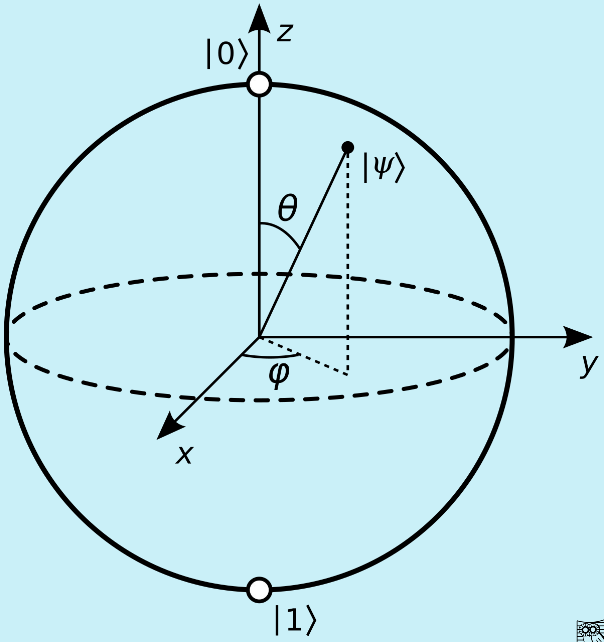
Quantum Neural Networks Need Checkpointing

