QuSAE: Quantum Structure-aware amplitude encoding with improved fidelity-complexity trade-off

Anonymous authors
Paper under double-blind review

Abstract

Efficiently loading classical data into quantum states is a prerequisite for translating the potential advantages of quantum computing into practical data-intensive applications. Yet exact amplitude encoding of generic data can require exponentially many gates, which is beyond the reach of current quantum devices. Approximate amplitude encoding alleviates this overhead by trading state fidelity for representation complexity. However, most existing methods treat the target state generically from a quantum perspective, overlooking structural regularities inherent in real-world data that could be exploited to achieve more efficient encoding. To address this gap, we propose a multiscale **Qu**antum **S**tructure-Aware **A**mplitude **E**ncoding (QuSAE) framework, which exploits the intrinsic multiscale organization of classical data to guide structured sparsification and efficient tensorization for approximate state preparation. This enables explicit control over approximation fidelity while reducing representation complexity. Comprehensive experiments across five benchmark datasets demonstrate that QuSAE exhibits favorable fidelity–complexity trade-off over competing methods, while preserving multiscale information in the resulting compact representations. More broadly, our results highlight structure-aware state construction as a general design principle for efficient quantum data encoding.

1 Introduction

Quantum computing has the potential to accelerate a wide range of computational problems that are challenging for classical methods (Harrow et al., 2009; Bravo-Prieto et al., 2023; Low & Su, 2026; Grinko et al., 2021; Claudon et al., 2025; Biamonte et al., 2017; Peral-García et al., 2024; Gil-Fuster et al., 2024). However, quantum resources remain severely limited in the current noisy intermediate-scale quantum (NISQ) era (Bharti et al., 2022), making such computational advantages difficult to realize in practice. This limitation places stringent demands on the resource efficiency of quantum state preparation (QSP), especially for practical quantum applications where classical data must be loaded into a quantum computer. Amplitude encoding provides an attractive approach by representing a high-dimensional classical vector with only a logarithmic number of qubits, yet preparing a generic amplitude-encoded state may require a gate count that grows exponentially with the number of qubits (Plesch & Brukner, 2011; Zhang & Yuan, 2024; Ben-Dov et al., 2024). Consequently, reducing the cost of amplitude encoding is a key step toward achieving practical quantum advantages.

Approximate amplitude encoding (AAE) offers a practical way to alleviate this bottleneck by allowing a controlled loss of encoding accuracy in exchange for a lower preparation complexity (Au-Yeung et al., 2024). Existing methods have made encouraging progress in resource-efficient circuit design and encoding schemes (Zheng et al., 2025; 2026), while increasing attention has been paid to the potential of exploiting intrinsic structure in classical data to further improve encoding performance (Purba et al., 2026). Images, time series, and scientific signals often exhibit strong correlations, sparsity, locality, or multiscale organization that can be potentially exploited during state preparation (Wang et al., 2024; Esteves et al., 2025; Tang et al., 2025). Tensor-network representations provide a natural framework for this purpose. In particular, matrix product states (MPS) can compactly represent structured quantum states using small bond dimensions and can be sys-

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.

tematically converted into state-preparation circuits (Schollwöck, 2011; Orús, 2014; Ran, 2020). Specifically, smooth functions and spectrally concentrated images have been shown to admit efficient low-bond MPS representations, with their bond requirements closely tied to intrinsic regularity and spectral decay (Holmes & Matsuura, 2020; Iaconis et al., 2024; Jobst et al., 2024). These observations suggest that exploiting intrinsic data structures offers a promising route toward a more efficient AAE.

However, existing MPS-based approximate encoding methods do not fully exploit this structural opportunity. The compactness of an MPS depends both on the representation of the input data and the organization of its amplitudes across the tensorized quantum register. Existing methods typically address these choices separately, focusing on either data representation or qubit ordering while keeping the other fixed or optimized heuristically (Jobst et al., 2024; Jeon et al., 2024). As a result, the compactness depends heavily on heuristic choices made prior to MPS construction, with their impact only observable after the resulting MPS is obtained. The intrinsic structure of the data is not systematically translated into a representation that is favorable for subsequent MPS construction, potentially leading to unnecessarily large complexity for a desired approximation accuracy. Hence, there exists a pressing need for a unified structure-aware encoding principle that jointly exploits data organization and tensorization to construct accurate yet compact amplitude-encoded states.

In this work, we propose **QuSAE**, a structure-aware approximate quantum amplitude encoding framework that exploits the multiscale organization of classical data for representation and tensorization. QuSAE first transforms the input into the wavelet domain to expose its multiscale structure as a hierarchical organization of coefficients, with the retained wavelet energy providing direct control over the approximation fidelity. It then organizes the retained coefficients over computational-basis indices according to their structural relationships, promoting a tensorization with reduced MPS bond demand. Based on this mechanism design, QuSAE jointly determines **what information is retained and how the retained information is tensorized**, enabling a prescribed encoding fidelity to be achieved with a more compact MPS representation. This structure-aware construction provides a principled route toward **improving the fidelity-complexity trade-off** of AAE.

Our main contributions are summarized as follows: (i) we propose QuSAE, a structure-aware encoding framework that jointly exploits sparse representation and effective mapping for compact MPS construction under a prescribed encoding quality; (ii) we establish fidelity guarantees for wavelet-tree truncation and derive a prefix-complexity upper bound that connects mapped support geometry to MPS bond demand; (iii) we provide numerical evidence across abundant real-world datasets that QuSAE improves the fidelity–complexity trade-off over direct MPS construction and unstructured baselines; (iv) we show that the proposed structure-aware encoding enables controllable compression scaling while preserving geometric information under the prescribed fidelity. Code is available at the given repository.

2 PRELIMINARIES

2.1 PROBLEM FORMULATION

Given a nonzero classical vector $\boldsymbol{x} = (x_0, \ldots, x_{N-1}) \in \mathbb{C}^N$ with $N = 2^n$, amplitude encoding represents it as the n-qubit state $|\psi(\boldsymbol{x})\rangle = \|\boldsymbol{x}\|_2^{-1} \sum_{j=0}^{N-1} x_j |j\rangle$. Exact preparation of a generic amplitude-encoded state may require a number of quantum gates exponential in n. We therefore consider approximate amplitude encoding. Given an allowable infidelity $\varepsilon \in [0, 1)$, AAE aims to construct an approximate state $\left|\widetilde{\psi}\right\rangle$ satisfying $F(\left|\widetilde{\psi}\right\rangle, |\psi(\boldsymbol{x})\rangle) = |\langle\psi(\boldsymbol{x})|\widetilde{\psi}\rangle|^2 \geq 1 - \varepsilon$, while reducing the complexity required for its representation and preparation.

In this work, we represent the approximate state using a MPS, $\left|\widetilde{\psi}\right\rangle = \sum_{\sigma_1,\ldots,\sigma_n \in \{0,1\}} A^{[1]\sigma_1} \cdots A^{[n]\sigma_n} |\sigma_1 \cdots \sigma_n\rangle$, where $A^{[i]\sigma_i} \in \mathbb{C}^{\chi_{i-1} \times \chi_i}$ and $\chi_0 = \chi_n = 1$. The internal dimensions $\boldsymbol{\chi} = (\chi_1, \ldots, \chi_{n-1})$ form the bond profile of the MPS. For a fixed tensorization, the minimum bond dimension required at the bipartition $1{:}i \,|\, i{+}1{:}n$ is given by the rank of the corresponding tensor unfolding, or equivalently by the Schmidt rank across that cut. The bond profile therefore characterizes the complexity of the MPS representation, with smaller

bond dimensions generally leading to a more compact representation. Refer to Appendix A for more details about MPS representation and complexity analysis.

Concretely, we quantify the MPS representation complexity as $\mathcal{C}_{\mathrm{MPS}}(\boldsymbol{\chi}) = \sum_{i=1}^{n} 2\chi_{i-1}\chi_i$, which measures the total number of entries in the MPS tensors. Let $\boldsymbol{\chi}(\left|\widetilde{\psi}\right\rangle)$ denote the minimal bond profile of $\left|\widetilde{\psi}\right\rangle$ under a given tensorization. The objective of MPS-based AAE is therefore

$$\min_{\left|\widetilde{\psi}\right\rangle} \quad \mathcal{C}_{\mathrm{MPS}}\Big(\boldsymbol{\chi}(\left|\widetilde{\psi}\right\rangle)\Big) \quad \text{s.t.} \quad F\Big(\left|\widetilde{\psi}\right\rangle, |\psi(\boldsymbol{x})\rangle\Big) \geq 1-\varepsilon. \tag{1}$$

2.2 Related work

AAE methods can be grouped into circuit-based and tensor-network (TN)-based approaches.

Circuit-based AAE. Circuit-based methods directly construct parameterized quantum circuits (PQCs) and optimize them to approximate the target state in Hilbert space, as in approximate amplitude encoding Nakaji et al. (2022) and its complex-valued extensions Mitsuda et al. (2024). Non-variational methods include automatic quantum circuit encoding (AQCE), which determines two-qubit unitaries through local singular value decomposition (SVD) updates (Shirakawa et al., 2024). More recently, AQER guides circuit growth by reducing entanglement and subsequently refines the preparation fidelity (Zhang et al., 2026). Although expressive and hardware-native, such variational loading schemes often face poor scalability for high-dimensional states, with trainability further challenged by barren-plateau phenomena McClean et al. (2018); Larocca et al. (2024).

TN-based AAE. TN-based methods approximate target amplitudes by tensor factorizations and derive loading circuits from the resulting tensors. MPS compilation includes sequential generation and iterative disentangling (Schön et al., 2005; Ran, 2020), with subsequent circuit optimization improving the approximation (Rudolph et al., 2024). Applications include probability distributions (Iaconis et al., 2024) and image datasets (Kiwit et al., 2025; Green & Wang, 2026). Related preparation protocols provide logarithmic-depth guarantees for normal MPS (Malz et al., 2024). Our method complements existing encoding strategies by providing a structure-aware perspective from the classical-data side and jointly considering the effects of data representation and qubit mapping on MPS construction.

3 Methods

3.1 Overall framework

The general objective in Eqn. (1) optimizes over an approximate quantum state $\left|\widetilde{\psi}\right\rangle$, without specifying how such a state should be constructed. QuSAE formulates this problem by parameterizing the approximate encoding with two coupled variables: a coefficient support set Ω, which determines **what information is retained**, and a qubit-index mapping π, which determines **how the retained information is tensorized**.

Specifically, Ω denotes the coefficient support retained after structured truncation in the wavelet domain, determining the structured sparsity and approximation error of the resulting state. A sparser support can tighten the unfolding-rank envelope across MPS bipartitions and potentially reduce the required bond dimensions. Given the retained support, π assigns these coefficients to computational-basis indices and hence determines how they are partitioned across successive MPS cuts. A common reindexing does not alter the retained coefficient values or the corresponding approximation fidelity, but can substantially change the unfolding ranks and thus the resulting bond profile. Therefore, while Ω primarily controls the fidelity of the approximation, both Ω and π jointly determine its MPS representation complexity. Denoting the bond profile induced by a support–mapping pair as $\boldsymbol{\chi}(\Omega, \pi)$ and the corresponding approximation fidelity as $F(\Omega)$, the MPS-based AAE objective can be instantiated as

$$\min_{\Omega,\pi} \quad \mathcal{C}_{\mathrm{MPS}}\big(\boldsymbol{\chi}(\Omega,\pi)\big) \quad \text{s.t.} \quad F(\Omega) \geq 1-\varepsilon. \tag{2}$$

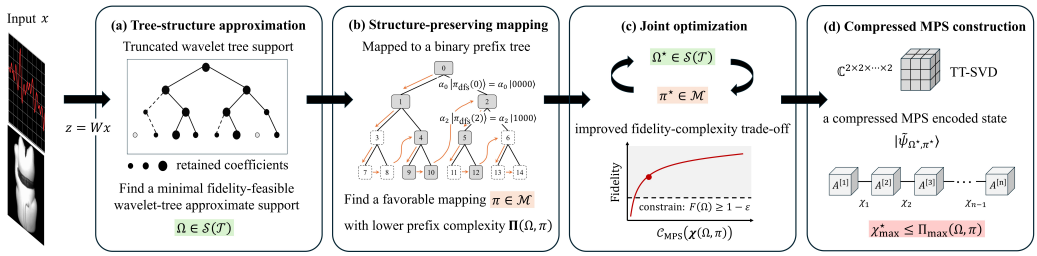

Figure 1: **Overview of QuSAE.** QuSAE consists of four main stages. (a) Tree-structured approximation. A structured sparse support is constructed under the prescribed fidelity. (b) Structure-preserving mapping. The retained support is reorganized to a binary prefix tree. (c) Joint optimization. The support and mapping are jointly refined to balance encoding fidelity and representation complexity. (d) Compressed MPS construction. The resulting state is represented as a compressed MPS with resource bounds.

