Quantum Machine Learning for High Energy Physics

Quantum Contrastive Learning

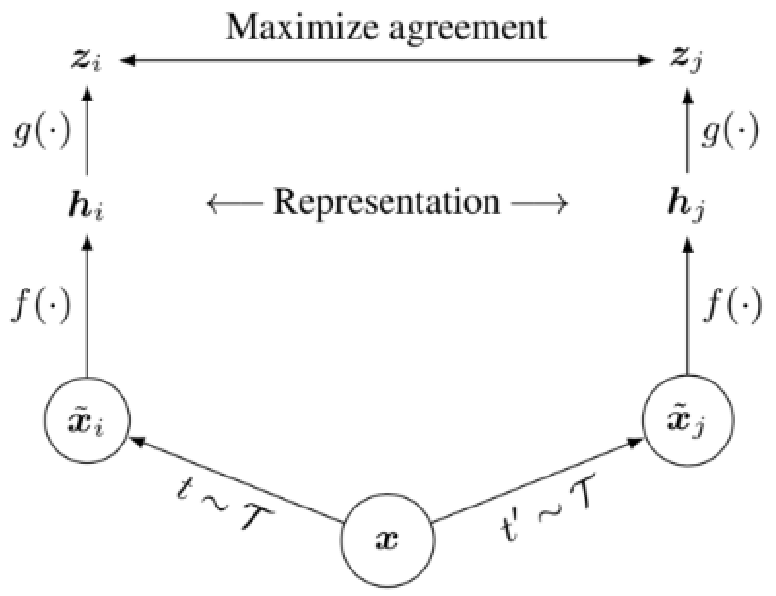
Amey Bhatuse

Google Summer of Code 2024
Final Evaluation

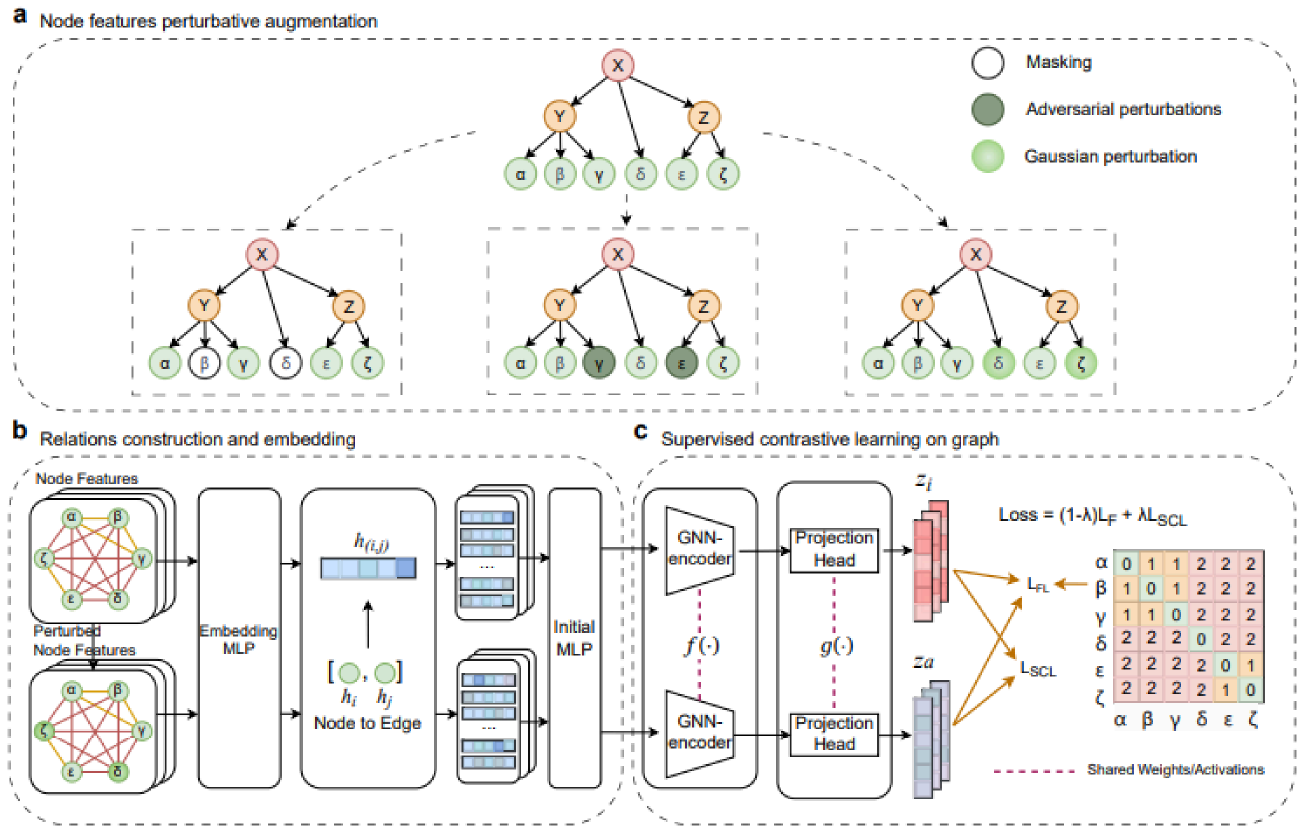
Goal

Learn better representations of high-energy physics data using quantum machine learning models and benchmark the model performance against pre-existing classical approaches

