# Quantum Diffusion Model for Quark and Gluon Jet Generation

**Mariia Baidachna[1*], Rey Guadarrama[2], Gopal Ramesh Dahale[3], Tom Magorsch[4], Isabel Pedraza[5], Konstantin T. Matchev[6], Katia Matcheva[6], Kyoungchul Kong[7], Sergei Gleyzer[6]**

[1]School of Science and Engineering, University of Glasgow, Glasgow, UK
[2]School of Physical and Mathematical Sciences, Benemérita Universidad Autónoma de Puebla, Puebla, Mexico
[3]EPFL Swiss Federal Technology Institute of Lausanne, Lausanne, Switzerland
[4]Technische Universität Dortmund, Dortmund, Germany
[5]CERN European Organization for Nuclear Research, Geneva, Switzerland
[6]Department of Physics and Astronomy, University of Alabama, Tuscaloosa, AL 35487, USA
[7]Department of Physics and Astronomy, University of Kansas, Lawrence, KS 66045, USA

**Abstract**

Diffusion models have demonstrated remarkable success in image generation, but they are computationally intensive and time-consuming to train. In this paper, we introduce a novel diffusion model that benefits from quantum computing techniques in order to mitigate computational challenges and enhance generative performance within high energy physics data. The fully quantum diffusion model replaces Gaussian noise with random unitary matrices in the forward process and incorporates a variational quantum circuit within the U-Net in the denoising architecture. We run evaluations on the structurally complex quark and gluon jets dataset from the Large Hadron Collider. The results demonstrate that the fully quantum and hybrid models are competitive with a similar classical model for jet generation, highlighting the potential of using quantum techniques for machine learning problems.

## Introduction

Denoising diffusion models (DDMs) have revolutionized the field of generative artificial intelligence (GenAI) by demonstrating their ability to generate high-quality images (Cao et al. 2024). They overcome the drawbacks of generative adversarial networks (GANs), which are prone to mode collapse, becoming a new state-of-the-art architecture for image generation (Dhariwal and Nichol 2021; Stypułkowski et al. 2024). Consequently, DDMs have been applied in many generative tasks for science from molecular biology to medical image synthesis to gravitational lensing (Xu et al. 2023; Baidachna et al. 2024; Yi et al. 2024; Pan et al. 2023; Reddy et al. 2024).

High-energy physics (HEP) experiments, such as those at the Large Hadron Collider (LHC), generate massive datasets of subatomic particle collisions. The pursuit of deriving robust insights from the datasets to answer fundamental questions about our Universe has hitherto relied on supervised pipelines using expert annotations or simulated labeling. However, these approaches suffer from multiple limitations: expensive data acquisition and reliance on human-labeled or simulated ground truth.

*Corresponding author: 2828197b@student.gla.ac.uk


This makes diffusion models an appealing candidate for unsupervised or weakly supervised generative modeling in HEP, where the ability to model complex, high-dimensional distributions—such as calorimeter and tracker outputs—can aid in data augmentation, anomaly detection, and simulation refinement. However, a major drawback of DDMs is their extensive computational cost. Training typically requires thousands of iterations over the noising and denoising processes (Song, Meng, and Ermon 2020; Kawar et al. 2022), making it challenging to scale these models for large physics datasets.

To overcome this bottleneck, new computational paradigms are needed. One promising direction is quantum machine learning (QML) (Schuld, Sinayskiy, and Petruccione 2015), which leverages the unique properties of quantum mechanics—such as superposition and entanglement—to accelerate certain classes of machine learning tasks. Quantum computers can, in principle, perform operations on exponentially large state spaces more efficiently than classical counterparts (Ciliberto et al. 2018), opening the door for hybrid quantum-classical models that could reduce inference time, improve denoising accuracy, or scale better with data dimensionality.

In this work, we investigate this intersection by evaluating classical, hybrid, and quantum diffusion models on a challenging HEP dataset: quark vs. gluon jet classification. By adapting the diffusion framework to jet detector data and integrating quantum modules in the denoising process, we aim to explore both the performance gains and the architectural trade-offs introduced by quantum enhancements in realistic scientific settings.

## Background

### Denoising Diffusion Model

There have been multiple variants of a DDM proposed, such as denoising diffusion probabilistic model (DDPM) (Sohl-Dickstein et al. 2015; Ho, Jain, and Abbeel 2020) and denoising diffusion implicit model (DDIM) (Song, Meng, and Ermon 2020), but we generalize these into DDMs with the common factors being a noising scheduling algorithm and a learned denoising process.