Directly solving Eqn. (2) is combinatorial. Without structural constraints, Ω may range over exponentially many possible supports. Even for a fixed support, the mapping π still admits factorial many possible coefficient orderings. Moreover, evaluating a candidate pair generally requires examining the ranks or singular spectra across multiple MPS bipartitions. Exhaustive joint optimization is therefore intractable even at moderate input dimensions.

QuSAE addresses this challenge by exploiting intrinsic data structure to restrict and coordinate the two search spaces. We first develop a structured tree approximation that efficiently determines Ω under a prescribed fidelity requirement. We then construct a structure-preserving mapping that organizes the retained coefficients to reduce MPS bond demand. Finally, we combine the two components through a joint support–mapping optimization. The overall framework of QuSAE is shown in Fig. 1.

3.2 Tree-structured approximation

We first address support selection by exploiting the multiscale organization of the input through the wavelet tree model (WTM). Given an input sample $\boldsymbol{x}$, an orthogonal discrete wavelet transform (DWT) W yields the coefficient vector $\boldsymbol{z} = W\boldsymbol{x}$. These wavelet coefficients are organized into a rooted hierarchy $\mathcal{T} = (V, E, r)$, where $V = \{0, \ldots, N-1\}$ is the coefficient index set, E specifies the parent–child relations, and $r \in V$ is the root index. Each node $j \in V$ carries coefficient z_j. For one-dimensional inputs, the detail coefficients form a binary coarse-to-fine tree, whereas for two-dimensional inputs the horizontal, vertical, and diagonal subbands form three orientation-specific quadtrees sharing the coarsest approximation root.

Significant wavelet coefficients of real-world data often persist across neighboring locations and parent–child scales, producing clustered branches and subtrees in this hierarchy (Mallat, 2009; Cohen et al., 2001; Baraniuk et al., 2010). QuSAE incorporates this structural prior by restricting the retained support to $\mathcal{S}(\mathcal{T})$, the family of rooted ancestor-closed subsets of V containing r. Thus, retaining a node requires retaining all of its ancestors. Each $\Omega \in \mathcal{S}(\mathcal{T})$ induces a retained subtree $\mathcal{T}[\Omega]$, and $\boldsymbol{z}_\Omega \in \mathbb{C}^N$ denotes the coefficient vector obtained by retaining z_j for $j \in \Omega$ and setting all remaining entries to zero. The following proposition connects this structured truncation directly to the encoding fidelity.

Proposition 1 (Fidelity guarantee)**.** *For any $\Omega \subseteq V$ with $\|\boldsymbol{z}_\Omega\|_2 > 0$ and any bijective index mapping $\pi : V \to \{0,1\}^n$, define $U_\pi |j\rangle = |\pi(j)\rangle$ and $|\psi_{\Omega,\pi}\rangle = U_\pi |\psi(\boldsymbol{z}_\Omega)\rangle$. Then*

$$F(|\psi_{\Omega,\pi}\rangle, U_\pi |\psi(\boldsymbol{z})\rangle) = F(|\psi(\boldsymbol{z}_\Omega)\rangle, |\psi(\boldsymbol{z})\rangle) = \frac{\|\boldsymbol{z}_\Omega\|_2^2}{\|\boldsymbol{z}\|_2^2} = 1 - \frac{\|\boldsymbol{z} - \boldsymbol{z}_\Omega\|_2^2}{\|\boldsymbol{z}\|_2^2}. \tag{3}$$

Denote this retained-energy ratio by $F(\Omega)$. For any normalized tensor-train SVD (TT-SVD) approximation $\left|\widetilde{\psi}\right\rangle$ with realization infidelity $\eta = 1 - F(\left|\widetilde{\psi}\right\rangle, |\psi_{\Omega,\pi}\rangle)$, the condition $F(\Omega) \geq 1 - \varepsilon$ implies

$$F\left(\left|\widetilde{\psi}\right\rangle, U_\pi |\psi(\boldsymbol{z})\rangle\right) \geq 1 - \varepsilon - \eta - 2\sqrt{\varepsilon\eta}. \tag{4}$$

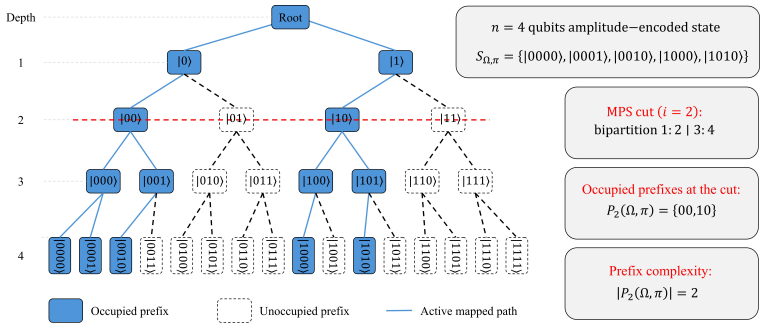

Figure 2: **Prefix-tree view of the mapped support.** Each active basis index corresponds to a root-to-leaf path. Filled nodes and solid paths denote occupied prefixes induced by $S_{\Omega,\pi}$. At MPS cut i, the occupied prefix and suffix sets determine the prefix complexity $\Pi_i(\Omega,\pi)$, which upper bounds the corresponding bond dimension $\chi_i(\Omega,\pi)$.

For exact TT-SVD, $\eta = 0$, and the final fidelity equals $F(\Omega)$.

Proposition 1 establishes the fidelity guarantee of the tree-structured approximation. Eqn. (3) converts the fidelity constraint into an additive retained-energy condition, while the second part shows that this certificate is invariant under any common index mapping and is affected only by subsequent TT-SVD realization error.

Hence, a support satisfying $F(\Omega) \geq 1-\varepsilon$ can be fidelity-certified before MPS construction. Accordingly, the tree-approximation stage obtains a minimum ancestor-closed support Ω_ε satisfying this certificate using an exact rooted tree-knapsack dynamic program, with $K_\varepsilon = |\Omega_\varepsilon|$ and computational complexity $\mathcal{O}(N^2)$. Implementation details are provided in Appendix C.2. This procedure is energy optimal within the structured family $\mathcal{S}(\mathcal{T})$, but does not necessarily yield the compact MPS representation. The retained hierarchy $\mathcal{T}[\Omega]$ also guides coefficient assignment to computational-basis indices, motivating the structure-aware mapping and joint optimization developed next.

3.3 Structure-preserving mapping

Having obtained a fidelity-controlled structured support, we next determine how the retained coefficients should be organized for compact MPS representation. To better understand this compression objective, we introduce a geometric view of the MPS bond demand by representing the mapped support as occupied leaves of a depth-n binary prefix tree, as illustrated in Fig. 2.

For a structured support $\Omega \in \mathcal{S}(\mathcal{T})$ and a valid mapping π, the mapped support is denoted as $S_{\Omega,\pi} = \{\pi(j) \mid j \in \Omega\}$. For $k \in S_{\Omega,\pi}$ with binary address $b_1 \cdots b_n$, the MPS cut $1{:}i \,|\, i{+}1{:}n$ separates the address into prefix $b_{1:i}$ and suffix $b_{i+1:n}$. We define the corresponding occupied prefix and suffix sets as $\mathcal{P}_i(\Omega,\pi) = \{b_{1:i} \mid k \in S_{\Omega,\pi}\}$ and $\mathcal{Q}_i(\Omega,\pi) = \{b_{i+1:n} \mid k \in S_{\Omega,\pi}\}$, which identify the potentially nonzero rows and columns of the i-th unfolding.

Proposition 2 (Prefix-complexity-induced bond bound)**.** *For any $\Omega \in \mathcal{S}(\mathcal{T})$ and any valid mapping π, the exact bond dimension across the i-th MPS cut satisfies*

$$\chi_i(\Omega,\pi) \leq \Pi_i(\Omega,\pi) \equiv \min\left\{|\mathcal{P}_i(\Omega,\pi)|,\ |\mathcal{Q}_i(\Omega,\pi)|\right\}. \tag{5}$$

The proof is provided in Appendix B.2. We denote the resulting prefix-complexity profile by $\mathbf{\Pi}(\Omega,\pi) = (\Pi_1(\Omega,\pi),\ldots,\Pi_{n-1}(\Omega,\pi))$, which provides a geometric and efficiently computable upper bound on the corresponding bond profile $\boldsymbol{\chi}(\Omega,\pi)$. This provides a tractable surrogate for explicit bond computation that can be evaluated in $\mathcal{O}(K\log N) = \mathcal{O}(nK)$ time for a support of size $K = |\Omega|$, avoiding repeated rank or SVD evaluations during mapping search.

This prefix-tree view also reveals why the tree-structured support is useful beyond coefficient sparsity. The ancestor-closed branches of $\mathcal{T}[\Omega]$ provide a favorable geometry for inducing shared paths in the binary prefix tree under a suitable mapping, which can reduce the prefix complexity. QuSAE therefore restricts the mapping search to a finite family $\mathcal{M}$ of structure-preserving mappings, rather than searching over arbitrary coefficient permutations. Each $\pi_\tau \in \mathcal{M}$ is induced by a complete node

ordering $\tau = (v_0, \ldots, v_{N-1})$, with $\pi_\tau(v_j) = \mathrm{bin}_n(j)$. Representative candidates include coarse-to-fine orderings, depth-first traversal, and other prescribed tree traversals, as detailed in Appendix C.3. Mappings with lower prefix complexity are favored in the subsequent joint optimization.

3.4 Joint support–mapping optimization

Support selection and coefficient mapping have coupled effects on the resulting MPS complexity. QuSAE therefore integrates fidelity-controlled tree approximation with the prefix-guided mapping searches through a joint support–mapping optimization, followed by TT-SVD construction of the final MPS-encoded states.

QuSAE formulates the joint optimization as

$$\begin{aligned} \min_{\pi \in \mathcal{M},\, \Omega \in \mathcal{S}(\mathcal{T})} \quad & \eta_K \frac{|\Omega|}{N} + \eta_\Pi \mathcal{C}_\Pi(\Omega, \pi) \\ \text{subject to} \quad & F(\Omega) \geq 1 - \varepsilon, \end{aligned} \tag{6}$$

where $\eta_K, \eta_\Pi \geq 0$ balance support sparsity and mapping demand. The prefix-based surrogate is defined as $\mathcal{C}_\Pi(\Omega, \pi) = \sum_{i=1}^{n-1} w_i \Pi_i(\Omega, \pi) / \min\{2^i, 2^{n-i}\}$, with $w_i \geq 0$. This normalized prefix budget provides a tractable surrogate for MPS bond demand in practical optimization.

In practice, QuSAE solves the constrained problem through alternating support and mapping updates. (i) During the support update, each sample selects its ancestor-closed support from a finite fidelity-feasible candidate pool under the current mapping using the sparsity–prefix objective. (ii) During the mapping update, a shared mapping is selected from $\mathcal{M}$ by minimizing the same calibration-averaged objective with the supports held fixed. Every accepted update decreases the fixed objective beyond a numerical tolerance. The coefficients η_K and η_Π are scanned across separate runs and remain fixed within each run to balance support sparsity and prefix complexity. The best fidelity-feasible iterate is retained, which yields a dataset-shared mapping $\pi^\star$ and sample-specific supports $\Omega^\star$. Overall, QuSAE avoids unrestricted enumeration by restricting both support and mapping to finite structured candidate sets. Further implementation details are provided in Appendix C.4.

For each optimized pair $(\Omega^\star, \pi^\star)$, TT-SVD constructs the MPS tensors $A^\star$ with bond profile $\boldsymbol{\chi}^\star = \boldsymbol{\chi}(\Omega^\star, \pi^\star)$, yielding the encoded state $\left|\widetilde{\psi}_{\Omega^\star, \pi^\star}\right\rangle$ without imposing a post-hoc bond cap. We use $\tau_{\mathrm{svd}} = 10^{-12}$ only to remove machine-precision singular values. The resulting realization infidelity η is therefore negligible, and Proposition 1 implies that the final encoded-state fidelity is effectively controlled by the prescribed tolerance ε. Since the orthogonal wavelet transform and the shared mapping are unitary changes of coordinates, QuSAE directly uses the mapped wavelet state as the encoding target, without applying an inverse DWT or inverse index mapping before downstream quantum tasks.

3.5 Resource analysis

We characterize the resource scaling of QuSAE through two structural conditions on the fidelity-controlled support and its mapped prefix geometry.