Christopher Kverne, Mayur Akewar, Yuqian Huo, Tirthak Patel, Janki Bhimani

ckver001@fiu.edu

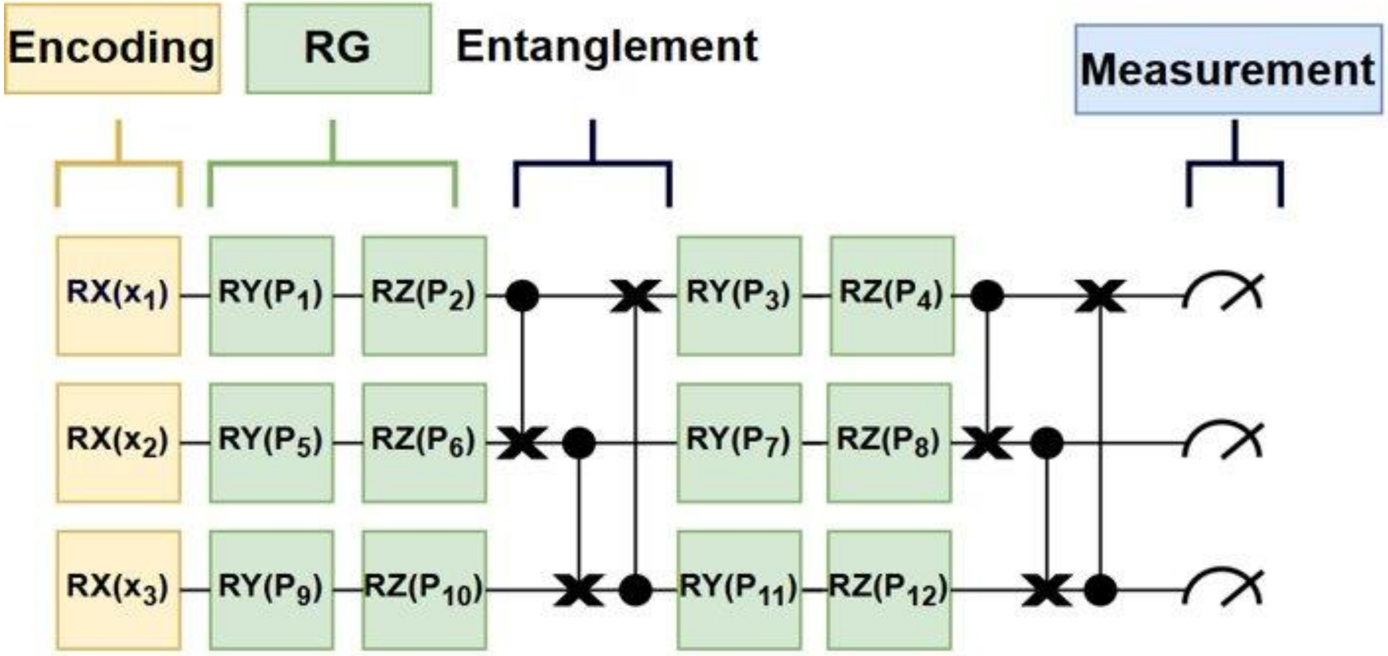
Quantum Computing Intro

- Qubits are Quantum Particles
- Manipulate Qubits
- Superposition + Entanglement
- When measured collapses into the $|0|$ or $|1|$ state
- Multiple measurement / circuit evaluations needed ("shots")



Quantum Machine Learning

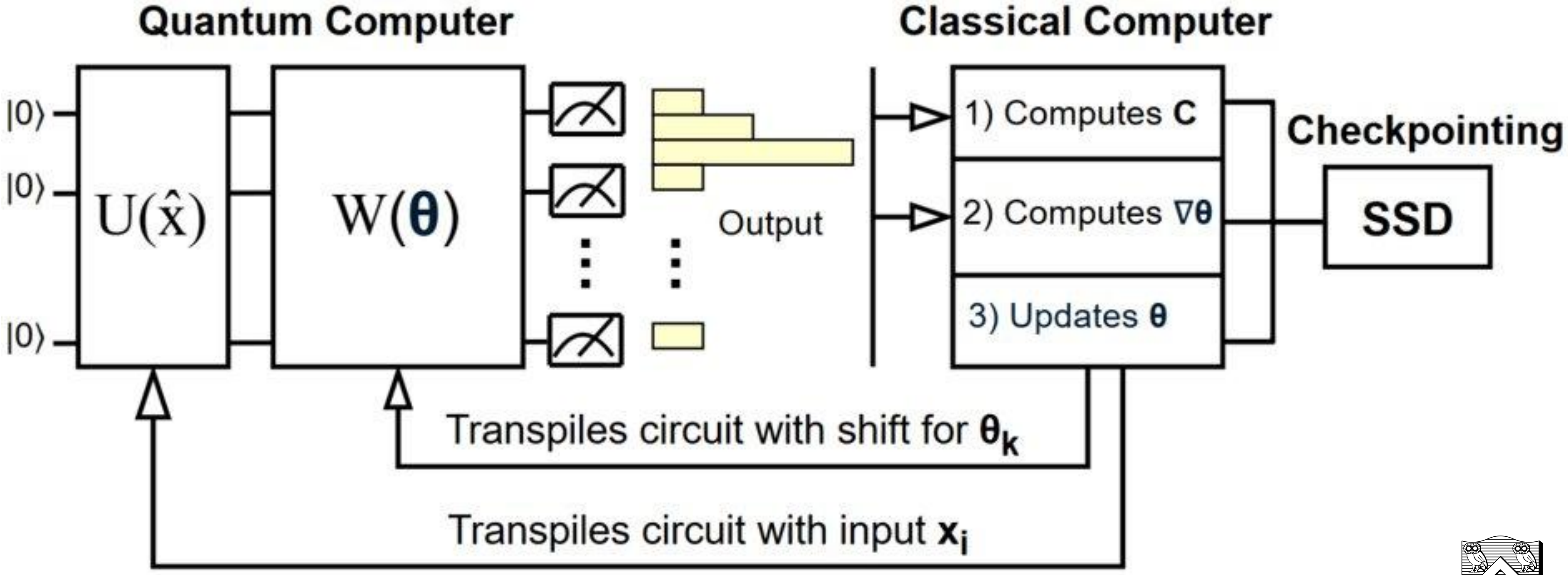
QNN = Quantum Circuit Capable of Universal Approximation
Trainable Gates (Rx, Ry, Rz) + Entanglement Gates (CNOT)



FIU

RICE

Quantum-Classical Training



FIU DaMRL RICE

Why is Quantum Checkpointing Important?

