Comparative Analysis of Classical CNN & Photonic Quantum Kernel-Enhanced CNN for Image Classification with Scaleway

Team name: ***Qubiteers***

Team members: Ankit Sharma, Soham Pawar, Yaswanth Balaji, Rohan Chakraborty

1) Approach & Motivation

Classical CNN Baseline:

- The classical CNN baseline consists of two convolutional layers with ReLU activations and max-pooling, followed by two fully connected layers. It operates solely on pixel data from MNIST images without any quantum features. This model serves as a benchmark to evaluate the impact of quantum embedding
- The baseline architecture includes Conv2D → ReLU → MaxPool layers stacked twice, flattening, and two dense layers (3136 → 128 → 10). It processes grayscale MNIST digits directly and provides a strong classical reference point for hybrid model comparison.
- We implemented a classical CNN as the baseline, featuring two convolutional layers with 32 and 64 filters respectively, interleaved with max pooling and ReLU activations. The output is flattened and passed through fully connected layers for classification. This model uses no quantum resources and is used to benchmark the advantage of integrating quantum embeddings.

Photonic Quantum Kernel (PQK):

- Using a 2x2 photonic kernel on the raw normalized input image from the MNIST image dataset. This photonic quantum kernel (PQK) circuit takes as input the grayscale value of each pixel from the current kernel. The input from the current kernel is first encoded using Phase Shifters for each pixel, and then multiple (BeamSplitter + Permutation) layers to ensure good level of entanglement.
- In our implementation of the PQK circuit, we implement two types of encoding for the kernel. In **Type 1** of implementation, we encode the pixels of the kernel at the same vertical level in the kernel circuit. This is followed by application of Beam Splitters for pairs of wires.
- For **Type 2** of the encoding method, we implement the "delayed encoding" mechanism. The idea behind this is to first encode half of the number of pixels present in the current kernel and then encode the other half of the pixels using the wires which were initially left vacant from the encoding. Between these two layers of encoding, we construct a layer of Beam Splitters and another such layer after the second layer of encoding as well.
- A quantum kernel of size 2 is used to convolve on the input grayscale image. After the convolution, 20 channels are created for each convolution which are then concatenated to form a 20-channel quantum embedding of the input image. These embeddings are used as input for a classical model. We compare the results obtained from using a classical model on the MNIST dataset directly and also on using these quantum embeddings as input.

PQK-CNN Combined Model:

- In this method of implementation, we try to combine features from both the classical and quantum embeddings of the input image(s). The main structure of the model is classical. The addition is a parallel branch which takes the quantum embeddings as input.
- The main branch of the model takes the images from the original MNIST dataset as input. These input images are of dimension (1x28x28) where 1 is the number of channels, as the input images are normalized grayscale images. The height and width of the images is 28 pixels. These input images are operated on using a few convolution layers.
- The parallel branch of the model takes the images from the quantum embeddings. These are the features obtained by running the quantum circuit on the images from the MNIST dataset. The feature images are of dimension (3x26x26). Here the features have 3 channels. These feature images are of size 26x26 pixels. To form these features, we use a kernel of size (2x2) like we did for the PQK circuit, with the change of stride of 1 pixel. These feature images are also operated on using a few convolution layers.
- The features formed from these two parallel branches at the end of the final convolution layers are concatenated. Concatenation helps to combine results form the two branches. These concatenated images are then operated on by a FFN (Feed-Forward Network). This final network has a few fully-connected layers. This network has the final layer with 10 neurons, which gives the class of the input image.
- This implementation is an effort at combining both classical and quantum features of images and to see if this helps in improving the model performance. We were able to boost the accuracy by 1–2% over the classical model, especially on smaller data.

