# TERMINAL-AWARE MEASUREMENT ALLOCATION FOR QUANTUM DEEP EQUILIBRIUM INFERENCE

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

## ABSTRACT

Quantum deep equilibrium models (QDEQs) construct implicit representations through repeated evaluations of a shared parameterized quantum circuit, but finite-shot measurements perturb the numerical information used by equilibrium solvers. We first characterize the measurement precision required to resolve Broyden secant differences at fixed query points, showing that sampling fluctuations can obscure shrinking residual differences near equilibrium. We then study damped Picard inference, which needs no secants, and allocate its shot budget by downstream impact. A locally linearized propagation model combining measurement covariance, readout, and state dynamics assigns each iteration a terminal impact coefficient, and calibration-averaged coefficients yield a shared schedule that is optimized offline, spends the budget exactly, and stays frozen at test time. On image classification benchmarks, the schedule reduces measurement-induced terminal squared error by 25.2%–33.9% relative to uniform allocation at equal budgets; downstream dynamics, rather than full covariance structure, account for most of the gain, and the preferred solver shifts from Picard to tuned Broyden as budgets grow. These results turn the measurement budget of quantum implicit models into an analyzable and controllable resource.

## 1 INTRODUCTION

Variational quantum algorithms (VQAs) combine parameterized quantum circuits with classical optimization, making their practical cost dependent on both circuit complexity and measurement resources (McClean et al., 2016; Wang et al., 2026). Quantum deep equilibrium models (QDEQs) (Schleich et al., 2024) build on the implicit representation paradigm of classical deep equilibrium models (DEQs) (Bai et al., 2019). In a QDEQ, a classical solver repeatedly evaluates a shared quantum mapping to compute a fixed-point representation, and gradients are then obtained by implicit differentiation at this fixed point rather than by backpropagating through every unrolled layer. This construction supports iterative computation without extending a single coherent quantum circuit and has demonstrated competitive performance against explicitly unrolled baselines (Schleich et al., 2024). Its resource efficiency, however, depends not only on the number of solver evaluations but also on the precision at which each evaluation is obtained.

Finite-shot sampling makes this precision dimension a practical bottleneck for inference. Even for an ideal circuit, a solver call estimates expectation values from finitely many measurement outcomes, introducing sampling uncertainty that decreases as more shots are acquired. Unlike a perturbation applied only to the final prediction, an intermediate measurement error changes the hidden state and hence the inputs and conditional measurement distributions of subsequent calls. The resulting inference process is a state-dependent stochastic recursion. The original QDEQ study identified forward-pass measurement overhead and shot noise as open questions (Schleich et al., 2024); addressing them requires relating the precision of individual observations to their cumulative effect on the terminal representation. This state dependence also affects solver choice. Broyden-type solvers build their secant information from differences between nearby residual evaluations (Broyden, 1965); under finite shots, these small differences are corrupted by sampling noise and lose relative accuracy (Shi et al., 2022). Consequently, reducing the number of solver evaluations does not by itself reduce the measurement cost, since the secant updates already require each residual to be estimated at high precision. Controlling computational accuracy has classical precedents in inexact Newton methods



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.

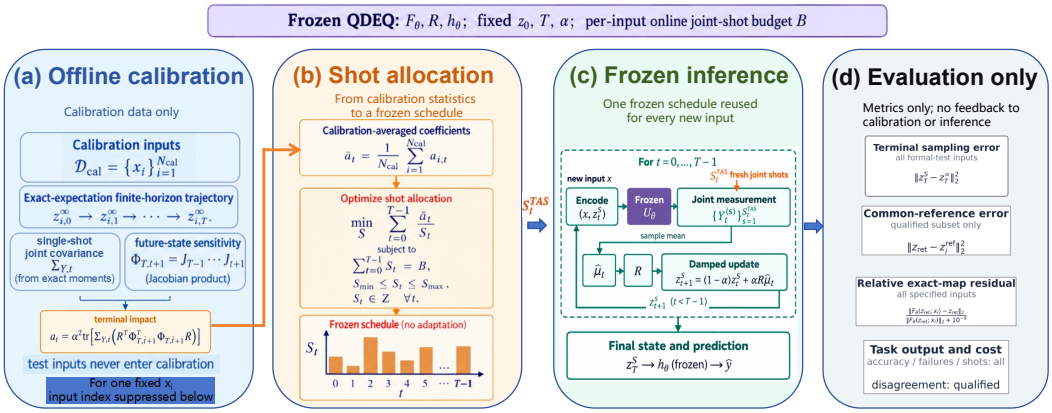



Figure 1: **Terminal-aware shot scheduling for frozen QDEQ inference.** (a) Offline calibration: exact finite-horizon trajectories provide single-shot joint covariances and future-state sensitivities, defining per-iteration terminal-impact coefficients (input index suppressed where unambiguous; $\Phi_{T:T} = I_d$). (b) Calibration-averaged coefficients yield a bounded, budget-exact integer schedule, frozen before testing. (c) Each new input reuses the schedule: iteration $t$ draws $S_t^{\text{TAS}}$ fresh joint shots, forms $\widehat{\mu}_t$, and applies the fixed readout and damped update. (d) Evaluation is post hoc and never feeds back into calibration, allocation, or inference.

and tunable-oracle optimization (Dembo et al., 1982); our focus is the finite-horizon inference of a frozen quantum operator: the allocation must account for how measurement perturbations propagate through subsequent state transitions and how those perturbations change future sampling statistics.

Figure 1 summarizes the separation between offline calibration, frozen shot allocation, finite-shot inference, and post-hoc evaluation.

To isolate the budget-allocation problem from online secant estimation, we study damped fixed-point iteration, retaining Broyden (Broyden, 1965) as a solver-level comparator. We formulate this problem as terminal-aware measurement cost control. We minimize measurement-induced terminal error under a fixed total shot budget, while keeping the horizon and damping fixed. Our central observation is that local measurement variance is insufficient to determine an evaluation's terminal impact, since subsequent updates can attenuate or amplify perturbations differently across directions. We capture these effects through terminal influence coefficients combining measurement covariance with products of state Jacobians. Under a linearized error-propagation model, this yields an exact terminal mean-squared error of the form $\sum_t a_t/S_t$ under a linear error model, together with bounds on the additional discrepancy for nonlinear inference. Together, these results turn the measurement budget of quantum equilibrium models from an implicit overhead into an analyzable and controllable resource. Our contributions are threefold:

- **Finite-shot solver analysis.** We characterize the root-mean-square measurement precision required to resolve secant differences at independently sampled fixed query points, establishing when shot noise corrupts Broyden updates near equilibrium.
- **Terminal-aware, budget-exact allocation.** For damped Picard inference, we derive a locally linearized terminal-error model that propagates joint-measurement covariance through the fixed readout and subsequent state dynamics. The resulting terminal influence coefficients expose when covariance information changes risk estimates without changing the optimal allocation, and define a bounded integer shot schedule that is computed offline, uses the prescribed budget exactly, and is frozen before testing.
- **Cross-setting empirical validation.** Across 11 frozen checkpoints spanning MNIST-4, FashionMNIST-4, and MNIST-10, the full-covariance schedule reduces measurement-induced terminal squared error by 25.2–33.9% relative to uniform allocation at equal budgets. The gains are primarily associated with downstream dynamics in these experiments: Full and Dynamics-only produce nearly identical schedules. A separate comparison under common shot-budget caps further reveals a budget-dependent Picard–Broyden ranking and a fixed-horizon Picard truncation floor.

### 1.1 Related Work

**Equilibrium models and stochastic iteration.** Stochastic approximation originates from recursive root finding under noisy observations (Robbins & Monro, 1951). Existing theoretical analyses study convergence of stochastic fixed-point iterations under contractive or nonexpansive mapping assumptions (Chen et al., 2020; Bravo & Cominetti, 2024), while stochastic equilibrium iterations can reduce the computational cost of equilibrium solving in inverse problems (Liu et al., 2022). Related equilibrium architectures impose monotonicity to obtain stable solvers and guarantees of a unique fixed point (Winston & Kolter, 2020). Our work takes a complementary perspective: we condition on a fully trained, frozen QDEQ with a fixed finite iteration horizon, and study how to allocate a prescribed total measurement budget across iterations to minimize terminal inference error.

**Quantum measurement allocation.** Adaptive shot allocation has been studied for variational quantum algorithms. iCANS (Kübler et al., 2020) distributes shots across optimization iterations and gradient components to reduce measurement cost in training. Metric-aware methods (van Straaten & Koczor, 2021) further partition samples between matrix and vector estimates required for each update. A separate line of work uses surrogate models to control observation cost. Orthogonal to measurement budgeting, error mitigation instead targets the bias of noisy expectation values and can be combined with, but does not substitute for, variance-reducing shot allocation (Temme et al., 2017). Gaussian process confidence regions (Anders et al., 2025) have been used to control observation cost for sequential minimal optimization and Bayesian parameter-shift gradients (Pedrielli et al., 2026). For observable estimation, grouping, factorization, and adaptive sampling can reduce measurement cost by exploiting compatible measurements and covariance structure (Huggins et al., 2021; Yen et al., 2023; Shlosberg et al., 2023). A recent preprint, AQKA uses downstream learner sensitivity to weight kernel entry acquisition under a shot budget (Xu et al., 2026). Adaptive-sampling and batch-size methods increase estimation precision as optimization progresses (Bollapragada et al., 2018; Smith et al., 2018). These connections motivate the validation-tuned Progressive and two-stage baselines in our formal comparison, allowing us to test whether terminal-aware coefficients add value beyond simple temporal heuristics.

## 2 Terminal-Aware Shot Scheduling

We consider the execution of a trained quantum model under a fixed measurement budget. The model supplies the map to be evaluated, whereas our method determines the precision of its successive evaluations. We first describe the representation and measurements involved in this computation, then specify the quantity used to compare shot allocations.

