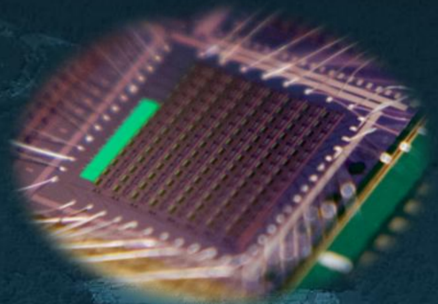


Cryogenic AI Software-Hardware Computing for Energy Apps

Shinjae Yoo
Key Collaborators: Huan-Hsin Tseng, Hsin-Yi Lin, David Park, Yihui Ray Ren (BNL), Samuel Yenchi Chen (Wells Fargo), Tzu-Chieh Wei (SBU), Junghoon Park and Jiook Cha (SNU), Jeff Shainline (Great Sky), Michael Churchill (PPPL)

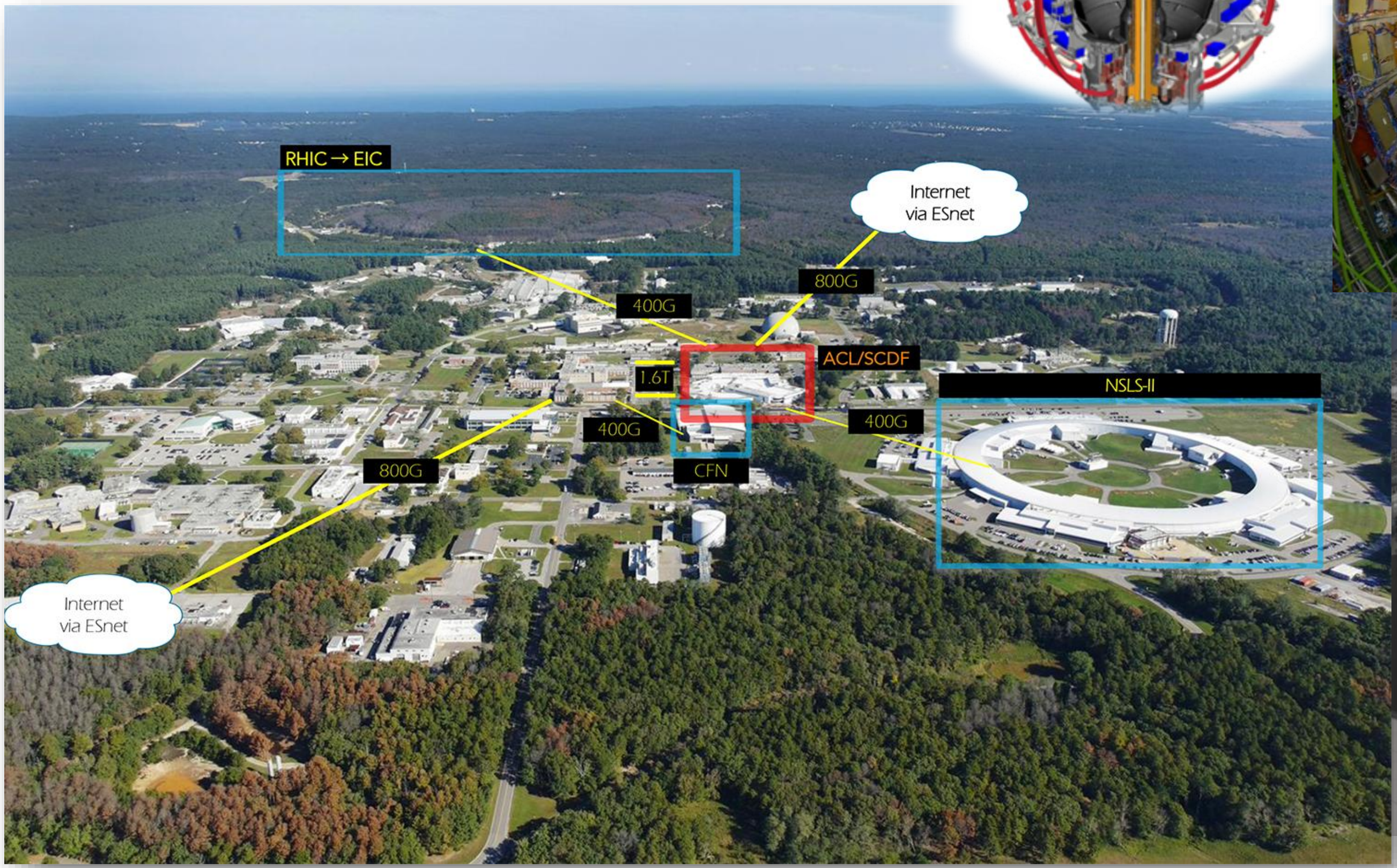
Artificial Intelligence Department
Computing and Data Science Directorate