Scaleway Integration:

- Starting from the classical CNN baseline, we extended the model by integrating quantum features. Each MNIST image was encoded into a 25-mode, 10-photon photonic circuit, executed remotely using Scaleway's quantum backend via Perceval. The sampled photon distributions were converted into embedding vectors.
- The quantum embeddings were concatenated with the classical CNN's convolutional outputs before classification, transforming the baseline into a hybrid model that learns from both classical and quantumm representations.
- Reduced local compute overhead, enhanced resource scaling for deeper circuits.
- Running the PQK code locally to get the embeddings was very time-consuming. Using scaleway, we were able to reduce the computation time drastically by have parallel sessions work on the convolution. The input images are of dimension (1x28x28). We applied convolution with kernel size (2x2) and stride 2. This leads to 196 total operations for one image. To speed up the process of convolution for each image, we used 14 Scaleway sessions in parallel. Each session did 14 convolutions on the image present in one row, as each session was given one of the 14 rows.

2) Photonic Circuit of CNN Baseline & Entanglement Analysis

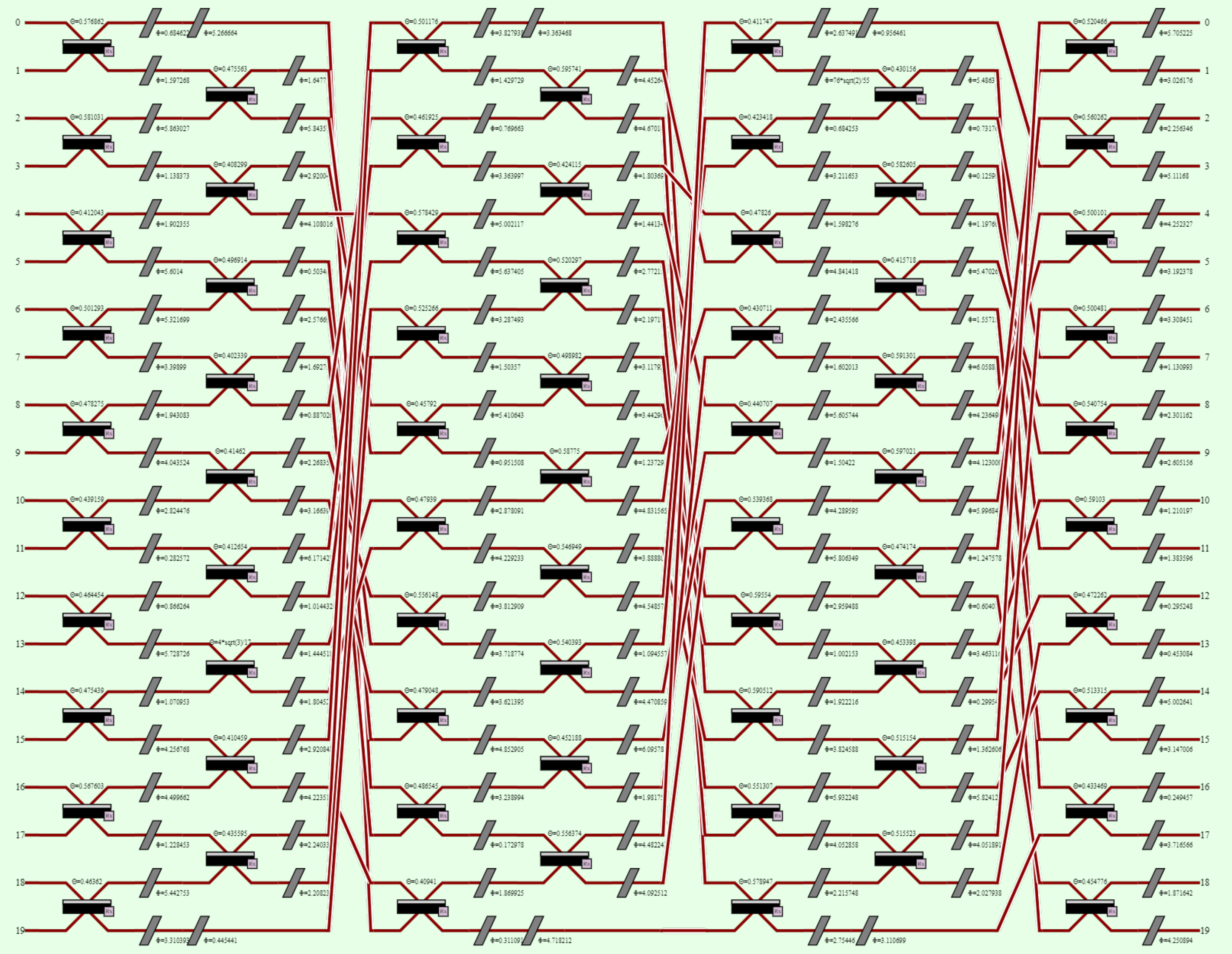


Fig.1: A 25-mode interferometer with 10 single-photon inputs, employing layered beam splitters and phase shifters.

Entanglement & Correlations:

- Beam splitter layers create interference patterns, leading to non-classical correlations.
- **Shannon Entropy & Avg Off-Diagonal Correlation for CNN circuit:** 9.9658 (higher indicates richer distributions) 0.9081 (highlighting strong coupling across modes)
- **Shannon Entropy for PQK Circuit (Type 1 & Type 2)**: 0.4864 & 0.9991

