SHARED ADDRESS LEARNING FOR AMPLITUDE ENCODING IN QUANTUM MACHINE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Quantum machine learning offers expressive Hilbert-space models, but loading classical data involves a trade-off between compactness and cost. Amplitude encoding is highly compact, yet state preparation can consume much of the entangling-gate budget on near-term hardware. This cost depends heavily on the address, which assigns data coordinates to computational-basis states and is typically fixed to a default order. Sharing one address across a dataset preserves all inner products and the ideal amplitude kernel, allowing us to reshape entanglement for efficient preparation. We prove an exponential fidelity separation between addresses on a family of maximally entangled states under limited gate budgets, and show that a correlation-based construction makes the family completely separable. Shared Address Learning (SAL) builds candidate addresses from the correlation structure of unlabeled data, refines them with a lightweight structural score, and selects one by shallow-circuit fidelity. Learned once, the address is applied to all inputs as classical preprocessing compatible with any loader. For images and pretrained vision and language features, SAL reaches equivalent preparation fidelity with significantly fewer entangling gates and consistently improves tensor-network and adaptive-compilation loaders. These gains extend to quantum kernel and quantum neural network (QNN) classifiers in simulation and persist for a QNN deployed on a superconducting processor.

1 INTRODUCTION

Quantum machine learning explores hybrid models leveraging quantum phenomena for computational advantages (Biamonte et al., 2017; Cerezo et al., 2022). For classical datasets, algorithms like quantum kernel methods (Havlíček et al., 2019; Lei et al., 2024; Thanasilp et al., 2024) and quantum neural networks (QNNs) (Pérez-Salinas et al., 2020; Wang et al., 2025b; Ye et al., 2025) rely on embedding maps translating classical data into quantum states (Schuld & Killoran, 2019). Amplitude encoding is particularly attractive, requiring logarithmically many qubits relative to the input dimension by mapping data (e.g., pixel intensities) directly to probability amplitudes (Jobst et al., 2024). Despite this compactness, exact preparation of generic states requires exponentially scaling circuit depths (Möttönen et al., 2005; Shende et al., 2006; Sun et al., 2023). The resulting entangling-gate counts induce prohibitive error rates (Preskill, 2018), often exhausting the hardware's coherence budget before learning begins. Consequently, approximate state preparation is a practical necessity. Techniques like matrix product state (MPS) approximations (Jobst et al., 2024) or variationally compiled loaders (Shirakawa et al., 2024; Zhang et al., 2026) show that bounded approximations preserve sufficient task-relevant information for classification.

A critical, overlooked degree of freedom in approximate preparation is the coordinate assignment provided to the loader. Standard practice assumes a rigid flattening (e.g., row-major for images), termed the *natural address*. This fixed assignment dictates amplitude partitioning, profoundly shaping the target state's entanglement profile. Reordering coordinates drastically alters entanglement complexity, even though the amplitude values remain identical. We exploit this freedom at the dataset level with what we call a *shared address*: a single coordinate permutation P applied to every input. Because P is orthogonal, it strictly preserves all pairwise inner products and the ideal amplitude kernel, while it can radically reduce hardware preparation resources. As Figure 1 illustrates, exchanging one pair of coordinates turns two different entangled inputs, one of them a Bell state,

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.

into product states simultaneously. We theoretically demonstrate that entire orthogonal families of maximally entangled states collapse into product states under an optimal address. For QNNs, this strategy interacts intricately with the ansatz architecture, requiring its impact on trainability and generalization to be evaluated alongside loading fidelities.

To systematically exploit this flexibility, we introduce *Shared Address Learning* (SAL), a scalable algorithmic framework that discovers effective coordinate permutations from unlabeled training data. Since combinatorial search is intractable, SAL leverages statistical amplitude correlations to identify promising binary index transformations. Starting from the natural order and a reference order based on the leading principal direction, SAL refines candidates with a lightweight structural score. Bounded-circuit fidelity evaluations on a held-out screening subset dictate the final selection. The selected permutation acts as a deterministic, reusable operation for unseen data. It remains completely agnostic to the state-preparation backend, allowing seamless integration with any approximate quantum compiler.

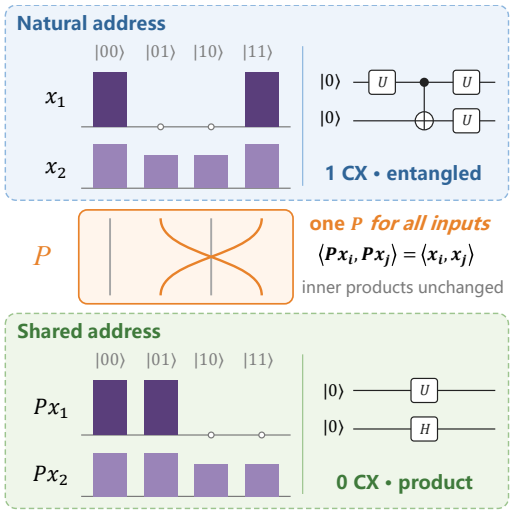


Figure 1: Shared address. Under the natural address, coordinate j of each input is stored on basis state $|j\rangle$, and both inputs shown are entangled, requiring one CX each. A single permutation matrix P, applied to the classical vectors before loading, exchanges the coordinates stored on $|01\rangle$ and $|11\rangle$, i.e., the index map $(a, b) \mapsto (a \oplus b, b)$. Both inputs become product states that need no CX, while all pairwise inner products are preserved (here $\langle x_1, x_2 \rangle = 0.8$). U denotes a single-qubit gate.

Our contributions span theoretical guarantees, algorithm design, and hardware deployment. We formalize the shared address and establish a rigorous theoretical separation in preparation fidelity between addresses, alongside a correlation-based recovery guarantee. Building on this, we present the SAL procedure, bridging classical data statistics with quantum bounded-circuit screening to identify hardware-efficient embeddings. Finally, we comprehensively evaluate entangling-gate savings and fidelity improvements across diverse loaders and data modalities. We demonstrate enhanced classification performance for quantum kernels and QNNs in simulation and for a QNN on superconducting hardware, complemented by ablations elucidating the structural information preserved by learned addresses.

2 Related Work and Preliminaries

2.1 Amplitude Encoding and State Preparation

Classical inputs are represented in quantum models via strategies like rotation angles or probability amplitudes (Schuld & Killoran, 2019). In amplitude encoding, a normalized vector $x \in \mathbb{R}^d$ ($d = 2^n$) is encoded into an n-qubit state $|x\rangle = \sum_{b\in\{0,1\}^n} x_b|b\rangle$, where b is a binary coordinate index. A simple product state factorizes into individual qubit states, requiring only single-qubit gates. However, correlations preventing factorization constitute entanglement, necessitating noisy two-qubit operations. Since exact preparation of generic targets requires exponentially many entangling gates, it is often prohibitive (Möttönen et al., 2005; Shende et al., 2006; Sun et al., 2023). Approximate methods optimize fidelity at fixed budgets via shallow variational encoding (Nakaji et al., 2022), MPS truncation (Ran, 2020; Jobst et al., 2024), automatic quantum circuit encoding (AQCE) (Shirakawa et al., 2024), or AQER (Zhang et al., 2026). While prior MPS work optimizes qubit tensor ordering individually per target (Jeon et al., 2024), we seek a single reordering shared across a dataset.

2.2 Input Mappings and Quantum Models

The input map dictates how quantum models process data (Havlíček et al., 2019; Schuld & Killoran, 2019; Kübler et al., 2021). Learned input representations include task-adapted quantum feature maps (Lei et al., 2024), neural amplitude embeddings for visual fields (Wang et al., 2025a), and product-

state node embeddings (Xiong et al., 2024). In amplitude encoding, circuit locality corresponds to index bits; local operations on neighboring qubits can couple spatially separated pixels, a dynamic crucial for quantum convolutional neural networks (Cong et al., 2019). Equivariant QNNs exploit this by designing assignments, circuits, and measurements around data symmetries (Meyer et al., 2023; West et al., 2023; Das et al., 2024). Recent quantum graph networks encode node and edge features while preserving node permutation equivariance (Ye et al., 2025). For a general QNN with unitary U_θ and observable O, a global permutation P modifies the prediction to:

$$f_{\theta,O}(Px) = \langle x|P^\dagger U_\theta^\dagger O U_\theta P|x\rangle. \tag{1}$$

The model family is preserved if $\{U_\theta^\dagger O U_\theta\}_\theta$ is closed under conjugation by P. Recent analyses connect QNN trainability to circuit architecture (Anschuetz, 2025) and examine prediction through task-dependent risk and classification margins (Du et al., 2023; Hur & Park, 2025). Since the precise analytical impact of coordinate reordering on general QNNs is difficult to quantify, we select addresses based on preparation fidelity and assess downstream QNN performance empirically.

2.3 Shared Addresses

We define a shared address as a single coordinate permutation P applied to every input before state preparation. We measure preparation cost by the budget G of logical controlled-NOT (CX) gates, as they generate entanglement and dominate execution errors (other blocks decompose into CX gates, Appendix C.3). Let $\mathcal{H}_G$ be the set of states reachable from a product state using arbitrary single-qubit gates and at most G CX gates. The optimal fidelity and its dataset average are:

$$F_G^\star(P;x) = \max_{|\psi\rangle\in\mathcal{H}_G} |\langle\psi|Px\rangle|^2, \qquad \overline{F}_G^\star(P) = \mathbb{E}_x F_G^\star(P;x). \tag{2}$$

Any common permutation satisfies $P^\dagger P = I$, so for normalized inputs x, y,

$$\langle Px, Py\rangle = \langle x, y\rangle, \qquad \|Px - Py\|_2 = \|x - y\|_2, \qquad k(Px, Py) = k(x, y), \tag{3}$$

where $k(x, y) = |\langle x, y\rangle|^2$ is the amplitude kernel. The second-moment matrices $M = \mathbb{E}[xx^\dagger]$ and $PMP^\dagger$ also have the same spectrum.

Approximate preparation perturbs this shared ideal kernel. Let F_i be the preparation fidelity for target x_i, K the ideal amplitude kernel of N targets, and $\widetilde{K}$ the kernel of the prepared states. Then

$$|K_{ij} - \widetilde{K}_{ij}| \le \sqrt{1 - F_i} + \sqrt{1 - F_j}, \qquad \frac{\|K - \widetilde{K}\|_F^2}{N^2} \le 4\bar{\epsilon}, \qquad \bar{\epsilon} = \frac{1}{N}\sum_i (1 - F_i). \tag{4}$$

Appendix A.2 derives these bounds and gives a stability bound for regularized kernel predictors. Sorting inputs independently, by contrast, can change the ideal kernel before preparation (Section 4.4).

While the ideal kernel remains invariant, the shared address's effect on preparation overhead can be exponentially large. The Bell-state input in Figure 1 is the smallest member of entire families of states that normally require entanglement across the same register partition, a requirement an optimal shared address removes entirely.

Theorem 1 (Shared-address separation)**.** *Let $n = 2r$, with registers A and B of size r. For $s \in \{-1, +1\}^r$, define the phased Bell-pair states*

$$|x_s\rangle = \bigotimes_{i=1}^{r} \frac{|0\rangle_{A_i}|0\rangle_{B_i} + s_i|1\rangle_{A_i}|1\rangle_{B_i}}{\sqrt{2}}. \tag{5}$$

One binary-linear address maps every $|x_s\rangle$ to a product state. Under the natural address, every $|\psi\rangle \in \mathcal{H}_G$ satisfies

$$|\langle\psi|x_s\rangle|^2 \le \min\{1, 2^{G-r}\}. \tag{6}$$

Proof sketch. The index map $(a, b) \mapsto (a \oplus b, b)$, where $\oplus$ is bitwise exclusive OR (XOR), sends $|x_s\rangle$ to $|0\rangle^{\otimes r} \otimes \bigotimes_i (|0\rangle + s_i|1\rangle)/\sqrt{2}$. Under the natural address, each target has 2^r equal Schmidt coefficients across $A|B$. Starting from a product state, each crossing CX at most doubles the Schmidt rank. A state of rank at most 2^G has a squared overlap of at most 2^{G-r} with the target (Appendix A.3). □

The 2^r sign choices yield mutually orthogonal states, perfectly preserving distinguishability. Under the natural address, each target requires r CX gates for exact preparation. Below this budget, fidelity falls exponentially. In sharp contrast, the reordered family requires only single-qubit gates. The next section leverages this binary index structure to guide practical address learning.

3 Shared Address Learning

Learning an address entails finding an optimal coordinate permutation, which requires searching over assignments and evaluating their preparation fidelity. The $d = 2^n$ coordinates admit $(2^n)!$ permutations, and evaluating even a single candidate involves fitting a bounded circuit to many targets. Exhaustive search is therefore prohibitive even for modest input dimensions. A practical construction must identify promising regions in this space before expending circuit fits on them.

SAL separates this task into proposal and selection. The proposal stage leverages relationships shared across inputs to construct a few starting addresses, which are then refined using a computationally inexpensive structural score. The selection stage compares their actual preparation fidelities. We first explain the binary transformations used to represent candidates and the data statistics that guide our initial choices. Figure 2 places these steps in the full procedure. The unlabeled training data serve three roles: a construction set D_{C} provides the statistics, a subset $D_{\mathrm{A}} \subset D_{\mathrm{C}}$ supplies the fast loader's topology, and a disjoint set D_{F} supplies the final fidelity scores.

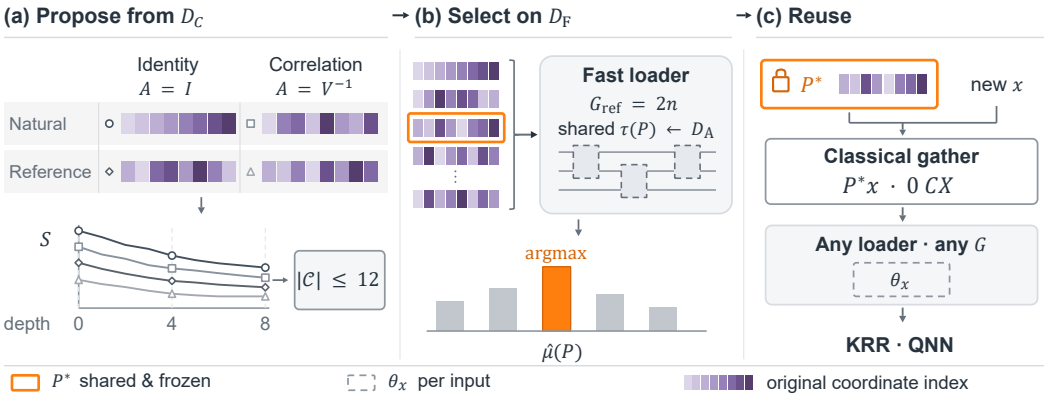


Figure 2: Shared Address Learning. (a) Natural and principal-reference coordinate orders are paired with identity and correlation-based binary maps. Local refinements provide candidate checkpoints. (b) Preparation fidelity on D_{F} selects an address, with the loader topology learned on D_{A}. (c) The address is reused on new inputs, each with its own preparation parameters.

3.1 Starting Addresses

Changing index bits. A coordinate exchange moves two amplitudes. It changes the Schmidt rank across any cut by at most two, so transforming a state in Theorem 1 into a product state requires at least $(2^r - 1)/2$ such exchanges. The separating map achieves this rearrangement efficiently with r XOR operations on index bits. Acting on bits moves many coordinates at once.

For example, replacing a two-bit index (j_1, j_2) by $(j_1 \oplus j_2, j_2)$, the map used in Figure 1, changes how its coordinates are assigned to qubits. The first output bit records whether the original bits differ, and applying the same rule again recovers the original index, ensuring a unique address. More generally, we let each output bit be an XOR of selected input bits and require the transformation to be invertible. These transformations are represented by invertible binary matrices in $\mathrm{GL}(n, 2)$, where arithmetic uses addition modulo two.

A candidate starts from a base coordinate order Q, giving $y = Qx$, and then applies $b = Aj$ to its binary indices:

$$(P_{Q,A}x)_b = y_{A^{-1}b}, \qquad A \in \mathrm{GL}(n, 2). \tag{7}$$

The inverse identifies the old coordinate assigned to each new address. If v_q is column q of A^{-1}, flipping output bit q pairs old indices j and $j \oplus v_q$. Thus, choosing A amounts to selecting which index direction each qubit represents. The directions must be linearly independent to yield an

invertible assignment. This family is much smaller than all coordinate permutations, yet it still contains $\prod_{k=0}^{n-1}(2^n - 2^k)$ maps. Data statistics guide the choice within it.

Correlated directions. Data statistics identify which directions are easy to represent locally. Consider a direction v along which paired amplitudes satisfy $y_{j\oplus v} = y_j$. If this direction is assigned to a single qubit, that qubit factors out as $(|0\rangle + |1\rangle)/\sqrt{2}$; a consistent sign reversal instead yields $(|0\rangle - |1\rangle)/\sqrt{2}$. Both inputs in Figure 1 satisfy $x_{j\oplus v} = x_j$ with $v = (1, 1)$; the map there assigns this direction to the second qubit, which therefore factors out as $(|0\rangle + |1\rangle)/\sqrt{2}$ for each of them.

We measure the strength of this relation along each direction using the binary autocorrelation:

$$c_y(v) = \sum_j y_j y_{j\oplus v}, \qquad \bar{c}_Q(v) = \frac{1}{|D_{\mathrm{C}}|} \sum_{x \in D_{\mathrm{C}}} |c_{Qx}(v)|. \tag{8}$$

The products retain relative signs within an input, while the absolute value allows a useful direction to have opposite correlation signs across different inputs. SAL scans nonzero directions in decreasing order of $\bar{c}_Q(v)$ and retains a direction whenever it increases the rank of those already selected. The first n retained directions form the columns of V, giving $A = V^{-1}$. On the separating family, this rule has an exact recovery guarantee.

Theorem 2 (Recovery by the correlation construction)**.** *For the family in Theorem 1, the correlation construction in the natural base maps every member to a product state, using any nonempty construction subset and exact statistics.*

All members occupy the same diagonal index set $H = \{(a, a) : a \in \mathbb{F}_2^r\}$. Their absolute correlations equal one along H and zero elsewhere, independently of the Bell-pair signs. The greedy rule therefore identifies the shared support first. In the resulting coordinates, the support and signs factorize across qubits, even for sign patterns absent from the construction set (Appendix A.4).

A principal-reference order. A complementary starting point leverages the principal components of the dataset. Let $M = \mathbb{E}[xx^\dagger] = \sum_j \lambda_j u_j u_j^\dagger$ denote the uncentered second-moment matrix. For a fixed preparation state ϕ,

$$\mathbb{E}_x |\langle\phi|Px\rangle|^2 = \sum_j \lambda_j |\langle\phi|Pu_j\rangle|^2, \qquad \overline{F}^\star_G(P) \geq \lambda_1 F^\star_G(P; u_1), \tag{9}$$

where $\overline{F}^\star_G(P)$ is the dataset average in Eq. (2). A circuit approximating Pu_1 supplies a feasible preparation for every input, providing this lower bound. When λ_1 is large, simplifying this shared leading direction raises a substantial fidelity floor. We estimate u_1 on D_{C}, fix its sign, and order its coordinates by sign group and decreasing magnitude. The resulting reference order Q_{ref} makes u_1 monotone within each sign group, providing a second base order for binary addresses.

We pair $Q \in \{I, Q_{\mathrm{ref}}\}$ with the identity and correlation maps in each order. These four starts include the natural address itself. The reference order organizes amplitudes using a dataset-wide profile, after which the binary map chooses which relationships become local to the qubits.

3.2 Local Refinement

The starting addresses capture broad regularities, which we then refine through small changes to the binary map. Specifically, adding row a of A to row b replaces output bit b by its XOR with bit a. This invertible move alters the relationship between the corresponding qubits, allowing a short sequence of row additions to explore the neighborhood of each starting address.

To evaluate these moves cheaply, we measure how close each qubit is to separating from the rest of the state. Its reduced state $\rho_q(y)$ has two eigenvalues summing to one; the smaller eigenvalue is zero when the reduced state is pure and increases with entanglement. Summing these values gives $S(y) = \sum_q \lambda_{\min}(\rho_q(y))$, which bounds the infidelity of the best product-state approximation $F^\star_{\mathrm{prod}}(y)$:

$$\max_q \lambda_{\min}(\rho_q(y)) \leq 1 - F^\star_{\mathrm{prod}}(y) \leq \min\{1, S(y)\}. \tag{10}$$

Thus, a small score guarantees an accurate product approximation. This is an ideal preliminary objective for a shallow loader: single-qubit gates prepare the product component, while the available entangling gates represent the remaining correlations.

