Improving Robustness of Quantum Computing with Machine Learning

A Hybrid Classical-Quantum Computer System Perspective

Hanrui Wang, Song Han

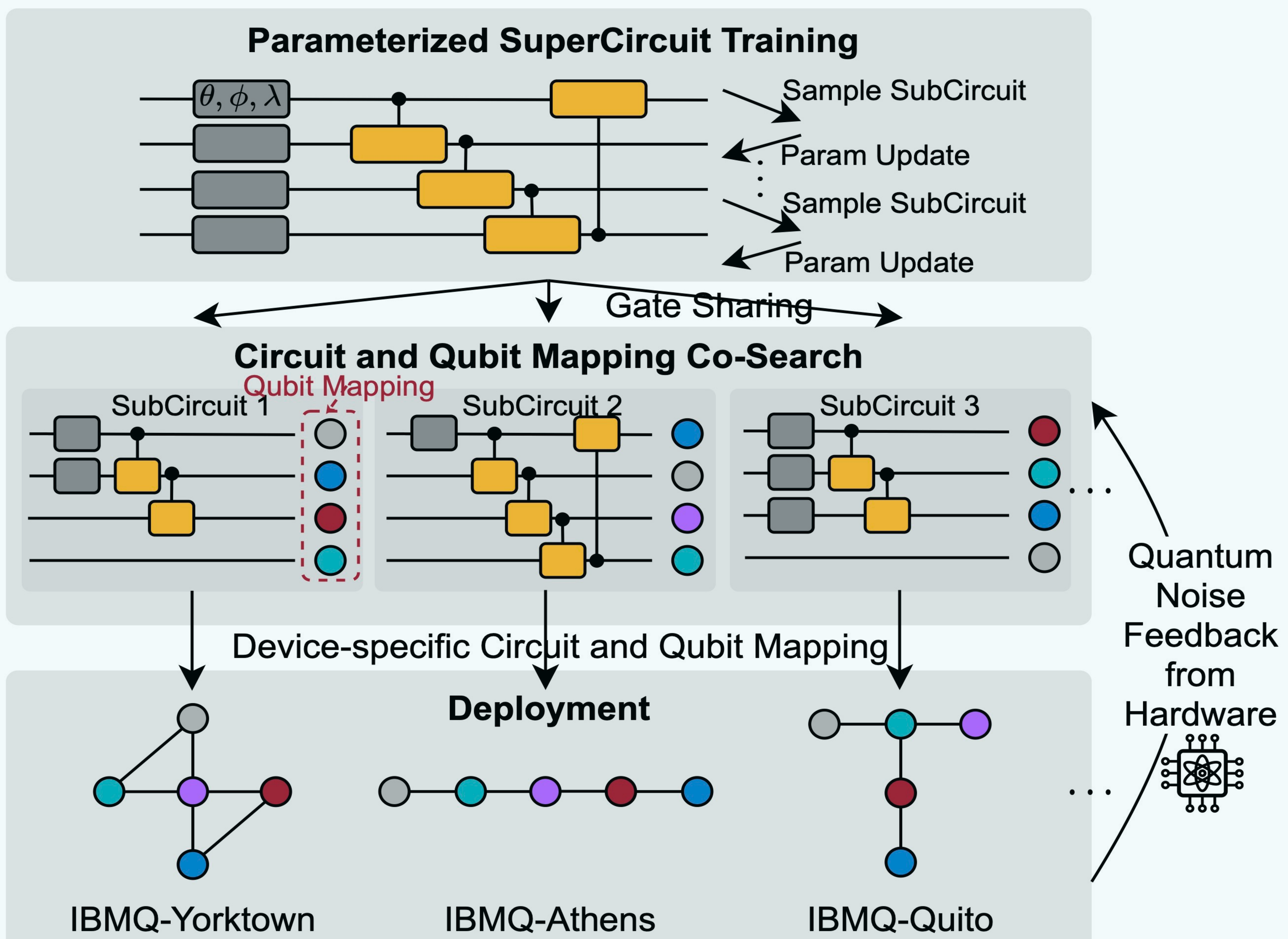