### 2.1 Quantum deep equilibrium models

A trained QDEQ implements a parameterized quantum map $F_\theta(z; x)$ through repeated circuit evaluations. At inference, the hidden representation is obtained by solving

$$z^\star = F_\theta(z^\star; x). \tag{1}$$

We consider the damped fixed-point iteration with hidden state $z_t \in \mathbb{R}^d$,

$$z_{t+1} = (1-\alpha)z_t + \alpha F_\theta(z_t; x), \tag{2}$$

where $0 < \alpha \leq 1$ and $T$ denotes the inference horizon. Our objective is to allocate a fixed shot budget $B$ across the $T$ quantum evaluations:

$$\mathbf{S} = (S_0, \ldots, S_{T-1}), \quad \sum_{t=0}^{T-1} S_t = B. \tag{3}$$

The detailed circuit measurement model and readout construction are provided in Appendix A.

**Finite-shot evaluation of the quantum map** Given a query pair $(z, x)$, state encoding and circuit execution prepare the quantum state $\rho_\theta(z, x)$. A computational-basis measurement returns a bitstring $b \in \{0,1\}^n$ sampled with Born probability $\langle b|\rho_\theta(z,x)|b\rangle$. Each bit yields a Pauli-$Z$ measurement outcome $Y_j = (-1)^{b_j}$ (Nielsen & Chuang, 2010). Here $Z_j$ denotes the Pauli-Z operator

acting on the j-th qubit. The expectation values assemble into the vector $\mu(z;x)$, which we map to the classical hidden dimension via a fixed linear readout:

$$\mu_j(z;x) = tr\big(\rho_\theta(z,x)Z_j\big), \qquad F_\theta(z;x) = R\mu(z;x), \quad R \in \mathbb{R}^{d\times n}. \tag{4}$$

Finite-shot measurements produce an unbiased stochastic estimate of the quantum map with conditional covariance

$$\mathbb{E}[\xi_t|\mathcal{F}_t] = 0, \qquad \mathrm{Cov}(\xi_t|\mathcal{F}_t) = \frac{C(z_t^S)}{S_t}. \tag{5}$$

The detailed joint measurement covariance and readout construction are provided in Appendix B. Even under perfect gate operations, finite-shot averaging introduces statistical fluctuations whenever the observables carry non-zero variance. At query t, we encode the current latent representation $z_t^S$ to prepare a quantum state and collect measurements from the circuit. We average $S_t$ independent joint measurement outcomes to get an estimate of the quantum map, which feeds into the damped Picard update (Kelley, 1995):

$$\widehat{\mu}_t = \frac{1}{S_t}\sum_{s=1}^{S_t} Y_t^{(s)}, \qquad z_{t+1}^S = (1-\alpha)z_t^S + \alpha R\widehat{\mu}_t. \tag{6}$$

We initialize with $z_0^S = z_0$, the all-zero vector, and run this iteration for T steps. The procedure outputs $z_T^S$, from which we predict the label $\hat{y} = \arg\max_k \big[h_\psi(z_T^S)\big]_k$. For our error analysis, define the estimation noise term $\xi_t = R\big[\hat{\mu}_t - \mu(z_t^S;x)\big]$. Let $\mathcal{F}_t$ denote the filtration capturing all information available prior to query t. The $S_t$ shots collected in query t are independent conditioned on $\mathcal{F}_t$, giving:

$$\mathbb{E}[\xi_t \mid \mathcal{F}_t] = 0, \qquad \mathrm{Cov}(\xi_t \mid \mathcal{F}_t) = \frac{C(z_t^S)}{S_t}. \tag{7}$$

Larger $S_t$ lowers the conditional variance of the estimate at iteration t. Since $z_t^S$ couples later measurement distributions to earlier measurement results, this conditional characterization does not demand unconditional independence across Picard iterations.

### 2.2 Measurement budget and terminal error

To evaluate the error caused by finite measurement shots, we compare against an ideal computation using exact quantum expectation values. Define the update map $G_\alpha(z) = (1-\alpha)z + \alpha F_\theta(z;x)$. The ideal noise-free iteration follows $z_{t+1}^\infty = G_\alpha(z_t^\infty)$, starting from $z_0^\infty = z_0$. The superscript $\infty$ denotes perfect measurement precision, not an infinite number of iterations. For integer bounds $1 \le S_{\min} \le S_{\max}$, we define the **feasible shot allocation set** and the **terminal risk** below. The feasible shot allocation set $\mathcal{S}_B$ collects all valid shot schedules under a fixed total measurement budget $B$. The terminal risk $\mathcal{R}_T(x;\mathbf{S})$ quantifies the expected squared terminal error of our noisy Picard iteration for a given input $x$ and shot schedule $\mathbf{S}$.

$$\mathcal{S}_B = \left\{ \mathbf{S} \in \mathbb{N}^T \;\middle|\; S_{\min} \le S_t \le S_{\max},\ \sum_{t=0}^{T-1} S_t = B \right\}, \qquad \mathcal{R}_T(x;\mathbf{S}) = \mathbb{E}_{\text{shot}} \left\| z_T^S - z_T^\infty \right\|_2^2. \tag{8}$$

This expectation averages over measurement sampling randomness for a fixed input x. Our goal is to find a shot allocation schedule that lowers this risk across all inputs, under the constraint $TS_{\min} \le B \le TS_{\max}$. The measurement budget B counts joint shots directly and does **not** include an extra factor of n. Directly optimizing this risk would require running many noisy quantum evaluations for every candidate shot schedule. The analysis below derives a linearized objective, which we can build during offline calibration instead. This objective isolates sampling error but does not remove the error from stopping after $T$ updates. For any fixed point selected for comparison,

$$\|z_T^S - z^*\|_2 \le \|z_T^S - z_T^\infty\|_2 + \|z_T^\infty - z^*\|_2. \tag{9}$$

The first term measures finite-shot error and the second measures truncation error. We compare schedules at fixed T and $\alpha$ against $z_T^\infty$, which isolates measurement-induced error from the finite-horizon truncation. Appendix C further distinguishes finite-horizon truncation, residual-based equilibrium accuracy, and prediction stability.

# 3 MEASUREMENT RESOLUTION AND ERROR PROPAGATION

The original QDEQ experiments used a Broyden solver (Schleich et al., 2024). Broyden's method constructs quasi-Newton updates from residual secant differences (Broyden, 1965). Under finite shots, however, the secant information that Broyden relies on becomes a measurement bottleneck, because the differences between nearby residuals shrink near equilibrium while the sampling noise stays roughly constant. Noise-aware finite-difference analysis exhibits a related trade-off between numerical resolution and evaluation noise (Shi et al., 2022), and stochastic quasi-Newton methods likewise design curvature pairs to avoid differencing pure noise in the gradients (Byrd et al., 2016; Berahas et al., 2016). This makes Broyden an awkward setting in which to study shot allocation, because the solver's own precision requirements dominate any budget trade-off. We therefore use damped Picard iteration, which requires only pointwise function evaluations as the inference loop, and ask how to distribute a fixed shot budget across its $T$ steps. This section first quantifies why secant-based updates are so shot-hungry, then traces how a single injection of shot noise propagates through Picard iterations to determine each step's contribution to the terminal error.

## 3.1 RESOLUTION OF RESIDUAL DIFFERENCES

Broyden's method builds its Jacobian approximation from differences between residual evaluations, so the relevant signal is not a single residual but the change between two queried states. Let $g(z) = F_\theta(z; x) - z$ and fix two query points $u, v$ before drawing measurement batches. The true residual difference is $\Delta g = g(v) - g(u)$, estimated as $\widehat{\Delta g} = \widehat{g}(v) - \widehat{g}(u)$. For $z \in \{u, v\}$, assume that $\widehat{g}(z)$ is the average of $S_z$independent single-shot evaluations, each of which is unbiased for $g(z)$and has covariance $C(z)$. Let $V(z) = \mathrm{tr}(C(z))$ denote the single-shot readout variance trace, and assume that the shot counts$S_u, S_v$ are fixed independently of the measurement outcomes.

**Proposition 1** (Frozen-query secant resolution). Under independent, unbiased evaluations at the two endpoints, the residual-difference estimate satisfies

$$\mathbb{E}\left\|\widehat{\Delta g} - \Delta g\right\|_2^2 = \frac{V(u)}{S_u} + \frac{V(v)}{S_v}. \tag{10}$$

The two endpoint errors are centered and independent, so their variances add. This establishes the absolute estimation error. To judge whether the difference remains useful, we compare its root-mean-square error with the magnitude of the true signal. For $\Delta g \neq 0$, equal per-endpoint counts $S_u = S_v = S$, and relative tolerance $\eta > 0$,

$$\left(\mathbb{E}\|\widehat{\Delta g} - \Delta g\|_2^2\right)^{1/2} \leq \eta\|\Delta g\|_2 \quad \Longleftrightarrow \quad S \geq \frac{V(u) + V(v)}{\eta^2\|\Delta g\|_2^2}. \tag{11}$$

Equivalently, the combined endpoint cost satisfies

$$S_u + S_v \geq \frac{2[V(u) + V(v)]}{\eta^2\|\Delta g\|_2^2}. \tag{12}$$

If the variance remains appreciable while the residual difference shrinks, maintaining this relative accuracy requires more shots. This explains why a fixed absolute query precision may be inadequate near equilibrium. The result concerns independently measured fixed endpoints. The proof, the optimal unequal-variance endpoint allocation, and the caveats for cached or adaptively selected endpoints are given in Appendix D.1.

## 3.2 PROPAGATION TO THE TERMINAL REPRESENTATION

Consider a small error introduced by query t. It first modifies the state $z_{t+1}^S$, and is then transformed through all subsequent iterative updates. Before reaching the terminal state, the error may be attenuated, rotated, or amplified. The corresponding sensitivity is given by the product of Jacobian matrices of subsequent updates. Assuming differentiability at each state point, we define along the exact solution:

$$J_t = DG_\alpha(z_t^\infty) = (1 - \alpha)I_d + \alpha DF_\theta(z_t^\infty), \quad \Phi_{T,t+1} = J_{T-1} \cdots J_{t+1}. \tag{13}$$

with $\Phi_{T,T} = I_d$. The matrix $J_t$ describes the Jacobian sensitivity of a single iteration update; $\Phi_{T,t+1}$ denotes the cumulative Jacobian product from $t+1$ up to the terminal step T. Let $e_t = z_t^S - z_t^\infty$, $C_t = C(z_t^\infty)$. Subtracting the exact iterative update from its noisy counterpart with sampling noise yields:

$$e_{t+1} = J_t e_t + \alpha \xi_t + r_t, \tag{14}$$

where these terms respectively propagate the existing error, inject the current measurement noise, and account for the bias induced by linear approximation $r_t = G_\alpha(z_t^\infty + e_t) - G_\alpha(z_t^\infty) - J_t e_t$.