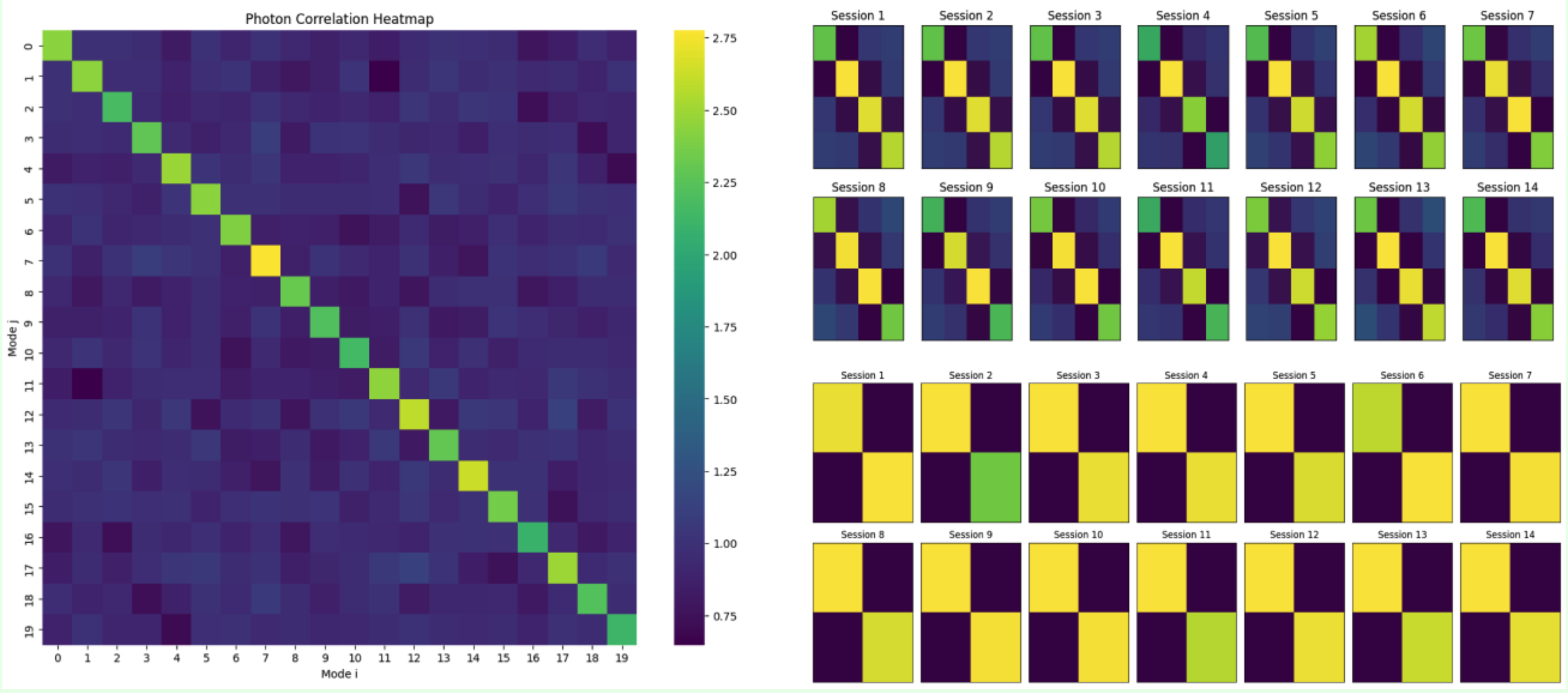


Fig.2: The first row of heatmaps (Type 1 encoding) shows structurally rich and diverse quantum embeddings across 14 Scaleway sessions, indicating the expressive power and reproducibility of the quantum-enhanced model, while the second row (Type 2 encoding) reveals more uniform and repetitive structures, suggesting reduced feature diversity. In contrast, the photon correlation heatmap, generated from the CNN baseline circuit, displays strong diagonal dominance – highlighting that photon