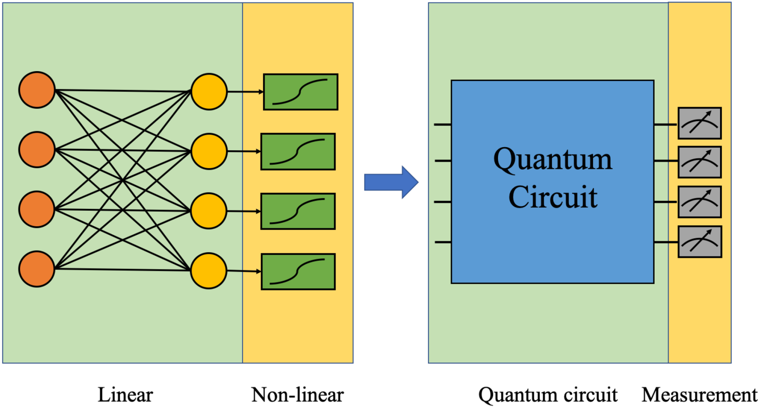
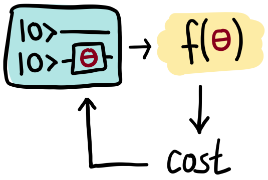
Hybrid Quantum-classical Neural Network

Qming

Hybrid Quantum-classical Neural Network



Paramterized quantum circuits (PQC)



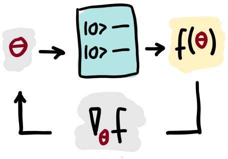
How to build?

- Preparation of a fixed initial state (e.g., the vacuum state or the zero state)
- A quantum circuit $U(\theta)$ parameterized by a set of trainable parameters θ
- Measurement of an observable A at the output

The forward pass function $f(\theta)$ represented by the parameterized quantum circuit is usually the expectation value of the observable A

$$f(\theta) = \langle 0|U^\dagger(\theta) A U(\theta)|0\rangle$$

Quantum backpropagation

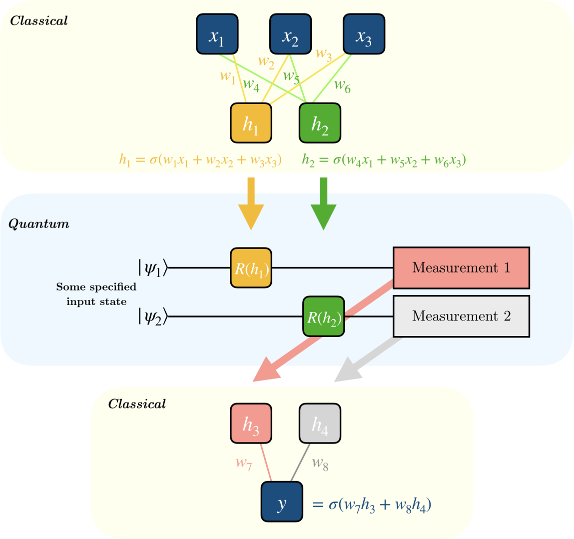


Gradient of the parameterized quantum circuit can be calculated in light of **parameter-shift rule**

$$\frac{d}{d\theta}f(\theta) = c\left[f(\theta + s) - f(\theta - s)\right]$$

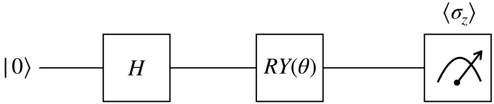
where constants c and s are determined by the transformation $U^{\dagger}(\theta)AU(\theta)$.

