Knowledge Distillation in Quantum Neural Network using Approximate Synthesis

ASP-DAC 2023

Mahabubul Alam, Satwik Kundu and Swaroop Ghosh
School of EECS, The Pennsylvania State University, USA

PennState
Lab of Green and Secure Integrated Circuit Systems (LOGICS)

Outline

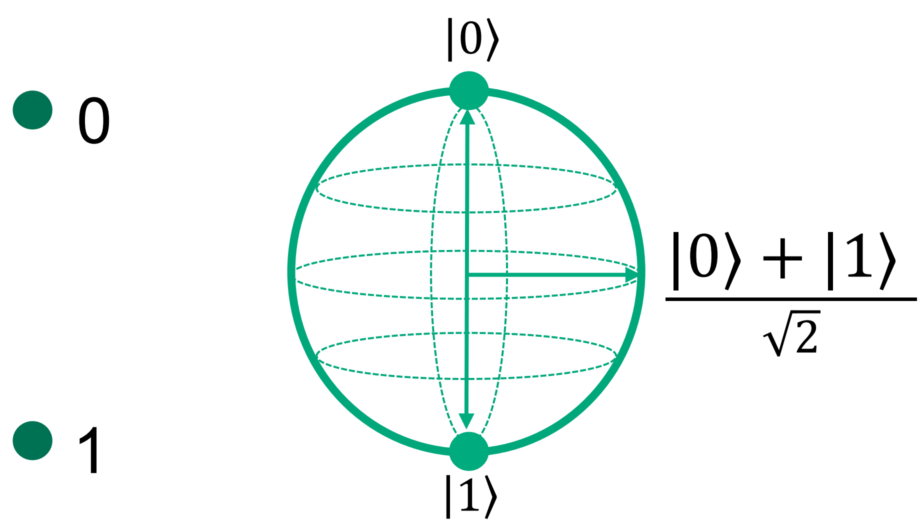
- Basics of Quantum Computing
- Quantum Neural Network (QNN)
- Quantum Circuit Compilation
- QNN adaptability challenges?
- Proposed Methodology
- Results
- Conclusion

Outline

- Basics of Quantum Computing
- Quantum Neural Network (QNN)
- Quantum Circuit Compilation
- QNN adaptability challenges?
- Proposed Methodology
- Results
- Conclusion

Qubits

- *Qubits:* building block of quantum computers
- Qubits can be at 0 or 1 or both simultaneously (i.e., superposition)
- Mathematically, single qubit can be described by a two-dimensional column vector $\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$.
- $|\alpha|^2$ and $|\beta|^2$ represents the probability of getting a 0 and 1 $(\text{hence}, |\alpha|^2 + |\beta|^2 = 1)$

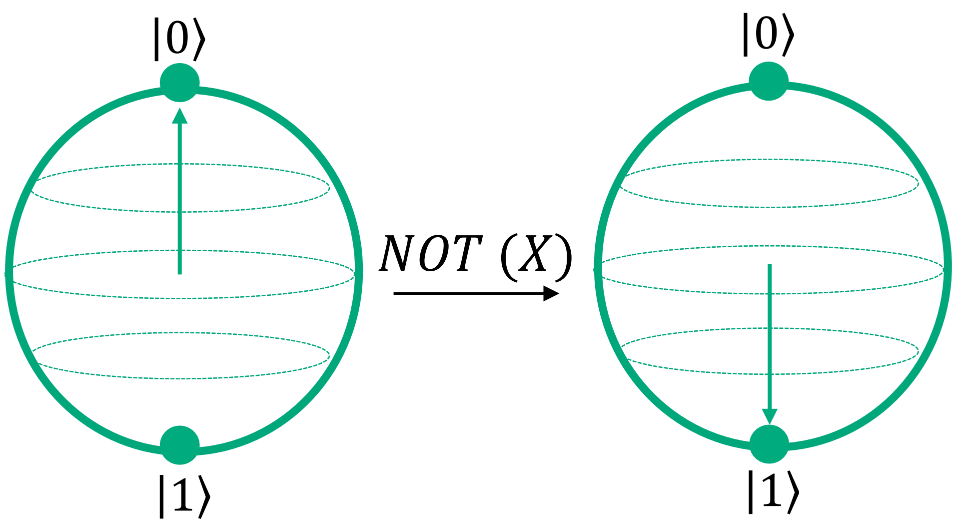


Classical Bit

Qubit

Quantum Gates

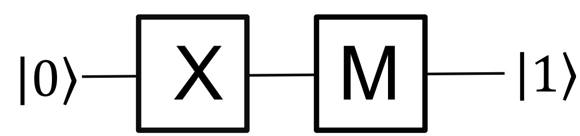
- *Quantum gates:* realized with (control) pulses (RF pulse, Laser pulse etc.)
- Realized gates (so far): 1-qubit and 2-qubit
- Quantum gates are mathematically represented using matrices.
- Example; NOT (X) gate → $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ & Hadamard (H) gate → $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$