**Theorem 1** (Terminal covariance propagation for finite-shot QDEQ inference). Fix input x and the shared shot schedule $\boldsymbol{S}$. Linearizing the dynamics on the reference trajectory yields the proxy recursion

$$\widetilde{e}_{t+1} = J_t \widetilde{e}_t + \alpha \widetilde{\xi}_t, \quad \widetilde{e}_0 = 0, \quad \mathrm{Cov}(\widetilde{\xi}_t \mid \widetilde{\mathcal{F}}_t) = C_t / S_t, \tag{15}$$

where $J_t$ and $C_t = C(z_t^\infty)$ are deterministic for a fixed input. Here $\widetilde{\xi}_t$ are square-integrable martingale differences with conditional covariance $C_t/S_t$, and the proxy increments satisfy $\mathbb{E}[\widetilde{\xi}_t \mid \widetilde{\mathcal{F}}_t] = 0$. Unrolling the recursion propagates each injected covariance to the terminal output:

$$\mathrm{Cov}(\widetilde{e}_T) = \sum_{t=0}^{T-1} \frac{\alpha^2}{S_t} \Phi_{T,t+1} C_t \Phi_{T,t+1}^\top. \tag{16}$$

Since $\mathbb{E}[\widetilde{e}_T] = 0$, the mean-squared error is $\mathbb{E}\|\widetilde{e}_T\|_2^2 = \mathrm{tr}\big(\mathrm{Cov}(\widetilde{e}_T)\big)$. Let $\Sigma_{Y,t} = \Sigma_Y(z_t^\infty)$ and define $H_t = R^\top \Phi_{T,t+1}^\top \Phi_{T,t+1} R$. Then

$$Q_T(x; \boldsymbol{S}) = \sum_{t=0}^{T-1} \frac{a_t(x)}{S_t}, \quad a_t(x) = \alpha^2 \mathrm{tr}(\Sigma_{Y,t} H_t) \geq 0. \tag{17}$$

We call $a_t$ the terminal impact of query $t$. It combines three components with distinct roles: joint measurement covariance describes which perturbations are generated; readout maps those perturbations into representation coordinates; and subsequent dynamics determine how much reaches the terminal state. The damping factor scales each injected perturbation by $\alpha$, hence its squared effect by $\alpha^2$. Appendix E gives the proof and propagation-indexing details of the linear identity, while Appendix E.2 analyzes nonlinear remainders and state-dependent covariance.

### 3.3 Measurement Covariance and Downstream Sensitivity

The terminal-impact coefficient $a_t$ integrates three distinct effects: local quantum measurement uncertainty, classical readout correlations, and the downstream dynamics that amplify or attenuate perturbations across the iterative trajectory.

Cross-observable covariance terms refine the prediction of terminal risk, but do not necessarily change the optimal shot schedule. A uniform positive rescaling of all temporal coefficients leaves the optimal allocation unchanged. In our audited repeated-readout setting, raw off-diagonal readout correlations exert no net influence on the final Euclidean risk. Our experiments are designed to disentangle these two separate effects: improved risk prediction fidelity and potential adjustment of the optimal allocation. The repeated-readout algebra and the precise covariance-ablation definitions are given in Appendix B.2 and Appendix E.

## 4 Terminal-Aware Shot Allocation

### 4.1 Calibration of Shared Temporal Weights

The terminal impact coefficient $a_t(x)$ derived in previous sections is input-dependent and requires future trajectory information for evaluation. It cannot be directly computed during online inference. To resolve this issue, we precompute a fixed set of temporal weights offline using a held-out calibration dataset. For each calibration sample $x_i$, we simulate the full $T$-step iterative trajectory and calculate the terminal impact $a_t(x_i)$ for every query step $t$. We then average these step-wise impacts across all calibration inputs to obtain a universal weight vector applicable to all test samples:

$$\bar{a}_t = \frac{1}{N_{\mathrm{cal}}} \sum_{i=1}^{N_{\mathrm{cal}}} a_t(x_i). \tag{18}$$

This averaged weight $\bar{a}_t$ characterizes the typical contribution of the $t$-th query to the final terminal error. A critical numerical property is that averaging per-sample impacts first and then solving the allocation problem yields a different and more desirable objective than averaging individually optimized sample allocations, since averaging does not commute with normalization operations.

The resulting linearized empirical risk for a universal allocation $\mathbf{S}$ is defined as

$$\widehat{Q}(\mathbf{S}) = \sum_{t=0}^{T-1} \frac{\bar{a}_t}{S_t}. \tag{19}$$

### 4.2 Finite-Budget Optimal Allocation

Based on the calibrated temporal weights, we formulate the shot allocation as a constrained optimization problem. We minimize the empirical linearized terminal error subject to a fixed total measurement budget and per-step shot bounds:

$$\min_{S_0,\dots,S_{T-1}} \sum_{t=0}^{T-1} \frac{\bar{a}_t}{S_t}, \quad \text{s.t.} \quad \sum_{t=0}^{T-1} S_t = B, \quad S_{\min} \le S_t \le S_{\max}. \tag{20}$$

The per-step lower bound $S_{\min} \ge 1$ ensures every iteration is measured, while $S_{\max}$ prevents excessive concentration of shots on single steps.

We first analyze the unconstrained continuous relaxation to provide intuitive insight. Without per-step bounds, the optimal continuous allocation follows a square-root rule:

$$S_t^{\text{cont}} = B \cdot \frac{\sqrt{\bar{a}_t}}{\sum_j \sqrt{\bar{a}_j}}. \tag{21}$$

For bounded integer allocations, we initialize every step at $S_{\min}$ and assign each remaining shot to the step with the largest marginal decrease $\bar{a}_t/[s(s+1)]$. Because these marginal gains are non-increasing, this rule solves the integer problem exactly—an instance of the classical marginal-allocation principle for separable, diminishing-returns objectives (Fox, 1966; Ibaraki & Katoh, 1988). Appendix F gives the proof, boundary cases, tie-breaking rule, and heap implementation, while Algorithm 1 summarizes calibration and frozen inference.

## 5 Experiments

### 5.1 Experimental setup

We reuse 11 accepted checkpoints: five MNIST-4 seeds and three each for FashionMNIST-4 and MNIST-10. Frozen calibration, validation, and formal-test indices are disjoint; each four-class checkpoint has 128 test inputs and each MNIST-10 checkpoint has 200. The allocation study fixes $T = 25$, $\alpha = 0.1$, five fresh-measurement repeats, and $B \in \{6.4, 12.8, 25.6, 51.2, 102.4\}\times 10^3$ joint shots per input. All seven Picard schedules spend exactly $B$ and are scored by $\|z_{25}^{\mathbf{S}} - z_{25}^{\infty}\|_2^2$. The solver study uses the same five budget *caps*; Broyden may stop below a cap. Reported intervals are descriptive, unadjusted Student-$t$ intervals across checkpoint means ($n = 5$ or 3). Appendices F.4 and G record the frozen method definitions, selection rules, evaluation protocol, and complete task–budget summaries.

### 5.2 Validation of terminal error propagation

For each test input and frozen schedule, we pair $Q_i = \sum_t a_{i,t}^{\text{full}}/S_t$ with its five-repeat mean terminal error. The saved $a_{i,t}^{\text{full}}$ already contains $\alpha^2$. This yields 56,840 matched input–schedule–budget conditions (284,200 rollouts). Test coefficients are used only for this offline diagnostic: they never generate a schedule, and no common-reference mask is applied.

Figure 2 shows aggregate agreement with substantial input-level scatter. Across 15 task–budget cells, measured/predicted task-mean ratios are 0.981–1.009; median per-condition $|\log_{10}(\text{measured}/\text{predicted})|$ is 0.104, 0.103, and 0.062 for MNIST-4, FashionMNIST-4, and MNIST-10, respectively (Appendix G, Table 4). This supports the proxy as an aggregate offline risk model rather than an exact per-input predictor; we next test it through fixed-budget allocation.

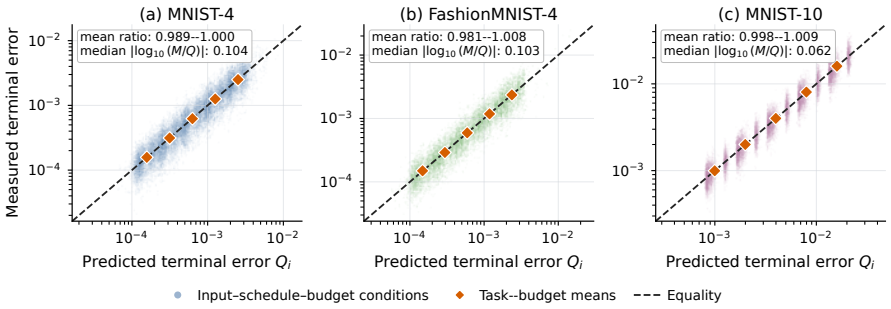



Figure 2: **Terminal propagation validation.** Faint points pair $Q_i$ with measured five-repeat terminal error for the same input, schedule, and budget; orange diamonds are task–budget means and the dashed line is equality. All 56,840 conditions are retained.