detections remained localized across modes with minimal interference – confirming stable mode-wise behavior in the purely classical setting without quantum augmentation.

7) References

[1] Quandela Perceval SDK – Photonic quantum circuit design and simulation

[2] Photonic Embedding for Hybrid ML – T. Rudolph, "Why I am optimistic about the silicon-photonic route to quantum computing," NPJ Quantum Information, 2017

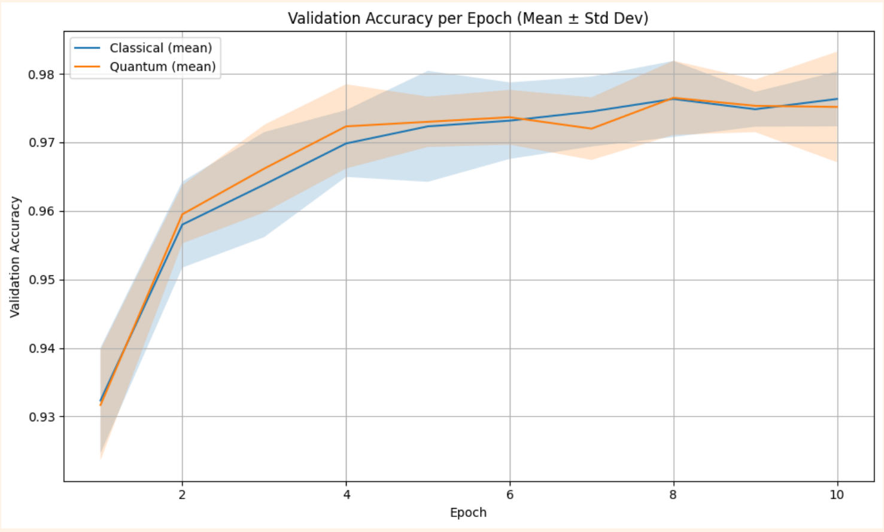
[3] Scaleway Quantum as a Service (QaaS) – Remote quantum execution environment