Crash

Popular (Long queues & short sessions)

Reproducibility

FIU DaMRL RICE

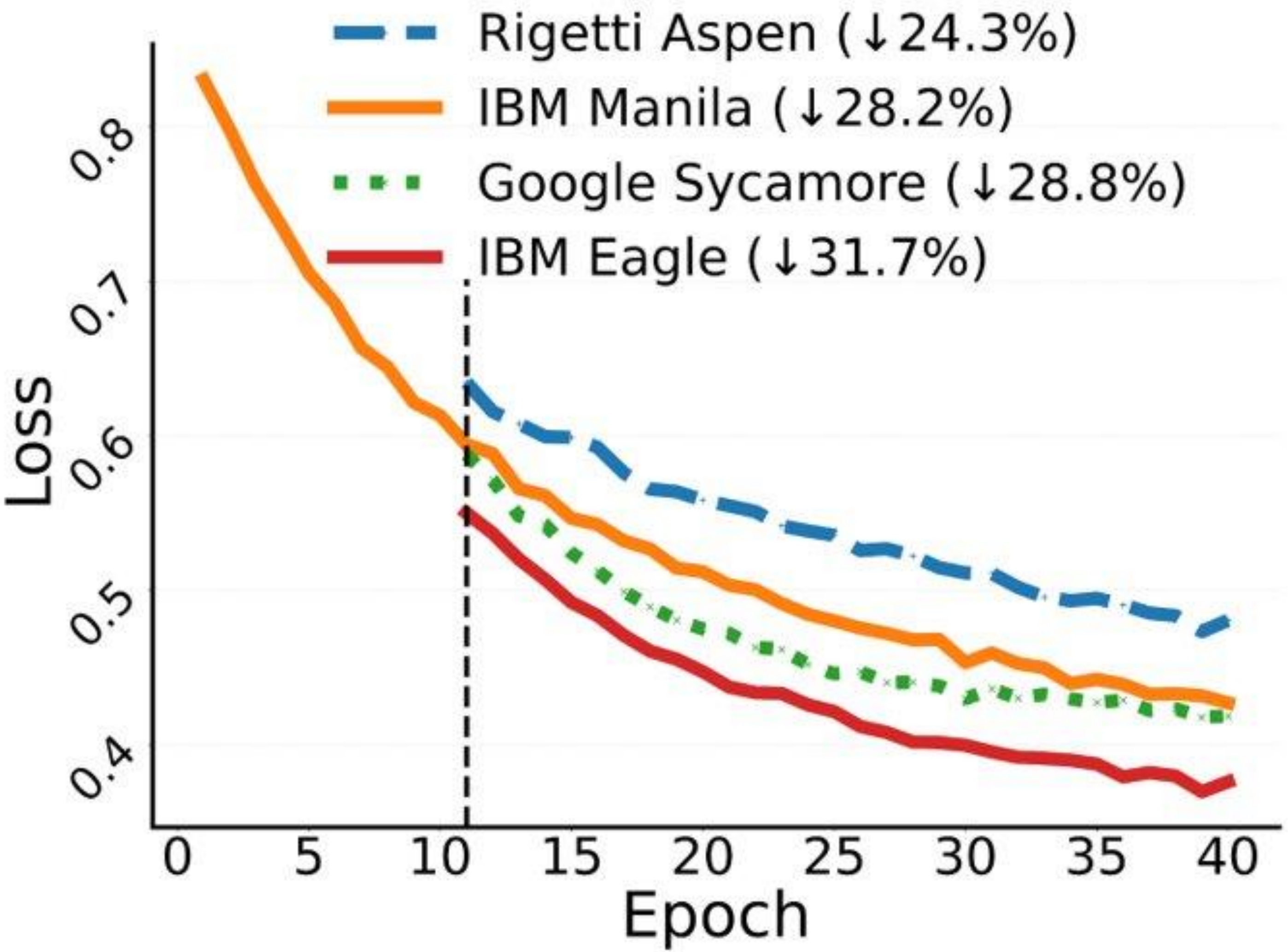
Problem: No one has defined what a Quantum Checkpoint is.

- What's the difference between quantum and classical checkpointing?
- What should a quantum checkpoint store?
- What happens when quantum checkpoints migrate across QC's?