The score also permits efficient local updates. Under a binary address, each reduced state depends on an autocorrelation and on the difference between the amplitude energies in two parity groups. Walsh transforms compute these statistics for all binary directions (O'Donnell, 2014). Because a row addition changes only two terms of S, all $n(n-1)$ moves can be compared directly from the stored statistics (Appendix B.1). Each step accepts the move that most lowers the mean score on D_{C}. A trajectory takes at most eight steps, stopping when the improvement falls below a tolerance.

We retain checkpoints at depths 0, 4, and 8. Because finite-budget preparation depends on how remaining correlations are distributed among qubit pairs, a loader might prefer an earlier checkpoint over the lowest-score one. Deduplicating the resulting permutations yields a bank $\mathcal{C}$ of at most twelve candidates. Circuit fitting then determines the most useful refinement.

3.3 Selection and Reuse

Each candidate is ultimately assessed by the preparation fidelity it achieves on D_{F}. For efficiency, a fast loader starts with a fitted real product state and applies L real two-qubit rotation blocks. Each block is an orthogonal matrix in $\mathrm{SO}(4)$, compiled with two CX gates (Wei & Di, 2012), giving $G = 2L$ (Appendix C.3). The qubit-pair sequence $\tau(P)$ is learned for each address on D_{A}. This gives every candidate a topology adapted to its coordinates. Product factors and block parameters are fitted separately for each target.

The mean fidelity on the screening set selects the final address:

$$\hat{\mu}(P) = \frac{1}{|D_{\mathrm{F}}|} \sum_{x \in D_{\mathrm{F}}} \widehat{F}_{2n,\tau(P)}(P; x), \qquad P^{\star} = \arg\max_{P \in \mathcal{C}} \hat{\mu}(P). \tag{11}$$

We use a reference budget of $G_{\mathrm{ref}} = 2n$ (i.e., n rotation blocks). At this shallow depth, the loader exploits pairwise structure beyond a product approximation, while remaining sensitive enough for the address to strongly influence preparation. The selection cost is one topology construction and $|D_{\mathrm{F}}|$ target fits per candidate.

For a new input, $P^{\star}$ is applied as a classical data reordering in $O(d)$ time. The reordered vector can then be passed to any loader at the desired budget. For instance, larger fast-loader circuits can extend the learned pair sequence using construction data, while independent loaders fit their own preparation parameters to each input. Algorithm 1 in Appendix B.2 gives the complete procedure.

4 Experiments

We evaluate preparation fidelity and gate requirements, then use the prepared states in kernel and QNN classifiers. Construction ablations examine coordinate and sign information, and structural measurements relate the observed gains to the encoded data.

4.1 Setup

Benchmarks and protocol. Eight benchmarks cover four representation types: grayscale handwriting (MNIST, KMNIST; $n = 10$; LeCun et al., 1998; Clanuwat et al., 2018), frozen ResNet50 (He et al., 2016) features (CIFAR-10, CIFAR-100; $n = 11$; Krizhevsky, 2009), frozen RoBERTa (Liu et al., 2019; Reimers & Gurevych, 2019) sentence embeddings (SST-2, AG News; $n = 10$; Socher et al., 2013; Zhang et al., 2015), and larger grayscale images (OrganAMNIST-64 at 64×64 and Galaxy10-128 at 128×128; $n = 12, 14$; Yang et al., 2023; Leung & Bovy, 2024). Inputs are ℓ_2-normalized, and encoders are fixed before address learning.

In each of the five independent runs, we draw unlabeled construction, screening and topology sets ($|D_{\mathrm{C}}| = 256$, $|D_{\mathrm{F}}| = 128$, $|D_{\mathrm{A}}| = 64$) from the training split. Addresses are selected by fast-loader fidelity and reused unchanged by all downstream loaders and models. The natural address uses the same loader, topology recipe and 64 fitting sweeps per target. Downstream training and validation splits are paired across addresses. Fast-loader evaluation uses 1000 fixed external targets (872 for

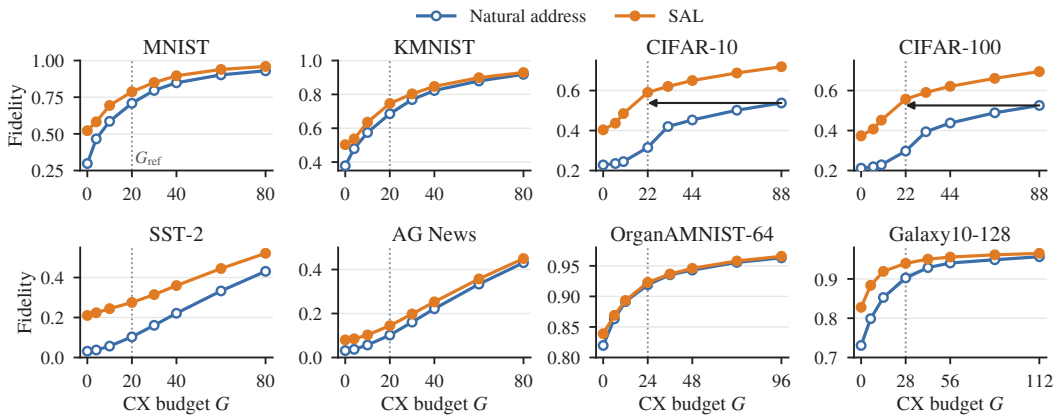


Figure 3: Fast-loader fidelity versus CX budget (mean over five runs; ±1 s.d. bands are mostly narrower than the lines). Dotted lines: G_{ref}. Arrows: on CIFAR, SAL at 22 CX exceeds the natural address at 88 CX.

Table 1: Mean loading fidelity at G_{ref} and KRR test accuracy (%) over five runs. Standard deviations are shown for approximate-kernel accuracy. Δ_{inf} is the relative reduction in mean infidelity. Both addresses share the ideal kernel.

			Loading fidelity at G_{ref}			KRR test accuracy (%)				
							G_{low}		G_{ref}	
Dataset	n	$G_{\text{low}}/G_{\text{ref}}$	Natural	SAL	Δ_{inf} (%)	Ideal	Natural	SAL	Natural	SAL
MNIST	10	4/20	0.709	**0.787**	27.0	96.72	75.82±0.40	**86.46**±0.76	94.04±0.83	**94.90**±0.39
KMNIST	10	4/20	0.685	**0.746**	19.4	82.96	51.14±0.48	**61.04**±1.15	75.04±1.64	**78.18**±1.04
CIFAR-10	11	6/22	0.315	**0.591**	40.3	89.92	53.36±0.66	**63.04**±2.02	68.08±0.79	**72.90**±2.09
CIFAR-100	11	6/22	0.297	**0.557**	36.9	67.10	23.30±0.70	**26.72**±0.93	33.68±1.08	**40.18**±0.87
SST-2	10	4/20	0.103	**0.275**	19.2	90.09	64.89±0.89	**67.20**±1.64	76.70±1.74	**79.56**±1.25
AG News	10	4/20	0.102	**0.144**	4.8	89.64	53.96±1.52	**61.12**±0.68	75.46±0.92	**76.90**±0.64
Organ-64	12	6/24	0.919	**0.923**	4.8	65.40	55.60±0.76	**57.56**±0.80	63.84±0.32	**64.16**±0.94
Galaxy-128	14	6/28	0.903	**0.940**	38.0	39.88	27.14±0.91	**37.14**±1.17	40.96±0.96	**42.66**±1.03

SST-2; Wang et al., 2019) at $G/n \in \{0, 0.5, 1, 2, 3, 4, 6, 8\}$, rounded to even CX counts; $G_{\text{ref}} = 2n$ and $G_{\text{low}} \approx n/2$ denote the reference and shallow budgets. We report means and standard deviations over runs. Detailed preprocessing and complete hyperparameters are collected in the appendices. We use JAX (Bradbury et al., 2018) for numerical computation, Keras (Chollet et al., 2015) for frozen visual features, and Qiskit (Javadi-Abhari et al., 2024) for hardware execution.

4.2 Preparation Cost

Loading performance. SAL improves mean preparation fidelity on all eight datasets throughout the measured budget range (Figure 3). At G_{ref} in run 0, SAL improves fidelity for 85–100% of individual targets within each dataset. The gains also reduce gate requirements at matched fidelity. At 22 CX, SAL raises CIFAR-10 fidelity from 0.315 to 0.591 and CIFAR-100 from 0.297 to 0.557, exceeding the natural address at 88 CX (0.538 and 0.525). The SAL product state on SST-2 matches the natural address at approximately 38 CX by linear interpolation, while on Galaxy10-128, SAL at 14 CX outperforms the natural address at 28 CX.

Independent loaders. Transfer to MPS (Ran, 2020), AQCE (Shirakawa et al., 2024) and AQER (Zhang et al., 2026) is evaluated on MNIST, CIFAR-10 and SST-2, using 50 fixed targets per dataset and three runs. Each loader retains its native budgets and optimizer. At the middle of its three budgets, AQCE fidelity improves from 0.728 to 0.831 on MNIST, 0.400 to 0.624 on CIFAR-10, and 0.138 to 0.320 on SST-2. MPS shows the same direction of improvement, including 0.369 to 0.609 on CIFAR-10. Across the fast loader and these independent loaders, SAL lowers mean infidelity in all 91 dataset–budget–loader comparisons (Figure 4a).

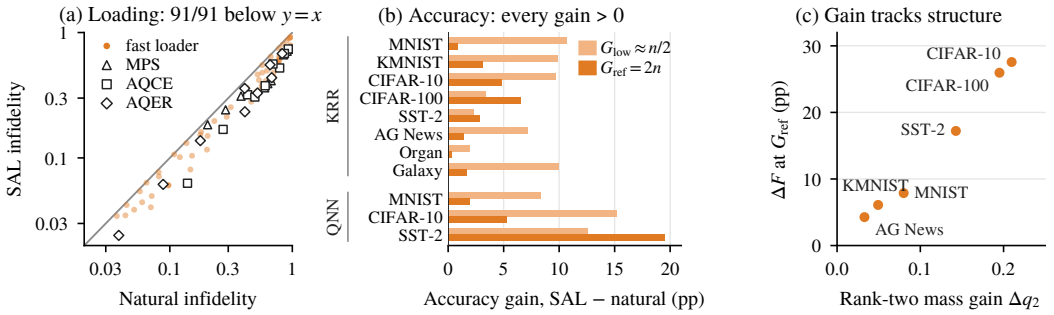


Figure 4: Consistency across loaders, models and datasets. (a) Infidelity under SAL versus the natural address in all 91 dataset–budget–loader comparisons: fast loader on eight datasets (dots) and MPS, AQCE, AQER on MNIST, CIFAR-10 and SST-2 (open markers). (b) Five-run mean accuracy gain at the shallow (light) and reference (dark) budgets. (c) Fidelity gain at G_{ref} versus the gain in rank-two mass over four balanced cuts, for the six datasets with $n \in \{10, 11\}$.

4.3 Downstream Learning

Kernel classifier. We use centered multiclass kernel ridge regression (KRR) with the amplitude kernel k, 5000 training and 1000 validation examples per run, and all native classes. Under approximate loading, kernel entries are squared overlaps of the prepared states. We fit centered one-hot labels with a centered Gram matrix, using a validation-selected ridge parameter shared across class outputs. Since a shared address preserves ideal overlaps, both addresses use the same ideal-kernel fit.

SAL is more accurate in all 32 comparisons across eight datasets and four scheduled budgets, $G/n \in \{0, 0.5, 1, 2\}$. Table 1 and Figure 4b report the shallow and reference budgets, with shallow gains of 10.64 points on MNIST, 9.68 on CIFAR-10 and 10.00 on Galaxy10-128. The gaps generally narrow toward G_{ref}. The ideal kernel provides a common reference; on Galaxy10-128, both approximate kernels exceed its accuracy at G_{ref} (40.96% and 42.66% vs. ideal 39.88%).

Trainable quantum readout. For MNIST, CIFAR-10 and SST-2, we train a three-layer hardware-efficient QNN (Kandala et al., 2017). Each layer applies R_Y and R_Z rotations to every qubit followed by a CX ring. Single-qubit Z expectations feed a linear head with one output per class. Circuit angles and the head are trained jointly by cross-entropy with Adam (Kingma & Ba, 2015), using validation loss for early stopping. The readout's $3n$ CX gates are not counted in G.

Models are trained separately on ideal natural, ideal SAL, approximate natural and approximate SAL inputs. Approximate inputs are evaluated at G_{low} and G_{ref}. SAL is more accurate in all six approximate-input dataset–budget groups. On MNIST at $G = 4$, it reaches 73.18% against 64.84%, while the ideal SAL model is 0.88 points below the ideal natural model. The models trained on ideal inputs establish a baseline for how much the readout natively prefers a given address. The accuracy drop from each ideal model to its approximate counterpart defines the approximation penalty. On CIFAR-10 at G_{ref}, the approximate gain is close to the ideal-address gain, whereas MNIST at G_{low} benefits mainly from a smaller approximation penalty.

Eight-qubit hardware. On `ibm_kingston`, MNIST digits 0–3 at 16×16 are encoded on eight qubits at $G = 4$. A trained two-layer readout uses R_Y and R_Z rotations with nearest-neighbor entangling gates and a final R_Y rotation on each qubit. Each of four disjoint bit pairs encodes a class index; class probabilities average their frequencies over 1024 shots per image, and the largest probability determines the prediction. Over three runs with 160 test images each, SAL raises accuracy from $57.92 \pm 2.60\%$ to $71.04 \pm 1.91\%$ and macro-F1 from $56.28 \pm 2.20\%$ to $69.06 \pm 3.85\%$.

4.4 Construction Information

For the coordinate and sign interventions, only the vectors supplied to candidate construction are altered; topology fitting, screening and evaluation use the original vectors (Table 2).

Permuting every construction row independently preserves its norm and amplitude multiset but removes common coordinate identities. The gain on SST-2 disappears, and those on MNIST and

Table 2: Construction-information ablations: loading fidelity (mean ± s.d., five runs) at G_{low} and G_{ref}. ρ is the fraction of construction rows with aligned coordinates. n/a: nonnegative inputs ($|x| = x$). —: not measured.

	MNIST		CIFAR-10		SST-2	
Construction input	G=4	G=20	G=6	G=22	G=4	G=20
Natural address	0.466±.000	0.709±.018	0.235±.000	0.315±.006	0.037±.000	0.103±.001
SAL	0.582±.007	0.787±.005	0.437±.002	0.591±.002	0.224±.009	0.275±.005
Aligned rows ($\rho = 0$)	0.465±.001	0.716±.004	0.238±.007	0.394±.014	0.037±.001	0.103±.001
Aligned rows ($\rho = 0.25$)	0.583±.005	0.791±.004	0.418±.004	0.573±.001	0.211±.006	0.264±.009
Magnitudes only ($|x|$)	n/a	n/a	n/a	n/a	0.049±.001	0.138±.002
$|D_{\text{C}}| = 32$	—	—	—	0.581±.004	—	0.246±.016

CIFAR-10 shrink, while restoring one quarter of aligned rows recovers most of the improvement. Relative signs also matter: constructing from magnitudes $|x|$ lowers SST-2 fidelity from 0.224 to 0.049 at $G = 4$, consistent with the sign dependence of the construction autocorrelation. Address construction needs few examples: 32 rows reach 0.581 on CIFAR-10, close to 0.591 with 256.

Sorting each input by sign and magnitude makes it nearly trivial to load (fidelity ≥ 0.992 at G_{ref}), while mapping distinct inputs with identical amplitude multisets to the same state. Even with ideal loading, sorted representations yield KRR accuracies of only 34.80%, 27.92% and 53.58% on MNIST, CIFAR-10 and SST-2, compared with 96.72%, 89.92% and 90.09% for any shared address. Sharing the permutation retains the original classification geometry.

4.5 Structural Analysis

A shared permutation preserves the second-moment spectrum while rearranging its principal directions across qubits. We examine the resulting structure through balanced register cuts and preparation error decomposition.

The rank-two Schmidt mass q_2 is the sum of the two largest squared singular values across a cut, equal to the best rank-two approximation fidelity. Averaged over 128 complete external targets and four fixed balanced cuts, its increase orders the six datasets with $n \in \{10, 11\}$ in the same way as their fidelity gains (Spearman $r_s = 1.00$ over datasets; Figure 4c). CIFAR-10 rises from 0.425 to 0.634, whereas AG News changes from 0.213 to 0.246. These measurements describe how correlations are distributed across the register; a single-cut spectrum alone does not determine the cost of a full circuit.

For error attribution, the four leading second-moment directions define a shared subspace. Infidelity decomposes exactly into errors within this subspace, within its complement, and in the terms coupling them. Handwriting gains involve all three terms, visual-feature gains mainly the residual and cross terms, and text gains the residual term.

The starting representation also determines the available headroom. OrganAMNIST-64 contains large uniform regions, with a permutation-invariant median uniform-state fidelity of 0.83. Its mean fidelity at G_{ref} rises only from 0.919 to 0.923. Galaxy10-128 has a median uniform-state fidelity of 0.73 and achieves a 38% reduction in remaining infidelity. The benefit of address learning therefore depends on the input structure alongside its dimension.

5 Conclusion

SAL improves amplitude preparation by reordering coordinates with a shared address, which preserves the ideal kernel of a dataset. SAL combines data-driven proposals, local structural refinement and circuit-based selection, with gains across loaders, downstream classifiers and a hardware evaluation. The selected address is reused as classical preprocessing, amortizing its construction cost over subsequent inputs. Its benefit depends on shared data structure and the preparation quality of the natural address. The present construction uses real amplitudes and a restricted address family; connectivity-aware selection, complex amplitudes and joint address–circuit design are natural extensions.

AI USE STATEMENT

In this work, we used generative AI tools to assist in writing the proofs in Appendices A and B, to implement our methods, to clean and reformat the datasets, to assist with parts of the experimental design, to analyze and interpret experimental results, and to translate text. We have not used generative AI tools to generate synthetic datasets, develop theoretical frameworks, formulate the mathematical claims or propose hypotheses, and qualitative data analysis is not applicable to this work. Additionally, we used generative AI tools to search for related literature, to help create figures, and to edit the paper for readability. We have reviewed all AI-assisted work. Additions to the proofs suggested by AI tools were checked line by line and kept only after the authors had verified them. AI-assisted code, including the data preprocessing, was reviewed and run by the authors, and all reported results and figures come from these runs. AI-assisted analyses were checked by the authors against the raw experimental outputs, and the final experimental design was decided by the authors. All translated and edited text was reviewed by the authors, and all references found with AI assistance were checked against the original sources. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The code for our method is included in the supplementary material. The datasets and preprocessing are described in Appendix C.1, and the experimental protocol and hyperparameters in Appendix C.2. The four loaders are described in Appendices C.3 and C.5, and the hardware experiments in Appendix E.3. Proofs of the theoretical results are given in Appendices A and B.

REFERENCES

Eric R. Anschuetz. A unified theory of quantum neural network loss landscapes. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/f3680ca31afcb9076846950dbc72c7e3-Abstract-Conference.html.

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

James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Yash Katariya, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: Composable transformations of Python+NumPy programs, 2018. URL https://github.com/jax-ml/jax.

M. Cerezo, Guillaume Verdon, Hsin-Yuan Huang, Lukasz Cincio, and Patrick J. Coles. Challenges and opportunities in quantum machine learning. *Nature Computational Science*, 2:567–576, 2022. doi: 10.1038/s43588-022-00311-3.

François Chollet et al. Keras, 2015. URL https://keras.io.

Tarin Clanuwat, Mikel Bober-Irizar, Asanobu Kitamoto, Alex Lamb, Kazuaki Yamamoto, and David Ha. Deep learning for classical Japanese literature. *arXiv preprint arXiv:1812.01718*, 2018.

Iris Cong, Soonwon Choi, and Mikhail D. Lukin. Quantum convolutional neural networks. *Nature Physics*, 15(12):1273–1278, 2019. doi: 10.1038/s41567-019-0648-8.

Sreetama Das, Stefano Martina, and Filippo Caruso. The role of data embedding in equivariant quantum convolutional neural networks. *Quantum Machine Intelligence*, 6:82, 2024. doi: 10.1007/s42484-024-00215-7.

Yuxuan Du, Yibo Yang, Dacheng Tao, and Min-Hsiu Hsieh. Problem-dependent power of quantum neural networks on multiclass classification. *Physical Review Letters*, 131(14):140601, 2023. doi: 10.1103/PhysRevLett.131.140601.

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

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

Tak Hur and Daniel K. Park. Understanding generalization in quantum machine learning with margins. In *International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 26338–26360, 2025. URL https://proceedings.mlr.press/v267/hur25a.html.

Ali Javadi-Abhari, Matthew Treinish, Kevin Krsulich, Christopher J. Wood, Jake Lishman, Julien Gacon, Simon Martiel, Paul D. Nation, Lev S. Bishop, Andrew W. Cross, Blake R. Johnson, and Jay M. Gambetta. Quantum computing with Qiskit, 2024. URL https://arxiv.org/abs/2405.08810. arXiv:2405.08810.

