SAME SYSTEM, FEWER MEASUREMENTS: STRUCTURED VARIATIONAL QUANTUM LINEAR SOLVERS

Anonymous authors
Paper under double-blind review

ABSTRACT

We introduce StructVQLS, a structure-based quantum linear solver exploiting recurring relations in widely used partial differential equation (PDE) discretizations for more accurate solutions with fewer measurements. On a 32768×32768 heat system, GPU circuit-simulator experiments achieve **61.7% lower mean equation residual** than merged exact Pauli at 512M measurements per run. **Stable 1% residual is reached with StructVQLS**, but not with Pauli within this budget in the evaluated cohort. In separate nonsymmetric convection-diffusion experiments, native sampling gives **50.6-60.9% lower mean residual** than merged Pauli across two advection weights. Separate measurement emulation achieves **75% fewer measurements with lower final residual** than merged Pauli in observed cross-budget comparisons. StructVQLS integrates exact operator completion, residual-preserving distributed coordination and weighted gradient estimation. Fixed stencil relations give grid-independent operator counts; coefficient weights control gradient noise. Common operator structure thus improves measurement efficiency.

1 INTRODUCTION

Variational quantum linear solvers (VQLS) fit $Ax = b$ through measured gradients (Bravo-Prieto et al., 2023). StructVQLS exploits common PDE structure to deliver **61.7% lower mean residual at equal measurement budgets** than merged Pauli in native circuit-simulator experiments at $N = 32768$. Separate native experiments on nonsymmetric convection-diffusion give **50.6-60.9% lower mean residual** than merged Pauli, supporting directed operator structure.

The key opportunity is recurring local coupling: finite differences repeat stencils; finite elements assemble mesh neighbors (PETSc Development Team, 2026). Refinement preserves a fixed stencil's relation types. For the same system and trial circuit, exact representations preserve the objective and exact gradient, but their coefficient weights govern finite-measurement gradient noise. StructVQLS uses this connection to improve learning.

Quantum notation as linear algebra. A circuit prepares normalized $|x(\theta)\rangle$; scale ρ gives $x = \rho|x(\theta)\rangle$. Kets are column vectors, bras their conjugate transposes; n qubits index $N = 2^n$ amplitudes. Measurements (*shots*) estimate inner products. Gates implement norm-preserving *unitaries*; an LCU (linear combination of unitaries), $A = \sum_\ell c_\ell U_\ell$, enables residual estimation. Pauli strings use a tensor-product dictionary; structured terms encode grid relations (Figure 1). Squared residuals give $O(L^2)$ term pairs before simplification.

Contributions. **(1) Exact structural representations.** Completion represents nonsymmetric stencils and directed interfaces with $2K + 1$ terms for K relations; weighted matching gives degree-controlled counts for symmetric meshes. Fixed lifts preserve residuals and derivatives. **(2) Distributed learning theory.** An exact row-residual identity, complete gradients and gradient-noise bounds connect structure to coordination and measurement allocation. **(3) Realized accuracy and efficiency gains.** GPU circuit sampling validates gains on large heat systems and nonsymmetric convection-diffusion. Measurement emulation extends source, scale and budget coverage; all eight native and 80 emulated upwind pairs favor StructVQLS.

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.

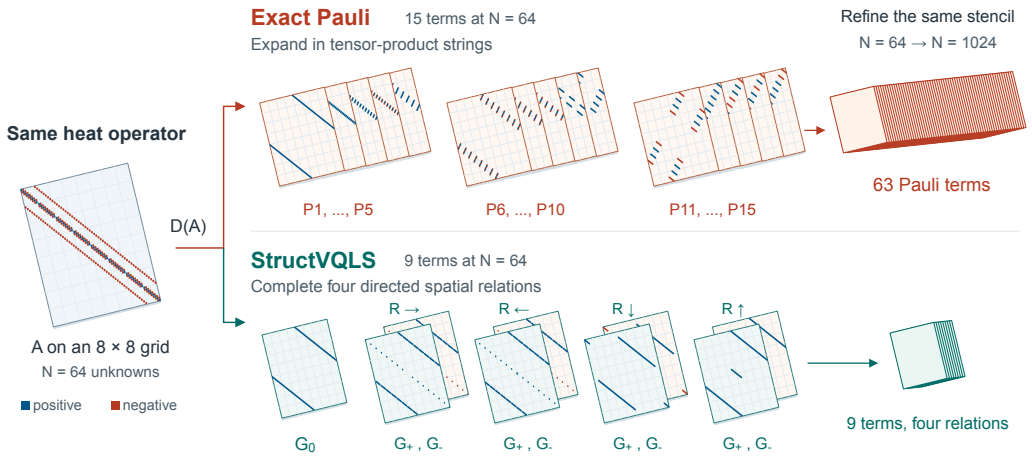

Figure 1: One heat operator, two exact dictionaries. For $\mu = 1$, $r_x = .25$, $r_y = .40$, refinement from $N = 64$ to 1024 increases Pauli terms from 15 to 63; StructVQLS retains nine by completing four spatial relations.

2 Related Work

Related foundations include quantum inversion (Harrow et al., 2009; Childs et al., 2017; Gilyén et al., 2019), variational methods (Cerezo et al., 2021a; Sato et al., 2021; Liu et al., 2021) and trainability (Yan et al., 2024; Gil-Fuster et al., 2025; Han et al., 2026).

Li et al. (2023) establish the $4d + 1$ self-inverse Poisson count; Liu et al. (2025) decompose A and A^2 for Poisson and banded Toeplitz systems; Shen et al. (2026) supply block-local coordination. StructVQLS connects these foundations to directed/nonsymmetric completion, an exact distributed residual identity and complete-gradient noise bounds. Our direct-square control uses Liu et al.'s A^2 decomposition with the same objective, circuit and optimizer. Sigma-basis methods (Gnanasekaran & Surana, 2024; 2026) provide compact operator access; D-VQLS (Lu et al., 2026) distributes variational solving; ME-VQLS (Liu & Wang, 2026) exploits real-overlap symmetry to group measurements. Our shared-objective controls distinguish structural representation from measurement grouping, charging output reconstruction.

3 Structured Decomposition and Distributed Learning

Inputs, output and optimization target. The inputs are relations and coefficients for A, a right-hand-side (RHS) preparation $b/\|b\|$, a trial circuit (*ansatz*) $|x(\theta)\rangle$, and S shots per gradient. Learn θ and $\rho = e^{\lambda}$ by minimizing $C(\theta, \lambda) = \|\rho A|x(\theta)\rangle - b\|^2/\nu$, with fixed $\nu = \|b\|^2 > 0$, reporting the learned-scale raw residual and aligned residual $R = \min_{c\in\mathbb{C}} \|cA|x(\theta)\rangle - b\|/\|b\|$. For one qubit, $|0\rangle, |1\rangle$ are basis vectors, $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, and X swaps the basis states. A *lift* adds a fixed auxiliary qubit while preserving the residual; Figure 2 shows the solver workflow.

Operator class and PDE realization. We target repeated local relations and symmetric weighted meshes. Backward Euler for anisotropic diffusion on tensor grids gives

$$A_{ii} = \mu I + \sum_{k=1}^{d} r_k L_k^{(k)}, \qquad r_k = \kappa_k \Delta t/h_k^2, \qquad N = \prod_{k=1}^{d} n_k. \tag{1}$$

Here L_k is the one-dimensional Dirichlet second-difference operator, $\mu > 0$ is the mass/reaction contribution, and directed face blocks transmit neighboring subdomain values.

The relation construction covers diffusion, reaction, higher-order and directed transport stencils with changing coefficients and inputs; weighted matching handles symmetric nonuniform edges.

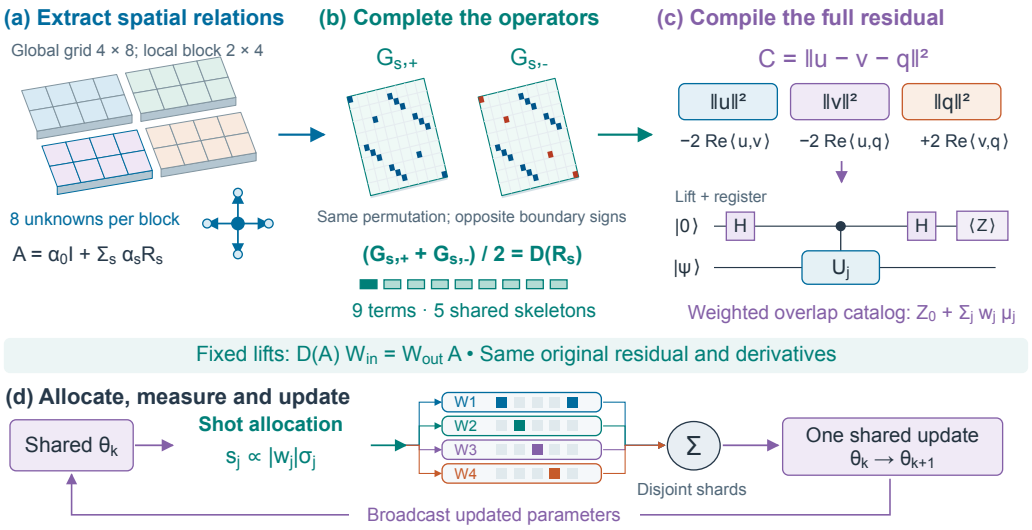

Figure 2: Solver workflow. (a,b) Complete grid relations with shared permutations. (c) Compile $\|u - v - q\|^2$ using fixed lifts, with $u = \rho Ax$, $v = b_i/m_i$, $q = Lz$. (d) Sum workers' disjoint measurement shards into one gradient update.

3.1 Exact completion and residual preservation

Let $A = \alpha_0 I + \sum_{s=1}^{K} \alpha_s R_s$, with nonzero partial signed permutations R_s: at most one ± 1 per row and column, with distinct unsigned supports (nonzero-entry locations) differing from the identity. Nonzero real coefficients and fixed symbolic relations specify every admissible grid, including truncated offsets and unequal forward/backward weights.

Theorem 1 (Fixed-relation completion)**.** *For $\alpha_0 \neq 0$, the Hermitian dilation $D(A) = \left[\begin{smallmatrix} 0 & A \\ A^\intercal & 0 \end{smallmatrix}\right]$ admits the indexed LCU*

$$D(A) = \alpha_0 G_0 + \sum_{s=1}^{K} \frac{\alpha_s}{2}(G_{s,+} + G_{s,-}), \qquad G_{s,\pm} = D(R_s) \pm P_s, \tag{2}$$

where P_s projects onto the inactive states of $D(R_s)$ and $G_0 = X \otimes I$. Every term is Hermitian, satisfies $G^2 = I$, and is a signed permutation. It has $2K + 1$ terms and $K + 1$ unsigned permutation skeletons for every N. Each pair satisfies $G_{s,-} = G_{s,+}(I - 2P_s)$.

Disjoint active and inactive supports give $D(R_s)^2 + P_s = I$; the artificial boundary terms cancel. Interiors and interfaces share one compiler: each relation uses one unsigned permutation and a boundary-sign switch.

Corollary 2 (Heat blocks and directed interfaces)**.** *For $A_d = \mu I + \sum_{k=0}^{d-1} r_k L_k^{(k)}$, with Dirichlet $L_k = \mathrm{tridiag}(-1, 2, -1)$, axis sizes $n_k \geq 2$, and $\mu, r_k > 0$, the $K = 2d$ directed nearest-neighbor relations give $4d + 1$ terms and $2d + 1$ skeletons. A directed face block $B = -r_k |n_k - 1\rangle\langle 0| \otimes I_{\text{face}}$ has $D(B) = r_k(G_+ + G_-)/2$ with two completed unitaries.*

Theorem 3 (Fixed-lift preservation)**.** *For symmetric A, $W_+ v = |+\rangle v$ satisfies $(X \otimes A)W_+ = W_+ A$. For any real square A, $W_{\text{in}} v = |1\rangle v$ and $W_{\text{out}} v = |0\rangle v$ satisfy $D(A)W_{\text{in}} = W_{\text{out}} A$. Consequently,*

$$\|Ax(\theta) - b - q\|^2 = \|D(A)W_{\text{in}} x(\theta) - W_{\text{out}}(b + q)\|^2, \tag{3}$$

with the analogous W_+ identity. All parameter derivatives are preserved.

Appendix A gives the full proofs.

Weighted local relations. Color a symmetric operator's edges into C weighted matchings T_r. For $s_r = \|T_r\| > 0$, unitaries $T_r/s_r \pm i\sqrt{I - (T_r/s_r)^2}$ sum to $2T_r/s_r$. Completing the diagonal gives at most $2C + 3$ terms, with $C \leq 2\Delta - 1$ for maximum degree Δ; angle and circuit costs remain mesh-dependent. Appendix G.5 proves this weighted construction.

3.2 COMPLETE DISTRIBUTED RESIDUALS

Partition A into row and column blocks. Row i has m_i agents; agent (i, j) owns A_{ij}, b_i/m_i, a solution replica $x_{ij} = \rho_{ij}|x(\theta_{ij})\rangle$, and a coordination vector $z_{ij} = \sigma_{ij}|z(\phi_{ij})\rangle$. Following Shen et al. (2026), let $q_{ij} = \sum_k (L_i)_{jk} z_{ik}$ for a symmetric row-graph Laplacian and $r_{ij} = A_{ij}x_{ij} - b_i/m_i - q_{ij}$.

Theorem 4 (Exact row-objective decomposition)**.** *Writing $y_{ij} = A_{ij}x_{ij} - b_i/m_i$ and $\bar{y}_i = m_i^{-1} \sum_j y_{ij}$ gives*

$$m_i \sum_j \|r_{ij}\|^2 = \Big\| \sum_j y_{ij} \Big\|^2 + m_i \sum_j \|(y_{ij} - \bar{y}_i) - q_{ij}\|^2. \tag{4}$$

Connected row graphs and unrestricted coordination vectors make the second term zero at the coordination optimum. With column consensus $x_{ij} = x_j$, the first term is the original block-row residual. A specified coordination ansatz contributes its nonnegative penalty; equal $m_i = m$ gives $m \sum_{ij} \|r_{ij}\|^2$.

Orthogonality of the centered difference and row mean proves the identity. For L terms and row degree Δ, the reference catalog has $Q_{\text{local}} = 2L^2 + 2L(\Delta + 2) + (\Delta + 1)(\Delta + 2)$ components, including action, source and coordination cross terms; $Q_{\text{local}}(9, 1) = 222$ for one-neighbor 2D heat.

Complete learning updates. Write $E_{ij} = \|r_{ij}\|^2$. The normalized distributed objective is $C_{\text{dist}} = \nu^{-1} \sum_{ij} m_i E_{ij}$, using the fixed global ν. For the equal-row-count experiments, $C_{\text{local}} = \sum_{ij} E_{ij}$ differs by a fixed factor absorbed into the step size. With solution parameters $u_{ij} = (\theta_{ij}, \rho_{ij})$, coordination parameters $v_{ij} = (\phi_{ij}, \sigma_{ij})$, and symmetric doubly stochastic mixing matrix W on a connected column graph, a synchronous round is

$$u_{ij}^+ = \sum_k W_{ik} u_{kj} - \eta \widehat{\nabla}_{u_{ij}} C_{\text{local}}, \qquad v_{ij}^+ = v_{ij} - \eta \widehat{\nabla}_{v_{ij}} C_{\text{local}}. \tag{5}$$

Gradients include every dependent residual. Row neighbors exchange coordination parameters and overlaps; column neighbors mix solution parameters in common circuit coordinates. Column consensus yields x; unequal row counts use weights m_i. Large-system learning uses *measurement sharding*: workers measure disjoint terms at shared parameters and sum gradients for Adam. Block-local controls test consensus and coupled residuals.

A compact grid-informed trial state. For 2D heat, each even-sized axis uses three self-inverse maps: E swaps $(0, 1), (2, 3), \ldots$; O_+ swaps $(1, 2), (3, 4), \ldots$ and fixes endpoints; O_- adds a minus sign at those endpoints. Thus $E + (O_+ + O_-)/2$ is exactly the nearest-neighbor adjacency. These data-coordinate maps preserve the lift. With normalized input $|b\rangle$, let $R_b = 2|b\rangle\langle b| - I$ and $V = e^{-i\pi R_b/4}$. One layer applies rotations about the six axis maps and R_b, giving

$$|x(\theta)\rangle = V U_D(\theta) \cdots U_1(\theta) V^\dagger |b\rangle, \qquad e^{-i\theta G/2} = \cos(\theta/2) I - i \sin(\theta/2) G. \tag{6}$$

Each U_ℓ has seven rotations: D layers use $7D$ angles; zero angles return $|b\rangle$. The source phase gives orthogonal tangent $(G - \langle b|G|b\rangle)|b\rangle/2$, real for real G, b. Preparation U_b implements $R_b = U_b(2|0\rangle\langle 0| - I)U_b^\dagger$; swaps and endpoint phases implement axis maps. All representations share this circuit and input-only initialization; generators obey the derivative rule below.

3.3 WEIGHTED ESTIMATION AND LEARNING

From a residual to measured overlaps. For $r = \rho \sum_\ell c_\ell U_\ell x - d$, expanding $\|r\|^2$ gives $\rho^2 \sum_{\ell,m} c_\ell^* c_m \langle U_\ell x | U_m x\rangle - 2\rho \operatorname{Re} \sum_\ell c_\ell \langle d | U_\ell x\rangle + \|d\|^2$. An overlap is an inner product: after vector normalization, auxiliary-qubit interference gives binary outcomes with its real or imaginary part as mean. Coefficients and preparation norms determine weights; derivative shifts give gradient weights.