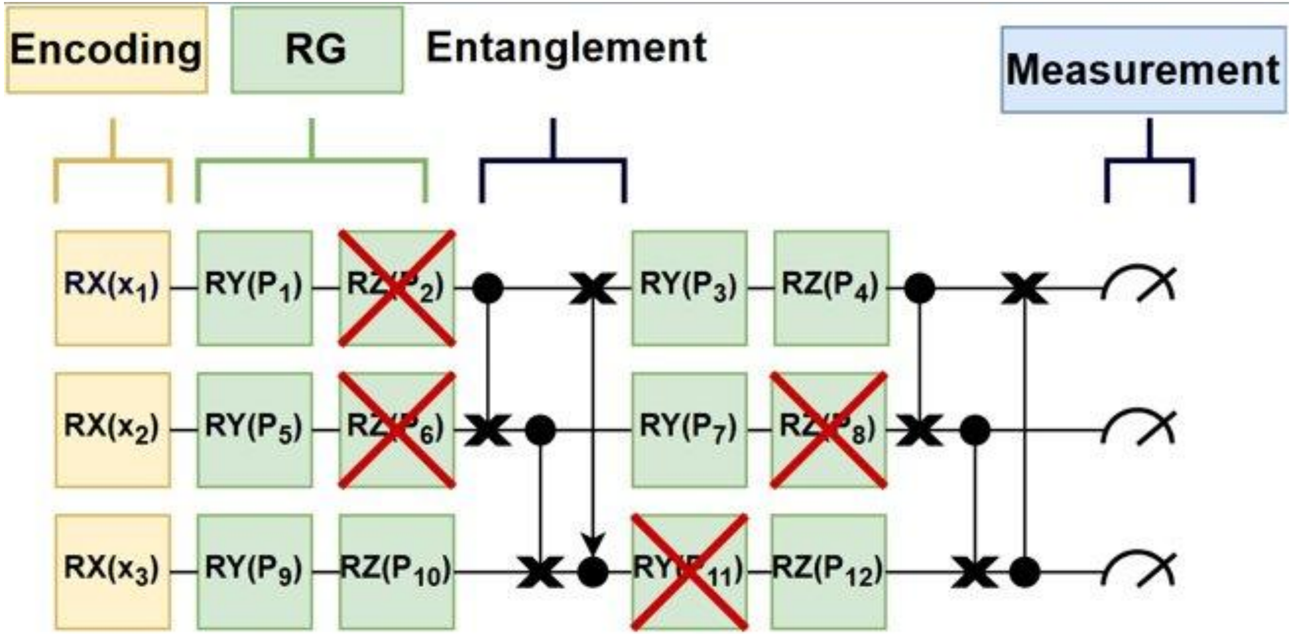
FIU DaMRL RICE

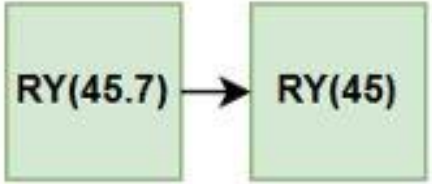
Experiment 1: QNN's on different quantum hardware.