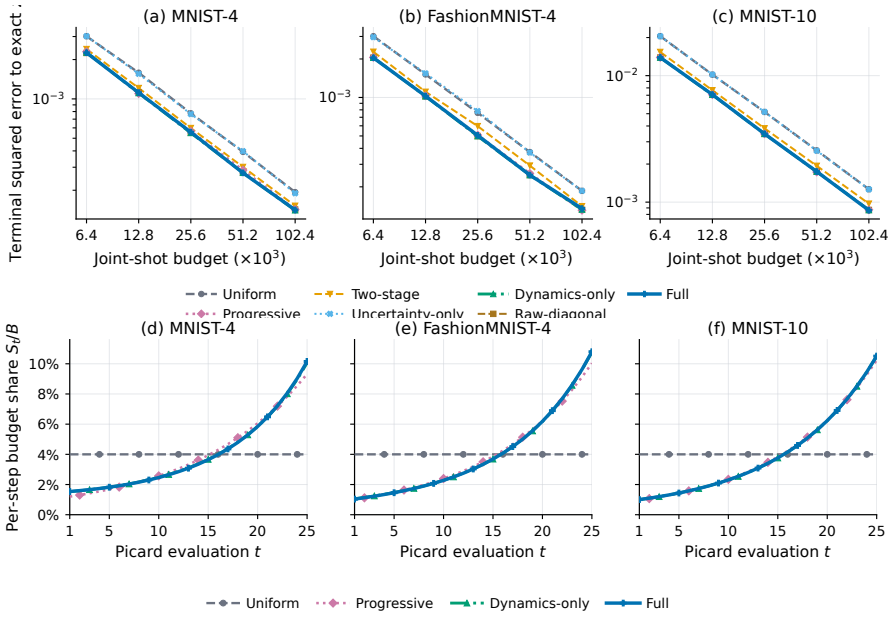



Figure 3: **Equal-budget allocation and mechanism. Top (a–c):** task means for seven schedule ablations. **Bottom (d–f):** mean $S_t/B$ at the representative frozen budget $B = 12{,}800$, averaged across checkpoints. Four schedules are retained below for legibility; Full and Dynamics-only nearly coincide.

### 5.3 Equal-budget shot allocation

Figure 3a–c compares the seven schedules as mechanism ablations. At equal budgets, Full lowers task-mean terminal error relative to Uniform in all 15 task–budget cells, by 25.2–33.9%, and paired seed-level 95% intervals for the absolute Uniform-minus-Full difference exclude zero throughout. Full does not, however, separate from Progressive or Dynamics-only: their paired contrasts generally cross zero, and their schedules nearly coincide (Figure 3d–f). The equal-budget gains thus appear to be driven by a simple temporal signal rather than the full covariance profile; Section 6.4 identifies which components are responsible.

### 5.4 Mechanism and ablation

Appendix F.4 gives the exact information used by each schedule and its frozen construction. Progressive and Two-stage compress the full tuning target through a geometric fit or two-segment selection; for Uncertainty-only, raw off-diagonals cancel under the frozen repeated readout $R$. The exact construction, validation-side selection rule, and frozen test-time use of each schedule are documented in Appendix F.4.

**Why these baselines.** These contrasts ask whether Full benefits from nonuniform timing (Uniform), propagation beyond local statistics (Uncertainty-only), measured covariance beyond fixed $I$ (Dynamics-only), raw cross-qubit covariance (Raw-diagonal), or a fully resolved profile rather than validation-tuned compressed profiles (Progressive/Two-stage). Uncertainty-only gives a median 0.04% reduction from Uniform across 15 cells (range $-3.09$–$2.04\%$); the Raw-diagonal, Dynamics-only, and Full medians are 31.39%, 31.51%, and 31.81%. Across 55 pairs, median $\|\mathbf{S}_{\mathrm{Full}} - \mathbf{S}_{\mathrm{Dyn}}\|_1/B = 0.50\%$ (maximum 1.77%; Figure 3d–f). This suggests that downstream propagation is the dominant scheduling signal here: Full shows no resolved gain over Dynamics-only, and Raw-diagonal task means are descriptively close.

### 5.5 Solver-level equilibrium inference

We compare original and tuned Broyden with four Picard variants under common shot-budget caps. The common reference qualifies 540/640, 310/384, and 442/600 designated inputs for MNIST-4, FashionMNIST-4, and MNIST-10, respectively. Reference-distance and prediction-disagreement metrics use only these frozen subsets, whereas exact-map residual, accuracy, failure, and actual-shot metrics retain all designated inputs. Full qualification details and operational metrics are reported in Appendix G, Tables 5 and 6. Means favor Full Picard at 6.4k caps and tuned Broyden at 102.4k, without a universal seed-level crossover. Exact-$T = 25$ floors are 0.01077, 0.00426, and 0.01871 for the three tasks; no finite-shot trajectory reaches the $10^{-3}$ residual threshold, so these are budget-capped approximate solves.

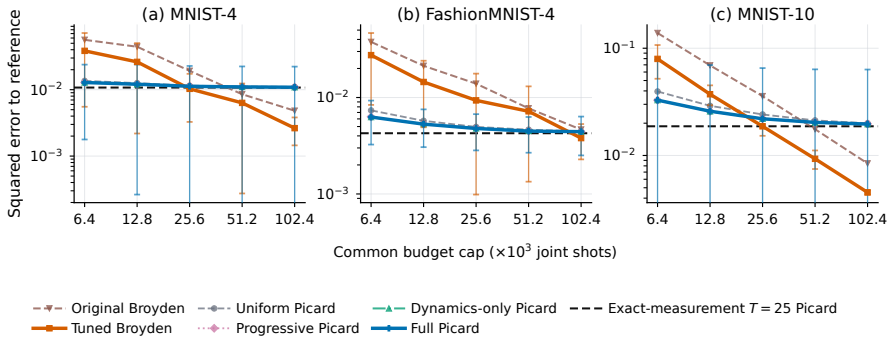



Figure 4: **Common-reference error under shared caps.** Lines show the two Broyden configurations and four Picard schedules; cross-seed intervals are highlighted for Full Picard and tuned Broyden. Dashed horizontal lines mark the exact-measurement $T = 25$ Picard floors. Broyden can use slightly fewer shots than the cap because shots are partitioned across sampled-map calls.

## 6 Conclusion

We studied how a fixed joint-shot budget should be allocated across repeated map evaluations during frozen QDEQ inference. A linear propagation proxy predicted task-level finite-shot terminal error across all 11 checkpoints, with measured-to-predicted ratios between 0.981 and 1.009, turning measurement noise into a quantity that can be analyzed and optimized before quantum execution. At equal budgets, the resulting terminal-aware schedule reduced measurement-induced terminal squared error by 25.2–33.9% relative to uniform allocation in all 15 task–budget cells. Two findings carry practical weight. First, the dominant allocation signal is downstream dynamics rather than full covariance structure: schedules using only dynamics matched the full policy within statistical resolution while differing by a median of 0.5% of the budget, so most of the achievable gain is available at a fraction of the calibration cost. Second, solver choice is budget-dependent: task means favor terminal-aware Picard at the tightest budgets and tuned Broyden at the most generous, while fixed-horizon Picard exhibits a truncation floor that no allocation can overcome. Our claims concern equal-budget terminal error under ideal finite-shot sampling in simulation; extending the analysis to hardware noise, adaptive training, and matched-quality budgeting is the natural next step. More broadly, these results reframe the measurement budget of quantum implicit models as a first-class, controllable resource rather than a fixed overhead.

### AI USE STATEMENT

In this work, we used generative AI tools for language editing and readability improvements, literature discovery, feedback on draft structure, and design inspiration for the figures. We did not use generative AI tools to generate synthetic data, develop theoretical models, formulate or prove mathematical claims, design experiments, implement methods, or interpret results. All AI-assisted text was reviewed, verified, and edited by the authors, and all suggested references were checked against the original sources. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

### REPRODUCIBILITY STATEMENT

The measurement assumptions, theoretical proofs, allocation algorithm, frozen evaluation protocol, and complete numerical summaries are provided in Appendices B–G. Algorithm 1 specifies the offline calibration and frozen-inference interfaces.

## REFERENCES

Christopher J. Anders, Kim A. Nicoli, Bingting Wu, Naima Elosegui, Samuele Pedrielli, Lena Funcke, Karl Jansen, Stefan Kühn, and Shinichi Nakajima. Adaptive observation cost control for variational quantum eigensolvers, 2025. URL https://arxiv.org/abs/2502.01704.

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. Deep equilibrium models. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/01386bd6d8e091c2ab4c7c7de644d37b-Paper.pdf.

Albert S Berahas, Jorge Nocedal, and Martin Takac. A multi-batch l-bfgs method for machine learning. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/8ebda540cbcc4d7336496819a46a1b68-Paper.pdf.

Raghu Bollapragada, Richard Byrd, and Jorge Nocedal. Adaptive sampling strategies for stochastic optimization. *SIAM Journal on Optimization*, 28(4), 2018. doi: 10.1137/17M1154679. URL https://doi.org/10.1137/17M1154679.

Mario Bravo and Roberto Cominetti. Stochastic fixed-point iterations for nonexpansive maps: Convergence and error bounds. *SIAM Journal on Control and Optimization*, 62(1):191–219, 2024. doi: 10.1137/22M1515550. URL https://doi.org/10.1137/22M1515550.

C. G. Broyden. A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation*, 19(92):577–593, 1965. doi: 10.1090/S0025-5718-1965-0198670-6.

R. H. Byrd, S. L. Hansen, Jorge Nocedal, and Y. Singer. A stochastic quasi-newton method for large-scale optimization. *SIAM Journal on Optimization*, 26(2):1008–1031, 2016. doi: 10.1137/140954362. URL https://doi.org/10.1137/140954362.

Zaiwei Chen, Siva Theja Maguluri, Sanjay Shakkottai, and Karthikeyan Shanmugam. Finite-sample analysis of contractive stochastic approximation using smooth convex envelopes. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/5d44ee6f2c3f71b73125876103c8f6c4-Abstract.html.

Ron S. Dembo, Stanley C. Eisenstat, and Trond Steihaug. Inexact newton methods. *SIAM Journal on Numerical Analysis*, 19(2):400–408, 1982. doi: 10.1137/0719025.

Bennett Fox. Discrete optimization via marginal analysis. *Manage. Sci.*, 13(3):210–216, November 1966. ISSN 0025-1909. doi: 10.1287/mnsc.13.3.210. URL https://doi.org/10.1287/mnsc.13.3.210.

