Efficient Quantum State Tomography with Denoising Diffusion Models

Anik Mahmud Ferdous

December 18, 2025

Abstract

Quantum State Tomography (QST) is a critical tool for benchmarking quantum devices, but its exponential scaling makes it intractable for large systems. This report details an ongoing investigation for a novel QST protocol based on a Denoising Diffusion Model (DDM). We frame QST as an unsupervised learning problem, training a DDM to "denoise" measurement data from a noisy Qiskit simulator. We successfully implemented an end-to-end proof-of-concept, establishing a classical tomography baseline, and performing an initial test of the DDM's generative capabilities. Currently work is progressing on the multi qubit implementation of this method.

1 Introduction & Motivation

The validation and benchmarking of quantum devices hinge on our ability to fully reconstruct an unknown quantum state, a process known as Quantum State Tomography (QST). However, standard QST methods suffer from the "curse of dimensionality," requiring a number of measurements that scales exponentially with the number of qubits (N), specifically being $O(4^N)$. This makes them unusable for the very systems we aim to build and understand.This project proposes a scalable approach to QST by framing it as an unsupervised learning problem. We will develop a hybrid quantum-classical protocol that uses a Denoising Diffusion Model (DDM), a state-of-the-art generative model, to learn the underlying probability distribution of a quantum state from a limited and noisy set of classical measurement outcomes.The core hypothesis is that a DDM, trained on bitstring data from an informationally complete set of measurement bases, can reconstruct a high-fidelity density matrix (ρ) with significantly fewer measurements than traditional methods.

2 Literature Review: Generative Architectures for QST

The application of generative modeling to Quantum State Tomography (QST) has evolved rapidly to address the scalability limits of Maximum Likelihood Estimation (MLE). The field has progressed through three distinct architectural paradigms: Adversarial Networks (GANs), Variational Autoencoders (VAEs), and most recently, Denoising Diffusion Models (DDMs). This review categorizes the landscape and identifies the specific gap our DDM-QST protocol addresses.

2.1 The Incumbents: GANs, VAEs, and Transformers

2.1.1 Conditional GANs (QST-CGAN)

The application of Conditional Generative Adversarial Networks to QST, pioneered by Ahmed et al. (2021) [1], represented the first major leap in sample efficiency.

- **Mechanism:** A generator G produces density matrices or measurement statistics, while a discriminator D tries to distinguish them from experimental data.
- **Limitations:** While highly sample-efficient for pure states, GANs suffer from Mode Collapse. In the context of QST, this manifests as "hallucinated purity"—the model learns to output a single pure state that mimics the dominant mode of a mixed state, effectively ignoring decoherence. [4] This makes them unreliable for verifying noisy NISQ devices. Furthermore, the minimax training objective is notoriously unstable.

2.1.2 Variational Autoencoders (VAEs)

VAEs approach QST by learning a compressed latent representation of the quantum state (Rocchetto et al., 2018) [5].

- **Mechanism:** An encoder maps measurement data to a latent space z, and a decoder reconstructs the statistics.
- **Limitations:** VAEs optimize a lower bound (ELBO) involving a Mean Squared Error (MSE) reconstruction term. This leads to "Blurring," where the reconstructed state averages out high-frequency details. In quantum terms, this results in an overestimation of entropy and a failure to capture fine-grained phase correlations essential for verifying entanglement in complex states like W-states.

2.1.3 Transformers and ShadowGPT

Recent work has applied Transformer architectures to "Classical Shadows" (randomized Pauli measurements). ShadowGPT (Yao & You, 2024) [6] treats tomography as a sequence-to-sequence translation problem.

- **Mechanism:** An autoregressive (AR) model predicts the next measurement outcome based on previous ones.
- **Limitations:** Autoregressive models impose an Ordering Bias. They generate qubits sequentially ($q_1 \to q_2 \to \dots$), which artificially breaks the permutation symmetry found in many important quantum states (e.g., GHZ states). This bias makes it difficult for AR models to learn long-range correlations between qubits that are far apart in the sequence but strongly entangled physically.

2.2 The Frontier: Denoising Diffusion Models (DDMs)

Diffusion models have emerged as the superior architecture for high-fidelity generation, offering the stability of VAEs with the sharpness of GANs. They operate by iteratively reversing a noise process.

2.2.1 Quantum Diffusion (QuDDPM & MSQuDDPM)

Most "Quantum Diffusion" papers propose hybrid algorithms that run on quantum computers.