MIT AI Hardware Program

qmlsys.mit.edu

Near-Term: Variational Quantum Algorithms

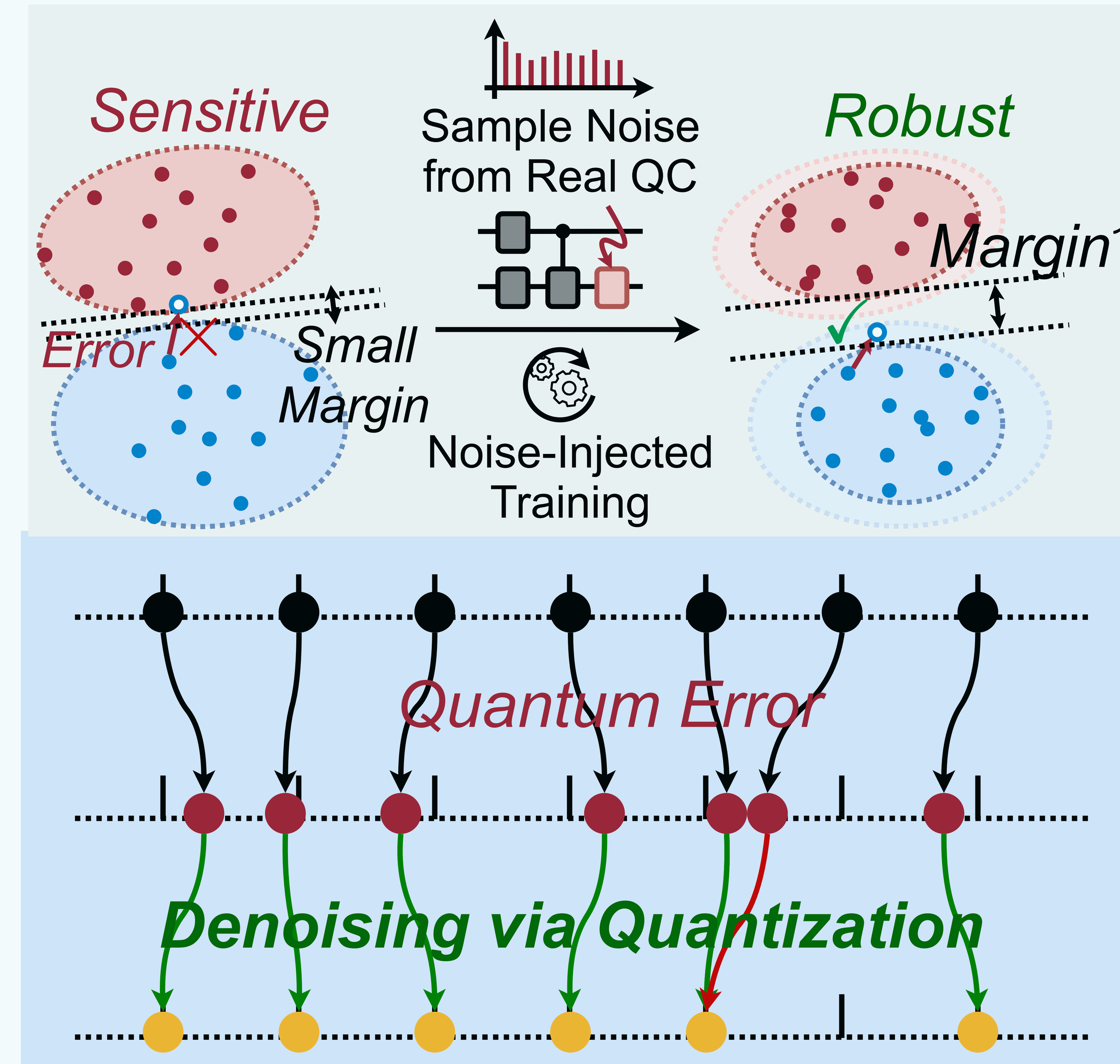
NAS for Robust Quantum Circuit Architecture