Hyeongjun Jeon, Kyungmin Lee, Dongkyu Lee, Bongsang Kim, and Taehyun Kim. Optimal qubit mapping search for encoding classical data into matrix product state representation with minimal loss, 2024. URL https://arxiv.org/abs/2406.06935v2. arXiv:2406.06935v2.

Bernhard Jobst, Kevin Shen, Carlos A. Riofrío, Elvira Shishenina, and Frank Pollmann. Efficient MPS representations and quantum circuits from the Fourier modes of classical image data. *Quantum*, 8: 1544, 2024. doi: 10.22331/q-2024-12-03-1544.

Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549:242–246, 2017. doi: 10.1038/nature23879.

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.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3519–3529, 2019.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

Jonas M. Kübler, Simon Buchholz, and Bernhard Schölkopf. The inductive bias of quantum kernels. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86:2278–2324, 1998.

Cong Lei, Yuxuan Du, Peng Mi, Jun Yu, and Tongliang Liu. Neural auto-designer for enhanced quantum kernels. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/4d9944ab3330fe6af8efb9260aa9f307-Abstract-Conference.html.

W. Henry Leung and Jo Bovy. Galaxy10 DECaLS. Zenodo, version 1, 2024. URL https://zenodo.org/records/10845026.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

Johannes Jakob Meyer, Marian Mularski, Elies Gil-Fuster, Antonio Anna Mele, Francesco Arzani, Alissa Wilms, and Jens Eisert. Exploiting symmetry in variational quantum machine learning. *PRX Quantum*, 4:010328, 2023. doi: 10.1103/PRXQuantum.4.010328.

Mikko Möttönen, Juha J. Vartiainen, Ville Bergholm, and Martti M. Salomaa. Transformation of quantum states using uniformly controlled rotations. *Quantum Information and Computation*, 5: 467–473, 2005.

Alex Murphy, Joel Zylberberg, and Alona Fyshe. Correcting biased centered kernel alignment measures in biological and artificial neural networks. *arXiv preprint arXiv:2405.01012*, 2024.

Kouhei Nakaji, Shumpei Uno, Yohichi Suzuki, Rudy Raymond, Tamiya Onodera, Tomoki Tanaka, Hiroyuki Tezuka, Naoki Mitsuda, and Naoki Yamamoto. Approximate amplitude encoding in shallow parameterized quantum circuits and its application to financial market indicators. *Physical Review Research*, 4:023136, 2022. doi: 10.1103/PhysRevResearch.4.023136.

Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. doi: 10.1017/CBO9781139814782.

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

John Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, 2018.

Shi-Ju Ran. Encoding of matrix product states into quantum circuits of one- and two-qubit gates. *Physical Review A*, 101:032310, 2020.

Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *EMNLP-IJCNLP*, pp. 3982–3992, 2019.

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

Vivek V. Shende, Stephen S. Bullock, and Igor L. Markov. Synthesis of quantum-logic circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 25:1000–1010, 2006.

Tomonori Shirakawa, Hiroshi Ueda, and Seiji Yunoki. Automatic quantum circuit encoding of a given arbitrary quantum state. *Physical Review Research*, 6:043008, 2024.

Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *EMNLP*, pp. 1631–1642, 2013.

Xiaoming Sun, Guojing Tian, Shuai Yang, Pei Yuan, and Shengyu Zhang. Asymptotically optimal circuit depth for quantum state preparation and general unitary synthesis. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 42(10):3301–3314, 2023. doi: 10.1109/TCAD.2023.3244885.

Gábor J. Székely and Maria L. Rizzo. Partial distance correlation with methods for dissimilarities. *The Annals of Statistics*, 42:2382–2412, 2014.

Supanut Thanasilp, Samson Wang, M. Cerezo, and Zoë Holmes. Exponential concentration in quantum kernel methods. *Nature Communications*, 15:5200, 2024. doi: 10.1038/s41467-024-49287-w.

Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1804.07461.

Shuteng Wang, Christian Theobalt, and Vladislav Golyanik. Quantum visual fields with neural amplitude encoding. In *Advances in Neural Information Processing Systems*, volume 38, 2025a. doi: 10.52202/085713-5297. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/e895155ee8e30f07f4e5e3d37475aa49-Abstract-Conference.html.

Xin Wang, Hanxiao Tao, and Rebing Wu. Predictive performance of deep quantum data re-uploading models. In *International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 64491–64524, 2025b.

Hai-Rui Wei and Yao-Min Di. Decomposition of orthogonal matrix and synthesis of two-qubit and three-qubit orthogonal gates. *Quantum Information and Computation*, 12:262–270, 2012.

Maxwell T. West, Martin Sevior, and Muhammad Usman. Reflection equivariant quantum neural networks for enhanced image classification. *Machine Learning: Science and Technology*, 4(3): 035027, 2023. doi: 10.1088/2632-2153/acf096.

Hao Xiong, Yehui Tang, Yunlin He, Wei Tan, and Junchi Yan. Node2ket: Efficient high-dimensional network embedding in quantum Hilbert space. In *International Conference on Learning Representations*, 2024.

Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. MedMNIST v2: A large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Scientific Data*, 10:41, 2023.

Xinyu Ye, Hao Xiong, Jianhao Huang, Ziang Chen, Jia Wang, and Junchi Yan. On designing general and expressive quantum graph neural networks with applications to MILP instance representation. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/ab59d149fc0c2c9039d3e3049f7914b1-Abstract-Conference.html.

Kaining Zhang, Xinbiao Wang, Yuxuan Du, Min-Hsiu Hsieh, and Dacheng Tao. AQER: A scalable and efficient data loader for digital quantum computers. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=0zKfU1rsXd.

Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-level convolutional networks for text classification. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

APPENDIX OVERVIEW

	Content	Page
A	Notation, kernel bounds, proofs of Theorems 1 and 2, score bounds	14
B	Walsh statistics, selection algorithms, candidate bank in practice, classical workload	17
C	Datasets, sample roles, hyperparameters, loaders and their circuits (Figures 6–9)	19
D	Complete loading curves, independent loaders, per-target results, example targets	23
E	Kernel classification, QNN (Figure 19), eight-qubit hardware (Figure 20)	29
F	Shared-coordinate, sign, sample-size and instance-wise ablations	32
G	Loader views, descriptor subspaces, exact infidelity decomposition, sample relations	35

A MATHEMATICAL FOUNDATIONS

A.1 NOTATION AND CONVENTIONS

Table 3 collects the notation used throughout the paper. Input coordinates are indexed by binary strings, with qubit 0 as the least significant bit. We use $\oplus$ for addition over $\mathbb{F}_2$.

Table 3: Notation.

Symbol	Meaning
$d = 2^n$, n	input dimension and number of qubits
x, $\lvert x\rangle$	unit-norm real vector and its amplitude state
P, $P^\star$	common address (coordinate permutation) and selected address
Q, A	outer coordinate base and invertible binary address matrix, $A \in \mathrm{GL}(n, 2)$
w_q, v_q	row q of A and column q of A^{-1}
$\mathcal{C}$	deduplicated candidate bank, $\lvert\mathcal{C}\rvert \le 12$
D_{C}, D_{F}, D_{A}	construction, screening and topology sets, $D_{\mathrm{A}} \subset D_{\mathrm{C}}$, $D_{\mathrm{C}} \cap D_{\mathrm{F}} = \varnothing$
G, L	logical CX budget and number of $\mathrm{SO}(4)$ blocks, $G = 2L$
$G_{\mathrm{ref}} = 2n$, $G_{\mathrm{low}} \approx n/2$	reference and shallow budgets
$\tau(P)$	shared ordered qubit-pair sequence learned for address P
$F_G^\star$, $\widehat{F}_{G,\tau}$	optimal and achieved fidelity (squared overlap)
ρ_q, S	one-qubit reduced state and structural score
$c_y(v)$, $z_y(w)$	binary autocorrelation and parity-energy statistic
M, λ_j, u_j	uncentered second moment, its eigenvalues and eigenvectors
Π_r, $\Pi_r^\perp = I - \Pi_r$	top-r descriptor projector and its complement
q_χ	rank-χ Schmidt mass across a bipartition
K, $\widetilde{K}$	ideal and approximate amplitude kernels
a_U	off-diagonal, U-centered kernel alignment
E_{c}, E_{r}, $E_\times$	common, residual and cross-subspace infidelity terms

The candidate bank, selected address and fast-loader pair sequence are learned from training data. The selected address and pair sequence are shared across evaluation inputs, with larger budgets extending the learned sequence. Product factors and block matrices are fitted separately for each target.

We measure fidelity by the squared overlap $F(x, z) = |x^\top z|^2$ between normalized real vectors and preparation cost by the logical CX count under the decompositions in Appendix C.3. The address is applied as a classical reordering and adds no CX gates.

A.2 KERNEL CONTROL

Proof of Eq. (4). Let $\rho_i = |x_i\rangle\langle x_i|$ and $\sigma_i = |\hat{x}_i\rangle\langle\hat{x}_i|$ for the prepared states $\hat{x}_i$. The ideal and approximate kernels are $K_{ij} = \mathrm{tr}(\rho_i\rho_j)$ and $\widetilde{K}_{ij} = \mathrm{tr}(\sigma_i\sigma_j)$, and their difference splits as

$$K_{ij} - \widetilde{K}_{ij} = \mathrm{tr}[(\rho_i - \sigma_i)\rho_j] + \mathrm{tr}[\sigma_i(\rho_j - \sigma_j)].$$

For a traceless Hermitian matrix D and an effect $0 \preceq E \preceq I$, $|\,\mathrm{tr}(DE)| \le \|D\|_1/2$, and the pure-state trace distance satisfies $\|\rho_i - \sigma_i\|_1/2 = \sqrt{1 - F_i}$. This gives the entrywise bound, which may

be capped at one because both kernels lie in $[0, 1]$. Writing $\epsilon_i = 1 - F_i$, Cauchy–Schwarz gives

$$\frac{1}{N^2}\|K - \widetilde{K}\|_F^2 \le \frac{1}{N^2}\sum_{i,j}\left(\sqrt{\epsilon_i} + \sqrt{\epsilon_j}\right)^2 \le \frac{2}{N}\sum_i \epsilon_i + \frac{2}{N^2}\Big(\sum_i \sqrt{\epsilon_i}\Big)^2 \le 4\bar{\epsilon}. \qquad \square$$

Preparation error therefore bounds the kernel perturbation, although the bound can be loose at low fidelity. Classification also depends on which pairwise relations are affected; Appendix G.6 examines these relations empirically.

Regularized kernel prediction. For centered positive semidefinite matrices K_c and $\widetilde{K}_c$, centered labels Y_c and $\gamma = N\lambda > 0$, let $\alpha = (K_c + \gamma I)^{-1} Y_c$ and $\widetilde{\alpha} = (\widetilde{K}_c + \gamma I)^{-1} Y_c$. The resolvent identity gives

$$\|\alpha - \widetilde{\alpha}\|_F \le \frac{\|K_c - \widetilde{K}_c\|_2 \|Y_c\|_F}{\gamma^2},$$

and for centered test rows k and $\tilde{k}$ the logits satisfy

$$\|k\alpha - \tilde{k}\widetilde{\alpha}\|_2 \le \frac{\|k - \tilde{k}\|_2 \|Y_c\|_F}{\gamma} + \frac{\|\tilde{k}\|_2 \|K_c - \widetilde{K}_c\|_2 \|Y_c\|_F}{\gamma^2}.$$

These bounds hold at fixed regularization and labels. If every logit changes by at most η, a prediction with a top-two margin above 2η remains unchanged. Kernel error and classification margins thus jointly determine prediction stability.

A.3 Shared-Address Separation

The separation follows from a Schmidt-rank bound for a maximally entangled target. Let $|\Omega\rangle = D^{-1/2}\sum_{j=1}^{D} e^{i\theta_j}|j\rangle|j\rangle$. Every normalized bipartite state $|\phi\rangle$ of Schmidt rank at most R satisfies $|\langle\phi|\Omega\rangle|^2 \le \min\{1, R/D\}$. To see this, reshape ϕ into a coefficient matrix B with $\|B\|_F = 1$ and $\operatorname{rank} B \le R$. The coefficient matrix of Ω is $D^{-1/2}V$ with V diagonal and unitary. The trace inequality and Cauchy–Schwarz over the nonzero singular values give

$$|\langle\phi|\Omega\rangle| = D^{-1/2}|\operatorname{tr}(B^\dagger V)| \le D^{-1/2}\|B\|_* \le \sqrt{R/D}\,\|B\|_F.$$

Squaring gives the bound, capped at one by normalization. Keeping and renormalizing any $R \le D$ Schmidt terms attains equality.

A gate acting entirely on one side of a bipartition preserves the Schmidt rank. A CX that crosses the cut is a sum of two product operators, $\mathrm{CX} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X$, so applied to a sum of R product terms it produces at most $2R$ terms. Starting from a product state, a circuit with $g_{A|B}$ crossing CX gates therefore has Schmidt rank at most $2^{g_{A|B}} \le 2^G$.

Proof of Theorem 1. Define $P|a, b\rangle = |a \oplus b, b\rangle$. This bijection is its own inverse and is binary-linear with block matrix $\left(\begin{smallmatrix} I & I \\ 0 & I \end{smallmatrix}\right)$. With $\chi_s(a) = \prod_i s_i^{a_i}$,

$$P|x_s\rangle = 2^{-r/2}\sum_a \chi_s(a)|0\rangle^{\otimes r}|a\rangle = |0\rangle^{\otimes r} \otimes \bigotimes_{i=1}^{r} \frac{|0\rangle + s_i|1\rangle}{\sqrt{2}}.$$

The same P serves every sign vector, so all reordered targets are product states. Under the natural address, each target has $D = 2^r$ equal Schmidt coefficients across $A|B$. The Schmidt-rank bound above therefore gives $F \le 2^G/2^r$, capped at one. $\square$

A.4 Recovery by the Correlation Construction

Proof of Theorem 2. Let $H = \{(a, a) : a \in \mathbb{F}_2^r\} \subset \mathbb{F}_2^{2r}$. Each state in Eq. (5) is supported on H with amplitude $2^{-r/2}\chi_s(a)$ at (a, a). If $v \notin H$, the cosets H and $H \oplus v$ are disjoint and $c_{x_s}(v) = 0$. If $v = (t, t) \in H$, multiplicativity of the character gives

$$c_{x_s}(v) = 2^{-r}\sum_{a \in \mathbb{F}_2^r} \chi_s(a)\chi_s(a \oplus t) = \chi_s(t).$$

Hence $|c_{x_s}(v)| = \mathbf{1}[v \in H]$ for every s, and averaging absolute correlations over any nonempty subset leaves this indicator unchanged.

All directions in H therefore rank ahead of all directions outside H, and the greedy independence rule selects an ordered basis $h_1, \dots, h_r$ of H before completing it to a basis $h_1, \dots, h_{2r}$ of $\mathbb{F}_2^{2r}$. With $V = [h_1, \dots, h_{2r}]$ and $A = V^{-1}$, the new address b reads the old index $j = Vb = \bigoplus_k b_k h_k$, so the support H occupies exactly the addresses with $b_{r+1} = \cdots = b_{2r} = 0$. Writing $h_i = (t_i, t_i)$ for $i \le r$, the amplitude at such an address carries the sign

$$\chi_s\Big(\bigoplus_{i=1}^{r} b_i t_i\Big) = \prod_{i=1}^{r} \chi_s(t_i)^{b_i},$$

and the reordered state is $\bigotimes_{i=1}^{r}(|0\rangle + \chi_s(t_i)|1\rangle)/\sqrt{2} \otimes |0\rangle^{\otimes r}$, a product state for every member. □

The result holds for any deterministic ordering of tied directions and for sign patterns absent from the construction subset. It applies to the natural-base correlation start retained at depth zero, before local refinement and fidelity screening.

A.5 Shared-Direction Lower Bound

The second moment $M = \mathbb{E}[xx^\dagger]$ is positive semidefinite with unit trace. For a fixed normalized ϕ and any shared permutation, linearity of the trace gives

$$\mathbb{E}_x |\langle \phi | Px \rangle|^2 = \operatorname{tr}\big(|\phi\rangle\langle\phi|\, PMP^\dagger\big) = \sum_j \lambda_j |\langle \phi | Pu_j \rangle|^2.$$

Every term is nonnegative. If a budget-G state ϕ_G approximates Pu_1 with fidelity f_1, then $\mathbb{E}_x F_G^\star(P; x) \ge \lambda_1 f_1$: the same ϕ_G is a feasible preparation for every target. Optimizing its overlap with Pu_1 gives Eq. (9).

A.6 Structural Score and Product-State Bounds

The one-qubit reduced states under Eq. (7) depend on the autocorrelation and the parity-energy statistic

$$z_y(w) = \sum_j (-1)^{w \cdot j} y_j^2. \tag{12}$$

Fix a base order and let y be a normalized real vector in that order. Under $b = Aj$, the two populations of output qubit q correspond to the parity groups defined by $w_q \cdot j$, whose energy difference is $z_y(w_q)$. Flipping output bit q changes the old index by $v_q = A^{-1} e_q$, so twice the off-diagonal entry of the reduced state is $c_y(v_q)$:

$$\rho_q = \frac{1}{2}\begin{pmatrix} 1 + z_y(w_q) & c_y(v_q) \\ c_y(v_q) & 1 - z_y(w_q) \end{pmatrix}, \qquad \lambda_\pm(\rho_q) = \frac{1 \pm \sqrt{c_y(v_q)^2 + z_y(w_q)^2}}{2}. \tag{13}$$

The off-diagonal entry is real because the targets are real. Summing λ_- over qubits gives

$$S(P_{Q,A} x) = \sum_q \frac{1 - \sqrt{c_{Qx}(v_q)^2 + z_{Qx}(w_q)^2}}{2}. \tag{14}$$

Positivity of the reduced state keeps the square root at most one.

Proof of Eq. (10). Let $\epsilon_q = \lambda_{\min}(\rho_q)$, let $|a_q\rangle$ be a top eigenvector of ρ_q, and let E_q be the corresponding one-qubit projector extended to the full space. The projectors commute, and the scalar inequality $1 - \prod_q e_q \le \sum_q (1 - e_q)$ for $e_q \in \{0, 1\}$ implies the operator inequality $I - \prod_q E_q \preceq \sum_q (I - E_q)$. Taking the expectation in y gives $1 - |\langle \bigotimes_q a_q | y \rangle|^2 \le \sum_q \epsilon_q$, and the best product state is at least as good as this one. For the lower bound, every product state is also a product across the cut $q|\bar{q}$, so its squared overlap with y is at most the largest squared Schmidt coefficient of that cut, $1 - \epsilon_q$. This holds for every q, hence $1 - F_{\text{prod}}^\star(y) \ge \max_q \epsilon_q$. □

A small score thus guarantees an accurate product approximation. At finite G, preparation fidelity also depends on how the remaining correlations are distributed among qubit pairs and how well the circuit is fitted. Candidates with similar scores can therefore have different loading fidelities.

B Address Construction

B.1 Walsh Statistics and Binary Updates

Let $H_{w,j} = (-1)^{w \cdot j}$ be the unnormalized Walsh–Hadamard matrix over $\mathbb{F}_2^n$, so that $H^2 = dI$. The parity statistic is $z_y = H(y \odot y)$. For the autocorrelation, let $h = Hy$; then

$$\frac{1}{d}\sum_w (-1)^{w \cdot v} h_w^2 = \frac{1}{d}\sum_{j,k} y_j y_k \sum_w (-1)^{w \cdot (v \oplus j \oplus k)} = \sum_j y_j y_{j \oplus v} = c_y(v),$$

because the inner sum equals d when its argument vanishes and zero otherwise. The full autocorrelation array is therefore $c_y = H[(Hy) \odot (Hy)]/d$. A butterfly implementation computes both arrays in $O(dn)$ operations per vector, after which candidate moves use the stored statistics. The zero direction is excluded when constructing an invertible basis.

Updating the binary map. For distinct a, b let $E_{ba} = I + e_b e_a^\top$ over $\mathbb{F}_2$. Since $(e_b e_a^\top)^2 = 0$ and addition equals subtraction, $E_{ba}^{-1} = E_{ba}$. A row addition $A' = E_{ba} A$ gives

$$A'^{-1} = A^{-1} E_{ba}, \qquad w'_b = w_b \oplus w_a, \qquad v'_a = v_a \oplus v_b,$$

and leaves all other rows of A and columns of A^{-1} unchanged. With $s_y(w, v) = \big(1 - \sqrt{z_y(w)^2 + c_y(v)^2}\big)/2$, the exact change of the score is

$$\Delta_{a \to b}(y) = s_y(w_a, v_a \oplus v_b) + s_y(w_b \oplus w_a, v_b) - s_y(w_a, v_a) - s_y(w_b, v_b).$$

SAL averages $\Delta_{a \to b}$ over D_{C}, compares all $n(n-1)$ ordered pairs, and accepts the smallest value only if it is below -10^{-10}. Each accepted move lowers the construction-set mean of S.