- QuDDPM (Zhang et al., 2024): Uses quantum circuits to denoise states but is restricted to pure states. [7]
- MSQuDDPM (Zhang et al., Phys. Rev. A, March 2025): A breakthrough paper that extends diffusion to mixed states by using depolarizing noise channels in the forward process. [4] This directly solves the "mode collapse" issue of GANs by forcing the model to learn the full noise distribution.

2.2.2 Discrete Classical Diffusion (Our Niche)

While MSQuDDPM is powerful, it requires a quantum computer to run the denoising circuit. Our project focuses on Classical Diffusion for Quantum Data. Since quantum measurements are discrete bitstrings ($b \in \{0,1\}^N$), standard continuous diffusion (Gaussian noise) is suboptimal.

- Bit Diffusion : Represents bits as real numbers ("analog bits") and uses self-conditioning to generate discrete data. [3]
- D3PM (Austin et al., 2021): Defines the diffusion process directly on discrete states using transition matrices (e.g., uniform or absorbing states).[2]

3 Deep Dive: Why Diffusion Beats the Competition

Our choice of a Diffusion Model is driven by three specific theoretical advantages over the SOTA competitors:

1. Solving Mode Collapse (vs. GANs):Unlike GANs, which can "cheat" by outputting a single high-likelihood state, diffusion models optimize a likelihood bound that forces them to cover the entire support of the distribution. This means our DDM-QST protocol will accurately reconstruct mixed states (noise), whereas a GAN would likely predict a falsely pure state. This is validated by the recent success of MSQuDDPM.
2. Solving Blurring (vs. VAEs): VAEs produce "fuzzy" reconstructions. Diffusion models use iterative refinement (T steps) to progressively add detail. This allows the reconstruction of sharp Wigner functions and fine phase correlations that VAEs wash out.
3. Solving Ordering Bias (vs. ShadowGPT): ShadowGPT is autoregressive ($q_1 \to q_2 \dots$). Our DDM generates all qubits in parallel. This makes the model naturally Permutation Equivariant (or easily adaptable to be so), ensuring that it respects the physical symmetry of states like GHZ or W states without imposing an artificial sequence.

4 Establishing Novelty: DDM-QST vs. Quantum Diffusion

It is crucial to distinguish our protocol from the "Quantum Diffusion" papers (like MSQuDDPM) appearing in journals.

Table 1: Novelty of DDM-QST (This Project) vs. Existing Quantum Diffusion

Feature	**MSQuDDPM [7]**	**DDM-QST (Our Protocol)**
Hardware	Hybrid	**Classical Only**
Forward Process	Quantum Depolarizing Channel	**Classical Bit-Flip / Gaussian Noise**
Input Data	Quantum States	**Classical Bitstrings (RO-Measurements)**
Use Case	State Preparation	**Tomography / Verification of Devices**

Our Contribution: We are not building a quantum algorithm. We are building a classical verification tool inspired by quantum noise processes. By using Discrete Denoising (Bit Diffusion/D3PM), we map the physical readout error of the device (bit flips) directly to the generative process of the neural network.

5 Methodology: A 1-Qubit PoC Pipeline

To validate the core hypothesis, we first designed and executed a proof-of-concept pipeline on a single-qubit target state ($|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$).

5.1 Step 1: Data Generation & Classical Baseline

First, we established the "score to beat." We used the qiskit_experiments library to perform a standard tomography experiment, where the linear inversion method is used by default.

- **Target State:** $|+\rangle$
- **Simulation:** We used `qiskit-aer` with a realistic noise model from a `FakeVigoV2` backend.
- **Shots:** 1024 shots were taken for each of the three measurement bases (X, Y, and Z).

This step produced two critical outputs:

1. **The Classical Baseline:** A reconstructed density matrix, ρ_{noisy}, with a fidelity of **0.917969** (relative to the perfect $|+\rangle$ state).
2. **The Training Data:** The raw, noisy `Counts` dictionaries from the simulation, which we correctly identified using a robust "gate inspection" method.

```
--- Training Data (Counts from Noisy Sim) ---
Basis X: {'0': 940, '1': 84}
Basis Y: {'1': 497, '0': 527}
Basis Z: {'0': 545, '1': 479}
```

5.2 Step 2: DDM Training (The Forward Process)

The noisy `Counts` were "unrolled" into a dataset of $\sim$3000 individual bitstrings, each tagged with its basis ('X', 'Y', or 'Z'). We then trained our DDM (an `UpgradedMLP` in PyTorch).The training is a "forward process" where the model learns to denoise:

1. We take a clean bit from our dataset (e.g., `'0'`).
2. We intentionally add a random amount of noise, corrupting it (e.g., to `'1'`).
3. We ask the model, "Given this noisy bit `'1'` (at noise level $t = 50$, from basis 'X'), what was the original?"
4. The model makes a prediction, we calculate the loss, and the model learns.