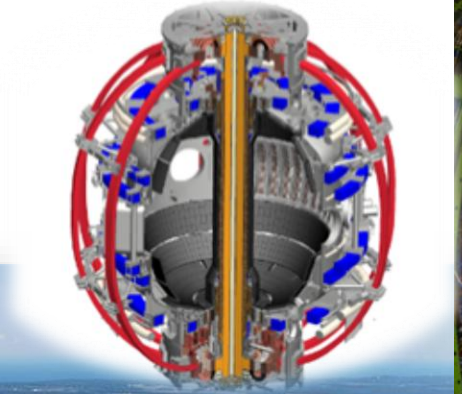
Brookhaven National Laboratory

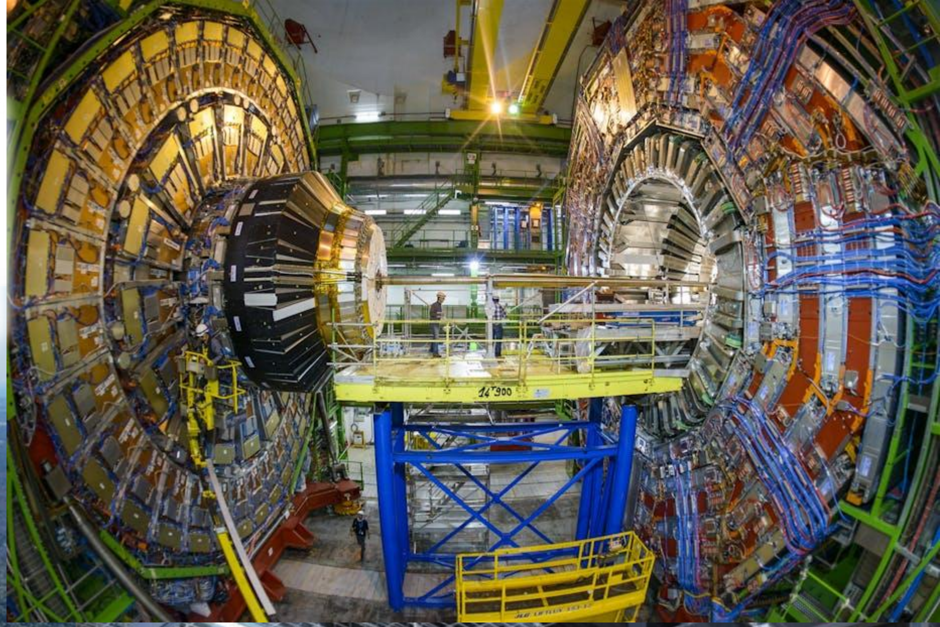
U.S. DEPARTMENT OF ENERGY

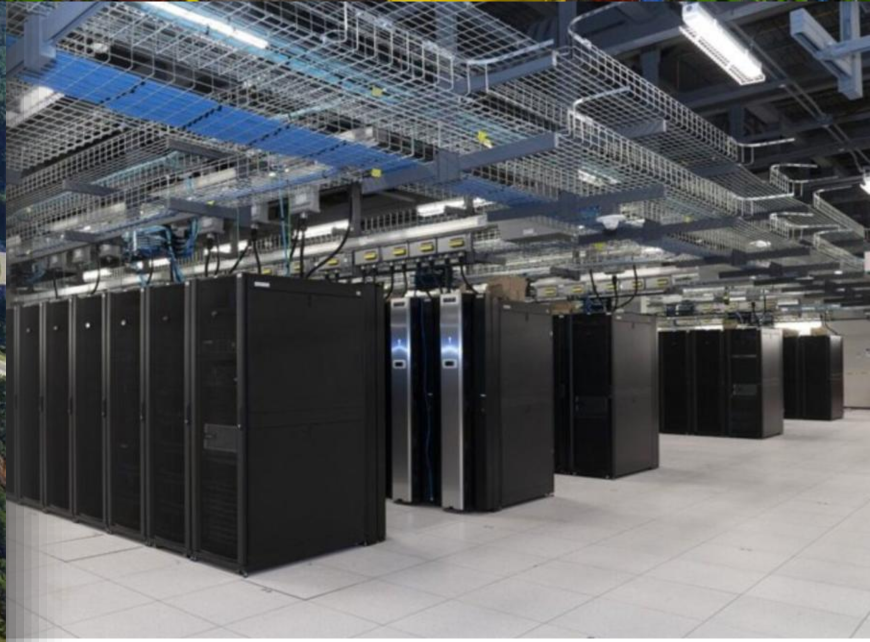
@BrookhavenLab

Motivation









Brookhaven National Laboratory

Two computing regimes, a shared co-design need

Quantum computing

Quantum states and parameterized gates

Classical AI selects circuits, measurements and training strategies.