Each row addition preserves invertibility. The identity start uses the original coordinate bits, while the correlation start ranks nonzero directions by $\mathbb{E}_{D_{\mathrm{C}}}|c_y(v)|$ and retains those that increase the binary rank. The first n retained directions form the columns of V, giving $A = V^{-1}$. The group has $|\mathrm{GL}(n, 2)| = \prod_{j=0}^{n-1}(2^n - 2^j)$ elements, far fewer than the $(2^n)!$ coordinate permutations but still too many for exhaustive search. Row additions connect the whole group; the eight-step refinement explores a small neighborhood of each start. Since the reference base can introduce a permutation outside this binary-linear family, deduplication compares the full gathers and records all originating bases, roots and checkpoints.

B.2 Candidate Generation and Selection

Algorithm 1 gives the full SAL procedure. Algorithms 2 and 3 detail candidate construction and fidelity-based selection, respectively.

Algorithm 1 Shared Address Learning

Require: Unlabeled $D_{\mathrm{C}}, D_{\mathrm{F}}$ with $D_{\mathrm{C}} \cap D_{\mathrm{F}} = \varnothing$; $D_{\mathrm{A}} \subset D_{\mathrm{C}}$

1: Build the natural and principal-reference bases from D_{C}; form the identity and correlation roots in each.
2: Refine each start by row additions that lower S; keep depths $0, 4, 8$ and deduplicate into $\mathcal{C}$.
3: For every $P \in \mathcal{C}$, learn $\tau(P)$ on D_{A}, fit all targets in D_{F} at $G_{\mathrm{ref}} = 2n$, and select $P^\star$ by Eq. (11).
4: **return** $P^\star$, frozen for all evaluation inputs, budgets, loaders and downstream models.

B.3 Candidate Bank in Practice

The bank contains at most $2 \times 2 \times 3 = 12$ distinct candidates, whose screening fidelities are shown in Figure 5. The preferred base depends on the representation: all runs select the natural base for handwriting and larger images, and the principal-reference base for visual and sentence features. Averaged over runs, the selected screening fidelity and the best fidelity in the other base are 0.778 versus 0.586 on MNIST and 0.593 versus 0.485 on CIFAR-10. The correlation start is selected in 31 of the 40 runs, while the preferred refinement depth varies across datasets. Table 4 gives the base, root and depth selected in each run.

Algorithm 2 Candidate bank construction

Require: Normalized construction vectors D_{C}, checkpoints $J = \{0, 4, 8\}$
Compute the dominant eigenvector u of the uncentered second moment and fix its sign at its largest-magnitude coordinate.
$Q_0 \leftarrow I$; $\quad Q_1 \leftarrow \mathrm{argsort}_j\big(\mathbf{1}[u_j < 0],\, -|u_j|,\, j\big)$; $\quad \mathcal{C} \leftarrow \varnothing$.
for $Q \in \{Q_0, Q_1\}$ **do**
    Compute c_{Qx} and z_{Qx} with Walsh transforms for all $x \in D_{\mathrm{C}}$.
    Build the roots $A_0 = I$ and $A_1 = V^{-1}$ from independent correlation directions.
    for $A \in \{A_0, A_1\}$ **do**
        Record the initial gather with its base, root and depth.
        for $t = 1, \ldots, 8$ **do**
            $(a, b) \leftarrow \arg\min_{a \neq b} \mathbb{E}_{D_{\mathrm{C}}} \Delta_{a \to b}(Qx)$.
            if $\mathbb{E}_{D_{\mathrm{C}}} \Delta_{a \to b} < -10^{-10}$ **then**
                $w_b \leftarrow w_b \oplus w_a$; $\quad v_a \leftarrow v_a \oplus v_b$.
            else
                stop the trajectory and reuse its final gather at the remaining checkpoints.
            end if
            If $t \in J$, record the current full gather.
        end for
    end for
end for
Deduplicate full gathers, retaining their bases, roots and depths, and **return** $\mathcal{C}$.

Algorithm 3 Selection by loading fidelity and budget reuse

Require: $\mathcal{C}$, $D_{\mathrm{A}} \subset D_{\mathrm{C}}$, D_{F} with $D_{\mathrm{F}} \cap D_{\mathrm{C}} = \varnothing$, $G_{\mathrm{ref}} = 2n$
for $P \in \mathcal{C}$ **do**
    Learn n shared pair positions $\tau(P)$ from PD_{A} only.
    Fit target-specific product factors and blocks to every Px, $x \in D_{\mathrm{F}}$.
    $\hat{\mu}(P) \leftarrow |D_{\mathrm{F}}|^{-1} \sum_{x \in D_{\mathrm{F}}} \widehat{F}_{2n,\tau(P)}(P; x)$.
end for
$P^\star \leftarrow \arg\max_P \hat{\mu}(P)$ with a deterministic tie order.
Extend the topology of $P^\star$ on D_{C} for larger budgets, keeping $P^\star$ fixed.
return $P^\star$ for downstream learning and for independent loaders.

Table 4: Selected start in each run: outer base (N: natural, R: principal reference), binary root (I: identity, C: correlation) and refinement depth. The correlation root is selected in 31 of the 40 runs.

Dataset	Run 0	Run 1	Run 2	Run 3	Run 4
MNIST	N·C·4	N·C·0	N·C·0	N·I·8	N·C·4
KMNIST	N·I·8	N·C·0	N·C·0	N·I·8	N·C·0
CIFAR-10	R·I·4	R·C·0	R·C·4	R·C·4	R·C·4
CIFAR-100	R·I·4	R·I·0	R·C·0	R·C·0	R·C·8
SST-2	R·C·0	R·C·0	R·C·0	R·I·8	R·I·8
AG News	R·C·4	R·I·0	R·C·4	R·C·4	R·C·4
OrganAMNIST-64	N·C·0	N·C·0	N·C·0	N·C·0	N·C·0
Galaxy10-128	N·C·0	N·C·0	N·C·4	N·C·0	N·C·0

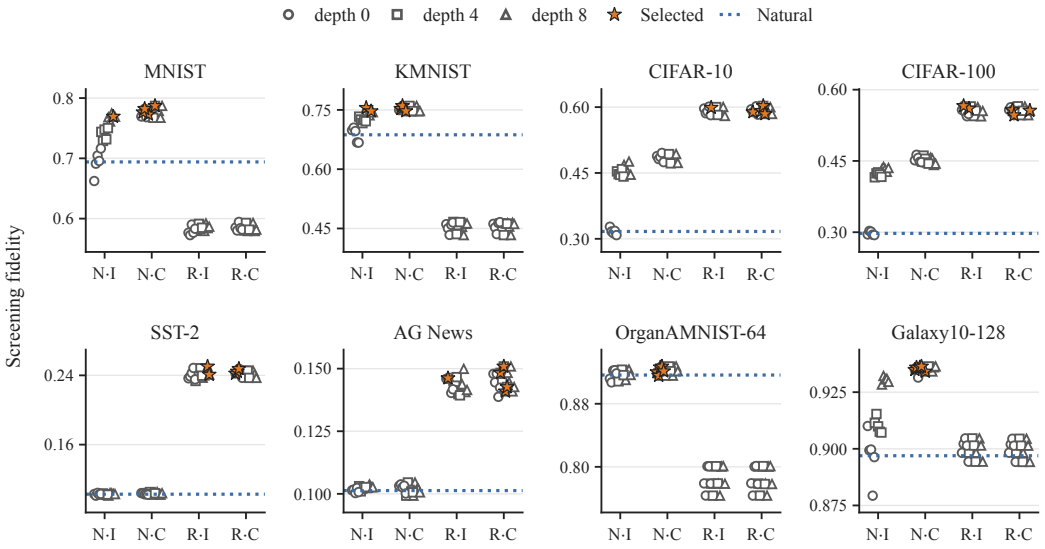


Figure 5: Candidate bank on the screening set, five runs. The horizontal position groups the four starts (N/R: natural or principal-reference base; I/C: identity or correlation root), the marker shape gives the refinement depth, stars mark the selected address and the dotted line is the natural address (mean over runs). The best start depends on the representation: reference-base starts fall far below the natural address on handwriting and OrganAMNIST-64, and lie above every natural-base start on the visual and sentence features.

B.4 Classical Workload

SAL operates on explicitly available classical vectors. It stores the learned address as a gather of d indices and applies it in $O(d)$ time per input. Computing the Walsh statistics takes $O(|D_C|\, dn)$ operations per base, and each refinement step compares $n(n-1)$ two-term updates averaged over D_C. The principal-reference order additionally requires the leading training direction and one coordinate sort. With the reported screening size, twelve distinct candidates require at most $12 \times 128 = 1536$ target fits at G_{ref}, together with topology construction. This address-learning cost is incurred once per dataset and run. Subsequent target-specific circuit fits are part of the loading procedure.

C Data, Protocols and Circuits

C.1 Datasets and Frozen Representations

Table 5 lists the vectors supplied to loading and learning.

Table 5: Dataset suite.

Dataset	d	n	Classes	Input representation
MNIST (LeCun et al., 1998)	1024	10	10	bilinear 32×32 grayscale pixels
KMNIST (Clanuwat et al., 2018)	1024	10	10	bilinear 32×32 grayscale pixels
CIFAR-10 (Krizhevsky, 2009)	2048	11	10	frozen ResNet50 pooled features
CIFAR-100 (Krizhevsky, 2009)	2048	11	100	frozen ResNet50 pooled features
SST-2 (Socher et al., 2013)	1024	10	2	frozen RoBERTa sentence features
AG News (Zhang et al., 2015)	1024	10	4	frozen RoBERTa sentence features
OrganAMNIST-64 (Yang et al., 2023)	4096	12	11	native 64×64 grayscale pixels
Galaxy10-128 (Leung & Bovy, 2024)	16384	14	10	128×128 grayscale galaxy images

Grayscale images. We resize the native MNIST and KMNIST images from 28×28 to 32×32 by bilinear interpolation, flatten them in row-major order and apply ℓ_2 normalization. This gives 1024 coordinates per image; Appendix E.3 describes the hardware variant. OrganAMNIST-64 uses the native 64×64 archive with its original split boundaries, row-major flattening and per-image normalization. Galaxy10-128 uses the duplicate-free Galaxy10 DECaLS release with 17,675 images in ten classes; each 256×256 RGB image is converted to grayscale, resized to 128×128 with Lanczos

interpolation, flattened row by row and normalized, keeping all 16,384 pixels. A fixed stratified split reserves 1000 external images, and each run draws 5000 training and 1000 internal-validation images from the remaining pool.

Frozen visual features. CIFAR-10 and CIFAR-100 use global-average-pooled features from the Keras ImageNet ResNet50 (He et al., 2016). We resize the RGB images bilinearly to 224×224, apply the prescribed BGR conversion and channel-mean subtraction, and retain all 2048 pooled coordinates before normalization. These features follow a ReLU and are nonnegative.

Frozen sentence features. SST-2 and AG News use the same `all-roberta-large-v1` sentence encoder (Reimers & Gurevych, 2019; Liu et al., 2019). Tokenization truncates at 256 tokens. We average the final hidden states over attention-mask positions and normalize the result to obtain 1024 signed coordinates. Duplicate training sentences and training copies of external sentences are removed before splitting. For SST-2, the 872 labeled validation sentences form the external set, while internal validation uses examples drawn from the official training data.

C.2 Experimental Protocol

Five outer seeds determine the stratified training and internal-validation draws and the associated address-learning subsets. All runs and conditions use the same fixed stratified external set. For MNIST and SST-2, we reserve 512 training examples for method development and exclude them from the main training pool. Tables 6 and 7 give the sample counts and fixed settings.

Table 6: Training and evaluation counts.

Dataset	Train	Int. val.	External
MNIST, KMNIST	5000	1000	1000
CIFAR-10/100	5000	1000	1000
SST-2	5000	1000	872
AG News	5000	1000	1000
OrganAMNIST-64	5000	1000	1000
Galaxy10-128	5000	1000	1000

Table 7: Fixed numerical and selection settings.

Quantity	Setting
$\lvert D_{\mathrm{C}}\rvert$ / $\lvert D_{\mathrm{F}}\rvert$ / $\lvert D_{\mathrm{A}}\rvert$	256 / 128 / 64
Starts / checkpoints	2×2; depths 0, 4, 8
Topology refinement	5 sweeps on D_{A}
Fitting	64 sweeps per target
Loading budgets G/n	0, 0.5, 1, 2, 3, 4, 6, 8
Kernel budgets G/n	0, 0.5, 1, 2
KRR ridge grid	$\{1, 3\} \times 10^{-5}, \ldots, 10^{-2}$
External loaders	3 runs, 50 fixed targets
Descriptors	512 training vectors, $r = 4$
Structural / relational	128 / 512 external vectors

In the sample-size study, construction sets of size $N_{\mathrm{C}} \in \{32, 64, 128, 256, 512\}$ are nested within each run. The 128 screening examples are disjoint from the largest construction set, and the topology subset has size $\min\{N_{\mathrm{C}}, 64\}$. The plotted construction-set size therefore corresponds to $N_{\mathrm{C}} + 128$ observations used for address learning.

For a metric $m_{s,i}$ from run s and target i, we first average over targets, $\bar{m}_s = N^{-1} \sum_i m_{s,i}$. Tables report $S^{-1} \sum_s \bar{m}_s$ and the sample standard deviation over runs. Since the external targets are fixed, variation in mean loading fidelity reflects the learned address and topology; Appendix D.3 reports variation across inputs. Gains are differences of condition means, and relative infidelity reduction is $(\overline{F}_{\mathrm{S}} - \overline{F}_{\mathrm{N}})/(1 - \overline{F}_{\mathrm{N}})$.

C.3 Fast Loader and Gate Accounting

For a target $y = Px$, the fast loader prepares

$$|\psi(v, U; \tau)\rangle = U_{L,\tau_L} \cdots U_{1,\tau_1} \bigotimes_{q=0}^{n-1} \big(v_{q,0}|0\rangle + v_{q,1}|1\rangle\big), \qquad U_\ell \in \mathrm{SO}(4), \quad \|v_q\| = 1.$$

The pair sequence τ is shared, while v and U are fitted separately for each target. Product factors allow negative amplitudes, so at $L = 0$ the loader fits a real product state. Initialization uses dominant one-qubit factors, which are then refined through overlap environments.

Fitting the circuit. With all other gates fixed, contracting the forward and backward states around a block gives a real 4×4 environment E, and the local objective $\mathrm{tr}(U^\top E)$ is linear in the block.

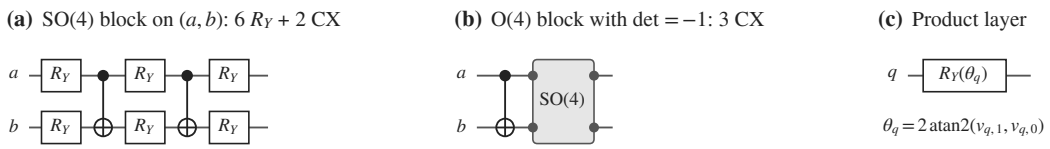


Figure 6: Gate-level form of the loader blocks in the implementation. (a) An $\mathrm{SO}(4)$ block compiles into six R_Y rotations and two CX gates with the first qubit of the pair as control. (b) An orthogonal block with determinant -1 adds one CX. (c) Each qubit of the signed product layer is prepared by one R_Y rotation.

If $E = W\Sigma V^\top$, its maximizer over $\mathrm{SO}(4)$ is $U^\star = W\,\mathrm{diag}(1,1,1,\det(WV^\top))V^\top$. We alternate product-factor and block updates in both directions for 64 sweeps, yielding the achieved fidelity $\widehat{F}_{G,\tau}$. Pair positions are chosen greedily from all unordered qubit pairs on D_{A}, with five refinement sweeps. Budgets above G_{ref} extend the n-block prefix using construction data.

Gate counts. Each real $\mathrm{SO}(4)$ block compiles into six R_Y rotations and two CX gates, with the first qubit of the pair as control (Wei & Di, 2012). Figure 6 shows the implemented sequence. We decompose blocks separately and count only CX gates, giving $G = 2L$ for L fast-loader blocks; for example, $G = 20$ corresponds to ten blocks. An orthogonal block with determinant -1, as used by the external loaders, is written as $O = (OC)C$ for a CX matrix C and costs three CX gates. Table 8 gives the conversion for each loader.

C.4 Learned Fast-Loader Circuits

Figures 7 and 8 show example topologies at G_{ref} and pair usage across all datasets and runs. On CIFAR-10 and CIFAR-100, SAL tends to connect qubits that are closer in index, with a usage-weighted distance of 1.9, compared with 4.1 and 3.7 under the natural address. The change is smaller on sentence features: 3.1 versus 3.7 on SST-2, and 3.2 versus 3.6 on AG News. The natural topologies for handwriting and Galaxy10-128 already use mostly neighboring qubits; SAL spreads their blocks slightly, with a mean distance of 2.7 versus 2.0 on MNIST.

C.5 Independent Loading Algorithms

Each independent loader receives the normalized target x or its fixed permutation $P^\star x$. We use real-state implementations of the cited procedures, with the same native budgets and numerical updates under both addresses. Figure 9 shows an example circuit for each loader.

Residual rank-two MPS layers. A rank-two tensor-train factorization is constructed by sequential singular-value truncation of the current residual target, and the resulting isometries are completed to real orthogonal two-qubit blocks in a fixed staircase of $n-1$ blocks (Ran, 2020). After a layer is constructed, its inverse is applied to obtain the next residual, and the new preparation layer is inserted at the corresponding inner position of the composed circuit. One, two and four layers are evaluated.

AQCE with residual-basis initialization. For each native block count B, two-qubit reduced density matrices determine descending eigenbases that initialize exactly B real $\mathrm{O}(4)$ blocks (Shirakawa et al., 2024). The removal sequence is reversed into preparation order to form C_0. With $r = C_0^\top x$ and $j^\star = \arg\max_j |r_j|$, the preparation starts from $C_0|j^\star\rangle$, where the basis state is prepared by X gates. With the basis fixed, 100 bidirectional all-pair polar sweeps refine the blocks and their pair assignments. The determinant of each fitted block decides whether its expansion uses two or three CX gates.

AQER disentangling and fidelity refinement. AQER first constructs a sequence of local rotations and R_{ZZ} gates that reduces the sum of one-qubit Rényi-2 entropies $E(y) = -\sum_q \log \mathrm{tr}[\rho_q(y)^2]$, then approximates the remaining state by a product and refines the reconstructed preparation circuit against fidelity (Zhang et al., 2026). Entropy minimization uses a zero-centered Nelder–Mead simplex with tolerances 10^{-4} and a limit of 1000 function evaluations. The final refinement uses 2000 Adam steps with learning rate 0.01, moments $(0.9, 0.99)$ and $\epsilon = 10^{-8}$. We evaluate $T = 20, 40, 80$ native R_{ZZ} gates with $G = 2T$; prefixes of the disentangling stage may be reused, and each budget receives its own refinement. Each R_{ZZ} gate compiles into CX·R_Z·CX.

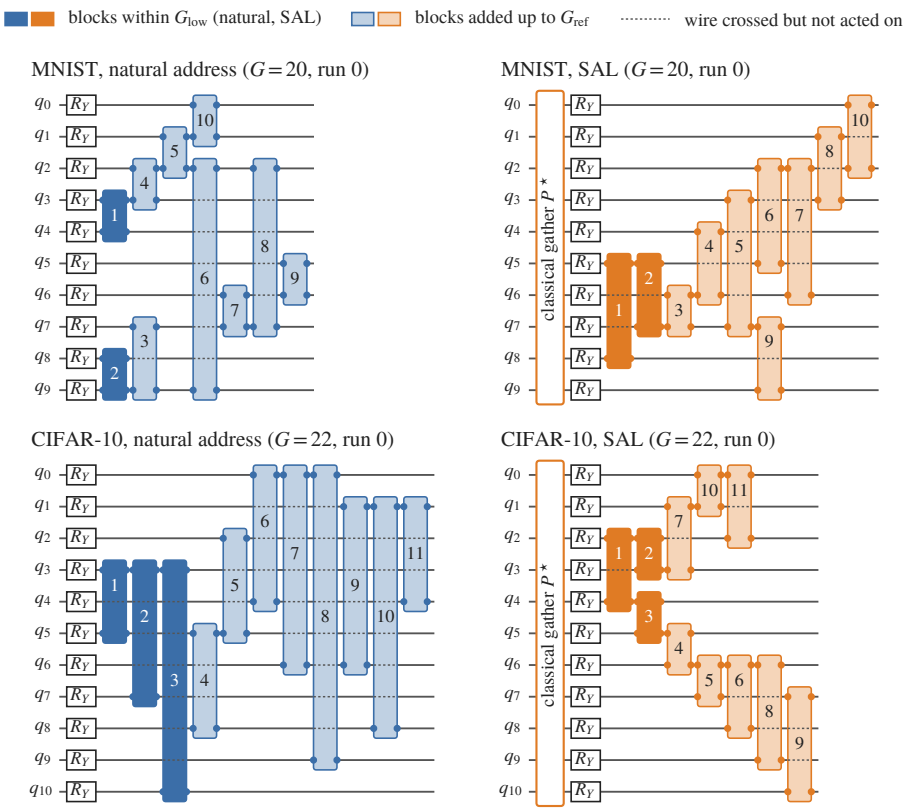


Figure 7: Shared $\mathrm{SO}(4)$ topologies learned for the fast loader at the reference budget (run 0). Each block acts on the two qubits marked by dots; dotted segments mark wires it crosses. Numbers give the application order after the product layer; dark blocks fit within G_{low}. Under SAL a classical gather $P^\star$ precedes the circuit.