Physical representation

$$X|0\rangle = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = |1\rangle$$

Mathematical representation



Circuit representation

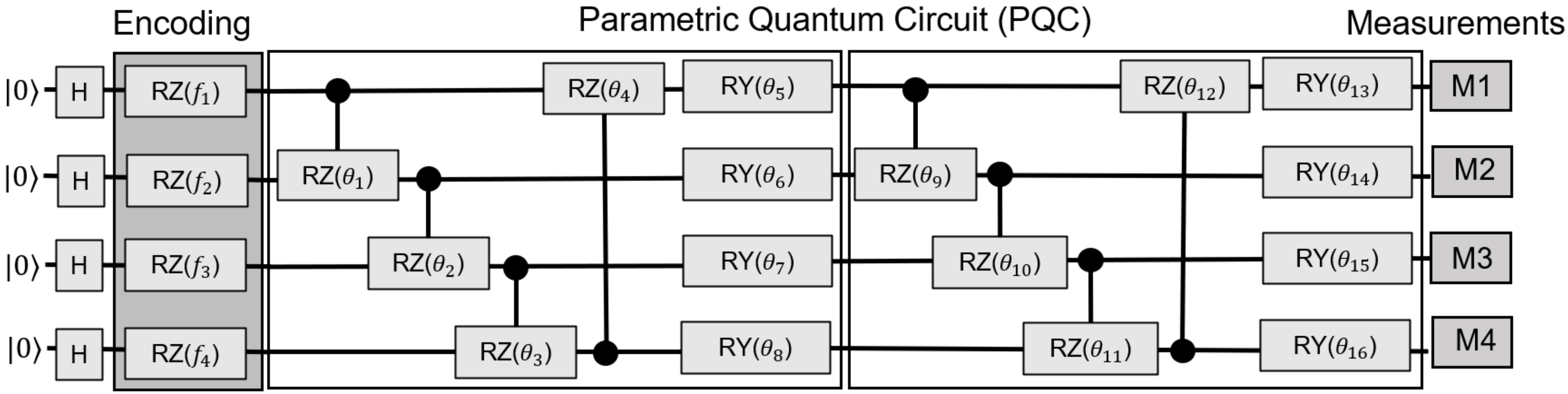
Quantum Noise

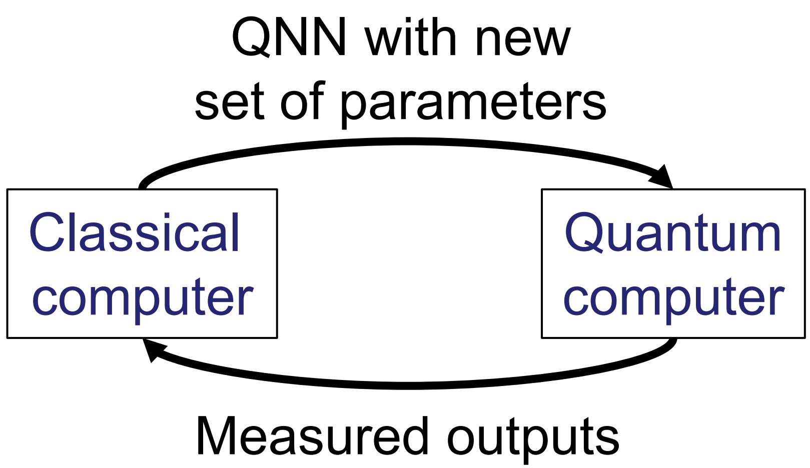
- *Quantum noises/errors* impedes quantum computers reliability/performance;
  - *Coherence errors:* qubits retain its state for a short period of time
  - *Gate errors:* gate operations are erroneous
  - *Measurement errors:* measured value are often erroneous
  - *Crosstalk errors:* Parallel execution of multiple gates on different qubits hinders each other's performance
- *High gate count → more gate error*
- *Deeper circuit → large execution time → more decoherence error*

Outline

- Basics of Quantum Computing
- Quantum Neural Network (QNN)
- Quantum Circuit Compilation
- QNN adaptability challenges?
- Proposed Methodology
- Results
- Conclusion

Quantum Neural Network





A classical optimizer (gradient-based/gradient-free) updates the trainable parameters ($\theta_1 - \theta_{16}$) to generate a desired output distribution.

Quantum Neural Network

Encoding Techniques

Angle Encoding

Amplitude Encoding

Basis Encoding

NEQR

...