A simple example of hybrid quantum-classical neural network (HQNN)



Paramterized quantum circuit with single qubit

1-qubit circuit with one trainable quantum parameter θ



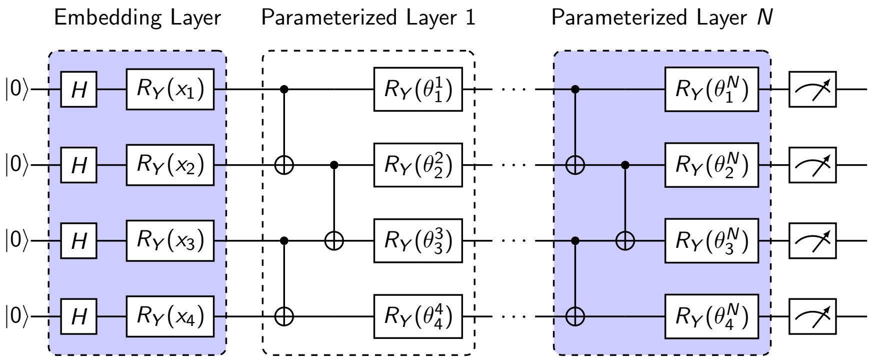
where H is Hadamard gate, $R_Y(\theta)$ is the operator performing a rotation of θ around the y-axis in the Bloch sphere, and σ_z represents the Pauli Z matrix or Z-gate. The corresponding forward function is $f(\theta) = \langle \sigma_z \rangle$ and it can be obtained by measuring the output in the z-basis:

$$\langle \sigma_z \rangle = \sum_i z_i p(z_i)$$

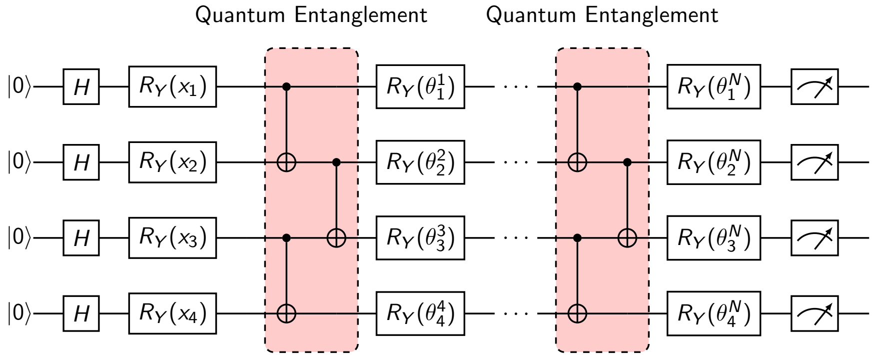
where z_i is the observable value and $p(z_i)$ is the corresponding probability. The gradient of $f(\theta)$ is

$$\frac{d}{d\theta} f(\theta) = \frac{1}{2}\left[f(\theta + \frac{\pi}{2}) - f(\theta - \frac{\pi}{2}) \right]$$

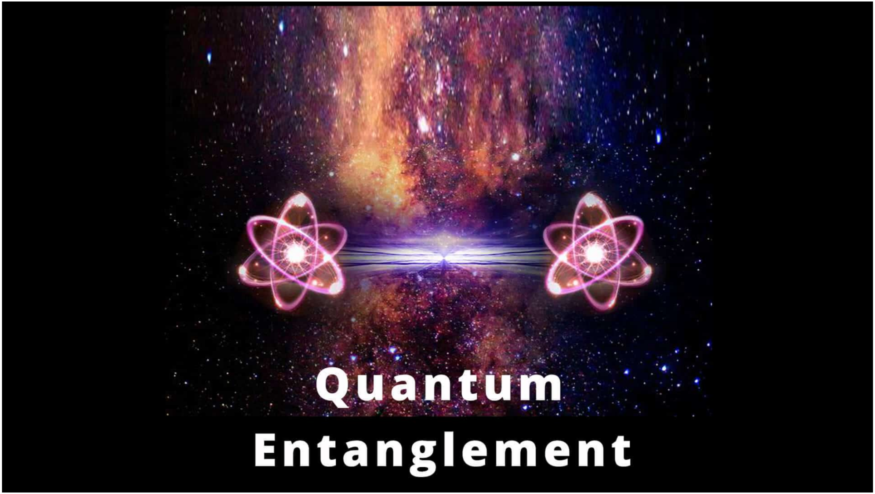
Paramterized quantum circuit with multiple qubits