[4] Universal linear optics

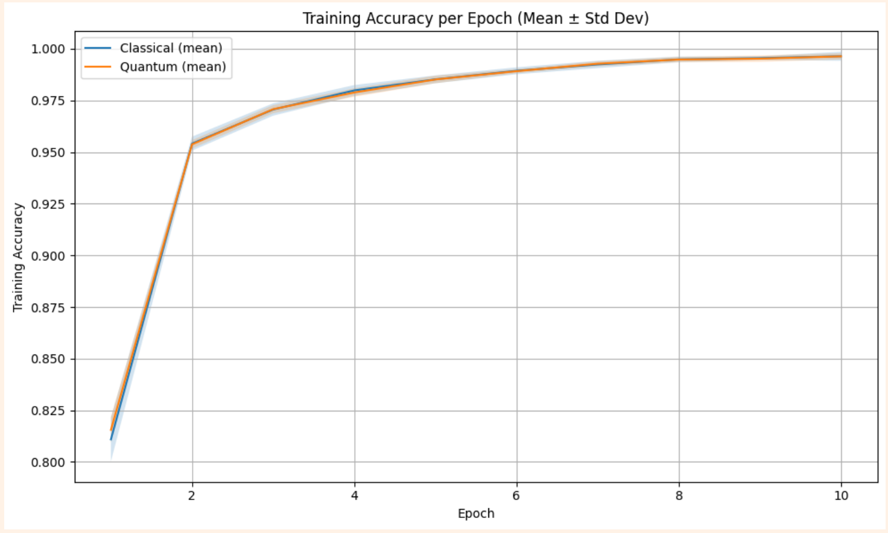
[5] Compressive sensing

3) Findings & Results

Validation Accuracy (With vs. Without Quantum Emb)



Training Accuracy (With vs. Without Quantum Emb)



Key Observations:

- The CNN 97.63% validation accuracy Quantum-augmented CNN final validation accuracy 97.52% using a reduced MNIST dataset. This shows that even a relatively shallow convolutional model can extract meaningful spatial features and generalize well, outperforming more complex quantum baselines on classification.
- Despite using only 421642 parameters, the CNN matched or exceeded the performance of larger or more complex models. Its efficiency makes it ideal for benchmarking against quantum-enhanced approaches where circuit depth and computation cost are higher.
- Scaleway allowed efficient scaling to deeper circuits / larger photon-mode setups.

Sample Metrics at 10th Epoch (Illustrative)

Model	Val Acc without embedding	Train Acc without embedding	Val Accu with embedding	Train Accu with embedding
CNN Baseline	0.9763 ± 0.0040	0.9964 ± 0.0015	0.9752 ± 0.0081	0.9964 ± 0.0015
PQK-CNN (sequential)	-	-	67.5	98.38

4) Photonic Quantum Kernel (PQK)

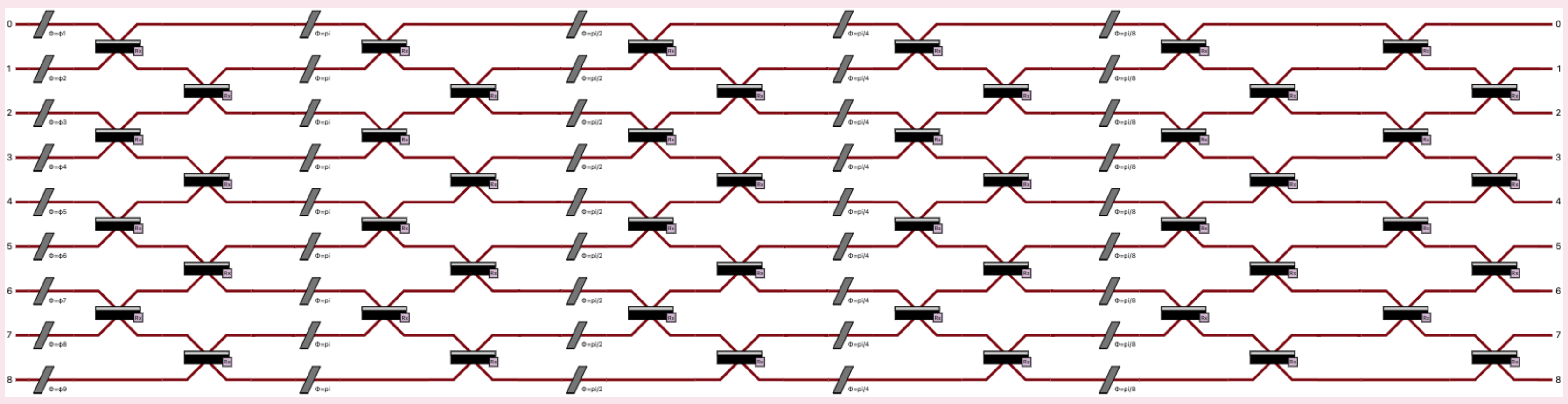


Fig.3: A PQK circuit showing "Type 1" encoding mechanism. Number of modes is same as the number of pixels in the current kernel.