Definition 5 (Finite-shot representation profile)**.** *For an exact scalar estimator $Z = Z_0 + \sum_{j=1}^M w_j \mu_j$, define $\mathfrak{C} = (L, M, \{w_j, \sigma_j^2, s_j, \tau_j\}_{j=1}^M)$: σ_j^2 is single-shot variance, s_j its allocation, and τ_j single-shot cost. M counts components after the stated exact simplifications; Z_0 contains known constants.*

Proposition 6 (Independent-component allocation)**.** *For independent outcomes in* $[-1, 1]$ *and positive allocations,*

$$\mathrm{Var}(\widehat{Z}) = \sum_j \frac{|w_j|^2\sigma_j^2}{s_j}, \quad V_{\mathrm{unif}} = \frac{M\sum_j |w_j|^2\sigma_j^2}{S}, \quad V_* = \frac{(\sum_j |w_j|\sigma_j)^2}{S}. \tag{7}$$

The continuous optimum uses $s_j \propto |w_j|\sigma_j$*. For* $\tau_j > 0$*, under serial time* $\sum_j \tau_j s_j = T$*, it uses* $s_j \propto |w_j|\sigma_j/\sqrt{\tau_j}$ *and gives* $V_{T,*} = (\sum_j |w_j|\sigma_j\sqrt{\tau_j})^2/T$*.*

Cauchy-Schwarz on $|w_j|\sigma_j/\sqrt{s_j}$ and $\sqrt{s_j}$ gives the optimum. Fresh post-pilot samples preserve conditional unbiasedness; integer budgets and worker assignment determine realized cost.

From symbolic coefficients to residual weights. Splitting each relation coefficient between its two completions preserves its absolute mass. Thus

$$\lambda_1 = |\alpha_0| + \sum_s |\alpha_s|, \qquad \lambda_1^{\mathrm{heat}} = \mu + 4\sum_k r_k. \tag{8}$$

The dimension-uniform bounds below hold for fixed r_k and bounded preparation scales. At fixed physical Δt, refinement can increase r_k: relation count stays fixed while coefficient mass changes. Write $r = \sum_\nu a_\nu |u_\nu\rangle$ with real a_ν and normalized u_ν, and let $H = \sum_\nu |a_\nu|$. Expanding $\|r\|^2$ gives total overlap weight $\|w\|_1 \le \sum_{\nu,\omega} |a_\nu a_\omega| = H^2$. Exact simplification cannot increase this norm; continuous coefficient-proportional allocation gives $\mathrm{Var}(\widehat{E}) \le H^4/S$ for $E = \|r\|^2$.

Corollary 7 (Complete-residual variance envelope)**.** *For normalized preparations, let* $a = |\rho|\lambda_1$*,* $b_* = \|b_i/m_i\| + \sum_k |(L_i)_{jk}\sigma_{ik}|$*, and* $H = a + b_*$*, where* $\lambda_1 = \sum_\ell |c_\ell|$*. The residual estimator has weight one-norm at most* H^2 *and continuous coefficient-proportional variance at most* H^4/S*. A solution angle occurring once in an involution rotation has separated-shift weight bound* $B_k = a^2 + \sqrt{2}ab_*$*; the log-scale bound is* $B_\lambda = 2aH$*. Under continuous allocation with* S/P *shots per coordinate,*

$$\mathbb{E}\|\widehat{g} - g\|^2 \le \frac{P}{S}\sum_{k=1}^{P} B_k^2. \tag{9}$$

Fixed relations, coefficients, graph weights and degree, bounded solution/RHS/coordination scales, and bounded gate multiplicity make each coordinate envelope independent of N*. The full-vector bound retains* P*.*

For the fixed normalized objective, divide derivative weights by ν and variance bounds by ν^2.

Frequency-resolved complete gradients. Fix $d = b_i/m_i + q_{ij}$ and ρ, and write $E = Q + F + \|d\|^2$, where $Q = \rho^2\|Ax\|^2$ and $F = -2\rho\,\mathrm{Re}\langle d|Ax\rangle$. For an angle appearing once in $e^{-i\theta G/2}$ with $G^2 = I$, Q has frequencies $0, 1$ and F has frequency $1/2$. Hence

$$\partial_\theta E = \frac{Q(\theta + \pi/2) - Q(\theta - \pi/2)}{2} + \frac{F(\theta + \pi/2) - F(\theta - \pi/2)}{2\sqrt{2}}. \tag{10}$$

These differences give $B_k = a^2 + \sqrt{2}ab_*$; $\partial_\lambda E = 2Q + F$. Coordination derivatives include the dependence of d on q. Repeated gate parameters contribute once per occurrence (Schuld et al., 2019; Wierichs et al., 2022).

Randomized allocation at large catalogs. Write $g_k = o_k + \sum_j w_{kj}\mu_{kj}$ with analytic o_k and binary $Y_{kj} \in \{-1, 1\}$ of mean μ_{kj}. For $b_k = \sum_j |w_{kj}| > 0$, each shot selects j with probability $|w_{kj}|/b_k$ and returns $Z = \mathrm{sign}(w_{kj})Y_{kj}$, avoiding per-term shot minima:

$$\widehat{g}_k = o_k + b_k\overline{Z}_k, \qquad \mathrm{Var}(\widehat{g}_k) = \frac{b_k^2 - (g_k - o_k)^2}{S_k}. \tag{11}$$

Summing over terms gives the mean; $Z^2 = 1$ gives variance, including term selection. Integer allocation reserves $S_k \ge 1$ per active coordinate, requiring $S \ge P_{\mathrm{active}}$, and rounds the remaining proportional shares to keep $\sum_k S_k = S$. Workers pool signed counts; $b_k = 0$ gives $g_k = o_k$ analytically.

Unbiased reuse of past counts. A *control variate*, an additive correction with known conditional mean, controls term-selection noise. Fix h_{kj} to the preceding update's term mean, zero if unobserved. With $p_{kj} = |w_{kj}|/b_k$, fresh $J_s \sim p_k$ and binary Y_s, use

$$\widehat{g}_k = o_k + \sum_j w_{kj} h_{kj} + \frac{b_k}{S_k} \sum_{s=1}^{S_k} \operatorname{sign}(w_{kJ_s})(Y_s - h_{kJ_s}). \tag{12}$$

Let $\delta_k = \sum_j w_{kj}(\mu_{kj} - h_{kj})$. Conditional on past data $\mathcal{F}_t$, the correction has mean δ_k, preserving unbiasedness. Since $Y^2 = 1$ and $|h_{kj}| \le 1$,

$$\operatorname{Var}(\widehat{g}_k \mid \mathcal{F}_t) = \frac{b_k \sum_j |w_{kj}|(1 - 2h_{kj}\mu_{kj} + h_{kj}^2) - \delta_k^2}{S_k} \le \frac{4b_k^2}{S_k}. \tag{13}$$

The bound uses $|Y - h| \le 2$; its coordinate sum bounds gradient mean-square error (MSE). All native reuse-on arms use preceding charged counts, initially $h = 0$ (Eq. (11)). Variance reduction depends on past means' agreement with current means.

Representation-controlled stochastic descent. For fixed θ_0, if $C \ge C_{\inf}$ is $\mathcal{L}$-smooth, estimates are conditionally unbiased with mean-square error at most V, and stochastic gradient descent (SGD) uses $\eta \le 1/\mathcal{L}$, then (Ghadimi & Lan, 2013)

$$\frac{1}{T} \sum_{t<T} \mathbb{E}\|\nabla C(\theta_t)\|^2 \le \frac{2(C(\theta_0) - C_{\inf})}{\eta T} + \mathcal{L}\eta V. \tag{14}$$

A uniform bound on $4\sum_k b_k^2/S_k$ supplies V with reuse; Eq. (9) bounds continuous allocation without reuse. Residual targets also depend on capacity, conditioning and gradient signal. SGD probes test noise; Adam runs measure residuals.

Direct squared-operator control. For 2D heat, write $S_k = T_k^+ + T_k^-$ and $a_0 = \mu + 2(r_x + r_y)$. Then

$$A^2 = a_0^2 I - 2a_0(r_x S_x + r_y S_y) + r_x^2 S_x^2 + r_y^2 S_y^2 + 2r_x r_y S_x S_y, \tag{15}$$

where $S_k^2 = 2I - P_{k,\mathrm{lo}} - P_{k,\mathrm{hi}} + (T_k^+)^2 + (T_k^-)^2$. Here $T_k^{\pm}$ are truncated shifts; $P_{k,\mathrm{lo/hi}}$ select boundary faces. Completing 16 relations gives 33 indexed terms with analytic identity. The linear residual retains nine-term A, testing direct product construction (Liu et al., 2025) with the same objective and circuit.

4 Experimental Evaluation

Evaluation design and metrics. We test large-system measurement efficiency, nonsymmetric native learning and cross-resolution structural predictions. Representation pairs share A, source, circuit, initialization, optimizer and per-gradient budget. S denotes StructVQLS; M, *merged Pauli*, combines repeated $A^\dagger A$ products and removes identities; D, *direct square*, compiles A^2 using Eq. (15). The raw residual $r = \sqrt{C} = \|A(\rho\psi) - b\|/\|b\|$ measures returned-vector mismatch ($r = .01$: 1%); R optimizes scale and phase. Native results use actual NVIDIA circuit-simulator samples. Measurement emulation samples the exact stated binary-outcome law without executing circuits; exact-gradient runs diagnose optimization. Residual checks use the original A, b and never supply sampled training updates.

4.1 Native measurements to 1% original-system residual

Frozen paired protocol. On a 128×256 heat grid ($N = 32768$), $A = I + .25L_x + .25L_y$, normalized $b \propto e^{\cdot 3x + .5y}$ and condition number approximately three, a depth-1 circuit uses seven angles and learned log scale. Four input-centered starts share $[-.03, .03]$ perturbations and 512 Adam $(.9, .9)$ updates, learning rate .003 then cosine decay to .00006, at 1M shots/update. S/M/D use reuse; S/D also run without it. Stable 1% requires five passing checkpoints ten updates apart and a passing endpoint; non-hits are censored at 512M shots (Appendix B).

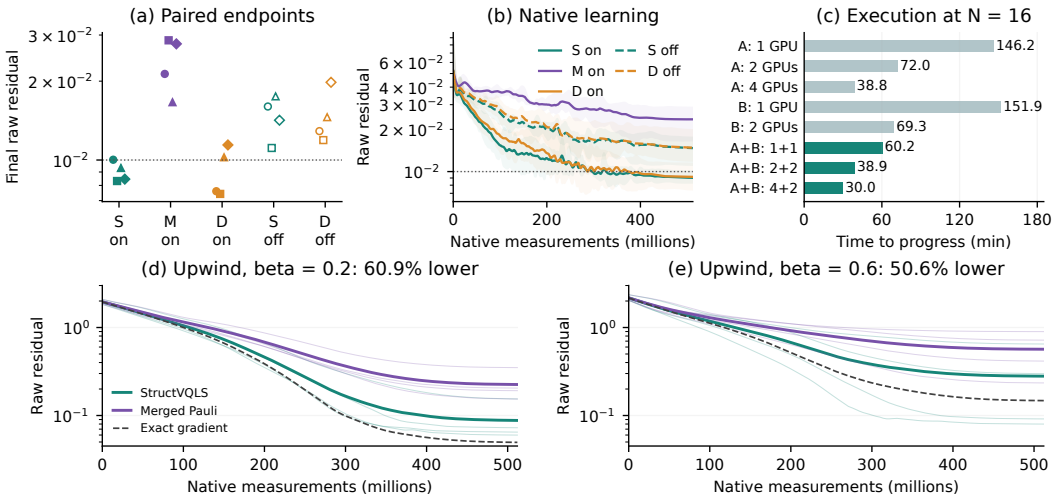

Figure 3: Native sampling. (a,b) Heat, $N = 32768$: four paired endpoints (shapes), mean curves/full ranges; reuse on/off (solid/dashed), 512M shots/arm. (c) Eight $N = 16$ layouts: time to common progress, including setup/transport. (d,e) Upwind, $N = 256$: all four starts/weight (thin), means (thick), exact controls aligned by update (dashed); 512M shots/arm. Reductions compare final means.

Precision at equal measurement budgets. With reuse, mean raw residuals are **.009030 for S and .02358 for M**, a **61.7% reduction**, consistent across paired endpoints. The paired mean residual difference, S minus M, is $-.01455$ (95% Student-t interval $[-.02458, -.004525]$). **StructVQLS reaches stable 1% residual**; merged Pauli does not within the same budget in this cohort. Direct square gives .009172; S-minus-D is $-.000142$ (95% interval $[-.003841, .003557]$; Appendix B).

Measurement reuse. At matched final-update points and allocations, reuse lowers S's conditional gradient variance by **79.1%**, preserving the mean. All S/D pairs improve, with paired 95% intervals excluding zero (Appendix B).

4.2 Native learning on nonsymmetric PDEs

This study extends the heat-system evidence to directed operator completion under native sampling. At $N = 256$, $A = I + .25L + \beta(I - T^-)$, $\beta = .2, .6$ and normalized $b \propto e^{.4x}$, four fresh pairs/weight share a depth-3 circuit and 512 Adam updates at 1M native shots/update. S uses five completed terms; M merges exact Pauli products. Both use complete gradients without reuse. At equal 512M budgets, mean raw residuals are **.08790 versus .2251** for $\beta = .2$ and **.2800 versus .5666** for $\beta = .6$: **60.9% and 50.6% lower**, consistently favoring S across paired starts. Paired S-minus-M 95% Student-t intervals are $[-.2050, -.06938]$ and $[-.4651, -.1080]$ (Figure 3d,e; endpoints and audits in Appendix J).

4.3 Matched-budget accuracy across sources and scales

Six sources (smooth, signed sine mixtures and localized Gaussians) each use two fresh starts at $N = 8192, 16384, 32768$ with the heat operator and a generic depth-5 R_y, R_z, R_y/controlled-Z circuit. The 196/211/226 variables include scale. The 108 sampled paths and 36 exact controls share 3200 Adam updates, with 400M shots per sampled path (Appendix D.1).

At $N = 8192, 16384, 32768$, structured residuals .2945/.2898/.2926 improve on merged Pauli .3989/.4430/.4665 by **26.2/34.6/37.3%**; paired difference intervals exclude zero at all sizes, and all 36 paired final residuals favor StructVQLS (Figure 4). Direct-square means are .2909/.2905/.3004; S-minus-D intervals span zero. No sampled path reaches stable .1. Exact controls diagnose shared optimization; the RHS itself gives $R \leq .05433$, and sparse conjugate gradients reach 10^{-10} in milliseconds. These are quantum-circuit learning comparisons; grid-informed circuits establish 1% precision (Section 4.1).

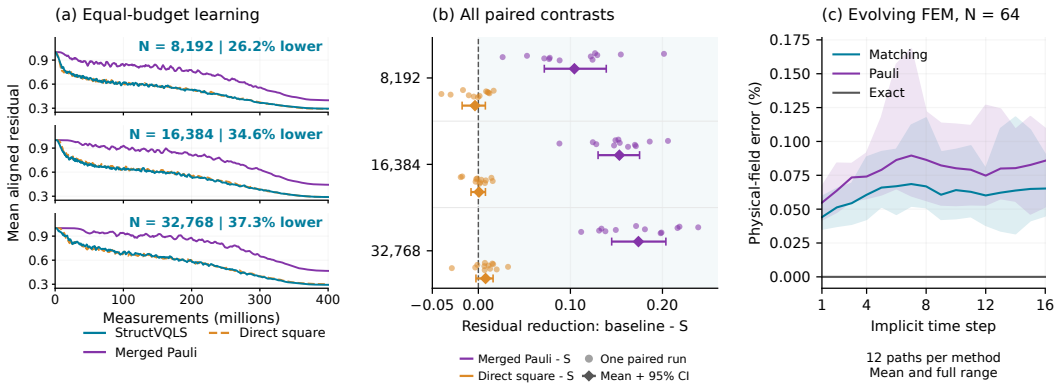

Figure 4: Measurement emulation. (a,b) Six sources, two starts each, 400M shots: mean curves, all paired reductions and 95% paired-bootstrap intervals (10000 source/restart resamples). S enters both contrasts. (c) Six evolving meshes, 12 warm/cold paths/method: mean/full range; 3.2768×10^{18} shots/path, with exact controls (Section 4.6).

