symmetry

IMPACT FACTOR 2.2

CITESCORE 5.3

Article

Rank-Aware Conditional Synthesis: Feasible Quantum Generative Modeling on Matrix Product State Manifolds

Dongkyu Lee, Won-Gyeong Lee, Hyunjun Hong and Ohbyung Kwon

Special Issue

Symmetry and Its Applications in Deep Learning and Artificial Intelligence Methods

Edited by

Dr. Sibo Qiao

MDPI

https://doi.org/10.3390/sym18040605

symmetry

MDPI

Article

Rank-Aware Conditional Synthesis: Feasible Quantum Generative Modeling on Matrix Product State Manifolds

Dongkyu Lee [1], Won-Gyeong Lee [2], Hyunjun Hong [3] and Ohbyung Kwon [4,*]

[1] Quantum AI Task, LG AI Research, LG Electronics Co., Ltd., 19 Yangjae-daero 11-gil, Seocho-gu, Seoul 06772, Republic of Korea; dongkyu44.lee@lge.com
[2] Department of Smart Farm Science, Kyung Hee University, Deogyeong-daero, Giheung-gu, Yongin-si 17104, Republic of Korea; lwk4537@khu.ac.kr
[3] Agricultural AI Lab, Neobio Inc., 20 Pangyo-ro 289beon-gil, Bundang-gu, Seongnam-si 13488, Republic of Korea; hyunjun@neobio.town
[4] Department of Bigdata Analytics, Kyung Hee University, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul 02447, Republic of Korea
* Correspondence: obkwon@khu.ac.kr

Abstract

Matrix Product States (MPSs) have become an indispensable symmetry-based representation for simulating quantum systems on near-term hardware by constraining entanglement entropy through a fixed bond dimension χ. This study identifies a critical "rank explosion" phenomenon that destabilizes this low-rank manifold during conditional quantum diffusion processes. We empirically demonstrate that the introduction of conditional guidance—essential for semantic control—injects global correlations that drive the effective Schmidt rank to increase by 4× (from $\chi = 4$ to 16), saturating the simulation limits and necessitating quantum circuits with approximately 1.8×10^3 Controlled-NOT (CNOT) gates. Such circuit depths fundamentally exceed the operational coherence budgets of Noisy Intermediate-Scale Quantum (NISQ) devices. To mitigate this structural instability, we propose Rank-Aware Conditional Synthesis (RACS), a sampling framework that maintains the latent trajectory within a prescribed MPS manifold through step-wise manifold projection and time-shift error correction. Experimental results on real-world semantic data reveal that RACS reduces reconstruction error, or Mean Squared Error (MSE) by 30.8% and enhances latent trajectory smoothness by 36.8% compared to conventional post hoc truncation. At a fixed hardware-efficient rank of $\chi = 4$, RACS achieves a +4.8% fidelity gain and exhibits superior robustness against depolarizing noise. By resolving the tension between conditional expressivity and entanglement constraints, RACS provides a principled, hardware-aware methodology for high-fidelity quantum generative modeling.

Keywords: quantum machine learning; matrix product states; conditional diffusion models; entanglement entropy; NISQ feasibility; tensor network manifolds; rank-aware synthesis

Academic Editor: Jie Yang and Sibo Qiao

Received: 2 March 2026
Revised: 26 March 2026
Accepted: 31 March 2026
Published: 2 April 2026

1. Introduction

Conditional generation is a defining requirement for practical generative modeling. Without the ability to enforce explicit conditions—such as class labels or semantic constraints—generative models remain of limited utility for downstream tasks including controllable synthesis, inverse design, and data augmentation. In classical machine learning, diffusion models have recently emerged as a leading framework for conditional generation, owing to their stable training objectives and effective guidance mechanisms [1–3].

Recent work has begun extending diffusion-based generative modeling into the quantum domain, proposing quantum-native and hybrid quantum–classical diffusion processes for learning complex probability distributions [4]. These advances naturally motivate the exploration of diffusion-based approaches within quantum machine learning (QML), where quantum resources promise expressive representations of complex, high-dimensional distributions [5,6]. Recent advances in quantum generative models have demonstrated the feasibility of learning complex distributions on near-term quantum devices, with competitive performance against classical counterparts [7].

Despite this promise, QML encounters a well-known and fundamental challenge: quantum state preparation and representation scale poorly when strong global correlations are present. Preparing an arbitrary n-qubit quantum state generally requires resources that scale exponentially with system size due to the intrinsic $O(2^n)$ degrees of freedom of the Hilbert space [8–10]. This state-preparation bottleneck has been consistently recognized as a significant barrier to near-term quantum advantage in generative modeling and learning pipelines [7,11,12].

Tensor-network representations, particularly Matrix Product States (MPS), provide a principled way to mitigate this bottleneck by exploiting structured entanglement patterns [13–15]. An MPS represents a quantum state as a chain of low-rank tensors, parameterized by a bond dimension χ that directly controls both expressivity and entanglement capacity. When correlations are predominantly local, MPS enables compact representations and efficient manipulation, making it widely adopted in classical many-body simulations and an attractive substrate for quantum state encoding [13–16]. Recent tensor-network-based learning methods have explored scalable training with adaptive bond-dimension control via SVD-based truncation during optimization [17].

However, χ is not merely a modeling parameter: it directly determines physical feasibility. As χ increases, the number of entangling gates required to realize the corresponding quantum circuit grows rapidly—often quadratically—quickly exceeding the depth tolerable on noisy intermediate-scale quantum (NISQ) devices [18,19].

In this work, we identify and study an under-explored but critical failure mode that arises when conditional guidance is applied to diffusion processes operating on MPS latent spaces. Conditional diffusion updates introduce guidance terms that encode global, condition-dependent correlations [2,3]. Recent work indicates that conditional guidance can introduce structural instability in generative processes despite improving controllability [20].

While effective for control, these terms fundamentally alter the entanglement structure of intermediate latent states. Through experiments on real handwritten-digit data, we empirically observe that conditional guidance consistently drives the effective Schmidt rank of latent MPS tensors to saturate the maximum allowed value, even when unconditional generation remains stable at the initial rank budget. In our setting, this manifests as an immediate $\sim 4\times$ rank increase (e.g., from $\chi = 4$ *to* $\chi = 16$), a phenomenon we refer to as rank explosion.

This phenomenon has direct and severe implications for hardware feasibility. Allowing rank explosion implies quantum circuits dominated by thousands of entangling gates, rendering execution on NISQ hardware effectively impossible due to accumulated noise and limited coherence times [18,19]. The expressivity of quantum models is fundamentally limited by entanglement constraints under realistic hardware settings [21].

A natural workaround is to allow the diffusion process to evolve freely and apply a single truncation step at the end. However, this approach remains inadequate. Our results show that such naive end truncation behaves as a catastrophic projection: once the

diffusion trajectory departs far from the low-rank manifold, late-stage truncation discards substantial information, leading to avoidable fidelity loss.