### Forward Diffusion Process

The forward diffusion process gradually adds Gaussian noise to the data, leading to a series of latent variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$. The forward process is defined as:

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t \mid \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}), \quad (1)$$

where $\beta_t$ is the variance schedule that controls the amount of noise added at each step $t$.

Starting from the original data distribution $q(\mathbf{x}_0)$, the joint distribution over the sequence of latent variables is given by:

$$q(\mathbf{x}_{1:T} \mid \mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t \mid \mathbf{x}_{t-1}), \quad (2)$$

and the marginal distribution of any $\mathbf{x}_t$ given $\mathbf{x}_0$ is:

$$q(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t \mid \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}), \quad (3)$$

where $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$.

### Reverse Diffusion Process

The reverse diffusion process is modeled as:

$$p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1} \mid \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \sigma_t^2 \mathbf{I}), \quad (4)$$

where $\boldsymbol{\mu}_\theta(\mathbf{x}_t, t)$ is a neural network parameterized by $\theta$ that is trained to predict the mean, and $\sigma_t^2$ is typically set to $\beta_t$ or some other small variance.

The training objective is to minimize the variational lower bound on the negative log-likelihood:

$$\begin{aligned} L_{\text{vlb}} = E_q \Bigg[ & D_{\text{KL}}(q(\mathbf{x}_T \mid \mathbf{x}_0) \,\|\, p(\mathbf{x}_T)) \\ & + \sum_{t=2}^{T} D_{\text{KL}}(q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0) \,\|\, p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)) \\ & - \log p_\theta(\mathbf{x}_0 \mid \mathbf{x}_1) \Bigg] \end{aligned} \quad (5)$$

where $D_{\text{KL}}$ denotes the Kullback-Leibler divergence, though the loss is often simplified to the mean squared error (MSE).

### Quantum Theory for Machine Learning

**Qubit States and Measurement:** Quantum bits, or qubits, can exist in superpositions of states, allowing them to represent more information and process it more efficiently than classical bits (Nielsen and Chuang 2002). Unlike a classical bit that can be either 0 or 1, a qubit can be in a superposition of both states simultaneously. Mathematically, the state of a qubit can be written as:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad (6)$$

where $|0\rangle$ and $|1\rangle$ are the basis states, and $\alpha$ and $\beta$ are complex numbers representing the probability amplitudes. These amplitudes are normalized so that:

$$|\alpha|^2 + |\beta|^2 = 1. \quad (7)$$

When a measurement is made on a qubit, the superposition collapses to one of the basis states. The probability of measuring the state $|0\rangle$ is $|\alpha|^2$, and the probability of measuring the state $|1\rangle$ is $|\beta|^2$.

One method of measurement relevant in quantum computing and QML is the Haar measurement. Haar measurement involves sampling unitary operations uniformly according to the Haar measure, which is the unique, invariant measure on the group of unitary matrices. This type of measurement is useful in QML because it provides a way to generate random quantum states and operations, which is important for algorithms that require randomization (Haar 1933).

**Variational Quantum Circuits in Machine Learning:** Variational quantum circuits (VQCs) are a central tool in QML used in various architectures (Chen et al. 2020; Griol-Barres et al. 2021; Romero and Aspuru-Guzik 2021; Shu et al. 2024; Comajoan Cara et al. 2024). A VQC is a parameterized quantum circuit where a predefined set of the gates depends on adjustable parameters. These parameters are represented by quantum gates in the unitary operations. The general form of a VQC can be represented as:

$$|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|0\rangle, \quad (8)$$

where $U(\boldsymbol{\theta})$ is a unitary operation that depends on the parameters $\boldsymbol{\theta}$.

In the context of machine learning, these parameters are optimized using classical optimization techniques to minimize the cost function that measures the performance of the circuit throughout training. The power of VQCs in machine learning arises from their ability to exploit superposition and entanglement to potentially represent and solve problems more efficiently than classical algorithms (Biamonte et al. 2017).

## Related Work

Multiple works have combined the classical DDM and quantum algorithms. The paper in (Zhang et al. 2024) proposed a fully quantum DDPM (QuDDPM) for generating an unknown quantum state distribution. Their method consists of applying random unitaries to quantum states, thus scrambling them into noise, and summing three error functions to optimize during training. The QuDDPM performed best compared to a quantum GAN and a quantum direct transport model.