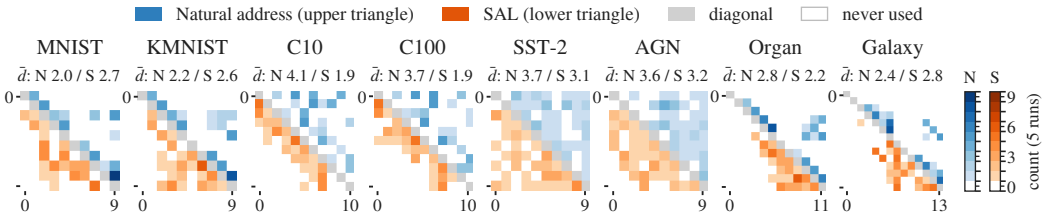


Figure 8: Qubit pairs used by the first n blocks at G_{ref}, summed over five runs; upper triangle: natural address, lower triangle: SAL, white: never used. Both color bars share one count scale. $\bar{d} = \sum_{i<j} c_{ij}|i-j| / \sum_{i<j} c_{ij}$ is the usage-weighted distance between the two qubits of a block (N: natural address; S: SAL).

Table 8: Resource units. Native counts are kept alongside the logical CX counts used on the horizontal axes.

Structure	Native count	Logical CX count
Fast $\mathrm{SO}(4)$ blocks	L	$2L$
Rank-two MPS layers	L_{MPS}	$2L_{\text{MPS}}(n-1)$
AQCE $\mathrm{O}(4)$ blocks	B	$2B + \#\{\det O < 0\}$
AQER R_{ZZ} gates	T	$2T$

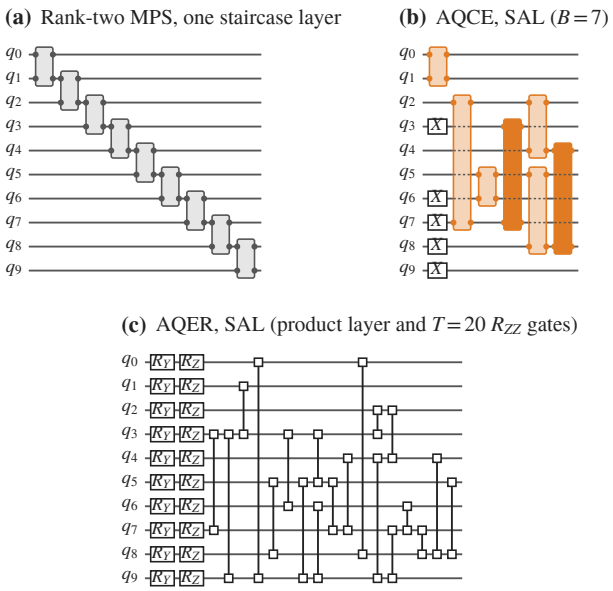


Figure 9: Independent loaders on MNIST ($n = 10$). (a) One layer of the rank-two MPS staircase. (b) AQCE circuit for one external target under the SAL address: basis preparation followed by seven $\mathrm{O}(4)$ blocks (dark: determinant -1). (c) AQER circuit for one external target under the SAL address: product layer and 20 R_{ZZ} interactions; the local rotations after each interaction are omitted.

D Complete Loading Results

D.1 Resource–Fidelity Curves

Table 9 gives the fast-loader fidelities at the eight budgets plotted in Figure 3. Together with the independent loaders, all 91 dataset–budget–loader comparisons favor SAL, as shown by their positions below the diagonal in Figure 4a. Figure 10 expresses the gains as matching natural-address budgets. On CIFAR-10 and CIFAR-100, the natural address needs $5.3n$ and $4.6n$ to match SAL at the nominal n budget (10 CX after rounding); even its largest measured budget, $8n$, does not match SAL at $2n$.

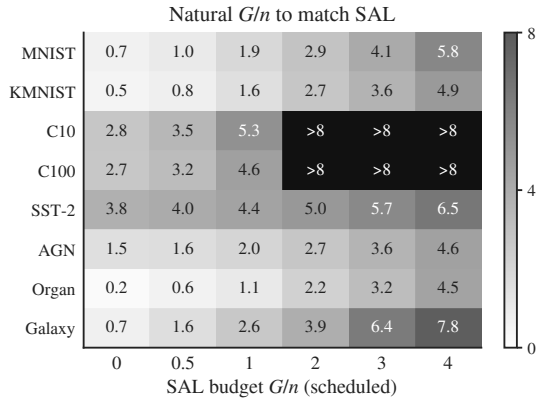


Figure 10: Natural-address budget G/n needed to reach the mean fidelity that SAL attains at each scheduled budget, interpolated linearly between measured natural budgets. Black cells (> 8) exceed the largest measured natural budget, $8n$.

For a SAL fidelity f bracketed by consecutive natural-address means $F_{\mathrm{N}}(G_j) \le f \le F_{\mathrm{N}}(G_{j+1})$, the matching budget is

$$G_{\mathrm{N}}(f) = G_j + (G_{j+1} - G_j)\frac{f - F_{\mathrm{N}}(G_j)}{F_{\mathrm{N}}(G_{j+1}) - F_{\mathrm{N}}(G_j)}.$$

We interpolate the five-run mean curves using the actual even CX counts. The columns in the figure retain the scheduled SAL budget densities: for $n = 11$, the nominal n and $3n$ budgets correspond to 10 and 32 CX. Fidelities beyond the measured natural-address curve are marked $> 8n$.

Table 9: Fast-loader fidelity at all eight budgets (mean ± s.d. over five runs). ΔF is the difference of mean fidelities in percentage points; Δ_{inf} is the relative reduction of remaining infidelity in percent. Natural entries with zero s.d. are deterministic: at small budgets the natural fit is identical across runs.

Dataset	G	Natural	SAL	ΔF	Δ_{inf}
MNIST	0	0.2979±.0000	0.5217±.0051	22.39	31.9
	4	0.4658±.0000	0.5820±.0067	11.62	21.8
	10	0.5855±.0028	0.6937±.0062	10.82	26.1
	20	0.7085±.0179	0.7873±.0053	7.88	27.0
	30	0.7964±.0048	0.8501±.0111	5.37	26.4
	40	0.8482±.0075	0.8960±.0078	4.78	31.5
	60	0.9025±.0026	0.9392±.0081	3.67	37.7
	80	0.9300±.0056	0.9596±.0069	2.96	42.3
KMNIST	0	0.3782±.0000	0.5028±.0023	12.46	20.0
	4	0.4791±.0098	0.5372±.0027	5.80	11.1
	10	0.5747±.0027	0.6356±.0042	6.10	14.3
	20	0.6853±.0172	0.7463±.0071	6.10	19.4
	30	0.7690±.0057	0.8031±.0074	3.41	14.8
	40	0.8228±.0052	0.8470±.0086	2.43	13.7
	60	0.8793±.0019	0.8986±.0119	1.93	16.0
	80	0.9186±.0022	0.9289±.0114	1.03	12.7
CIFAR-10	0	0.2279±.0000	0.4034±.0017	17.55	22.7
	6	0.2351±.0000	0.4368±.0015	20.17	26.4
	10	0.2446±.0001	0.4848±.0017	24.02	31.8
	22	0.3152±.0065	0.5909±.0018	27.57	40.3
	32	0.4201±.0128	0.6206±.0034	20.05	34.6
	44	0.4529±.0009	0.6501±.0026	19.72	36.0
	66	0.5016±.0015	0.6878±.0026	18.62	37.4
	88	0.5378±.0018	0.7191±.0020	18.13	39.2
CIFAR-100	0	0.2118±.0000	0.3731±.0019	16.13	20.5
	6	0.2187±.0000	0.4073±.0027	18.86	24.1
	10	0.2286±.0002	0.4521±.0017	22.34	29.0
	22	0.2973±.0003	0.5569±.0012	25.96	36.9
	32	0.3940±.0014	0.5904±.0010	19.64	32.4
	44	0.4377±.0025	0.6212±.0023	18.35	32.6
	66	0.4887±.0018	0.6609±.0012	17.23	33.7
	88	0.5255±.0017	0.6949±.0015	16.95	35.7

Dataset	G	Natural	SAL	ΔF	Δ_{inf}
SST-2	0	0.0317±.0000	0.2103±.0134	17.87	18.4
	4	0.0375±.0002	0.2238±.0086	18.62	19.4
	10	0.0569±.0003	0.2442±.0055	18.73	19.9
	20	0.1031±.0013	0.2753±.0047	17.22	19.2
	30	0.1609±.0018	0.3145±.0054	15.36	18.3
	40	0.2212±.0010	0.3597±.0057	13.85	17.8
	60	0.3329±.0006	0.4449±.0051	11.20	16.8
	80	0.4308±.0004	0.5210±.0043	9.02	15.8
AG News	0	0.0315±.0000	0.0800±.0015	4.86	5.0
	4	0.0370±.0001	0.0850±.0016	4.80	5.0
	10	0.0566±.0003	0.1026±.0015	4.59	4.9
	20	0.1018±.0009	0.1445±.0013	4.27	4.8
	30	0.1607±.0014	0.1978±.0015	3.71	4.4
	40	0.2216±.0005	0.2528±.0015	3.12	4.0
	60	0.3334±.0004	0.3572±.0008	2.38	3.6
	80	0.4310±.0007	0.4498±.0008	1.87	3.3
Organ-64	0	0.8197±.0000	0.8390±.0000	1.93	10.7
	6	0.8632±.0045	0.8689±.0018	0.57	4.2
	12	0.8914±.0017	0.8934±.0022	0.19	1.8
	24	0.9190±.0022	0.9229±.0011	0.39	4.8
	36	0.9354±.0012	0.9366±.0015	0.13	1.9
	48	0.9430±.0005	0.9459±.0012	0.28	5.0
	72	0.9556±.0014	0.9579±.0006	0.23	5.2
	96	0.9631±.0008	0.9656±.0008	0.25	6.9
Galaxy-128	0	0.7307±.0000	0.8278±.0021	9.71	36.0
	6	0.7989±.0004	0.8836±.0005	8.47	42.1
	14	0.8527±.0002	0.9192±.0003	6.65	45.1
	28	0.9027±.0112	0.9397±.0009	3.70	38.0
	42	0.9288±.0012	0.9504±.0005	2.15	30.3
	56	0.9406±.0022	0.9557±.0007	1.51	25.4
	84	0.9489±.0003	0.9614±.0008	1.25	24.4
	112	0.9566±.0005	0.9652±.0006	0.86	19.8

On MNIST, SAL retains an advantage throughout the measured range and at 60 CX exceeds the natural address at 80 CX. Although the absolute gap narrows, the reduction in remaining infidelity reaches 42.3% at the largest budget. KMNIST shows a weaker high-budget effect: the mean gain remains positive at 80 CX, but two of the five runs favor the natural address. An address selected at the reference budget therefore need not be the best at every depth.

The two CIFAR datasets share an encoder but differ in image distribution and class count. For both, the fidelity gap grows from the product-state point to the reference budget, then narrows slowly. SAL at 22 CX exceeds the natural address at 88 CX. At the largest budget, it reduces the remaining infidelity by 39.2% on CIFAR-10 and 35.7% on CIFAR-100.

SST-2 and AG News also share a fixed encoder, and their natural-address curves nearly coincide. SAL produces different gains: SST-2 improves by 0.172 at the reference budget and by 0.090 at 80 CX, while AG News shows smaller improvements throughout. Both SAL curves continue to rise over the final 20-CX interval, by 0.076 on SST-2 and 0.093 on AG News. Absolute fidelities nevertheless remain modest at the largest measured budget.

Among the larger images, OrganAMNIST-64 starts from a natural product-state fidelity of 0.820 and reaches 0.963 at 96 CX, leaving gains of 0.1 to 1.9 points for SAL. Galaxy10-128 improves at every budget, from 0.731 to 0.828 at zero CX and from 0.903 to 0.940 at the reference budget. SAL at 14 CX already exceeds the natural address at 28 CX.

D.2 Independent Loaders

Figure 11 and Table 10 report every native budget on 50 fixed targets. SAL improves mean fidelity in each comparison. Natural-address fits are deterministic, while the learned SAL addresses vary across

runs. For AQCE, the fitted block determinants produce slightly different CX counts under the two addresses at the same native budget.

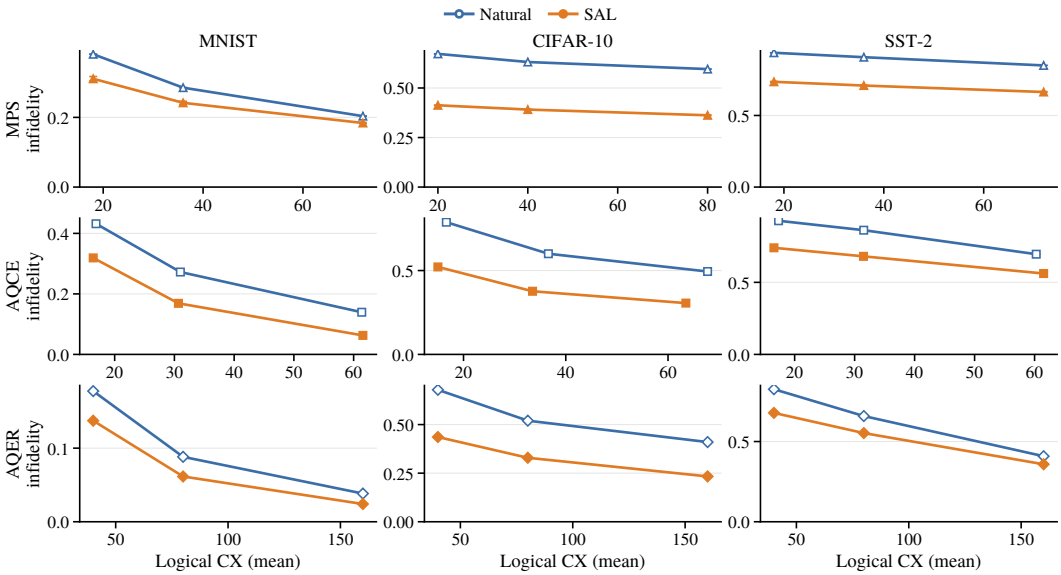


Figure 11: Independent loaders at matched native budgets. The horizontal axis is the mean CX count; error bars show ± 1 s.d. over three runs.

Table 10: Independent-loader fidelity (mean $\pm$ s.d., three runs). $G_{\mathrm{N}}/G_{\mathrm{S}}$ gives CX counts after decomposition; native budgets follow Appendix C.5.

Dataset	Loader	Native budget	$G_{\mathrm{N}}/G_{\mathrm{S}}$	Natural	SAL
MNIST	MPS	1	18.00/18.00	0.6188±.0000	0.6892±.0065
MNIST	MPS	2	36.00/36.00	0.7147±.0000	0.7582±.0045
MNIST	MPS	4	72.00/72.00	0.7964±.0000	0.8161±.0017
MNIST	AQCE	7	16.88/16.41	0.5681±.0000	0.6810±.0024
MNIST	AQCE	13	31.00/30.68	0.7277±.0000	0.8312±.0019
MNIST	AQCE	27	61.36/61.55	0.8606±.0000	0.9372±.0021
MNIST	AQER	20	40.00/40.00	0.8225±.0000	0.8627±.0014
MNIST	AQER	40	80.00/80.00	0.9118±.0000	0.9385±.0002
MNIST	AQER	80	160.00/160.00	0.9617±.0000	0.9759±.0001
CIFAR-10	MPS	1	20.00/20.00	0.3278±.0000	0.5873±.0012
CIFAR-10	MPS	2	40.00/40.00	0.3689±.0000	0.6088±.0014
CIFAR-10	MPS	4	80.00/80.00	0.4046±.0000	0.6377±.0019
CIFAR-10	AQCE	7	16.64/14.99	0.2118±.0000	0.4785±.0025
CIFAR-10	AQCE	15	36.60/33.50	0.3998±.0000	0.6235±.0048
CIFAR-10	AQCE	29	67.74/63.49	0.5057±.0000	0.6940±.0048
CIFAR-10	AQER	20	40.00/40.00	0.3211±.0000	0.5640±.0043
CIFAR-10	AQER	40	80.00/80.00	0.4799±.0000	0.6708±.0013
CIFAR-10	AQER	80	160.00/160.00	0.5900±.0000	0.7667±.0025
SST-2	MPS	1	18.00/18.00	0.0621±.0000	0.2649±.0026
SST-2	MPS	2	36.00/36.00	0.0931±.0000	0.2910±.0030
SST-2	MPS	4	72.00/72.00	0.1505±.0000	0.3361±.0029
SST-2	AQCE	7	17.32/16.55	0.0728±.0000	0.2604±.0047
SST-2	AQCE	13	31.52/31.48	0.1382±.0000	0.3196±.0032
SST-2	AQCE	27	60.28/61.51	0.3047±.0000	0.4387±.0012
SST-2	AQER	20	40.00/40.00	0.1739±.0000	0.3214±.0021
SST-2	AQER	40	80.00/80.00	0.3411±.0000	0.4473±.0025
SST-2	AQER	80	160.00/160.00	0.5915±.0000	0.6417±.0014

D.3 Per-Target Results

Preparation fidelity varies substantially across targets. Figure 12 pairs the two addresses at G_{ref}, with one point per external target. In run 0, SAL improves 98.2% of MNIST targets, 96.1% of KMNIST, all CIFAR targets, 99.4% of SST-2, 90.3% of AG News, 85.0% of OrganAMNIST-64 and 99.1% of Galaxy10-128. At G_{low}, the smallest fractions are 81.4% on KMNIST and 75.5% on OrganAMNIST-64.

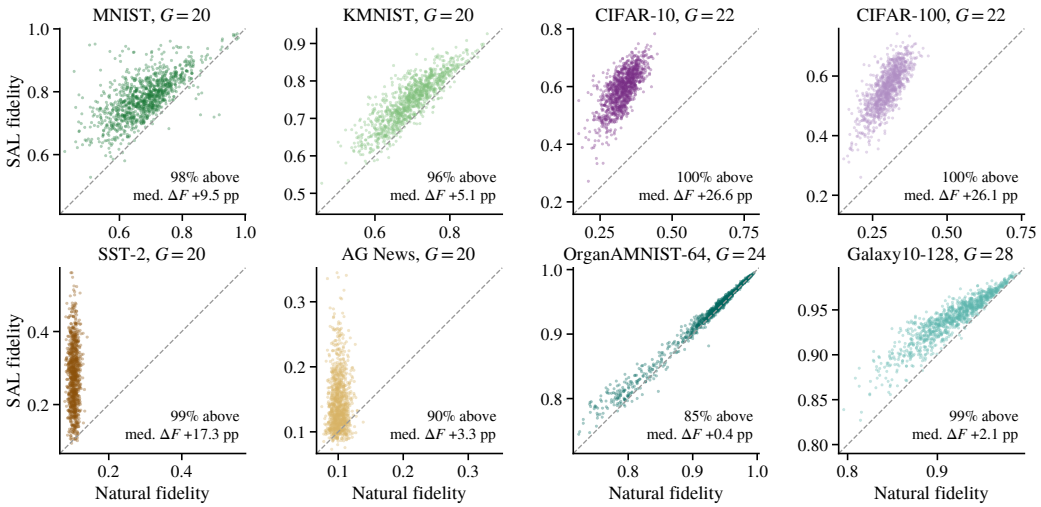


Figure 12: Per-target fidelity at the reference budget (run 0). Each point is one external target; each panel gives the fraction of targets above the diagonal and the median paired gain.