To address this, we propose Rank-Aware Conditional Synthesis (RACS), a sampling framework designed to preserve conditional controllability while maintaining proximity to a target low-rank MPS manifold throughout generation. Rather than permitting uncontrolled rank growth followed by abrupt compression, RACS enforces rank constraints *step-wise* during sampling. At each diffusion step, we apply conditional guidance, project the updated latent back onto the admissible MPS manifold, and compensates for projection-induced distortion using a correction mechanism inspired by recent time- or variance-corrected diffusion samplers [22]. This strategy transforms a single catastrophic truncation into many small, recoverable corrections, yielding smoother trajectories and improved fidelity under fixed rank budgets.

RACS integrates algorithmic design with physical constraints. By maintaining a low bond dimension throughout sampling, RACS dramatically reduces the required entangling-gate count while preserving or improving output quality compared to naive truncation under the same resource limits. In this sense, RACS provides a concrete mechanism for bridging conditional generative modeling and NISQ feasibility, addressing a central obstacle to practical quantum generative pipelines.

While recent advances in classical tensor-network machine learning have explored adaptive bond-dimension procedures and truncation schemes, these methods primarily focus on static state preparation or unconditional dynamics. Unlike prior post hoc truncation approaches that passively allow unbounded entanglement growth before catastrophic compression, RACS proposes an active, step-wise intervention. By critically comparing our dynamic rank control against these static baseline methods, we demonstrate that resolving the rank explosion inherently triggered by conditional guidance requires preserving both semantic conditionality and structural geometry.

This paper provides three core contributions. First, we empirically demonstrate that rank explosion under conditional guidance is a real and systematic phenomenon on real data, not an artifact of synthetic benchmarks. Second, we introduce RACS, a rank-aware conditional sampling framework that stabilizes diffusion trajectories on low-rank MPS manifolds. Third, we establish the hardware relevance of rank-aware synthesis by explicitly connecting rank control to entangling-gate counts, fidelity retention, and robustness under realistic noise models. Together, these results suggest that rank-aware control is not optional but essential for conditional quantum generation in the NISQ era.

2. Materials and Methods

In this section, we present the proposed Rank-Aware Conditional Synthesis (RACS) framework. We begin by defining the mathematical foundation of Matrix Product State (MPS) manifolds and the associated rank constraints. We then describe the real-world semantic dataset used in this study and the quantum encoding pipeline that maps classical data to low-rank quantum states. Finally, we formalize latent diffusion on MPS, analyze the structural origin of rank explosion under conditional guidance, and detail the RACS algorithm for rank-controlled conditional generation. All experiments were conducted using PyTorch (Meta AI, Menlo Park, CA, USA) and NumPy (NumPy Developers, USA). The MNIST dataset was obtained via TensorFlow Datasets (Google LLC, Mountain View, CA, USA). Quantum circuit simulations were performed using Qiskit (IBM Quantum, Armonk, NY, USA). Computations were executed on an NVIDIA GPU (NVIDIA Corporation, Santa Clara, CA, USA).

2.1. Preliminaries: MPS Manifolds and Rank Constraints

Among tensor-network representations, Matrix Product States (MPS) provide a compact representation of quantum systems with structured entanglement. While advanced architectures like Tree Tensor Networks (TTN) or MERA offer superior expressivity for hierarchical or critical systems, MPS remains the most practical choice for near-term generative modeling [23]. This is because the 1D linear structure of MPS natively maps to the linear or grid-like qubit connectivity of most superconducting NISQ processors (e.g., IBM Quantum), avoiding the massive SWAP-gate overhead required to embed tree-like structures. For an n-qubit system, a quantum state $|\psi\rangle$ can be written in MPS form as:

$$|\psi\rangle = \sum_{i_1,\ldots,i_n} A_{i_1}^{[1]} A_{i_2}^{[2]} \cdots A_{i_n}^{[n]} |i_1,\ldots,i_n\rangle, \tag{1}$$

where each tensor $A_{i_k}^{[k]} \in C^{\chi_{k-1}\times\chi_k}$ denotes the bond dimension between sites k and $k+1$. The maximum bond dimension $\chi = \max_k \chi_k$ controls the expressivity and entanglement capacity of the representation.

We define the MPS manifold:

$$M\chi = \{\, |\,\psi\rangle \,|\, \textit{bond dimension} \le \chi \,\}, \tag{2}$$

as the set of all quantum states representable with bond dimension at most χ.

This constraint has direct physical implications. When an MPS is compiled into a quantum circuit, the number of required entangling gates typically scales as $O(n \cdot \chi^2)$. Consequently, even moderate increases in χ can result in circuit depths that exceed the coherence limits of noisy intermediate-scale quantum (NISQ) devices. Maintaining latent states within $\mathcal{M}_\chi$ throughout generation is therefore essential for hardware feasibility [24].

2.2. Materials: Real-World Semantic Data and Quantum Encoding

To study structural instabilities arising in conditional quantum generation, we use real-world image data instead of synthetic or randomly generated quantum states. Specifically, we use the MNIST handwritten digit dataset, which exhibits strong and structured spatial correlations.

Rationale.

Unlike random vectors in Hilbert space, natural images contain meaningful global and local correlations induced by semantic content. When mapped to quantum states, these correlations translate into nontrivial entanglement patterns. MNIST therefore serves as an effective material for probing how conditional information interacts with structured entanglement and whether such interactions destabilize low-rank tensor-network representations.

Preprocessing.

Original 28×28 grayscale images are resized to 16×16 resolution to balance semantic preservation and computational tractability. Each image is flattened into a vector of length $N = 256$, corresponding exactly to the Hilbert space dimension of an n = 8-qubit system $\left(2^8 = 256\right)$.

Quantum encoding.

Each image vector $v \in R^{256}$ is normalized and mapped to a quantum state using amplitude encoding:

$$|\psi\rangle = \sum_{i=0}^{2^n-1} v_i |i\rangle. \tag{3}$$

This encoding preserves global correlations and provides a direct mapping from classical semantic structure to quantum-state entanglement. All encoded states are subsequently converted into MPS form with a prescribed initial bond dimension χ_0.

2.3. Latent Diffusion on Matrix Product States

We consider a generative process defined directly on the continuous parameters of MPS tensors. Let z_t denote the flattened vector formed by concatenating all entries of the MPS tensors at diffusion time step t.

The forward diffusion process $q(z_t|z_0)$ incrementally perturbs the latent representation with Gaussian noise, while the reverse process $p_\theta(z_{t-1}|z_t)$ learns to denoise z_t and reconstruct a low-rank quantum state. Conditioning information c is incorporated via a guidance term in the reverse process, following standard conditional diffusion paradigms.

2.4. The Origin of Rank Explosion Under Conditional Guidance

During conditional reverse diffusion, the latent update can be written as:

$$z'_t = z_t + w \cdot g_c(z_t), \tag{4}$$

where $g_c = \nabla_{z_t} \log p(c|z_t)$ denotes the conditional guidance vector and w is the guidance weight.

Interpreting z_t and g_c back in tensor form, this update corresponds to a sum of tensor networks. For matrices and tensors, the rank of a sum satisfies

$$rank(X + Y) \leq rank(X) + rank((Y). \tag{5}$$

In the MPS context, the guidance term g_c encodes global correlations required to satisfy the condition c. As a result, adding g_c to the current latent state increases the effective Schmidt rank across multiple bipartitions. Repeated application of conditional guidance leads to cumulative rank inflation, rapidly pushing the latent state outside the target manifold $\mathcal{M}_\chi$.