- Leverage a SuperCircuit to search for robust circuit with high efficiency
- 95% 2-class, 85% 4-class, and 32% 10-class acc on real quantum machine
- Achieve more accurate VQE eigenvalue than UCCSD baseline



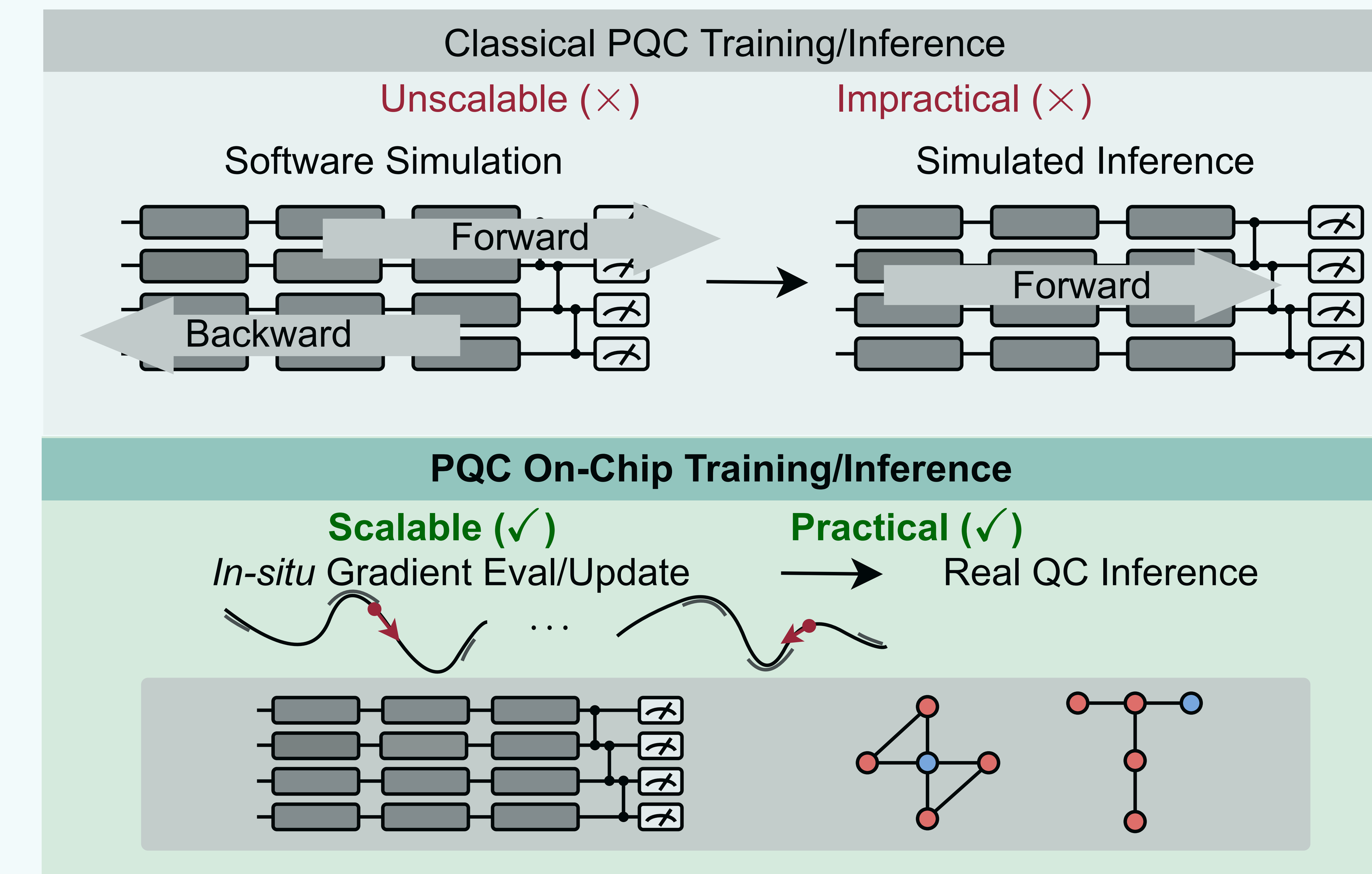
Quantization and Noise-Injection for Robustness

