

# QAMP 2025

*Enhancing Quantum Diffusion Models for Complex Image Generation*

Jeongbin Jo

jeongbin033@yonsei.ac.kr

December 8, 2025

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	What's the potential benefit of replacing neural network with PQCs? . . . . .	2
1.2	How does the proposed QDM compare to just classical DM? . . . . .	2
1.3	Diffusion Model . . . . .	3
1.3.1	Classical Diffusion Model . . . . .	3
1.3.2	Forward Process: Diffusion Process . . . . .	3
1.3.3	Reverse Process: Generative Process . . . . .	3
1.3.4	Loss Function . . . . .	3
1.3.5	Quantum Diffusion Model . . . . .	4
1.3.6	Forward Process: Depolarizing Channel . . . . .	4
1.3.7	Reverse Process: Quantum Denoising . . . . .	4
1.3.8	Loss Function: Fidelity or Trace Distance . . . . .	5
<b>2</b>	<b>Implementation</b>	<b>5</b>
2.1	Architecture . . . . .	5

# 1 Intrduction

## 1.1 What’s the potential benefit of replacing neural network with PQCs?

### Parameter Efficiency (Model Compression)

The authors state in their conclusion that "quantum diffusion models require fewer parameters with respect to their classical counterpart". For example, their 'conditioned latent' model, which generated all 10 MNIST digits, used only 1110 parameters. This is a massive reduction compared to the millions of parameters typical in classical U-Net architectures used for diffusion.

### Expressivity and Scalability

The paper suggests that for a "full quantum case," it is "possible to generate distributions whose features scale exponentially with the number of qubits". This implies PQCs have a fundamentally greater expressive power.

### Future Potential

The long-term benefit is the potential to "approximate probability distributions that are not classically tractable", including "quantum datasets" themselves.

## 1.2 How does the proposed QDM compare to just classical DM?

### Hybrid Architecture

The paper’s QDM is a hybrid model. The forward process (adding noise) is performed classically. The backward process (denoising) is performed by a PQC.

### PQC as a Denoising Operator

Instead of having the PQC predict the noise  $\epsilon_\theta$  (like a classical U-Net), this model trains the PQC to be a direct denoising operator  $P(\theta, t)$  such that  $**P(\theta, t)|x_t\rangle = |x_{t-1}\rangle**$ . This avoids the "impractical" cycle of decoding/encoding quantum states at every timestep.

### Use of Complex Noise

Because PQC operations can make state coefficients complex, the authors found that using complex Gaussian noise ( $\epsilon = \epsilon_r + i\epsilon_i$ ) during the classical forward process led to a "significant improvement in performance". This is a notable departure from classical models.

## Comparable Performance (in Latent Space)

The paper’s ‘Latent QDM’ (which uses a classical autoencoder first) achieved a competitive FID of 38.2 and was able to generate ”samples of similar quality to those generated by classical algorithms”.

## 1.3 Diffusion Model

### 1.3.1 Classical Diffusion Model

Diffusion models are a class of generative models in machine learning inspired by non-equilibrium thermodynamics. The goal is to learn a diffusion process that destroys structure in data, and then learn a reverse process that restores structure to generate new data samples from noise.<sup>[1]</sup>

### 1.3.2 Forward Process: Diffusion Process

The forward process is defined by a Markov chain that gradually adds Gaussian noise to the data according to a variance schedule  $\beta_1, \dots, \beta_T$ .

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t\mathbf{I}\right) \quad (1)$$

Using the notation  $\alpha_t = 1 - \beta_t$  and  $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$ , we can sample  $x_t$  at any arbitrary time step  $t$  directly from  $x_0$  in closed form:

$$q(x_t|x_0) = \mathcal{N}\left(x_t; \sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)\mathbf{I}\right) \quad (2)$$

This property allows for efficient training by eliminating the need to iterate through all previous timesteps.

### 1.3.3 Reverse Process: Generative Process

The generative process is defined as the reverse Markov chain. Since the exact reverse posterior  $q(x_{t-1}|x_t)$  is intractable, we approximate it using a neural network with parameters  $\theta$ :

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \quad (3)$$

Here,  $\mu_\theta(x_t, t)$  is the predicted mean, and  $\Sigma_\theta(x_t, t)$  is the covariance (often fixed to  $\beta_t\mathbf{I}$  or learned).