William J. Huggins, Jarrod R. McClean, Nicholas C. Rubin, Zhang Jiang, Nathan Wiebe, K. Birgitta Whaley, and Ryan Babbush. Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers. *npj Quantum Information*, 7(1):23, 2021. doi: 10.1038/s41534-020-00341-7. URL https://doi.org/10.1038/s41534-020-00341-7.

Toshihide Ibaraki and Naoki Katoh. *Resource Allocation Problems: Algorithmic Approaches*. MIT Press, 1988.

C. T. Kelley. *Iterative Methods for Linear and Nonlinear Equations*, volume 16 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1995. doi: 10.1137/1.9781611970944. URL https://doi.org/10.1137/1.9781611970944.

Jonas M. Kübler, Andrew Arrasmith, Lukasz Cincio, and Patrick J. Coles. An adaptive optimizer for measurement-frugal variational algorithms. *Quantum*, 4:263, 2020. doi: 10.22331/q-2020-05-11-263.

Jiaming Liu, Xiaojian Xu, Weijie Gan, shirin shoushtari, and Ulugbek Kamilov. Online deep equilibrium learning for regularization by denoising. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 25363–25376. Curran Associates, Inc., 2022. doi: 10.52202/068431-1839. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/a2440e23f6a8c037eff1dc4f1156aa35-Paper-Conference.pdf.

Jarrod R. McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, 2016. doi: 10.1088/1367-2630/18/2/023023. URL https://doi.org/10.1088/1367-2630/18/2/023023.

Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 10th anniversary edition, 2010. doi: 10.1017/CBO9780511976667. URL https://doi.org/10.1017/CBO9780511976667.

Samuele Pedrielli, Christopher J. Anders, Lena Funcke, Karl Jansen, Kim A. Nicoli, and Shinichi Nakajima. Bayesian parameter shift rule in variational quantum eigensolvers. In *International Conference on Learning Representations (ICLR)*, 2026.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, 1951. doi: 10.1214/aoms/1177729586. URL https://doi.org/10.1214/aoms/1177729586.

Philipp Schleich, Marta Skreta, Lasse B. Kristensen, Rodrigo A. Vargas-Hernández, and Alán Aspuru-Guzik. Quantum deep equilibrium models. In *Advances in Neural Information Processing Systems*, volume 37, pp. 31940–31967. Curran Associates, Inc., 2024. doi: 10.52202/079017-1004. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/386432c7534eec9a1cd7cbeea90d7e9f-Paper-Conference.pdf.

Hao-Jun Michael Shi, Yuchen Xie, Melody Qiming Xuan, and Jorge Nocedal. Adaptive finite-difference interval estimation for noisy derivative-free optimization. *SIAM Journal on Scientific Computing*, 44(4):A2302–A2321, 2022. doi: 10.1137/21M1452470. URL https://arxiv.org/abs/2110.06380.

Ariel Shlosberg, Andrew J. Jena, Priyanka Mukhopadhyay, Jan F. Haase, Felix Leditzky, and Luca Dellantonio. Adaptive estimation of quantum observables. *Quantum*, 7:906, 2023. doi: 10.22331/q-2023-01-26-906. URL https://arxiv.org/abs/2110.15339.

Samuel L. Smith, Pieter-Jan Kindermans, Chris Ying, and Quoc V. Le. Don't decay the learning rate, increase the batch size. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=B1Yy1BxCZ.

Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. Error mitigation for short-depth quantum circuits. *Physical Review Letters*, 119(18):180509, 2017. doi: 10.1103/PhysRevLett.119.180509.

Barnaby van Straaten and Bálint Koczor. Measurement cost of metric-aware variational quantum algorithms. *PRX Quantum*, 2(3):030324, August 2021. doi: 10.1103/PRXQuantum.2.030324. URL https://doi.org/10.1103/PRXQuantum.2.030324.

Zhirao Wang, Junxiang Huang, Runyu Ye, Qingyu Li, Qi-Ming Ding, Yiming Huang, Ting Zhang, Yumeng Zeng, Jianshuo Gao, Xiao Yuan, and Yuan Yao. A review of variational quantum algorithms: Insights into fault-tolerant quantum computing, 2026. URL https://arxiv.org/abs/2604.07909.

Ezra Winston and J. Zico Kolter. Monotone operator equilibrium networks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 10718–10728, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/798d1c2813cbdf8bcdb388db0e32d496-Abstract.html.

Jian Xu, Chao Li, Delu Zeng, John Paisley, and Qibin Zhao. AQKA: Active quantum kernel acquisition under a shot budget. arXiv preprint arXiv:2605.14672, 2026. URL https://arxiv.org/abs/2605.14672.

Tzu-Ching Yen, Aadithya Ganeshram, and Artur F. Izmaylov. Deterministic improvements of quantum measurements with grouping of compatible operators, non-local transformations, and covariance estimates. *npj Quantum Information*, 9:14, 2023. doi: 10.1038/s41534-023-00683-y.

# Appendix of "Terminal-Aware Measurement Allocation for Quantum Deep Equilibrium Inference"

**Algorithm 1** Terminal-aware calibration and frozen-model inference

**Require:** Frozen map $F_\theta$, prediction head $h_\psi$, readout $R$, calibration set $\mathcal{D}_{\mathrm{cal}}$, initialization $z_0$, horizon $T$, damping $\alpha$, budget $B$, and bounds $S_{\min}, S_{\max}$

**Ensure:** Frozen integer schedule $\mathbf{S}^{\mathrm{TAS}}$ and, for a new input $x$, its final representation, scores, predicted label, and actual shot count

Check that $TS_{\min} \le B \le TS_{\max}$
$\bar{a}_t \leftarrow 0$ for every $t \in \{0, \dots, T-1\}$
**for all** $x_i \in \mathcal{D}_{\mathrm{cal}}$ **do**
    Simulate the exact trajectory $\{z_{i,t}^{\infty}\}_{t=0}^{T}$
    Compute $J_{i,t} = D_z G_\alpha(z_{i,t}^{\infty}; x_i)$ and the single-shot covariance $\Sigma_{Y,i,t}$ for all $t$
    $P \leftarrow I_d$
    **for** $t = T-1, T-2, \dots, 0$ **do**
        $\bar{a}_t \leftarrow \bar{a}_t + \dfrac{\alpha^2}{|\mathcal{D}_{\mathrm{cal}}|} \operatorname{tr}\big[\Sigma_{Y,i,t}(PR)^\top (PR)\big]$
        $P \leftarrow P J_{i,t}$ ▷ update only after scoring query $t$
    **end for**
**end for**
Initialize $S_t^{\mathrm{TAS}} \leftarrow S_{\min}$ for all $t$
**while** $\sum_t S_t^{\mathrm{TAS}} < B$ **do**
    Select an eligible query

$$t^\star \in \arg\max_{t:\, S_t^{\mathrm{TAS}} < S_{\max}} \frac{\bar{a}_t}{S_t^{\mathrm{TAS}}(S_t^{\mathrm{TAS}} + 1)}$$

    $S_{t^\star}^{\mathrm{TAS}} \leftarrow S_{t^\star}^{\mathrm{TAS}} + 1$
**end while**
Freeze $\mathbf{S}^{\mathrm{TAS}} = (S_0^{\mathrm{TAS}}, \dots, S_{T-1}^{\mathrm{TAS}})$
**Inference on a new input** $x$
$z \leftarrow z_0$
**for** $t = 0, 1, \dots, T-1$ **do**
    Prepare the circuit at $(z, x)$ and acquire $S_t^{\mathrm{TAS}}$ fresh joint outcomes
    Form their joint sample mean $\widehat{\mu}_t$
    $z \leftarrow (1-\alpha) z + \alpha R \widehat{\mu}_t$
**end for**
**return** $z$, $h_\psi(z)$, $\arg\max_k [h_\psi(z)]_k$, and $\sum_t S_t^{\mathrm{TAS}} = B$

# A QDEQ Definition and Measurement Model

For input $x \in \mathcal{X}$, let $z \in \mathbb{R}^d$ be the classical hidden representation. A trained parameterized quantum circuit with input encoding and readout implements an exact-expectation map. A classification head $h_\psi$ maps the representation to K class scores. Following the DEQ formulation and its quantum counterpart (Bai et al., 2019; Schleich et al., 2024), the QDEQ representation and prediction are defined as

$$z^\star = F_\theta(z^\star; x), \qquad \widehat{y}^\star(x) = \arg\max_{1 \le k \le K} [h_\psi(z^\star(x))]_k. \tag{22}$$

Unlike explicit unrolled quantum networks, QDEQ reuses a shared quantum mapping during iterative inference. We focus on the measurement precision required by this fixed-point computation. At inference, we solve the fixed-point equation $z^\star = F_\theta(z^\star; x)$ by numerical root finding. The classical state $z_t \in \mathbb{R}^d$ is updated iteratively, and the predicted class $\hat{y}^*(x)$ is obtained from the fixed point $z^*(x)$ via the classification head $h_\psi$. The inference procedure is parameterized by an initial state $z_0$, iteration horizon T, and damping coefficient $0 < \alpha \le 1$. We learn a shared shot schedule $\mathbf{S} = (S_0, \ldots, S_{T-1})$, where $S_t$ denotes the number of measurement shots allocated at iteration t. This schedule is built using a held-out calibration set $\mathcal{D}_{\text{cal}} = \{x_i\}_{i=1}^{N_{\text{cal}}}$, subject to a fixed total shot budget B per test sample.

# B Measurement Assumptions and Readout Construction

Unless otherwise stated, all expectations condition on one input, one frozen model, and one schedule fixed before test inference. Exact and sampled trajectories use the same deterministic initialization. Shots within a query are conditionally independent and identically distributed, and each new query uses fresh samples from the circuit evaluated at the current state. The readout is a fixed deterministic linear map.

## B.1 Conditional moments and temporal orthogonality

One shot comprises state preparation, circuit run, and full readout over all n qubits, producing a single joint outcome vector rather than n independent, separately budgeted measurements. The linear readout may expand this vector dimension without introducing additional quantum information. For the four-qubit configuration, each expectation is repeated four times, yielding $d = 16$ and $R = I_4 \otimes \mathbf{1}_4$.