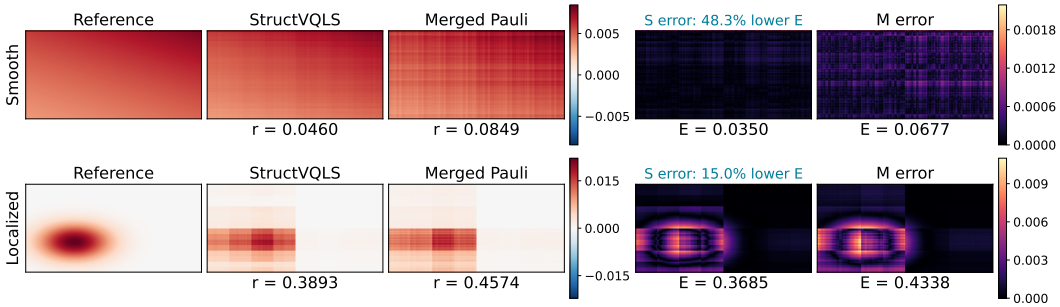

Figure 5: Spatial accuracy at $N = 32768$: 256 emulated updates, 512M shots, input-centered depth-5 circuit, first registered start/source. Columns show reference, real S/M solutions and complex absolute errors; scales are shared within rows. $r = \sqrt{C}$, $E = \|x - x_{\text{ref}}\|/\|x_{\text{ref}}\|$. Percentages compare shown samples; cohort outcomes are in Section 4.3.

Upwind convection-diffusion with advection weights $.2, .6$ uses depth-3 circuits, 512 Adam updates, 1M emulated measurements/update and eight paired starts per setting. At 1D $N = 256, 512, 1024$ and 2D $16 \times 16, 32 \times 32$ grids, paired final raw residuals consistently favor S; mean reductions are **59.4-70.1%** and **35.1-45.8%**, respectively. At $N = 512$, S with 128M/512M total measurements outperforms M with 512M/2048M: **75% fewer measurements and lower residual** across paired weight/budget comparisons. Paired 95% intervals exclude zero; budget contrasts reuse the same 16 starts (Appendix I).

The spatial cohort crosses smooth/localized sources, three starts, sizes and budgets. At $N = 32768$ and 512M shots, smooth S/M means are .04536/.07238 with stable 5% residual attained by S but not M; localized means are .3873/.4538. Figure 5 shows solution error E and mismatch r (Appendix E).

4.4 Structural mechanism and distributed correctness

Grid-independent term counts characterize the representation; execution costs also depend on circuits and training. Seven families (diffusion, higher-order and nonsymmetric stencils) use 5-17 structured versus 30-2045 Pauli terms at $N = 1024$. Across 28 noise cases (200k full-gradient shots, three seeds, 500 estimates/seed), equal-count Toeplitz and higher-count nine-point controls isolate weights' role (Figure 6). At $N = 32$, a charged 5% adaptive pilot still gives S/M gradient RMSE .03273/.03591. Across three families, $N = 32, 64$, five angles and 200k/1M shots, S has lower gradient RMSE in all coefficient-allocation and most adaptive comparisons (mean ratios .546/.517; Appendix H.1). The six-size 3D study favors S at $N = 128, 256$ under three allocations; 216 gradient cells give measured/predicted MSE 1.0006. Complete 8-qubit circuits cost 943.86/1468.58 expected CX

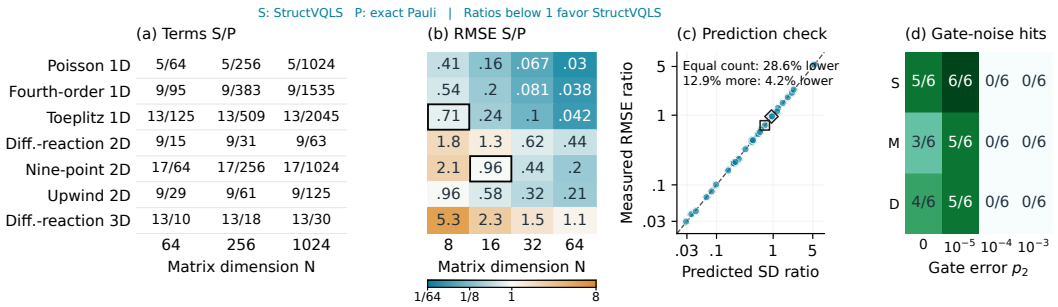

Figure 6: Structural weights and noise response. (a) Exact terms. (b) All 28 gradient RMSE ratios, structured/Pauli, with uniform allocation. (c) Predicted standard-deviation versus measured RMSE ratios; dashed: equality. Outlines mark equal/more-component controls (Toeplitz $N = 8$/nine-point $N = 16$). (d) Separate $N = 16$ gate-noise cohort: stable 1% successes out of six starts, 10.24B emulated shots per trajectory, $p_1 = p_2/10$.

gates/sample for S/M, including preparation/readout. In 45 shared-gradient runs, coupled block-local updates preserve objective/gradient parity to $4.44 \times 10^{-16}/2.44 \times 10^{-15}$, with aligned residual at most .001319 (Appendix F).

4.5 Shared objectives and distributed execution

At $N = 16$, 100 paired starts share a real depth-5 circuit, normalized local objective $C_L = 1 - \langle\psi|A^\dagger U_b P_0 U_b^\dagger A|\psi\rangle/\langle\psi|A^\dagger A|\psi\rangle$, where $P_0 = n^{-1}\sum_j (|0\rangle\langle 0|)_j$ and $U_b|0\rangle = b/\|b\|$. This measures local RHS agreement. After 1024 Adam updates, raw 1% success rates at 1M, 10M and 100M shots per update are **78%, 95% and 99%** for structured grouping, **78%, 100% and 99%** for merged-equivalent (ME) grouping, and **9%, 81% and 84%** for ordered complex Pauli; exact controls succeed at every start. Complete all-to-all circuits cost 46.80M versus 16.92M CX gates per 1M-shot update for S/ME; across three native starts, S reaches the target less often than ME (Appendix G).

Time to common early progress. Eight native layouts share $N = 16$, four blocks, depth 3, one start and 400 updates (64B shots each). All reach raw residual $\leq .75$ at two checkpoints 50 updates apart by update 350. Including setup and transport, six GPUs take 29.95 minutes versus 146.24 on one ($4.88\times$), with final residuals .585-.592 (Figure 3c; Appendix G).

4.6 Circuit geometry, gate noise and weighted mesh instances

At $N = 256$, condition number 100 and real alternating-entangler depth 18, six starts compare 4096 metric versus Adam updates. The metric rescales gradients using inner products of Ax derivatives. It attains stable 1% residual in most starts, while Adam does not, both with exact gradients and at 10^{18} emulated shots/update. A full-chain real circuit has certified residual floor .0467 (Appendix G.3). For compiled gate noise at $N = 16$, 72 depth-3 trajectories cross six starts, three representations and four depolarizing levels over 1024 updates at 10M emulated shots/update. Of 12 zero-noise successes, one loses stable 1% at 10^{-5}, all at higher levels (Figure 6d; Appendix G). On evolving domains, six fresh finite-element meshes (49 vertices padded to 64) give 12 warm/cold paths per method (exact/matching/Pauli). All attain 1% across 16 implicit solves and in final field error, with classically supplied RHSs (Figure 4c; Appendix G.5).

5 Conclusion

At $N = 32768$, native heat-system training lowers mean residual by **61.7%** versus merged Pauli at 512M shots. Stable 1% residual is reached with StructVQLS but not Pauli in this cohort within that budget. Native nonsymmetric training at $N = 256$ lowers mean residual by **50.6-60.9%**. At $N = 512$, observed emulated budget comparisons use **75% fewer measurements with lower residual**, supporting structure-based measurement efficiency across common PDE discretizations.

AI Assistance Statement

The authors used large language models to assist with drafting and revising manuscript sections, improving clarity and organization, and identifying related literature. AI tools also assisted with experimental planning, code development, experiment execution, and analysis of numerical results. The authors are responsible for verifying the cited sources, mathematical claims, experimental procedures, and reported results, and take full responsibility for the final manuscript.

Reproducibility. Supplementary code, data and protocols support reproduction.

References

Andrew Arrasmith, Lukasz Cincio, Rolando D. Somma, and Patrick J. Coles. Operator sampling for shot-frugal optimization in variational algorithms. *arXiv preprint arXiv:2004.06252*, 2020. URL https://arxiv.org/abs/2004.06252.

Carlos Bravo-Prieto, Ryan LaRose, M. Cerezo, Yigit Subasi, Lukasz Cincio, and Patrick J. Coles. Variational quantum linear solver. *Quantum*, 7:1188, 2023. doi: 10.22331/q-2023-11-22-1188.

Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995. doi: 10.1137/0916069.

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

M. Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J. Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*, 12:1791, 2021b. doi: 10.1038/s41467-021-21728-w.

Andrew M. Childs, Robin Kothari, and Rolando D. Somma. Quantum algorithm for systems of linear equations with exponentially improved dependence on precision. *SIAM Journal on Computing*, 46 (6):1920–1950, 2017. doi: 10.1137/16M1087072.

Saeed Ghadimi and Guanghui Lan. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013. doi: 10.1137/120880811.

Elies Gil-Fuster, Casper Gyurik, Adrián Pérez-Salinas, and Vedran Dunjko. On the relation between trainability and dequantization of variational quantum learning models. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/pdf?id=TdqaZbQvdi.

András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 193–204, 2019. doi: 10.1145/3313276.3316366.

Abeynaya Gnanasekaran and Amit Surana. Efficient variational quantum linear solver for structured sparse matrices. *arXiv preprint arXiv:2404.16991*, 2024. URL https://arxiv.org/abs/2404.16991.

Abeynaya Gnanasekaran and Amit Surana. Efficient quantum access model for sparse structured matrices using linear combination of "things". *Physical Review A*, 113(2):022437, 2026. doi: 10.1103/x4b2-9d94. URL https://arxiv.org/abs/2507.03714.

Jason Han, Nicholas S. DiBrita, Daniel Leeds, Jianqiang Li, Jason Ludmir, and Tirthak Patel. Layerwise federated learning for heterogeneous quantum clients using Quorus. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/pdf?id=ZwwFuVQv64.

Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103:150502, 2009. doi: 10.1103/PhysRevLett.103.150502.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015. URL https://arxiv.org/abs/1412.6980.

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.

Hui-Min Li, Zhi-Xi Wang, and Shao-Ming Fei. Variational quantum algorithms for Poisson equations based on the decomposition of sparse Hamiltonians. *Physical Review A*, 108:032418, 2023. doi: 10.1103/PhysRevA.108.032418.

Hai-Ling Liu, Yu-Sen Wu, Lin-Chun Wan, Shi-Jie Pan, Su-Juan Qin, Fei Gao, and Qiao-Yan Wen. Variational quantum algorithm for the Poisson equation. *Physical Review A*, 104:022418, 2021. doi: 10.1103/PhysRevA.104.022418.

Xiaoqi Liu, Yuedi Qu, Ming Li, and Shu-qian Shen. Variational quantum algorithm for the Poisson equation based on the banded Toeplitz systems. *Communications in Theoretical Physics*, 77:045101, 2025. URL https://arxiv.org/abs/2504.14828.

Yunya Liu and Pai Wang. Measurement-efficient variational quantum linear solver for carleman-linearized nonlinear dynamics. *arXiv preprint arXiv:2605.15366*, 2026. URL https://arxiv.org/abs/2605.15366.

Chao Lu, Pooja Rao, Muralikrishnan Gopalakrishnan Meena, and Kalyana Chakaravarthi Gottiparthi. Distributed variational quantum linear solver. *arXiv preprint arXiv:2604.14435*, 2026. URL https://arxiv.org/abs/2604.14435.

Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9:4812, 2018. doi: 10.1038/s41467-018-07090-4.

PETSc Development Team. DM basics. PETSc documentation, 2026. URL https://petsc.org/release/manual/dmbase/. Accessed September 23, 2026.

Yuki Sato, Ruho Kondo, Satoshi Koide, Hideki Takamatsu, and Nobuyuki Imoto. Variational quantum algorithm based on the minimum potential energy for solving the Poisson equation. *Physical Review A*, 104:052409, 2021. doi: 10.1103/PhysRevA.104.052409.

Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99:032331, 2019. doi: 10.1103/PhysRevA.99.032331.

Tong Shen, Zeru Zhu, and Ji Liu. Distributed variational quantum linear solver. *arXiv preprint arXiv:2604.01426*, 2026. URL https://arxiv.org/abs/2604.01426.

David Wierichs, Josh Izaac, Cody Wang, and Cedric Yen-Yu Lin. General parameter-shift rules for quantum gradients. *Quantum*, 6:677, 2022. doi: 10.22331/q-2022-03-30-677.

Ge Yan, Hongxu Chen, Kaisen Pan, and Junchi Yan. Rethinking the symmetry-preserving circuits for constrained variational quantum algorithms. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/pdf?id=SL7djdVpde.

A Complete Theory, Derivatives, and Worker Reduction

A.1 Structured completion and fixed lifts

The operator class is

$$A = \alpha_0 I + \sum_{s=1}^{K} \alpha_s R_s, \tag{16}$$

$$D(A) = \begin{bmatrix} 0 & A \\ A^{\mathsf{T}} & 0 \end{bmatrix}. \tag{17}$$

with nonzero real coefficients. Each nonzero R_s is a partial signed permutation. Its unsigned support differs from the identity and from every other registered relation support. Grid sizes are admissible when these conditions hold. The statements concern a constructed indexed LCU; exact algebraic simplification can further reduce its active size.

Proof of Theorem 1. Write $S_s = D(R_s)$ and let P_s project onto the states not incident to its matching. The partial-permutation property gives $S_s^{\mathsf{T}} = S_s$, $S_s^2 = I - P_s$, and $S_s P_s = P_s S_s = 0$. Thus $G_{s,\pm} = S_s \pm P_s$ are symmetric and square to I. Every active row has the single signed matching entry, and every inactive row has a single diagonal entry ± 1, so each completion is a signed-permutation unitary. Their half-sum is S_s, proving the reconstruction. Including $G_0 = D(I)$ gives the $2K + 1$ indexed terms. Within a pair only inactive diagonal signs differ. The distinct unsigned relation supports, also distinct from the identity, give exactly $K + 1$ unsigned skeletons. Finally, $(S_s + P_s)(I - 2P_s) = S_s - P_s$; the same calculation on the left proves commutation with the sign layer. □

For the heat specialization,

$$A_d = \mu I + \sum_{k=0}^{d-1} r_k L_k^{(k)}, \qquad L_k = 2I_k - T_k^{+} - T_k^{-}, \tag{18}$$

where $T_k^{\pm}$ are truncated nearest-neighbor shifts. With $a_0 = \mu + 2\sum_k r_k$, the explicit representation is

$$X \otimes A_d = a_0 G_0 - \sum_{k=0}^{d-1} \frac{r_k}{2} \sum_{s\in\{+,-\}} \sum_{t\in\{+,-\}} G_{k,s,t}. \tag{19}$$

Proof of Corollary 2. There are two nonzero directed shifts on every axis of size at least two. Substitution into Theorem 1 gives $L = 4d + 1$ terms and $2d + 1$ skeletons. For a directed interface,

$$B_f^{(k)} = -r_k\big(|n_k - 1\rangle\langle 0| \otimes I_{\text{face}}\big), \qquad B_b^{(k)} = (B_f^{(k)})^{\mathsf{T}}, \tag{20}$$

the normalized operator B/r_k is itself one partial signed permutation. Its dilation is therefore the half-sum of two completions. □

Lemma 8 (Boundary-sign factorization)**.** *For heat axis k and orientation s, let $P_{k,s}^{\partial}$ project onto the inactive lifted boundary states. Then*

$$D_{k,s}^{\partial} = I - 2P_{k,s}^{\partial}, \tag{21}$$

$$G_{k,s,-} = G_{k,s,+} D_{k,s}^{\partial} = D_{k,s}^{\partial} G_{k,s,+}. \tag{22}$$

The negative diagonal support of $D_{k,s}^{\partial}$ is exactly the fixed-point set of the unsigned skeleton and has $2N/n_k$ entries.

Proof. The factorization follows from the disjoint-support identities in the proof of Theorem 1. Every active lifted state is in a transposition between lift halves, so only inactive states are fixed. There is one unmatched boundary face of size N/n_k in each lift half. □

Corollary 9 (Permutation reuse)**.** *The $4d$ completed heat terms use $2d$ distinct unsigned permutations. Including G_0 gives $2d+1$ skeletons; the two-dimensional nine-term LCU uses five. A compiler that separates permutation synthesis and diagonal sign predicates replaces duplicated skeleton work by*

$$\sum_{k,s} 2C_{\mathrm{perm}}(k,s) \quad \longmapsto \quad \sum_{k,s} \big[C_{\mathrm{perm}}(k,s) + C_{\partial}(k,s)\big]. \tag{23}$$

Proof. The factor $D^{\partial}_{k,s}$ changes signs without changing nonzero positions. One skeleton per orientation is therefore sufficient; two orientations on each of d axes give the count. Equation (23) is the work model for the stated compiler and its selected gate lowering. □

Theorem 10 (Two-term directed-interface completion)**.** *Either directed interface in Eq.* (20) *satisfies*

$$D(B) = \frac{r_k}{2}(G_+ + G_-) \tag{24}$$

for Hermitian self-inverse signed permutations $G_\pm$.