Paramterized quantum circuit with multiple qubits



Quantum entanglement



Quantum entanglement is a physical phenomenon that occurs when a pair or group of particles are generated, interact, or share spatial proximity in a way such that the quantum state of each particle of the pair or group cannot be described independently of the state of the others, including when the particles are separated by a large distance.

Entangled state example

$$|\Psi_{ab}\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$$

- $|\Psi_{ab}\rangle$ **can not** be expressed as $|\Psi_{ab}\rangle = |\Psi_a\rangle \otimes |\Psi_b\rangle$
- Particles a and b are always in the **same** state no matter how far they are from each other
- A completely quantum phenomenon that classical system **can not** produce

Unentangled state example

$$|\Psi_{ab}\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|01\rangle$$

- $|\Psi_{ab}\rangle$ **can** be expressed as $|\Psi_{ab}\rangle = |\Psi_a\rangle \otimes |\Psi_b\rangle$ where

$$\begin{aligned} |\Psi_a\rangle &= |0\rangle \\ |\Psi_b\rangle &= \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle \end{aligned}$$

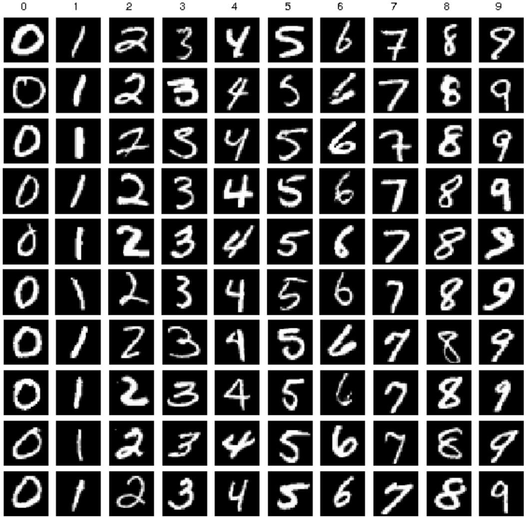
- Particles a and b are in **independent** states

Experiments

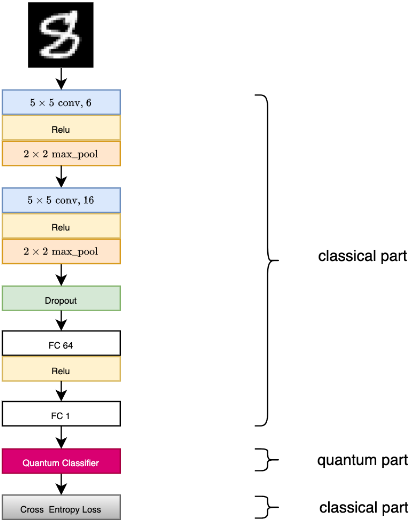
- **Experiment 1: Quantum Classifier**
- **Experiment 2: Quantum Activation**
- **Experiment 3: Quantum Entanglement**

Experiment 1: Quantum Classifier

- **Dataset**: MNIST (handwritten digits)



- **Model Architecture**:

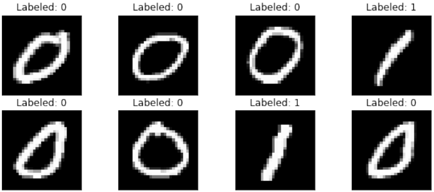


Result

Model performance

Train loss	Test loss	Test acc
-0.9870	-0.9847	100%

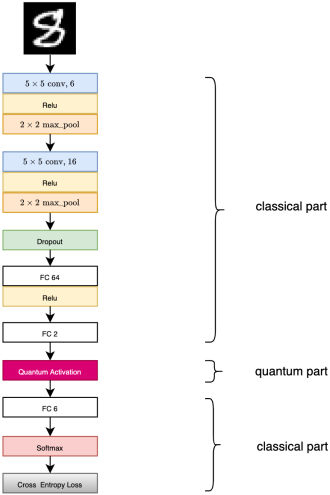
Sample classification results



Experiment 2: Quantum Activation

Task 1: multi-class classification

- **Dataset**: MNIST
- **Model Architecture**:

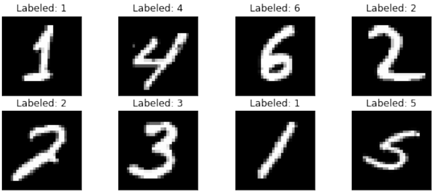


Result

Model performance

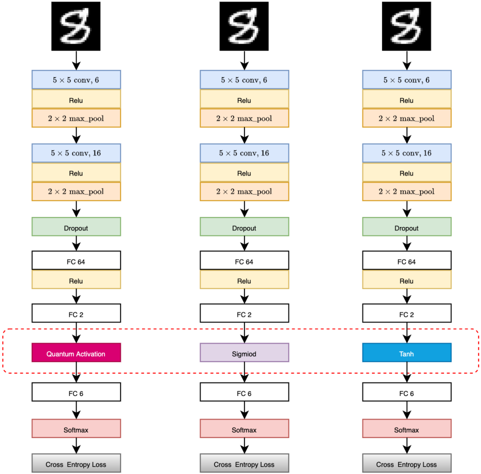
Test loss	Test acc
0.4324	0.986%

Sample classification results

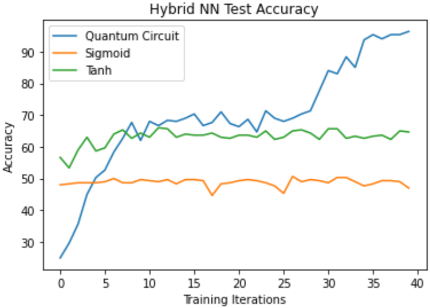


Task 2: comparison between quantum and classical activation functions

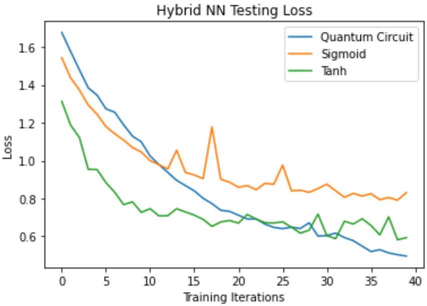
- **Dataset**: MNIST
- **Model architectures**:



- **Result**



Hybrid NN Test Accuracy
Quantum Circuit
Sigmoid
Tanh
Accuracy
Training Iterations