Measurement

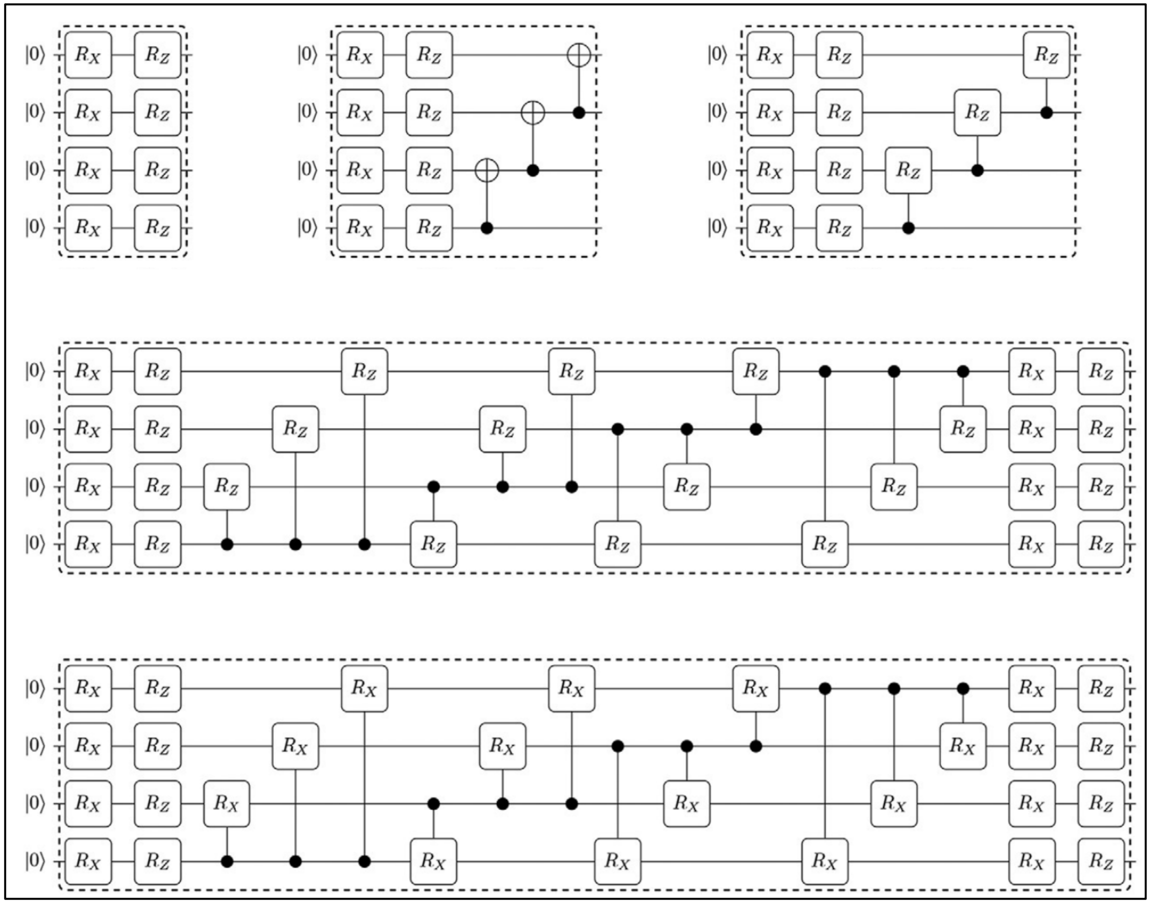
Pauli X Basis

Pauli Y Basis

Pauli Z Basis

...

Example of few 4-qubit PQCs [1]



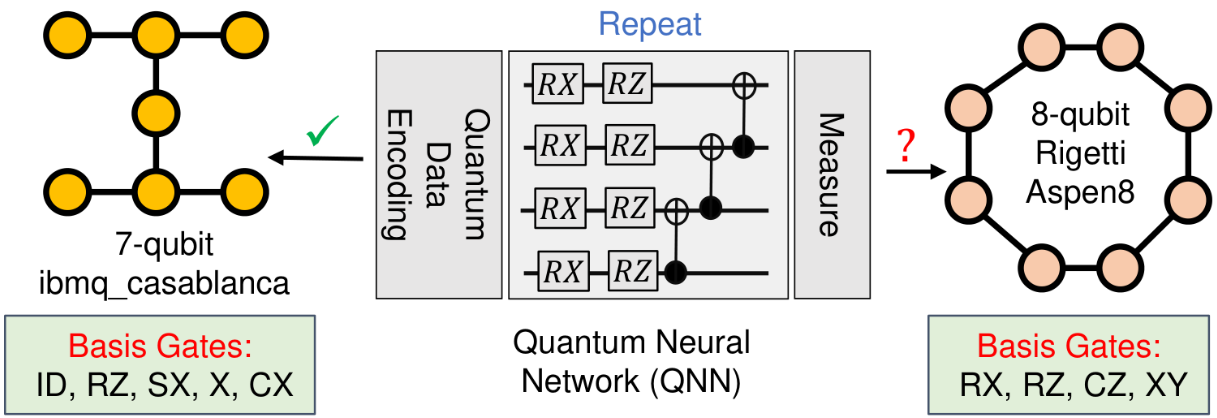
Sukin Sim et al. "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms." *Advanced Quantum Technologies* 2, no. 12 (2019)

Outline

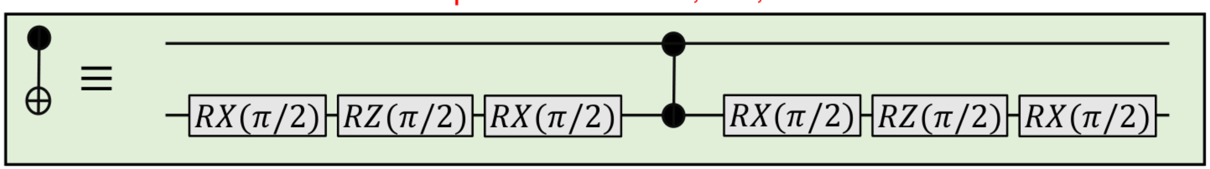
- Basics of Quantum Computing
- Quantum Neural Network (QNN)
- Quantum Circuit Compilation
- QNN adaptability challenges?
- Proposed Methodology
- Results
- Conclusion