- Insert noise during the training of parameters to improve robustness
- Perform quantization of measurement outcomes for denoising
- Improves accuracy by up to 43%



Quantum On-Chip Training for Better Scalability

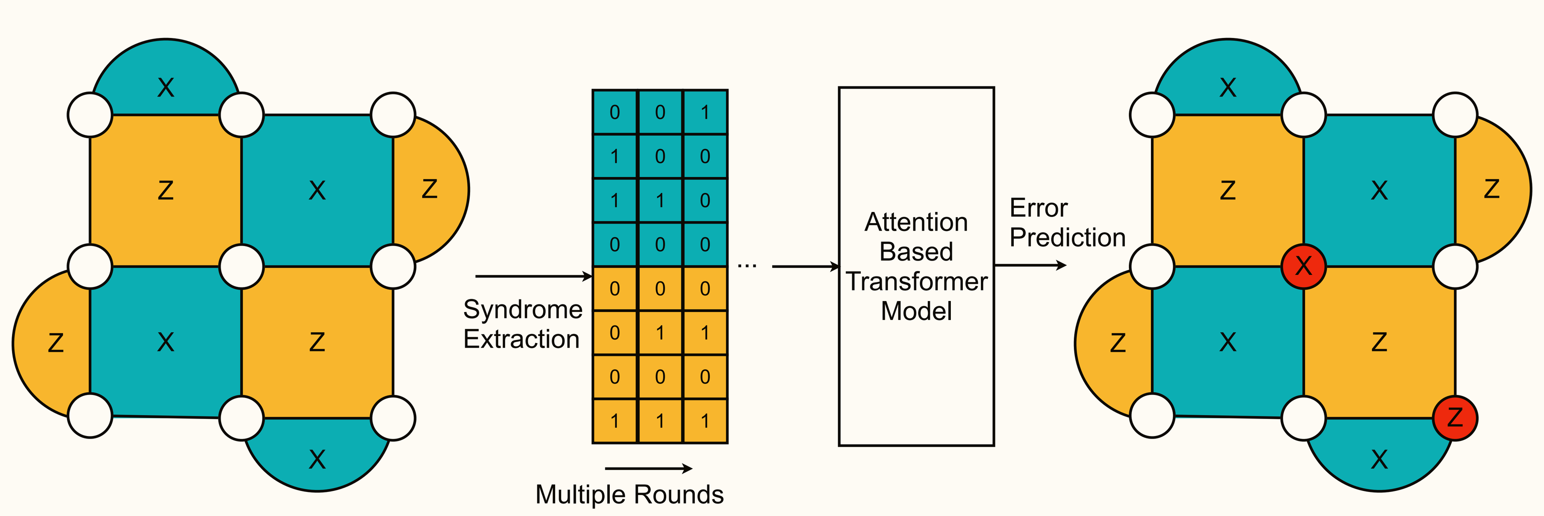
- First experimental demonstration of parameter shift rule on real quantum machines (IBMQ)
- Propose gradient pruning to further remove unreliable gradients and speedup training process
- Over 90% and 60% accuracy for 2-class and 4-class image classification