This reveals a structural tension: conditional generation demands global correlation, while MPS efficiency relies on constrained entanglement.

2.5. Rank-Aware Conditional Synthesis (RACS)

To resolve this conflict, we propose Rank-Aware Conditional Synthesis (RACS), which enforces rank constraints dynamically throughout the reverse diffusion process rather than relying on a single truncation step at the end.

Each reverse diffusion step proceeds in three stages:

Step 1: Conditional Guidance (Exploration).
$z'_t = z_t + w \cdot g_c(z_t)$, which improves alignment with the target condition but may increase rank.
Step 2: Manifold Projection (Constraint).
The guided latent state is projected back onto the admissible MPS manifold via SVD-based truncation:

$$\tilde{z}_t = \prod_{\mathcal{M}_\chi}(z'_t), \tag{6}$$

ensuring $rank \leq \chi$ at the cost of truncation error.
Step 3: Time-Shift Correction (Compensation).
Truncation removes part of the accumulated variance, effectively altering the diffusion noise schedule. Mathematically, SVD-based manifold projection discards the smallest singular values to satisfy the rank constraint rank $\leq \chi$. This intrinsically reduces the total norm of the latent state. Since diffusion models rely on a strict forward process where the variance

at step t must exactly match a predefined schedule, this truncation-induced variance loss breaks the schedule consistency. To compensate, we apply a time-shift correction inspired by time-corrected diffusion samplers:

$$\hat{z}_t = TimeCorrect(\tilde{z}_t), \tag{7}$$

where the effective diffusion time is adjusted to restore consistency with the assumed noise level.
By repeating this procedure, RACS converts a single catastrophic truncation into a sequence of micro-corrections, maintaining high fidelity within a fixed rank budget.

To clarify the formal connections in Algorithm 1, let n denote the total number of qubits, χ the target bond dimension, and w the conditional guidance scale. The dominant computational overhead in RACS arises from the manifold projection (Step 2). Truncating the updated latent state z'_t back to $\mathcal{M}_\chi$ requires a sequence of Singular Value Decompositions (SVD). For an MPS of n sites, this projection scales asymptotically as $O(n\chi^3)$. By dynamically enforcing this constraint, RACS bounds the total classical simulation complexity to $O(T \cdot n\chi^3)$ over T timesteps, directly preventing the intractable $O(n\chi^3_{\max})$ scaling that occurs when the rank explodes. Beyond computational efficiency, this step-wise projection also provides theoretical rank stability bounded by the Eckart-Young-Mirsky theorem. At each step t, the projection $z_t = \Pi_{\mathcal{M}_\chi}(z'_t)$ guarantees that the truncation error $||z'_t - z_t||^2_F$ is minimized for the given χ. However, this projection strictly reduces the variance of the latent state. Without correction, this variance deficit accumulates, causing the diffusion trajectory to mathematically deviate from the evidence lower bound (ELBO) of the target distribution. The time-shift correction, empirically verified in Section 3.4, acts as a variance-matching mechanism. It restores the local noise consistency, ensuring that the trajectory remains asymptotically bounded near the ideal forward process.

Algorithm 1: Rank-Aware Conditional Synthesis (RACS).

Input:

- Initial latent state $z_T \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$
- Conditional guidance function $g_c(\cdot)$
- Guidance weight w
- Target bond dimension χ
- Diffusion timesteps T

Output:

- Low − rank conditional MPS sample $z_0 \in \mathcal{M}_\chi$

Procedure:

$z_t \longleftarrow z_T$ //Initialize with original predictions
for $t = T, T-1, \ldots, 1$ *do*
  $z'_t \longleftarrow z_t + w \cdot g_c(z_t)$ //Step 1: Conditional guidance
  $\tilde{z}_t \longleftarrow \prod_{\mathcal{M}_\chi}(z'_t)$ //Step 2: Manifold projection
  $\hat{z}_t \longleftarrow TimeCorrect(\tilde{z}_t, t)$ //Step 3: Time-shift correction
  $z_{t-1} \longleftarrow \hat{z}_t$
end for
return z_0

2.6. Practical Considerations: The Double Defense System

In practice, RACS operates synergistically with structural optimizations. Optimal Qubit Mapping acts as a passive defense by reordering qubits to minimize initial entanglement entropy, reducing the pressure on rank control [25]. RACS then serves as an active defense, dynamically correcting rank growth induced by conditional guidance.

Together, this double-defense system significantly reduces the computational overhead of rank truncation while maximizing fidelity and hardware feasibility of the generated quantum states.

3. Results

We evaluate RACS framework on real handwritten digit data to examine how conditional guidance alters the entanglement structure of quantum states and whether rank-aware control improves practical feasibility. The experiments focus on:

(i) Identifying rank explosion during conditional diffusion sampling,
(ii) Understanding how this structural inflation manifests even on real, low-rank data,
(iii) Demonstrating the need for explicit rank control for stable and resource-efficient quantum state synthesis.

3.1. Rank Explosion on Real MNIST Digits

We examine whether conditional guidance intrinsically amplifies entanglement complexity even when the target data are real and semantically structured. Following the experimental setup in Table 1, we track the evolution of the effective Schmidt rank along the reverse diffusion trajectory, comparing unconditional sampling against class-conditional sampling on real digit instances.

Table 1. Experimental setup, baselines, and evaluation metrics.

Category	Item	Description
Dataset	Dataset name	MNIST (real-world handwritten digits)
	Data type	Grayscale image, semantic real data
	Classes	10 (digits 0–9)
Preprocessing	Image resolution	Resized from 28×28 to 16×16
	Flattening	256-dimensional vector
	Qubit mapping	$n = 8$ qubits $(2^8 = 256)$
	Encoding scheme	Amplitude encoding
Latent Representation	Tensor format	Matrix Product State (MPS)
	Initial bond dimension	$\chi = 4$
	Maximum analysis rank	$\chi_{max} = 16$
Generative Model	Framework	Latent diffusion on MPS parameters
	Sampling modes	Unconditional/Conditional (class-guided)
Baselines	Baseline 1	Unconditional diffusion
	Baseline 2	Conditional diffusion + naive end-truncation
	Baseline 3	Conditional diffusion + projection-only
Proposed Method	Method name	Rank-Aware Conditional Synthesis (RACS)
	Key components	Optimal Qubit Mapping + Rank Correction Sampler
Evaluation Metrics	Structural metrics	Effective rank, entanglement entropy
	Hardware metrics	CNOT gate count
	Quality metrics	MSE, reconstruction fidelity
	Stability metrics	Latent trajectory smoothness
	Robustness metrics	Noise robustness (depolarizing channel)

Figure 1 shows direct evidence of structural inflation under conditional generation. As illustrated in Figure 1a, unconditional sampling keeps the state in the initial low-rank

regime, with $\chi = 4$ throughout the diffusion steps. In contrast, conditional sampling diverges from the efficient MPS manifold: the effective rank saturates at the simulation limit $\chi = 16$ within the earliest steps and remains maximized for the remainder of the trajectory. This corresponds to a $4\times$ increase in effective rank compared to the initial state.