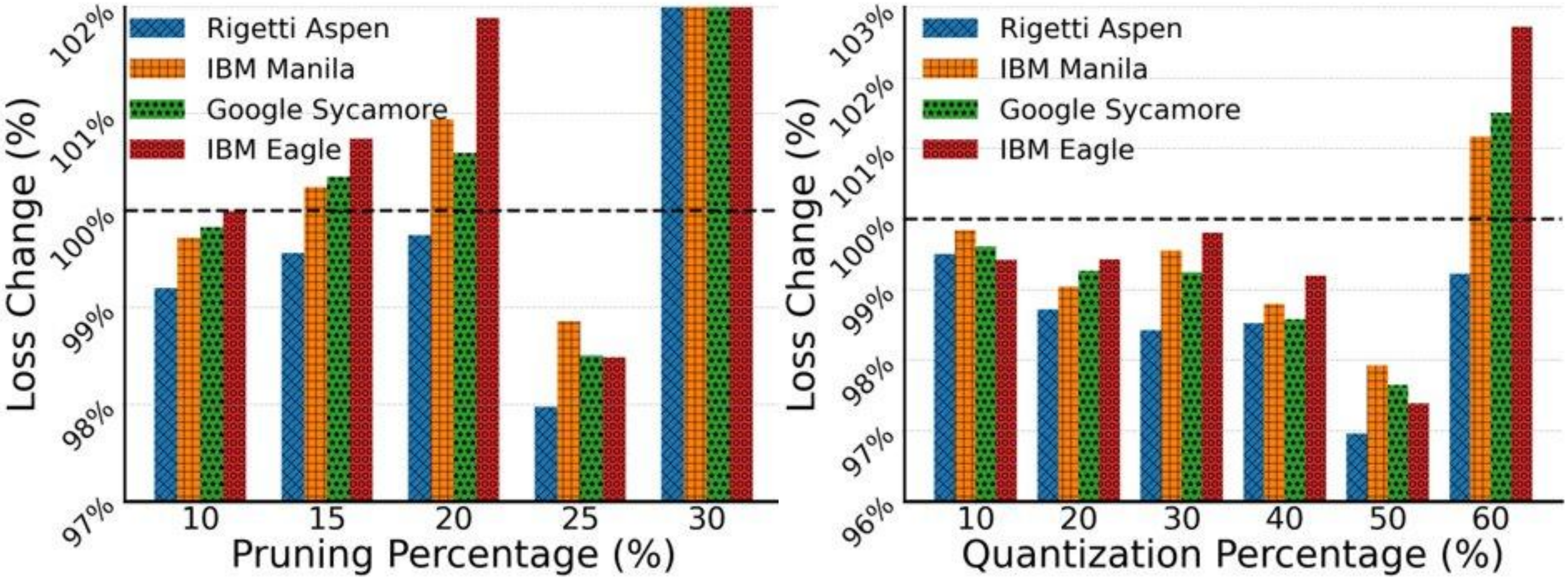
Single-qubit gate errors (e.g., Rx gate errors)
Two-qubit gate errors (e.g., *CNOT* gate errors)
$T1$ relaxation times (energy decay)
$T2$ dephasing times (loss of coherence)
Readout errors (measurement)


Rigetti Aspen (↓24.3%)
IBM Manila (↓28.2%)
Google Sycamore (↓28.8%)
IBM Eagle (↓31.7%)
Loss
0.4
0.5
0.6
0.7
0.8
0
5
10
15
20
25
30
35
40
Epoch