Proof. Let $S = D(B)/r_k$, let P project onto its inactive states, and set $G_\pm = S \pm P$. The completion proof above applies, including the negative sign of the active interface entries. □

Proof of Theorem 3. The symmetric identity uses $X|+\rangle = |+\rangle$. The directed identity follows by multiplying $D(A)$ by the vector $|1\rangle v$. All three fixed maps are isometries. Applying the corresponding output isometry to $Ax(\theta) - b - q$ therefore preserves its norm. This is a pointwise identity in every trainable parameter, including any parameters of q, so all existing derivatives agree. No additional variational degree of freedom is introduced by the lift. □

A.2 Distributed residual and complete catalog

Let L_i be the Laplacian of a symmetric row graph. With the notation of the main text,

$$q_{ij} = \sum_k (L_i)_{jk} z_{ik}, \tag{25}$$

$$r_{ij} = A_{ij}x_{ij} - b_i/m_i - q_{ij}, \tag{26}$$

$$C_{\mathrm{local}} = \sum_{ij} \|r_{ij}\|^2, \tag{27}$$

$$\sum_j r_{ij} = \sum_j A_{ij}x_{ij} - b_i, \tag{28}$$

$$C_{\mathrm{weighted}} = \sum_i m_i \sum_j \|r_{ij}\|^2. \tag{29}$$

This uses the row/column coordination construction of Shen et al. (2026). The pointwise identity below quantifies the coordination penalty associated with the chosen coordination ansatz.

Proof of Theorem 4. Suppress the row index and write $y_j = \bar{y} + (y_j - \bar{y})$. The tuples $(y_j - \bar{y})_j$ and $(q_j)_j$ both sum to zero; the latter property follows from the zero column sums of L. Their difference is orthogonal to the constant tuple $(\bar{y})_j$. Hence

$$\sum_j \|y_j - q_j\|^2 = m\|\bar{y}\|^2 + \sum_j \|(y_j - \bar{y}) - q_j\|^2.$$

Multiply by m and use $m\bar{y} = \sum_j y_j$. For a connected symmetric graph, the range of L is the zero-sum subspace, so unrestricted coordination realizes $q_j = y_j - \bar{y}$. If solution replicas also agree in each column, $\sum_j y_j$ is the original block-row residual. For unequal row counts m_i, the displayed weighting is needed. □

For a symmetric column Metropolis matrix W, one implemented synchronous block-local update has the form

$$\theta_{ij}^{x,+} = \sum_k W_{ik}\theta_{kj}^x - \eta\,\partial_{\theta_{ij}^x} C_{\mathrm{local}}, \qquad \rho_{ij}^{+} = \sum_k W_{ik}\rho_{kj} - \eta\,\partial_{\rho_{ij}} C_{\mathrm{local}}, \tag{30}$$

$$\phi_{ij}^{+} = \phi_{ij} - \eta\,\partial_{\phi_{ij}} C_{\mathrm{local}}, \qquad \sigma_{ij}^{+} = \sigma_{ij} - \eta\,\partial_{\sigma_{ij}} C_{\mathrm{local}}. \tag{31}$$

Derivatives include every objective term depending on that variable. Row neighbors exchange coordination parameters and reduced overlap statistics; column neighbors exchange common-coordinate solution parameters. A measurement-distributed global-gradient round instead broadcasts one shared classical parameter point, evaluates disjoint shards of the global derivative catalog, reduces one gradient, and applies one optimizer update. Both exchange classical data. The block-local iteration and diagnostic consensus replicas used beside a shared update are distinct protocols.

Proposition 11 (Reference local-cost catalog)**.** *For L operator terms and Δ row neighbors, the declared reference catalog contains*

$$Q_{\mathrm{local}}(L,\Delta) = 2L^2 + 2L(\Delta+2) + (\Delta+1)(\Delta+2) \tag{32}$$

scalar real/imaginary components before further exact simplification.

Proof. The categories contain $2L^2$ action-norm components, $2L$ action-RHS components, $2L(\Delta+1)$ action-coordination components, $\Delta(\Delta+1)$ coordination-pair components, and $2(\Delta+1)$ RHS-coordination components. Their sum gives the formula. □

Coordination self-norms are known classical constants; Hermitian symmetry leaves $\Delta(\Delta+1)$ off-diagonal real/imaginary components. The action-norm reference retains its ordered $2L^2$ tally before exact simplification. This reference count and the executed active catalog serve different purposes.

For a single fixed-vector residual $C = \|\rho Ax - d\|^2$,

$$C = \rho^2 \sum_{\ell,m} c_\ell^* c_m \langle O_\ell x | O_m x\rangle - 2\rho\ \mathrm{Re}\sum_\ell c_\ell \langle d | O_\ell x\rangle + \|d\|^2. \tag{33}$$

Its unsimplified cost catalog is $M_{\mathrm{cost}} = 2L^2 + 2L$. The notation $M_{\mathrm{coord}} = 2L^2$ in the static action-gradient experiments identifies their particular scalar probe catalog. M_{grad} is the actual complete-gradient active count after derivative construction and simplification. Coincident Pauli products, conjugate partners, identity contributions and zero coefficients determine the simplifications applied to each reference catalog.

For P angle coordinates, the four-cost rule below requires up to $(4P+2)Q_{\mathrm{local}}$ tasks including base and updated objectives. Componentwise differentiation can use a smaller plan, whose active count is recorded explicitly. For two diagonal agents with $L = 9$ and two directed-interface agents with $L = 2$, each with $\Delta = 1$, a base objective contains $2(222) + 2(26) = 496$ declared tasks.

A.3 Correct derivatives of the complete residual

In this subsection, C denotes the unnormalized local residual E of the main text. The global relative objective divides by the fixed $\nu = \|b\|^2$; the distributed objective also applies the row weights m_i.

Fix an angle occurring once as $R(\theta) = \exp(-i\theta G/2)$, where $G^2 = I$, with every other angle and amplitude held fixed. For the prepared state, the product rule gives $\partial_\theta x(\theta) = x(\theta+\pi)/2$. The quadratic part $Q(\theta) = \rho^2\langle x|A^\dagger A|x\rangle$ has frequencies in $\{0,1\}$, while the linear part $F(\theta) = -2\rho\,\mathrm{Re}\langle d|Ax\rangle$ has frequencies in $\{1/2\}$. Therefore

$$Q'(\theta) = \tfrac{1}{2}[Q(\theta+\pi/2) - Q(\theta-\pi/2)], \tag{34}$$

$$F'(\theta) = \tfrac{1}{2\sqrt{2}}[F(\theta+\pi/2) - F(\theta-\pi/2)] = \tfrac{1}{4}[F(\theta+\pi) - F(\theta-\pi)]. \tag{35}$$

The source d can include fixed coordination vectors. The same construction applies to a coordination angle while the remaining vectors are held fixed. Parameters occurring in several gates are differentiated occurrence by occurrence and summed.

For $C = Q + F + C_0$, a rule using complete costs is

$$C'(\theta) = \tfrac{1}{2}[C(\theta + \pi/2) - C(\theta - \pi/2)] + \frac{1 - \sqrt{2}}{4}[C(\theta + \pi) - C(\theta - \pi)]. \tag{36}$$

To verify it, apply the right side to $\exp(i\omega\theta)$. Its multiplier is $i[\sin(\omega\pi/2)+(1-\sqrt{2})\sin(\omega\pi)/2] = i\omega$ for $\omega = 0, \pm 1/2, \pm 1$. This is a finite-Fourier instance of generalized parameter shift (Wierichs et al., 2022). The first difference multiplies the half-frequency derivative by $\sqrt{2}$; the second difference restores its exact coefficient. The frequency-resolved rule defines the complete derivative target.

An equivalent direct differentiated-overlap construction is

$$\partial_{\theta_k} C = 2\rho^2 \operatorname{Re}\langle Ax | A\partial_{\theta_k} x\rangle - 2\rho \operatorname{Re}\langle d | A\partial_{\theta_k} x\rangle. \tag{37}$$

For the log scale $\lambda = \log \rho$,

$$\partial_\lambda C = 2\rho^2 \langle Ax | Ax\rangle - 2\rho \operatorname{Re}\langle d | Ax\rangle. \tag{38}$$

The fixed normalization C/ν, with $\nu = \|b\|^2$, divides scalar and derivative weights by ν. The effective local vector $d = b_i/m_i + q_{ij}$ can depend on coordination parameters; its norm is not used as a trainable denominator. All dependence through q_{ij} remains in the residual derivatives.

A.4 Allocation, complete weights, and dimension dependence

Operator sampling (Arrasmith et al., 2020) and adaptive shot optimizers (Kübler et al., 2020) use measurement budgets as optimization variables. The following calculation applies independent-component allocation to the complete residual and derivative catalogs, including their coefficient weights and execution costs.

Proof of Proposition 6. Independence gives Eq. (7). For $a_j = |w_j|\sigma_j$, Cauchy-Schwarz gives $(\sum_j a_j)^2 \le (\sum_j a_j^2/s_j)(\sum_j s_j)$, with equality for

$$s_j^* = S\frac{|w_j|\sigma_j}{\sum_k |w_k|\sigma_k}, \tag{39}$$

$$V_* = \frac{(\sum_j |w_j|\sigma_j)^2}{S}. \tag{40}$$

For time allocation apply the same inequality to $a_j/\sqrt{s_j}$ and $\sqrt{\tau_j s_j}$:

$$(\sum_j a_j\sqrt{\tau_j})^2 \le \Big(\sum_j a_j^2/s_j\Big)T.$$

Equality gives $s_j = T(a_j/\sqrt{\tau_j})/\sum_k a_k\sqrt{\tau_k}$. Components with zero weighted variance are interpreted by a limiting allocation or evaluated as known constants when their means are known. □

The formulas above are continuous optima. For integer $S > M$, reserving one shot per active component and assigning $1 + \lfloor (S-M)p_j \rfloor$, then distributing the remainder, gives $s_j \ge (S-M)p_j$. Thus choosing $p_j \propto |w_j|$ yields the rigorous integer bound $\|w\|_1^2/(S-M)$; replacing p_j by normalized $|w_j|\sigma_j$ gives $V_\sigma^2/(S-M)$. The experiments use their actual integer allocations in Eq. (7). The condition $S \ge M$ belongs to this minimum-one-shot policy. Randomized importance-sampling estimators can operate under other budget conditions.

For adaptive estimation, pilot outcomes determine the final allocation. Final estimates use fresh samples, so conditional expectation given the pilot equals Z. Total variance is the expectation of the conditional variance; the variance of the conditional mean is zero. Pilot shots are charged to the total budget; fresh final samples form the reported means. The oracle optimum supplies the known-variance benchmark for evaluating the pilot policy.

Proposition 12 (Structured coefficient norms)**.** *The general completion has* $\lambda_1 = |\alpha_0| + \sum_s |\alpha_s|$. *For the heat block in Eq.* (18)*,*

$$\lambda_1 = \mu + 4\sum_k r_k, \qquad \lambda_2^2 = (\mu + 2\sum_k r_k)^2 + \sum_k r_k^2.$$

A directed axis-k interface has $\lambda_1 = r_k$ *and* $\lambda_2 = r_k/\sqrt{2}$.

Proof. Each relation coefficient is divided between two completions of magnitude $|\alpha_s|/2$. A heat axis contributes four coefficients of magnitude $r_k/2$. Taking their first and second norms gives the formulas. □

Proof of Corollary 7. All coefficients are real. Expand a local residual as $r = \sum_\nu a_\nu |u_\nu\rangle$ with normalized u_ν. The solution-action coefficients contribute total absolute mass $a = |\rho|\lambda_1$. Normalizing the RHS and coordination states contributes $b_* = \|b_i/m_i\| + \sum_k |(L_i)_{jk}\sigma_{ik}|$. Thus $\sum_\nu |a_\nu| = H$. The ordered real-overlap expansion of $\|r\|^2$ has total absolute weight at most $\sum_{\nu,\omega} |a_\nu a_\omega| = H^2$. Removing known terms and combining identical expectations cannot increase this norm. Coefficient-proportional allocation gives variance at most H^4/S in the continuous allocation model.

For a solution angle, the action-norm part has weight mass at most a^2. Its two shifts, each with prefactor $1/2$, retain that bound. The cross term with RHS and coordination has weight mass at most $2ab_*$. Its two shifts have prefactor $1/(2\sqrt{2})$, giving at most $\sqrt{2}ab_*$. Hence $B_k = a^2 + \sqrt{2}ab_*$. Equation (38) gives $B_\lambda \le 2a^2 + 2ab_* = 2aH$.

For an angle in a coordination state with coefficient mass $c = |(L_i)_{jk}\sigma_{ik}|$, the same argument gives $c^2 + \sqrt{2}c(H - c)$; its self-norm term is in fact constant and can be removed. Its log-scale bound is $2cH$. For a parameter shared by finitely many gate occurrences, summing the occurrencewise derivative expansions bounds the weight mass by the sum of their bounds. For direct scale rather than log-scale optimization, replace the affected amplitude mass by its scale derivative in Eq. (37). These observations give bounded coordinate envelopes whenever the stated scales and gate multiplicities are bounded.

For unbiased coordinate estimates, $\mathbb{E}\|\widehat{g} - g\|^2 = \sum_k \mathrm{Var}(\widehat{g}_k)$. Giving coordinate k budget S_k yields $\sum_k B_k^2/S_k$; setting $S_k = S/P$ proves Eq. (9). If $B_k \le B$, this is P^2B^2/S for a fixed complete-gradient budget, or PB^2/s when every coordinate receives a fixed budget s. Across a fixed number of weighted agents, their coordinate weight bounds sum before applying the same result. □

For a normalized objective C/ν, the cost and gradient variance bounds are divided by ν^2. Dimension-uniform conclusions therefore require ν bounded away from zero as well as the amplitude bounds. Finite-shot independence concerns repeated measurement outcomes at a fixed parameter state. The initialization law separately specifies the distribution of exact gradient signals for the learning experiment.

The algebraic heat limit keeps r_k fixed. In a backward Euler discretization, $r_k = \kappa_k \Delta t/h_k^2$. Fixed r_k under proportional mesh refinement corresponds to a time step decreasing quadratically with the spatial mesh. If physical Δt is fixed while h_k decreases, r_k, the coefficient norms, and the variance envelope change accordingly. This distinction is part of the parameter path for the theorem.

A.5 Operational trainability and conditional descent

Circuit ensembles and cost locality govern the distribution of exact gradients (McClean et al., 2018; Cerezo et al., 2021b). Our representation profile quantifies the conditional measurement noise at a fixed parameter state; the operational target below combines this estimator with a specified ansatz, initialization law, and residual threshold.

Fix a problem $A, b \neq 0$, an ansatz, optimizer, initialization law, measurement policy, and stopping rule. Define the original-system residual

$$R(\theta) = \min_{\alpha \in \mathbb{C}} \frac{\|\alpha A|x(\theta)\rangle - b\|}{\|b\|}. \tag{41}$$

Let B_t count every logical shot consumed through evaluation t, including allocation pilots and any shot-based stopping evaluation. For target ε and failure probability δ, define

$$S_{\varepsilon,\delta} = \inf\left\{ B : \Pr\left[\min_{t: B_t \le B} R(\theta_t) \le \varepsilon \right] \ge 1 - \delta \right\}. \tag{42}$$

The probability is over the specified initialization and measurement law; when instances are sampled it also uses the explicitly specified instance distribution. Set the infimum to infinity when the target

is unattainable. Finite campaigns report the observed success fraction and budget-censored hitting times.

Theorem 13 (Conditional smooth-SGD descent)**.** *For fixed initialization θ_0, assume C is bounded below by $C_{\inf}$ and has $\mathcal{L}$-Lipschitz gradient on a region containing the SGD iterates and update segments. Given the history $\mathcal{F}_t$, assume $\mathbb{E}[\widehat{g}_t|\mathcal{F}_t] = g_t = \nabla C(\theta_t)$ and $\mathbb{E}[\|\widehat{g}_t - g_t\|^2|\mathcal{F}_t] \le V_t$. For $\theta_{t+1} = \theta_t - \eta\widehat{g}_t$ and $0 < \eta \le 1/\mathcal{L}$,*

$$\mathbb{E}[C(\theta_{t+1})|\mathcal{F}_t] \le C(\theta_t) - \frac{\eta}{2}\|g_t\|^2 + \frac{\mathcal{L}\eta^2}{2}V_t, \tag{43}$$

$$\frac{1}{T}\sum_{t<T}\mathbb{E}\|g_t\|^2 \le \frac{2(C(\theta_0) - C_{\inf})}{\eta T} + \frac{\mathcal{L}\eta}{T}\sum_{t<T}\mathbb{E}V_t. \tag{44}$$

Proof. The smoothness inequality is $C(\theta_t - \eta\widehat{g}_t) \le C(\theta_t) - \eta\langle g_t, \widehat{g}_t\rangle + (\mathcal{L}\eta^2/2)\|\widehat{g}_t\|^2$. Conditional unbiasedness gives expected inner product $\|g_t\|^2$ and expected squared norm at most $\|g_t\|^2 + V_t$. The coefficient of $\|g_t\|^2$ is $-\eta(1 - \mathcal{L}\eta/2) \le -\eta/2$. Taking expectations, summing, and using the lower bound proves the result. Uniform $V_t \le V$ gives Eq. (14). This is the usual smooth nonconvex stochastic-descent argument (Ghadimi & Lan, 2013), with its noise term supplied by the representation profile. □

The theorem quantifies stationarity for the stated smooth-SGD dynamics. For the residual target in Eq. (42), the problem specification additionally includes ansatz attainability and the relation between gradient size and objective error. The measured learning protocols use Adam (Kingma & Ba, 2015) and L-BFGS-B (Byrd et al., 1995), with source instances and stopping rules fixed as part of the evaluation design.