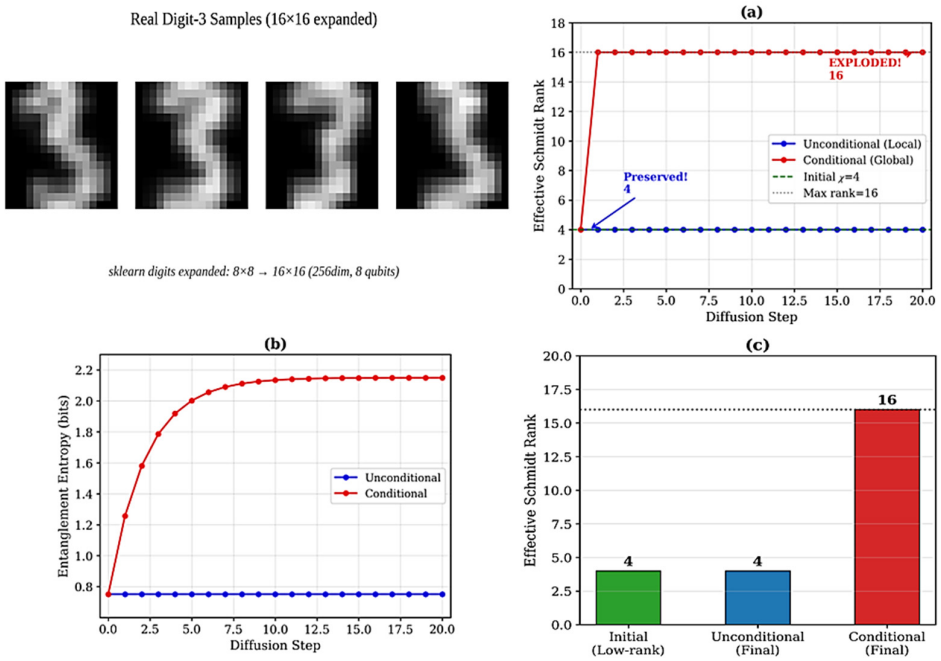

Figure 1. Rank explosion under conditional diffusion sampling on real handwritten digits. (**a**) Rank remains low for unconditional sampling ($\chi = 4$) but quickly saturates under conditional guidance ($\chi = 16$). (**b**) Entropy rises only in the conditional case. (**c**) Final ranks confirm persistent rank inflation.

The entropy dynamics in Figure 1b reinforce this rank explosion. While unconditional sampling stays in a low-entropy regime, conditional guidance induces a sharp entropy increase, suggesting that the guidance signal introduces global correlations that exceed the entanglement limit imposed by the fixed bond dimension. Finally, Figure 1c confirms that this structural inflation persists to the final sample: unconditional diffusion returns a low-rank state consistent with the initial manifold, whereas conditional diffusion yields a high-rank state occupying the largest accessible subspace under $\chi_{max} = 16$.

A quantitative summary of this structural inflation is reported in Table 2. Overall, these results confirm that rank explosion is not a synthetic artifact, but a fundamental structural consequence of conditional guidance on real semantic data—thereby motivating explicit rank-aware control mechanisms such as RACS.

Table 2. Quantitative summary of rank explosion under conditional diffusion.

Setting	Initial Rank (χ_0)	Peak Rank (χ_{max})	Final Rank (χ_T)	Inflation (χ_{max}/χ_0)	Entropy Trend
Unconditional	4	4	4	1.0×	Flat (low)
Conditional	4	16	16	4.0×	Rapid increase (high)

3.2. Truncation Failure: Post Hoc Compression Destroys Conditional Structure

Rank truncation is commonly used to keep tensor-network models feasible, but applying it only once at the end of sampling is insufficient under conditional diffusion. The core issue is that class guidance rapidly drives the intermediate state outside the low-rank MPS manifold, accumulating global correlations that cannot be captured at small bond

dimension. As a result, an end-of-process SVD truncation does not recover a feasible state, but instead removes essential structure, leading to irreversible degradation.

Figure 2a quantifies this failure by showing reconstruction fidelity after end-truncation under different rank budgets. When the target bond dimension is small, fidelity collapses sharply—0.78 at $\chi = 2$ and 0.88 at $\chi = 4$—demonstrating that the generated conditional state cannot be compressed into the low-rank subspace without distortion. Fidelity increases only when the rank budget becomes large (0.97 at $\chi = 8$, 1.00 at $\chi = 16$), indicating that naïve truncation "works" only at ranks already impractical for NISQ implementation.

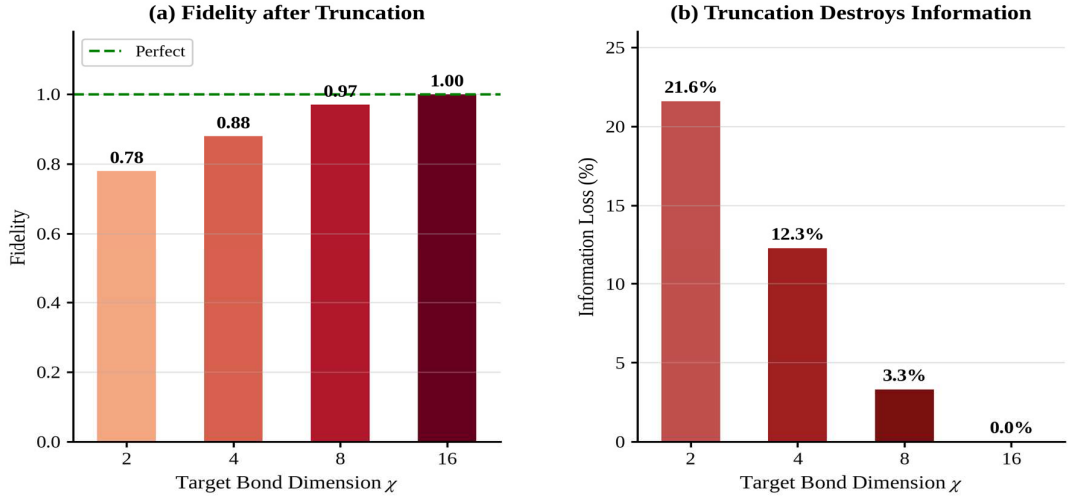

Figure 2. Post hoc rank truncation irreversibly destroys conditional structure. (**a**) Fidelity drops sharply at small bond dimensions ($\chi = 2,\ 4$) and recovers only at larger χ. (**b**) Truncation induces substantial information loss in the low-rank regime.

To directly characterize how much structure is destroyed by truncation, Figure 2b reports the information loss induced by compression. The loss is substantial in the low-rank regime, reaching 21.6% at $\chi = 2$ and 12.3% at $\chi = 4$, while decreasing at higher ranks (3.3% at $\chi = 8$, 0.0% at $\chi = 16$). These results confirm that end-truncation is not a benign approximation: under realistic rank constraints, it behaves as a catastrophic projection that removes the very correlations introduced by conditional guidance.

Table 3 summarizes the same outcome numerically (fidelity and information loss) to provide a compact comparison across rank budgets. Overall, this experiment establishes that post hoc truncation cannot repair rank explosion. Instead, feasibility must be enforced during sampling, motivating step-wise rank control strategies such as RACS.