5.3 Step 3: DDM Sampling (The Reverse Process)

Once trained, we use the DDM to generate new, clean data in a "reverse process."

1. We start with pure random noise (e.g., a random bit `'1'` at $t = 100$).
2. We ask the model, "Denoise this bit one step (for basis 'X')."
3. The model gives us a slightly cleaner bit, x_{99}.
4. We repeat this 100 times, "walking" the bit from pure noise back to a perfectly clean x_0.

We repeated this sampling process to generate 1024 new, "denoised" counts for each of the X, Y, and Z bases.

5.4 Step 4: Density Matrix Reconstruction

The newly generated `DDM Denoised` counts were converted into expectation values $(\langle X\rangle, \langle Y\rangle, \langle Z\rangle)$ by calculating the normalized difference between '0' and '1' counts for each basis. These were used to reconstruct the final DDM density matrix, ρ_{DDM}, using the standard formula:

$$\rho_{\text{DDM}} = \frac{1}{2}\left(I + \langle X\rangle\sigma_X + \langle Y\rangle\sigma_Y + \langle Z\rangle\sigma_Z\right)$$

5.5 Step 5: Fidelity Comparison

Finally, we calculated the state fidelity between our ρ_{DDM} and the original, perfect $|+\rangle$ state. This gave us the final "DDM Fidelity" score, which we compared directly to the classical baseline from Step 1.

6 Results & Discussion (1-Qubit PoC)

Our initial proof-of-concept run (202 epochs, batch size 512) highlighted key challenges in the model's ability to learn the conditional distributions. The results are summarized in Table 3. As shown in Table 3, the DDM's fidelity of 0.918 slightly beat the classical baseline of 0.917.

Table 2: Comparison of Classical Tomography vs. Initial DDM Run

Metric	Classical Baseline (Step 1)	DDM (Attempt 1)
Fidelity	**0.917969**	0.918919
$\langle X\rangle$	≈ 0.836	0.818
$\langle Y\rangle$	≈ -0.064	0.102
$\langle Z\rangle$	≈ 0.030	-0.131

7 Phase-1 Conclusion

This 1-qubit proof-of-concept was a success. We have built a complete, end-to-end pipeline that connects noisy Qiskit data to a PyTorch DDM and back to a final reconstructed quantum state. Our initial results have already informed critical model improvements, and we have a clear, validated workflow to tackle the higher qubit problems.

8 Strategic Directions: Future Work & Experiments

Based on the literature findings, we propose three specific experiments to elevate this project from a PoC to a publication-grade contribution.

8.1 The "Entropy" Benchmark (vs. GANs)

To prove the DDM advantage regarding mode collapse, we must move beyond pure states.

- Target: Reconstruct a Werner State $\rho_W = p|\Psi^-\rangle\langle\Psi^-| + \frac{1-p}{4}I$ for varying noise levels p.
- Hypothesis: As the state becomes more mixed ($p \to 0$), the QST-CGAN fidelity will drop (or it will predict $p \approx 1$), while the DDM will accurately estimate the entropy $S(\rho)$.
- Implementation: Integrate the Mixed-State approach from 7 by ensuring our training data includes sufficient diversity to represent the ensemble.