Following the notion of s-tree compressibility (Cohen et al., 2001; Baraniuk et al., 2010), we define the budget-constrained support family as $\mathcal{S}_K(\mathcal{T}) = \{\Omega \in \mathcal{S}(\mathcal{T}) : |\Omega| \leq K\}$ and the optimal tree-approximation error as $e_{\mathrm{tree}}(K) = \min_{\Omega \in \mathcal{S}_K(\mathcal{T})} \|\boldsymbol{z} - \boldsymbol{z}_\Omega\|_2^2 / \|\boldsymbol{z}\|_2^2$. We consider an input to be s-tree-compressible if $e_{\mathrm{tree}}(K) \leq \gamma_{\mathrm{tree}} K^{-2s}$ for some $s > 0$ and $\gamma_{\mathrm{tree}} > 0$. By Eqn. (3), the minimum fidelity-compatible budget $K_\varepsilon = \min\{K : e_{\mathrm{tree}}(K) \leq \varepsilon\}$ satisfies $K_\varepsilon = \mathcal{O}(\varepsilon^{-1/(2s)})$.

Motivated by the hierarchical supports produced by tree approximation, which can induce shared prefix paths under a suitable structure-preserving mapping, we further consider a prefix-compressible regime satisfying $\Pi_{\max}(\Omega, \pi) := \max_{1 \leq i \leq n-1} \Pi_i(\Omega, \pi) \lesssim \gamma_\pi |\Omega|^\beta$ with $0 < \beta < 1$. Here, β characterizes the effective prefix-growth rate, while γ_π reflects the mapping-dependent degree of path sharing.

For s-tree-compressible data under a prefix-compressible mapping, Proposition 2 yields

$$\chi^\star_{\max} := \max_i \chi_i(\Omega^\star, \pi^\star) \leq \Pi_{\max}(\Omega^\star, \pi^\star) \lesssim \gamma_{\pi^\star} \gamma_{\mathrm{tree}}^{\beta/(2s)} \varepsilon^{-\beta/(2s)} = \widetilde{\gamma}\, \varepsilon^{-\beta/(2s)}, \tag{7}$$

where $\widetilde{\gamma} = \gamma_{\pi^\star}\gamma_{\mathrm{tree}}^{\beta/(2s)}$. Thus, the maximum MPS bond demand grows sublinearly with the fidelity-controlled support size. Using the representative $\mathcal{O}(n\chi_{\max}^2)$ cost of sequential MPS preparation (Schön et al., 2005; Jobst et al., 2024), Eqn. (7) further implies that the required two-qubit gate count is expected to scale as $G_{2q}(\varepsilon) = \mathcal{O}(n\varepsilon^{-\beta/s})$ with the prescribed fidelity requirement for a given dataset. These conditional bounds characterize a structure-aware regime in which QuSAE exploits the intrinsic structure of real-world data to achieve favorable theoretical scaling of MPS preparation resources.

4 EXPERIMENTS

In this section, we conduct numerical simulations to evaluate QuSAE on representative image and signal datasets against multiple reference methods, emphasizing its advantages in improving the fidelity–complexity trade-off. We further isolate the three stages of QuSAE to examine their respective contributions, and reconstruct the final encoded states for complementary assessment of information preservation. Additional results on input-side structural analysis, resource-scaling behavior, and sequential quantum circuit realization are provided in Appendix D.

4.1 SETUP

Datasets. We evaluate QuSAE on three image datasets, MNIST (LeCun et al., 1998), COIL-20 (Nene et al., 1996), and CelebA (Liu et al., 2015), and two one-dimensional signal datasets, MIT-BIH (Moody & Mark, 2001; Goldberger et al., 2000) and STEAD (Mousavi et al., 2019). MNIST is resized to 32×32 ($N = 2^{10}$); COIL-20 and CelebA use 128×128 (2^{14}) and 64×64 (2^{12}) images, while MIT-BIH and STEAD use 2^8- and 2^{11}-point segments, respectively. For each dataset, we uniformly sample 100 examples per class or use all available samples when sample size is limited. Detailed preprocessing, tree construction, and preliminary analysis are provided in the Appendix C.1.

Evaluation metrics. For the final MPS-encoded state, we evaluate encoding quality using state fidelity F and representation complexity $\mathcal{C}_{\mathrm{MPS}}(\boldsymbol{\chi})$. For image datasets, we further compute the gradient magnitude similarity deviation (GMSD) (Xue et al., 2014) between the original and reconstructed images as a complementary metric for structural information preservation.

Reference methods. We compare QuSAE with three representative baselines: (i) **Standard MPS** applies TT-SVD directly under a fixed heuristic mapping, which has been shown to be efficient for approximating structured classical data (Jobst et al., 2024); (ii) **Permuted MPS** reproduces the optimal qubit-mapping method of Jeon et al. (Jeon et al., 2024), which searches over qubit permutations to reduce MPS truncation loss; (iii) **Low-rank state preparation (LRSP)** performs bipartite Schmidt decomposition followed by rank truncation (Araujo et al., 2024), providing a complementary baseline to the sequential truncation employed by MPS.

4.2 EXPERIMENTAL RESULTS

QuSAE consistently improves the fidelity–complexity trade-off over reference methods. For each dataset, we encode the same set of samples using QuSAE, standard MPS, permuted MPS, and LRSP, and record the resulting fidelity and realized bond profile. QuSAE is evaluated over a predefined grid of target infidelities ε, with each operating point obtained by solving the corresponding fidelity-constrained support–mapping optimization problem. The reference methods instead sweep the maximum bond-dimension cap $\chi_{\max}$ over its feasible range. Fig. 3 summarizes the resulting fidelity–complexity trade-offs across the five datasets. Standard MPS generally outperforms LRSP on these structured real-world datasets, while permuted MPS further reduces MPS truncation loss by qubit ordering. However, the gain from permutation alone narrows for more complex image data, decreasing from 1.66 percentage on MNIST to 0.82 on COIL-20 and only 0.67 on CelebA, while the improvement on one-dimensional data is even less pronounced. Moreover, permuted MPS requires up to $n!$ sample-wise permutation searches, whereas QuSAE restricts the search to structured candidates with $\mathcal{O}(nK)$ prefix evaluation at the dataset level. Overall, QuSAE reaches a prescribed fidelity with lower MPS representation complexity, with the advantage being most evident in the medium- to high-fidelity regime.

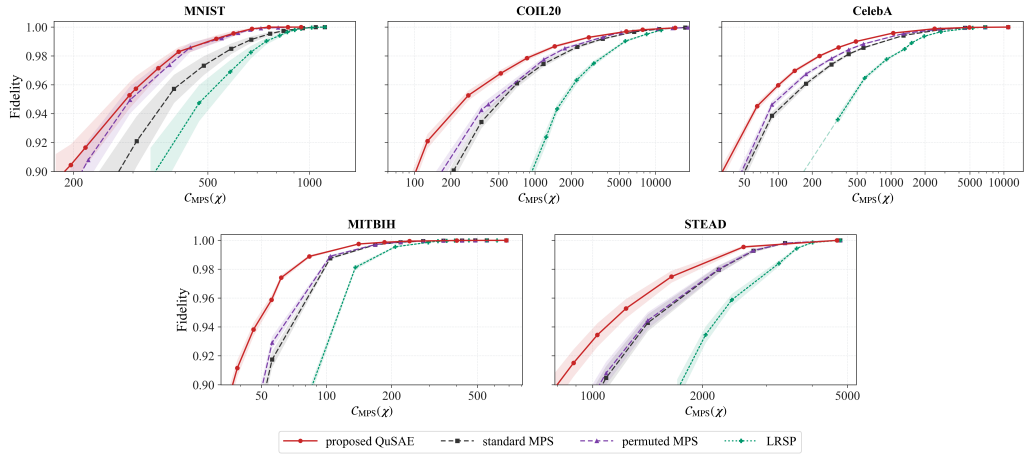

Figure 3: **Fidelity–complexity trade-offs across five datasets.** State fidelity is plotted against the MPS representation complexity for QuSAE, standard MPS, permuted MPS, and LRSP. Markers denote operating points with $F \geq 0.9$, and shaded regions indicate sample-level variability. Points located in the upper left indicate a better trade-off.

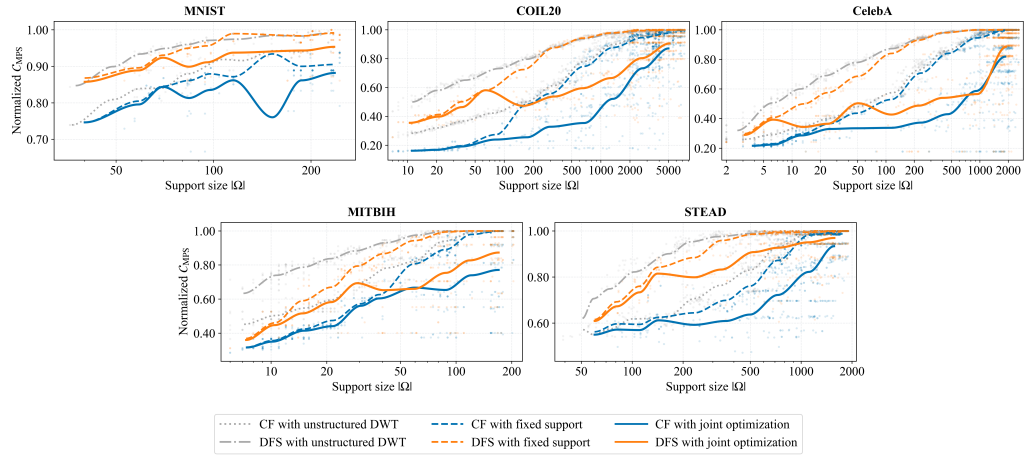

Figure 4: **Ablation experiment results.** Blue and orange denote the coarse-to-fine (CF) and depth-first-search (DFS) mapping strategies, respectively. Light-colored points represent the normalized MPS representation complexity of individual samples, and the curves show the corresponding mean values across samples at each support size. Solid curves correspond to joint optimization of support and mapping, dashed curves use fixed support, and gray curves show the unstructured-DWT baselines under the corresponding mappings.

Optimized tree-structured supports and structure-preserving mappings contribute to compact MPS representations. We further isolate the three stages of our method to assess their individual contributions. As shown in Fig. 4, different mapping strategies yield different representation complexities for the same retained support, highlighting the importance of mapping optimization. The gray unstructured-DWT curves generally exhibit higher complexity than their structured counterparts, indicating that a structured representation promotes shared prefix paths and more compact MPS representations. Moreover, the dashed curves remain above or overlap the corresponding solid curves, supporting the benefit of jointly considering representation and mapping rather than optimizing them separately.

QuSAE better preserves image gradient structures under compressed MPS encoding. For image data, we reconstruct the encoded amplitudes only for evaluation and visualization. We use

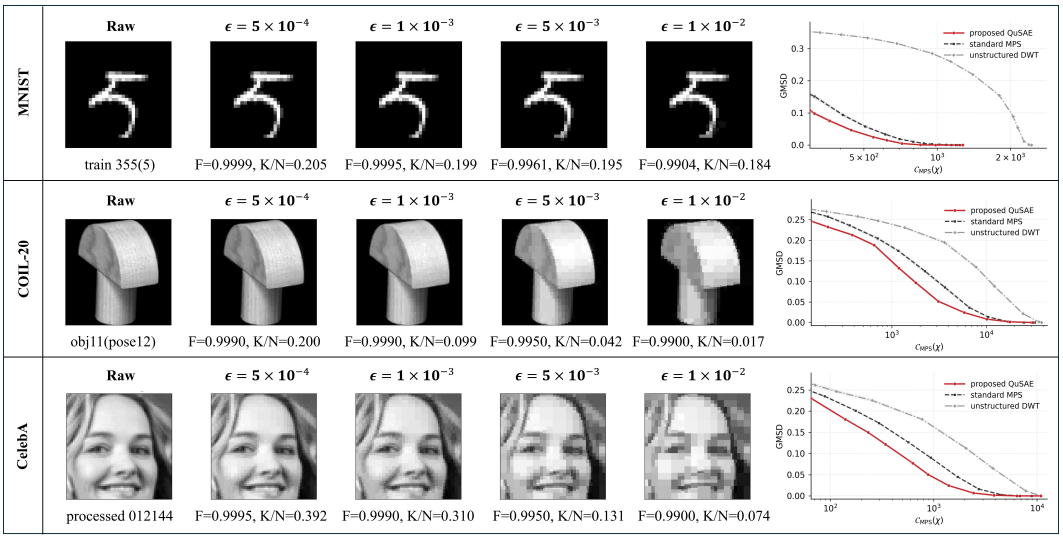

Figure 5: **Reconstruction quality evaluation of QuSAE.** Each row presents a randomly selected sample from MNIST, COIL-20, and CelebA, respectively. The middle panels show reconstructions of the encoded amplitude states under different fidelity tolerances, together with the realized fidelity F and normalized support size K/N. The right panels compare GMSD against the MPS representation complexity for QuSAE, standard MPS, and unstructured DWT.