Quantum Circuit Compilation

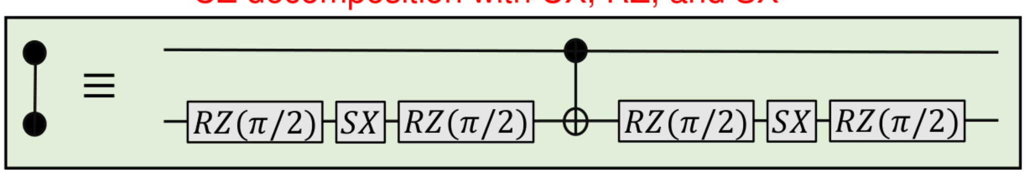
Different quantum hardware can have different set of basis gates.



CX Decomposition with CZ, RZ, and RX

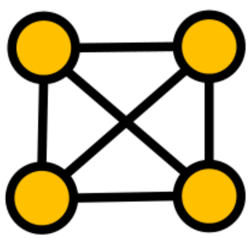


CZ decomposition with CX, RZ, and SX

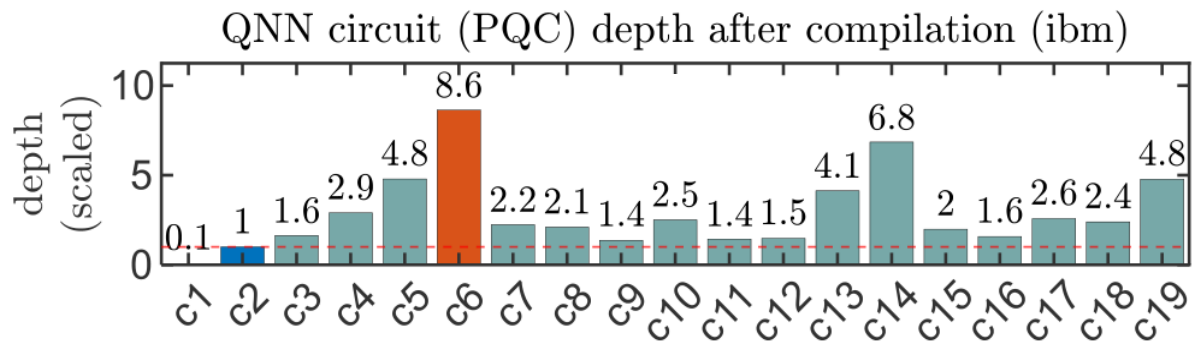


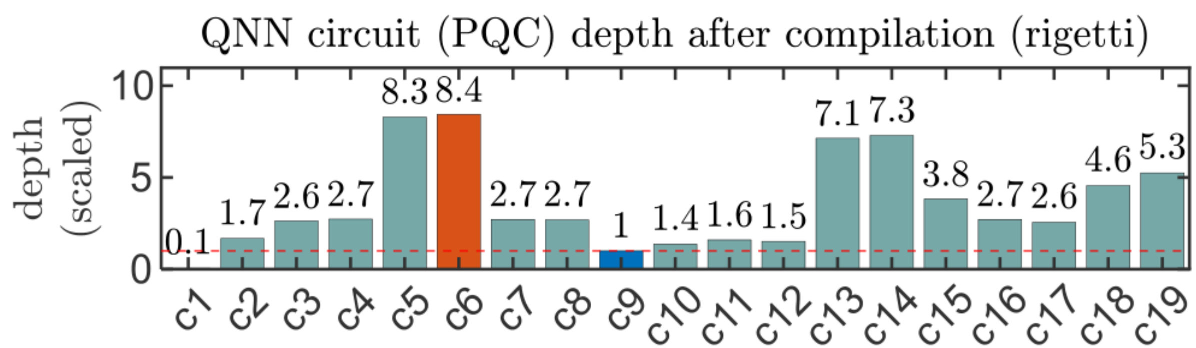
CX and CZ gate decomposition on Rigetti and IBM devices, respectively.

Circuit Depth (after compilation)



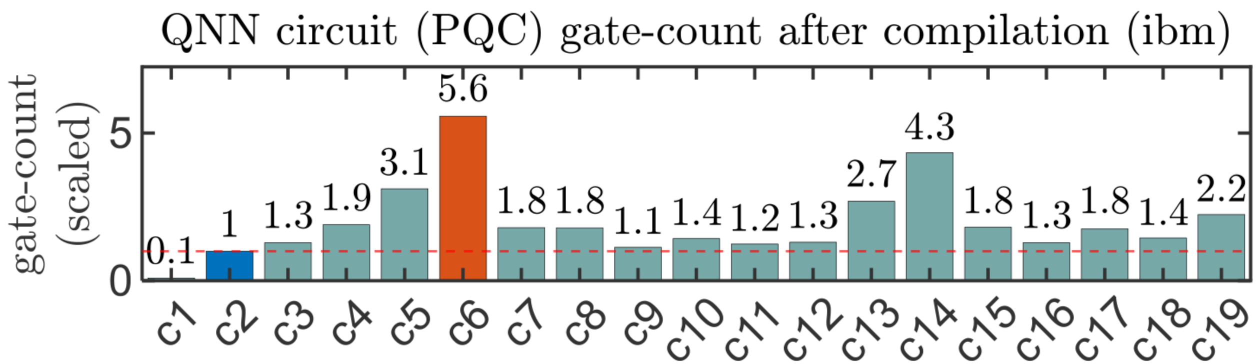
Hypothetical
4-qubit
Hardware
Coupling Graph