The entries of this joint measurement vector are generally correlated. To capture these dependencies, define $\Sigma_Y(z)$ as the covariance matrix for a single-shot outcome vector, and $C(z)$ as its covariance after applying the readout map. Omitting the fixed input x for compact notation,

$$\big[\Sigma_Y(z)\big]_{jk} = \langle Z_j Z_k \rangle_z - \mu_j(z)\mu_k(z), \quad C(z) = R\Sigma_Y(z)R^\top. \tag{23}$$

These covariances describe statistics of a single circuit run, prior to averaging across repeated shots. Off-diagonal terms encode correlations between jointly measured observables (Yen et al., 2023).

Let $\mathcal{F}_t$ contain all randomness revealed before query $t$. Conditioned on $\mathcal{F}_t$, the current state $z_t^S$ and its Born distribution are fixed. Consequently,

$$\mathbb{E}[\widehat{\mu}_t \mid \mathcal{F}_t] = \mu(z_t^S; x), \qquad \mathrm{Cov}(\widehat{\mu}_t \mid \mathcal{F}_t) = \frac{\Sigma_Y(z_t^S; x)}{S_t}. \tag{24}$$

For $\xi_t = R[\widehat{\mu}_t - \mu(z_t^S; x)]$,

$$\mathbb{E}[\xi_t \mid \mathcal{F}_t] = 0, \qquad \mathrm{Cov}(\xi_t \mid \mathcal{F}_t) = \frac{C(z_t^S; x)}{S_t}, \quad C(z; x) = R\Sigma_Y(z; x)R^\top. \tag{25}$$

For $s < t$, $\xi_s$ is measurable with respect to $\mathcal{F}_t$, and hence

$$\mathbb{E}[\xi_t \xi_s^\top] = \mathbb{E}\big[\mathbb{E}[\xi_t \mid \mathcal{F}_t]\xi_s^\top\big] = 0. \tag{26}$$

This is temporal second-moment orthogonality. It does not assert statistical independence of complete noisy queries or trajectories.

For $M \geq 2$ independent raw joint outcomes at one fixed query, the unbiased estimator of the *single-shot* covariance is

$$\widehat{\Sigma}_Y = \frac{1}{M-1} \sum_{s=1}^{M} (Y^{(s)} - \overline{Y})(Y^{(s)} - \overline{Y})^\top. \tag{27}$$

The covariance of an $S$-shot sample mean is instead $\Sigma_Y / S$. These two quantities should not be conflated. Estimating the local covariance from finite data also does not supply the future Jacobians needed for terminal propagation.

### B.2 Repeated readout does not create independent measurements

For a square matrix $A$, define $\mathcal{D}(A) = \mathrm{Diag}(\mathrm{diag}\, A)$. In the four-qubit models, each jointly measured Pauli-$Z$ expectation is repeated four times by the fixed classical readout:

$$R = I_4 \otimes \mathbf{1}_4 \in \mathbb{R}^{16 \times 4}, \qquad R^\top R = 4 I_4, \qquad \mathrm{rank}(R) = 4. \tag{28}$$

Even when the raw single-shot covariance is diagonal,

$$R\, \mathrm{Diag}(v_1, v_2, v_3, v_4) R^\top = \mathrm{blockdiag}\big(v_1 \mathbf{1}_4 \mathbf{1}_4^\top, \ldots, v_4 \mathbf{1}_4 \mathbf{1}_4^\top\big), \tag{29}$$

which contains nonzero off-diagonal entries within each repeated block. Therefore, $\mathcal{D}(R \Sigma_Y R^\top)$ removes correlations introduced by deterministic duplication as well as correlations between distinct raw outcomes.

To remove only raw cross-qubit covariance while retaining the actual deterministic readout, our raw-diagonal ablation uses

$$C^{\mathrm{rawdiag}} = R \mathcal{D}(\Sigma_Y) R^\top. \tag{30}$$

This replacement is used only when constructing offline coefficients. The finite-shot rollouts continue to draw fresh joint bitstrings from the original sampling process.

For MNIST-10, the frozen readout is

$$R = I_{10} \otimes \mathbf{1}_{10} \in \mathbb{R}^{100 \times 10}, \qquad R^\top R = 10 I_{10}, \qquad \mathrm{rank}(R) = 10. \tag{31}$$

Both readout interfaces were checked from the frozen model configurations rather than inferred from hidden-state dimension alone.

## C Finite-Horizon Error and Equilibrium Metrics

Let $z^\star$ denote a selected fixed point, let $e_T = z_T^S - z_T^\infty$, and let $b_T = z_T^\infty - z^\star$. The exact decomposition is

$$\mathbb{E}\|z_T^S - z^\star\|_2^2 = \mathbb{E}\|e_T\|_2^2 + \|b_T\|_2^2 + 2\langle \mathbb{E}[e_T], b_T \rangle. \tag{32}$$

The cross term need not vanish for nonlinear dynamics. The always-valid root-mean-square bound is

$$\big(\mathbb{E}\|z_T^S - z^\star\|_2^2\big)^{1/2} \leq \big(\mathbb{E}\|z_T^S - z_T^\infty\|_2^2\big)^{1/2} + \|z_T^\infty - z^\star\|_2. \tag{33}$$

Thus finite-shot error and finite-horizon truncation are distinct.

If $G_\alpha$ is a contraction with factor $q < 1$ on a complete invariant domain containing the trajectory and fixed point, then

$$\|z_T^\infty - z^\star\|_2 \leq q^T \|z_0 - z^\star\|_2. \tag{34}$$

A sufficient condition is that $F_\theta$ is $L_F$-Lipschitz with $L_F < 1$ on the domain, in which case $G_\alpha = (1-\alpha) I + \alpha F_\theta$ is Lipschitz with factor at most $1 - \alpha + \alpha L_F$. These sufficient conditions need not hold globally for a trained QDEQ, so Eq. (34) is not assumed in the formal empirical comparisons.

If $F_\theta$ is $L_F$-Lipschitz with $L_F < 1$ on a domain containing $z$ and $z^\star$, then

$$\|z - z^\star\|_2 \leq \frac{\|F_\theta(z; x) - z\|_2}{1 - L_F}. \tag{35}$$

Without such a contraction condition, a small exact-map residual is not a certified distance to a fixed point. This motivates reporting common-reference distance, exact-map residual, classification accuracy, and prediction disagreement as separate quantities.

## D Proof and Scope of Frozen-Query Secant Resolution

**Proposition D.1.** *Proposition 1 exposes a structural cost of secant-based solvers under sampled evaluations. Every Broyden update is built from a residual difference $\Delta g$, and resolving that difference at fixed relative accuracy costs $\Theta(\eta^{-2}\|\Delta g\|_2^{-2})$ shots. The signal $\|\Delta g\|_2$ vanishes at the fixed point while the per-shot variance does not, so each Broyden step becomes more expensive precisely as the iteration converges: for smooth $g$ with a nonsingular Jacobian at $z_\star$, $\|\Delta g\|_2 = \Theta(\|v-u\|_2)$, and the per-update measurement budget grows as $\Theta(\|v-u\|_2^{-2})$. Iteration counts from classical convergence analysis therefore understate the true sampling cost of secant acceleration, and any fixed absolute query precision is provably inadequate near equilibrium. Damped Picard avoids this particular precision requirement because it uses a sampled map value directly. The replacement does not eliminate sampling error or guarantee a lower total cost. It shifts the allocation problem to controlling the errors introduced by successive map estimates. Their importance depends on the updates remaining before the output is returned. Let $g(z) = F_\theta(z;x) - z$ and fix query points $u, v$ before drawing measurement batches. Define*

$$\Delta g = g(v) - g(u), \qquad \widehat{\Delta g} = \widehat{g}(v) - \widehat{g}(u).$$

*Write $\widehat{g}(u) = g(u) + \varepsilon_u$ and $\widehat{g}(v) = g(v) + \varepsilon_v$. At fixed endpoints, subtracting the known states in $g = F_\theta - I$ adds no randomness. Under independent endpoint batches,*

$$\mathbb{E}[\varepsilon_u] = \mathbb{E}[\varepsilon_v] = 0, \qquad \mathrm{Cov}(\varepsilon_u) = \frac{C(u)}{S_u}, \qquad \mathrm{Cov}(\varepsilon_v) = \frac{C(v)}{S_v}.$$

*Therefore,*

$$\mathbb{E}\|\widehat{\Delta g} - \Delta g\|_2^2 = \mathbb{E}\|\varepsilon_v - \varepsilon_u\|_2^2 \tag{36}$$

$$= \mathrm{tr}\left(\frac{C(v)}{S_v}\right) + \mathrm{tr}\left(\frac{C(u)}{S_u}\right) \tag{37}$$

$$= \frac{V(v)}{S_v} + \frac{V(u)}{S_u}, \tag{38}$$

*where $V(z) = \mathrm{tr}\, C(z)$.*

*For equal per-endpoint counts $S_u = S_v = S$ and $\Delta g \neq 0$,*

$$\left(\mathbb{E}\|\widehat{\Delta g} - \Delta g\|_2^2\right)^{1/2} \leq \eta\|\Delta g\|_2 \quad \Longleftrightarrow \quad S \geq \frac{V(u)+V(v)}{\eta^2\|\Delta g\|_2^2}. \tag{39}$$

*Hence the corresponding* total *endpoint shot count satisfies*

$$S_u + S_v = 2S \geq \frac{2[V(u)+V(v)]}{\eta^2\|\Delta g\|_2^2}. \tag{40}$$

*The factor of two in Eq.* (40) *is required whenever the reported resource counts the combined shots at both endpoints.*

*For completeness, unequal endpoint allocation can reduce total cost when the two endpoint variances differ. Let $R_\eta = \eta^2\|\Delta g\|_2^2$. Ignoring integrality, minimizing $S_u + S_v$ subject to $V(u)/S_u + V(v)/S_v \leq R_\eta$ gives*

$$S_u^\star = \frac{\sqrt{V(u)}[\sqrt{V(u)} + \sqrt{V(v)}]}{R_\eta}, \tag{41}$$

$$S_v^\star = \frac{\sqrt{V(v)}[\sqrt{V(u)} + \sqrt{V(v)}]}{R_\eta}, \tag{42}$$

$$S_u^\star + S_v^\star = \frac{[\sqrt{V(u)} + \sqrt{V(v)}]^2}{R_\eta}. \tag{43}$$