### 1.3.4 Loss Function

Training is performed by optimizing the variational lower bound (VLB) on the negative log-likelihood.

$$\mathcal{L} = \mathbb{E}[-\log p_\theta(\mathbf{x}_0)] \leq \mathbb{E}_q \left[ -\log \frac{p_\theta(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \right] \quad (4)$$

This can be rewritten as a sum of KL divergence terms:

$$\mathcal{L} = \mathbb{E}_q \left[ \underbrace{D_{KL}(q(\mathbf{x}_T|\mathbf{x}_0)||p(\mathbf{x}_T))}_{L_T} + \sum_{t>1} \underbrace{D_{KL}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}} \underbrace{- \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0} \right] \quad (5)$$

In practice, Ho et al. found that a simplified objective function yields better sample quality. The objective is to predict the noise  $\epsilon$  added to  $x_0$ :

$$\mathcal{L}_{\text{simple}}(\theta) = \mathbb{E}_{t, x_0, \epsilon} [\|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, t)\|^2] \quad (6)$$

where  $\epsilon \sim \mathcal{N}(0, \mathbf{I})$  and  $\epsilon_\theta$  is a function approximator (e.g., U-Net).

### 1.3.5 Quantum Diffusion Model

While classical diffusion models operate on probability distributions over classical data, Quantum Diffusion Models (QDMs) extend this concept to the Hilbert space of quantum states. The goal is to generate quantum states (density matrices) from a maximally mixed state.

### 1.3.6 Forward Process: Depolarizing Channel

Instead of adding Gaussian noise, the forward process in QDM is typically modeled as a standard depolarizing channel acting on a density matrix  $\rho$ . For a system of  $n$  qubits with dimension  $d = 2^n$ , the state at step  $t$  is given by:

$$\rho_t = \mathcal{E}_t(\rho_{t-1}) = (1 - p_t)\rho_{t-1} + p_t \frac{I}{d} \quad (7)$$

where  $p_t \in [0, 1]$  is the depolarization probability (noise schedule). Similar to the classical case, we can express  $\rho_t$  directly from the initial state  $\rho_0$ . Let  $\alpha_t = \prod_{s=1}^t (1 - p_s)$ , then:

$$\rho_t = \alpha_t \rho_0 + (1 - \alpha_t) \frac{I}{d} \quad (8)$$

As  $t \rightarrow T$ ,  $\alpha_T \rightarrow 0$ , and the state converges to the maximally mixed state  $\rho_T \approx \frac{I}{d}$ , which contains no information about  $\rho_0$ .

### 1.3.7 Reverse Process: Quantum Denoising

The reverse process aims to restore the quantum state from the noise. This is modeled by a parameterized quantum circuit (PQC), denoted as a unitary operator  $U(\theta)$ . The discrete reverse step can be approximated as:

$$\rho_{t-1} \approx \mathcal{D}_\theta(\rho_t, t) = U(\theta_t)\rho_t U^\dagger(\theta_t) \quad (9)$$

For more complex generative tasks, the reverse process may involve ancillary qubits and measurements to simulate non-unitary maps.

### 1.3.8 Loss Function: Fidelity or Trace Distance

Unlike the KL divergence used in classical models, quantum models often use Fidelity or Hilbert-Schmidt distance to measure the closeness between the generated state and the target state. A common objective is to maximize the overlap with the target pure state  $|\psi\rangle$ :

$$\mathcal{L}(\theta) = 1 - \mathbb{E} [\langle \psi | \rho_{\text{gen}}(\theta) | \psi \rangle] \quad (10)$$

Alternatively, for density matrices, we minimize the trace distance or maximize the quantum fidelity  $F(\rho, \sigma) = (\text{tr} \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}})^2$ .

## 2 Implementation

### 2.1 Architecture

## References

- [1] *Diffusion Model*. URL: [https://en.wikipedia.org/wiki/Diffusion\\_model](https://en.wikipedia.org/wiki/Diffusion_model) (visited on 12/07/2025).