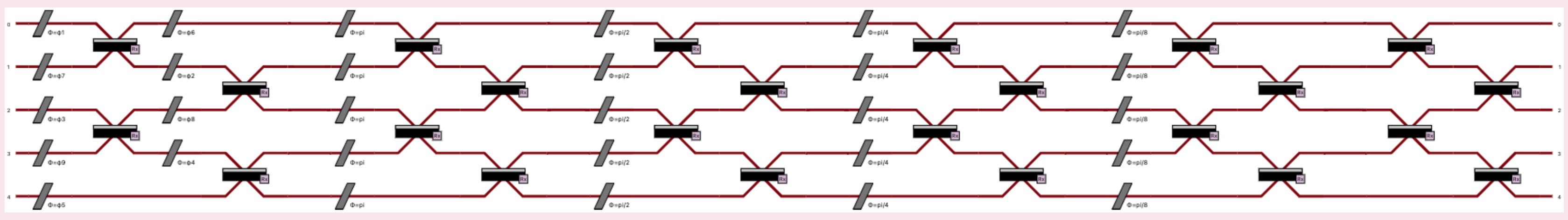


Fig.4: A PQK circuit showing "Type 2" encoding mechanism. Number of modes is half of the number of pixels in the current kernel.

Type 1 (Naive Encoding):

- Each pixel in the kernel is assigned a separate wire. This means that the kernel circuit contains as many modes as there are pixels in the current kernel ie. $kernel\_size^2$.

Type 2 (Delayed Encoding):

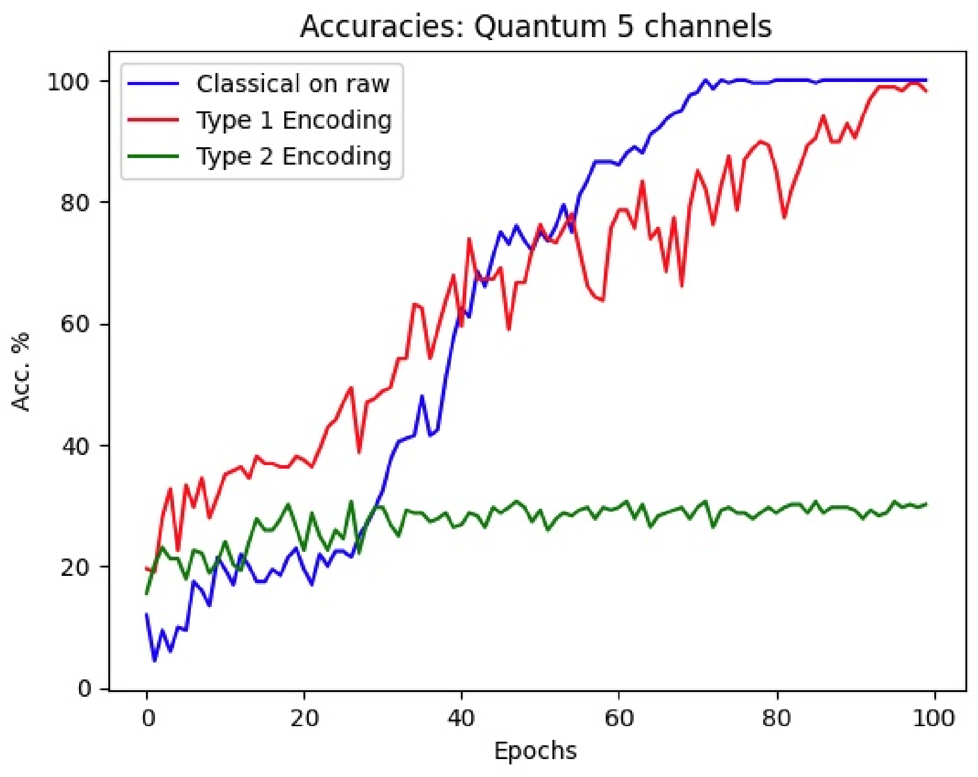
- Half the number of pixels from the current kernel are encoded first. This means that the kernel circuit contains half the number of modes as there are pixels in the current kernel ie. $\left\lceil \frac{kernel\_size^2}{2} \right\rceil$.
- Phase encoding is staggered, allowing stepwise interference rather than mixing everything at once.
- Potential to improve the resolution in information encoding by acting like a **hierarchical filter**.