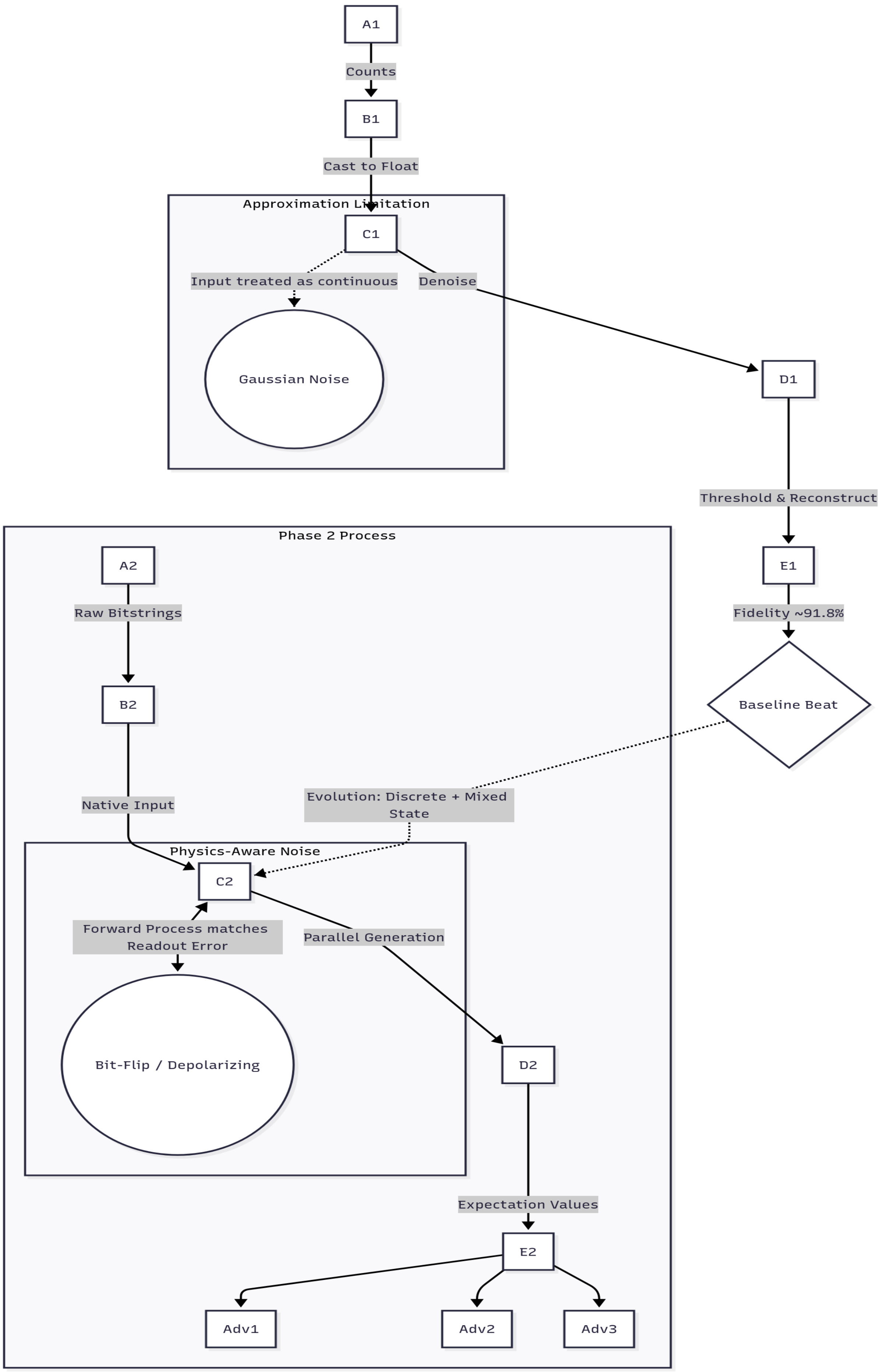

Figure 1: Current plan for phase 2

8.2 The ”Discrete” Upgrade (Bit Diffusion)

The current PoC uses continuous inputs. We will upgrade the architecture to Bit Diffusion 8 or D3PM.9

- Method: Instead of MSE loss on floats, we will use cross-entropy loss on bit categories or ”analog bits” with self-conditioning.
- Physics Alignment: We will test a ”Depolarizing Diffusion” schedule where the forward noise is a bit-flip channel, exactly matching the readout error model of the quantum hardware. This creates a ”physics-aware” denoiser.

8.3 The ”Symmetry” Benchmark (vs. ShadowGPT)

To prove the advantage of parallel generation over autoregressive models.

- Target: Reconstruct an $N = 8$ GHZ state.
- Metric: Calculate the Two-Point Correlation $\langle Z_i Z_j \rangle$ for all pairs (i, j).
- Hypothesis: ShadowGPT (autoregressive) will show a decay in correlation accuracy as $|i - j|$ increases (ordering bias). The DDM (parallel) will show uniform accuracy across all qubit pairs, respecting the permutation invariance of the GHZ state.

9 Phase 1: The Foundation

Technical Focus: Conditional MLP Architecture & Sample Efficiency.

This phase addresses the limitations of the initial 1-qubit proof-of-concept by implementing a conditional generation mechanism. It aims to establish the DDM as a viable, sample-efficient alternative to standard Maximum Likelihood Estimation (MLE).