GMSD to quantify local gradient-structure distortion, where a lower value indicates better preservation of edges and other high-frequency structures (Xue et al., 2014). The reconstructed images in Fig. 5 show that QuSAE retains visually sharper object boundaries and local edge structures despite increasingly sparse representations. Numerically, QuSAE also achieves lower GMSD at the same representation complexity than both standard and unstructured counterparts. These results indicate the potential of structure-aware encoding to preserve multiscale structural features, ons, which may be beneficial for downstream quantum learning.

5 Conclusion

In this work, we proposed QuSAE, a structure-aware framework for fidelity-controlled quantum amplitude encoding. QuSAE integrated wavelet-tree approximation with structure-preserving mapping, where retained wavelet energy controls encoding fidelity and prefix complexity provides a tractable upper bound on MPS bond demand. Experiments on representative image and signal datasets demonstrated improved fidelity–complexity trade-offs over reference methods, while reconstruction evaluation further indicates improved preservation of structural information in images. These results establish a practical connection between structured classical representations and resource-efficient quantum state preparation.

AI use statement

Generative AI tools were used to assist with the preparation and revision of this manuscript, including language polishing, improving readability and clarity, and rephrasing author-written text. They were not used to generate research ideas, develop theoretical results or proofs, design experiments, implement methods, process data, or interpret experimental results. All scientific content, mathematical derivations, experimental design, analyses, and conclusions were produced and verified by the authors. The authors reviewed all AI-assisted edits and take full responsibility for the final content of this work.

Reproducibility statement

We are committed to supporting the independent verification and reproduction of the results presented in this work. We have provided the source code and supporting experimental material on an anonymous platform for peer review. Upon acceptance, we will make these reproducibility-related materials publicly accessible.

References

Israel F. Araujo, Carsten Blank, Ismael C. S. Araújo, and Adenilton J. da Silva. Low-rank quantum state preparation. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 43(1):161–170, 2024. doi: 10.1109/TCAD.2023.3297972.

Rhonda Au-Yeung, Bruno Camino, Ojas Rathore, and Viv Kendon. Quantum algorithms for scientific computing. *Reports on Progress in Physics*, 87(11):116001, 2024. doi: 10.1088/1361-6633/ad85f0.

Richard G. Baraniuk, Volkan Cevher, Marco F. Duarte, and Chinmay Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 56(4):1982–2001, 2010. doi: 10.1109/TIT.2010.2040894.

Matan Ben-Dov, David Shnaiderov, Adi Makmal, and Emanuele G. Dalla Torre. Approximate encoding of quantum states using shallow circuits. *npj Quantum Information*, 10(1):65, 2024. doi: 10.1038/s41534-024-00858-1.

Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S Kottmann, Tim Menke, et al. Noisy intermediate-scale quantum algorithms. *Reviews of Modern Physics*, 94(1):015004, 2022.

Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017. doi: 10.1038/nature23474.

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

Baptiste Claudon, Jean-Philip Piquemal, and Pierre Monmarché. Quantum speedup for non-reversible markov chains. *Nature Communications*, 16:10732, 2025. doi: 10.1038/s41467-025-65761-5.

Albert Cohen, Wolfgang Dahmen, Ingrid Daubechies, and Ronald DeVore. Tree approximation and optimal encoding. *Applied and Computational Harmonic Analysis*, 11(2):192–226, 2001. doi: 10.1006/acha.2001.0336.

Carlos Esteves, Mohammed Suhail, and Ameesh Makadia. Spectral image tokenizer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 17181–17190, 2025. doi: 10.1109/ICCV51701.2025.01596.

Elies Gil-Fuster, Jens Eisert, and Carlos Bravo-Prieto. Understanding quantum machine learning also requires rethinking generalization. *Nature Communications*, 15:2277, 2024. doi: 10.1038/s41467-024-45882-z.

Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000. doi: 10.1161/01.CIR.101.23.e215.

Josh Green and Jingbo Wang. Quantum encoding of functions and images with matrix product states. *Physical Review A*, 113(5):052616, 2026. doi: 10.1103/1mn6-j57y.

Dmitry Grinko, Julien Gacon, Christa Zoufal, and Stefan Woerner. Iterative quantum amplitude estimation. *npj Quantum Information*, 7:52, 2021. doi: 10.1038/s41534-021-00379-1.

Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103(15):150502, 2009. doi: 10.1103/PhysRevLett.103.150502.

Adam Holmes and A. Y. Matsuura. Efficient quantum circuits for accurate state preparation of smooth, differentiable functions. In *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pp. 169–179. IEEE, 2020. doi: 10.1109/QCE49297.2020.00030.

Jason Iaconis, Sonika Johri, and Elton Yechao Zhu. Quantum state preparation of normal distributions using matrix product states. *npj Quantum Information*, 10:15, 2024. doi: 10.1038/s41534-024-00805-0.

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. *Physics Letters A*, 516:129642, 2024. doi: 10.1016/j.physleta.2024.129642.

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.

Florian J. Kiwit, Bernhard Jobst, Andre Luckow, Frank Pollmann, and Carlos A. Riofrío. Typical machine learning datasets as low-depth quantum circuits. *Quantum Science and Technology*, 10 (4):045035, 2025. doi: 10.1088/2058-9565/ae0123.

Martín Larocca, Supanut Thanasilp, Samson Wang, Kunal Sharma, Jacob Biamonte, Patrick J. Coles, Lukasz Cincio, Jarrod R. McClean, Zoë Holmes, and M. Cerezo. A review of barren plateaus in variational quantum computing. *arXiv preprint arXiv:2405.00781*, 2024.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.

Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3730–3738, 2015. doi: 10.1109/ICCV.2015.425.

Guang Hao Low and Yuan Su. Quantum linear system algorithm with optimal queries to initial state preparation. *Quantum*, 10:2041, 2026. doi: 10.22331/q-2026-03-23-2041.

Stéphane Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 3 edition, 2009. ISBN 9780123743701.

Daniel Malz, Georgios Styliaris, Zhi-Yuan Wei, and J. Ignacio Cirac. Preparation of matrix product states with log-depth quantum circuits. *Physical Review Letters*, 132(4):040404, 2024. doi: 10.1103/PhysRevLett.132.040404.

Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):4812, 2018. doi: 10.1038/s41467-018-07090-4.

Naoki Mitsuda, Tatsuhiro Ichimura, Kouhei Nakaji, Yohichi Suzuki, Tomoki Tanaka, Rudy Raymond, Hiroyuki Tezuka, Tamiya Onodera, and Naoki Yamamoto. Approximate complex amplitude encoding algorithm and its application to data classification problems. *Physical Review A*, 109(5):052423, 2024.

George B. Moody and Roger G. Mark. The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, 2001. doi: 10.1109/51.932724.

S. Mostafa Mousavi, Yixiao Sheng, Weiqiang Zhu, and Gregory C. Beroza. STanford EArthquake Dataset (STEAD): A global data set of seismic signals for AI. *IEEE Access*, 7:179464–179476, 2019. doi: 10.1109/ACCESS.2019.2947848.

Kouhei Nakaji, Satoko 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(2):023136, 2022. doi: 10.1103/PhysRevResearch.4.023136.

Sameer A. Nene, Shree K. Nayar, and Hiroshi Murase. Columbia object image library (COIL-20). Technical Report CUCS-005-96, Department of Computer Science, Columbia University, 1996. URL https://cave.cs.columbia.edu/repository/COIL-20.

Román Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics*, 349:117–158, 2014. doi: 10.1016/j.aop.2014.06.013.

Ivan V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011. doi: 10.1137/090752286.

David Peral-García, Juan Cruz-Benito, and Francisco José García-Peñalvo. Systematic literature review: Quantum machine learning and its applications. *Computer Science Review*, 51:100619, 2024. doi: 10.1016/j.cosrev.2024.100619.

Martin Plesch and Časlav Brukner. Quantum-state preparation with universal gate decompositions. *Physical Review A*, 83(3):032302, 2011. doi: 10.1103/PhysRevA.83.032302.

Sadia Afrin Purba, M. A. Al Mamun, Iyad Obeid, and Joseph Picone. No quantum advantage without structure: A controlled benchmark of quantum machine learning on classical data. In *2026 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, December 2026. To appear.

Shi-Ju Ran. Encoding of matrix product states into quantum circuits of one- and two-qubit gates. *Physical Review A*, 101:032310, 2020. doi: 10.1103/PhysRevA.101.032310.

Manuel S. Rudolph, Jing Chen, Jacob Miller, Atithi Acharya, and Alejandro Perdomo-Ortiz. Decomposition of matrix product states into shallow quantum circuits. *Quantum Science and Technology*, 9(1):015012, 2024. doi: 10.1088/2058-9565/ad04e6.

Ulrich Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics*, 326(1):96–192, 2011. doi: 10.1016/j.aop.2010.09.012.

C. Schön, E. Solano, F. Verstraete, J. I. Cirac, and M. M. Wolf. Sequential generation of entangled multiqubit states. *Physical Review Letters*, 95:110503, 2005. doi: 10.1103/PhysRevLett.95.110503.

Tomonori Shirakawa, Hiroshi Ueda, and Seiji Yunoki. Automatic quantum circuit encoding of a given arbitrary quantum state. *Physical Review Research*, 6(4):043008, 2024. doi: 10.1103/PhysRevResearch.6.043008.

Anzheng Tang, Jun-Bo Wang, Yijin Pan, Wence Zhang, Yijian Chen, Hongkang Yu, and Rodrigo C. de Lamare. Spatially non-stationary XL-MIMO channel estimation: A three-layer generalized approximate message passing method. *IEEE Transactions on Signal Processing*, 73:356–371, 2025. doi: 10.1109/TSP.2024.3512575.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y. Zhang, and Jun Zhou. TimeMixer: Decomposable multiscale mixing for time series forecasting. In *International Conference on Learning Representations*, 2024.

Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C. Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 23(2):684–695, February 2014. doi: 10.1109/TIP.2013.2293423.

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

Xiao-Ming Zhang and Xiao Yuan. Circuit complexity of quantum access models for encoding classical data. *npj Quantum Information*, 10:42, 2024. doi: 10.1038/s41534-024-00835-8.

Jin Zheng, Qing Gao, and Maciej Ogorzałek. Approximate quantum amplitude encoding with parameterized quantum circuits. In *2025 IEEE International Conference on Quantum Control, Computing and Learning (qCCL)*, pp. 155–160. IEEE, 2025. doi: 10.1109/QCCL65142.2025.11159050.

Jin Zheng, Qing Gao, Yi Du, Jinhu Lü, and Maciej Ogorzałek. Design of an approximate quantum encoding model based on quantum circuits. *Automatica*, 190:113052, 2026. doi: 10.1016/j.automatica.2026.113052.

A Additional preliminaries

MPS construction. An n-qubit system is described in the Hilbert space $(\mathbb{C}^2)^{\otimes n}$ with computational basis $\{|j\rangle\}_{j=0}^{N-1}$, where $N = 2^n$. For the binary representation $j = (b_1 \cdots b_n)_2$, we identify $|j\rangle \equiv |b_1 \cdots b_n\rangle = |b_1\rangle \otimes \cdots \otimes |b_n\rangle$, with $b_i \in \{0, 1\}$. Given a coefficient vector $\boldsymbol{z} \in \mathbb{C}^N$, a bijective index mapping $\pi : \{0, \ldots, N-1\} \to \{0, 1\}^n$ assigns each coefficient to a computational-basis string and induces the mapped state $|\psi(\boldsymbol{z}, \pi)\rangle = U_\pi |\psi(\boldsymbol{z})\rangle$, where $U_\pi |j\rangle = |\pi(j)\rangle$. Since U_π is unitary, applying the same mapping preserves inner products and state fidelity. The mapped coefficients define an order-n tensor $C_\pi(\boldsymbol{z}) \in \mathbb{C}^{2\times\cdots\times 2}$, whose unfolding across the bipartition $1{:}i \,|\, i{+}1{:}n$ is denoted by $Z_\pi^{(i)}(\boldsymbol{z}) \in \mathbb{C}^{2^i \times 2^{n-i}}$. Tensor-train singular value decomposition (TT-SVD) constructs the MPS sequentially from these unfoldings. Retaining all nonzero singular values yields an exact MPS, whereas truncation reduces the realized bond dimensions at the cost of approximation error. Two truncation settings are commonly used. A sufficiently small singular-value threshold removes only numerically negligible components; in our implementation, $\tau_{\mathrm{svd}} = 10^{-12}$ is used solely to discard machine-precision singular values, yielding an effectively exact MPS up to numerical precision. Alternatively, imposing a maximum bond dimension $\chi_{\max}$ retains only the leading singular components whenever the local rank exceeds this cap. Such resource-driven truncation does not directly enforce a prescribed state fidelity, since the resulting error depends on the local singular spectra, and the achieved fidelity must therefore be evaluated after MPS construction.