Table 3. Numerical summary of fidelity and information loss after end truncation.

Target Bond Dimension (χ)	Fidelity After End-Truncation	Information Loss (%)
2	0.78	21.6
4	0.88	12.3
8	0.97	3.3
16	1.00	0.0

3.3. Circuit Feasibility and NISQ Robustness

The truncation results in Section 3.2 show that low-rank compression applied only at the end of sampling disrupts conditional structure. Beyond reconstruction quality, this structural mismatch directly dictates hardware feasibility, as an inflated bond dimension requires quantum circuits that are too deep for current NISQ architectures. In this section,

we quantify how rank inflation increases circuit cost, and we further evaluate the stability of low-rank sampling under realistic noise.

Figure 3a,b report the compilation cost of the generated MPS as a function of bond dimension. As the rank increases, the circuit depth grows approximately as $O(n\log\chi)$, while the number of entangling gates scales more steeply, approximately as $O(n\chi^2)$. Consequently, allowing conditional sampling to saturate $\chi = 16$ leads to a growth in entangling operations, reaching $\sim 1.8 \times 10^3$ CNOT gates. This strictly violates real-world hardware constraints, as current NISQ processors are heavily bound by limited coherence times and restricted qubit connectivity. Such an inflated bond dimension requires highly non-local entangling operations that rapidly decohere physical qubits. By constraining the trajectory to $\chi = 4$, RACS aligns the generative process with these practical real-world limitations. In contrast, constraining the sampling trajectory to a feasible manifold (e.g., $\chi = 4$) reduces the entangling-gate requirement to roughly 10^2 CNOTs, shifting the generation regime from theoretically valid but unexecutable to physically runnable.

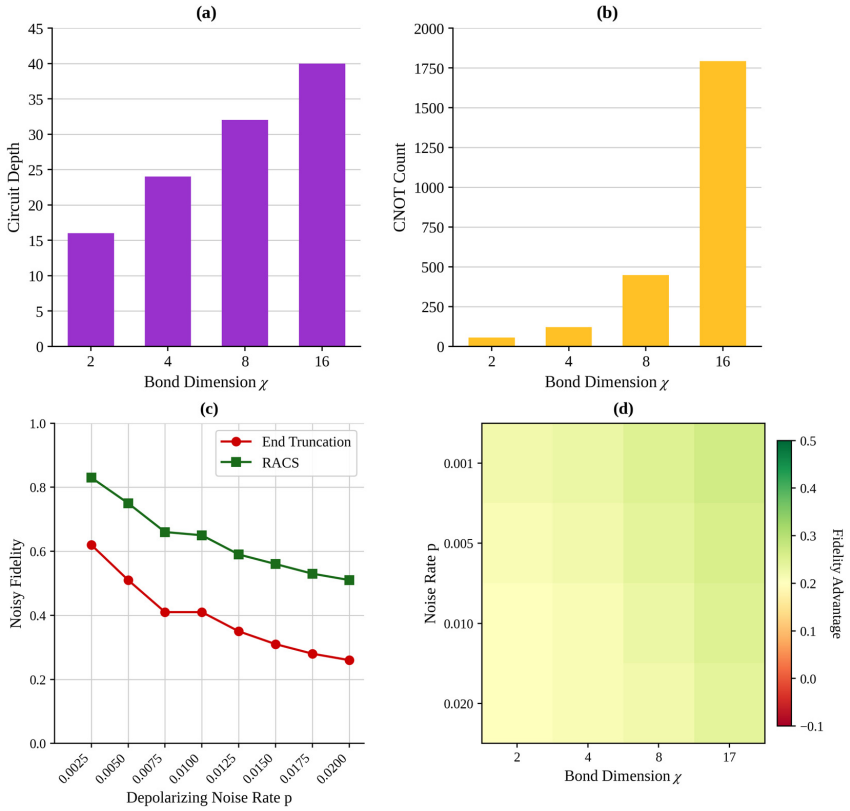

Figure 3. Circuit feasibility and NISQ robustness as a function of bond dimension. (**a**) Circuit depth increases with χ, reflecting growing compilation complexity; (**b**) entangling-gate count grows much more steeply, indicating the hardware cost of rank inflation; (**c**) RACS outperforms end truncation under depolarizing noise at $\chi = 4$; (**d**) RACS gains are strongest in the low-rank regime.

To approximate execution on platforms like IBM Qiskit or Google Cirq without the prohibitive cost of running $\sim 1.8 \times 10^3$ CNOTs on physical hardware (which would currently yield near-zero fidelity), we evaluate robustness under a hardware-agnostic depolarizing noise channel. This noise model serves as a standard proxy for the cumulative gate errors present in real NISQ devices [26]. Figure 3c compares end-truncation and RACS at a fixed feasible rank $\chi = 4$ across increasing depolarizing error rates p. While both methods degrade with noise, RACS consistently maintains higher fidelity across all tested regimes. For example, at low noise ($p = 0.001$), RACS preserves fidelity at approximately 0.83, whereas end-truncation remains around 0.62. Under more severe corruption ($p = 0.02$), RACS still

retains a usable reconstruction level (≈ 0.52), while end-truncation collapses more sharply (≈ 0.27). This gap indicates that step-wise rank control does not only reduce circuit cost, but also yields intrinsically more noise-tolerant quantum states by preventing unstable manifold departure during sampling.

Finally, Figure 3d summarizes the advantage landscape across both bond dimension and noise rate. The heatmap shows that the fidelity benefit of RACS is most pronounced in the resource-limited regime (small χ), precisely where NISQ feasibility is most constrained. Rank-aware control is therefore integral to both structural fidelity and hardware feasibility. Controlling rank during synthesis ensures that conditional generation remains physically executable and robust, providing a viable pathway for implementing complex generative tasks on NISQ devices.

3.4. Mechanistic Evidence and Practical Benefits of RACS

To confirm that the performance gain of RACS reflects a genuine structural mechanism rather than coincidence, we conduct a series of diagnostic analyses examining (i) visual error localization, (ii) component-wise contribution, (iii) synergy with optimal qubit mapping, and (iv) latent-space stability. As summarized in Figures 4–8, these analyses consistently indicate that RACS improves generation quality by preventing catastrophic rank collapse while maintaining a physically smoother diffusion trajectory under strict low-rank budgets.

3.4.1. Visual Error Localization: Difference Maps Confirm Structure Preservation

While fidelity metrics quantify the average similarity between reconstructions and references, they do not directly reveal where truncation damages the signal. To provide localized evidence, Figure 4 visualizes pixel-level difference maps between the original digit and outputs obtained by naive end-truncation and RACS.

As shown in Figure 4, naive end-truncation introduces spatially distributed artifacts, particularly along high-curvature stroke regions and boundary transitions. In contrast, RACS suppresses these deviations and yields a visibly cleaner reconstruction. Quantitatively, the mean squared error drops from 0.02749 (naive truncation) to 0.01902 (RACS), corresponding to a 30.8% reduction in reconstruction error. This confirms that RACS preserves digit morphology more reliably under low-rank constraints rather than merely increasing global similarity.