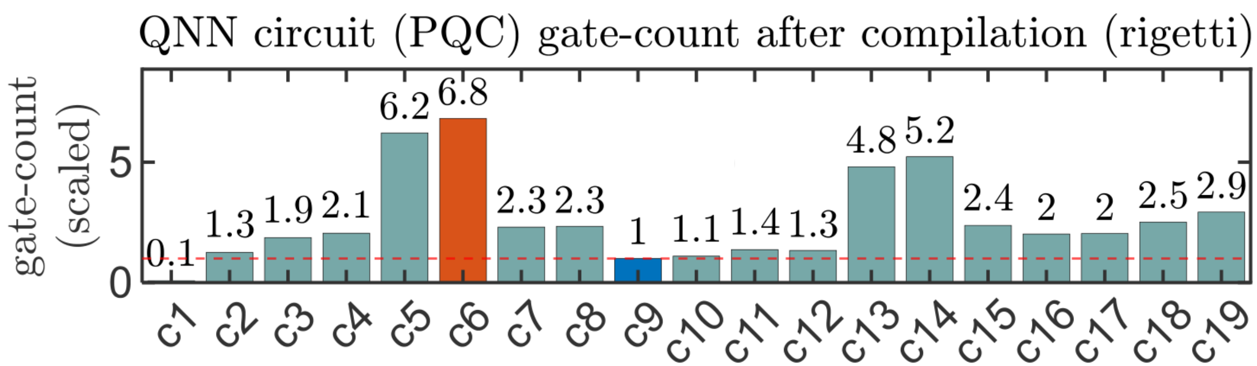




Sukin Sim et al. "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms." *Advanced Quantum Technologies* 2, no. 12 (2019)

Circuit Gate-count (after compilation)





Sukin Sim et al. "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms." *Advanced Quantum Technologies* 2, no. 12 (2019)

Outline

- Basics of Quantum Computing
- Quantum Neural Network (QNN)
- Quantum Circuit Compilation
- QNN adaptability challenges?
- Proposed Methodology
- Results
- Conclusion

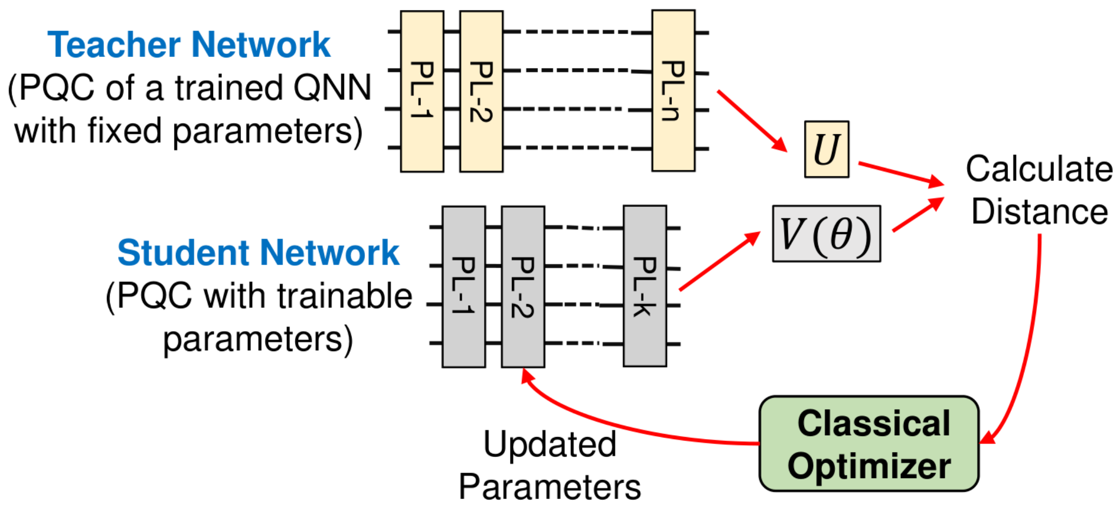
QNN adaptability challenges?

- Noise adaptability
  - Near-term quantum devices have a limited number of qubits and suffer from various errors.
  - Noises build up quickly as a circuit is scaled up.
  - Thus, a deep QNN may perform poorly on an actual hardware compared to a shallower network.
- Hardware adaptability
  - Different quantum hardware may exhibit varying degrees of noise and may support a completely different set of basis gates.
  - As a result, porting a QNN designed for one piece of hardware to another is problematic.

Outline

- Basics of Quantum Computing
- Quantum Neural Network (QNN)
- Quantum Circuit Compilation
- QNN adaptability challenges?
- Proposed Methodology
- Results
- Conclusion

Overview



In traditional knowledge distillation, a pre-trained large DNN is used as a guide to train a smaller DNN without sacrificing much performance.