Experiment 2: Circuit Optimization on QNN's







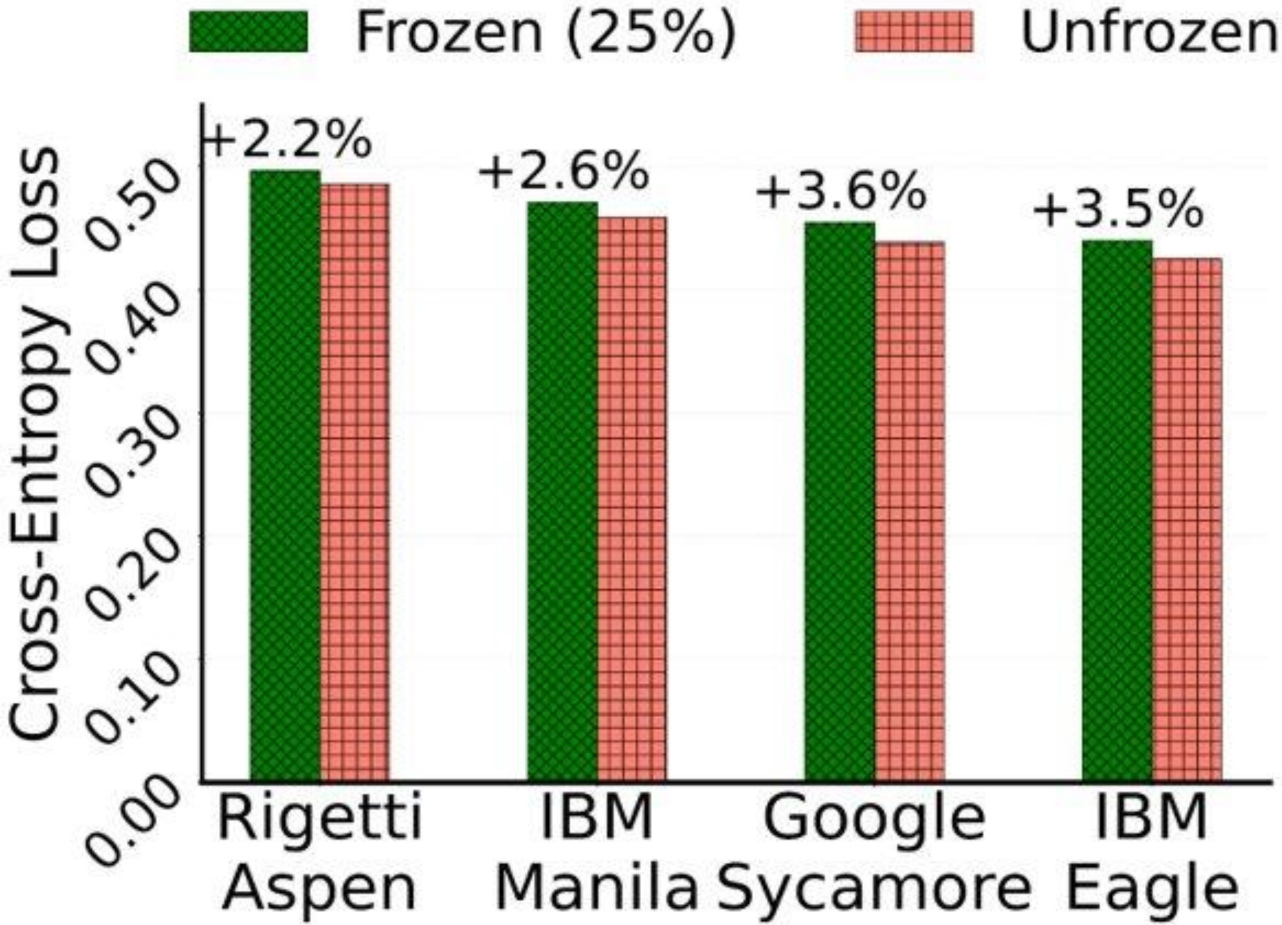
Loss Change (%)
102%
101%
100%
99%
98%
97%
Rigetti Aspen
IBM Manila
Google Sycamore
IBM Eagle
10
15
20
25
30
Pruning Percentage (%)
Loss Change (%)
103%
102%
101%
100%
99%
98%
97%
96%
Rigetti Aspen
IBM Manila
Google Sycamore
IBM Eagle
10
20
30
40
50
60
Quantization Percentage (%)

Experiment 3: Parameter Freezing

$$\frac{\partial C(\boldsymbol{\theta})}{\partial \theta_k} = \varsigma_k \left[C\left(\boldsymbol{\theta} + \frac{\pi}{4\varsigma_k}\mathbf{e}_k\right) - C\left(\boldsymbol{\theta} - \frac{\pi}{4\varsigma_k}\mathbf{e}_k\right)\right]$$

Freezing k parameter reduces the number of forward passes per input by 2k (Speeds up training significantly)

FIU DaMRL RICE


Frozen (25%)
Unfrozen
Cross-Entropy Loss
0.00
0.10
0.20
0.30
0.40
0.50
+2.2%
+2.6%
+3.6%
+3.5%
Rigetti Aspen
IBM Manila
Google Sycamore
IBM Eagle

Quantum Checkpoints Need Hardware Data!

FIU DaMRL RICE

Why?