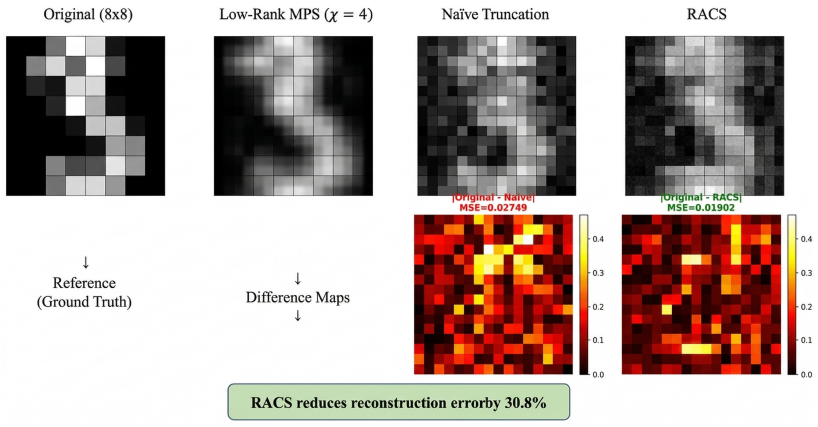

Figure 4. Difference maps reveal localized error suppression by RACS. Naïve truncation produces widespread artifacts, while RACS localized errors and preserves structure.

3.4.2. Component-Wise Contribution: Ablation Study of RACS

To identify which components drive the improvements, we conduct an ablation study in Figure 5 by disabling each correction module while keeping the diffusion backbone fixed.

Figure 5a demonstrates that structure preservation improves from 0.847 (None) to 0.877 (Full RACS), showing consistent gains when rank-aware corrections are introduced. Figure 5b further shows that reconstruction error decreases substantially when either (i) projection-based stabilization or (ii) temporal correction is applied.

Most importantly, Figure 5c decomposes the fidelity gain into two main positive contributors:

- Projection effect (+0.021): stabilizes rank distortion geometrically, preventing abrupt truncation artifacts.
- Time-correction effect (+0.033): enforces stepwise consistency, preventing accumulated rank drift across diffusion steps.

The synergy term is slightly negative (−0.024), suggesting that the two mechanisms partially overlap in what they correct (i.e., they are not orthogonal). This behavior is consistent with known optimization landscape phenomena in shallow parametrized quantum circuits [27,28]. Nevertheless, the full combination remains the most stable configuration overall, supporting the use of both correction types rather than relying on a single mechanism.

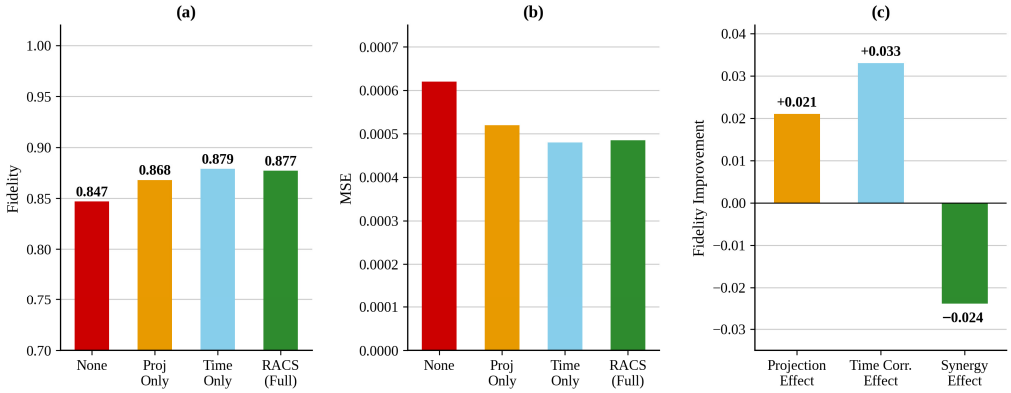

Figure 5. Component-Wise Contribution: Ablation Study of RACS. (**a**) Fidelity across different ablation settings. (**b**) Corresponding reconstruction error (MSE). (**c**) Decomposition of the fidelity gain into projection and time-correction effects.

3.4.3. Double Defense Mechanism: Optimal Qubit Mapping + RACS Synergy

Beyond stepwise rank correction, Figure 6 shows that optimal qubit mapping and RACS form a complementary double defense strategy. The first defense is passive: reducing initial entanglement through qubit ordering. The second defense is active: correcting rank explosion during diffusion.

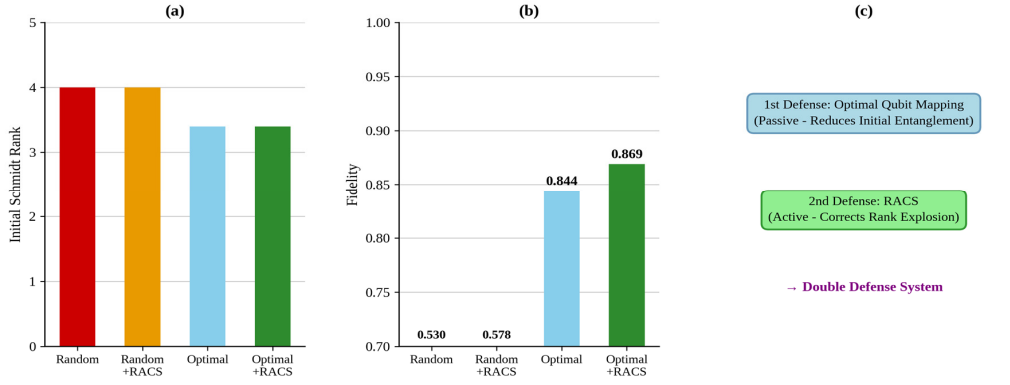

Figure 6. Double Defense Mechanism: Optimal Qubit Mapping + RACS Synergy. (**a**) Initial effective Schmidt rank under different qubit mappings. (**b**) Fidelity under rank-bounded generation. (**c**) Conceptual illustration of the two-stage defense system, where the first stage (optimal qubit mapping)

passively reduces initial entanglement and the second stage (RACS) actively suppresses rank growth during diffusion. As shown in (**a**), random qubit mappings initialize the system at an effective Schmidt rank of 4, whereas optimal mapping reduces this baseline entanglement to 3, lowering the initial complexity before diffusion begins. (**b**) shows that this reduced entanglement directly translates into stronger preservation under rank-bounded generation: fidelity improves from 0.844 (optimal only) to 0.869 (optimal + RACS). (**c**) summarizes this interaction as a two-stage robustness design: mapping reduces the entanglement pressure at initialization, while RACS continuously prevents rank blow-up during sampling.

This synergy is important int practice because it reduces both initial and dynamic entanglement demands, which is essential for low-rank implementation.

3.4.4. Latent-Space Stability: RACS Maintains a Smoother Diffusion Path

A core hypothesis of RACS is that naive truncation destabilizes diffusion dynamics by allowing uncontrolled rank growth followed by abrupt collapse, producing discontinuous latent evolution. Figure 7 directly evaluates this by projecting latent trajectories into a PCA space and measuring trajectory smoothness using the path length metric.