Another work in (De Falco et al. 2024) proposed a hybrid quantum-classical DDM. The algorithm involved a quantum denoising U-Net and classical noising and optimization. Good results were achieved on the MNIST dataset. However, tests on MNIST are not generalizable to other data, and a physics-conscious approach is required for the more complex Quark-Gluon data.

Other recent work such as (Schuld and Killoran 2019; Beer et al. 2020) explores the theoretical and practical aspects of hybrid QML, including training strategies for parameterized quantum circuits and data encoding methods.

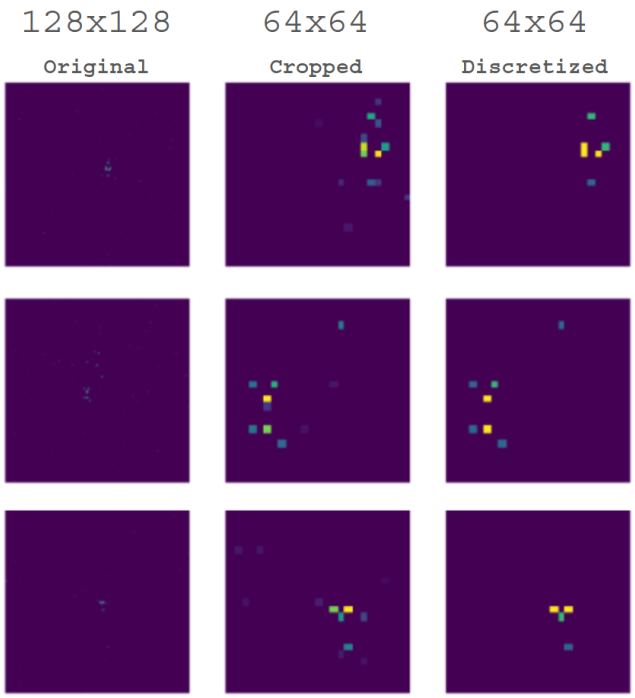


Figure 1: Preprocessing of three of the original 128x128 samples into 64x64 samples.

However, a practical fusion of quantum denoising with physically grounded diffusion processes, especially for use in high-energy physics, is still an emerging research area.

### Dataset and Preprocessing Description

In this work, we generate quark and gluon jet data from the open-source LHC Compact Muon Solenoid (CMS) detector data. The data for our model contains two classes that follow different distributions: hits from quark and gluon jets. Each sample is captured by three CMS subdetectors: electromagnetic calorimeter (ECAL), hadronic calorimeter (HCAL), and the reconstructed tracks as described in (Andrews et al. 2020). The image is cropped and minimally discretized as shown in Figure 1. This was done to increase quantum simulation computing speed and shift the focus to experimentally deriving the best architecture.

## Methodology

In this section, we outline the methods used to construct fully quantum, hybrid, and fully classical models. The pipeline in Figure 2 shows all the models and their combinations.

### Quantum Embedding

The first step of the pipeline is embedding classical data into quantum. We implemented angle encoding with $Rx$ rotation gates in groups of four pixels on the ECAL-detected images. Each sample is split into four channels, as shown in Figure 3.

### Forward Quantum Diffusion

The forward diffusion process was inspired by the scrambling implementation in (Zhang et al. 2024) using Haar random unitaries to mimic random noise application, as shown in Figure 3. Since the choice of forward scrambling does not significantly impact model performance, as the authors in (Bansal et al. 2024) have shown, an arbitrary noising transformation can be used. We chose to use the Haar measure for the quantum model as it allows for the unitary matrix transforms to scale with increasing resolution and dataset size. The final unitary is applied to each encoded channel, avoiding costly calculations associated with each timestep.

### Denoising Quantum Network

The goal of the quantum models is to learn the parameters of the quantum circuits throughout training by minimizing loss. Similar to classical DDMs, the denoising neural network is trained on learned parameters with the MSE loss function. The hybrid model uses a quantum strongly entangling layer surrounded by fully convolutional layers, and the fully quantum model relies only on the quantum layers. The circuit in the quantum layer, consisting of rotation and strongly entangling gates, is kept uniform across all models, with the number of layers treated as a tunable parameter.