Hybrid NN Testing Loss
Quantum Circuit
Sigmoid
Tanh
Loss
Training Iterations

Justification

Justification

- gradient vanishing

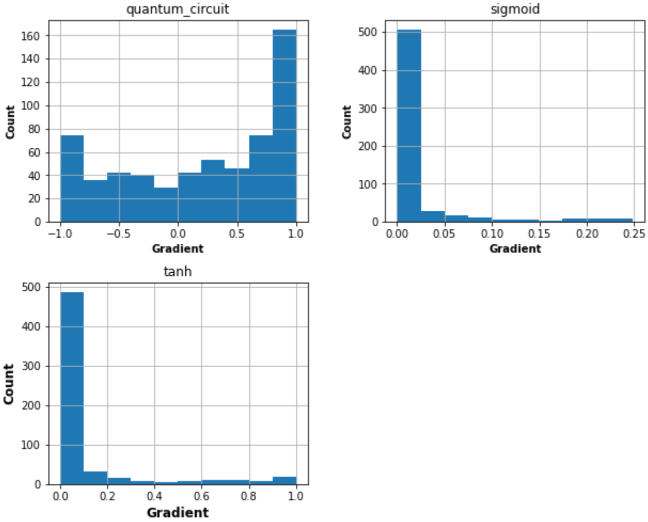
Justification

- gradient vanishing
- gradients of sigmoid and tanh functions

Justification

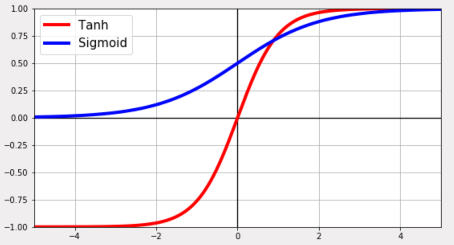
- gradient vanishing
- gradients of sigmoid and tanh functions
- gradient of quantum activation funciton

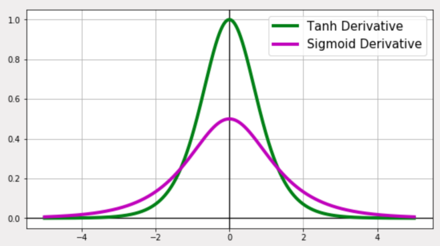
gradient vanishing



gradients of three activation functions calculated at the early stage of the training

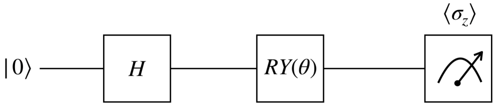
gradients of sigmoid and tanh





They both have a very small active or unsaturated region, namely the region with larger gradients

gradient of quantum activation function



The quantum activation function $f(\theta)$ in our case is defined by the parameterized quantum circuit with one qubit

$$f(\theta) \quad = \quad \langle 0|HR_Y^\dagger(\theta)ZR_Y(\theta)H|0\rangle \tag{1}$$

where $R_Y(\theta)$ is the operator performing a rotation of θ around the y-axis in the Bloch sphere and it is generated by the Y Pauli matrix (Y-gate)

$$R_Y(\theta) = e^{-i\frac{1}{2}\theta Y}. \tag{2}$$

In light of the relation

$$e^{-i\theta G} = I\cos(\theta) - iG\sin(\theta) \tag{3}$$

where G is a Hermitian and unitary operator, $R_Y(\theta)$ can be rewritten as

$$R_Y(\theta) = I\cos(\frac{\theta}{2}) - iY\sin(\frac{\theta}{2}) \tag{4}$$

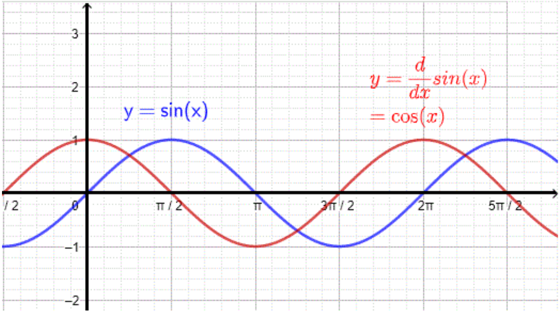
Using

$$\begin{aligned} H|0\rangle &= |+\rangle = \frac{\sqrt{2}}{2}(|0\rangle + |1\rangle) \\ Y|+\rangle &= -i|-\rangle \\ Y|-\rangle &= i|+\rangle \\ Y^{\dagger} &= Y \end{aligned}$$

and (4), the quantum activation function $f(\theta)$ can be calculated as

$$\boxed{f(\theta) = -\sin(\theta)} \tag{5}$$

which is periodic function and has an infinite number of active regions. This explains why it is more likely for the quantum activation function $f(\theta)$ to avoid gradient vanishing problem.