The available improvement depends partly on how well the natural address already admits a product approximation. Targets with high natural product-state fidelity tend to show smaller gains (Figure 13a). For nonnegative images, the uniform-state fidelity $(\sum_j x_j)^2/d$ is invariant to permutation and provides a lower bound on the best product-state fidelity. Its median is 0.17 on MNIST and 0.25 on KMNIST, compared with 0.83 on OrganAMNIST-64 and 0.73 on Galaxy10-128. Both addresses on OrganAMNIST-64 lie close to this bound (panel b), whereas Galaxy10-128 leaves more room for improvement at small budgets.

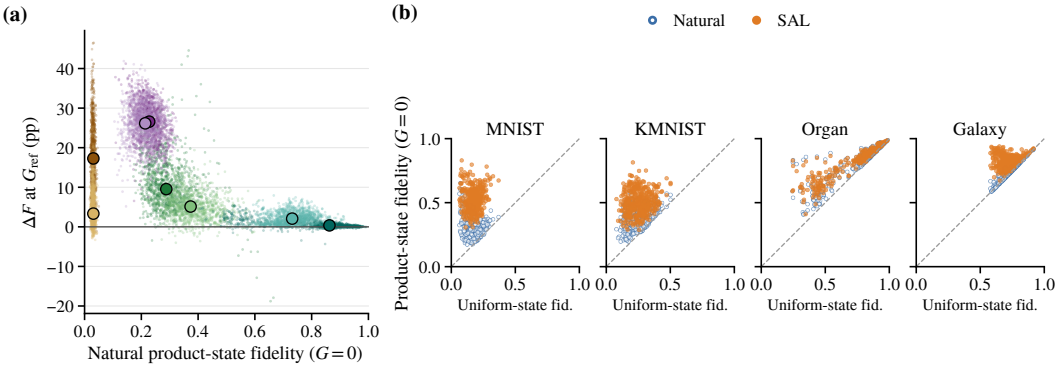


Figure 13: Natural-address preparation and fidelity gains. (a) Paired gain at the reference budget versus the natural product-state fidelity of the same target; colors as in Figure 12, large markers are dataset medians. (b) For nonnegative images, product-state fidelity under both addresses versus the uniform-state fidelity $(\sum_j x_j)^2/d$, which is the same for every address; images with nearly uniform backgrounds are already well approximated at $G = 0$.

Figures 14–17 show representative targets from each pixel-image dataset. We take the run-0 targets at the 10th, 30th, 50th, 70th and 90th percentiles of the paired fidelity gain at G_{ref} and display their prepared states at $G = 0$, G_{low} and G_{ref}. SAL is ahead at G_{ref} in all five rows for MNIST, KMNIST and Galaxy10-128, and in four of five for OrganAMNIST-64. Feature representations are covered by the per-target comparisons in Figure 12. To display SAL states, we invert the gather, align the global sign with the target and show negative amplitudes as zero. Fidelities use the full normalized states before these display transformations.

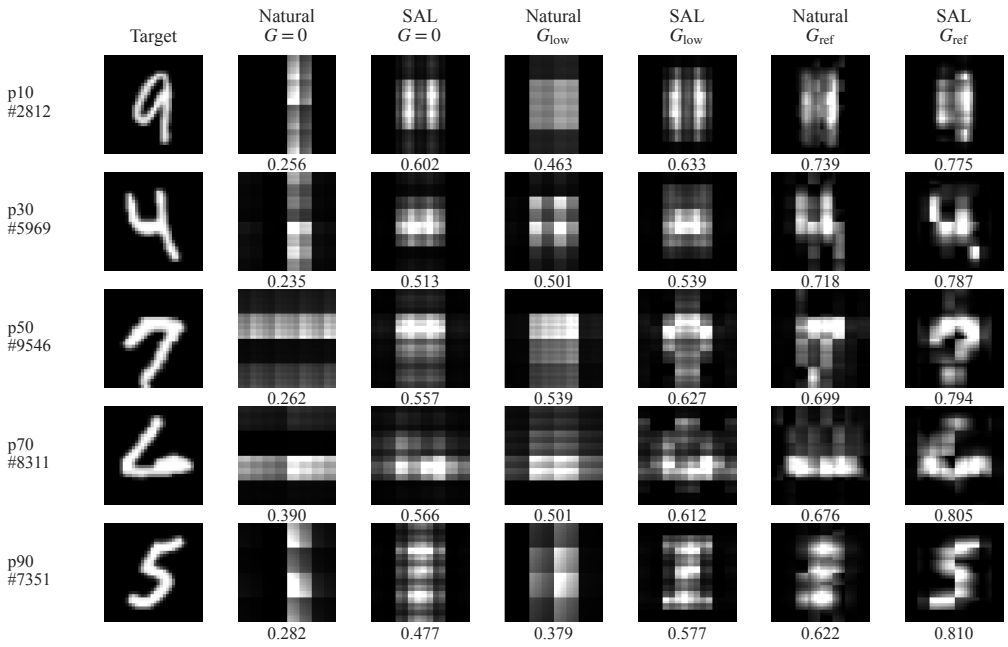


Figure 14: MNIST: external targets at the 10th, 30th, 50th, 70th and 90th percentiles of the paired fidelity gain at G_{ref} (run 0), with the states prepared at $G = 0$, G_{low} and G_{ref} under both addresses; numbers give fidelity. Pixel order; negative amplitudes are displayed as zero.

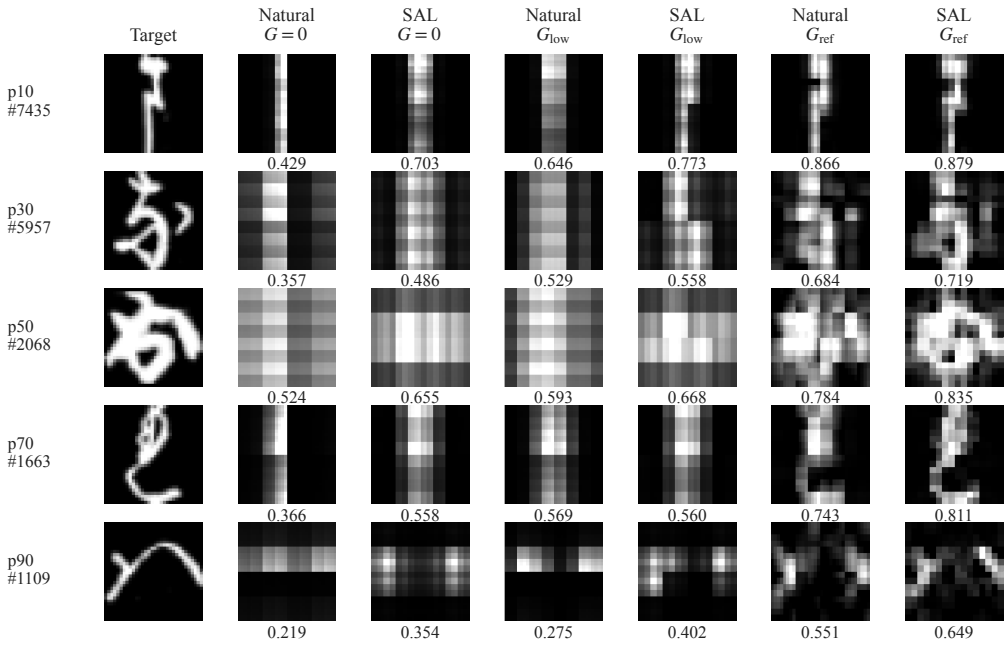


Figure 15: KMNIST, as in Figure 14.

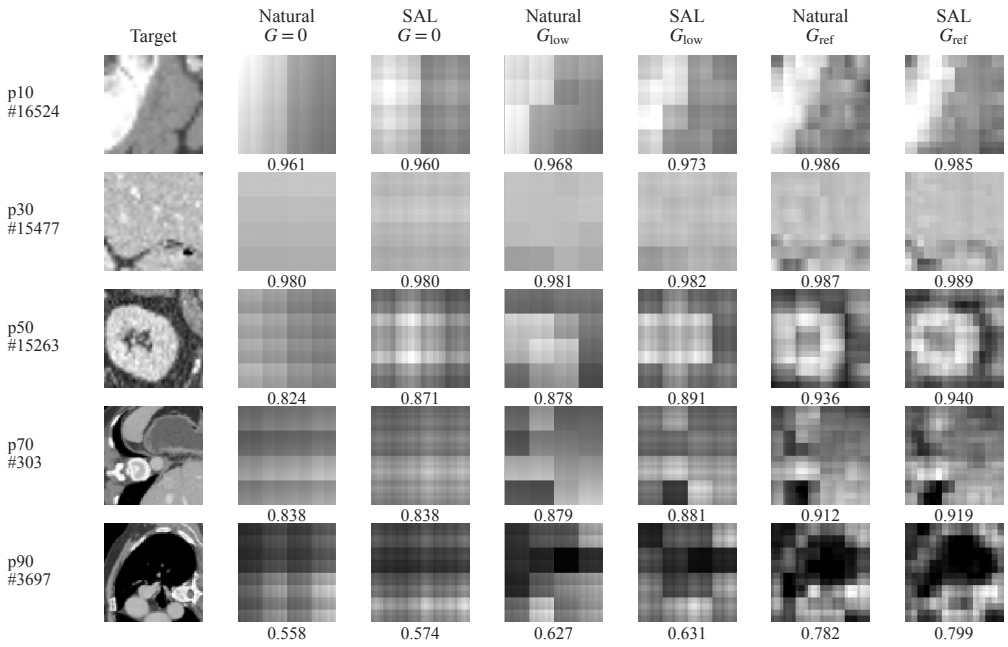


Figure 16: OrganAMNIST-64, as in Figure 14.

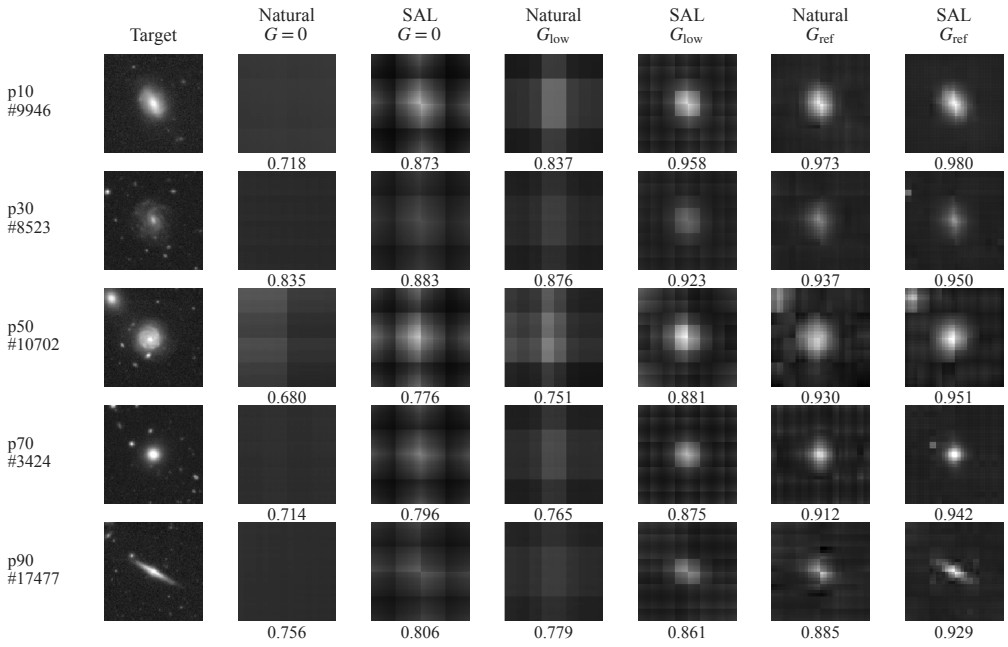


Figure 17: Galaxy10-128, as in Figure 14.

E Downstream Models

E.1 Kernel Classification

Kernel classification uses all native classes. Since the ideal amplitude kernel is shared by both addresses, its classifier is fitted once per outer training split. We fit centered one-hot labels by multiclass ridge regression and choose a single ridge parameter for all outputs on the internal validation set, using the grid in Table 7. Prediction restores the training label mean and takes the class with the largest score. Validation and external examples use the training centering: a test row k becomes $k_c = k - \bar{k}\mathbf{1}^\top - \bar{K}_{\mathrm{col}} + \bar{K}_{\mathrm{all}}\mathbf{1}^\top$.

Tables 11 and 12, together with Figure 18, report all four kernel budgets. We compare accuracy at matched CX counts, consistent with selecting addresses by bounded-circuit fidelity. Eq. (4) bounds the kernel perturbation, while Appendix G.6 examines the pairwise relations it changes. On Galaxy10-128, both approximate-kernel classifiers exceed the ideal-kernel accuracy at G_{ref}, with SAL ahead of the natural address. Approximate loading can therefore improve classification even though it perturbs the ideal kernel.

Table 11: KRR test accuracy (%), mean ± s.d. over five runs.

Dataset	G	Ideal	Natural	SAL
MNIST	0	96.72±0.54	70.00±0.61	**80.88**±1.05
	4	96.72±0.54	75.82±0.40	**86.46**±0.76
	10	96.72±0.54	88.22±1.67	**92.78**±0.19
	20	96.72±0.54	94.04±0.83	**94.90**±0.39
KMNIST	0	82.96±0.67	45.52±0.50	**52.80**±0.51
	4	82.96±0.67	51.14±0.48	**61.04**±1.15
	10	82.96±0.67	65.88±2.39	**73.42**±1.22
	20	82.96±0.67	75.04±1.64	**78.18**±1.04
CIFAR-10	0	89.92±0.40	43.72±0.26	**54.50**±3.66
	6	89.92±0.40	53.36±0.66	**63.04**±2.02
	10	89.92±0.40	60.10±0.98	**65.40**±3.23
	22	89.92±0.40	68.08±0.79	**72.90**±2.09
CIFAR-100	0	67.10±0.60	14.36±0.65	**17.52**±0.73
	6	67.10±0.60	23.30±0.70	**26.72**±0.93
	10	67.10±0.60	28.06±0.68	**31.58**±0.87
	22	67.10±0.60	33.68±1.08	**40.18**±0.87
SST-2	0	90.09±0.92	64.52±1.54	**64.84**±1.47
	4	90.09±0.92	64.89±0.89	**67.20**±1.64
	10	90.09±0.92	70.23±1.05	**72.48**±1.89
	20	90.09±0.92	76.70±1.74	**79.56**±1.25
AG News	0	89.64±0.24	52.54±0.46	**56.32**±0.63
	4	89.64±0.24	53.96±1.52	**61.12**±0.68
	10	89.64±0.24	63.10±2.79	**69.86**±1.04
	20	89.64±0.24	75.46±0.92	**76.90**±0.64
Organ-64	0	65.40±1.40	47.92±0.57	**49.56**±0.74
	6	65.40±1.40	55.60±0.76	**57.56**±0.80
	12	65.40±1.40	59.32±1.01	**61.32**±0.54
	24	65.40±1.40	63.84±0.32	**64.16**±0.94
Galaxy-128	0	39.88±0.74	19.76±0.67	**33.98**±0.62
	6	39.88±0.74	27.14±0.91	**37.14**±1.17
	14	39.88±0.74	36.40±0.65	**42.38**±0.48
	28	39.88±0.74	40.96±0.96	**42.66**±1.03

Table 12: KRR macro-F1 (%) for the same predictions.

Dataset	G	Ideal	Natural	SAL
MNIST	0	96.68±0.54	69.30±0.59	**80.26**±1.19
	4	96.68±0.54	75.46±0.42	**86.15**±0.77
	10	96.68±0.54	88.00±1.66	**92.65**±0.20
	20	96.68±0.54	93.96±0.86	**94.83**±0.42
KMNIST	0	82.93±0.67	45.38±0.65	**52.83**±0.64
	4	82.93±0.67	51.20±0.58	**61.11**±1.03
	10	82.93±0.67	65.85±2.42	**73.36**±1.25
	20	82.93±0.67	75.02±1.61	**78.11**±1.01
CIFAR-10	0	89.89±0.41	42.89±0.22	**54.31**±3.69
	6	89.89±0.41	53.22±0.62	**62.76**±2.17
	10	89.89±0.41	59.91±1.02	**65.07**±3.45
	22	89.89±0.41	67.89±0.78	**72.80**±2.13
CIFAR-100	0	65.99±0.89	10.50±0.44	**14.25**±0.94
	6	65.99±0.89	19.65±0.67	**23.56**±1.09
	10	65.99±0.89	24.94±0.76	**28.16**±0.63
	22	65.99±0.89	31.18±1.22	**37.26**±0.68
SST-2	0	90.08±0.92	63.32±1.79	**63.42**±1.89
	4	90.08±0.92	64.21±1.11	**66.17**±2.23
	10	90.08±0.92	69.77±1.05	**72.11**±2.03
	20	90.08±0.92	76.43±1.81	**79.46**±1.24
AG News	0	89.65±0.24	52.41±0.47	**56.19**±0.85
	4	89.65±0.24	53.84±1.43	**61.09**±0.70
	10	89.65±0.24	63.06±2.74	**69.91**±1.22
	20	89.65±0.24	75.35±0.91	**77.01**±0.60
Organ-64	0	62.66±1.83	38.49±0.64	**41.69**±1.32
	6	62.66±1.83	47.99±0.69	**50.74**±1.17
	12	62.66±1.83	54.94±1.02	**57.33**±0.48
	24	62.66±1.83	60.14±0.75	**60.93**±0.51
Galaxy-128	0	34.68±1.17	10.27±0.27	**28.56**±0.39
	6	34.68±1.17	22.19±0.97	**30.91**±1.15
	14	34.68±1.17	30.79±0.67	**35.59**±0.65
	28	34.68±1.17	34.73±0.77	**36.69**±1.08

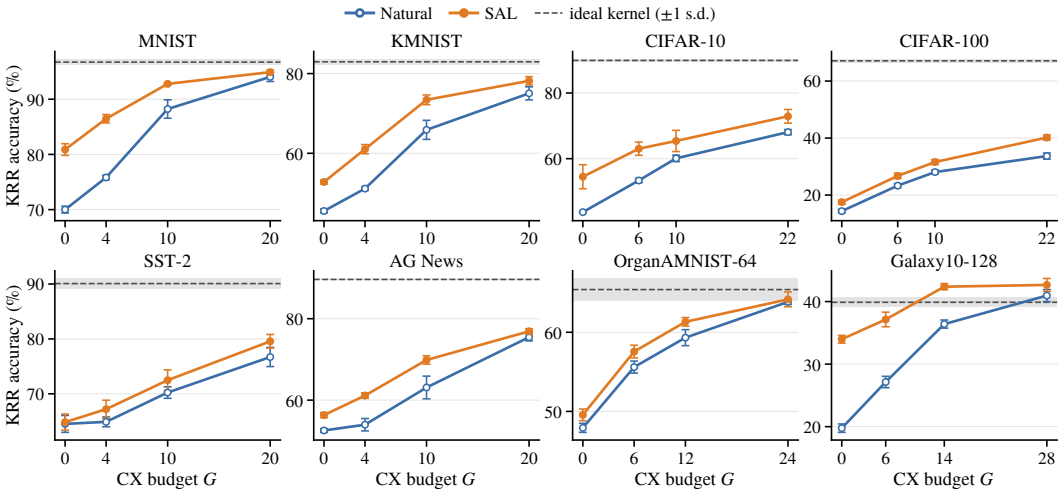


Figure 18: KRR test accuracy versus loading budget. Grey band: ideal kernel ± 1 s.d.; error bars: ± 1 s.d. over five runs.

E.2 Quantum Neural Network

MNIST, CIFAR-10 features and SST-2 use the same QNN template (Figure 19). In each layer $\ell = 1, 2, 3$, every qubit receives R_Y followed by R_Z, then CX gates act on $(0, 1), (1, 2), \ldots, (n-2, n-1), (n-1, 0)$ in that order. Writing the full unitary as U_θ, the prediction for an input state z is

$$h_q(z;\theta) = \langle z|U_\theta^\dagger Z_q U_\theta|z\rangle, \qquad f(z) = Wh(z;\theta) + b, \qquad \hat{y} = \arg\max_c f_c(z).$$

The readout circuit has $3n$ CX gates, counted separately from the preparation budget G, together with $6n$ rotation angles and $nC + C$ head parameters: 170 trainable parameters for MNIST, 186 for CIFAR-10 and 82 for SST-2.