MPS complexity analysis. For the tensor $C_\pi(\boldsymbol{z})$, the minimum exact bond dimension across cut i is $\chi_i(\boldsymbol{z}, \pi) = \mathrm{rank}(Z_\pi^{(i)}(\boldsymbol{z}))$, equivalently the Schmidt rank of $|\psi(\boldsymbol{z}, \pi)\rangle$ across the same bipartition (Oseledets, 2011). The corresponding bond profile is $\boldsymbol{\chi}(\boldsymbol{z}, \pi) = (\chi_1(\boldsymbol{z}, \pi), \ldots, \chi_{n-1}(\boldsymbol{z}, \pi))$, with $\chi_0(\boldsymbol{z}, \pi) = \chi_n(\boldsymbol{z}, \pi) = 1$. When only coefficients indexed by a support Ω are retained, we use the shorthand $\chi_i(\Omega, \pi) := \chi_i(\boldsymbol{z}_\Omega, \pi)$ and $\boldsymbol{\chi}(\Omega, \pi) := \boldsymbol{\chi}(\boldsymbol{z}_\Omega, \pi)$. For a realized bond profile $\boldsymbol{\chi} = (\chi_1, \ldots, \chi_{n-1})$, we quantify the raw MPS representation complexity by $\mathcal{C}_{\mathrm{MPS}}(\boldsymbol{\chi}) = \sum_{i=1}^{n} 2\chi_{i-1}\chi_i$, which counts the total number of entries in the MPS tensors. For a coefficient vector with support size $K_{\mathrm{z}} = \|\boldsymbol{z}\|_0$, sparsity gives the mapping-independent bound $\chi_i(\boldsymbol{z}, \pi) \le \overline{\chi}_i = \min\{K_{\mathrm{z}}, 2^i, 2^{n-i}\}$, with $\overline{\chi}_0 = \overline{\chi}_n = 1$. Consequently, $\mathcal{C}_{\mathrm{MPS}}(\boldsymbol{\chi}(\boldsymbol{z}, \pi)) \le \sum_{i=1}^{n} 2\overline{\chi}_{i-1}\overline{\chi}_i = \mathcal{O}(nK_{\mathrm{z}}^2)$. This bound provides only a sparsity-induced envelope: supports with the same cardinality can yield different unfolding ranks under different coefficient arrangements. Hence, sparsity limits the possible bond demand, whereas support geometry and index mapping jointly determine the realized bond profile.

Sequential state preparation. An MPS in canonical form can be interpreted as a sequence of local isometries and systematically converted into a sequential state-preparation circuit (Schön et al., 2005). The i-th MPS tensor has size $2\chi_{i-1}\chi_i$, so the bond profile directly determines the dimensions of the local isometries involved in state preparation. Under a uniform bound $\chi_i \le \chi_{\max}$, both MPS storage and the size of the sequential isometric representation scale as $\mathcal{O}(n\chi_{\max}^2)$. Reducing the realized bond dimensions therefore leads to more compact state-preparation constructions. We emphasize, however, that $\mathcal{C}_{\mathrm{MPS}}$ measures representation size rather than the exact compiled two-qubit gate count, which additionally depends on the chosen isometry decomposition and circuit-synthesis procedure.

B PROOFS OF THEORETICAL RESULTS

B.1 PROOF OF PROPOSITION 1

Proof. Since π is bijective, $U_\pi^\dagger U_\pi = I$, and hence

$$F(|\psi_{\Omega,\pi}\rangle, U_\pi |\psi(\boldsymbol{z})\rangle) = F(|\psi(\boldsymbol{z}_\Omega)\rangle, |\psi(\boldsymbol{z})\rangle). \tag{8}$$

By the definition of $\boldsymbol{z}_\Omega$,

$$\langle\psi(\boldsymbol{z}_\Omega)| |\psi(\boldsymbol{z})\rangle = \frac{\sum_{j\in\Omega} |z_j|^2}{\|\boldsymbol{z}_\Omega\|_2 \|\boldsymbol{z}\|_2} = \frac{\|\boldsymbol{z}_\Omega\|_2}{\|\boldsymbol{z}\|_2}, \tag{9}$$

$$\|\boldsymbol{z}\|_2^2 = \|\boldsymbol{z}_\Omega\|_2^2 + \|\boldsymbol{z} - \boldsymbol{z}_\Omega\|_2^2. \tag{10}$$

Together, these identities establish Eqn. (3).

For normalized pure states, let $P(|\phi\rangle, |\varphi\rangle) := \sqrt{1 - F(|\phi\rangle, |\varphi\rangle)}$. Its triangle inequality and $F(\Omega) \geq 1 - \varepsilon$ give

$$\begin{aligned} P\Big(\Big|\widetilde{\psi}\Big\rangle, U_\pi |\psi(\boldsymbol{z})\rangle\Big) &\leq P\Big(\Big|\widetilde{\psi}\Big\rangle, |\psi_{\Omega,\pi}\rangle\Big) + P(|\psi_{\Omega,\pi}\rangle, U_\pi |\psi(\boldsymbol{z})\rangle) \\ &\leq \sqrt{\eta} + \sqrt{\varepsilon}. \end{aligned} \tag{11}$$

Squaring yields Eqn. (4).

Finally, $\eta = 0$ implies $\Big|\widetilde{\psi}\Big\rangle = e^{\mathrm{i}\theta} |\psi_{\Omega,\pi}\rangle$ for some $\theta \in \mathbb{R}$, so

$$F\Big(\Big|\widetilde{\psi}\Big\rangle, U_\pi |\psi(\boldsymbol{z})\rangle\Big) = F(\Omega). \tag{12}$$

□

B.2 PROOF OF PROPOSITION 2

Proof. Fix $\Omega \in \mathcal{S}(\mathcal{T})$ with $\|\boldsymbol{z}_\Omega\|_2 > 0$, a valid mapping π, and $1 \leq i < n$. Let $Z_{\Omega,\pi}^{(i)} \in \mathbb{C}^{2^i \times 2^{n-i}}$ denote the unfolding of the unnormalized mapped coefficient vector. By the definition of $S_{\Omega,\pi}$, $[Z_{\Omega,\pi}^{(i)}]_{p,q} = 0$ whenever $p \notin \mathcal{P}_i(\Omega,\pi)$ or $q \notin \mathcal{Q}_i(\Omega,\pi)$. The Schmidt-rank characterization gives

$$\begin{aligned} \chi_i(\Omega,\pi) &= \operatorname{rank}\Big(Z_{\Omega,\pi}^{(i)}\Big) \\ &\leq \min\big\{|\mathcal{P}_i(\Omega,\pi)|,\, |\mathcal{Q}_i(\Omega,\pi)|,\, 2^i,\, 2^{n-i}\big\} = \Pi_i(\Omega,\pi). \end{aligned} \tag{13}$$

For fixed (Ω,π), suppress this dependence and define $d_i := \min\{2^i, 2^{n-i}\}$ and $g_i := \Pi_i - \chi_i$. Let $G_i = (\mathcal{P}_i, \mathcal{Q}_i, E_i)$ be the bipartite graph with $E_i := \{(p,q) : pq \in S_{\Omega,\pi}\}$, and let ν_i denote its maximum matching size. A nonzero determinant of an $\ell \times \ell$ minor contains a nonzero permutation term, which identifies a matching of size ℓ. Hence,

$$\chi_i \leq \nu_i \leq \min\{|\mathcal{P}_i|, |\mathcal{Q}_i|\} = \Pi_i \leq d_i. \tag{14}$$

Therefore,

$$\begin{aligned} g_i &= (\Pi_i - \nu_i) + (\nu_i - \chi_i), \\ \Pi_i - \nu_i \geq 0, &\qquad \nu_i - \chi_i \geq 0. \end{aligned} \tag{15}$$

Since $\|\boldsymbol{z}_\Omega\|_2 > 0$, every unfolding is nonzero, so $\chi_i \geq 1$ and

$$0 \leq g_i \leq \Pi_i - 1 \leq d_i - 1. \tag{16}$$

For $\Omega = V$ and $z_j = 1$ for all $j \in V$,

$$\begin{aligned} Z^{(i)} &= \mathbf{1}_{2^i} \mathbf{1}_{2^{n-i}}^\top, \\ \chi_i = 1, &\qquad \nu_i = \Pi_i = d_i, \qquad g_i = d_i - 1. \end{aligned} \tag{17}$$

Thus, Eqn. (16) is sharp.

By the minor characterization of matrix rank, $g_i = 0$ holds if and only if there exist $R_i \subseteq \mathcal{P}_i$ and $C_i \subseteq \mathcal{Q}_i$ such that

$$|R_i| = |C_i| = \Pi_i, \qquad \det Z^{(i)}[R_i, C_i] \neq 0. \tag{18}$$

In particular, suppose that $G_i[R_i, C_i]$ has a unique perfect matching $\mathcal{L}_i$, with $|R_i| = |C_i| = \Pi_i$, and that $[Z^{(i)}]_{p,q} \neq 0$ for every $(p, q) \in \mathcal{L}_i$. Then

$$\det Z^{(i)}[R_i, C_i] = \pm \prod_{(p,q)\in\mathcal{L}_i} [Z^{(i)}]_{p,q} \neq 0, \tag{19}$$

which yields $\chi_i = \Pi_i$.

More generally, choose a maximum matching $\mathcal{L}_i$ with endpoint sets R_i, C_i, where $|R_i| = |C_i| = \nu_i$. For $\boldsymbol{c} = (c_j)_{j\in\Omega} \in \mathbb{C}^{|\Omega|}$, define $Z^{(i)}(\boldsymbol{c})$ by $[Z^{(i)}(\boldsymbol{c})]_{p,q} = c_{\pi^{-1}(pq)}$ on E_i and zero elsewhere, and set $\Delta_i(\boldsymbol{c}) := \det Z^{(i)}(\boldsymbol{c})[R_i, C_i]$. Taking $c_{\pi^{-1}(pq)} = 1$ on $\mathcal{L}_i$ and $c_{\pi^{-1}(pq)} = 0$ on $E_i \setminus \mathcal{L}_i$ gives $\Delta_i(\boldsymbol{c}) = \pm 1$. Thus $\Delta_i \not\equiv 0$, and

$$\mathcal{A}_i := \{\boldsymbol{c} \in \mathbb{C}^{|\Omega|} : \Delta_i(\boldsymbol{c}) = 0\} \subsetneq \mathbb{C}^{|\Omega|}. \tag{20}$$

For $(z_j)_{j\in\Omega} \notin \mathcal{A}_i$, the selected minor is nonsingular. Together with Eqn. (14), this gives

$$\chi_i = \nu_i, \qquad g_i = \Pi_i - \nu_i. \tag{21}$$

Consequently, outside $\mathcal{A}_i$, $\chi_i = \Pi_i$ holds if and only if $\nu_i = \Pi_i$. □

B.3 Feasibility and attainment of the joint optimization

Assume $\boldsymbol{z} \neq 0$ and that the finite structure-preserving mapping family $\mathcal{M}$ is nonempty. For any $0 \leq \varepsilon < 1$, the optimization in Eqn. (6) has a nonempty feasible set and attains a minimizer.

Proof. Define $\mathcal{F}_\varepsilon := \{(\Omega, \pi) \in \mathcal{S}(\mathcal{T}) \times \mathcal{M} : F(\Omega) \geq 1 - \varepsilon\}$. The full support V belongs to $\mathcal{S}(\mathcal{T})$, and Proposition 1 gives $F(V) = 1$. Thus $(V, \pi) \in \mathcal{F}_\varepsilon$ for every $\pi \in \mathcal{M}$, and $1 \leq |\mathcal{F}_\varepsilon| \leq 2^N|\mathcal{M}| < \infty$.

Let $J(\Omega, \pi) := \eta_K|\Omega|/N + \eta_\Pi \mathcal{C}_\Pi(\Omega, \pi)$. Since $0 \leq \Pi_i(\Omega, \pi) \leq \min\{2^i, 2^{n-i}\}$,

$$0 \leq J(\Omega, \pi) \leq \eta_K + \eta_\Pi \sum_{i=1}^{n-1} w_i < \infty. \tag{22}$$

Therefore, J attains a minimum on $\mathcal{F}_\varepsilon$.