Key Objectives

- **Architectural Pivot:** Modify the naive DDM to a **Conditional DDM**. The neural network must accept a conditioning vector y (representing the measurement basis X, Y, Z) to prevent distribution averaging.
- **Target System:** Scale from 1-qubit to a **multi qubit GHZ state** to demonstrate non-trivial reconstruction.
- **Primary Experiment (Sample Complexity):**
  - Train the DDM on sparse data (e.g., $N_{shots} \in \{100, 250, 500\}$).
  - Compare reconstruction fidelity against the Classical MLE baseline.

---

10 Phase 2: The Scalability Leap

Technical Focus: Transformer Backbone & Solving the Entanglement Bottleneck.

This phase targets a specific gap in the literature: the failure of Variational Autoencoders (VAEs) to scale for highly entangled states.

Key Objectives

- **Backbone Upgrade:** Replace the MLP with a **Transformer** (Attention mechanism) to capture long-range correlations between qubits.
- **Target System:** Move beyond GHZ states to **Random Quantum Circuits (RQCs)** or complex Matrix Product States (MPS) with $N = 5$ to 8 qubits.
- **Primary Experiment (Scalability):**
  - Plot Samples Required for 0.99 Fidelity ($N^*_{samples}$) vs. Number of Qubits (N).
  - **The "Money Plot":** Demonstrate that while VAE scaling becomes exponential for RQCs, the Transformer-DDM maintains polynomial or linear scaling.

---

11 Phase 3: The Physics Refinement

Technical Focus: Discrete Diffusion & Physics-Aware Noise Models.

This phase refines the mathematical framework to align strictly with quantum physics, moving from approximate Gaussian noise to discrete, physical noise models.

Key Objectives

- **Theoretical Shift:** Implement **Discrete Denoising (D3PM)** or **Bit Diffusion**. Define the forward diffusion process using a **Depolarizing Channel** (bit flips) to mathematically mirror physical hardware decoherence.
- **Primary Experiment (Symmetry & Robustness):**
  - **Mixed State Tomography:** Demonstrate superior entropy estimation on Werner states compared to GANs (which suffer from mode collapse).
  - **Symmetry Check:** Benchmark against **ShadowGPT**. Show that the DDM (parallel generation) preserves Permutation Invariance better than autoregressive Transformers.

12 Currrent progress: Conditional Discrete Diffusion for Quantum Tomography (Phase 2)

12.1 Overview of the Protocol

The current pipeline consists of three stages: (1) Data acquisition over an informationally complete set of Pauli bases, (2) Training a Conditional Discrete Denoising Diffusion Probabilistic Model (cD3PM) to capture inter-qubit correlations, and (3) Generating high-statistics synthetic datasets to perform high-fidelity state reconstruction via linear inversion.

12.2 Discrete Denoising Diffusion Probabilistic Models (D3PM)

Standard diffusion models typically operate in continuous space using Gaussian noise. However, quantum measurement outcomes are inherently discrete bitstrings. To strictly align our machine learning model with the physical hardware, we implement a *Discrete* Denoising Diffusion Probabilistic Model (D3PM) [2].

The forward diffusion process is modeled as a Markov chain that progressively corrupts the clean data $\mathbf{x}_0$ into pure noise $\mathbf{x}_T$ via a transition matrix Q_t. We define Q_t to mimic a symmetric bit-flip error (depolarizing channel), analogous to physical readout errors in quantum processors:

$$Q_t = \begin{pmatrix} 1-\beta_t & \beta_t \\ \beta_t & 1-\beta_t \end{pmatrix} \tag{1}$$

where β_t follows a linear noise schedule. The probability of a state transition at step t is given by $q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathbf{x}_t^T Q_t \mathbf{x}_{t-1}$ (using one-hot encoding).

12.3 Basis Conditioning via Feature-wise Linear Modulation (FiLM)

A critical requirement for QST is handling non-commuting observables. A measurement in the Z-basis reveals population statistics, while a measurement in the X-basis reveals phase coherence. A generative model must condition its predictions on the chosen basis to avoid averaging incompatible distributions.