Training. We train by multiclass cross-entropy with Adam, using a learning rate of 0.01, batches of 64 and at most 50 epochs. Validation cross-entropy selects the saved model, with early-stopping patience of ten epochs. Circuit angles are initialized from a zero-mean normal distribution with standard deviation 0.1, head weights with standard deviation $1/\sqrt{n}$, and biases at zero. Ideal natural, ideal SAL, approximate natural and approximate SAL inputs are trained independently over five runs. Each ideal model is trained once per address and run and reused for both displayed budgets, since ideal inputs are independent of the loading budget.

These two ideal conditions account for the interaction between address and readout. The ideal kernel depends only on inner products, whereas the QNN evaluates $\langle x|P^\dagger U_\theta^\dagger Z_q U_\theta P|x\rangle$. Its model family is preserved when the variational observables are closed under conjugation by P (Eq. (1)). Comparing each approximate condition with its corresponding ideal condition separates this address effect from the approximation penalty.

Approximation penalty. Write $A_{\mathrm{id,N}}$, $A_{\mathrm{id,S}}$, $A_{\mathrm{ap,N}}$ and $A_{\mathrm{ap,S}}$ for the four accuracies. The approximate difference decomposes exactly as

$$A_{\mathrm{ap,S}} - A_{\mathrm{ap,N}} = \big(A_{\mathrm{id,S}} - A_{\mathrm{id,N}}\big) + \big[(A_{\mathrm{id,N}} - A_{\mathrm{ap,N}}) - (A_{\mathrm{id,S}} - A_{\mathrm{ap,S}})\big].$$

The first term is the ideal-address accuracy difference; the second is the reduction in approximation penalty. On MNIST at the shallow budget, ideal SAL is 0.88 points below ideal natural, while approximate SAL is 8.34 points above approximate natural, giving a penalty reduction of about 9.22 points. On SST-2 at $G = 20$, the approximate gain is 19.49 points and the ideal-address difference is 1.88 points, leaving about 17.61 points in the second term. CIFAR-10 behaves differently at $G = 22$: its 5.26-point approximate gain is close to the 5.24-point ideal-address difference, while the shallow-budget gain also includes a substantial reduction in approximation penalty. SAL improves every approximate condition examined, with the source of the gain depending on the dataset and budget.

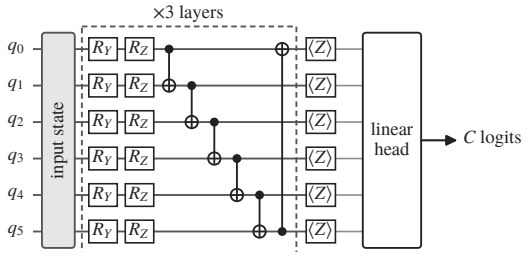


Figure 19: QNN template, shown for $n = 6$: three layers of R_Y and R_Z rotations on every qubit followed by a ring of CX gates, the Z expectation of every qubit, and a linear head that returns C logits.

Table 13: QNN test accuracy and macro-F1 under the four input conditions (mean ± s.d. over five training runs). Ideal inputs are independent of G; their entries repeat across the two budgets.

		Ideal input		Approximate input	
Dataset	G	Natural	SAL	Natural	SAL
Accuracy (%)					
MNIST	4	77.50±0.93	76.62±1.24	64.84±1.78	**73.18**±0.99
MNIST	20	77.50±0.93	76.62±1.24	74.12±2.86	**76.08**±2.20
CIFAR-10	6	54.90±2.08	60.14±0.55	32.92±1.03	**48.14**±2.92
CIFAR-10	22	54.90±2.08	60.14±0.55	42.74±1.42	**48.00**±4.33
SST-2	4	79.50±2.39	81.38±1.97	53.53±1.33	**66.10**±3.03
SST-2	20	79.50±2.39	81.38±1.97	52.96±2.82	**72.45**±1.15
Macro-F1 (%)					
MNIST	4	76.84±0.89	75.92±1.33	63.29±1.96	**71.72**±1.37
MNIST	20	76.84±0.89	75.92±1.33	73.11±3.13	**75.09**±2.50
CIFAR-10	6	54.59±2.19	59.89±0.63	31.68±1.18	**47.62**±3.22
CIFAR-10	22	54.59±2.19	59.89±0.63	41.94±1.40	**47.36**±4.61
SST-2	4	79.45±2.42	81.30±2.00	47.68±3.07	**64.76**±3.92
SST-2	20	79.45±2.42	81.30±2.00	45.34±8.26	**72.11**±1.16

E.3 Eight-Qubit Hardware Experiment

Data and training. MNIST digits 0 to 3 are resized bilinearly to 16×16 and ℓ_2-normalized, which gives 256 amplitudes on eight qubits. Training, internal validation, development and test sets contain 1600, 400, 160 and 160 images with equal class counts; the test images come from the official test partition. The preparation budget $G = 4$ and the pair-vote readout were chosen on the development set. SAL uses $|D_{\mathrm{C}}| = 256$, $|D_{\mathrm{F}}| = 128$, $|D_{\mathrm{A}}| = 64$, checkpoints at depths 0, 4 and 8, selection at $G_{\mathrm{ref}} = 16$ and 64 fitting sweeps. The readout has two layers of R_Y and R_Z rotations with nearest-neighbor entangling gates, followed by a final R_Y on each qubit, for 40 trainable parameters (Figure 20). Each address condition is trained over three runs by categorical cross-entropy with batches of 64, a learning rate decaying from 0.02 to 0.002, gradient clipping at 5 and at most 200 epochs; internal-validation loss selects the checkpoint.

Pair-vote readout. For an eight-bit outcome $z = (b_7 \dots b_0)$, pair (b_{2k}, b_{2k+1}) encodes the integer $b_{2k} + 2b_{2k+1}$. With $N_x(z)$ the count of z among $S = 1024$ shots,

$$\hat{p}(c \mid x) = \frac{1}{4S} \sum_{k=0}^{3} \sum_{z} N_x(z)\, \mathbf{1}\{b_{2k} + 2b_{2k+1} = c\}, \qquad c \in \{0, 1, 2, 3\}. \tag{15}$$

The four pairs contribute equally, and prediction takes the largest probability, resolving ties toward the smallest class index.

Hardware execution took place on 23 September 2026. Logical qubits $0, \dots, 7$ map to the physical chain $154, 153, \dots, 147$. Each circuit uses 1024 shots, and class probabilities are computed directly from raw counts. The software environment is Qiskit 2.5.2 with Qiskit IBM Runtime 0.49.0.

Preparation uses two $\mathrm{SO}(4)$ blocks, or four logical CX gates, under either address. In the representative circuits of Figure 20, the natural-address blocks act on logical pairs $(2, 3)$ and $(6, 7)$, while the SAL blocks act on $(4, 6)$ and $(4, 5)$.

Table 14 reports each run, and Figure 21 shows the pooled confusion matrices. Under the natural address, 79.2% of digit-0 images are assigned to class 3 and only 15.8% are correctly recognized.

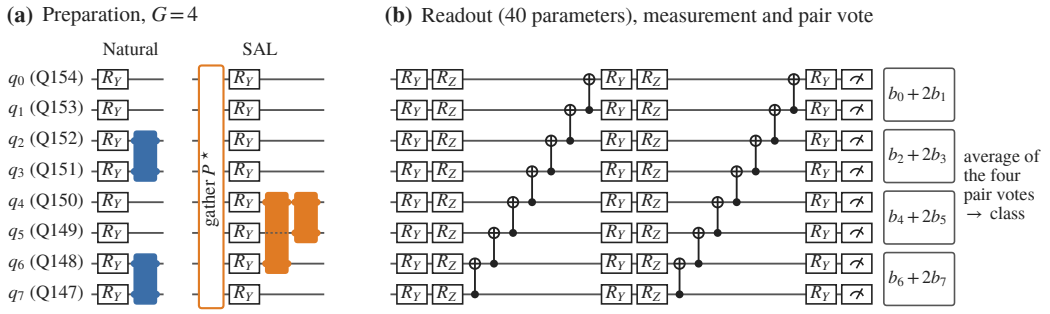


Figure 20: Representative eight-qubit hardware circuit. (a) Preparation at $G = 4$ under both addresses; wires give logical qubits and, in parentheses, physical qubits of `ibm_kingston`. SAL gathers the 256 amplitudes classically before preparation. (b) Trained readout, measurement and the fixed pair vote of Eq. (15).

Table 14: Hardware accuracy and macro-F1 (%) on 160 test images per run, computed from raw counts.

	Accuracy		Macro-F1	
Run	Natural	SAL	Natural	SAL
0	60.00	73.12	58.17	72.44
1	55.00	69.38	53.86	64.87
2	58.75	70.62	56.79	69.88
Mean ± s.d.	57.92±2.60	71.04±1.91	56.28±2.20	69.06±3.85

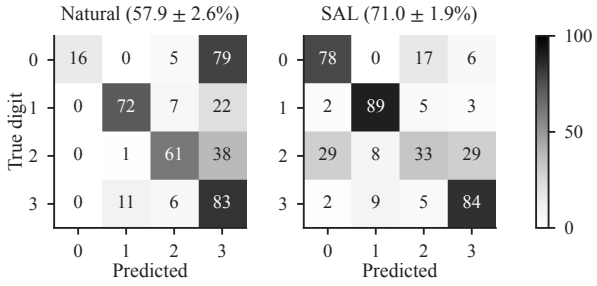


Figure 21: Eight-qubit hardware: confusion matrices of the test predictions pooled over three runs (row-normalized, %); titles give the mean ± s.d. accuracy over runs.

SAL recognizes 77.5% of zeros, 89.2% of ones and 84.2% of threes. Its weakest class is digit 2, with 33.3% accuracy and frequent confusion with 0 and 3.

F Design Ablations

We examine the information used by address construction and compare shared with instance-wise addresses on MNIST, CIFAR-10 and SST-2, covering pixels, visual features and signed sentence features. All experiments use five runs and the main shallow and reference budgets, with the same construction procedure, loader and input dimensions.

F.1 Shared-Coordinate Information

For each construction row x_i we sample one independent uniform coordinate permutation R_i. A fixed random ordering of the rows gives nested retained sets I_ρ with fractions $\rho \in \{0, 0.25, 0.5, 1\}$, and candidate generation receives

$$\tilde{x}_i^{(\rho)} = \begin{cases} x_i, & i \in I_\rho, \\ R_i x_i, & i \notin I_\rho. \end{cases}$$

Each row retains its signed amplitude multiset, dimension and norm. The permutation changes coordinate-specific statistics only for candidate construction: both bases, both roots and all checkpoints are generated from these modified rows. Topology fitting uses the original construction vectors,

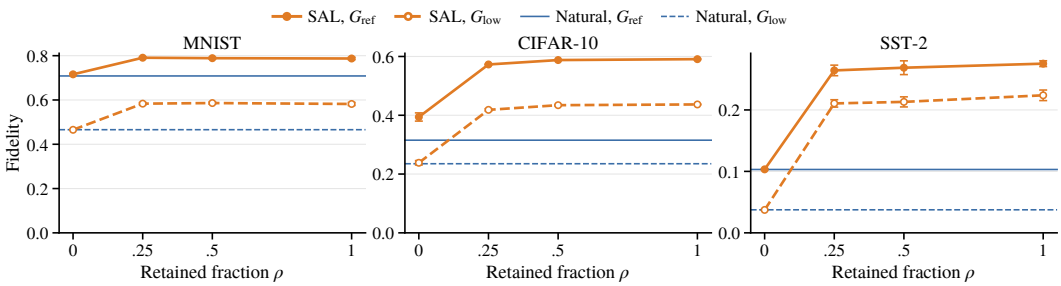


Figure 22: Loading fidelity as the fraction ρ of construction rows with aligned coordinates grows (solid: G_{ref}; dashed: G_{low}; horizontal lines: natural address; mean $\pm$ s.d. over five runs).

screening uses the original D_{F}, and external targets keep their original amplitudes and coordinate order.

Removing all coordinate alignment ($\rho = 0$) nearly eliminates the reference-budget gain on MNIST, eliminates it on SST-2 and reduces it by 71% on CIFAR-10 (Figure 22, Table 15). CIFAR-10 and SST-2 recover progressively as aligned rows are restored; one quarter of aligned rows recovers 91% and 94% of the gain lost at $\rho = 0$. MNIST reaches a plateau at $\rho = 0.25$, with small differences thereafter.

Table 15: Shared-coordinate ablation (mean $\pm$ s.d., five runs). Shallow budgets $G = 4, 6, 4$ and reference budgets $G = 20, 22, 20$.

Dataset	ρ	G_{low}	G_{ref}
MNIST	0	0.4653±.0011	0.7158±.0045
	0.25	0.5834±.0049	0.7908±.0035
	0.5	0.5861±.0047	0.7888±.0036
	1	0.5820±.0067	0.7873±.0053
CIFAR-10	0	0.2383±.0075	0.3941±.0140
	0.25	0.4184±.0038	0.5728±.0011
	0.5	0.4343±.0021	0.5878±.0012
	1	0.4368±.0015	0.5909±.0018
SST-2	0	0.0373±.0006	0.1033±.0008
	0.25	0.2106±.0060	0.2643±.0086
	0.5	0.2131±.0082	0.2687±.0112
	1	0.2238±.0086	0.2753±.0047

Table 16: Construction-set size at G_{ref}; totals include the 128 screening examples.

Dataset	$\lvert D_{\text{C}}\rvert$	Total	Natural	SAL
CIFAR-10	32	160	0.3183±.0077	0.5807±.0035
	64	192	0.3152±.0065	0.5864±.0008
	128	256	0.3152±.0065	0.5896±.0007
	256	384	0.3152±.0065	0.5909±.0018
	512	640	0.3152±.0065	0.5917±.0009
SST-2	32	160	0.1023±.0005	0.2464±.0161
	64	192	0.1025±.0011	0.2628±.0127
	128	256	0.1025±.0011	0.2696±.0091
	256	384	0.1025±.0011	0.2752±.0047
	512	640	0.1025±.0011	0.2725±.0010

F.2 Signed-Amplitude Information

For the signed SST-2 features, the magnitude-only condition constructs candidates from the entrywise absolute values of D_{C}. Topology fitting, screening and evaluation retain the original signed vectors. Compared with magnitude-only construction, signed construction raises fidelity by 17.44 points at $G = 4$ and 13.73 points at $G = 20$ (Table 2, Figure 23b). This agrees with the role of relative signs in $c_y(v) = \sum_j y_j y_{j \oplus v}$: SAL takes the absolute value after computing each example's correlation, whereas replacing y by $|y|$ changes the correlation itself. For nonnegative pixel and ResNet inputs, $|x| = x$.

F.3 Construction Data

We vary construction-set size on CIFAR-10 and SST-2 while keeping the encoder, classification splits and external targets fixed. Within each run, construction prefixes of size 32, 64, 128, 256 and 512 are nested and disjoint from a fixed 128-example screening set. The budget remains $G_{\text{ref}} = 2n$. With 32 construction examples, the mean gain reaches 95.2% of the 256-example gain on CIFAR-10 and 83.4% on SST-2 (Table 16, Figure 23a). The topology subset contains at most 64 examples, while the reference and correlation statistics use the full construction set. Natural-address fits are repeated for each condition.

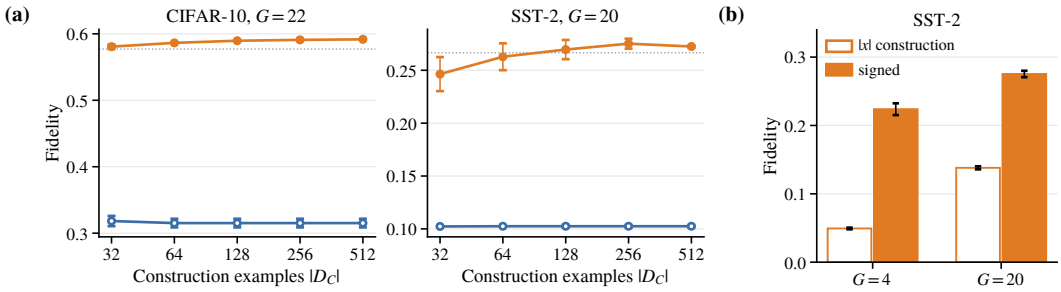


Figure 23: (a) Construction-set size with a fixed 128-example screening set; the dotted line marks 95% of the gain at $|D_C| = 256$. (b) Signed versus magnitude-only construction on SST-2.

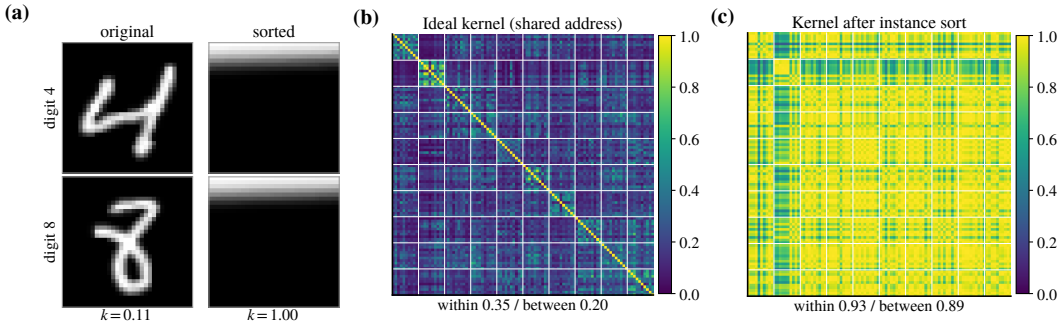


Figure 24: Instance-wise sorting destroys the kernel geometry. (a) Two MNIST test digits of different classes whose sorted states nearly coincide, with their amplitude kernel k before and after sorting. (b, c) Amplitude kernels of 100 test digits ordered by class under a shared address and after instance-wise sorting; numbers give the mean within-class and between-class kernel.

F.4 Shared versus Instance-Wise Addresses

Instance-wise sorting orders each example by sign group and decreasing magnitude, placing nonnegative amplitudes first and resolving ties by the original index. Training, validation and test examples are each sorted independently. We fit separate KRR classifiers to these representations, selecting the ridge parameter on the sorted internal-validation set from the original grid. For approximate loading, the topology is learned from sorted topology examples and circuits use the original fitting procedure. Fidelity is measured against the target of each condition.

A shared permutation preserves all pairwise inner products. Instance-wise sorting instead maps inputs with the same signed amplitude multiset to the same representation. Figure 24 illustrates the resulting loss on MNIST: two digits from different classes have amplitude kernel 0.11 before sorting and 1.00 after sorting, as their sorted states nearly coincide. Across 100 test digits, the mean within-class and between-class kernels change from 0.35 and 0.20 to 0.93 and 0.89. This loss is present in the ideal representation, before approximate loading (Table 17). At the reference budget, sorted MNIST, CIFAR-10 and SST-2 targets load with fidelities 0.9998, 0.9992 and 0.9924, yet their classifiers remain close to the degraded ideal classifiers. SAL preserves the original ideal kernel and reaches approximate accuracies of 94.90%, 72.90% and 79.56%, respectively (Table 18). Easy preparation alone therefore does not ensure that the representation retains the relations needed for classification.

Table 17: Ideal-kernel prediction and geometry under shared and instance-wise addresses (mean ± s.d., five runs). The relative kernel error $\|K_s - K\|_F / \|K\|_F$ uses 512 external examples; every shared address reproduces the ideal kernel exactly.

Dataset	Address	Accuracy (%)	Macro-F1 (%)	Kernel error
MNIST	Natural	96.72±0.54	96.68±0.54	0.000
	Instance sort	34.80±1.04	28.90±1.21	2.484
	SAL	96.72±0.54	96.68±0.54	0.000
CIFAR-10	Natural	89.92±0.40	89.89±0.41	0.000
	Instance sort	27.92±0.73	26.42±0.99	2.131
	SAL	89.92±0.40	89.89±0.41	0.000
SST-2	Natural	90.09±0.92	90.08±0.92	0.000
	Instance sort	53.58±1.36	48.74±2.82	7.322
	SAL	90.09±0.92	90.08±0.92	0.000

Table 18: Loading and prediction under shared and instance-wise addresses. Fidelity refers to each condition's own target; accuracy and macro-F1 in percent (mean ± s.d., five runs).