For $1 \leq K \leq N$, define $\mathcal{S}_K(\mathcal{T}) := \{\Omega \in \mathcal{S}(\mathcal{T}) : |\Omega| \leq K\}$ and $e_{\mathrm{tree}}(K) := \min_{\Omega\in\mathcal{S}_K(\mathcal{T})} \|\boldsymbol{z} - \boldsymbol{z}_\Omega\|_2^2/\|\boldsymbol{z}\|_2^2$. If $e_{\mathrm{tree}}(K) \leq \varepsilon$, a minimizing support Ω_K satisfies $|\Omega_K| \leq K$ and $F(\Omega_K) = 1 - e_{\mathrm{tree}}(K) \geq 1 - \varepsilon$. Hence $(\Omega_K, \pi) \in \mathcal{F}_\varepsilon$ for every $\pi \in \mathcal{M}$.

For the remaining comparisons, write $\xi = (\Omega, \pi)$ and take all minima over the same feasible set $\mathcal{F}_\varepsilon$. Let $f(\xi) := \mathcal{C}_{\mathrm{MPS}}(\boldsymbol{\chi}(\xi))$, $d_i := \min\{2^i, 2^{n-i}\}$ for $0 \leq i \leq n$, and $\chi_0 = \chi_n = \Pi_0 = \Pi_n = 1$. Define

$$\mathcal{U}_\Pi(\xi) := 2\sum_{i=1}^{n} \Pi_{i-1}(\xi)\Pi_i(\xi), \qquad C_{\max} := 2\sum_{i=1}^{n} d_{i-1}d_i. \tag{23}$$

Proposition 2 gives

$$2n \leq f(\xi) = 2\sum_{i=1}^{n} \chi_{i-1}(\xi)\chi_i(\xi) \leq \mathcal{U}_\Pi(\xi) \leq C_{\max}. \tag{24}$$

Let $g_i(\xi) := \Pi_i(\xi) - \chi_i(\xi)$, with $g_0 = g_n = 0$. Suppressing the dependence on ξ inside the summation,

$$\begin{aligned} \Gamma(\xi) &:= \mathcal{U}_\Pi(\xi) - f(\xi) \\ &= 2\sum_{i=1}^{n} \left(\chi_{i-1}g_i + \chi_i g_{i-1} + g_{i-1}g_i\right). \end{aligned} \tag{25}$$

Consequently,

$$0 \le \Gamma(\xi) \le C_{\max} - 2n. \tag{26}$$

The upper bound is attained for $\Omega = V$ and $z_j = 1$ for all $j \in V$, for which $\chi_i = 1$ and $\Pi_i = d_i$. Since every $\chi_i \ge 1$, Eqn. (25) also shows that $\Gamma(\xi) = 0$ holds if and only if $g_i(\xi) = 0$ for every $1 \le i < n$.

Let $\xi^\star \in \arg\min_{\xi \in \mathcal{F}_\varepsilon} f(\xi)$ and $\xi_U \in \arg\min_{\xi \in \mathcal{F}_\varepsilon} \mathcal{U}_\Pi(\xi)$. By $\mathcal{U}_\Pi(\xi_U) \le \mathcal{U}_\Pi(\xi^\star)$,

$$\begin{aligned} 0 &\le f(\xi_U) - f(\xi^\star) \\ &\le \Gamma(\xi^\star) - \Gamma(\xi_U) \\ &\le \Gamma(\xi^\star) \le C_{\max} - 2n. \end{aligned} \tag{27}$$

The objective J in Eqn. (6) uses a linear prefix aggregate rather than $\mathcal{U}_\Pi$. For any $a > 0$, define

$$\begin{aligned} \mathcal{E}_a(\xi) &:= J(\xi) - a\mathcal{U}_\Pi(\xi) \\ &= \eta_K \frac{|\Omega|}{N} + \eta_\Pi \sum_{i=1}^{n-1} w_i \frac{\Pi_i(\xi)}{d_i} - 2a \sum_{i=1}^{n} \Pi_{i-1}(\xi)\Pi_i(\xi), \end{aligned} \tag{28}$$

and $\omega_a := \max_{\xi \in \mathcal{F}_\varepsilon} \mathcal{E}_a(\xi) - \min_{\xi \in \mathcal{F}_\varepsilon} \mathcal{E}_a(\xi)$. Then

$$J(\xi) = af(\xi) + a\Gamma(\xi) + \mathcal{E}_a(\xi). \tag{29}$$

Let $\widehat{\xi} \in \arg\min_{\xi \in \mathcal{F}_\varepsilon} J(\xi)$. Using $J(\widehat{\xi}) \le J(\xi^\star)$ in Eqn. (29),

$$\begin{aligned} 0 &\le f(\widehat{\xi}) - f(\xi^\star) \\ &\le \Gamma(\xi^\star) - \Gamma(\widehat{\xi}) + \frac{\mathcal{E}_a(\xi^\star) - \mathcal{E}_a(\widehat{\xi})}{a} \\ &\le \Gamma(\xi^\star) + \frac{\omega_a}{a}. \end{aligned} \tag{30}$$

Together with Eqn. (24),

$$0 \le f(\widehat{\xi}) - f(\xi^\star) \le \min\left\{ C_{\max} - 2n,\, \Gamma(\xi^\star) + \frac{\omega_a}{a} \right\}. \tag{31}$$

For a feasible approximate minimizer with objective suboptimality tolerance $\delta_{\text{opt}} \ge 0$ satisfying $J(\widehat{\xi}) \le \min_{\xi \in \mathcal{F}_\varepsilon} J(\xi) + \delta_{\text{opt}}$, the second term in this bound becomes $\Gamma(\xi^\star) + (\omega_a + \delta_{\text{opt}})/a$. This objective-gap tolerance is distinct from the update-acceptance tolerance δ_J used by the practical solver.

If $\Pi_i(\xi) = \chi_i(\xi)$ for every $\xi \in \mathcal{F}_\varepsilon$ and every $1 \le i < n$, then $\Gamma = 0$ throughout $\mathcal{F}_\varepsilon$, and

$$\arg\min_{\xi \in \mathcal{F}_\varepsilon} \mathcal{U}_\Pi(\xi) = \arg\min_{\xi \in \mathcal{F}_\varepsilon} f(\xi). \tag{32}$$

For the objective J, additionally suppose that $\mathcal{E}_a(\xi) = b$ for all $\xi \in \mathcal{F}_\varepsilon$, for some $a > 0$ and $b \in \mathbb{R}$. Then $J(\xi) = af(\xi) + b$, and

$$\arg\min_{\xi \in \mathcal{F}_\varepsilon} J(\xi) = \arg\min_{\xi \in \mathcal{F}_\varepsilon} f(\xi). \tag{33}$$

The condition on $\mathcal{E}_a$ is additional: $\Pi_i = \chi_i$ alone does not identify the linear, sparsity-regularized objective J with the parameter-count objective f. □

B.4 Kernel-target alignment stability

Proposition. Consider m samples with normalized wavelet-domain amplitude states $\{|\psi(\boldsymbol{z}_a)\rangle\}_{a=1}^m$, and apply the same mapping π to all samples. Define $|\phi_a\rangle = U_\pi |\psi(\boldsymbol{z}_a)\rangle$, and let $\left|\widehat{\psi}_a\right\rangle$ denote the corresponding normalized QuSAE realization with $\delta_a = 1 - F(\left|\widehat{\psi}_a\right\rangle, |\phi_a\rangle)$. Let $\mathbf{K}$ and $\widehat{\mathbf{K}}$ be the fidelity-kernel matrices, with $\mathbf{K}_{ab} = |\langle\phi_a|\phi_b\rangle|^2$ and $\widehat{\mathbf{K}}_{ab} = |\langle\widehat{\psi}_a|\widehat{\psi}_b\rangle|^2$. Then

$$|\mathbf{K}_{ab} - \widehat{\mathbf{K}}_{ab}| \le \sqrt{\delta_a} + \sqrt{\delta_b}, \qquad \|\mathbf{K} - \widehat{\mathbf{K}}\|_F \le 2\sqrt{m \sum_{a=1}^{m} \delta_a}. \tag{34}$$

Let $H = I_m - \frac{1}{m}\mathbf{1}\mathbf{1}^\top$, $\mathbf{K}_c = H\mathbf{K}H$, $\widehat{\mathbf{K}}_c = H\widehat{\mathbf{K}}H$, and $Y_c = HYH$ for a target matrix Y with $\|Y_c\|_F > 0$. Define the centered kernel-target alignment (KTA) $\mathrm{KTA}_c(\mathbf{K}, Y) = \langle \mathbf{K}_c, Y_c\rangle_F / (\|\mathbf{K}_c\|_F \|Y_c\|_F)$, and $\kappa_c = \min\{\|\mathbf{K}_c\|_F, \|\widehat{\mathbf{K}}_c\|_F\} > 0$. Then

$$\left|\mathrm{KTA}_c(\widehat{\mathbf{K}}, Y) - \mathrm{KTA}_c(\mathbf{K}, Y)\right| \le \frac{4\sqrt{m\sum_{a=1}^m \delta_a}}{\kappa_c}. \tag{35}$$

Proof. Write $\rho_a = |\phi_a\rangle\langle\phi_a|$ and $\widehat{\rho}_a = \left|\widehat{\psi}_a\right\rangle\left\langle\widehat{\psi}_a\right|$. Since $\mathbf{K}_{ab} = \mathrm{Tr}(\rho_a\rho_b)$ and $\widehat{\mathbf{K}}_{ab} = \mathrm{Tr}(\widehat{\rho}_a\widehat{\rho}_b)$,

$$\begin{aligned} |\mathbf{K}_{ab} - \widehat{\mathbf{K}}_{ab}| &\le |\mathrm{Tr}[(\rho_a - \widehat{\rho}_a)\rho_b]| + |\mathrm{Tr}[\widehat{\rho}_a(\rho_b - \widehat{\rho}_b)]| \\ &\le D_{\mathrm{tr}}(\rho_a, \widehat{\rho}_a) + D_{\mathrm{tr}}(\rho_b, \widehat{\rho}_b) = \sqrt{\delta_a} + \sqrt{\delta_b}, \end{aligned} \tag{36}$$

where $D_{\mathrm{tr}}(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1$ is the trace distance, with $\|\cdot\|_1$ denoting the trace norm. The final equality uses the pure-state identity $D_{\mathrm{tr}}(|\phi\rangle\langle\phi|, |\varphi\rangle\langle\varphi|) = \sqrt{1 - F(|\phi\rangle, |\varphi\rangle)}$. Therefore,

$$\|\mathbf{K} - \widehat{\mathbf{K}}\|_F^2 \le \sum_{a,b=1}^m (\sqrt{\delta_a} + \sqrt{\delta_b})^2 \tag{37}$$

$$\le 4m\sum_{a=1}^m \delta_a, \tag{38}$$

which proves the second inequality in Eqn. (34).

Because H is an orthogonal projector, $\|H\|_2 = 1$, so $\|\mathbf{K}_c - \widehat{\mathbf{K}}_c\|_F \le \|\mathbf{K} - \widehat{\mathbf{K}}\|_F$. For nonzero matrices X_1 and X_2, $\|X_1/\|X_1\|_F - X_2/\|X_2\|_F\|_F \le 2\|X_1 - X_2\|_F / \min\{\|X_1\|_F, \|X_2\|_F\}$. Applying this inequality to $\mathbf{K}_c$ and $\widehat{\mathbf{K}}_c$, followed by Cauchy–Schwarz against $Y_c/\|Y_c\|_F$, yields

$$\left|\mathrm{KTA}_c(\widehat{\mathbf{K}}, Y) - \mathrm{KTA}_c(\mathbf{K}, Y)\right| \le \frac{2\|\widehat{\mathbf{K}}_c - \mathbf{K}_c\|_F}{\kappa_c} \le \frac{4\sqrt{m\sum_a \delta_a}}{\kappa_c}. \tag{39}$$

□

Corollary. If sample a satisfies $F(\Omega_a) \ge 1 - \varepsilon_a$ and has TT-SVD realization infidelity η_a, Proposition 1 gives $\delta_a \le (\sqrt{\varepsilon_a} + \sqrt{\eta_a})^2$. Hence

$$\left|\mathrm{KTA}_c(\widehat{\mathbf{K}}, Y) - \mathrm{KTA}_c(\mathbf{K}, Y)\right| \le \frac{4}{\kappa_c}\sqrt{m\sum_{a=1}^m (\sqrt{\varepsilon_a} + \sqrt{\eta_a})^2}. \tag{40}$$

If $\varepsilon_a \le \varepsilon$ and $\eta_a \le \eta$ for all a, the bound reduces to $4m(\sqrt{\varepsilon} + \sqrt{\eta})/\kappa_c$; for exact TT-SVD it becomes $4m\sqrt{\varepsilon}/\kappa_c$.