Long-Term: Quantum Error Correction

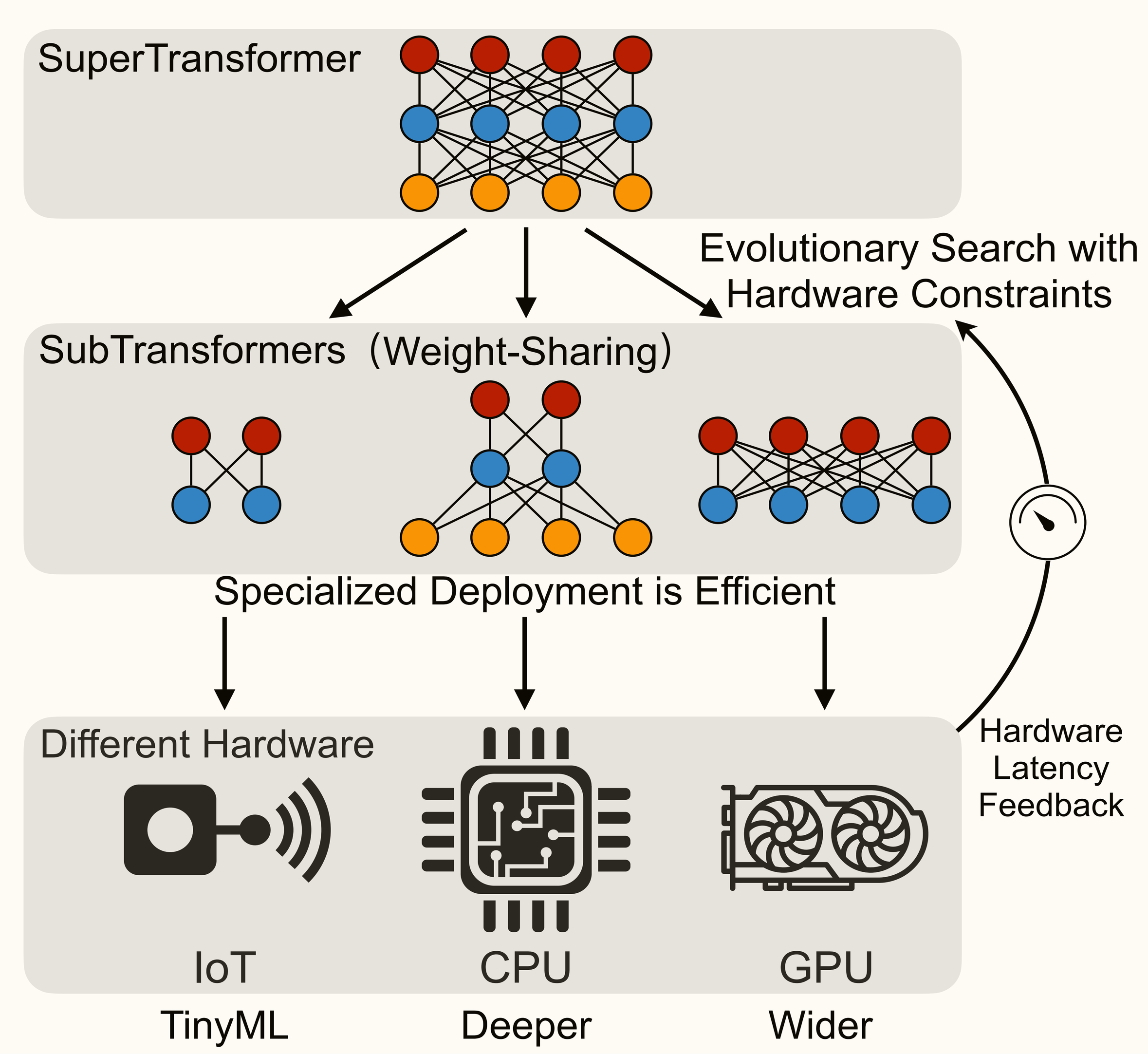
Transformer for Quantum Error Correction Code Decoder