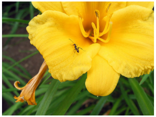
Experiment 3: Quantum Entanglement

- **Dataset**

- **Dataset**
  - MNIST

- **Dataset**
  - MNIST
  - Hymenoptera

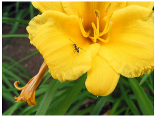
bee



ant

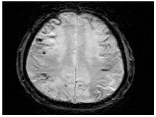
- **Dataset**
  - MNIST
  - Hymenoptera

bee

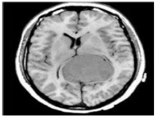


ant

  - Brain Tumor



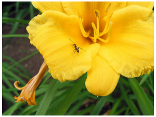
no tomor



meningioma tumor

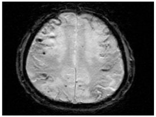
- **Dataset**
  - MNIST
  - Hymenoptera

bee

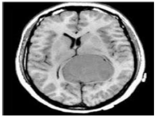


ant

  - Brain Tumor



no tomor



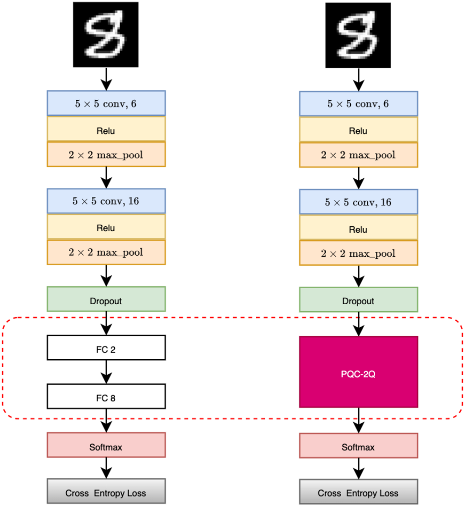
meningioma tumor

  - Chessman

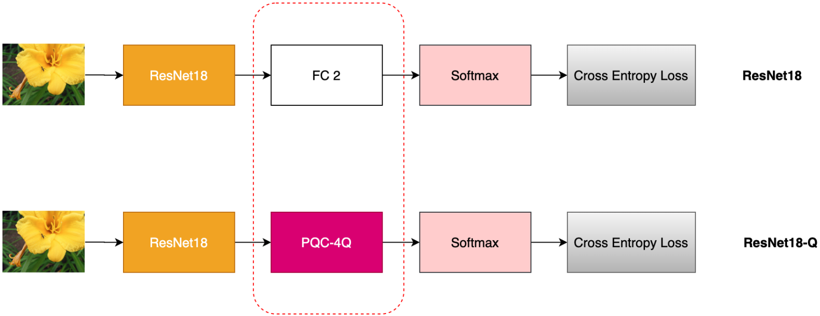
king

bishop

- **Model Architecture**:



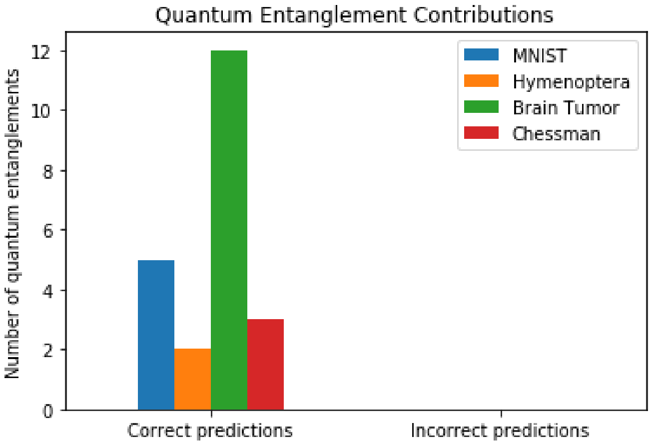
LeNet **LeNet-Q**


ResNet18
FC 2
Softmax
Cross Entropy Loss
ResNet18
ResNet18
PQC-4Q
Softmax
Cross Entropy Loss
ResNet18-Q