*Positive integer counts must be rounded or solved using the corresponding discrete allocation.*

*The identity is an RMS statement, not a tail guarantee. Markov's inequality gives*

$$\Pr\left(\|\widehat{\Delta g} - \Delta g\|_2 \geq \eta\|\Delta g\|_2\right) \leq \frac{V(u)/S_u + V(v)/S_v}{\eta^2\|\Delta g\|_2^2}. \tag{44}$$

*Cached estimates, correlated endpoint batches, or adaptively selected query points require retaining the corresponding cross-covariance terms. The frozen-query result does not by itself analyze an entire adaptive Broyden trajectory.*

# E Terminal Propagation and Proxy Discrepancy

## E.1 The fixed-coefficient linear proxy

Consider

$$\widetilde{e}_{t+1} = J_t \widetilde{e}_t + \alpha \widetilde{\xi}_t, \qquad \widetilde{e}_0 = 0, \qquad \mathrm{Cov}(\widetilde{\xi}_t \mid \widetilde{\mathcal{F}}_t) = \frac{C_t}{S_t}. \tag{45}$$

Define

$$\Phi_{T,t+1} = J_{T-1} \cdots J_{t+1}, \qquad \Phi_{T,T} = I.$$

Unrolling the recursion gives

$$\widetilde{e}_T = \alpha \sum_{t=0}^{T-1} \Phi_{T,t+1} \widetilde{\xi}_t. \tag{46}$$

The martingale-difference property and Eq. (26) eliminate all cross-time covariance terms. Therefore,

$$\mathrm{Cov}(\widetilde{e}_T) = \alpha^2 \sum_{t=0}^{T-1} \Phi_{T,t+1} \frac{C_t}{S_t} \Phi_{T,t+1}^\top. \tag{47}$$

Since $\mathbb{E}[\widetilde{e}_T] = 0$,

$$Q_T(x; S) = \mathbb{E}\|\widetilde{e}_T\|_2^2 \tag{48}$$

$$= \sum_{t=0}^{T-1} \frac{a_t(x)}{S_t}, \qquad a_t(x) = \alpha^2 \operatorname{tr}(\Sigma_{Y,t} H_t)\,, \tag{49}$$

$$H_t = R^\top \Phi_{T,t+1}^\top \Phi_{T,t+1} R.$$

Both $\Sigma_{Y,t}$ and $H_t$ are positive semidefinite, so $a_t(x) \geq 0$.

The indexing follows the order in which measurement error enters the state. Noise from query $t$ enters state $t+1$ and is subsequently propagated by $J_{t+1}, \ldots, J_{T-1}$. For $T = 3$, the three propagation products are $J_2 J_1$, $J_2$, and $I$.

## E.2 Nonlinear remainder and random covariance

For the actual nonlinear recursion, define $e_t = z_t^S - z_t^\infty$ and

$$e_{t+1} = J_t e_t + \alpha \xi_t + r_t, \qquad r_t = G_\alpha(z_t^\infty + e_t) - G_\alpha(z_t^\infty) - J_t e_t. \tag{50}$$

Let

$$Q_T^{\mathrm{act}} = \sum_{t=0}^{T-1} \frac{\alpha^2}{S_t} \operatorname{tr}\left[\Phi_{T,t+1} \mathbb{E} C(z_t^S) \Phi_{T,t+1}^\top\right], \tag{51}$$

$$D_T = \sum_{t=0}^{T-1} \|\Phi_{T,t+1}\|_2 \left(\mathbb{E}\|r_t\|_2^2\right)^{1/2}. \tag{52}$$

Then

$$|\mathcal{R}_T(x; S) - Q_T(x; S)| \leq \left|Q_T^{\mathrm{act}} - Q_T(x; S)\right| + 2\sqrt{Q_T^{\mathrm{act}}}\, D_T + D_T^2. \tag{53}$$

To prove the bound, unroll Eq. (50) as

$$e_T = L_T + U_T, \qquad L_T = \alpha \sum_t \Phi_{T,t+1} \xi_t, \qquad U_T = \sum_t \Phi_{T,t+1} r_t.$$

The actual measurement increments remain martingale differences even though their conditional covariance depends on the random state. Hence $\mathbb{E}\|L_T\|_2^2 = Q_T^{\mathrm{act}}$. Minkowski's inequality gives $(\mathbb{E}\|U_T\|_2^2)^{1/2} \leq D_T$. Cauchy–Schwarz yields

$$\left|\mathbb{E}\|e_T\|_2^2 - Q_T^{\mathrm{act}}\right| \leq 2\sqrt{Q_T^{\mathrm{act}}}\, D_T + D_T^2,$$

and adding $|Q_T^{\mathrm{act}} - Q_T|$ proves Eq. (53).

The first term in Eq. (53) measures covariance change along the stochastic trajectory. The remaining terms measure nonlinear state propagation. Conditional unbiasedness of individual map estimates does not imply an unbiased terminal representation.

If $DG_\alpha$ is $L_t$-Lipschitz in operator norm along the line segment between $z_t^\infty$ and $z_t^S$, Taylor's theorem gives

$$\|r_t\|_2 \le \frac{L_t}{2}\|e_t\|_2^2. \tag{54}$$

This identifies a local small-error regime but is not a global convergence theorem for arbitrary trained QDEQs.

### E.3 Raw covariance contribution

Define the raw-diagonal terminal coefficient

$$a_t^{\mathrm{rawdiag}} = \alpha^2 \operatorname{tr}(\mathcal{D}(\Sigma_{Y,t})H_t)\,. \tag{55}$$

The omitted raw cross-covariance contribution is

$$a_t - a_t^{\mathrm{rawdiag}} = \alpha^2 \sum_{j\neq k} (\Sigma_{Y,t})_{jk}(H_t)_{kj}. \tag{56}$$

It can be positive or negative. For the four-qubit repeated readout, $H_{T-1} = R^\top R = 4I_4$, so raw cross-qubit covariance does not affect the Euclidean terminal-risk contribution of the final query. It may affect earlier queries after downstream dynamics mix directions.

If two coefficient profiles differ by a common positive scalar, $\bar{a}_t^{(1)} = c\bar{a}_t^{(2)}$ for all $t$, their bounded integer objectives have the same minimizing schedules. Covariance can therefore improve risk prediction without changing the selected allocation.

## F Bounded Allocation and Frozen Inference

### F.1 Continuous relaxation

For nonnegative calibration-averaged coefficients $c_t = \bar{a}_t$, consider

$$\min_{S_0,\dots,S_{T-1}} \sum_{t=0}^{T-1} \frac{c_t}{S_t} \quad \text{subject to} \quad \sum_t S_t = B, \qquad S_{\min} \le S_t \le S_{\max}. \tag{57}$$

If all coefficients are positive and no bound is active, Cauchy–Schwarz gives

$$\left(\sum_t \sqrt{c_t}\right)^2 = \left(\sum_t \sqrt{\frac{c_t}{S_t}}\sqrt{S_t}\right)^2 \le B\sum_t \frac{c_t}{S_t}. \tag{58}$$

Equality holds precisely when $S_t \propto \sqrt{c_t}$, giving

$$S_t^{\mathrm{cont}} = B\frac{\sqrt{c_t}}{\sum_j \sqrt{c_j}}. \tag{59}$$

With active bounds, the KKT conditions give

$$S_t^{\mathrm{cont}} = \operatorname{clip}\left(\sqrt{\frac{c_t}{\lambda}}, S_{\min}, S_{\max}\right), \tag{60}$$

where $\lambda > 0$ is chosen so that $\sum_t S_t = B$.

When $c_t = 0$, the lower bound prevents an unmeasured iteration. Zero-weight coordinates remain at $S_{\min}$ until all positive-weight coordinates have reached $S_{\max}$. If all weights are zero, every feasible allocation is optimal and a deterministic tie rule is used.

The shared calibration objective follows from exchanging finite sums:

$$\frac{1}{N_{\mathrm{cal}}} \sum_i \sum_t \frac{a_{i,t}}{S_t} = \sum_t \frac{\bar{a}_t}{S_t}. \tag{61}$$

Square roots and normalization do not commute with averaging, so averaging sample-specific normalized schedules does not generally solve the shared calibration objective.

### F.2 Exact bounded integer allocation

At current count $s$, assigning one additional shot to query $t$ reduces the objective by

$$\Delta_t(s) = \frac{c_t}{s} - \frac{c_t}{s+1} = \frac{c_t}{s(s+1)}. \tag{62}$$

For each query this marginal gain is non-increasing in $s$. Start from $S_t = S_{\min}$ and repeatedly increment the feasible coordinate with the largest current $\Delta_t(S_t)$ until the total is $B$.

Any feasible allocation selects a prefix of the marginal-gain sequence associated with each query. Because every sequence is non-increasing, the greedy procedure selects the largest prefix-consistent collection of $B - TS_{\min}$ increments. No other feasible allocation can have a larger total risk reduction, so the returned integer schedule is globally optimal. The optimum need not be unique when marginal gains tie; increasing time index is used as a deterministic tie rule.

A max-heap implements this rule in

$$O(T + (B - TS_{\min}) \log T)$$

time and $O(T)$ memory. This computation is performed offline.

### F.3 Terminal-aware calibration and inference algorithm

The order in the backward pass is essential: the coefficient for query $t$ is computed using the current $P = \Phi_{T,t+1}$ before the update $P \leftarrow PJ_t$.

### F.4 Definitions of the seven allocation mechanisms

All seven methods use the same horizon, damping, shot bounds, and exact integer total budget. Formal-test coefficients do not generate or tune any schedule.