- In the long term, we need quantum error correction code to further surpass quantum error
- QEC reduces logical error rates by introducing redundancy – encoding the quantum information to multiple qubits
- QEC requires a powerful decoder to process the syndromes obtained from syndrome qubit and predict the errors on the data qubit
- We propose a Transformer based decoder to process the Surface Code error syndromes to achieve high accuracy decoding than the traditional non-ML based decoders such as Union-Find



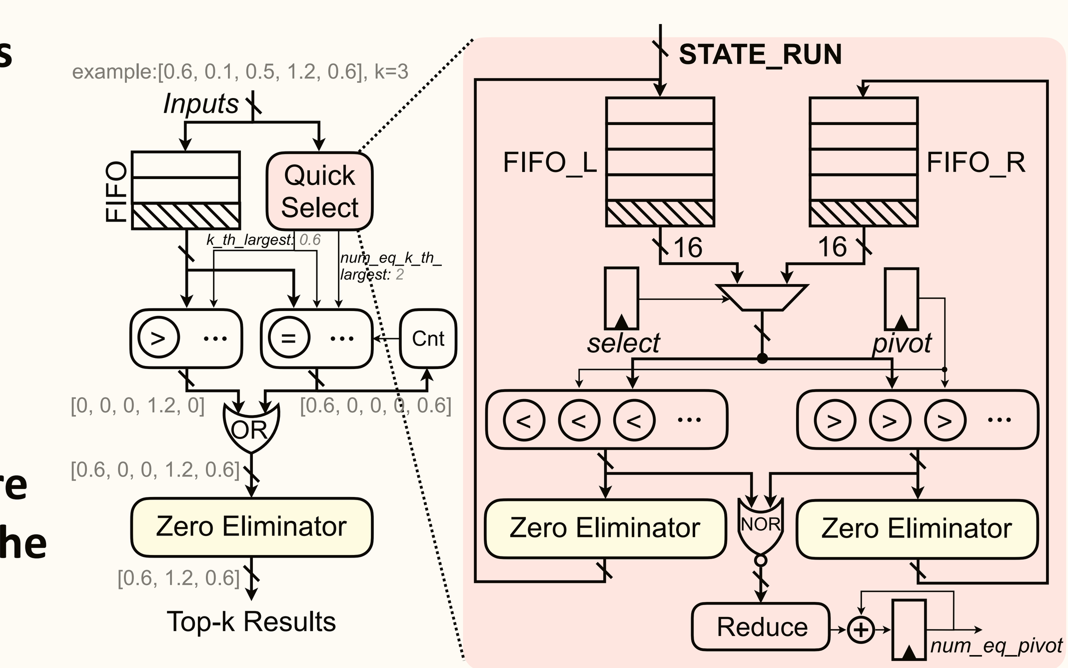
Hardware-Aware Transformer for QEC

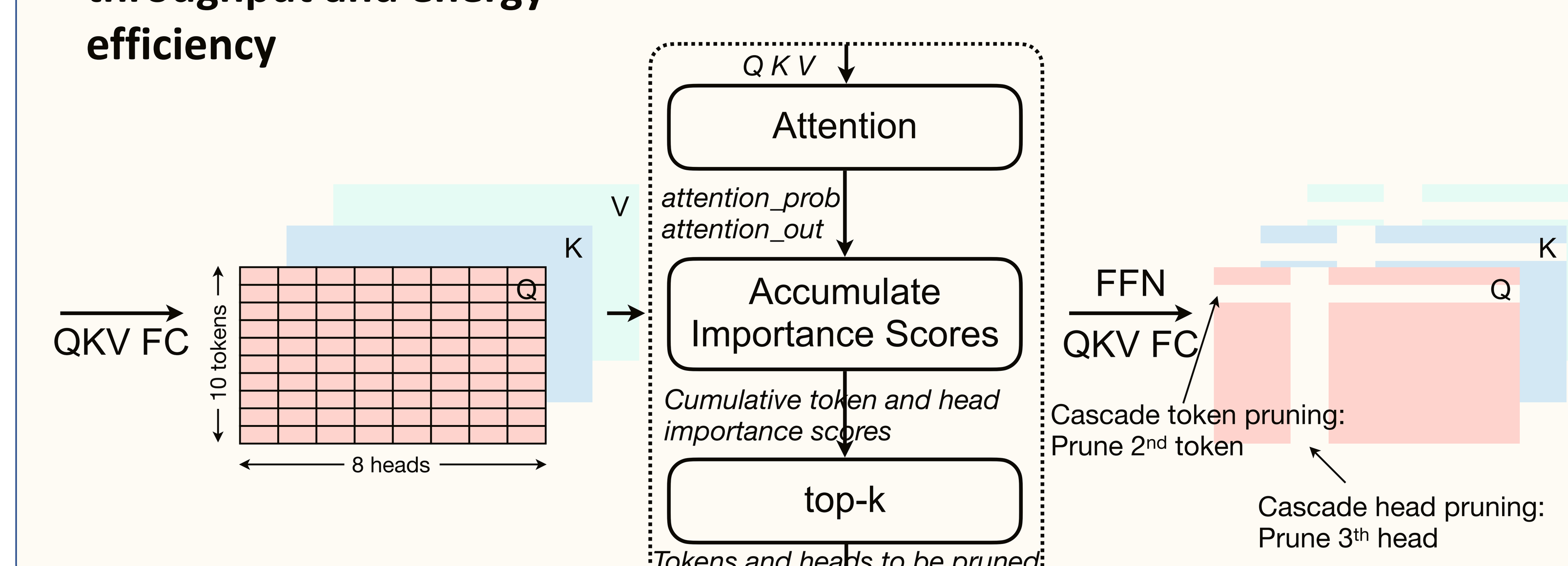
- The Transformer based decoder needs to be very efficiency to process the error syndromes in real-time
- We propose hardware aware transformer to find the most efficient model according to the quantum hardware decoherence time requirements



SpAtten: Syndrome Token Pruning Transformer Hardware Accelerator

- Not all syndrome tokens are created equal
- We propose token pruning to remove unimportant tokens to accelerate the Transformer processing
- We also design hardware accelerator to support the Transformer for higher throughput and energy efficiency