Dataset	G	Address	Fidelity	Accuracy	Macro-F1
MNIST	4	Natural	0.4658±.0000	75.82±0.40	75.46±0.42
		Instance sort	0.9958±.0000	34.14±1.23	27.77±1.42
		SAL	0.5820±.0067	86.46±0.76	86.15±0.77
MNIST	20	Natural	0.7085±.0179	94.04±0.83	93.96±0.86
		Instance sort	0.9998±.0000	35.12±0.50	29.10±0.78
		SAL	0.7873±.0053	94.90±0.39	94.83±0.42
CIFAR-10	6	Natural	0.2351±.0000	53.36±0.66	53.22±0.62
		Instance sort	0.9207±.0000	28.66±0.82	26.87±1.13
		SAL	0.4368±.0015	63.04±2.02	62.76±2.17
CIFAR-10	22	Natural	0.3152±.0065	68.08±0.79	67.89±0.78
		Instance sort	0.9992±.0002	27.38±0.74	25.49±1.14
		SAL	0.5909±.0018	72.90±2.09	72.80±2.13
SST-2	4	Natural	0.0375±.0002	64.89±0.89	64.21±1.11
		Instance sort	0.9494±.0003	51.12±0.27	34.82±1.47
		SAL	0.2238±.0086	67.20±1.64	66.17±2.23
SST-2	20	Natural	0.1031±.0013	76.70±1.74	76.43±1.81
		Instance sort	0.9924±.0067	51.42±0.78	38.87±6.56
		SAL	0.2753±.0047	79.56±1.25	79.46±1.24

G Structural and Relational Analysis

G.1 Amplitude Structure

For a vector $v \in \mathbb{R}^{2^n}$, we display the coefficient matrix of the contiguous balanced cut, with rows indexed by the high $\lceil n/2 \rceil$ address bits and columns by the low $\lfloor n/2 \rfloor$ bits. Under row-major flattening and the least-significant-bit convention for qubit 0, the natural-address matrix of a pixel image is the image itself. The singular values $s_1 \geq s_2 \geq \cdots$ give the cumulative Schmidt mass $q_\chi = \sum_{k \leq \chi} s_k^2$, which bounds the fidelity attainable by a rank-χ state across this cut. A circuit with g crossing CX gates therefore has fidelity at most q_{2^g}.

Figure 25 shows the run-0 target with the median paired gain at G_{ref} in each dataset, together with Schmidt masses averaged over 128 external targets and then over five runs. The addresses rearrange the same entries into different coefficient matrices. On this contiguous cut, mean rank-two mass ($\chi = 2$) rises from 0.42 to 0.64 on CIFAR-10, from 0.41 to 0.61 on CIFAR-100 and from 0.21 to 0.36 on SST-2; the two addresses remain within 0.02 on MNIST and KMNIST. The bottom panels show the tail mass $1 - q_\chi$ on logarithmic axes, with lower curves indicating a more accurate rank-χ approximation. This cut is used for display; Appendix G.2 averages four fixed balanced cuts on the six datasets with $n \in \{10, 11\}$.

Figure 26 shows the dominant shared direction u_1 under the three orders. On MNIST, it resembles an average digit and already has rank-two mass 0.99 in the natural order. On CIFAR-10 and SST-2, the principal-reference order turns a noisy profile into a smooth one with rank-two mass close to one.

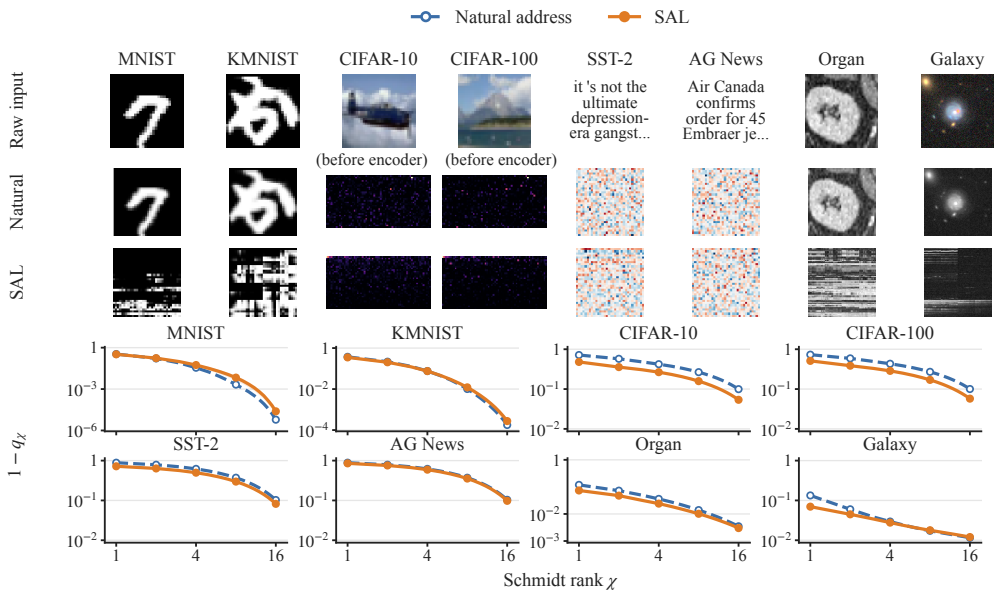


Figure 25: Amplitude structure under the two addresses. Top: one external target per dataset, the run-0 target with the median paired fidelity gain at G_{ref}; CIFAR shows the image before the frozen encoder. Middle: its amplitudes as the coefficient matrix of the contiguous balanced cut under both addresses (identical entries and color scale within a column). Bottom: Schmidt tail mass $1 - q_\chi$ across that cut on logarithmic axes, averaged over 128 external targets and then over five runs; bands show ± 1 run s.d. Lower values indicate more mass captured at rank χ.

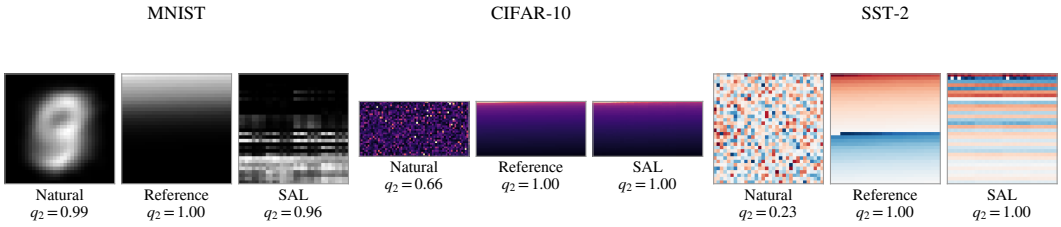


Figure 26: Dominant second-moment direction u_1 under the natural order, the principal-reference order and the selected SAL address (run 0), with its rank-two Schmidt mass across the contiguous balanced cut.

G.2 Descriptor Subspaces and Fixed Cuts

The leading direction may be easy to prepare even when other high-energy directions remain difficult. We therefore examine a shared subspace through the projector $\Pi_r = \sum_{j \le r} u_j u_j^\dagger$ and the projected target $x_c = \Pi_r x$. The projector is invariant to basis rotations within the selected subspace, avoiding dependence on the orientation of nearly degenerate modes within it. We estimate this subspace from 512 training observations disjoint from address construction and screening, using $r = 4$ as the primary rank and $r = 1$ and $r = 16$ for sensitivity analysis. Each projected state is normalized before measuring its Schmidt spectrum; its original mass $\|x_c\|^2$ is recorded separately to distinguish shared energy from the complexity of its tensor arrangement.

Schmidt mass across fixed cuts. For a bipartition $A|B$, reshape a normalized state into $X_{A|B}$ with singular values $s_1 \ge s_2 \ge \cdots$. The best normalized rank-χ approximation has fidelity $q_\chi(X_{A|B}) = \sum_{j \le \chi} s_j^2$, by the trace inequality used in Appendix A.3. A circuit with g crossing CX gates has Schmidt rank at most 2^g, giving the fidelity bound q_{2^g}. Since a circuit must realize all cuts simultaneously, we use balanced-cut rank-two mass as a structural diagnostic. Table 19 fixes the four cuts; each run averages over them, and we report the mean over five runs. The 128 structural targets are a subset of the 512 relational targets.

Table 19: Left-qubit sets of the four balanced cuts; the right side is the complement.

n	A_1	A_2	A_3	A_4
10	$\{0,1,2,3,4\}$	$\{0,2,4,6,8\}$	$\{0,6,7,8,9\}$	$\{0,1,5,7,9\}$
11	$\{0,1,2,3,4\}$	$\{0,2,4,6,8\}$	$\{1,2,3,4,5\}$	$\{0,1,2,3,9\}$

G.3 Shared-Structure Measurements

We estimate the second-moment spectrum from the 512 descriptor observations and define its effective rank as $\exp[-\sum_j \lambda_j \log \lambda_j]$ for unit-trace M. Tables 20 and 21, together with Figure 27a, report rank-two mass for the common component and complete target. On CIFAR-10 and CIFAR-100, substantial shared energy accompanies a large change in tensor arrangement: the first-mode energies are 0.545 and 0.505, and common-component rank-two mass increases by about 0.31.

MNIST illustrates the value of examining several shared directions. The principal-reference order makes the first direction nearly rank two (Figure 26), but its common-component rank-two mass remains below that of the natural order. SAL selects the natural base and improves both the common projection and the complete target. The reference base is more effective on frozen features, consistent with the screening results in Figure 5. Common-component rank-two mass improves in all twenty run–cut comparisons for CIFAR-10, CIFAR-100, SST-2 and AG News; on MNIST, the improvement holds after averaging over cuts.

Table 20: Shared energy and common-component rank-two quality. N, R and S denote the natural order, the principal-reference order and the selected SAL address. The spectrum is the same under every address.

Dataset	λ_1	$\sum_{j\le 4}\lambda_j$	r_{eff}	$q_{2,\mathrm{N}}$	$q_{2,\mathrm{R}}$	$q_{2,\mathrm{S}}$
MNIST	0.452	0.597	16.84	0.8655	0.8533	0.9295
KMNIST	0.353	0.504	31.08	0.9493	0.8261	0.9632
CIFAR-10	0.545	0.632	18.39	0.5977	0.9038	0.9051
CIFAR-100	0.505	0.584	23.96	0.5902	0.9000	0.9001
SST-2	0.198	0.287	100.25	0.2206	0.7338	0.7346
AG News	0.080	0.167	161.65	0.2102	0.5591	0.5597

Table 21: Rank-two quality of the complete targets and the reference-budget fidelity gain.

Dataset	$q_{2,\mathrm{N}}$	$q_{2,\mathrm{S}}$	ΔF at G_{ref} (pp)
MNIST	0.7316	0.8116	7.881
KMNIST	0.7239	0.7732	6.100
CIFAR-10	0.4245	0.6342	27.569
CIFAR-100	0.4089	0.6040	25.959
SST-2	0.2140	0.3567	17.218
AG News	0.2128	0.2457	4.269

G.4 Loading-Error Decomposition

Let x and z be normalized pure states in the same address and $D = xx^\dagger - zz^\dagger$. Since $\mathrm{tr}[(xx^\dagger)^2] = \mathrm{tr}[(zz^\dagger)^2] = 1$, we have $\frac{1}{2}\|D\|_F^2 = 1 - |x^\dagger z|^2$. Write $\Pi^\perp = I - \Pi$ for the complementary projector. The blocks $\Pi D\Pi$, $\Pi D\Pi^\perp$, $\Pi^\perp D\Pi$ and $\Pi^\perp D\Pi^\perp$ are orthogonal in the Frobenius inner product and $\Pi^\perp D\Pi = (\Pi D\Pi^\perp)^\dagger$, hence

$$1 - F = E_{\mathrm{c}} + E_{\mathrm{r}} + E_{\times}, \qquad E_{\mathrm{c}} = \tfrac{1}{2}\|\Pi D\Pi\|_F^2, \quad E_{\mathrm{r}} = \tfrac{1}{2}\|\Pi^\perp D\Pi^\perp\|_F^2, \quad E_{\times} = \|\Pi D\Pi^\perp\|_F^2. \tag{16}$$

Each term is nonnegative. Subtracting the SAL errors from the natural-address errors therefore gives an exact attribution of the fidelity gain.

The terms can be evaluated without forming D. Set $a = \Pi x$, $b = \Pi z$, $c = \Pi^\perp x$, $e = \Pi^\perp z$, with masses $p = \|a\|^2$, $q = \|b\|^2$ and overlaps $\alpha = a^\dagger b$, $\beta = c^\dagger e$. Then

$$E_{\mathrm{c}} = \tfrac{1}{2}\big(p^2 + q^2 - 2|\alpha|^2\big), \qquad E_{\mathrm{r}} = \tfrac{1}{2}\big\{(1-p)^2 + (1-q)^2 - 2|\beta|^2\big\},$$
$$E_{\times} = p(1-p) + q(1-q) - 2\Re(\alpha\bar{\beta}).$$

Since $x^\dagger z = \alpha + \beta$, these terms sum to $1 - |\alpha + \beta|^2$ and require only projections and inner products.

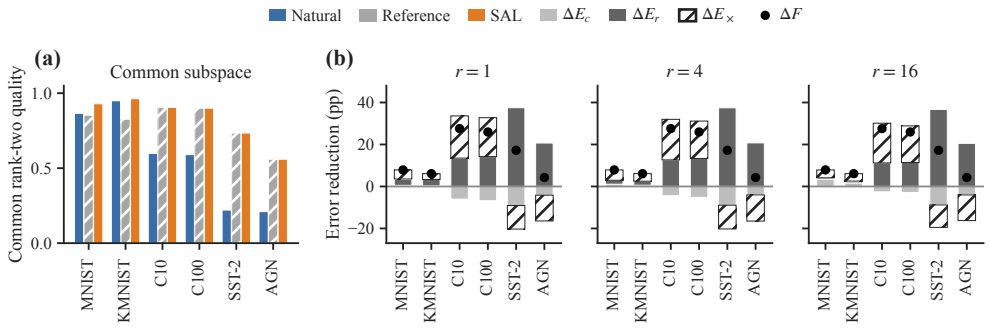


Figure 27: (a) Common-component rank-two quality under the natural order, the principal-reference order and SAL. (b) Signed infidelity decomposition at the reference budget for descriptor ranks $r = 1, 4, 16$; positive values mean smaller error under SAL, and the three terms sum to ΔF (dot).

To distinguish changes in subspace mass from changes in direction, define $f_c = |\alpha|^2/(pq)$ when $pq > 0$ and $f_r = |\beta|^2/[(1-p)(1-q)]$ when $(1-p)(1-q) > 0$. This gives

$$E_c = \tfrac{1}{2}(p-q)^2 + pq(1-f_c), \qquad E_r = \tfrac{1}{2}(p-q)^2 + (1-p)(1-q)(1-f_r).$$

Residual-block error can thus decrease through changes in subspace mass, better directional agreement, or both. The cross term measures agreement in the coupling between the two subspaces. For each target, we invert the selected address in classical analysis before applying the same training projector, then average the three terms over targets and runs.

G.5 Attribution Across Ranks

Table 22 and Figure 27b give the decomposition at G_{ref} for $r = 1, 4, 16$, with positive values indicating smaller error under SAL. On MNIST and KMNIST, all three error terms decrease, and the cross term contributes most at the primary rank. CIFAR-10 and CIFAR-100 show large reductions in residual and cross errors alongside a smaller increase in common-component error. On text, the residual error falls sharply while the common and cross errors increase, with changes in subspace mass contributing strongly. These patterns persist across the three descriptor ranks.

Table 22: Exact attribution of the reference-budget fidelity gain (percentage points) at three descriptor ranks. $\Delta E_c + \Delta E_r + \Delta E_\times = \Delta F$ up to rounding; positive values mean smaller error under SAL.

Dataset	r	ΔF	ΔE_c	ΔE_r	$\Delta E_\times$
MNIST	1	+7.881	+0.536	+3.019	+4.326
	4	+7.881	+1.083	+1.863	+4.934
	16	+7.881	+3.498	+0.629	+3.754
KMNIST	1	+6.100	+0.299	+3.047	+2.754
	4	+6.100	+0.586	+1.917	+3.597
	16	+6.100	+1.710	+0.865	+3.525
CIFAR-10	1	+27.569	−6.043	+13.501	+20.110
	4	+27.569	−4.415	+12.654	+19.329
	16	+27.569	−2.564	+11.207	+18.925
CIFAR-100	1	+25.959	−6.792	+14.217	+18.534
	4	+25.959	−5.191	+13.295	+17.855
	16	+25.959	−2.906	+11.401	+17.464
SST-2	1	+17.218	−9.071	+37.509	−11.220
	4	+17.218	−8.916	+37.459	−11.325
	16	+17.218	−8.806	+36.699	−10.675
AG News	1	+4.269	−4.150	+20.694	−12.275
	4	+4.269	−4.012	+20.763	−12.481
	16	+4.269	−4.025	+20.526	−12.231

G.6 Relations Between Samples

Amplitude kernels have unit diagonal. We remove these self-similarities and apply U-centering to examine relations between distinct samples (Székely & Rizzo, 2014). For a symmetric K with zero

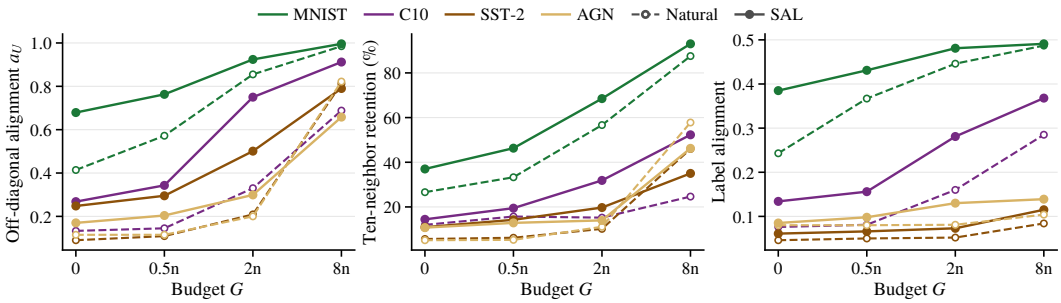


Figure 28: Relational diagnostics across budgets: off-diagonal kernel alignment, ten-neighbor retention and label alignment (dashed: natural address; solid: SAL).

diagonal and $m > 3$,

$$(UK)_{ij} = K_{ij} - \frac{K_{i\cdot}}{m-2} - \frac{K_{\cdot j}}{m-2} + \frac{K_{\cdot\cdot}}{(m-1)(m-2)} \quad (i \neq j), \qquad (UK)_{ii} = 0,$$

and the alignment is $a_U = \langle UK, U\widetilde{K}\rangle_F / (\|UK\|_F \|U\widetilde{K}\|_F)$. This kernel alignment (Kornblith et al., 2019; Murphy et al., 2024) compares the centered relations. We also measure neighbor retention by averaging $|N_k(i;K) \cap N_k(i;\widetilde{K})|/k$ over queries, excluding each query itself, with $k = 10$. Label alignment is the standard centered kernel alignment (CKA; Kornblith et al., 2019) between $\widetilde{K}$ and a class-balanced label kernel $K_y = YY^\top$, where $Y_{ic} = \mathbf{1}[y_i = c]/\sqrt{m_c}$ and m_c is the number of samples in class c: $\mathrm{CKA}(\widetilde{K}, K_y) = \langle \Theta\widetilde{K}\Theta, \Theta K_y \Theta\rangle_F / (\|\Theta\widetilde{K}\Theta\|_F \|\Theta K_y \Theta\|_F)$ with $\Theta = I - \mathbf{1}\mathbf{1}^\top/m$. All measurements use 512 external vectors per run.

Table 23 and Figure 28 report kernel alignment, ten-neighbor retention and label alignment. All three improve at the reference budget on the four representative datasets, and kernel alignment and neighbor retention improve at every measured budget up to G_{ref}. At $8n$, however, SAL lowers both measures on SST-2 and AG News while still improving preparation fidelity and label alignment. Loading and relational gains therefore agree most closely at small-to-intermediate budgets.

Table 23: Relational diagnostics on 512 external vectors per run. a_U: off-diagonal kernel alignment with the ideal kernel; NN10: ten-neighbor retention (%); label: kernel–label alignment. The last row per dataset is the largest loading budget.

		a_U		NN10		Label	
Dataset	G	Natural	SAL	Natural	SAL	Natural	SAL
MNIST	0	0.414	0.679	26.60	36.99	0.243	0.385
	4	0.572	0.763	33.28	46.30	0.367	0.431
	20	0.855	0.924	56.65	68.50	0.446	0.481
	80	0.985	0.996	87.53	93.02	0.487	0.491
CIFAR-10	0	0.133	0.268	11.97	14.44	0.076	0.134
	6	0.145	0.343	15.70	19.45	0.081	0.156
	22	0.330	0.750	15.21	31.87	0.160	0.281
	88	0.688	0.912	24.64	52.29	0.285	0.368
SST-2	0	0.090	0.248	5.64	10.91	0.046	0.061
	4	0.109	0.295	6.20	14.24	0.050	0.066
	20	0.209	0.501	10.18	19.72	0.052	0.073
	80	0.812	0.790	45.95	35.01	0.084	0.115
AG News	0	0.115	0.170	5.10	10.74	0.080	0.085
	4	0.115	0.204	5.25	12.86	0.080	0.098
	20	0.200	0.299	11.11	13.98	0.081	0.130
	80	0.822	0.658	57.80	46.23	0.104	0.139