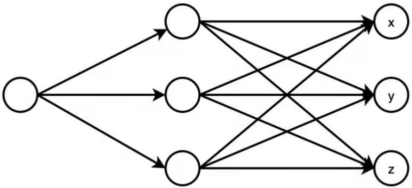
- **Result**

Dataset	Method	Test Acc
MNIST	LeNet	0.8750
	LeNet-Q	0.8925
	LeNet-Q-Ent	**0.9150**
Hymenoptera	ResNet18	0.9477
	ResNet18-Q	0.9608
	ResNet18-Q-Ent	**0.9673**
Brain Tumor	ResNet18	0.8727
	ResNet18-Q	0.8950
	ResNet18-Q-Ent	**0.9364**
Chessman	ResNet18	0.8824
	ResNet18-Q	0.8824
	ResNet18-Q-Ent	**0.9118**

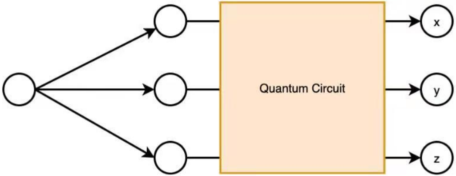
Hybrid quantum-classical neural nets generally perform better than classical nets. In particular, HQNNs with quantum entanglement (ResNet18-Q-Ent) achieve best performances.



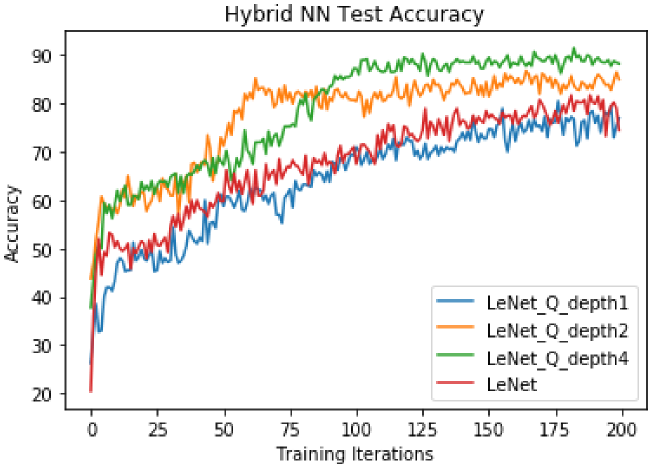
All quantum entanglements have contributed to correct predictions



In classical neural networks, neurons x, y and z are independently calculated from each other.



In hybrid quantum-classical neural networks, neurons x, y and z can interact with each other thanks to the quantum entanglement.



Models with higher quantum depth have lower test error

Device	Batch Time
Pennylane default_qubit simulator	0.3093s
Qiskit qasm_simulator	0.8687s
Amazon SV1 simulator	26.4140s
Rigetti Aspen-9	86.5870s

Batch times are obtained from the measurement in experiments of training across different quantum devices the model *ResNet18-Q-Ent* (4 qubits, 4 parameters, 2 quantum layers) with batch size of 4 . All experiments are conducted using Amazon Braket Service. **It can be seen that remote simulators (e.g. SV1) do not have the advantage on small circuits since the latency times of communicating the circuit to AWS dominate over simulation times**

Conclusions

Conclusions

- Thanks to Pennylane, it is feasible to integrate techniques from machine learning and quantum computing.

Conclusions

- Thanks to Pennylane, it is feasible to integrate techniques from machine learning and quantum computing.

- Parameterized quantum circuits might be considered as a good option for activation functions in the neural network models as they help avoid gradient vanishing.

Conclusions

- Thanks to Pennylane, it is feasible to integrate techniques from machine learning and quantum computing.
- Parameterized quantum circuits might be considered as a good option for activation functions in the neural network models as they help avoid gradient vanishing.
- Quantum entanglement can help potentially improve performance of classical neural network models in the context of image recognition.