C Implementation details

C.1 Dataset preprocessing

We evaluate QuSAE on three image datasets, MNIST, COIL-20, and CelebA, and two one-dimensional signal datasets, MIT-BIH and STEAD. MNIST is resized from 28×28 to 32×32, while COIL-20 and CelebA use 128×128 and 64×64 images, respectively. MIT-BIH and STEAD are represented by fixed-length segments of 256 and 2048 samples. These dimensions correspond to 10, 14, 12, 8, and 11 qubits for MNIST, COIL-20, CelebA, MIT-BIH, and STEAD, respectively. All samples are converted to real-valued vectors and ℓ_2-normalized before the wavelet transform. For each class, we uniformly sample 100 examples according to the task labels used in the experiments, or use all available samples when fewer than 100 are available.

Before the main experiments, we evaluate four orthogonal wavelet families, Haar, db2, db4, and sym4, together with four representative structure-preserving mappings. For each wavelet–mapping

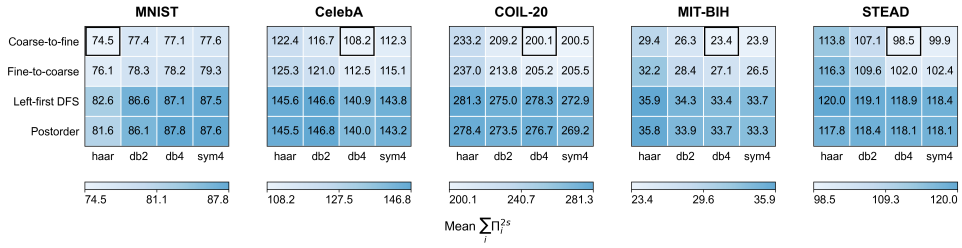

Figure 6: **Preliminary sensitivity analysis over wavelet families and representative mappings.** Columns correspond to Haar, db2, db4, and sym4, while rows correspond to coarse-to-fine, fine-to-coarse, left-first DFS, and postorder mappings. Each cell reports the mean prefix-complexity score under matched support-error tolerances, with lower values indicating more favorable prefix organization. Black boxes mark the best score for each dataset.

Table 1: **Dataset and wavelet configurations.** The selected wavelet family is fixed for all subsequent experiments on each dataset.

Dataset	Input size	Qubits	Wavelet
MNIST	32×32	10	Haar
COIL-20	128×128	14	db4
CelebA	64×64	12	db4
MIT-BIH	256	8	db4
STEAD	2048	11	db4

pair, we compute the mean prefix-complexity sum as the selection score under matched support-error tolerances. Lower values indicate more favorable prefix organization and thus lower bond demand. This preliminary analysis is used only to select the dataset-specific wavelet transform, while the candidate mappings remain subject to optimization in the subsequent joint support–mapping optimization.

As shown in Fig. 6, Haar gives the lowest prefix-complexity score for MNIST, whereas db4 is preferred for COIL-20, CelebA, MIT-BIH, and STEAD. These wavelet families are fixed in the subsequent experiments.

C.2 Wavelet-tree construction

For each input $\boldsymbol{x}$, an orthogonal DWT produces $\boldsymbol{z} = W\boldsymbol{x}$. All transforms use periodization boundary handling and a full dyadic decomposition. For one-dimensional inputs, the coarsest approximation coefficient forms the root and the detail coefficients are organized into a binary coarse-to-fine hierarchy. For two-dimensional inputs, the coarsest approximation coefficient is shared by the horizontal, vertical, and diagonal detail branches, each forming an orientation-specific quadtree across scales. Within each scale and sibling set, the implementation uses a fixed deterministic coefficient ordering. These relations define $\mathcal{T} = (V, E, r)$, while admissible structured supports belong to $\mathcal{S}(\mathcal{T})$.

Given a prescribed infidelity tolerance ε, the fidelity-controlled tree stage solves

$$\Omega_\varepsilon \in \operatorname*{arg\,min}_{\Omega \in \mathcal{S}(\mathcal{T})} |\Omega| \quad \text{subject to} \quad F(\Omega) \geq 1 - \varepsilon, \tag{41}$$

with minimum budget $K_\varepsilon = |\Omega_\varepsilon|$. The optimization is solved exactly by a rooted tree-knapsack dynamic program. Let $e_j = |z_j|^2$, and let $D_j(k)$ denote the maximum retained energy obtainable by selecting exactly k nodes from the subtree rooted at j, conditioned on retaining j. For children $c \in \mathrm{ch}(j)$,

$$D_j(k) = e_j + \max_{\substack{\sum_{c \in \mathrm{ch}(j)} k_c = k-1 \\ k_c \geq 0}} \sum_{c \in \mathrm{ch}(j)} \widetilde{D}_c(k_c), \tag{42}$$

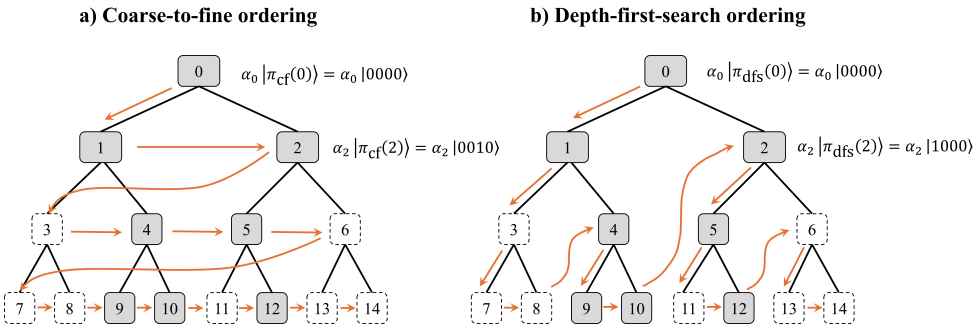

Figure 7: Representative mapping families. The integer labels denote canonical wavelet-tree node IDs. The natural wavelet ordering preserves the deterministic flattened coefficient layout of the transform, while the coarse-to-fine ordering stably sorts the same coefficient nodes by scale before location. These mappings illustrate how wavelet-tree structure can be transferred to the MPS prefix tree.

where $\widetilde{D}_c(0) = 0$, $\widetilde{D}_c(k) = D_c(k)$ for $k \geq 1$, and infeasible states are assigned $-\infty$. For a leaf j, $D_j(1) = e_j$. The tables are evaluated bottom-up with budgets restricted by subtree size. The root table $D_r(k)$ gives the maximum retained energy at each cardinality, and the smallest k satisfying $D_r(k) \geq (1-\varepsilon)\|\boldsymbol{z}\|_2^2$ is K_ε. Backtracking through the maximizing child allocations recovers the corresponding support. For the bounded-degree wavelet trees considered here, sequential child-table merging over the complete budget range requires $\mathcal{O}(N^2)$ time.

The exact dynamic program is used for the fidelity-only tree-support problem in Eqn. (41). Once prefix complexity is included in the joint objective, the support update is no longer separable into the same tree-knapsack recurrence and is handled by the structured heuristic described in Appendix C.4.

C.3 Structure-preserving mappings

QuSAE restricts the mapping search to a finite family $\mathcal{M}$ of structure-preserving mappings. Each $\pi_\tau \in \mathcal{M}$ is induced by a complete node ordering $\tau = (v_0, \ldots, v_{N-1})$ through $\pi_\tau(v_j) = \mathrm{bin}_n(j)$, where $\mathrm{bin}_n(j)$ is the n-bit representation of j. The ordering is derived from the hierarchy of $\mathcal{T}$, and deterministic spatial and sibling tie-breaking is fixed across all samples.

The mapping candidate family used in the experiments contains coarse-to-fine ordering, fine-to-coarse ordering, left-first depth-first traversal, and postorder traversal. Fig. 7 illustrates two representative coefficient layouts. Coarse-to-fine and fine-to-coarse candidates order nodes primarily by tree depth in ascending and descending order, respectively, while the traversal-based candidates follow the parent–child connectivity of $\mathcal{T}$. For two-dimensional inputs, the same construction is applied to the three orientation-specific branches sharing the common root. These candidates define a finite search space that preserves the wavelet hierarchy while allowing different prefix organizations of the retained support.

C.4 Joint support–mapping solver

The practical solver performs a finite, structure-restricted search over ancestor-closed supports and the mapping family $\mathcal{M}$. For sample m, we write $F(\Omega; \boldsymbol{z}^{(m)}) = \|\boldsymbol{z}_\Omega^{(m)}\|_2^2 / \|\boldsymbol{z}^{(m)}\|_2^2$ and use the penalized objective

$$\mathcal{J}_m(\Omega, \pi) = \eta_K \frac{|\Omega|}{N} + \eta_\Pi \mathcal{C}_\Pi(\Omega, \pi) + \lambda_F [1 - F(\Omega; \boldsymbol{z}^{(m)}) - \varepsilon]_+, \tag{43}$$

where $[a]_+ = \max\{a, 0\}$ and $\mathcal{C}_\Pi(\Omega, \pi) = \sum_{i=1}^{n-1} w_i \Pi_i(\Omega, \pi) / \min\{2^i, 2^{n-i}\}$. Here, η_K controls support sparsity, η_Π weights the prefix-based bond surrogate, and $\lambda_F \geq 0$ penalizes fidelity violations. To cover different sparsity–prefix trade-offs without continuous hyperparameter tuning, we scan $(\eta_K, \eta_\Pi) \in \{(0,1), (0.01,1), (0.1,1), (1,1), (10,1), (100,1), (1,0)\}$, with each scalarization setting treated as a separate run and kept fixed throughout the corresponding optimization.

Algorithm 1 QuSAE joint support–mapping optimization and MPS construction

Require: Calibration set $\mathcal{D}_{\mathrm{map}} = \{\boldsymbol{x}^{(m)}\}_{m=1}^{M_{\mathrm{d}}}$, orthogonal DWT W, wavelet tree $\mathcal{T} = (V, E, r)$, mapping family $\mathcal{M}$, fidelity tolerance ε, fixed weights $\eta_K, \eta_\Pi, \{w_i\}$, improvement tolerance $\delta_J \geq 0$, and positive integer iteration limit $T_{\max}$

for $m = 1, \ldots, M_{\mathrm{d}}$ **do**
  Compute $\boldsymbol{z}^{(m)} \leftarrow W\boldsymbol{x}^{(m)}$
  Generate an ancestor-closed support pool $\mathcal{G}_m \subset \mathcal{S}(\mathcal{T})$, including V
  Retain $\mathcal{G}_m^{\varepsilon} \leftarrow \{\Omega \in \mathcal{G}_m : F(\Omega; \boldsymbol{z}^{(m)}) \geq 1 - \varepsilon\}$
end for
Initialize $\pi \leftarrow \pi_0 \in \mathcal{M}$ and $\boldsymbol{\Omega} \leftarrow (V, \ldots, V)$
for $t = 1, \ldots, T_{\max}$ **do**
  $J_{\mathrm{old}} \leftarrow \overline{\mathcal{J}}(\pi, \boldsymbol{\Omega})$
  Support update: with π fixed, select $\widehat{\Omega}_m \in \arg\min_{\Omega \in \mathcal{G}_m^{\varepsilon}} \mathcal{J}_m(\Omega, \pi)$ for every m
  Set $\widehat{\boldsymbol{\Omega}} \leftarrow (\widehat{\Omega}_1, \ldots, \widehat{\Omega}_{M_{\mathrm{d}}})$
  if $\overline{\mathcal{J}}(\pi, \widehat{\boldsymbol{\Omega}}) < \overline{\mathcal{J}}(\pi, \boldsymbol{\Omega}) - \delta_J$ **then**
    $\boldsymbol{\Omega} \leftarrow \widehat{\boldsymbol{\Omega}}$
  end if
  Mapping update: with $\boldsymbol{\Omega}$ fixed, select $\widehat{\pi} \in \arg\min_{\pi' \in \mathcal{M}} \overline{\mathcal{J}}(\pi', \boldsymbol{\Omega})$
  if $\overline{\mathcal{J}}(\widehat{\pi}, \boldsymbol{\Omega}) < \overline{\mathcal{J}}(\pi, \boldsymbol{\Omega}) - \delta_J$ **then**
    $\pi \leftarrow \widehat{\pi}$
  end if
  if $\overline{\mathcal{J}}(\pi, \boldsymbol{\Omega}) \geq J_{\mathrm{old}} - \delta_J$ **then**
    break
  end if
end for
Set $\pi^\star \leftarrow \pi$ and $\Omega_m^\star \leftarrow \Omega_m$ for every m
for $m = 1, \ldots, M_{\mathrm{d}}$ **do**
  Form the normalized mapped support state $\left|\psi_{\Omega_m^\star, \pi^\star}\right\rangle$
  Apply TT-SVD with $\tau_{\mathrm{svd}} = 10^{-12}$, without an additional bond cap, to obtain $A_m^\star$ and $\left|\widetilde{\psi}_{\Omega_m^\star, \pi^\star}\right\rangle$
  Record $\boldsymbol{\chi}_m^\star = \boldsymbol{\chi}(\Omega_m^\star, \pi^\star)$ and evaluate $\mathcal{C}_{\mathrm{MPS}}(\boldsymbol{\chi}_m^\star)$
end for
return Shared mapping $\pi^\star$, sample-specific supports $\{\Omega_m^\star\}$, MPS tensors and encoded states, bond profiles, and MPS complexities

The complete procedure is detailed in Algorithm 1. For each prescribed tolerance ε, candidate supports are first filtered by the fidelity constraint. All support updates are performed strictly over candidates satisfying $F(\Omega; \boldsymbol{z}^{(m)}) \geq 1 - \varepsilon$, so the penalty term in Eqn. (43) vanishes throughout the alternating optimization. Each scalarization setting therefore alternates between support and mapping updates under the same weighted sparsity–prefix objective subject to the prescribed fidelity constraint. In this hard-feasible mode, λ_F does not affect the updates.