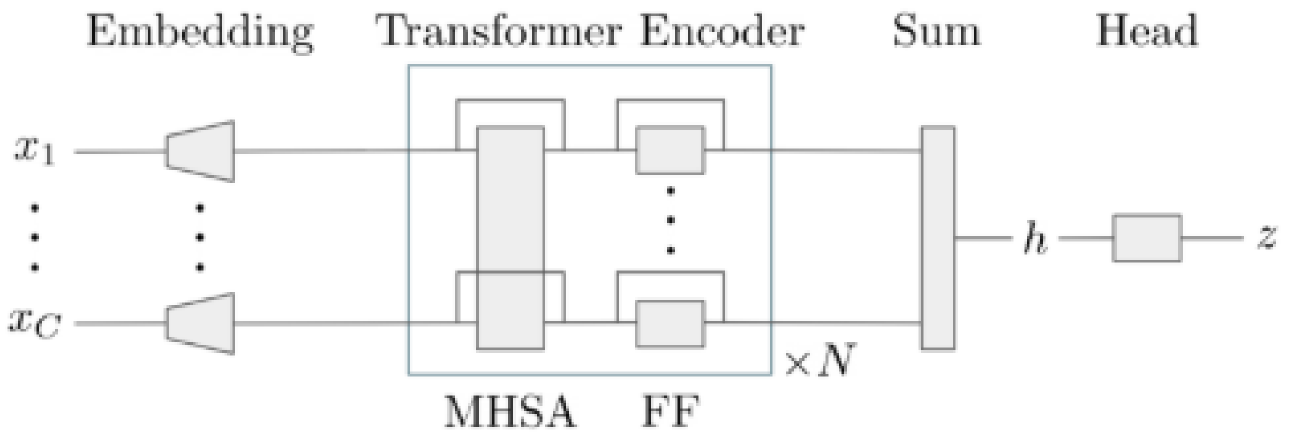
Contrastive Learning of HEP data



A simple framework for contrastive learning of visual representations. (Image source: Chen et al, 2020)