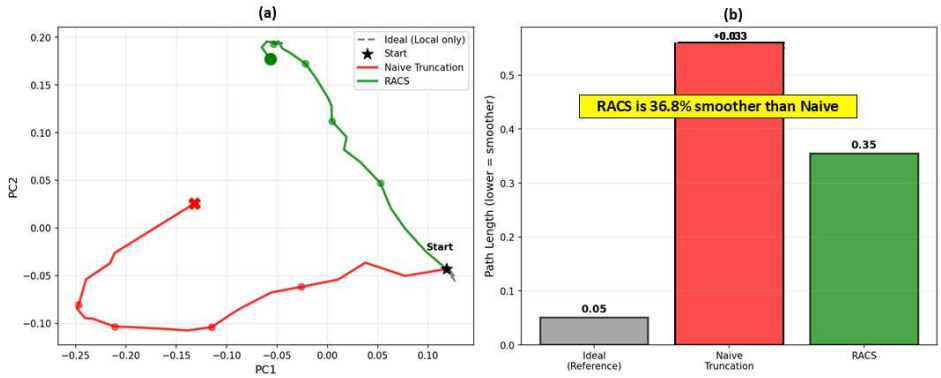

Figure 7. Latent-Space Stability: RACS Maintains a Smoother Diffusion Path. (**a**) PCA projection of latent diffusion trajectories. (**b**) Trajectory path length, where shorter paths indicate smoother diffusion.

As shown in Figure 7a, naive truncation exhibits an irregular and drifting trajectory that deviates from the ideal low-rank reference path. In contrast, RACS follows a smoother, more coherent path that remains closer to the ideal latent structure. Figure 7b quantifies this effect: the trajectory path length decreases from 0.56 (naive truncation) to 0.35 (RACS), corresponding RACS is 36.8% smoother.

This stability is not merely visual; it supports the view that RACS preserves diffusion continuity and prevents destructive truncation shocks, which explains why it produces more faithful reconstructions under identical rank budgets.

3.4.5. Consolidated Summary: RACS as a Controlled Alternative to Naïve Truncation

Finally, Figure 8 presents a joint comparison of rank dynamics and final reconstruction outcomes. Figure 8a highlights the fundamental failure mode: naive truncation allows the rank to explode to the maximum (16) early in diffusion, whereas RACS keeps the process controlled under the target budget.

This difference appears clearly in the final generation quality. At $\chi = 4$, Figure 8b reports a final fidelity improvement from 0.591 (naive) to 0.619 (RACS), corresponding to a +4.8% gain under the same strict rank constraint. Figure 8c further shows that this advantage is strongest in the low-to-mid χ regime where resource constraints are tight, demonstrating that RACS is most effective under NISQ-relevant budgets. The final summary panel (Figure 8) emphasizes the key message: RACS converts catastrophic truncation into controlled stepwise correction, enabling better fidelity without requiring large χ.

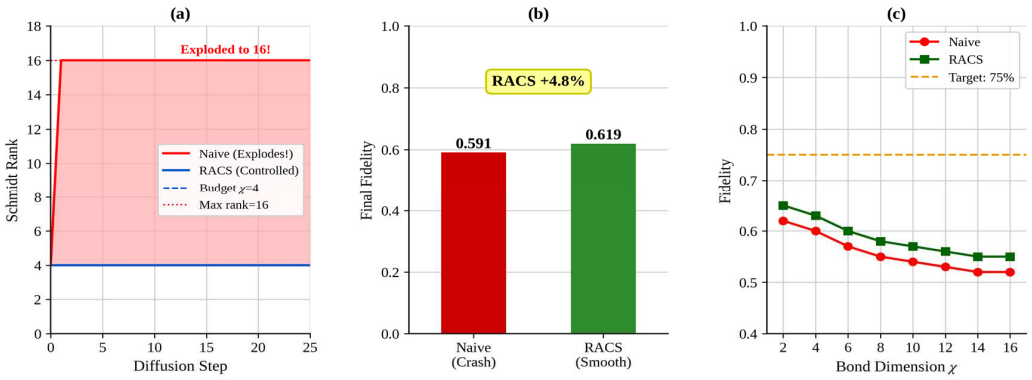

Figure 8. Consolidated Summary: RACS is the controlled Alternative to Native Truncation. (**a**) Rank trajectories show uncontrolled growth with truncation but stable control with RACS. (**b**) Final fidelity improves at $\chi = 4$ with RACS. (**c**) RACS consistently outperforms truncation in the low-to-mid χ regime.

4. Discussion

The results in Section 3 together show that naive truncation is incompatible with conditional latent diffusion on MPS representations, because it permits uncontrolled rank growth during sampling and enforces a single catastrophic collapse at the end. This failure mode manifests consistently across real handwritten digit generation, where conditional diffusion rapidly reaches the maximal effective rank and accumulates unrecoverable information loss under a fixed low-rank budget.

RACS addresses this instability by converting rank truncation from a one-shot operation into a controlled step-wise correction process. Across all analyses, the advantage of RACS is most prominent in the low-to-mid bond dimension regime, where the rank budget is tight and thus highly relevant to near-term resource limitations. In this regime, RACS improves reconstruction fidelity while simultaneously stabilizing entanglement growth and maintaining a smoother latent trajectory, indicating that the method functions as a structural correction consistent with the diffusion dynamics rather than a simple heuristic.

From a hardware perspective, the observed circuit complexity trends further support the practical motivation of rank-aware sampling. Since circuit depth and entangling gate counts grow rapidly with the bond dimension, improving fidelity under smaller χ directly translates into improved feasibility for NISQ-oriented deployment. Notably, RACS preserves quality without requiring χ to scale to the maximal rank, indicating that meaningful generation is possible under significantly reduced entanglement resources.

In addition, the synergy analysis suggests that robustness can be enhanced through a two-stage strategy: (i) reducing the initial entanglement pressure via optimal qubit mapping and (ii) actively suppressing rank explosion during diffusion using RACS. This "double defense" mechanism offers a general strategy for stabilizing tensor-network-based generative modeling when both representational capacity and hardware constraints must be satisfied simultaneously.

Despite these improvements, the present study focuses on rank-limited MPS diffusion in a relatively small-scale image domain. The 8-qubit system was strategically selected because it allows for the exact, unapproximated computation of the full state vector and effective Schmidt rank—metrics that become classically intractable to trace precisely for practical systems of 50+ qubits. While demonstrating practical quantum advantage ultimately requires scaling to complex datasets (e.g., CIFAR-10, quantum chemistry) and larger tensor networks, the 'rank explosion' phenomenon under conditional guidance is a fundamental structural bottleneck, not a dataset-specific artifact. In fact, as systems scale to higher-resolution data or complex domains, conditional guidance will inject even stronger global correlations, making rank-aware control exponentially more critical to maintain

hardware feasibility. Future work will explore scaling behavior to these broader conditional generation settings, as well as integrating more adaptive rank budgets and noise-aware correction rules. Nevertheless, the results suggest that rank-aware conditional sampling is a necessary ingredient for reliable tensor-network diffusion, and that RACS provides a principled and practical solution for controlling entanglement-driven failure modes.