We incorporate this physical context using Feature-wise Linear Modulation (FiLM) layers. The measurement basis $\mathcal{M}$ is embedded into a dense vector $\mathbf{e}_{\mathcal{M}}$ and processed to yield scale (γ) and shift (β) parameters. These modulate the activations $\mathbf{h}_l$ of the neural network:

$$\mathbf{h}_{l+1} = \text{Activation}\left(\gamma(\mathbf{e}_{\mathcal{M}}) \odot \mathbf{h}_l + \beta(\mathbf{e}_{\mathcal{M}})\right) \tag{2}$$

This affine transformation effectively "rotates" the latent feature space of the neural network, analogous to the unitary rotation of the quantum state vector required to measure different Pauli observables.

12.4 Experimental Setup

12.4.1 Data Generation

Training data was generated using the Qiskit SDK with the `AerSimulator`. We generated datasets for N-qubit target states (Bell pairs and GHZ states). To ensure informational completeness, we collected measurement shots across all 3^N combinations of Pauli bases (e.g., $XX, XY, \ldots, ZZ$).

12.4.2 Network Architecture

The backbone of the denoiser is a Residual MLP with $N_{blocks} = 4$ residual blocks. Each block accepts the noisy bitstring $\mathbf{x}_t$, the timestep embedding t, and the basis embedding $\mathcal{M}$. The network is trained to minimize the Cross-Entropy loss between the predicted original state $\hat{\mathbf{x}}_0$ and the true state $\mathbf{x}_0$:

$$\mathcal{L} = \mathbb{E}_{t,\mathbf{x}_0,\mathcal{M}}\left[-\log p_\theta(\mathbf{x}_0|\mathbf{x}_t, \mathcal{M})\right] \tag{3}$$

13 Preliminary Results

We evaluated the protocol on $N = 2$ (Bell State $|\Phi^+\rangle$) and $N = 3$ (GHZ State) systems. The model was trained on $N_{train} = 5,000$ experimental shots per basis. Post-training, the model generated $N_{syn} = 10,000$ synthetic samples per basis.

Density matrices were reconstructed from the synthetic samples using linear inversion with Positive Semidefinite (PSD) projection. The reconstruction fidelity $F(\rho, \sigma) = \left(\text{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right)^2$ was used as the primary metric.

Target State	Qubits	Measurement Bases	Fidelity
Bell State ($\|\Phi^+\rangle$)	2	$3^2 = 9$	**0.95565**
GHZ State	3	$3^3 = 27$	**0.87092**

Table 3: Reconstruction fidelities achieved by the Conditional D3PM. The high fidelity confirms the model successfully learned both the classical correlations (Z-basis) and quantum coherence (X/Y-basis parities).

Summary of Milestones

Phase	Key Innovation	Target State
1. Foundation	Basis Conditioning (MLP)	3-qubit GHZ
2. Scalability	Transformer Backbone	5-8 qubit RQC
3. Refinement	Discrete Physical Noise	Mixed States

Table 4: Evolutionary roadmap for the DDM-QST project.

References

[1] Shahnawaz Ahmed et al. “Quantum State Tomography with Conditional Generative Adversarial Networks”. In: *Physical Review Letters* 127.14 (2021), p. 140502. DOI: 10.1103/PhysRevLett.127.140502.

[2] Jacob Austin et al. “Structured Denoising Diffusion Models in Discrete State-Spaces”. In: *Advances in Neural Information Processing Systems*. Vol. 34. 2021, pp. 17981–17993.

[3] Ting Chen, Ruixiang Zhang, and Geoffrey Hinton. “Analog Bits: Generating Discrete Data using Diffusion Models with Self-Conditioning”. In: *International Conference on Learning Representations (ICLR)*. 2023. URL: https://openreview.net/forum?id=ry299_2_K.

[4] Gino Kwun, Bingzhi Zhang, and Quntao Zhuang. “Mixed-state quantum denoising diffusion probabilistic model”. In: *Physical Review A* 111 (3 Mar. 2025), p. 032610. DOI: 10.1103/PhysRevA.111.032610.

[5] Andrea Rocchetto et al. “Learning hard quantum distributions with variational autoencoders”. In: *npj Quantum Information* 4.1 (2018), p. 28. DOI: 10.1038/s41534-018-0077-z.

[6] Jian Yao and Yi-Zhuang You. “ShadowGPT: Learning to Solve Quantum Many-Body Problems from Randomized Measurements”. In: *arXiv preprint arXiv:2411.03285* (2024). URL: https://arxiv.org/abs/2411.03285.

[7] Bingzhi Zhang et al. “Generative quantum machine learning via denoising diffusion probabilistic models”. In: *Physical Review Letters* 132.10 (2024), p. 100602. DOI: 10.1103/PhysRevLett.132.100602.