A.6 Generic worker scheduling

Theorem 14 (Independent measurement-graph execution)**.** *At a fixed parameter point, let M independent component batches have execution times at most $\tau_{\max}$ and classical reduction cost T_{reduce}. Round-robin assignment to G workers covers each batch exactly once, with task counts differing by at most one, and*

$$T_G \le \lceil M/G\rceil\tau_{\max} + T_{\text{reduce}}. \tag{45}$$

Proof. The residue classes of the ordered task indices modulo G partition the catalog, with cardinalities $\lfloor M/G\rfloor$ or $\lceil M/G\rceil$. Each worker's assigned work takes at most the displayed compute time. Classical reduction then completes the result. □

Shot-weighted batch costs determine loads and runtimes for each assignment. For pilot-based allocations, final-stage load-aware ownership is chosen after the pilot. Logical critical load is the pilot maximum plus the final maximum. Greedy assignment by known shot load and divisible shot batches therefore provide stronger load controls than count balancing alone.

B Independent Native Replication and Measurement Reuse

This cohort fixes four fresh independent starts before execution and compares S/reuse-on, M/reuse-on, D/reuse-on, S/reuse-off and D/reuse-off at each start. It uses the same operator, source, depth-1 seven-angle circuit, eight-coordinate budget and Adam schedule as Section 4.1. The twenty arms each complete 512 updates and 512M actual NVIDIA GPU circuit-simulator measurements. Pair labels 1-4 correspond to registered start identifiers 0-3. Empirical decimals are rounded for display; comparisons, uncertainty and threshold decisions use unrounded records. The initial two-pair study in Appendix C is a separate cohort; its outcomes are not pooled into the statistics below.

Table 1: All four new native pairs, reuse on. Every arm uses 512M shots. Stable costs require the complete checkpoint window and final-state check; > 512M denotes no stable success within budget.

	Final raw residual r			Shots to stable 1%		
Pair	S	M	D	S	M	D
1	.01003	.02131	.007602	> 512M	> 512M	330M
2	.008314	.02864	.007424	340M	> 512M	240M
3	.009325	.01662	.01025	470M	> 512M	> 512M
4	.008454	.02776	.01142	340M	> 512M	> 512M

B.1 Complete paired endpoints and estimator contrasts

Reuse-off fixes every historical term mean h to zero while retaining the same weighted categorical estimator. Reuse-on uses only the preceding update's charged measurements. Every gradient coordinate is estimated at every update; no exact gradient enters the sampled optimizer. The reported output is the final state. Stable .01 success requires five passing checkpoints, ten updates apart, and a passing final state. Table 2 contains every registered arm; Table 1 gives all reuse-on target costs. All eight reuse-off arms finish above the target.

Table 2: All twenty final raw residuals at 512M native measurements per arm. On/off denotes past-count reuse. The four starts are the independent units.

Pair	S/on	M/on	D/on	S/off	D/off
1	0.01003	0.02131	0.007602	0.01600	0.01289
2	0.008314	0.02864	0.007424	0.01112	0.01192
3	0.009325	0.01662	0.01025	0.01750	0.01457
4	0.008454	0.02776	0.01142	0.01419	0.01980

With reuse, S/M/D mean residuals are 0.009030/0.02358/0.009172, with sample standard deviations 0.0008004/0.005673/0.001976. S lowers the four paired residuals relative to M by 53.0%, 71.0%, 43.9% and 69.5%; 61.7% is the reduction of the cohort means. The stable-success counts are 3/4, 0/4 and 2/4. S in Pair 1 returns .01003 and therefore remains a non-hit under the registered rule.

Table 3: Paired final-residual contrasts across four starts. Intervals are two-sided 95% Student-t intervals with three degrees of freedom. Interaction is the S-minus-D contrast with reuse minus that without reuse.

Contrast	Mean	95% interval
S minus M, reuse on	-0.01455	$[-0.02458, -0.004525]$
S minus D, reuse on	-0.000142	$[-0.003841, 0.003557]$
S minus D, reuse off	-0.000093	$[-0.006614, 0.006427]$
Interaction	-0.000049	$[-0.004655, 0.004558]$
S, on minus off	-0.005673	$[-0.009182, -0.002163]$
D, on minus off	-0.005624	$[-0.008629, -0.002619]$

The paired comparisons support the benefit of reuse for both structured estimators. Their observed mean residuals are close, with S-minus-D difference intervals spanning zero as reported in Table 3. Every non-hit retains its full measurement budget as a censoring boundary; a censored value is not a successful target cost. The four-start uncertainty calculation concerns final-residual differences, separately from the reported empirical success fractions.

B.2 Same-point conditional variance and count audit

For each arm, the first and final sampled updates are independently checked against exact conditional moments. The variance ratio holds parameters, coordinate budgets and term-selection probabilities fixed, changing only the historical baseline in Eq. (13). Thus it isolates the estimator mechanism from differences between trained endpoints.

Table 4: Final-update conditional variance divided by the same-point zero-baseline variance. All listed arms use reuse; off-arm ratios are one.

Pair	S	M	D
1	0.2094	0.8425	0.2066
2	0.2094	0.8427	0.2065
3	0.2093	0.8425	0.2068
4	0.2092	0.8426	0.2070

All 10240 updates pass independent reconstruction from integer measurement counts, historical corrections and optimizer states. The maximum count-based gradient reconstruction discrepancy is 1.27×10^{-14}; conditional expectation discrepancies at the checked updates are below 4.33×10^{-14}. These numerical checks establish the executed estimator semantics independently of the measured learning outcomes.

C Initial Compact-circuit Precision: Protocol and Full Results

The two native pairs share the smooth-input system and compact circuit/estimator in Eqs. (6) and (12). They are two fresh starts of that system.

C.1 Circuit construction, source and initialization

On a grid with n_x, n_y interior points, coordinates are $x_i = (i+1)/(n_x+1)$ and $y_j = (j+1)/(n_y+1)$ and $b_{ij} \propto \exp(.3x_i + .5y_j)$. Binary indices factor this source into single-bit amplitude ratios: if bit q contributes 2^q to an axis index with exponent coefficient c, its amplitude ratio is $\exp(c2^q/(n_k+1))$. A single $R_y(2\arctan(\exp(c2^q/(n_k+1))))$ prepares that bit, with the normalization supplied by the product state. This gives exact source preparation with one rotation per data bit. The source reflection is preparation, a phase about the all-zero basis state, and inverse preparation.

The data-generator maps E, O_+, O_- are permutations with boundary signs. E exchanges adjacent even/odd indices; O_+ exchanges odd/even interior indices and fixes the endpoints; O_- negates those endpoints. All are Hermitian involutions. The native E rotation acts on the least-significant axis bit. The $O_\pm$ rotations conjugate it by reversible index shifts and add controlled endpoint corrections. Source reflection computes the all-zero predicate onto a work qubit, applies the source phase and uncomputes the predicate. Controlled and adjoint forms preserve the same maps and return work qubits to zero. No dense per-term matrix gate is used in the native grid runtime.

Each layer applies E, O_+, O_- on the column axis, then on the row axis, then the source-reflection rotation. Fixed $V^\dagger$ before and V after all layers cancel at zero angles. The orthogonal part of the derivative at that point is $(G - \langle b|G|b\rangle)b/2$, which connects the grid structure to real solution corrections while retaining a source-phase direction. The native study uses depth one, radius .03 independent angle perturbations, and initial log scale zero. Both input-only seeds are retained in the frozen configurations. Their original raw residuals .03880/.05095 both exceed .01. No trained vector, inverse solve or exact gradient enters initialization or a sampled update.

C.2 Complete gradients with past-count correction

Write $g_k = o_k + \sum_j w_{kj}\mu_{kj}$, set $b_k = \sum_j |w_{kj}|$, and allocate the total 1M shots across all eight coordinates using the coefficient-aware integer rule. Per-coordinate term selection has full support on nonzero weights. For fixed previous empirical means h_{kj} and $\delta_k = \sum_j w_{kj}(\mu_{kj} - h_{kj})$, conditional variance of Eq. (12) is

$$\frac{1}{S_k}\left[b_k \sum_j |w_{kj}|(1 - 2h_{kj}\mu_{kj} + h_{kj}^2) - \delta_k^2\right]. \tag{46}$$

This follows by expanding $(Y-h)^2$ using $Y^2 = 1$ and averaging over term selection. The first update sets $h = 0$, recovering Eq. (11). Previous counts remain fixed until all current coordinates have been reduced, so the correction is conditionally unbiased. Unobserved terms have zero baseline;

the deterministic correction sums over the full baseline support. Its current coefficient weights can change with learned scale. No present exact expectation supplies h.

Each generator satisfies $G^2 = I$. The quadratic residual uses the ordinary $\pm\pi/2$ difference and the linear overlap uses the same shifts with factor $1/(2\sqrt{2})$, as in Eq. (10). The learned log-scale derivative is analytic in the same measured residual components. All eight coordinates receive fresh counts at each update. Each paired experiment uses one GPU. Compiled circuits are reused, so a selected measurement task and a compiled circuit are distinct quantities.

C.3 Complete outcomes and target costs

The common 512-update Adam schedule holds .003 for 30% of its duration, then decays to .00006, with $\beta_1 = \beta_2 = .9$ and $\epsilon = 10^{-8}$. All arms finish 512 updates and 512M measurements; the last state is the output. Stable success requires five scheduled checkpoints at raw residual at most .01 and a passing final state. Checkpoints are ten updates apart; first crossings are checked at every update.

Table 5: Complete 512M-shot outcomes and target events. First and stable costs are millions of measurements; no crossing occurs anywhere on either Pauli trajectory. Every arm finishes all 512 updates.

Pair	Arm	Final raw	Final aligned	Best raw	First 1%	Stable 1%
1	S	.006940	.006926	.006940	183	230
1	M	.02816	.02792	.02816	> 512	> 512
2	S	.007372	.007363	.007200	234	290
2	M	.02165	.02149	.02165	> 512	> 512

Structured stable-hit costs are 230M/290M, while neither Pauli trajectory crosses .01 by 512M. The Pauli/structured cost-ratio lower bounds are 512/230 and 512/290, reported conservatively as > 2.22 and > 1.76. Final raw residual reductions are 75.4%/66.0% at equal shots. Matched exact-gradient controls return .006501/.006490 under the same initial vectors and schedule. RHS-only raw/aligned residuals are .03971/.03895. These are two observed pairs under the fixed protocol.

Time-to-target observations. A common 6000-active-second observation window includes compilation, measurement, optimization and residual diagnostics; queue idle time is excluded. Recorded updates inside this window establish the time bounds below, while the equal-shot analysis uses all 512 updates of every arm. S first crossings occur at updates 183/234; its stable windows end at 230/290, at 760.2/1062.4 seconds. M has no crossing by its last in-window observations, at 5775.4/5918.1 seconds (updates 33/35). Comparing M's first crossing against S's confirmed stable window gives conservative time-ratio lower bounds $> 7.59/> 5.57$.

C.4 Matched direct-square control

The direct-square control in Eq. (15) shares each pair's original-system inputs, initial vector, compact circuit, 512-update Adam schedule, 1M-shot complete-gradient budget and past-count correction. Its draw seeds were fixed after the S/M study, before D execution. It adds one baseline arm to each observed pair, without adding independent starts. The quadratic catalog contains 33 indexed terms including the analytic identity; the linear catalog retains nine terms. The seven angles give 615 shifted measurement tasks, compared with S675. First-draw D/S variance ratios are 1.027/1.027 with zero count baselines; subsequent variance uses each run's observed count history.

Table 6: Complete native direct-square outcomes: 512 updates and 512M shots per arm. First and stable costs are millions of measurements. All displayed residuals use the final returned state; best raw residual is separate.

Pair	Final raw r	Final aligned R	Best raw r	First 1%	Stable 1%
1	0.007213	0.007186	0.006929	256	310
2	0.008119	0.008109	0.007903	182	350

All 1024 D updates use actual NVIDIA circuit-simulator measurements and pass independent reconstruction of counts, correction, optimizer and original state. Checkpoint recovery preserves parameters, Adam moments, previous counts, random streams and schedule position.

C.5 Independent compact-circuit measurement emulation

This separately frozen validation contains four fresh paired starts and four matched exact controls at each $N = 8192$ (64×128, depth 2) and $N = 32768$ (128×256, depth 1). All cases use $r_x = r_y = .25$, normalized smooth exponential input, independently sampled angles in $[-.03, .03]$, and initial log scale zero. There are 14/7 angles plus scale. Adam uses $\beta = (.9, .9)$ and $\epsilon = 10^{-8}$. At $N = 8192$, 2048 updates hold rate .001 for 30% of updates then decay to .00002; at $N = 32768$, 512 updates hold .003 for 30% then decay to .00006. S/M share every setting within a pair. Draw seeds are independent of development and native runs.

Each sampled update uses 2M emulated measurements for the complete gradient, with exact coefficient-mixture binary marginals from Eq. (11); no past-count correction is applied. Total budgets are 4.096B/1.024B per sampled trajectory. Exact controls use no measurements. This cohort tests independent initialization under the raw estimator and is reported separately from the native corrected estimator. Success requires five consecutive saved checkpoints at $r \leq .01$ and a passing final state, with checkpoints every ten updates and at the endpoint.

Table 7: All 24 independent compact-circuit validation cases. Entries are final raw residuals; parentheses give stable-hit update when the final state also passes. A dash means no stable final success.

N	Start	StructVQLS	Merged Pauli	Exact control
8192	1	.009605 (730)	.01538 (-)	.009156 (140)
	2	.009549 (1160)	.01163 (-)	.009158 (160)
	3	.01028 (-)	.01432 (-)	.009156 (140)
	4	.009338 (440)	.009407 (1720)	.009155 (110)
32768	1	.009500 (430)	.01680 (-)	.006502 (100)
	2	.009504 (460)	.03182 (-)	.006493 (110)
	3	.01052 (-)	.02521 (-)	.006495 (110)
	4	.009280 (420)	.02638 (-)	.006498 (120)

S attains stable final success in 3/4 starts at each size, versus M 1/4 and 0/4; exact controls attain 4/4 at both sizes. Every paired final residual favors S. At $N = 8192$, start 3 has an earlier passing five-checkpoint window ending at update 1280 but fails the final-state requirement, and is counted as unsuccessful. Its final residual .01028 and the unsuccessful $N = 32768$ residual .01052 are retained. All 24 update chains pass the source-side audit; independent reconstruction verifies input-only initial vectors, final original-system residuals and every reported stable window.

C.6 Validation and evidence accounting

Independent small-system matrix-exponential, controlled/adjoint and work-qubit checks verify the data maps, source phases and separate gradient frequencies. Full-coordinate finite differences of the matched exact controls at steps 0/256/512 have maximum error 1.18×10^{-10}. For the S/M comparison, count reduction, lag-one provenance, Adam transitions, seeds, original-state residuals and observer gradients are independently reconstructed for all 2048 updates and 2.048B shots, including the reused structured trajectories. The 1024 D updates are audited separately in Section C.4. Maximum count-gradient, raw/aligned residual and observer-gradient discrepancies are 1.62×10^{-14}, 1.88×10^{-15} and 1.01×10^{-14}, respectively. Independent state reconstruction uses explicit pair blocks and source projections, then evaluates the original sparse heat system. Exact observers never supply sampled training updates. At the first shared point, analytic full-gradient noise traces give merged/structured ratios 12.55/12.55. Selected tasks number 675 versus 104704/104813. Independent reconstruction verifies these first-draw quantities; later count histories enter the general variance formula above.

Frozen configurations and paired count records support independent reconstruction of every reported native and emulated outcome.

The smooth system also admits an axis-factor classical approximation: solving its two one-dimensional factors and taking their product gives raw residual .0003539 at $N = 32768$. The native measurements compare quantum-circuit training methods on this fixed-coefficient system.

C.7 Common-depth and common-budget size control

The additional scale control holds circuit depth 1, 512 Adam updates and 1M emulated shots per sampled update fixed at $N = 8192, 16384, 32768$. Each size has four new paired starts using the same smooth exponential source, $r_x = r_y = .25$, input-centered radius .03 and native-study optimizer schedule. Exact/S/M/D share the initial state and circuit. S/M/D use Eq. (11) without past-count reuse; exact controls use deterministic gradients. These are 48 complete trajectories, including twelve exact controls. This protocol varies grid size with the circuit depth and total sampling budget held constant.

Table 8: Common-depth size control. Final raw residual, mean ± sample standard deviation across four starts, with stable .01 success counts; exact controls use no measurements.

N	Method	Final raw residual, mean ± SD	Stable .01
8192	Exact	$0.009167 \pm 1.6 \times 10^{-6}$	4/4
	S	0.01080 ± 0.0014	2/4
	M	0.02060 ± 0.0069	0/4
	D	0.01139 ± 0.0011	0/4
16384	Exact	$0.007491 \pm 5.7 \times 10^{-6}$	4/4
	S	0.01027 ± 0.0023	3/4
	M	0.02474 ± 0.0017	0/4
	D	0.01214 ± 0.003	0/4
32768	Exact	$0.006494 \pm 2.4 \times 10^{-6}$	4/4
	S	0.01226 ± 0.0029	1/4
	M	0.02982 ± 0.008	0/4
	D	0.01015 ± 0.0012	1/4

The mean residual comparison favors S over M at every size. The two structured estimators have size-dependent ordering under this common policy; the native reuse study separately tests the historical-count mechanism.