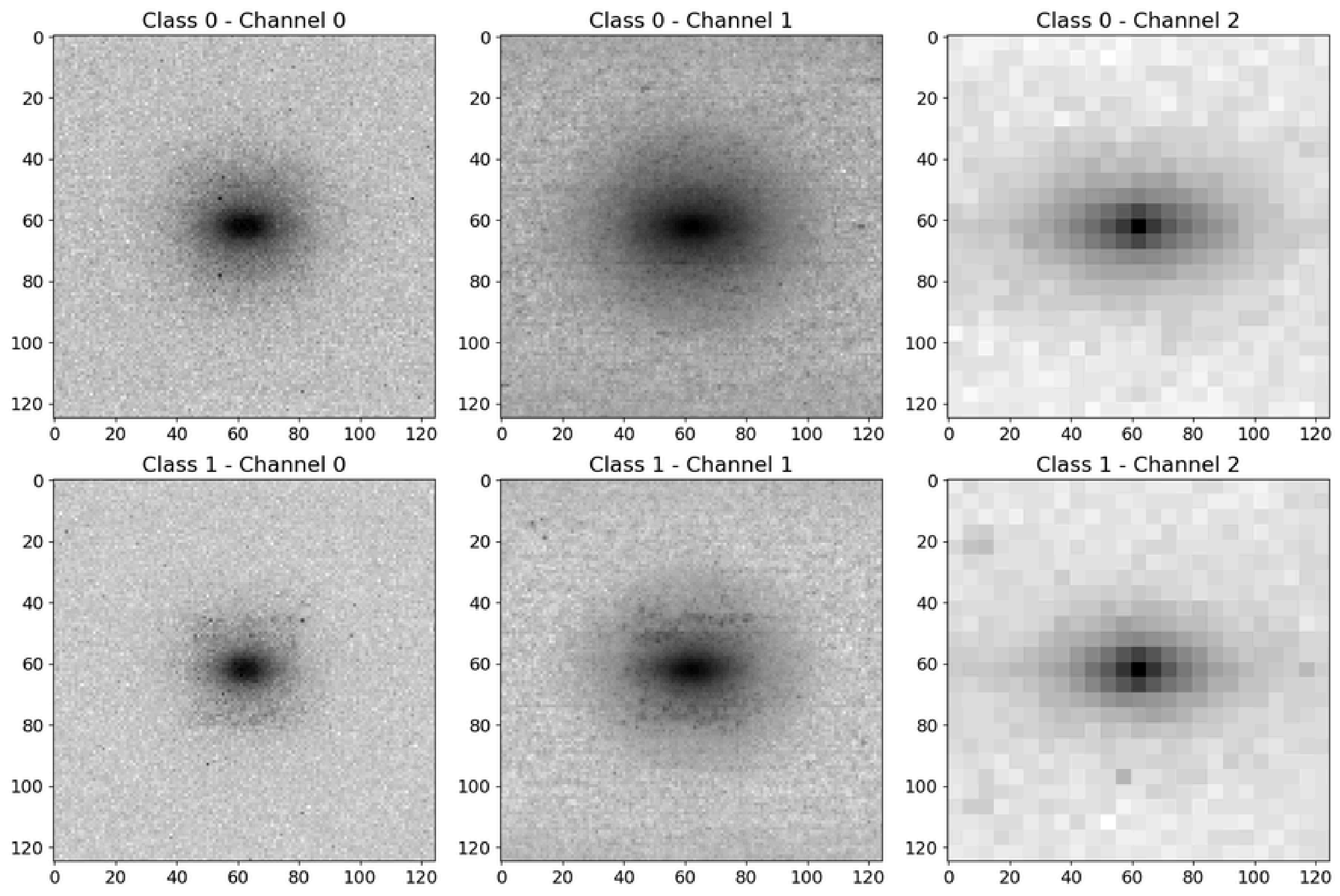


PASCL - Particle Decay Reconstruction



JetCLR - Jet classification

Working with Jet Images

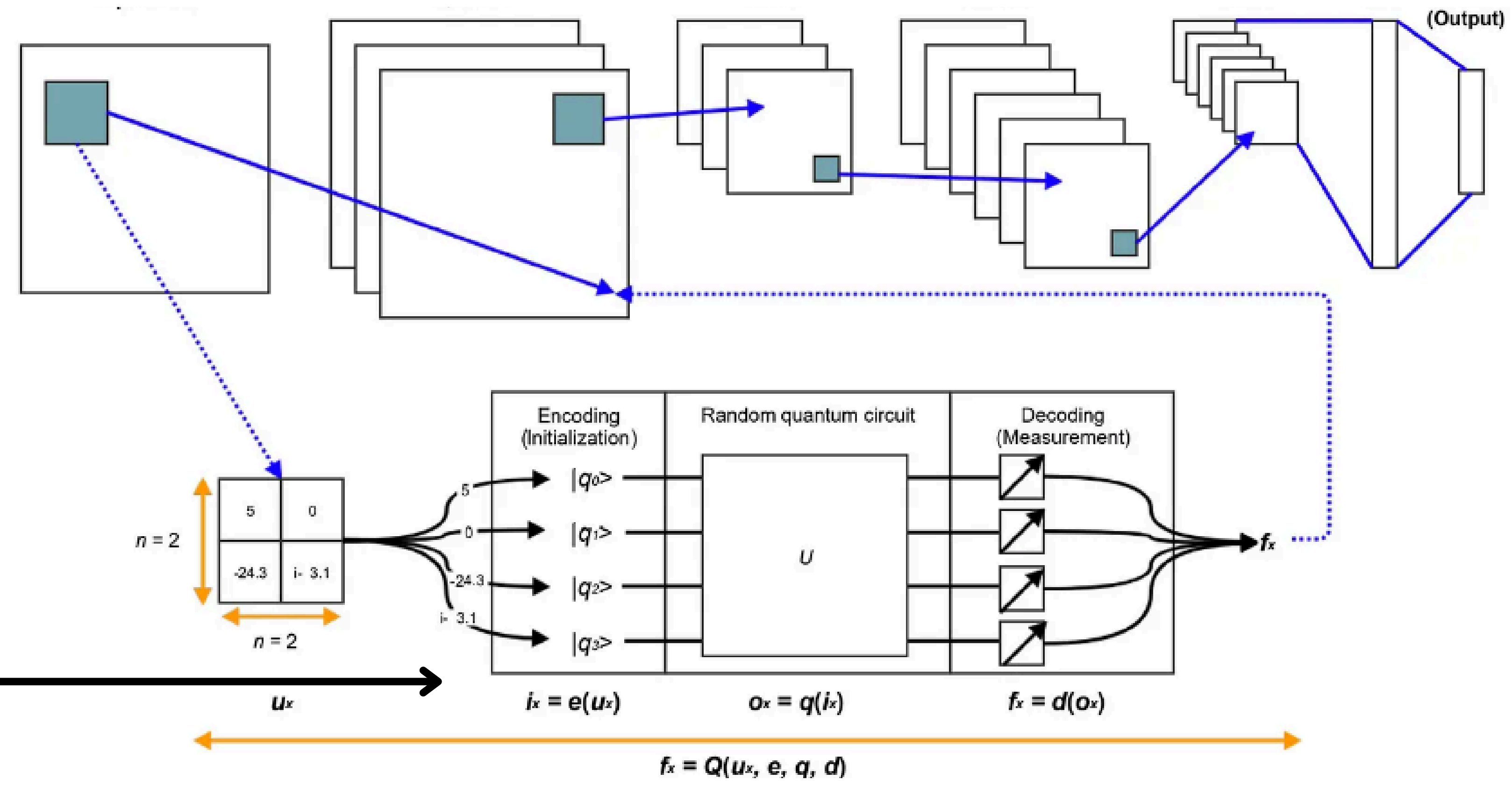


Average of jet images for each channel
Top - Gluon, Bottom - Quark

Data Augmentation

- Random Horizontal flips
- Random Vertical flips
- Random Rotations
- Z-score Normalization

Encoder - Quantum Convolutional NN



Data-Reuploading circuits (DRC)

HYBRID