We avoid this costly training technique by employing approximate synthesis to mimic the pre-trained QNN behavior.

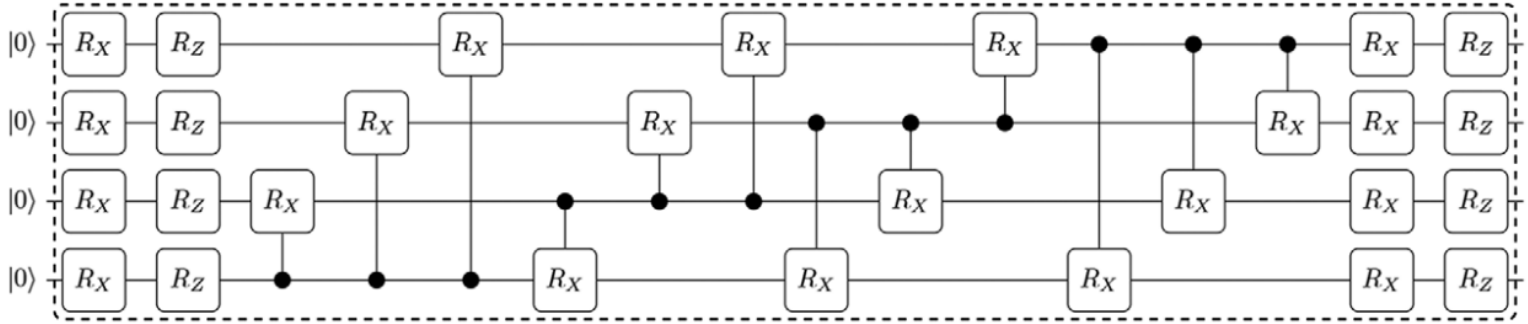
Approximate Synthesis

- An n-qubit quantum circuit can be represented by a unitary matrix U of size $2^n \times 2^n$.
- Idea is to find a new circuit with a different set of gates with unitary matrix representation (V).
- A synthesis algorithm tries to minimize the distance between U and V ($||U-V||$) using some metric.
- We use the following metric as cost function:

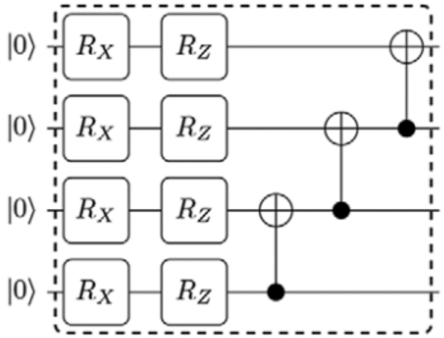
$$d = 1 - \frac{Tr(U^{-1}V)}{\dim(U)}$$

- Dual annealing optimizer is used to minimize the distance d.

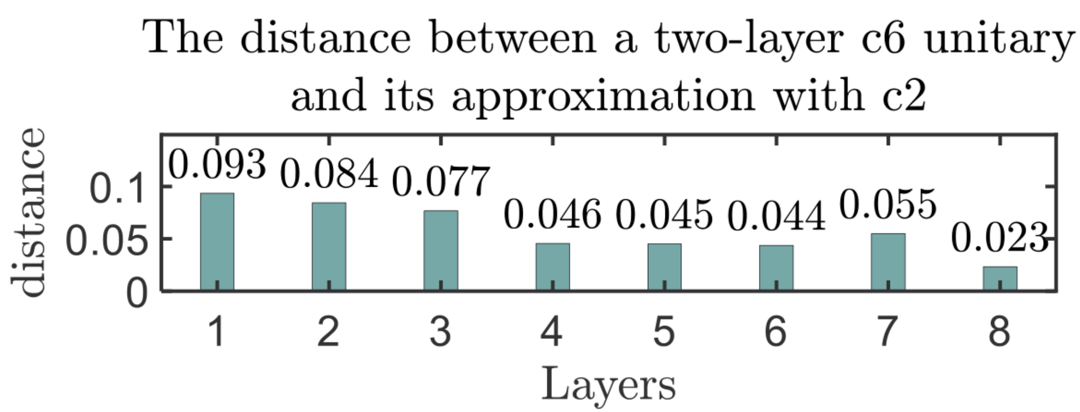
Approximate Synthesis



Circuit 6 [1]