## Experiments

In each experiment, we compare the performance using the Fréchet Inception Distance (FID) function, originally introduced in (Heusel et al. 2017) for evaluating GANs, along with the loss function. The FID is defined as:

$$\text{FID}(x, g) = \|\mu_x - \mu_g\|_2^2 + \text{Tr}(\Sigma_x + \Sigma_g - 2(\Sigma_x\Sigma_g)^{1/2}),$$

where $\mu_x$ and $\mu_g$ are the mean feature vectors of the real and generated images, respectively, and $\Sigma_x$ and $\Sigma_g$ are their corresponding covariance matrices. Though we do not currently have access to quantum computers, we can simulate the behavior of quantum systems using simulators like Pennylane (Bergholm et al. 2018) and compare the results to classical models.

First, we present the loss and FID functions of the classical, hybrid, and quantum models, respectively, in Figure 4. Training on 50 epochs seems to be sufficient to reach convergence for all models.

| Model Type | FID Score |
|---|---|
| Classical | 2.09 |
| Hybrid | 1.62 |
| Quantum | 2.11 |

Table 1: FID scores for classical, hybrid, and quantum models. FID score was calculated for the final generated distribution of 1k images against randomly selected 1k original images. The starting FID for each model of 64x64 images is around 3000. The similarity between distributions increases significantly with training.

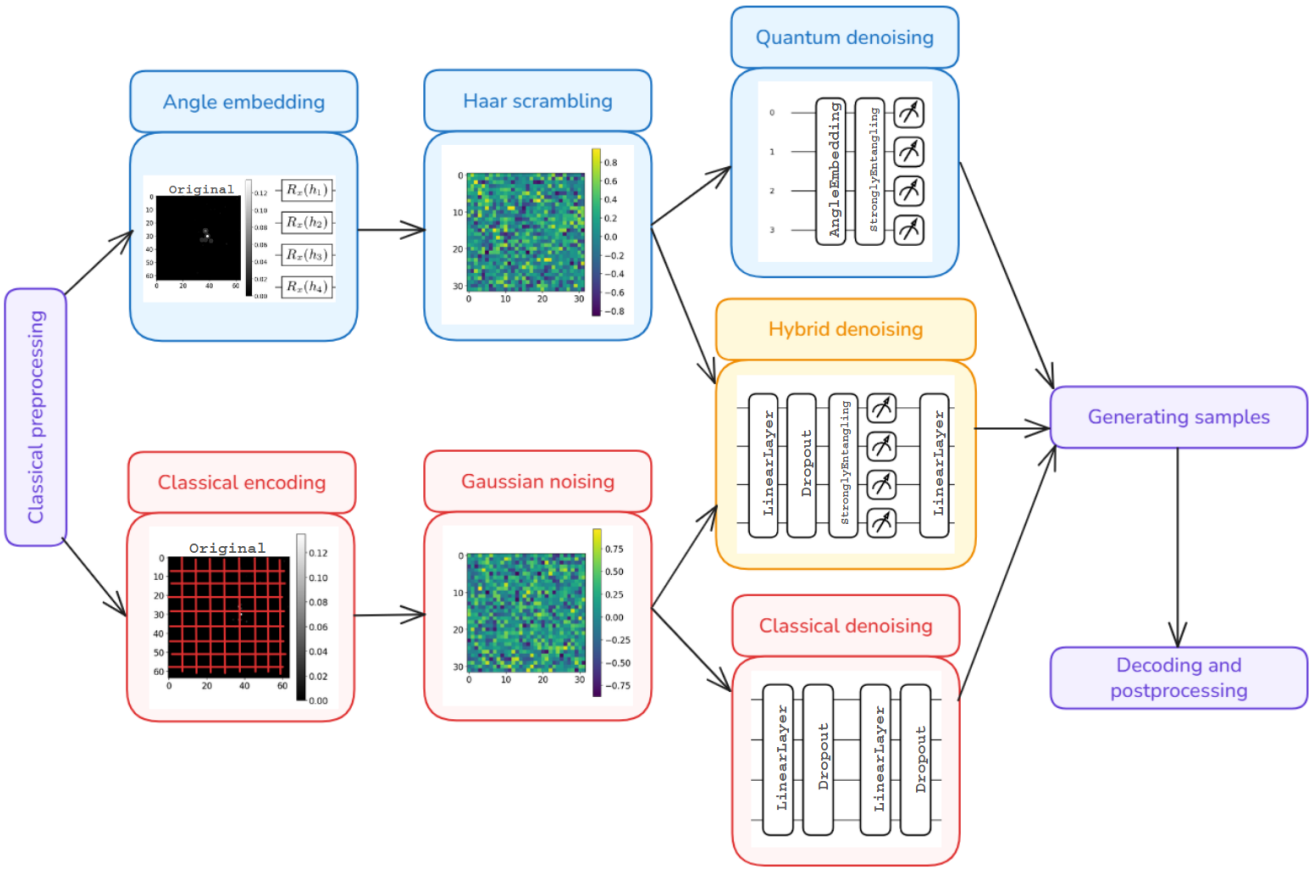


Figure 2: The pipeline that the data goes through with all the possible classical-quantum combinations of the forward and backward diffusion process.