- Quantum Convolution layers followed by a classical Linear layer
- Output - N-dimensional vector

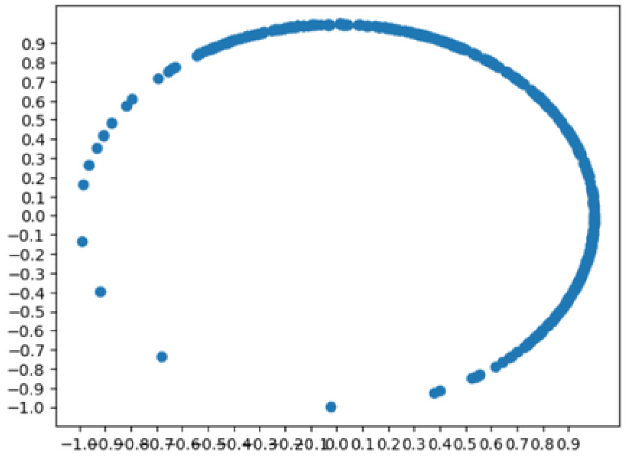
FULLY-QUANTUM

- Quantum Convolution layers followed by a parameterised quantum circuit
- Output - 2^n-dimensional quantum state vector

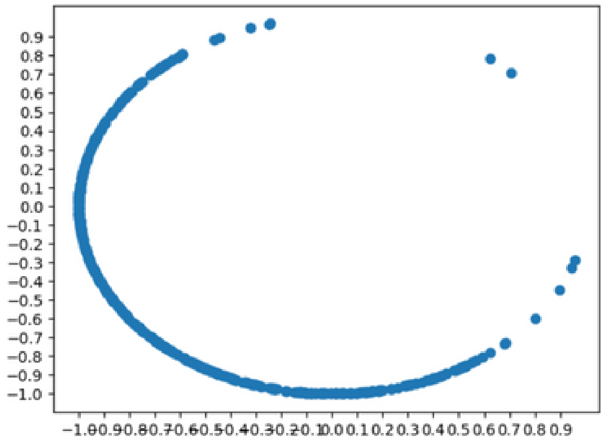
Image credits - https://arxiv.org/pdf/1904.04767

MNIST

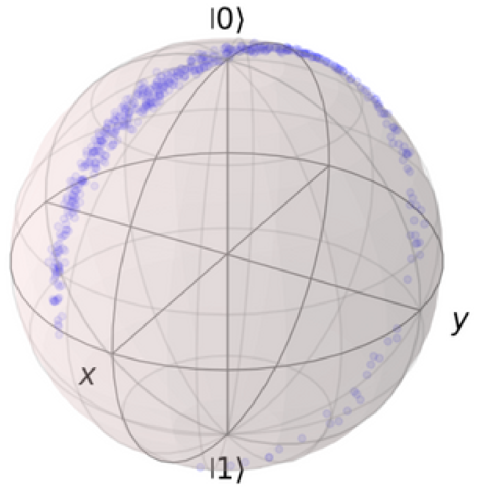
Downstream Classification Accuracy > 90%