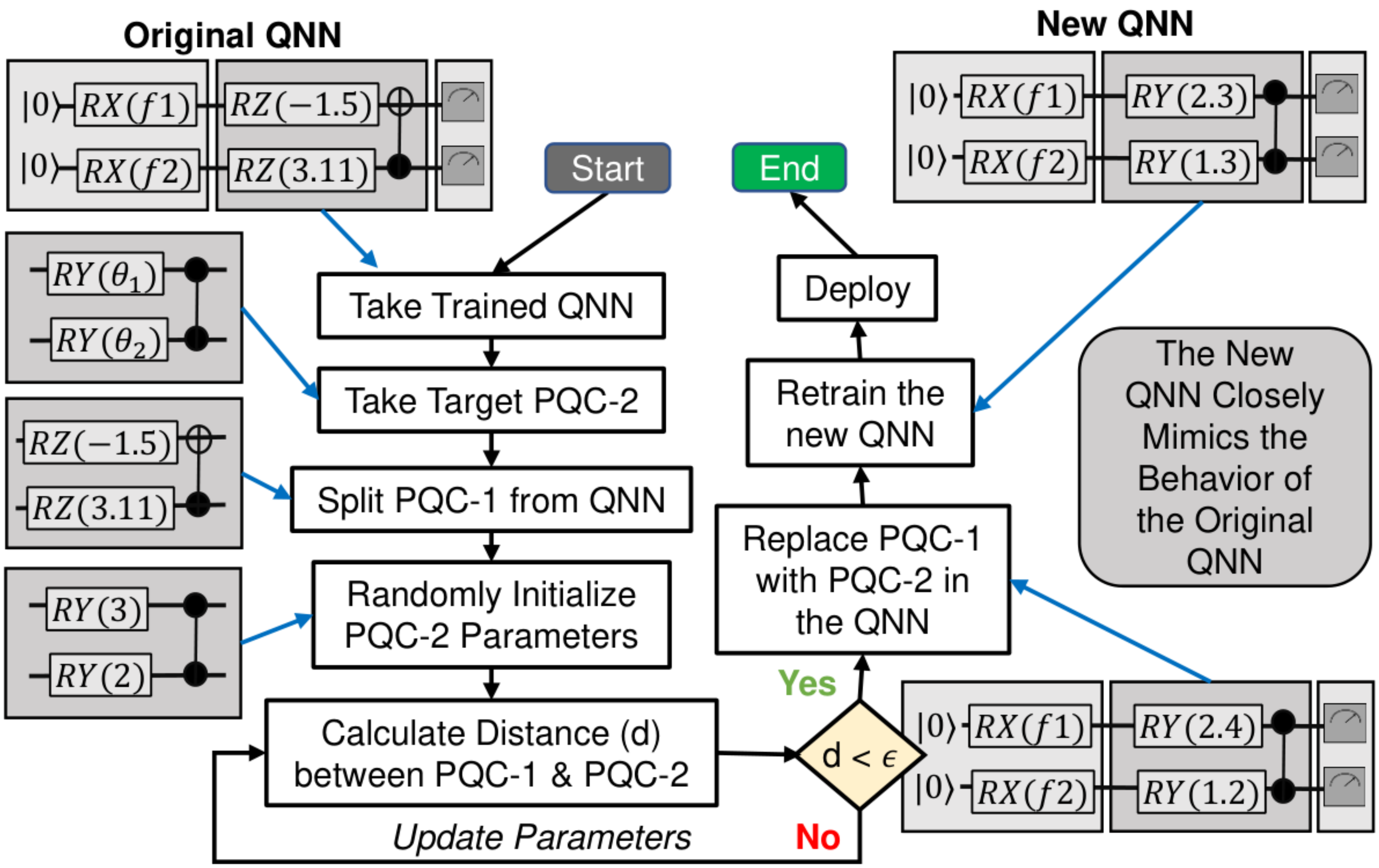


Circuit 2 [1]



[1] Sukin Sim et al. "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms." *Advanced Quantum Technologies* 2, no. 12 (2019)