Reference

[1] H. Wang, Y. Ding, J. Gu, Y. Lin, D. Z. Pan et al., "Quantumnas: Noise-adaptive search for robust quantum circuits," in HPCA 2022.
[2] H. Wang, J. Gu, Y. Ding et al., "Quantumnat: Quantum noise-aware training with noise injection, quantization and normalization," DAC, 2022.
[3] H. Wang, Z. Li, J. Gu, Y. Ding, D. Z. Pan, and S. Han, "Qoc: Quantum on-chip training with parameter shift and gradient pruning," DAC, 2022.
[4] H. Wang, Z. Wu, Z. Liu, H. Cai, L. Zhu et al., "Hat: Hardware-aware transformers for efficient natural language processing," ACL, 2020.
[5] H. Wang, Z. Zhang, and S. Han, "Spatten: Efficient sparse attention architecture with cascade token and head pruning," in HPCA 2021, 2021.
[6] H. Wang*, Z. Zhang*, S. Han and W. J. Dally, "Sparch: Efficient architecture for sparse matrix multiplication," in HPCA. IEEE, 2020.
[7] H. Wang et al., "Transformer for quantum circuit reliability prediction (torchquantum case study for robust quantum circuits)," ICCAD, 2022.
[8] H. Wang*, Z. Liang*, J. Cheng, Y. Ding, H. Ren, X. Qian et al., "Variational quantum pulse learning," in QCE, 2022.