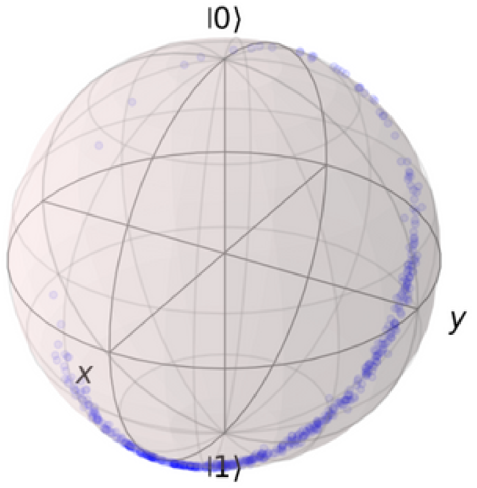
0s



8s



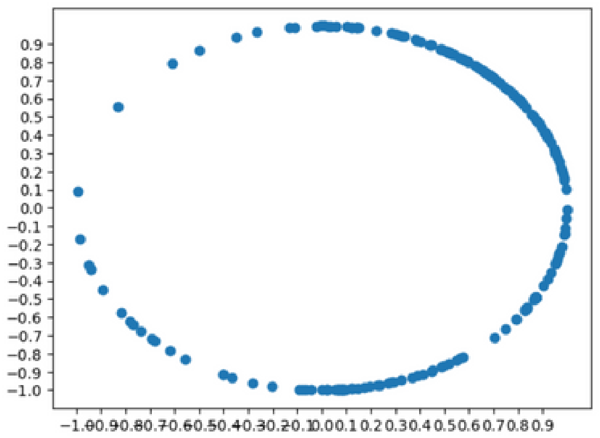
0s



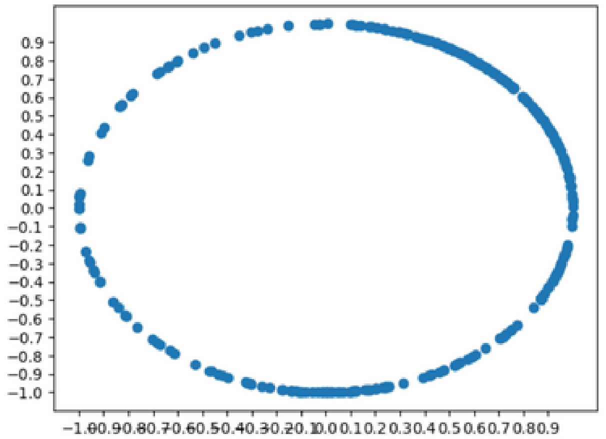
8s

Quark Gluon

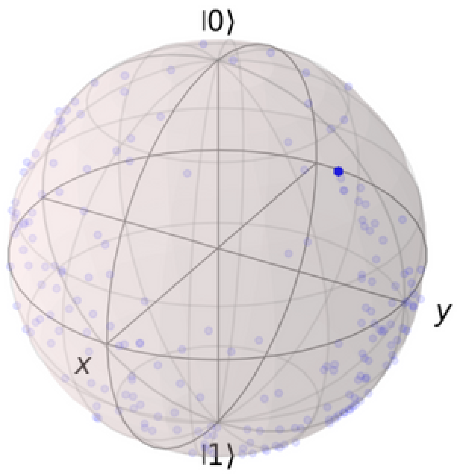
Downstream Classification Accuracy - 50-60 %



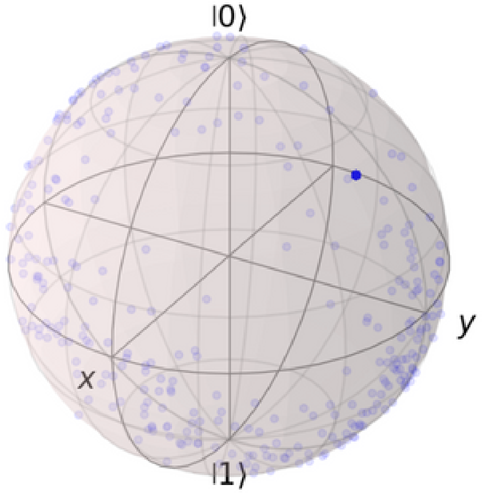
Quark



Gluon



Quark



Gluon

Key Learnings

- Data augmentations play more significant role than earlier assumed
- Jet images (that we use) are probably a less than good candidate for quantum contrastive learning

Jets as Graphs

Jet Observables

Original

$p_{T,\alpha}^{(i)}, y_{\alpha}^{(i)}, \phi_{\alpha}^{(i)}$ ⟵ feature vector for each particle

Derived (Roy's work from last year's GSoC)

Transverse Mass Per Multiplicity ($m_{\alpha,T}^{(i)}$)

$$m_{\alpha,T}^{(i)} = \sqrt{m_{\alpha}^{(i)^2} + p_{\alpha,T}^{(i)^2}}$$

Energy Per Multiplicity ($E_{\alpha}^{(i)}$)