Methodology

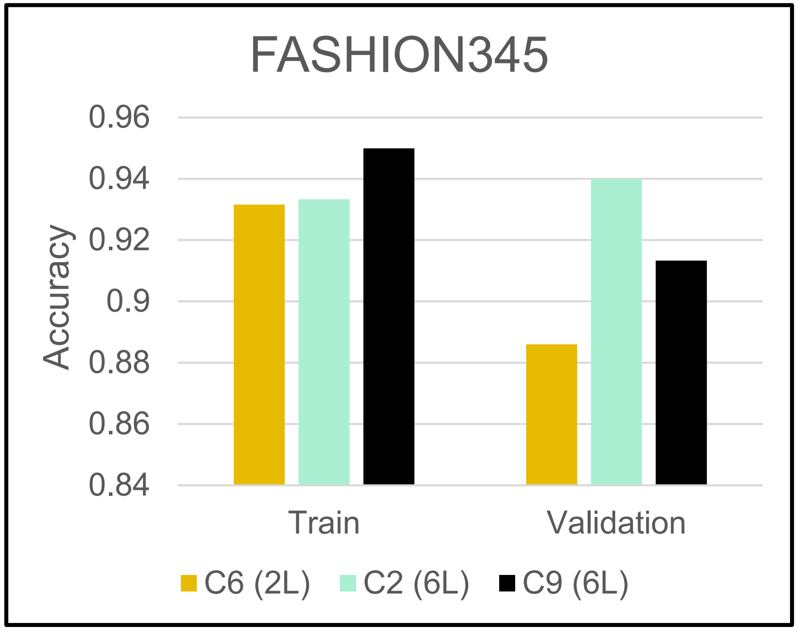


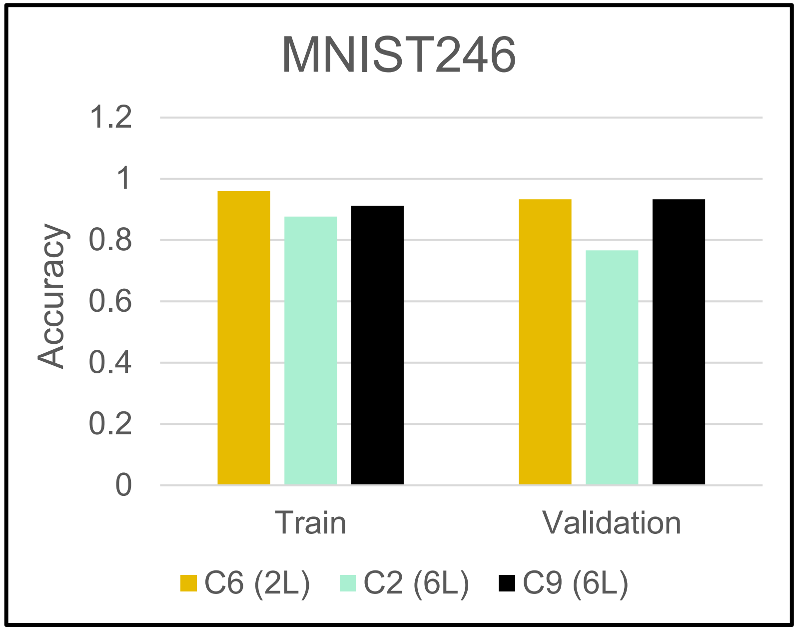
Outline

- Basics of Quantum Computing
- Quantum Neural Network (QNN)
- Quantum Circuit Compilation
- QNN adaptability challenges?
- Methodology
- Results
- Conclusion

Results

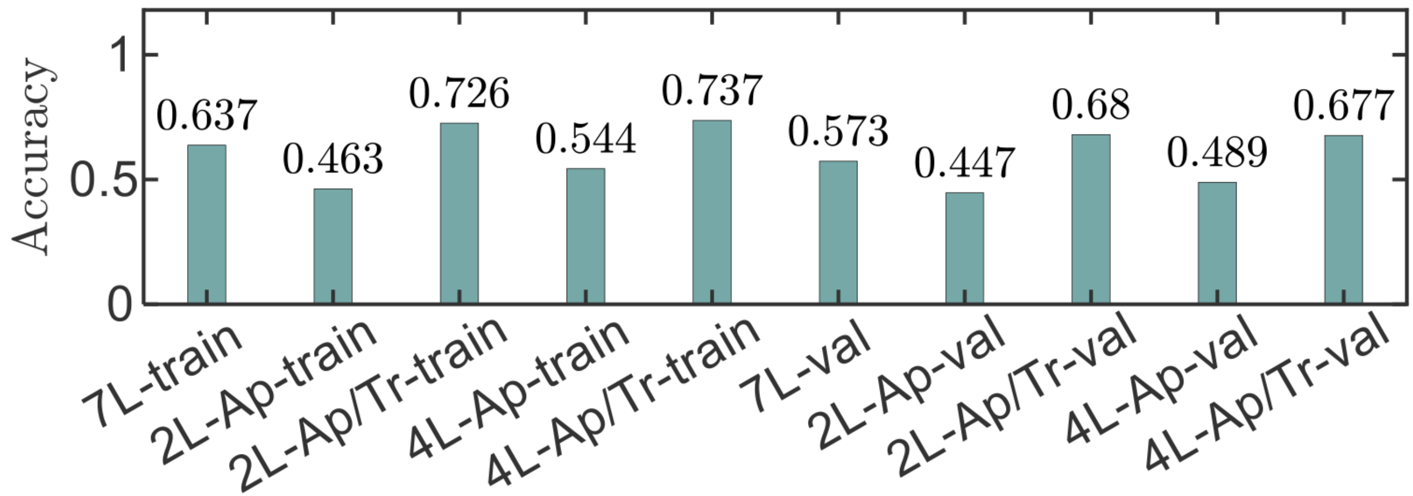
- Datasets: 6 (3-class Classification)
  - Created from MNIST and Fashion-MNIST
  - Latent dimension: 8 (reduced using convolutional autoencoder)
- Qubits in QNN: 4
- Approximated 2-layer C6 PQC's with 6-layer C2 and C9 PQC's (Noiseless simulations).





Results

Also approximated 7-layer C15 PQC with 2 & 4-layer of the same PQC. (Noisy simulations).



Outline

- Basics of Quantum Computing
- Quantum Neural Network (QNN)
- Quantum Circuit Compilation
- QNN adaptability challenges?
- Methodology
- Results
- Conclusion

Conclusion

- We introduced knowledge distillation in QNNs using approximate synthesis to compress a pre-trained QNN.
- Approximation error can degrade the performance of QNNs which can be compensated by training for a few epochs.
- Conducted comprehensive numerical analysis on multiple datasets with and without noise to evaluate the efficacy of the proposed method.
- Through empirical analysis, we demonstrate ≈71.4% reduction in circuit layers, and still achieve ≈16.2% better accuracy under noise.

Thank you!
We acknowledge and appreciate the support from our sponsors:

PennState
Institute for Computational and Data Sciences

NSF

DARPA

SRC
Semiconductor Research Corporation