Examples today
Hybrid variational circuits in simulation

Superconducting neuromorphic computing

Classical neural dynamics in physical circuits

AI trains programmable weights and adapts to measured dynamics.

Examples today
Measured devices and a recurrent network

Device physics shapes the learning problem

Device variation

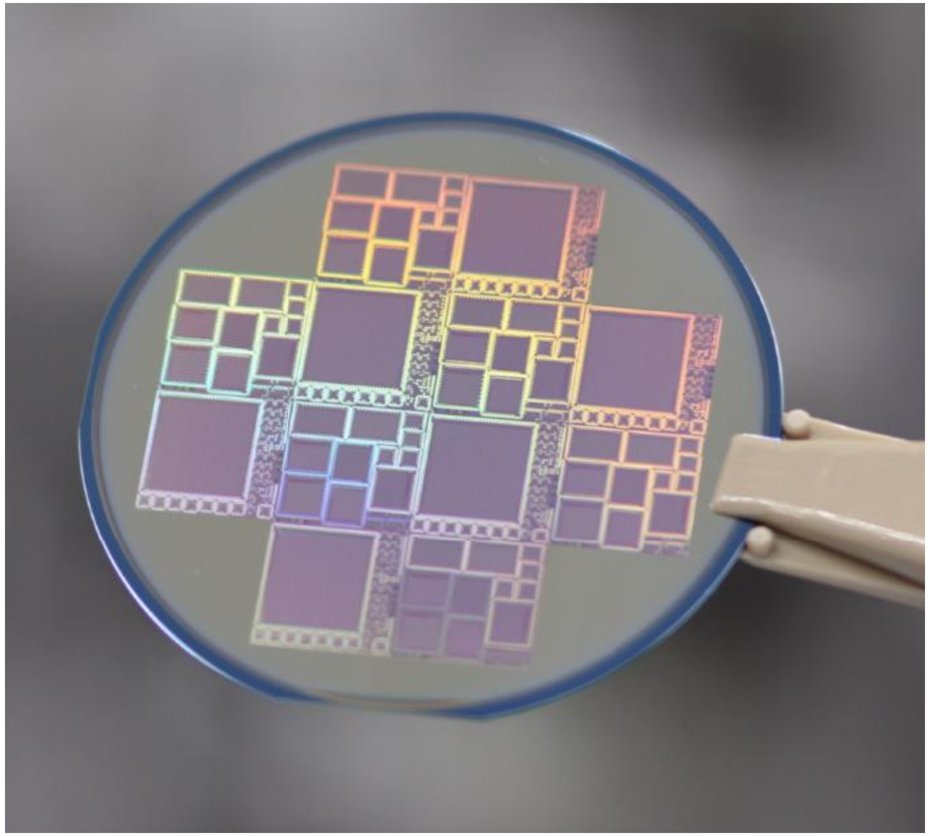
Measured behavior guides calibration and robust training.