D Six-source Validation: Protocol and Results

The validation uses $N = 8192, 16384, 32768$ on 64×128, 128×128 and 128×256 grids. All representations share the sparse heat matrix $A = I + .25L_x + .25L_y$, normalized source, depth-5 complex circuit, fixed plus lift and optimized log scale. Condition numbers stay near three.

D.1 SOURCES AND FROZEN TRAINING PROTOCOL

The six sources use interior coordinates $x = (j+1)/(n_x+1)$, $y = (i+1)/(n_y+1)$, row-major flattening and unit Euclidean normalization:

$$
\begin{aligned}
f_1(x,y) &= \sin(\pi x)\sin(\pi y)(1 + .3x + .2y),\\
f_2(x,y) &= \exp(.3x + .5y),\\
f_3(x,y) &= \sin(2\pi x)\sin(\pi y) + .35\sin(\pi x)\sin(2\pi y),\\
f_4(x,y) &= \sin(3\pi x)\sin(2\pi y) + .4\sin(\pi x)\sin(\pi y),\\
f_5(x,y) &= \exp\left(-\frac{(x-.28)^2 + (y-.35)^2}{2(.14)^2}\right),\\
f_6(x,y) &= \exp\left(-\frac{(x-.72)^2}{2(.11)^2} - \frac{(y-.68)^2}{2(.19)^2}\right).
\end{aligned}
$$

In order, these define `smooth_tilt`, `smooth_exponential`, `modal_21_12`, `modal_32_11`, `localized_left` and `localized_right`. They are analytic inputs, independent of A and the solution circuit.

Each of two fresh starts per source and size samples angles uniformly in $[-1, 1]$ and starts log scale at zero. A circuit layer applies R_y, R_z, R_y on each qubit followed by a controlled-Z chain, giving 196/211/226 variables including scale. All 108 sampled and 36 matched exact-control trajectories complete 3200 Adam updates with $\beta = (.9, .999)$ and $\epsilon = 10^{-8}$. The learning rate is .05 for the first half, then follows a cosine to .0001. Per-update shots increase linearly and sum to exactly 400M per sampled trajectory, averaging 125k per update; coefficient-proportional integer allocation covers all coordinates and four logical workers. The final raw parameters define the returned circuit; reported residual R then aligns scale and phase. The shared setting and output rule were selected by pooled performance using three of these six source functions in separate development runs at $N = 1024, 8192$, before validation. Both per-method selections agree with that shared setting. Validation starts are fresh and never reselect settings or outputs.

Stable .1 success requires all last five checks, spaced ten updates apart, to pass. Non-hits retain the full budget; exact controls have no shot charge. Pointwise 95% intervals use 10000 paired hierarchical bootstrap replicates, resampling the six sources and then their paired starts. These intervals condition on the evaluated source families and fixed coefficient regime.

D.2 MEASUREMENT EMULATION AND NUMERICAL CHECKS

For each coordinate, sampling the scalar term with probability proportional to its absolute coefficient makes its signed measurement Z binary. Its mean is $m = (g-o)/b$ and its second moment is one, so the rescaled gradient estimate from S draws has variance $(b^2 - (g-o)^2)/S$. Marginalizing the term index gives exactly a binomial count with success probability $(1+m)/2$. Sparse adjoint derivatives provide that probability for the measurement emulator. Only the resulting sampled gradient enters Adam. Allocation uses coefficient envelopes alone; the exact derivative and variance never choose an allocation or an update. Independent per-worker binomial counts sum to the same full-gradient law. This law includes term-selection noise; it is distinct from allocating a fixed number of shots to every individual component.

Small dense tests compare states, residuals, every signed completion, symbolic Pauli coefficients and full gradients through $N = 128$. Five thousand explicit term/outcome draws verify the marginal law. Independent depth-5 directional derivative checks at all three target sizes have maximum error 1.43×10^{-9}. Checkpoint/resume reproduces optimizer state, random-number state, parameters and all histories exactly. Numerical checks and circuit-execution checks are recorded separately.

D.3 COMPLETE VALIDATION OUTCOMES

Table 9: Scale-up mean aligned residuals at 400M shots and matched exact controls. Parentheses give stable .1 successes out of twelve. All sampled methods use the same independently selected setting and randomized policy.

N	Structured	Merged Pauli	Direct square	Exact gradient
8192	.2945 (0)	.3989 (0)	.2909 (0)	.1673 (7)
16384	.2898 (0)	.4430 (0)	.2905 (0)	.1675 (7)
32768	.2926 (0)	.4665 (0)	.3004 (0)	.1726 (6)

Table 10: Complete scale-up source accounting: aligned residual of the final iterate, averaged over the two frozen starts per source and representation. All sampled methods receive 400M shots; exact controls use matched updates.

N	Source	Structured	Merged Pauli	Direct square	Exact
8192	smooth_tilt	.2186	.3828	.2185	.0766
8192	smooth_exponential	.1273	.2693	.1365	.0450
8192	modal_21_12	.3924	.4799	.4038	.2896
8192	modal_32_11	.4867	.5387	.4516	.4194
8192	localized_left	.2837	.3579	.2791	.0971
8192	localized_right	.2585	.3646	.2557	.0760
16384	smooth_tilt	.2351	.3856	.2168	.0560
16384	smooth_exponential	.1436	.3396	.1500	.0393
16384	modal_21_12	.4000	.5390	.4005	.3410
16384	modal_32_11	.4769	.5850	.4888	.4069
16384	localized_left	.2416	.4004	.2411	.0951
16384	localized_right	.2416	.4084	.2458	.0669
32768	smooth_tilt	.2291	.4240	.2403	.0706
32768	smooth_exponential	.1538	.3814	.1613	.0348
32768	modal_21_12	.3991	.5564	.4079	.3400
32768	modal_32_11	.4718	.6120	.4862	.3923
32768	localized_left	.2674	.3922	.2593	.1067
32768	localized_right	.2342	.4326	.2475	.0911

The main-text paired intervals exclude zero against merged Pauli and include zero against direct square at all three sizes. All sampled stable .1 counts are zero. The modal sources also give the largest exact-control residuals, separating optimization effects from sampling noise. RHS-only aligned residuals are at most .05433; sparse conjugate-gradient references reach 10^{-10}, with the slowest measured solve taking .006646 seconds. The reported resource comparisons concern quantum-circuit solvers.

E SPATIAL ACCURACY: PROTOCOL AND COMPLETE OUTCOMES

This study supplies the solution/error heatmap in Figure 5. It uses the same heat matrix and generic depth-5 circuit as the six-source study, with input-centered initialization and a 256-update schedule.

Input-centered precision protocol. We retain the same operator, depth-5 ansatz and full randomized gradient, using smooth $b \propto \exp(.3x + .5y)$ and localized $b \propto \exp(-((x - .28)^2 + (y - .35)^2)/(2 \cdot .14^2))$ on interior coordinates. The initial product state matches the single-bit marginals of $|b|^2$; independent angle perturbations have radii .03/.1 for smooth/localized inputs. This uses the supplied RHS without solving $Ax = b$. The shared setting is fixed by pooled mean raw residual on separate development cases: Adam $\beta = (.9, .9)$, 256 updates and rate .001, decaying after 30% to .00005. Frozen validation crosses both sources, three sizes, three new paired starts and 32k/250k/2M shots per update, with exact controls: 126 trajectories. All 196/211/226 coordinates are updated, and the returned iterate is used. Precision requires both raw $\sqrt{C}$ and aligned residual to pass; stable success requires five checks ten updates apart and a passing returned state.

The input product is prepared in the final two layers using cancelling controlled-Z chains. Initial log scale is zero; every angle, including R_z, receives a perturbation. Adam uses $\epsilon = 10^{-8}$, and exact controls share the same initialization and complete schedule. The sampled updates use the uncorrected binary-mixture law described in Appendix D. Checkpoints lie on $0, 10, 20, \ldots, 250$; update 256 must pass but does not replace a scheduled check. Independent scalar gates and a separately assembled sparse operator reconstruct all 3402 saved validation states with zero recorded residual discrepancy. The main heatmap fixes the first registered restart for each source at $N = 32768$ and 2M shots/update; it does not rank starts by their outcomes.

Table 11: Mean raw residual and stable .1 counts (parentheses) across three fresh starts after 256 updates. S/M: structured/merged Pauli. Figure 7 includes all budgets and full ranges.

	32k shots/update		2M shots/update	
N	S	M	S	M
Smooth exponential input				
8192	.09585 (3)	.1335 (0)	.05660 (3)	.06683 (3)
16384	.09380 (3)	.1351 (0)	.04892 (3)	.06994 (3)
32768	.1013 (1)	.1428 (0)	.04536 (3)	.07238 (3)
Localized input				
8192	.5345 (0)	.6342 (0)	.3989 (0)	.4423 (0)
16384	.4940 (0)	.6172 (0)	.3635 (0)	.4205 (0)
32768	.5408 (0)	.6757 (0)	.3873 (0)	.4538 (0)

Accuracy and target cost. At $N = 32768$, 2M shots/update gives stable .05 success in 3/3 structured and 0/3 merged runs. First stable passage costs 220/400/420M shots; merged remains censored at 512M. At 32k shots/update, stable .1 is attained in 3/3 structured and 0/3 merged runs at both smaller sizes, within 8.192M shots. No sampled validation run reaches stable .01. Exact smooth-input means .04898/.04155/.03653 separate optimization from sampling error. Returning b alone gives .05647/.04579/.03971: these comparisons measure improvement over the paired stochastic solver, and threshold passage alone need not improve this inexpensive baseline. Localized inputs remain harder under the same short schedule.

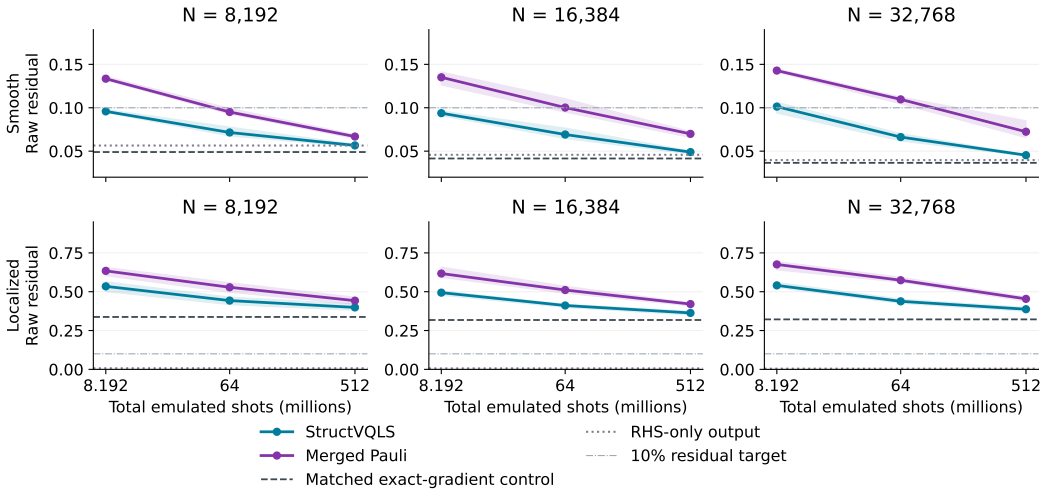

Figure 7: Measurement-emulated precision versus budget for both sources: means and full ranges over all three paired starts. Points are final outputs of separate budget experiments, not successive training checkpoints. Exact controls share initialization and 256 updates; RHS-only returns b. Sampled methods share settings, returned-state rules and the full-gradient binary-mixture law.

Optimizer robustness. The secondary merged-Pauli setting uses $\beta_2 = .999$, selected on separate development cases. It reuses every validation initial vector, the .001 learning rate, schedule, budget

and last-iterate rule, while drawing new measurement streams. Its 54 sampled cases and 18 matched exact controls are complete. The shared .9 setting is selected by pooled mean raw residual; using pooled mean squared residual instead selects .999. Figure 8 reports both settings over all 18 source/size/budget cells. The comparison assesses robustness to estimator-specific tuning; new measurement streams mean that differences between the two M cohorts do not isolate the moment coefficient alone. All 18 source/size/budget means favor structured, by 11.5-42.8%; the smooth $N = 32768$ stable .05 counts at 512M shots remain 3/3 versus 0/3.

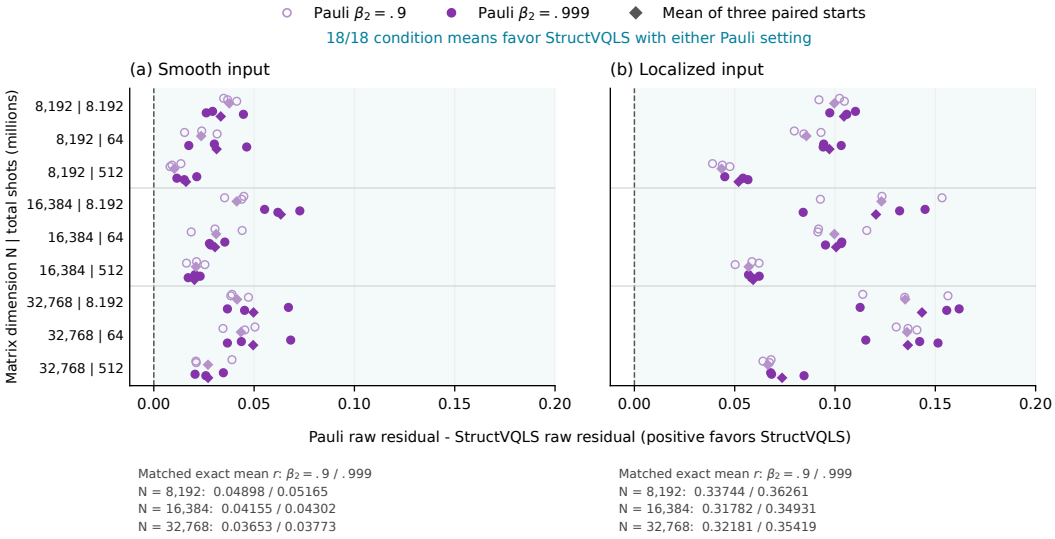

Figure 8: Optimizer robustness across all 18 source/size/budget conditions. Dots show each paired final raw-residual reduction, Pauli minus StructVQLS; diamonds show means. Positive favors StructVQLS. Both Pauli settings reuse the same three structured starts per condition, yielding 108 contrasts from 162 sampled runs. The .999 setting uses new measurement streams. Footer values retain both matched exact-control means. All sampled cases use measurement emulation.

F Measurement Mechanism and Distributed Correctness

These checks support Section 4.4: exact operator counts, weighted gradient-noise prediction and reduction of the original residual.

F.1 Operator families and finite-measurement noise

Seven stencil families at four resolutions give the 28 cases in Figure 6. At angle .43 and 200k complete-gradient shots, three fixed seeds generate 500 estimates each per representation and allocation. The main figure and Table 12 use uniform allocation; coefficient-aware and variance-aware checks use fresh samples under their stated allocations. Family (d, K) pairs are $(1,2),(1,4),(1,6),(2,4),(2,8),(2,4),(3,6)$, giving $2K+1$ completed terms. The $N = 256, 1024$ structured profiles use no dense per-term matrices. Exact Pauli coefficients are extracted recursively with tolerance 10^{-12}. The minimum-one-shot status in Table 13 refers to a fixed-component coordinate plan requiring $2L_P^2$ shots; the randomized native estimator does not require visiting every term.

Noise at learning states. The 36 exact-control trajectories from Appendix D supply 72 initial/final anchors. Three representations and 32 fresh complete-gradient draws per anchor give 216 cells and 6912 draws, each with 250k shots. The mean measured/predicted squared-error ratio is 1.001. Mean ratios at $N = 8192, 16384, 32768$ are .9991/.9999/1.003; median paired M/S variance ratios are 8.4488/9.9744/11.6247, while D/S medians are 1.0276 at each size. These compare estimators at identical states, separately from the final outcomes along their different training trajectories.

Table 12: All 28 structured-stencil cases under uniform allocation. S/P denotes structured completion followed by exact Pauli. RMSE averages three registered seeds with 500 estimates each at 200000 logical shots.

Family	Grid	N	Terms S/P	Active S/P	Uniform RMSE S/P
Shifted Poisson 1D	8	8	5/8	50/128	.03299/.07965
	16	16	5/16	50/512	.02788/.1731
	32	32	5/32	50/2048	.02444/.3663
	64	64	5/64	50/8192	.02155/.7254
Fourth-order diffusion 1D	8	8	9/11	162/242	.09711/.1812
	16	16	9/23	162/1058	.08358/.4091
	32	32	9/47	162/4418	.06989/.8673
	64	64	9/95	162/18050	.06386/1.682
Banded Toeplitz, bandwidth 3, 1D	8	8	13/13	338/338	.1168/.1635
	16	16	13/29	338/1682	.09530/.4036
	32	32	13/61	338/7442	.08476/.8391
	64	64	13/125	338/31250	.07231/1.728
Diffusion-reaction 2D	2×4	8	9/5	162/50	.03393/.01882
	4×4	16	9/7	162/98	.03599/.02841
	4×8	32	9/11	162/242	.03026/.04847
	8×8	64	9/15	162/450	.02925/.06643
Nine-point elliptic 2D	2×4	8	17/8	578/128	.1882/.09007
	4×4	16	17/16	578/512	.1943/.2029
	4×8	32	17/32	578/2048	.1670/.3823
	8×8	64	17/64	578/8192	.1666/.8192
Upwind convection-diffusion 2D	2×4	8	9/9	162/162	.07718/.08014
	4×4	16	9/13	162/338	.07974/.1385
	4×8	32	9/21	162/882	.07074/.2181
	8×8	64	9/29	162/1682	.06896/.3234
Diffusion-reaction 3D	$2 \times 2 \times 2$	8	13/4	338/32	.05679/.01068
	$2 \times 2 \times 4$	16	13/6	338/72	.05488/.02357
	$2 \times 4 \times 4$	32	13/8	338/128	.05425/.03617
	$4 \times 4 \times 4$	64	13/10	338/200	.05428/.04780

Table 13: Operator scaling through $N = 1024$. Pauli columns give exact coefficient-only term counts. $M_{\mathrm{coord}} = 2L_P^2$ at $N = 1024$. A/R denotes admitted or preflight-rejected under 200000 logical shots. Pauli strings are represented symbolically.