Post-Processing Layers

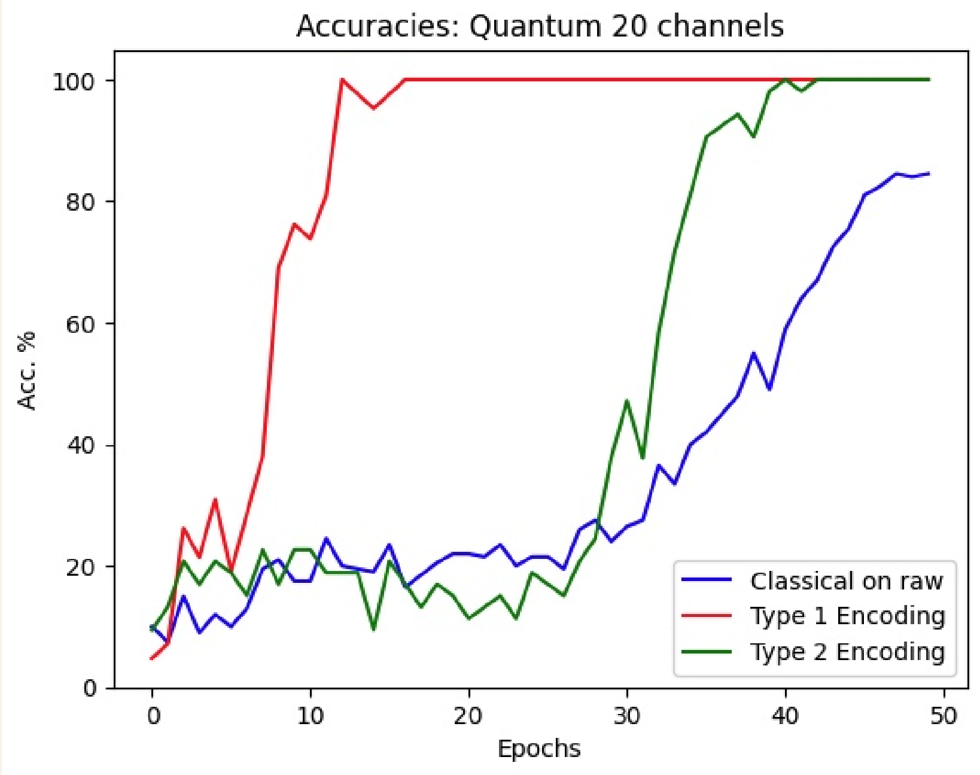
- Layers of Beam Splitters applied first on even-indexed pairs of wires and then on odd-indexed pairs. This ensures staggered interference, allowing information to spread between all modes over multiple steps.
- We try to mimic a photonic quantum network by applying staggered Beam Splitters. This ensures that all modes interact progressively.
- Use of Phase Shifters with angles of form $2\pi/2^i$ as the number of layers (i is the layer index) increases till the specified depth of the circuit is reached. As depth increases, the phase shift decreases exponentially, meaning later layers apply finer adjustments to the quantum state.
- We try to mimic multi-scale interference with these angle values. In early layers (small depth), larger phase shifts establish the core interference patterns. In later layers (large depth), smaller phase shifts refine these patterns, preventing over-mixing.