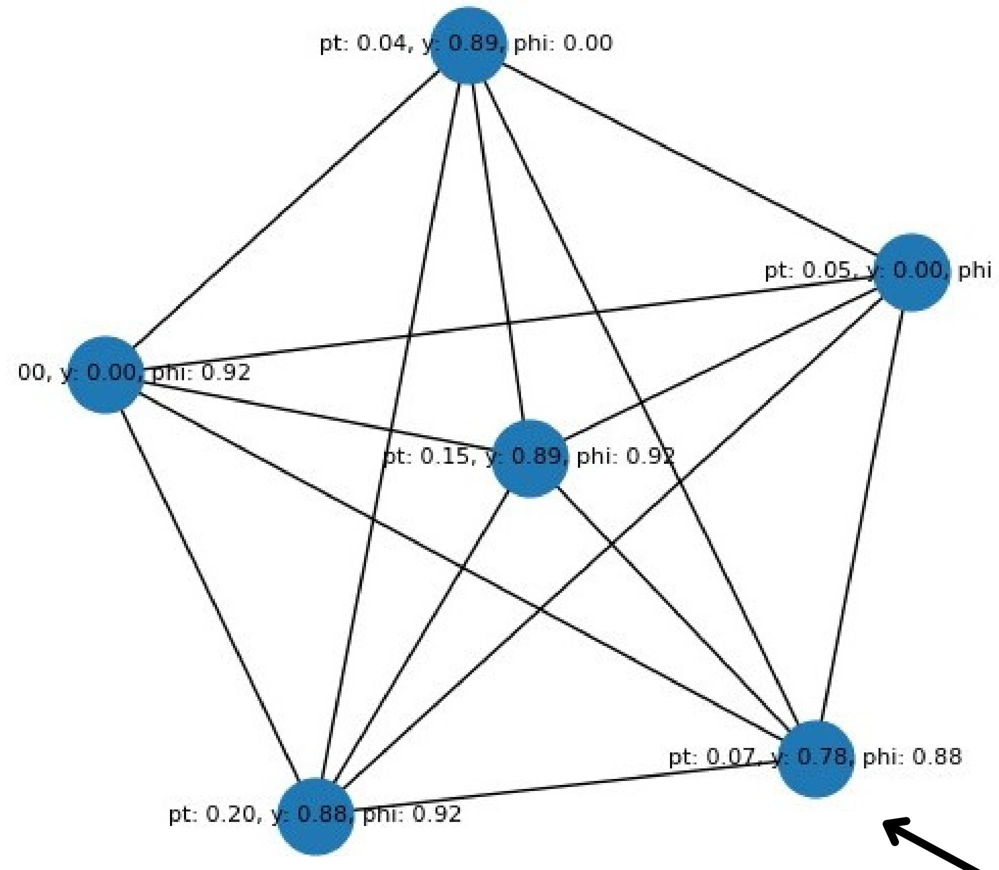
$$E_{\alpha}^{(i)} = m_{\alpha,T}^{(i)} \cosh y_{\alpha}^{(i)}$$

Kinematic Momenta Components Per Multiplicity ($\vec{p}_{\alpha}^{(i)} = (p_{x,\alpha}^{(i)}, p_{y,\alpha}^{(i)}, p_{z,\alpha}^{(i)})$)