The mapping stage uses the mapping-design samples as a calibration set. Writing $\boldsymbol{\Omega} = (\Omega_1, \ldots, \Omega_{M_{\mathrm{d}}})$ for the current supports of the M_{d} calibration samples, candidate mappings in $\mathcal{M}$ are evaluated using the calibration-averaged objective

$$\overline{\mathcal{J}}(\pi, \boldsymbol{\Omega}) = \frac{1}{M_{\mathrm{d}}} \sum_{m=1}^{M_{\mathrm{d}}} \mathcal{J}_m(\Omega_m, \pi).$$

The shared mapping is updated only when a candidate decreases this objective beyond the prescribed numerical tolerance. The support and mapping updates are alternated until no further improvement is obtained or the iteration limit is reached. For each scalarization setting, the best fidelity-feasible configuration under the same objective yields a dataset-shared mapping $\pi^\star$, while the corresponding optimized supports $\Omega_m^\star$ remain sample specific. Ties are resolved using a fixed candidate ordering.

Table 2: **Tree-compressibility under the selected wavelets.** Normalized minimum budgets K_ε/N are reported as percentages; larger s indicates faster tree-approximation error decay.

Dataset	$K_{10^{-1}}/N$ (%)	$K_{10^{-2}}/N$ (%)	$K_{10^{-3}}/N$ (%)	$K_{10^{-4}}/N$ (%)	s
MNIST	6.56	13.84	18.04	20.19	1.53
COIL-20	0.24	3.01	11.03	25.81	0.83
CelebA	0.18	4.35	19.63	45.38	0.79
MIT-BIH	3.80	9.00	20.98	47.82	1.20
STEAD	15.30	33.98	51.93	68.02	1.02

An unchanged sweep certifies that neither update block improves the objective by more than δ_J; reaching the iteration limit alone does not provide this certificate. Overall, the solver can be viewed as a structure-restricted alternating discrete search. It does not guarantee the global optimum of Eqn. (6), but avoids unrestricted enumeration by restricting both support and mapping searches to finite structured candidate sets.

D Additional experimental results

D.1 Tree-compressibility statistics

Table 2 complements the scaling curves in Fig. 8 by reporting the normalized minimum fidelity-compatible budgets at representative tolerances.

All fitted exponents are positive, supporting the s-tree-compressibility condition used in the resource analysis. At $\varepsilon = 10^{-2}$, COIL-20 and CelebA retain less than 5% of the coefficients, while MNIST exhibits the fastest fitted decay with $s = 1.53$ and retains only 20.19% even at $\varepsilon = 10^{-4}$. The one-dimensional datasets require larger budgets under stringent tolerances, particularly STEAD, illustrating the data dependence of the achievable structured compression.

D.2 Resource scaling analysis

QuSAE exhibits sublinear growth of the prefix-based bond upper bound.

Fig. 8 examines the two structural conditions underlying the conditional resource bounds. For tree approximation, a smaller budget indicates that fewer wavelet-tree coefficients are required to satisfy the same fidelity tolerance, while a larger s corresponds to faster approximation-error decay. The fitted tree exponents and representative fidelity-compatible budgets are summarized in Table 2.

These results support the s-tree-compressibility assumption over the tested range and show that QuSAE can construct substantially smaller fidelity-feasible structured supports before MPS encoding.

After structure-preserving mapping, all fitted prefix exponents are approximately 0.5 or below, indicating sublinear growth of the prefix-based bond upper bound with the retained support size over the tested range. Under the conditions of Eqn. (7), the fitted s and β correspond to the predicted exponents $\beta/(2s) \approx 0.10$–0.31 for minimum-support or budget-controlled fidelity-feasible encodings. Together, these observations provide empirical support for the structural conditions behind QuSAE and are consistent with its conditional resource-scaling mechanism.

The theoretical scaling results are conditional on several structural assumptions. For $0 < \varepsilon < 1$, consider the minimum fidelity-feasible tree support Ω_ε, with $|\Omega_\varepsilon| = K_\varepsilon$. Combining tree and prefix compressibility requires the prefix bound to hold on these same supports under the shared mapping $\pi^\star$, with s, β, γ_{tree}, and $\gamma_{\pi^\star}$ independent of ε over the considered tolerance range. Exact MPS realizations of these supports therefore yield fidelity-feasible encodings with the stated resource-scaling exponents. The same exponents apply to other fidelity-feasible supports Ω satisfying $|\Omega| \leq c_{\mathrm{bud}} K_\varepsilon$, for an ε-independent constant $c_{\mathrm{bud}} \geq 1$, provided that the prefix bound also holds for those supports; this budget factor changes only the prefactor. Accordingly, Eqn. (7) applies to the joint output $(\Omega^\star, \pi^\star)$ only under these support-budget and prefix conditions, rather than unconditionally for the practical solver. The finite-range fits do not themselves establish the support-budget condition

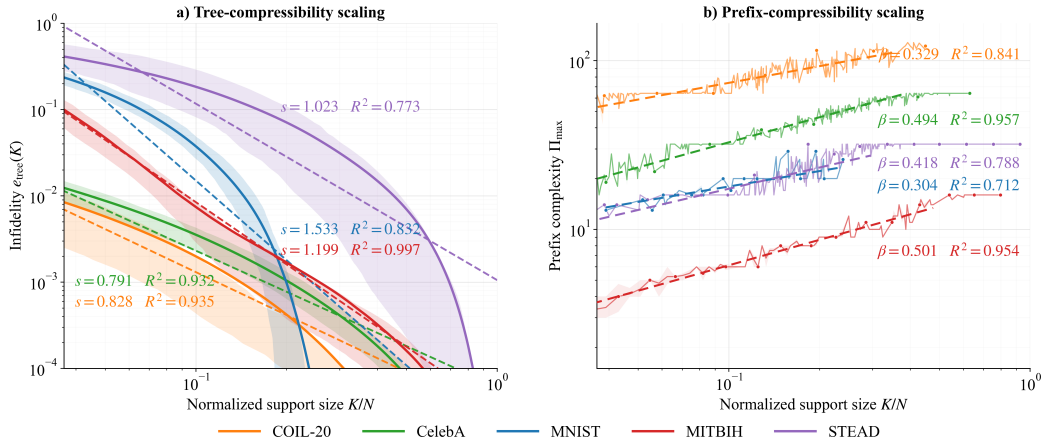

Figure 8: **Empirical resource-scaling analysis of QuSAE. (a)** Tree-compressibility scaling of the normalized approximation error $e_{\text{tree}}(K)$ with the retained support fraction K/N. Dashed lines show the fitted law $e_{\text{tree}}(K) \propto K^{-2s}$. **(b)** Prefix-compressibility scaling of $\Pi_{\max}$ with K/N, with dashed lines denoting fits to $\Pi_{\max} \propto K^{\beta}$. Solid curves show the empirical results and shaded regions indicate sample-level variability.

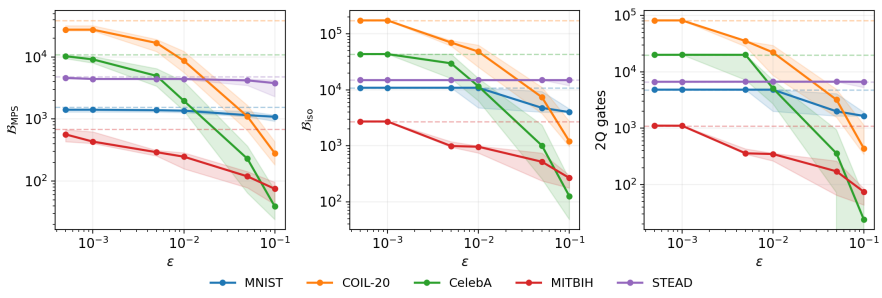

Figure 9: Sequential-preparation resource budgets across infidelity tolerances. The left and right panels report $\mathcal{B}_{\text{MPS}}$ and $\mathcal{B}_{\text{iso}}$, respectively. Solid curves show median QuSAE budgets over ten samples per class, with shaded bands indicating sample variability. Dashed horizontal lines denote the corresponding raw amplitude-state baselines.

for that output. For numerical TT-SVD, final-state fidelity additionally follows the realization-error bound in Proposition 1.

D.3 Sequential quantum circuits for MPS-encoded states

A small-sample experiment is conducted on the same representative real-world datasets to test whether the reduction in MPS representation complexity translates into lower sequential-circuit resource requirements. For each realized bond profile $\boldsymbol{\chi}$, we evaluate the tensor-level budget $\mathcal{B}_{\text{MPS}}(\boldsymbol{\chi}) = \mathcal{C}_{\text{MPS}}(\boldsymbol{\chi}) = \sum_{i=1}^{n} 2\chi_{i-1}\chi_i$ and the qubit-embedded local-isometry budget $\mathcal{B}_{\text{iso}}(\boldsymbol{\chi}) = \sum_{i=1}^{n} 4^{q_i}$, where $\chi_0 = \chi_n = 1$, $a_i = \lceil \log_2 \chi_i \rceil$, and $q_i = 1 + \max\{a_{i-1}, a_i\}$. These quantities characterize the resources implied by the realized bonds rather than backend-optimized gate counts.

The fidelity F reported here is evaluated between the final encoded state $\left|\widetilde{\psi}_{\Omega^\star,\pi^\star}\right\rangle$ and the full mapped target $U_{\pi^\star}|\psi(\boldsymbol{z})\rangle$. Fig. 9 shows that increasing the infidelity tolerance generally lowers $\mathcal{B}_{\text{MPS}}$, with dataset-dependent reductions in $\mathcal{B}_{\text{iso}}$. At $\varepsilon = 5 \times 10^{-3}$ ($F \approx 0.995$), the tensor-level budgets on COIL-20, CelebA, and MIT-BIH reach 0.433, 0.453, and 0.429 times their respective

raw amplitude-state baselines; the corresponding qubit-embedded budget ratios on COIL-20 and MIT-BIH are 0.402 and 0.365. Relaxing the fidelity requirement yields further resource reductions. At $\varepsilon = 10^{-2}$ ($F \approx 0.99$), the tensor-level budget ratios on COIL-20 and CelebA decrease to 0.224 and 0.180. For MIT-BIH, at $\varepsilon = 5 \times 10^{-2}$, $\mathcal{B}_{\mathrm{MPS}}$ and $\mathcal{B}_{\mathrm{iso}}$ reach 0.141 and 0.109 times the raw baseline, respectively. The plateaus in $\mathcal{B}_{\mathrm{iso}}$ reflect the integer local block widths q_i, which change only when the realized bonds cross qubit-register thresholds. For MNIST, retaining only 16.7% of the coefficients at $\varepsilon = 5 \times 10^{-3}$ leaves $\mathcal{B}_{\mathrm{iso}}$ at the raw baseline. STEAD likewise exhibits support reduction without substantially lowering its bond bottleneck or padded block widths. Thus, sequential-preparation savings depend on the realized bond profile and its qubit-register embedding, rather than support size alone.

It is worth noting that the circuit study here uses a basic sequential local-isometry construction as a conservative realization baseline. Recent progress in MPS and shallow state-preparation circuits has introduced more optimized decompositions, including shallow MPS-to-circuit constructions (Rudolph et al., 2024), log-depth MPS preparation schemes (Malz et al., 2024), and variational shallow encoding methods (Ben-Dov et al., 2024). These advances suggest that the compact MPS-encoded states may admit further reductions under more hardware-efficient MPS compilers. Thus, the present circuit audit should be viewed as a first executable validation of the representation-to-circuit translation, rather than the limit of possible circuit optimization.