Family	Struct. terms	Pauli $N = 64$	Pauli $N = 256$	Pauli $N = 1024$	M_{coord}	Shot status
Shifted Poisson 1D	5	64	256	1024	2097152	R
Fourth-order diffusion 1D	9	95	383	1535	4712450	R
Banded Toeplitz, bandwidth 3, 1D	13	125	509	2045	8364050	R
Anisotropic diffusion-reaction 2D	9	15	31	63	7938	A
Nine-point elliptic 2D	17	64	256	1024	2097152	R
Upwind convection-diffusion 2D	9	29	61	125	31250	A
Anisotropic diffusion-reaction 3D	13	10	18	30	1800	A

The equal-count Toeplitz and greater-count nine-point controls isolate weights from term count. The complete table also retains the small-grid and 3D cases in which structured RMSE is higher than Pauli RMSE.

Merged products and adaptive allocation. The $N = 32$ optimized-point comparison uses four coefficient cases and five starts, giving twenty fixed probes with 46 gradient coordinates. Each estimator receives 1M shots and ten independent draws per probe. Pauli products retain their phases before merging; identity contributions are analytic. Adaptive allocation charges an independent 5% pilot and uses fresh final samples. Table 14 reports all three allocations. Logical efficiency $S/(4C)$ uses total shots S and the sum C of the pilot and final maximum worker loads; it is distinct from physical GPU speedup. Canonical and greedy ownership use the same samples.

Table 14: Optimized $N = 32$ gradient probes at 1M shots: 20 points and ten seeds, 46 coordinates. Adaptive efficiency is $S/(4C)$ with pilot and final critical loads added.

	Gradient RMSE			Adaptive efficiency	
Estimator	Uniform	Coeff.	Adaptive	Canonical	Greedy
StructVQLS	0.07778	0.03174	0.03273	0.9833	1.0000
Pauli pair	0.09497	0.03563	0.03671	0.7638	1.000
Pauli merged	0.08670	0.03486	0.03591	0.7294	1.0000

F.2 Representation profiles across size and compiled circuits

The 3D diffusion-reaction follow-up uses shapes $(2, 2, 2)$, $(2, 2, 4)$, $(2, 4, 4)$, $(4, 4, 4)$, $(4, 4, 8)$ and $(4, 8, 8)$, the original fixed-state probe and angle .43. Each representation/allocation cell uses 200k shots, three independent sampling seeds and 500 estimates per seed. Exact operator and derivative agreement is checked before sampling. Uniform, coefficient-proportional and exact-variance oracle allocations use their actual integer budgets. The oracle is a diagnostic with known moments.

Table 15: Complete 3D sequence: S/Pauli ratios of predicted gradient standard deviation and measured RMSE, pooling squared errors over three sampling seeds. A value below one favors S.

	Uniform		Coefficient		Variance oracle	
N	Pred.	Meas.	Pred.	Meas.	Pred.	Meas.
8	5.3034	5.3412	2.6900	2.5853	3.5519	3.7022
16	2.3036	2.3198	1.5552	1.5686	1.8022	1.8010
32	1.4727	1.4662	1.1515	1.1381	1.2272	1.2131
64	1.1745	1.1780	1.0432	1.0555	1.0884	1.1190
128	0.7168	0.7096	0.7511	0.7777	0.7292	0.7419
256	0.5269	0.5055	0.6394	0.6405	0.5994	0.6076

Prediction tracks the measured ordering throughout the sequence. The crossover at $N = 128, 256$ occurs under all three allocations, illustrating how weights and observable variances determine measurement efficiency for an exact representation.

For complete-circuit accounting, we compile the source preparation, seven-angle trial state, source reflection, controlled measurement operators and readout to single-qubit rotations and controlled-NOT gates (CX). All-to-all and linear connectivity use the same compilation seed and optimization level. No dense unitary synthesis is used. The four-data-qubit ideal circuits reproduce the independent target gradients to numerical precision. The table gives expected CX gates per sampled measurement under the coefficient-weighted complete-gradient allocation.

Table 16: Complete compiled measurement circuits: coefficient-weighted CX count per sample, including preparation. Data-qubit count n sets $N = 2^n$.

n	Connectivity	S	M	D
4	all-to-all	226.62	242.33	228.31
4	linear	502.29	541.47	512.48
6	all-to-all	537.49	713.00	548.43
6	linear	1541.31	1819.86	1590.86
8	all-to-all	943.86	1468.58	976.34
8	linear	2853.27	4211.24	3008.56

This compiled-resource metric accounts for each representation's complete measurement circuit. Multiplying by a shot budget estimates gate execution volume; circuit-simulator wall time is recorded separately from this quantity.

F.3 Block-local and shared-gradient checks

The block-local trajectory uses four agents on a 4×4 spatial grid ($N = 16$), one initialization, 100 synchronous updates and learning rate .01. Its complete local objective decreases from 3.7963 to .1642, original relative residual from .4239 to .4027, and column disagreement from .2574 to .00213. This directly tests the block-local update.

A separate reduction check partitions the 4×8 operator into four 16×16 blocks. Four coefficient cases with five starts use one shared 45-angle vector and log scale. L-BFGS-B uses 800 iterations, gradient and function tolerances $10^{-9}/10^{-13}$, and 40 line-search steps. Maximum assembled objective/gradient errors are $9.09 \times 10^{-13}/7.28 \times 10^{-11}$. Diagnostic replicas use Metropolis weight $1/3$ for self and two neighbors and check mean preservation and scaled-state consensus separately from optimization.

The manufactured-source check crosses grids 4×8, 8×16 and 16×16, coefficient pairs $(.10, .10), (.25, .25), (.10, .40)$, and five starts. A depth-3 target defines $b = Ax_\star$; initial angles receive normal perturbations of standard deviation .15. Four diagnostic replicas start with RMS disagreement .10. Shared-gradient L-BFGS-B uses 500 iterations with the same tolerances and line-search allowance. These are reduction and consensus checks, separate from the block-local trajectory and from learning on independently specified analytic sources.

Table 17: Manufactured-solution four-worker parity and executability study. Each row contains three coefficient pairs and five restarts. Time is median elapsed time per trajectory.

N	Angles	Target met	Worst residual	Min. improv.	Max grad. error	Time s
32	45	15/15	.001291	99.780%	6.58×10^{-16}	.917
128	63	15/15	.000727	99.801%	1.11×10^{-15}	3.989
256	72	15/15	.001319	99.739%	2.44×10^{-15}	6.568

All 45 runs satisfy the registered residual target and scaled-state consensus check. Maximum objective/gradient parity errors are 4.44×10^{-16}/2.44×10^{-15}, and final scaled-state consensus error is zero. All runs use the full 500-iteration cap; the optimizer's success flag under its stricter gradient tolerance remains false and is separate from the residual outcome.

G Shared-objective, Execution and Robustness Studies

G.1 Normalized objectives and complete access costs

The normalized-objective cohort uses $A = I + .25L_x + .25L_y$ on a 4×4 grid, normalized smooth exponential RHS and a real depth-5 R_y/controlled-Z state with 20 angles and the common local objective $C_L = 1 - \langle\psi|A^\dagger U_b P U_b^\dagger A|\psi\rangle / \langle\psi|A^\dagger A|\psi\rangle$, where $P = n^{-1}\sum_{j=1}^{n}(|0\rangle\langle 0|)_j$ and $U_b|0\rangle = b/\|b\|$. The four layouts use ordered complex Pauli channels, optimized real ordered Pauli, ME real grouping, or structured real grouping. All share 100 fresh initializations uniform in $[-.3, .3]$, 1024 Adam updates with moments $(.9, .999)$ and cosine learning rate .01 to 10^{-5}, the same problem and per-update shot budgets. The signed terminal scalar is estimated with two independent 1M-shot measurements; success is the final original-system raw residual at most .01, including scalar-readout error. Exact aligned residuals are diagnostics. Both numerator and denominator are measured at the base and shifted states; quotient gradients have finite-estimation bias, and no exact denominator enters a sampled update. The exact arm and four layouts at three budgets give 1300 trajectories. Real ordered Pauli and ME have identical trajectories under the declared binary-mixture estimator and are not independent replications.

Table 18: Complete normalized-objective endpoint outcomes. Entries are raw successes out of 100; brackets are pointwise 95% Wilson intervals. Training budgets exclude the separately charged 2M-shot output reconstruction.

Layout	1M/update	10M/update	100M/update
Ordered complex extension	9 [.048,.162]	81 [.722,.875]	84 [.756,.899]
Real ordered Pauli	78 [.689,.850]	100 [.963,1]	99 [.946,.998]
ME real grouped	78 [.689,.850]	100 [.963,1]	99 [.946,.998]
Structured real grouped	78 [.689,.850]	95 [.888,.978]	99 [.946,.998]

Exact controls attain 100/100. Increasing the training budget leaves the terminal readout budget fixed. ME is an equation-level reconstruction of Liu & Wang (2026); the ordered complex method explicitly adds finite-shot training to the layout of Lu et al. (2026), whose released training uses exact expectations. These comparisons isolate representation and estimator components under a shared objective, rather than reproduce the original papers' complete optimization pipelines. The structured grouping arm is not a reproduction of Shen's distributed consensus algorithm.

Complete pair circuits include the common state preparation at each draw; a full gradient uses 41 input states, including 40 parameter shifts. Compilation uses single-qubit gates and CX with a fixed optimization level and arbitrary-input semantics. Expected CX volume weights each compiled circuit by its actual sampling probability. The terminal scalar circuits are counted separately. Effective-observable and controlled-operator audits agree to 1.48×10^{-15}.

Table 19: Unique pair types and expected logical CX gates per 1M-shot update. Gate volumes are compiled cost models, not measured device execution.

Layout	Pair types	Qubits	All-to-all CX	Linear CX
Ordered complex extension	490	5	16.4722M	21.4488M
Real ordered Pauli	238	5	16.9185M	23.3767M
ME real grouped	133	5	16.9185M	23.2749M
Structured real grouped	216	6	46.8021M	90.1099M

The three-start native follow-up freezes starts 300-302 before execution and retains all four layouts, 1024 updates, 1M shots per update and 2M terminal shots. All twelve trajectories pass independent audits. Success counts are respectively 0/3, 1/3, 2/3 and 1/3 in the table's method order. These are descriptive paired outcomes, separate from the 100-start measurement-emulation cohort.

G.2 Cross-host time to a frozen progress level

The execution experiment fixes $N = 16$, four blocks, depth 3, start 900, structured decomposition and the original 8000-update learning-rate schedule, executing its first 400 updates. Every layout uses 160M native GPU simulator shots per update. The target is raw residual at most .75 at two successive checkpoints separated by 50 updates; all confirm it at update 350. Each trajectory executes its complete 64-billion-shot budget. The time boundary includes setup and active work through confirmation, including network transport, and excludes later audit and archival work.

Table 20: All eight native simulator layouts. Host labels distinguish execution environments; no cross-platform hardware equivalence is assumed.

Layout	GPUs	Minutes to progress	Final raw residual
Host A	1	146.24	.5920
Host A	2	72.01	.5846
Host A	4	38.78	.5895
Host B	1	151.91	.5920
Host B	2	69.27	.5846
Hosts A+B	1+1	60.25	.5846
Hosts A+B	2+2	38.89	.5895
Hosts A+B	4+2	29.95	.5867

Independent replay checks numerical updates, timing boundaries and returned measurement counts, totaling 512 billion shots. The fixed-start experiment measures parallel feasibility at a common early-progress level. Final 1% convergence and weak scaling are different estimands.

G.3 Optimization witnesses and circuit invariants

The conditioning study fixes $N = 256$, condition number 100, depth 18 and 4096 updates. Six held-out starts 600-605 share initial parameters across exact/sampled metric and exact/sampled Adam arms. The real circuit alternates entangling edges between layers. Write $d_i = (\partial_i x)/s_i$, where $s_i = \rho/2$ for angles and $s_i = \rho$ for log scale. The metric is $H_{ij} = \mathrm{Re}\langle Ad_i, Ad_j\rangle/\|b\|^2$. Its eigenvalues are floored at $10^{-6}\|A\|_2^2/\|b\|^2$ before solving against $g_i/(2s_i)$ and dividing the result by s_i. The update is norm-capped before the frozen learning-rate multiplier. Sampled updates allocate 10% of 10^{18} shots to the metric and the rest to the gradient. Adam spends the same full budget on gradients. Schedules are frozen after development, not retuned on held-out starts.

Table 21: Every held-out conditioning endpoint. All non-hits are retained.

Start	Exact metric	Exact Adam	Sampled metric	Sampled Adam
600	.007918	.1021	.002804	.1025
601	.01151	.1027	.002618	.1026
602	.001423	.1037	.004794	.1037
603	.002367	.1024	.01013	.1024
604	.002259	.1028	.002458	.1028
605	.003084	.09727	.001602	.09709

Stable success requires five passing .01 checkpoints ten updates apart and a passing endpoint. Metric/Adam success is 5/6 versus 0/6 in both modes; pointwise Wilson intervals are [.436,.970] and [0,.390]. Over all six pairs, metric-minus-Adam mean residual differences are -.09708 exact and -.09777 sampled, with paired percentile intervals [-.1003,-.09377] and [-.09988,-.09519], enumerating all 6^6 resamples. Twelve sampled arms charge 4.9152×10^{22} emulated shots; native training shots are zero. These budgets establish attainable solutions under the intervention rather than practical sampling efficiency.

A separate real R_y/full-chain-CZ family has an all-depth algebraic obstruction. Numbering bits from zero, its outputs obey $x^{\mathsf{T}} P x = 0$ for $P = Y_1 Y_3$. The invariant follows from two Pauli-product spaces, generated by Y_j on even sites and CY_jC on odd sites, and vice versa, where C is the full-chain CZ operator. Local real rotations preserve their combined span and C exchanges the spaces; every nonidentity product has zero expectation in the all-zero input. For the registered input let $H = A^{\mathsf{T}}A + P/2 \succeq I/2$ and $c = A^{\mathsf{T}}b$. For any witness w, completion of the square gives the rigorous original residual certificate

$$\|Ax - b\|^2 \geq \|Aw - b\|^2 + \tfrac{1}{2} w^{\mathsf{T}} P w - 2\|Hw - c\|^2.$$

Exact rational evaluation of the stored binary64 inputs and witness yields relative residual greater than .0467. This certificate covers all depths, angles and output scales of this circuit family; the alternating circuit is a different family. A plateau by itself supplies no such certificate. Complementary rank controls use the fact that each crossing CZ at a bipartition doubles Schmidt rank at most: depth L has rank at most 2^L. If s_j are target Schmidt singular values, then $r \geq \sigma_{\min}(A)\sqrt{\sum_{j>2^L} s_j^2}/\|b\|$. Thirty-two shallow/deeper pairs have certified shallow exclusions and explicit deeper witnesses, with 0/32 versus 32/32 target attainment.

G.4 Matched compiled-circuit noise

Every noise trajectory uses $N = 16$, depth 3, 1024 updates and 10M emulated shots per update. Six paired starts cross S, Pauli and direct square with $p_2 \in \{0, 10^{-5}, 10^{-4}, 10^{-3}\}$ and $p_1 = p_2/10$. Compiled circuit density-matrix means, including preparation and controlled gates, define the binary count law. All 72 trajectories and update replays pass independent audits, with maximum discrepancy 8.88×10^{-15}.

Table 22: Stable 1% noise outcomes out of six starts per cell.

Representation	$p_2 = 0$	10^{-5}	10^{-4}	10^{-3}
Structured	5	6	0	0
Pauli	3	5	0	0
Direct square	4	5	0	0

Stable success uses five .01 checkpoints spaced ten updates and a passing endpoint. Only the 12 successful compiled zero-noise controls enter matched noise-loss attribution. Among their 36 nonzero-noise comparisons, 25 lose attainment: one at 10^{-5} and twelve at each higher level. Independent finite-shot draws explain why a small nonzero-noise cell can have more successes without establishing a beneficial noise effect. Median mean gradient-bias norms across methods are .00331, .0323 and .261 at increasing nonzero levels. At the two higher levels, target cost is censored at 10.24 billion

shots for every arm; no finite target gate cost is inferred. All 737.28 billion shots are emulated. Weighted logical CX volume is a compiled cost proxy, distinct from actual device time.