5. Conclusions

This study examined a critical structural failure mode in conditional diffusion over tensor-network quantum states, namely rank explosion, in which conditional guidance rapidly drives the generated state beyond the efficient low-rank MPS manifold. Through real handwritten digit experiments, we showed that conditional sampling induces a drastic increase in effective Schmidt rank and entanglement complexity, which in turn leads to severe information loss when naïve end-stage truncation is applied.

To mitigate this issue, we proposed Rank-Aware Conditional Synthesis (RACS), a stepwise rank correction framework that actively constrains entanglement growth throughout the reverse diffusion trajectory. Empirical results demonstrated that RACS consistently improves reconstruction fidelity under the same rank budget, suppresses truncation-induced artifacts, and produces a significantly smoother latent trajectory compared to naïve truncation. Furthermore, by keeping the bond dimension small while preserving output quality, RACS substantially reduces compiled quantum circuit cost and enhances feasibility under NISQ-relevant resource constraints.

Overall, these findings indicate that rank-aware sampling is not optional but essential for stable and physically realizable conditional generation on tensor-network quantum states. RACS provides a principled mechanism to prevent manifold departure during diffusion, enabling conditional synthesis that is simultaneously structurally consistent, hardware-aware, and robust. Future work will explore scaling to higher-dimensional data, adaptive rank budgeting strategies, and broader classes of quantum generative tasks beyond handwritten digit distributions. Furthermore, integrating rank-controlled quantum states within distributed architectures—such as securing generated resources for Quantum Key Distribution (QKD) networks—presents a highly promising direction for practical quantum communication.

Author Contributions: Methodology, D.L. and W.-G.L.; Software, D.L.; Investigation, W.-G.L.; Resources, O.K.; Data curation, H.H.; Writing—original draft, D.L.; Writing—review & editing, O.K.; Visualization, H.H.; Supervision, O.K.; Project administration, O.K.; Funding acquisition, O.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2025S1A5C3A01010737).

Data Availability Statement: The data presented in this study are openly available in https://www.tensorflow.org/datasets/catalog/mnist (accessed on 30 March 2026).

Conflicts of Interest: Author Dongkyu Lee was employed by Quantum AI Task, LG AI Research, LG Electronics. Co., Ltd. Author Hyunjun Hong was employed by Neobio Inc. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851.
2. Dhariwal, P.; Nichol, A. Diffusion models beat GANs on image synthesis. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 8780–8794.
3. Ho, J.; Salimans, T. Classifier-free diffusion guidance. *arXiv* **2022**, arXiv:2207.12598. [CrossRef]

4. Zhang, B.; Xu, P.; Chen, X.; Zhuang, Q. Generative quantum machine learning via denoising diffusion probabilistic models. *arXiv* **2024**, arXiv:2310.05866. [CrossRef] [PubMed]
5. Schuld, M.; Sinayskiy, I.; Petruccione, F. An introduction to quantum machine learning. *Contemp. Phys.* **2015**, *56*, 172–185. [CrossRef]
6. Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum machine learning. *Nature* **2017**, *549*, 195–202. [CrossRef]
7. Hibat-Allah, M.; Mauri, M.; Carrasquilla, J.; Perdomo-Ortiz, A. A framework for demonstrating practical quantum advantage: Comparing quantum against classical generative models. *Commun. Phys.* **2024**, *7*, 68. [CrossRef]
8. Möttönen, M.; Vartiainen, J.J.; Bergholm, V.; Salomaa, M.M. Quantum circuits for general multiqubit gates. *Phys. Rev. Lett.* **2004**, *93*, 130502. [CrossRef]
9. Shende, V.V.; Bullock, S.S.; Markov, I.L. Synthesis of quantum-logic circuits. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **2006**, *25*, 1000–1010. [CrossRef]
10. Markov, I.L.; Shi, Y. Simulating quantum computation by contracting tensor networks. *SIAM J. Comput.* **2008**, *38*, 963–981. [CrossRef]
11. Cerezo, M.; Verdon, G.; Huang, H.Y.; Cincio, L.; Coles, P.J. Challenges and opportunities in quantum machine learning. *Nat. Comput. Sci.* **2022**, *2*, 567–576. [CrossRef] [PubMed]
12. Lloyd, S.; Weedbrook, C. Quantum generative adversarial learning. *Phys. Rev. Lett.* **2018**, *121*, 040502. [CrossRef] [PubMed]
13. Schollwöck, U. The density-matrix renormalization group in the age of matrix product states. *Ann. Phys.* **2011**, *326*, 96–192. [CrossRef]
14. Orús, R. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Ann. Phys.* **2014**, *349*, 117–158. [CrossRef]
15. Verstraete, F.; Murg, V.; Cirac, J.I. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Adv. Phys.* **2008**, *57*, 143–224. [CrossRef]
16. Vidal, G. Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.* **2003**, *91*, 147902. [CrossRef]
17. Rieser, H.-M.; Köster, F.; Raulf, A.P. Tensor networks for quantum machine learning. *Proc. R. Soc. A* **2023**, *479*, 20230218. [CrossRef]
18. Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2018**, *2*, 79. [CrossRef]
19. Kandala, A.; Mezzacapo, A.; Temme, K.; Takita, M.; Brink, M.; Chow, J.M.; Gambetta, J.M. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **2017**, *549*, 242–246. [CrossRef]
20. Yang, L.; Ding, S.; Cai, Y.; Yu, J.; Wang, J.; Shi, Y. Guidance with Spherical Gaussian Constraint for Conditional Diffusion. *arXiv* **2024**, arXiv:2402.03201. [CrossRef]
21. Holmes, Z.; Sharma, K.; Cerezo, M.; Coles, P.J. Connecting entanglement and expressibility in quantum circuits. *PRX Quantum* **2022**, *3*, 010313. [CrossRef]
22. Jung, H.; Lee, D.; Park, Y.; Schmid, L.; Kim, B.; Yun, S.-Y.; Jo, J.; Shin, J. Conditional synthesis of 3D molecules with time correction sampler. *arXiv* **2024**, arXiv:2411.00551. [CrossRef]
23. Benedetti, M.; Garcia-Pintos, D.; Perdomo-Ortiz, A.; Leyton-Ortega, V. A generative modeling approach for benchmarking and training shallow quantum circuits. *npj Quantum Inf.* **2019**, *5*, 45. [CrossRef]
24. Farhi, E.; Neven, H. Classification with quantum neural networks on near term processors. *arXiv* **2018**, arXiv:1802.06002. [CrossRef]
25. Jeon, H.; Lee, K.; Lee, D.; Kim, B.; Kim, T. Optimal qubit mapping search for encoding classical data into matrix product state representation with minimal loss. *arXiv* **2024**, arXiv:2406.06935. [CrossRef]
26. Temme, K.; Bravyi, S.; Gambetta, J.M. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.* **2017**, *119*, 180509. [CrossRef]
27. Cerezo, M.; Sone, A.; Volkoff, T.; Cincio, L.; Coles, P.J. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nat. Commun.* **2021**, *12*, 1791. [CrossRef]
28. Knill, E.; Leibfried, D.; Reichle, R.; Britton, J.; Blakestad, R.B.; Jost, J.D.; Langer, C.; Ozeri, R.; Seidelin, S.; Wineland, D.J. Randomized benchmarking of quantum gates. *Phys. Rev. A* **2008**, *77*, 012307. [CrossRef]