$$p_{x,\alpha}^{(i)} = p_{T,\alpha}^{(i)} \cos\phi_{\alpha}^{(i)}$$

$$p_{y,\alpha}^{(i)} = p_{T,\alpha}^{(i)} \sin\phi_{\alpha}^{(i)}$$

$$p_{z,\alpha}^{(i)} = m_{T,\alpha}^{(i)} \sinh y_{\alpha}^{(i)}$$



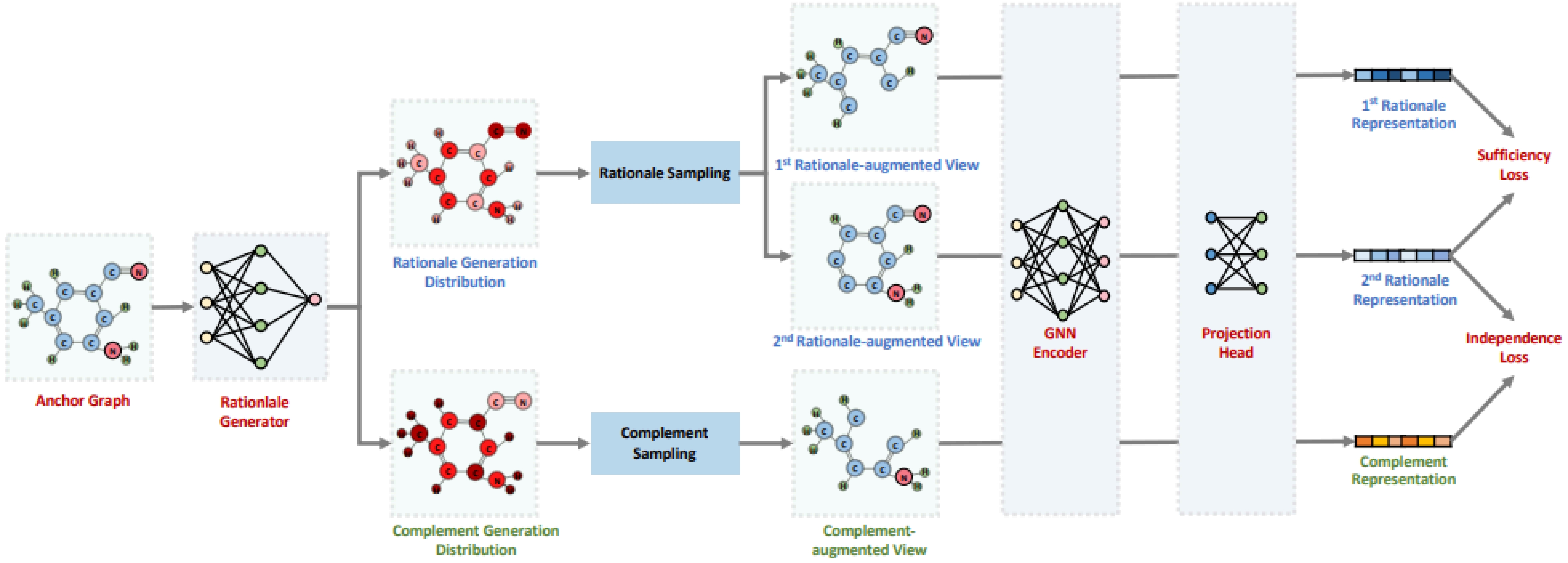
particles as nodes

Let Invariant Rationale Discovery Inspire Graph Contrastive Learning, Li et al.

Hypothesis -
Discriminatory information ----> Rationale (Causal part) (Here - a subset of nodes in the graph)

Perform augmentation to ***highlight*** rationale

- Encoder trains predominantly on the rationally important part leading to better representations
- Spurious Correlations (Non-Causal part) are ignored!



Non-Causal part

Augmentation and Rationale Generation

Node score from rationale generator

↓

Discrete probability distribution signifying each node's importance

↓

Randomly sample nodes (and their corresponding edges) under this distribution

IRC Safety

"Symmetries, Safety, and Self-Supervision" - Barry et al. (JetCLR paper)

Infrared safe $\eta' \sim \mathcal{N}\left(\eta, \frac{\Lambda_{\text{soft}}}{p_T}\right)$ and $\phi' \sim \mathcal{N}\left(\phi, \frac{\Lambda_{\text{soft}}}{p_T}\right)$ Λsoft = 100 MeV

Collinear safe $p_{T,a} + p_{T,b} = p_T$ $\eta_a = \eta_b = \eta$

$\phi_a = \phi_b = \phi$

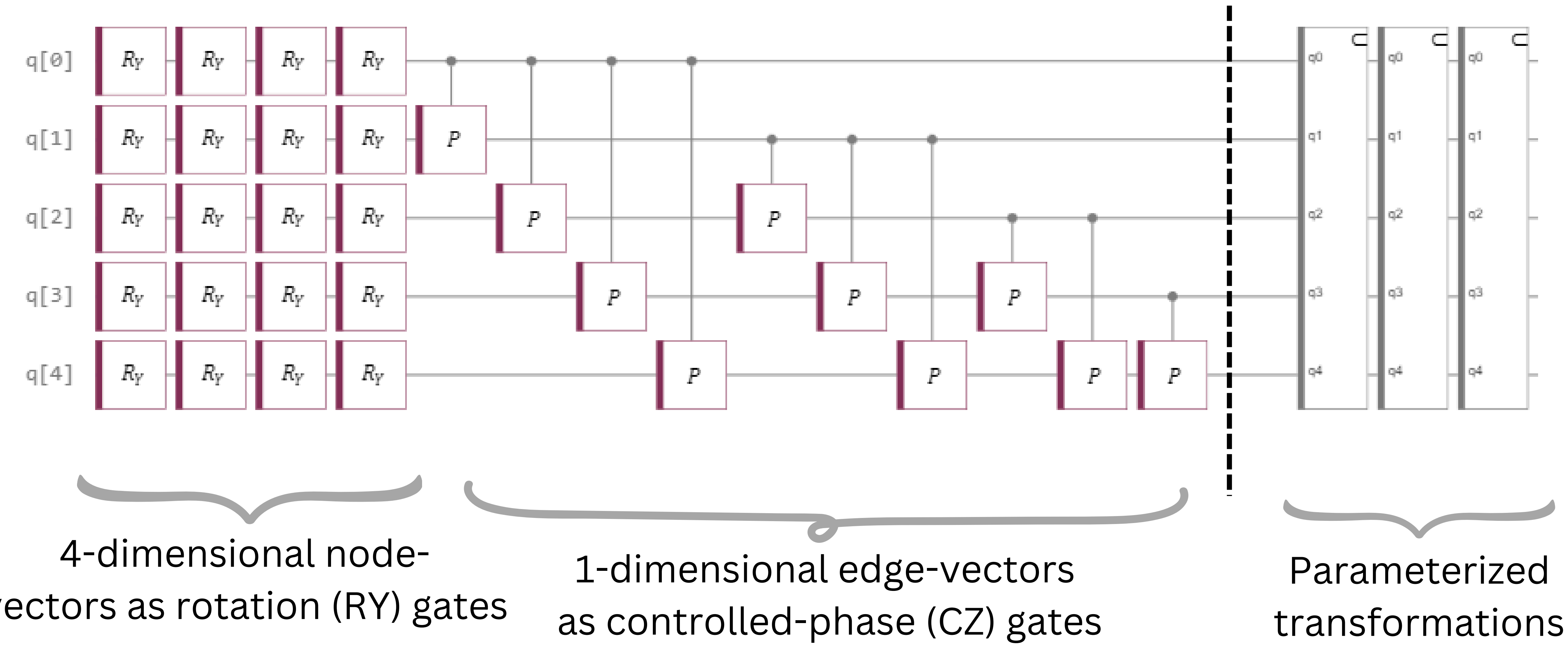
Backbone - ParticleNet, 10000 samples, 10 particles