## Discussion

All models exhibited a significant decrease in loss, approaching near-zero values, which is indicative of effective learning. The FID scores followed a similar downward trajectory, with final values of 2.09, 1.62, and 2.11. These results show that the performance of the models that leverage quantum circuits is comparable in relation to a structurally similar classical model. This implies that all or some parts of deep neural network computations can be offloaded to faster quantum processors to reduce training time without a performance trade-off. In fact, the hybrid model may be more expressive than its counterparts due to the optimization of several classical layers in addition to quantum, but this marginal difference in FID may be the result of random noise and needs to be investigated more closely.

A potential reason for the plateau in FID scores across all models could be the sparsity of the data, where each sample contains only a few non-zero points. This sparsity may prevent the complete removal of noise in some of the encoded channels. A possible solution is a post-processing step where only the most prominent values are retained in the decoded data, increasing the confidence in jet locations. Taking the most likely jet values, we calculated the structural similarity metric to augment the quantitative results of the FID.

Additionally, it is important to note that all quantum models in this work were simulated, which introduces a significant computational overhead compared to their classical counterparts. Since the simulations involve complex matrix manipulations, especially at scale, training is an order of magnitude slower. As a result, it is substantially more challenging to do simple architectural experimentation and hyperparameter tuning with quantum models. Despite this inherent disadvantage, quantum models show strong performance with minimal tuning, highlighting the potential of QML.

## Conclusion

This work marks significant progress in leveraging quantum computing for machine learning applications, specifically through the development and evaluation of a quantum-enhanced DDM. By comparing classical, hybrid, and fully quantum approaches on a non-trivial high-energy physics dataset, we demonstrate the potential of quantum components to contribute meaningfully to generative modeling, particularly in scientific domains.

Looking ahead, we aim to extend our framework to generate all three jet channels for each sample, enabling richer and more physically complete simulations. Further exploration