Table 1: Information used by the seven schedule mechanisms.

| Method | Local uncertainty | Downstream dynamics | Raw cross-covariance |
|---|---|---|---|
| Uniform | – | – | – |
| Progressive | via target | via full fit | via target |
| Two-stage | via target | via full selection | via target |
| Uncertainty-only | ✓ | – | cancels under $R$ |
| Dynamics-only | fixed $I$ | ✓ | – |
| Raw-diagonal terminal | ✓ | ✓ | – |
| Full terminal | ✓ | ✓ | ✓ |

Table 2: Frozen mechanism definitions for the equal-budget allocation study.

| Method | Frozen construction |
|---|---|
| Uniform | Equal shots at all 25 queries, with the integer remainder assigned deterministically. |
| Progressive | A geometric temporal profile fitted by least squares to $\log \sqrt{\mathbb{E}_i[a_{i,t}^{\mathrm{full}}]}$ on the disjoint tuning split. |
| Two-stage | All 24 split points are enumerated on the tuning split; the two segment levels are constructed from the corresponding segment coefficients and the split minimizing the tuning surrogate is frozen. |
| Uncertainty-only | Uses $c_t = \alpha^2 \operatorname{tr}(R\Sigma_{Y,t}R^\top)$, omitting downstream propagation. |
| Dynamics-only | Uses identity covariance in raw measurement space: $c_t = \alpha^2 \operatorname{tr}(H_t)$. |
| Raw-diagonal terminal | Uses $c_t = \alpha^2 \operatorname{tr}[\mathcal{D}(\Sigma_{Y,t})H_t]$. |
| Full terminal-aware | Uses $c_t = \alpha^2 \operatorname{tr}(\Sigma_{Y,t}H_t)$. |

# G Frozen Protocol and Additional Empirical Evidence

## G.1 Tasks, splits, budgets, and aggregation

The allocation experiment fixes $T = 25$, $\alpha = 0.1$,

$$B \in \{6{,}400, 12{,}800, 25{,}600, 51{,}200, 102{,}400\},$$

and five fresh-measurement repeats per formal test input. Every Picard schedule spends exactly the stated joint-shot budget. The allocation target is $\|z_{25}^{S} - z_{25}^{\infty}\|_2^2$, where both trajectories use the same frozen model, input, initialization, horizon, and damping.

Reported intervals are descriptive, two-sided, unadjusted Student-$t$ 95% intervals across checkpoint-seed means: five seeds for MNIST-4 and three for each other task. They do not represent uncertainty over new architectures, new dataset splits, or future checkpoints.

## G.2 Formal propagation validation

For each formal input and each frozen method schedule,

$$Q_i(S) = \sum_t \frac{a_{i,t}^{\mathrm{full}}}{S_t}$$

is paired with the five-repeat mean measured terminal squared error. The saved coefficients already contain $\alpha^2$. This produces 56,840 matched input–schedule–budget conditions and 284,200 noisy rollouts. Formal-test coefficients are used only for this offline diagnostic: they never generate or select a schedule, and no common-reference qualification mask is applied.

## G.3 Common-reference qualification and finite-horizon floor

The solver comparison uses two independently configured exact Picard references. The primary candidate uses $\alpha = 0.1$ for at most 500 updates; the crosscheck uses $\alpha = 0.05$ for at most 1,000 updates. Both must reach relative exact-map residual at most $10^{-6}$ and agree in relative state distance to at most $5 \times 10^{-5}$.

Of the 332 unqualified inputs, 331 have a nonconverged independent crosscheck and one has a nonconverged primary candidate. None is classified as two distinct successfully converged candidate states. The qualification rule was not modified after inspection.

## G.4 Solver metrics and actual shot use

All 243,600 finite-shot solver rows are finite. None reaches the original relative exact-map residual threshold $10^{-3}$. Every finite-shot Broyden run reaches its selected maximum-update count, and every Picard run ends at the fixed horizon. These are therefore budget-capped approximate solves rather than certified converged fixed points.

Table 3: Frozen task-level protocol. Calibration and tuning are disjoint validation-side subsets; formal evaluation uses the frozen official-test subset.

| Task | Seeds | Qubits | Hidden dim. | Cal./tune | Test |
|---|---|---|---|---|---|
| MNIST-4 | 5 | 4 | 16 | 128/128 | 128 |
| FashionMNIST-4 | 3 | 4 | 16 | 128/128 | 128 |
| MNIST-10 | 3 | 10 | 100 | 200/200 | 200 |

Table 4: Formal propagation validation across all 11 checkpoints and seven schedules. Mean ratio is the ratio of task-level measured and predicted means. The interval is the two-sided Student-$t$ 95% interval for the geometric paired seed ratio. Abs. log error is the median per-input $|\log_{10}(\text{measured}/\text{predicted})|$.

| Task | Shots | Mean ratio | Seed ratio [95% CI] | Abs. log error |
|---|---|---|---|---|
| MNIST-4 | 6,400 | 0.993 | 0.993 [0.974, 1.012] | 0.101 |
| MNIST-4 | 12,800 | 0.997 | 0.997 [0.977, 1.018] | 0.101 |
| MNIST-4 | 25,600 | 0.989 | 0.990 [0.954, 1.028] | 0.104 |
| MNIST-4 | 51,200 | 0.995 | 0.994 [0.968, 1.021] | 0.112 |
| MNIST-4 | 102,400 | 1.000 | 1.001 [0.974, 1.028] | 0.103 |
| FashionMNIST-4 | 6,400 | 0.992 | 0.988 [0.888, 1.100] | 0.104 |
| FashionMNIST-4 | 12,800 | 0.996 | 0.997 [0.975, 1.019] | 0.100 |
| FashionMNIST-4 | 25,600 | 1.000 | 1.001 [0.945, 1.060] | 0.104 |
| FashionMNIST-4 | 51,200 | 0.981 | 0.982 [0.960, 1.004] | 0.103 |
| FashionMNIST-4 | 102,400 | 1.008 | 1.008 [1.002, 1.014] | 0.104 |
| MNIST-10 | 6,400 | 1.003 | 1.003 [0.980, 1.027] | 0.064 |
| MNIST-10 | 12,800 | 1.009 | 1.009 [0.983, 1.035] | 0.065 |
| MNIST-10 | 25,600 | 1.004 | 1.004 [0.996, 1.012] | 0.061 |
| MNIST-10 | 51,200 | 1.004 | 1.004 [0.989, 1.019] | 0.063 |
| MNIST-10 | 102,400 | 0.998 | 0.998 [0.984, 1.012] | 0.060 |

Picard spends exactly the stated cap. Broyden partitions the cap across a fixed maximum number of sampled-map calls and can leave a small integer remainder unused, which explains why its actual use can fall slightly below the cap. The validation-selected Broyden maximum is five updates in 48 of the 55 checkpoint–budget cells and ten updates in the remaining seven.

At 6,400 shots, Full Picard has lower task-mean common-reference squared distance than tuned Broyden for all three tasks. At 102,400 shots, tuned Broyden has the lower task-mean reference distance for all three. The high-budget Picard values approach the exact $T = 25$ controls in Table 5, exposing a deterministic finite-horizon floor. The classification columns are descriptive and do not establish a general classification-accuracy improvement.

### G.5 Offline calibration cost and empirical scope

The frozen offline preparation reports approximately 32.13 GPU-hours across the 11 checkpoints. This cost includes exact trajectories, state Jacobians, and measurement moments for calibration, tuning, and formal-test diagnostics. It is separate from the online joint-shot budget of one inference trajectory. Formal-test coefficients are diagnostic only.

The experiments isolate ideal finite-shot statistical sampling noise. They do not model gate noise, state-preparation error, assignment bias, temporal device drift, or another hardware-dependent bias model. Increasing the number of shots reduces variance under the assumed sampling law; it does not remove systematic hardware bias. Formal horizon/damping selection is also outside the frozen comparison: all reported allocation results use $T = 25$ and $\alpha = 0.1$.

Table 5: Common-reference coverage and the exact $T = 25$ Picard control. Reference distance is reported only on qualified inputs.

| Task | Qualified/designated | Exact $T = 25$ ref. $\ell_2^2$ | Exact $T = 25$ residual |
|---|---|---|---|
| MNIST-4 | 540/640 (84.4%) | 0.01077 | 0.0818 |
| FashionMNIST-4 | 310/384 (80.7%) | 0.00426 | 0.0663 |
| MNIST-10 | 442/600 (73.7%) | 0.01871 | 0.0797 |
| Total | 1,292/1,624 (79.6%) | – | – |

Table 6: Frozen solver outcomes at the lowest and highest common shot-budget caps. Reference squared distance and prediction disagreement use the original qualified subset. Relative exact-map residual, accuracy, and actual joint-shot use include all designated inputs. Means are across checkpoints with five measurement repeats.

| Task | Method, cap | Ref. $\ell_2^2$ | Residual | Accuracy | Disagree. | Shots |
|---|---|---|---|---|---|---|
| MNIST-4 | Full, 6.4k | 0.0128 | 0.1016 | 0.901 | 0.034 | 6,400 |
| | Tuned Broyden, 6.4k | 0.0385 | 0.1914 | 0.887 | 0.049 | 6,396 |
| | Full, 102.4k | 0.0109 | 0.0834 | 0.901 | 0.033 | 102,400 |
| | Tuned Broyden, 102.4k | 0.0026 | 0.0484 | 0.909 | 0.010 | 102,396 |
| FashionMNIST-4 | Full, 6.4k | 0.0063 | 0.0891 | 0.809 | 0.024 | 6,400 |
| | Tuned Broyden, 6.4k | 0.0275 | 0.1924 | 0.802 | 0.058 | 6,394 |
| | Full, 102.4k | 0.0044 | 0.0679 | 0.809 | 0.017 | 102,400 |
| | Tuned Broyden, 102.4k | 0.0038 | 0.0588 | 0.813 | 0.019 | 102,398 |
| MNIST-10 | Full, 6.4k | 0.0328 | 0.1273 | 0.622 | 0.292 | 6,400 |
| | Tuned Broyden, 6.4k | 0.0795 | 0.2180 | 0.485 | 0.482 | 6,396 |
| | Full, 102.4k | 0.0196 | 0.0838 | 0.706 | 0.125 | 102,400 |
| | Tuned Broyden, 102.4k | 0.0045 | 0.0542 | 0.701 | 0.135 | 102,396 |