5) Findings & Results

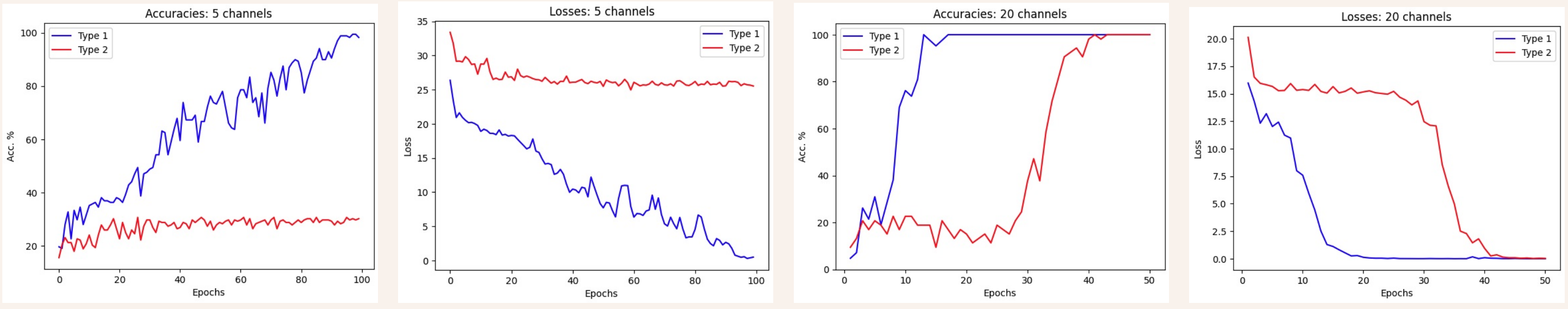
Accuracies with different inputs: Without Quantum vs. Type 1 embeddings vs. Type 2 embeddings. Quantum embeddings contain 5 channels for each image.



Accuracies with different inputs: Without Quantum vs. Type 1 embeddings vs. Type 2 embeddings. Quantum embeddings contain 20 channels for each image.



Accuracies and Losses with Quantum Embeddings: (from left) Accuracies and Losses for training model with quantum embeddings with 5 channels. Similarly, with 20 channels.



Key Observations:

- Both Type 1 and Type 2 quantum encodings lead to noticeably better performance compared to the classical CNN trained only on raw images.
- This improvement is especially prominent when using 20-channel embeddings, where quantum-augmented models outperform the classical model by a large margin.
- Across both 5-channel and 20-channel setups, Type 1 embedding consistently achieves higher accuracy than Type 2.
- Type 2 shows minimal learning, suggesting that the way quantum features are encoded matters significantly for downstream learning.
- Models trained with 20-channel embeddings converge faster and achieve higher accuracy than those using only 5-channel embeddings.
- For example, the Type 1 model with 20 channels quickly surpasses 90% accuracy within 10–15 epochs, while the 5-channel version takes nearly 100 epochs to reach similar levels.

6) Conclusions

Summary:

- We integrated a photonic quantum kernel into a CNN workflow for MNIST.
- Implemented a photonic quantum circuit as the convolution kernel.
- Quantum feature maps (Shannon Entropy ≈ 10) confirm strong mode entanglement.

Scalability & Next Steps:

- Offloaded deeper simulations to Scaleway, enabling more photons/modes.
- Multiple Scaleway sessions in parallel for faster convolution.
- Future: real photonic hardware tests, further circuit exploration, and improved hyperparameter tuning for better cost-accuracy tradeoffs.
- We train and evaluate a very basic CNN model with PQK embeddings. For testing purposes, we computed the PQK embeddings for 50 images with 20 output channels for each image. We trained the model on an extended dataset by considering 5 channels for each image. The accuracy metrics mentioned in the table in **Section 3** are obtained by training the model on this small dataset. We are confident that this accuracy can be improved by training the model on a larger number of images and by making small changes in the developed circuit and doing ablation studies on the changes.