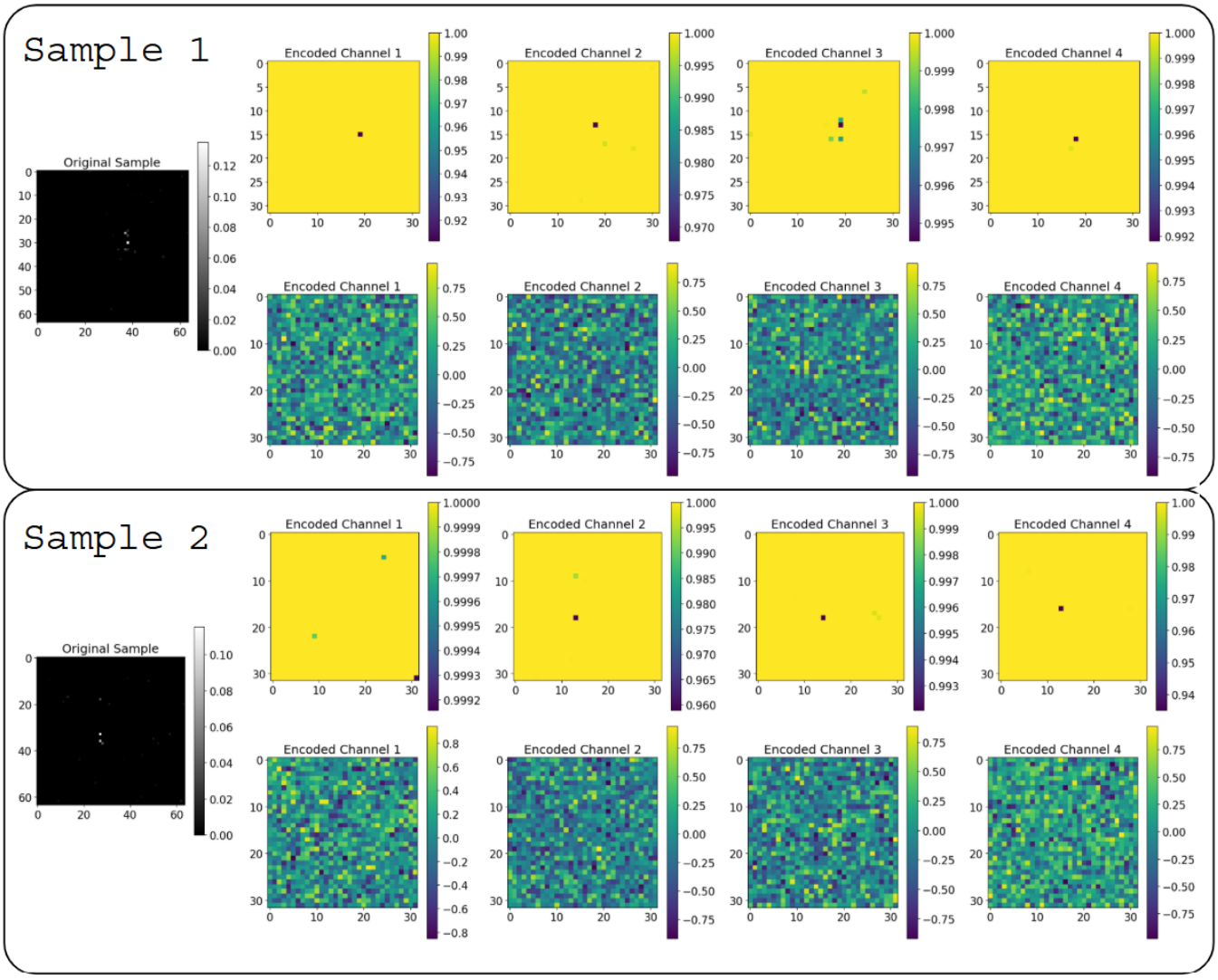


Figure 3: Two samples of the original raw image encoded into four channels using angle embedding and fully scrambled using Haar measure.

