37% of the f_1 range, so no single fixed rule (in particular no readout flip) is optimal in both branches. For leakage, logistic regression predicts m from $f_{1..7}$ at 0.485 accuracy (chance 0.5), so the m-channel carries no information available to the m-free baselines.

To make the comparison family-honest we scanned 12 fixed hand rules on the same head geometry as the learned rule, with $\alpha_m \in \{0, \pi\}$ per branch on the measured qubit, $\beta_m \in \{0, \pi/2, \pi, 3\pi/2\}$ per branch on the readout qubit, and cross combinations ($\alpha = (0, \pi)$ with $\beta = (0, 0), (\pi, \pi)$, and both flip directions). Each member trains θ_1, θ_2 for 600 steps with 3 seeds; the best member ($\alpha = (0, \pi)$, the readout-flip rule, $\beta = 0$) reaches 0.920/0.294 CE and is what the main text reports. The learned rule beats the best member by 2.5 points and the family mean by 8 points.

The same construction on CIFAR-10 cat-vs-dog gives a measured bit worth only 0.006 bits about the label. Learned reaches 0.830 (depth 2, 3 seeds, no spread) while the best family member is the

closed rule, no conditioned rotation, $\alpha = \beta = 0$, (0.845). The channel carries so little information that conditioning pays nothing; the mechanism needs a measured bit with real label information, and the MNIST result (measured bit worth 0.175 bits) is the informative operating point.

Training. Depth R=2, Adam lr 0.05, 600 steps, 3 seeds, same exact branch decomposition as the grid; wall time ≈ 85 s per run. Readout qubit r=1 carries the feature f_1 whose threshold defines the label, so the per-branch readout rotation can implement the branch threshold directly; the learned betas converge to $(+\pi/2, -\pi/2)$ on the readout qubit (branch-relative π flip, the coarse structure of the label convention) with a small branch-asymmetric offset that absorbs the $\tau_0 \neq \tau_1$ difference the fixed family cannot express, and the learned alphas converge to zero (the measured qubit is reset; touching it adds nothing). This matches the pattern that on reset-classification tasks the rule lives in the readout channel.

J WHAT THE TRAINED RULES LOOK LIKE

The trained rule parameters reveal a coherent structure. On ground-state-energy cells the α channel converges, in 7 of 9 cells, to the pattern $(\alpha_0, \alpha_1) \approx (0, -\pi/3)$, a signed rotation of the measured qubit conditioned on the outcome, of magnitude $\pi/3$ rather than the manual π (the processing layer absorbs the remaining correction). The β channel stays cell-dependent and small ($|\beta| \lesssim 1$), with no consistent sign across architectures. On phase-classification cells the roles swap, and the β channel (readout qubit) carries the response with $(\beta_0, \beta_1) \approx (\pm\pi/2, \mp\pi/2)$ in 6 of 9 cells (antiparallel readout rotations, the label-carrying move), while α shrinks toward 0 where the measured qubit is reset. On XOR cells the solution is heterogeneous across seeds (several (α, β) combinations achieve $\mathrm{CE} = 0$), consistent with a flat optimum in an exactly-solvable task. The full angle tables are emitted by the training scripts and accompany the code.

K WHETHER THE REPORTED NUMBERS ARE TRUSTWORTHY

The simulator is validated by (i) exact diagonalization of the TFIM Hamiltonian against sparse eigensolvers; (ii) finite-difference gradient checks on every parameter group; (iii) the variational red line $E \geq E^\star$ asserted per run; (iv) state-norm checks after every gate layer; (v) Bell-state and cross-site $Z \otimes Z$ correlation identities.