	Classical RG	Quantum RG
Trainable parameters	1073	45
AUC	0.752	0.758

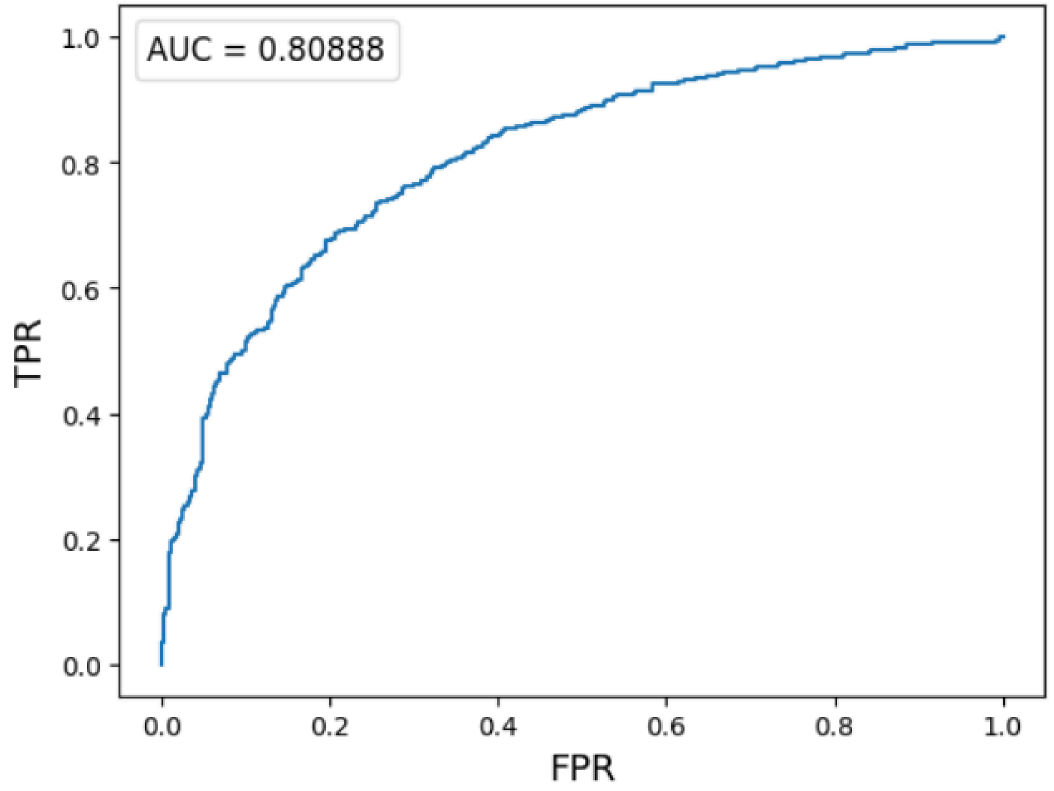
Quantum Rationale Generator

QNN

- Angle embedding node vectors (1 qubit per node)
- Use entangling layers for parameterization
- Output probability distribution over basis states
- Select probabilities of basis states with hamming weight 1 (For eg. 001, 010, 100) and normalize



Test Accuracy ~ 73%



In Progress...

- Gathering results for various model (different backbones) and data configurations
- Trying to work out the theoretical explanation for the performance of QRG and comparing it with the classical counterpart.

Thank you!