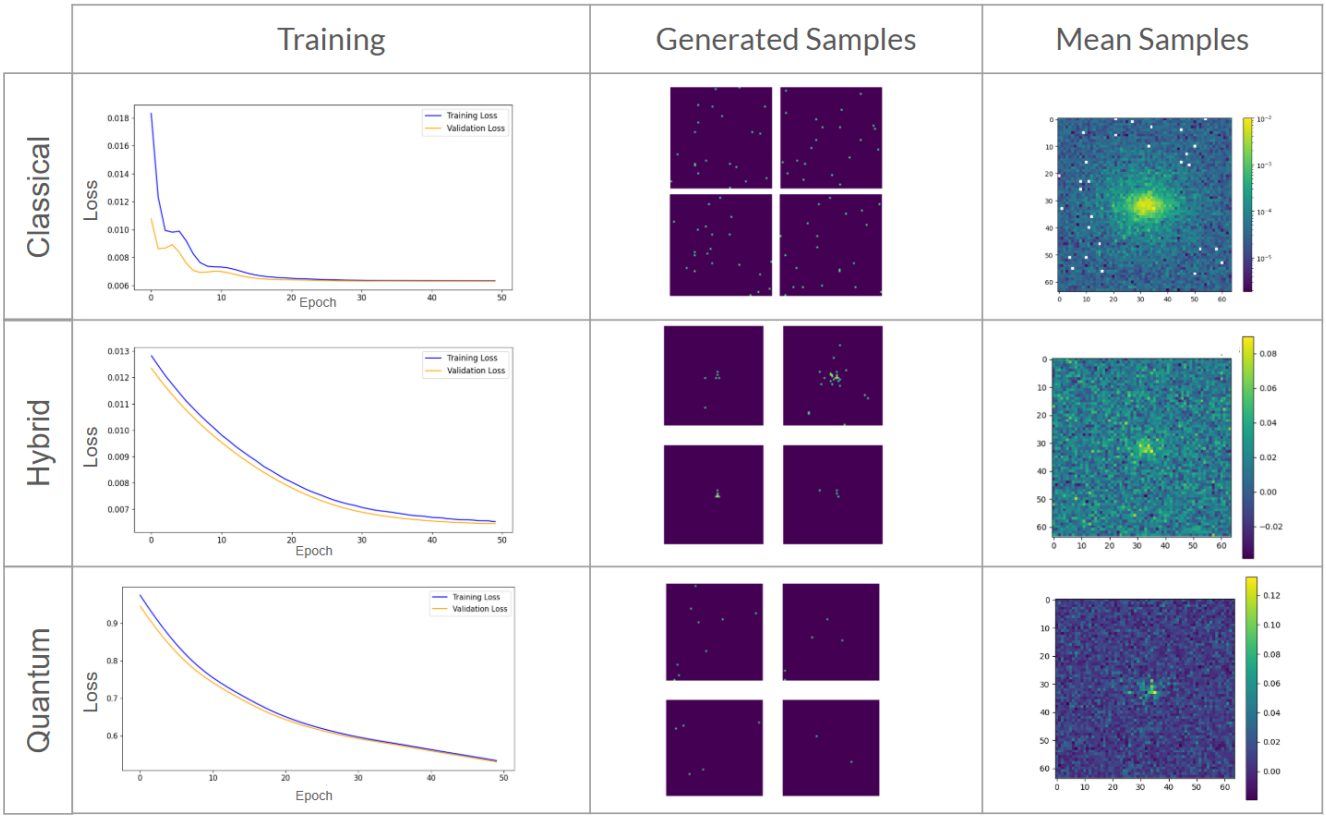


Figure 4: The loss graphs and generated samples of fully classical, hybrid, and fully quantum models. For all models, MSE and Adam optimizer was used to reach convergence. The far right column shows the mean generated samples for each distribution.

of different parametric quantum circuits and alternative forward scrambling strategies—such as Gaussian transformations in place of Haar unitaries—may yield deeper insights into how architectural choices influence quantum generative learning performance.

In parallel, deploying the quantum model on actual hardware with a limited number of qubits will be a key step toward assessing scalability. Such experiments will allow us to investigate the effects of quantum noise, hardware connectivity, and gate fidelity on model accuracy, offering a more realistic evaluation of feasibility in near-term quantum devices. Before testing on limited quantum computers, reliable results had to be shown through simulations. This work marks the threshold at that point in QML. Through these directions, we aim to advance the integration of quantum computing into practical, high-impact generative modeling pipelines for science.

## Data Availability

The code is open-source and can be found here: https://github.com/mashathepotato/GSoC-Quantum-Diffusion-Model. Please reach out to the corresponding author for any clarifications.

## Acknowledgements

This research was funded by Google Summer of Code (GSoC) 2024 and 2025 under the ML4Sci research group and used resources from the National Energy Research Scientific Computing Center (NERSC) for computing power.

## References

Andrews, M.; Alison, J.; An, S.; Burkle, B.; Gleyzer, S.; Narain, M.; Paulini, M.; Poczos, B.; and Usai, E. 2020. End-to-end jet classification of quarks and gluons with the CMS Open Data. *Nuclear instruments and methods in physics research section A: accelerators, spectrometers, detectors and associated equipment*, 977: 164304.

Baidachna, M.; Fatima, H.; Omran, R.; Ghadban, N.; Imran, M. A.; Taha, A.; and Mohjazi, L. 2024. Mirror, Mirror on the Wall: Automating Dental Smile Analysis with AI in Smart Mirrors. *Computing&AI Connect*, 1(1): 1–10.

Bansal, A.; Borgnia, E.; Chu, H.-M.; Li, J.; Kazemi, H.; Huang, F.; Goldblum, M.; Geiping, J.; and Goldstein, T. 2024. Cold diffusion: Inverting arbitrary image transforms without noise. *Advances in Neural Information Processing Systems*, 36.

Beer, K.; Bondarenko, D.; Farrelly, T.; Osborne, T.; Salzmann, R.; Scheiermann, D.; and Wolf, R. 2020. Training

deep quantum neural networks. *Nature Communications*, 11(1): 808.

Bergholm, V.; Izaac, J.; Schuld, M.; Gogolin, C.; Ahmed, S.; Ajith, V.; Alam, M. S.; Alonso-Linaje, G.; AkashNarayanan, B.; Asadi, A.; et al. 2018. Pennylane: Automatic differentiation of hybrid quantum-classical computations. *arXiv preprint arXiv:1811.04968*.

Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; and Lloyd, S. 2017. Quantum machine learning. *Nature*, 549(7671): 195–202.