1) If optimization is done on a QNN for system A system migrating it to system B is difficult
2) If you download a checkpoint but suddenly see different values than reported (accuracy/loss) the hardware data can give insights into why.
3) Parameter Freezing on a QNN for system A needs to be adjusted for system B

Proposed Checkpoint

Classical

1) Parameters
2) Optimizer State

Quantum

3) Number of Shots
4) Hardware Info

Size of Quantum Checkpoint

Classical

1) Parameters:
 l*q*g
2) Optimizer State:
2*l*q*g (Adam)

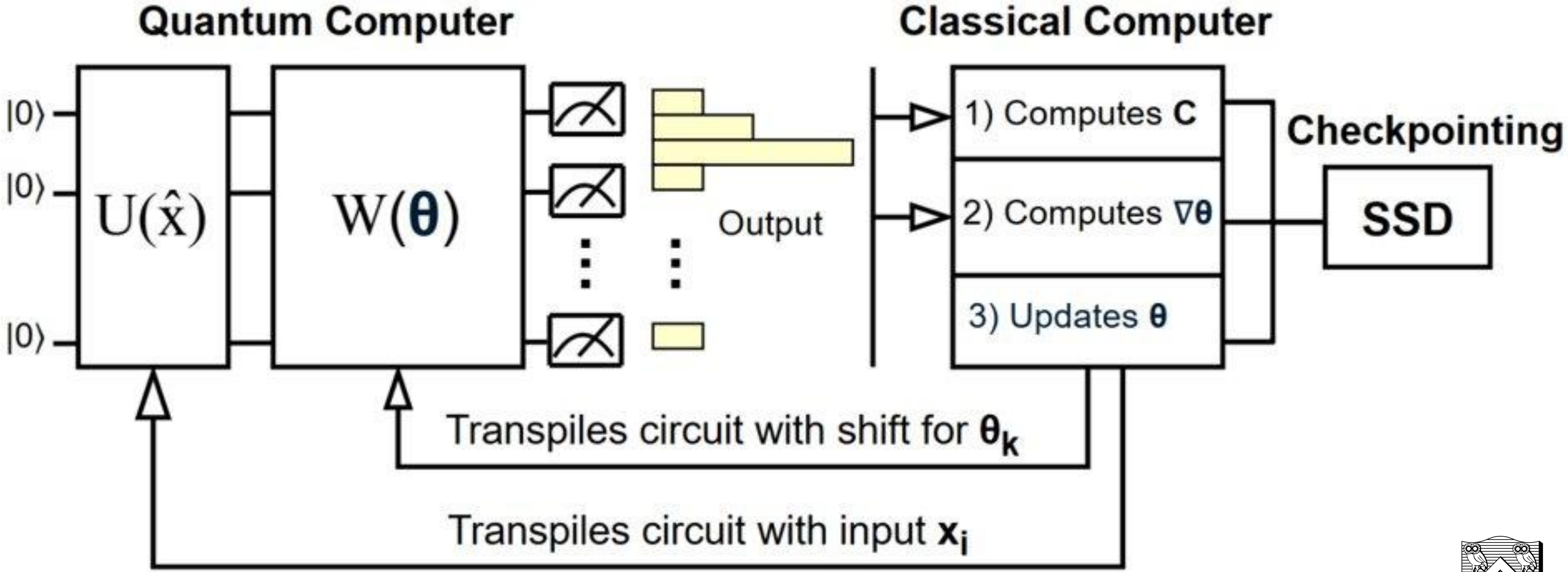
Quantum

3) Number of Shots
int (e.g., 1000)
4) Hardware Info
4q + e:
4q: q single, q readout, q T1, q T2
e = q(q-1)/2 two qubit errors (typically O(q)

FIU DaMRL

Additional hardware data is small (as no. of parameters is typically much larger than no. of qubits)

Overhead of Quantum Checkpoint



FIU DaMRL RICE

Quantum Checkpoints are fast (Asynchronous checkpointing is easy)

Future Work

How do we adjust pruning, quantization, and freezing when migrating checkpoints?

How can we checkpoint between different types of QC (i.e. superconducting computers to neutral-atom computers. (these have different trainable parameters).

FIU DaMRL

RICE

Q & A

https://github.com/Damrl-lab/Quantum-Checkpointi
ng

FIU

DaMRL

RICE

Acknowledgements

This work was supported by the following NSF grants: CSR2402328, CAREER-2338457, CSR-2406069, CSR-2323100, and HRD-2225201, as well as the DOE Quantum Testbed Finder Award DE-SC0024301.