G.5 Weighted local relations and dynamic finite elements

Let $A = D + \sum_{r=1}^{C} T_r$ be a real symmetric matrix, with each T_r supported on a matching in the off-diagonal graph. Its disjoint two-by-two blocks imply $\|T_r\| = \max_{(i,j)} |A_{ij}| = s_r$. For $s_r > 0$ set $B_r = T_r/s_r$ and $Q_r = (I - B_r^2)^{1/2}$. Because B_r, Q_r commute, $(B_r + iQ_r)(B_r - iQ_r) = I$ and their sum is $2B_r$. Thus each matching is the exact sum of two scaled unitaries. Centering D at $c_0 = (\max D + \min D)/2$ gives a diagonal contraction with the same two-term completion; the remaining $c_0 I$ adds at most one term. Zero-radius terms are omitted. Greedy edge coloring uses at most $2\Delta - 1$ colors because an edge meets at most $2\Delta - 2$ already colored edges. This proves the $2C + 3$ bound with no quantization of continuous coefficients. Controlled implementation uses pair-dependent rotations and diagonal phases, including their controlled global phases. Angle tables and generic two-level routing can grow with mesh size even at bounded degree.

The validation uses six independent finite-element meshes, seeds 400-405, 49 physical vertices padded to $N = 64$, 16 time steps and paired warm/cold starts. Each inner solve has depth 12 and 2048 metric updates with relative curvature floor 10^{-4}. Sampled methods allocate 10% of 10^{14} shots per update to the metric. The classical outer driver supplies complete complex RHS vectors and exact norms, propagating padding leakage. The trajectory reference uses the identical time discretization; a separate continuous-time reference diagnoses time-discretization error.

Table 23: All dynamic trajectories. Physical and padded-state errors measure different outputs and are reported separately.

Method	Success	Max inner residual	Max physical error	Max padded error
Exact	12/12	$3.04\,10^{-11}$	$2.80\,10^{-13}$	$9.48\,10^{-11}$
Weighted matching	12/12	.004901	.000905	.02822
Pauli	12/12	.005366	.001096	.03276

Success requires stable 1% residual at every inner solve and at most 1% final physical-field error. Meshes, not individual time steps, are the independent units; all six pass both starts for each method, with pointwise Wilson interval [.610,1]. Each sampled trajectory charges 3.2768×10^{18} emulated shots. All-update reconstruction checks FEM assembly, propagated inputs, counts, metric/gradient updates and circuit endpoints, with maximum discrepancy 9.56×10^{-11}. Separate compiled transport checks execute 277632 CPU and 277632 GPU simulator shots over 1446 programs and 4338 bindings. These transport counts are separate from the dynamic training, which uses no native circuit shots.

H Rapid Structure Checks: Allocation, Directed Learning and Precision

H.1 Cross-family allocation at fixed gradient probes

The scalar-gradient probes use anisotropic diffusion-reaction, nine-point elliptic and directed upwind convection-diffusion stencils on 4×8 and 8×8 grids. Each exact representation uses the same operator and five fixed angles $.10, .25, .43, .70, 1.00$. At each 200k or 1M total-shot budget, three seeds independently average 200 binary-component draws. Coefficient allocation spends the full budget on its final estimate. Adaptive allocation spends 5% on an independent coefficient-allocated pilot, uses its measured component variances to allocate the remainder and charges both stages to the stated total. The pilot does not enter the final gradient. Every active component receives at least one pilot shot. A cell's RMSE pools the three seed-level squared errors; Table 24 averages the ten cell ratios over two budgets and five angles for each family and size. The largest exact-gradient discrepancy is 9.77×10^{-15}; mean ratios of measured to predicted squared error across method-policy combinations lie between .998 and 1.011. These are fixed-point derivative probes, not optimizer trajectories.

Table 24: Cross-family scalar-gradient RMSE, S/M. Each row covers ten cells over two total-shot budgets and five angles. Counts give cells with lower S RMSE. The adaptive pilot is included in the shot budget.

Operator	N	Coeff. S/M	Coeff. wins	Adaptive S/M	Adaptive wins
Anisotropic diffusion	32	.677	10/10	.810	7/10
	64	.548	10/10	.602	10/10
Nine-point elliptic	32	.761	10/10	.765	9/10
	64	.523	10/10	.291	10/10
Directed upwind	32	.436	10/10	.406	10/10
	64	.329	10/10	.225	10/10
Mean over 60 cells		.546	60/60	.517	56/60

The four adaptive cells with S/M above one all use $N = 32$ and 200k shots: anisotropic diffusion at angles $.10, .25, .70$ has ratios 1.116, 1.137, 1.058, and nine-point elliptic at angle .70 has ratio 1.041. All other adaptive cells favor S. Full seed-level records retain all 720 representation-level estimates.

H.2 Nonsymmetric upwind learning

The implicit one-dimensional operator is $A = I + .25L + \beta(I - T_-)$, where T_- is the directed superdiagonal shift. At $N = 64, 128$ and $\beta = .2, .6$, exact directed completion and merged Pauli reconstruct the same A for a normalized smooth exponential source. Four paired starts per setting share a depth-3 $R_y R_z R_y$/controlled-Z circuit, learned log scale and 512 Adam updates. S/M draw 1M binary-mixture measurements per complete gradient; exact-gradient controls share the circuit. CPU emulation samples the estimator law without circuit execution. Original A, b residuals are recomputed; independent finite differences agree to 6.65×10^{-10} and all binary oracle means lie in $[-1, 1]$.

Table 25: Upwind learning at 512M emulated measurements per sampled trajectory. Entries are mean final raw original-system residual over four paired starts; S wins count paired final residuals below M.

N	β	Exact control	S	M	S wins
64	.2	.05995	.06796	.09430	4/4
64	.6	.1488	.1864	.3095	4/4
128	.2	.05464	.08256	.1561	4/4
128	.6	.1908	.3111	.5596	4/4

No arm reaches stable 1%; this protocol tests fixed-budget learning.

H.3 Native precision thresholds on the existing cohort

The four audited native reuse-on pairs of Appendix B are reanalyzed without new execution. The original 1% costs are unchanged; 2% and 5% are post-hoc. Stable hits require five passing checkpoints ten updates apart and a passing endpoint; non-hits are censored at 512M actual GPU circuit-simulator measurements, at 1M per update.

Table 26: Stable raw-residual thresholds on the same four native pairs. Costs are millions of measurements; > 512 denotes censoring.

Target	S hits	M hits	S costs	M costs
5%	4/4	4/4	50, 40, 40, 40	50, 40, 40, 40
2%	4/4	1/4	140, 110, 160, 100	> 512, > 512, 390, > 512
1%	3/4	0/4	> 512, 340, 470, 340	> 512, > 512, > 512, > 512

I LARGER NONSYMMETRIC SYSTEMS AND MEASUREMENT BUDGETS

Frozen protocol. We use $A = I + \sum_{k=1}^{d}[.25L_k + \beta(I - T_{-,k})]$, with open boundaries, $\beta = .2, .6$ and fixed coefficients across sizes. Directed completion has $4d + 1$ terms; merged Pauli uses exact coefficients of A and $A^\dagger A$. The normalized sources are $e^{\cdot 4x}$ in 1D and $e^{\cdot 4x + .2y}$ in 2D. For 1D $N = 256, 512, 1024$ and 2D grids $16 \times 16, 32 \times 32$, each setting has eight new paired starts, a depth-3 $R_y R_z R_y$/controlled-Z circuit, initial angles uniform in $[-.3, .3]$, and initial log scale zero. All arms use 512 Adam updates with $(\beta_1, \beta_2) = (.9, .9)$ and learning rate .003 for the first 30%, then cosine decay to .00006. S/M use 1M emulated measurements per complete gradient; exact-gradient controls share every initial vector and update schedule. At $N = 512$, added .25M/4M per-update budgets reuse the same 16 initial vectors and exact controls. The 304 distinct trajectories comprise 224 sampled paths and 80 exact controls; the base-budget study has 80 paired starts, while extra budgets add comparisons on existing starts. Independent development starts check identities and runtime before validation, with no validation-based retuning.

Measurement semantics and numerical checks. CPU emulation draws the coefficient-randomized binary-mixture law of Eq. (11), with coefficient-proportional coordinate allocation. Exact derivatives enter only the classical measurement-distribution oracle; sampled Adam receives its binomial draw. Total emulated measurements are 151.552B. Independent coordinate stencils, lifted signed-permutation actions and Pauli $A, A^\dagger A$ actions agree; profiles also match the preceding dense construction at $N = 64, 128$. Independently rebuilt original-system residuals agree to 2.22×10^{-16}; final finite-difference gradients agree to 1.01×10^{-10}. Every per-update budget and paired initial vector is audited.

Table 27: All larger upwind settings. Final raw residuals are means over eight paired starts; S is lower than M in 8/8 pairs in every row. Intervals are per-setting descriptive 95% Student-t intervals for paired S-minus-M residuals, with seven degrees of freedom. Exact controls repeat across budgets.

Grid	β	Shots/update	Exact	S	M	Paired 95% interval
256	0.2	1M	0.03631	0.07719	0.2544	$[-0.2298, -0.1247]$
256	0.6	1M	0.1434	0.2631	0.6484	$[-0.4463, -0.3244]$
512	0.2	0.25M	0.04417	0.2280	0.7270	$[-0.5412, -0.4569]$
512	0.2	1M	0.04417	0.1127	0.3770	$[-0.3148, -0.2138]$
512	0.2	4M	0.04417	0.06848	0.1703	$[-0.1326, -0.07093]$
512	0.6	0.25M	0.1138	0.4477	0.9778	$[-0.5804, -0.4798]$
512	0.6	1M	0.1138	0.2447	0.7274	$[-0.5345, -0.4308]$
512	0.6	4M	0.1138	0.1583	0.4101	$[-0.2961, -0.2074]$
1024	0.2	1M	0.06634	0.1818	0.6006	$[-0.4700, -0.3676]$
1024	0.6	1M	0.1128	0.3206	0.9052	$[-0.6197, -0.5494]$
16×16	0.2	1M	0.1367	0.2096	0.3240	$[-0.1476, -0.08114]$
16×16	0.6	1M	0.2732	0.3589	0.5528	$[-0.2309, -0.1569]$
32×32	0.2	1M	0.2255	0.4123	0.7604	$[-0.3774, -0.3187]$
32×32	0.6	1M	0.2220	0.5474	0.9515	$[-0.4547, -0.3534]$

Observed quarter-budget comparisons. Figure 9 displays the three measured budgets at $N = 512$. Table 28 compares S at 128M against M at 512M, and S at 512M against M at 2048M. S returns a lower original-system residual in all eight pairs of each contrast. These four contrasts share 16 distinct initializations. They establish 75% fewer measurements with lower endpoint residual at the observed budgets and fixed update count; target costs are computed only at recorded checkpoints.

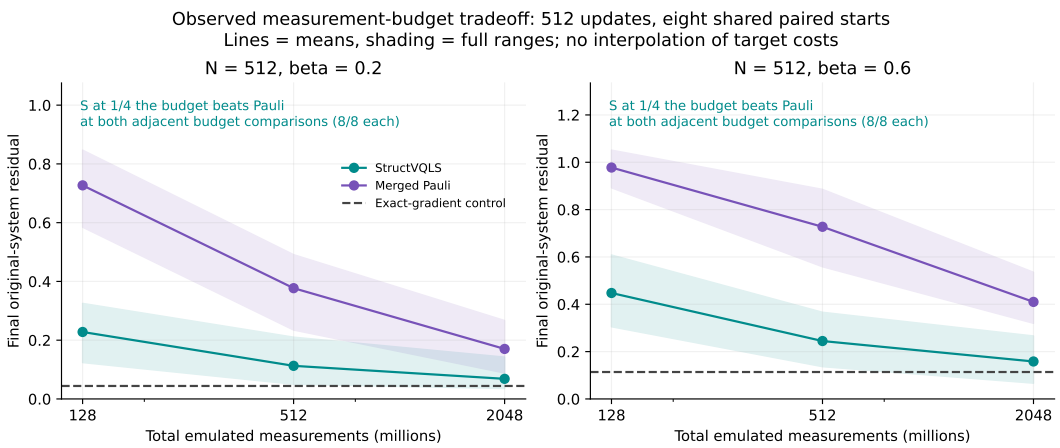

Figure 9: Measurement-emulated upwind learning at $N = 512$: eight paired starts per advection weight, 512 updates, means and full ranges. Exact-gradient controls share the initializations. Each adjacent cross-budget comparison uses the same eight starts; all favor S despite its fourfold smaller budget.

Table 28: Observed quarter-budget contrasts. Budgets are total emulated measurements; every row favors S in all eight pairs.

β	S budget	M budget	S residual	M residual	Paired 95% interval
0.2	128M	512M	0.2280	0.3770	$[-0.1747, -0.1235]$
0.2	512M	2048M	0.1127	0.1703	$[-0.07289, -0.04215]$
0.6	128M	512M	0.4477	0.7274	$[-0.3185, -0.2409]$
0.6	512M	2048M	0.2447	0.4101	$[-0.2047, -0.1259]$

Prespecified target outcomes. Raw-residual targets $.1, .2, .4$ are fixed before validation. Stable success requires five passing checkpoints ten updates apart and a passing returned state. Non-hits are censored at the full trajectory budget. Table 29 retains all thresholds and all controls; no threshold is selected from the validation results.

Table 29: Stable hits out of eight starts at each prespecified raw-residual target. Each triplet is exact/S/M. Exact controls are shared across budgets.

Grid	β	Shots/update	$r \le .1$	$r \le .2$	$r \le .4$
256	0.2	1M	8/7/0	8/8/2	8/8/8
256	0.6	1M	5/0/0	6/3/0	8/7/1
512	0.2	0.25M	7/0/0	8/3/0	8/8/0
512	0.2	1M	7/3/0	8/7/0	8/8/3
512	0.2	4M	7/7/1	8/8/5	8/8/8
512	0.6	0.25M	4/0/0	6/0/0	8/3/0
512	0.6	1M	4/0/0	6/3/0	8/8/0
512	0.6	4M	4/3/0	6/5/0	8/8/3
1024	0.2	1M	6/2/0	7/6/0	8/8/0
1024	0.6	1M	3/0/0	7/1/0	8/6/0
16×16	0.2	1M	0/0/0	8/4/1	8/8/6
16×16	0.6	1M	0/0/0	3/0/0	6/6/0
32×32	0.2	1M	0/0/0	5/0/0	7/3/0
32×32	0.6	1M	0/0/0	4/0/0	7/1/0

J NATIVE SAMPLING FOR NONSYMMETRIC CONVECTION-DIFFUSION

Frozen protocol. The open-boundary operator is $A = I + .25L + \beta(I - T^-)$, where L is the one-dimensional second-difference matrix and T^- has ones on the superdiagonal; $N = 256$, $\beta = .2, .6$, and $b_j \propto \exp(.4j/(N+1))$. For each weight, four fresh starts are shared by S, M and exact-gradient controls. The depth-3 circuit uses R_y, R_z, R_y rotations and CZ entanglers, angles uniform in $[-.3, .3]$, and initial log scale zero. Adam uses $(\beta_1, \beta_2) = (.9, .9)$, $\epsilon = 10^{-8}$, and learning rate .003 for the first 30% of 512 updates, then cosine decay to .00006. Each sampled full gradient uses exactly 1M actual GPU circuit-simulator measurements, totaling 512M per trajectory and 8.192B across all 16. No past-count reuse is applied. Eight exact-gradient paths use the same initializations and optimizer, with no sampled measurements.

Circuit realization and audit. S uses five signed-permutation completion terms for the directed lift $B = \begin{bmatrix} 0 & A \\ A^\dagger & 0 \end{bmatrix}$, input $|1\rangle|\psi\rangle$ and target $|0\rangle|b\rangle$. M uses exact Pauli representations of A and $A^\dagger A$, merging products and removing the identity; complex phases enter the controlled unitary. The exponential RHS has an exact rotation preparation. Both estimators use complete quadratic and linear parameter-shift terms and the analytic log-scale offset. Classical exact observers evaluate original-system residuals without supplying sampled updates. Small-system circuit and gradient checks precede target-size gates. Independent replay of integer counts, allocations and Adam states passes for all 16 trajectories; original-residual replay error is zero. All registered outcomes are retained.

Table 30: All native upwind endpoints after 512M measurements per arm. r is the learned-scale original residual and R the aligned residual. Pairs are fresh for this native cohort.

β	Pair	r_S	r_M	R_S	R_M
0.2	1	0.05991	0.1538	0.05980	0.1362
0.2	2	0.07262	0.2048	0.07249	0.1728
0.2	3	0.06475	0.1916	0.06473	0.1758
0.2	4	0.1543	0.3501	0.08046	0.2435
0.6	1	0.09147	0.4154	0.09093	0.3525
0.6	2	0.2992	0.7182	0.1326	0.6805
0.6	3	0.07994	0.2346	0.07985	0.2043
0.6	4	0.6495	0.8980	0.6194	0.8979

Paired outcomes. S wins all four pairs at each weight. Mean S/M raw residuals are $.08790/.2251$ at $\beta = .2$ and $.2800/.5666$ at $\beta = .6$, reductions of 60.9% and 50.6%. Descriptive 95% Student-t intervals for the paired mean S-minus-M difference are $[-.2050, -.06938]$ and $[-.4651, -.1080]$, computed over four starts per weight. Figure 3d,e retains all learning curves and the matched exact-gradient means. These results evaluate directed operator structure under native sampling at equal measurement budgets.