Cao, H.; Tan, C.; Gao, Z.; Xu, Y.; Chen, G.; Heng, P.-A.; and Li, S. Z. 2024. A Survey on Generative Diffusion Models. *IEEE Transactions on Knowledge and Data Engineering*, 36(7): 2814–2830.

Chen, S. Y.-C.; Yang, C.-H. H.; Qi, J.; Chen, P.-Y.; Ma, X.; and Goan, H.-S. 2020. Variational quantum circuits for deep reinforcement learning. *IEEE access*, 8: 141007–141024.

Ciliberto, C.; Herbster, M.; Ialongo, A. D.; Pontil, M.; Rocchetto, A.; Severini, S.; and Wossnig, L. 2018. Quantum machine learning: a classical perspective. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2209): 20170551.

Comajoan Cara, M.; Dahale, G. R.; Dong, Z.; Forestano, R. T.; Gleyzer, S.; Justice, D.; Kong, K.; Magorsch, T.; Matchev, K. T.; Matcheva, K.; et al. 2024. Quantum Vision Transformers for Quark–Gluon Classification. *Axioms*, 13(5): 323.

De Falco, F.; Ceschini, A.; Sebastianelli, A.; Saux, B. L.; and Panella, M. 2024. Towards Efficient Quantum Hybrid Diffusion Models. *arXiv preprint arXiv:2402.16147*.

Dhariwal, P.; and Nichol, A. 2021. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34: 8780–8794.

Griol-Barres, I.; Milla, S.; Cebrián, A.; Mansoori, Y.; and Millet, J. 2021. Variational quantum circuits for machine learning. an application for the detection of weak signals. *Applied Sciences*, 11(14): 6427.

Haar, A. 1933. Der Massbegriff in der Theorie der kontinuierlichen Gruppen. *Annals of mathematics*, 34(1): 147–169.

Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; and Hochreiter, S. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Kawar, B.; Elad, M.; Ermon, S.; and Song, J. 2022. Denoising Diffusion Restoration Models. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 23593–23606. Curran Associates, Inc.

Nielsen, M. A.; and Chuang, I. L. 2002. *Quantum Computation and Quantum Information*. Cambridge University Press.

Pan, S.; Wang, T.; Qiu, R. L.; Axente, M.; Chang, C.-W.; Peng, J.; Patel, A. B.; Shelton, J.; Patel, S. A.; Roper, J.; et al. 2023. 2D medical image synthesis using transformer-based denoising diffusion probabilistic model. *Physics in Medicine & Biology*, 68(10): 105004.

Reddy, P.; Toomey, M. W.; Parul, H.; and Gleyzer, S. 2024. A Conditional Diffusion Model for Super-Resolution of Gravitational Lensing Data. *arXiv preprint arXiv:2406.08442*.

Romero, J.; and Aspuru-Guzik, A. 2021. Variational quantum generators: Generative adversarial quantum machine learning for continuous distributions. *Advanced Quantum Technologies*, 4(1): 2000003.

Schuld, M.; and Killoran, N. 2019. Quantum machine learning in feature Hilbert spaces. *Physical Review Letters*, 122(4): 040504.

Schuld, M.; Sinayskiy, I.; and Petruccione, F. 2015. An introduction to quantum machine learning. *Contemporary Physics*, 56(2): 172–185.

Shu, R.; Xu, X.; Yung, M.-H.; and Cui, W. 2024. Variational Quantum Circuits Enhanced Generative Adversarial Network. *arXiv preprint arXiv:2402.01791*.

Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; and Ganguli, S. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, 2256–2265. PMLR.

Song, J.; Meng, C.; and Ermon, S. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.

Stypułkowski, M.; Vougioukas, K.; He, S.; Zieba, M.; Petridis, S.; and Pantic, M. 2024. Diffused Heads: Diffusion Models Beat GANs on Talking-Face Generation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 5091–5100.

Xu, D.; Tan, J. C.; Hsu, C.-J.; and Zhu, Y. 2023. Denoising diffusion probabilistic models to predict the density of molecular clouds. *The Astrophysical Journal*, 950(2): 146.

Yi, K.; Zhou, B.; Shen, Y.; Liò, P.; and Wang, Y. 2024. Graph denoising diffusion for inverse protein folding. *Advances in Neural Information Processing Systems*, 36.

Zhang, B.; Xu, P.; Chen, X.; and Zhuang, Q. 2024. Generative quantum machine learning via denoising diffusion probabilistic models. *Physical Review Letters*, 132(10): 100602.