Connectivity

Couplers and optical routing constrain model size and communication.

Readout and control

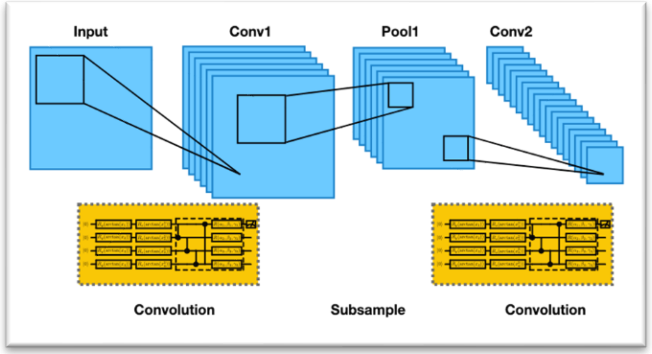
Sampling, conversion and cryogenic I/O shape useful throughput.



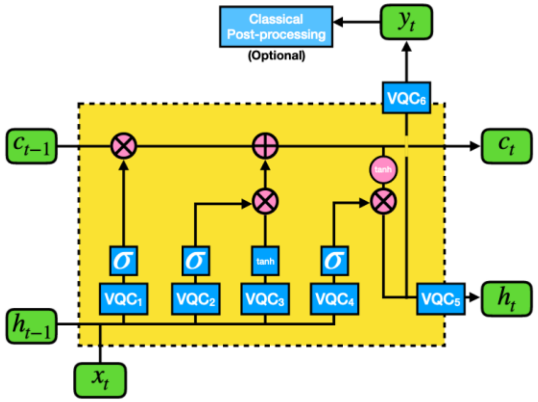
Great Sky wafer photograph

Quantum for AI Overview

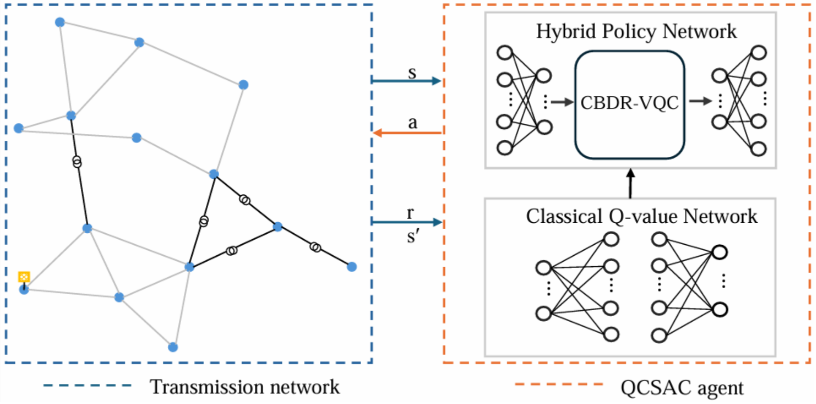
Quantum CNN



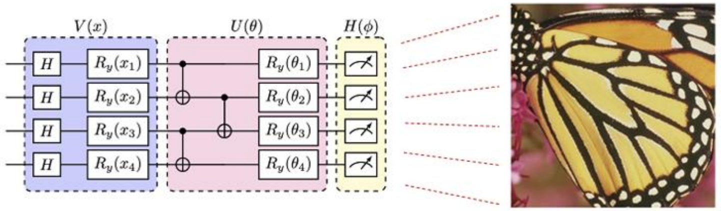
Quantum LSTM



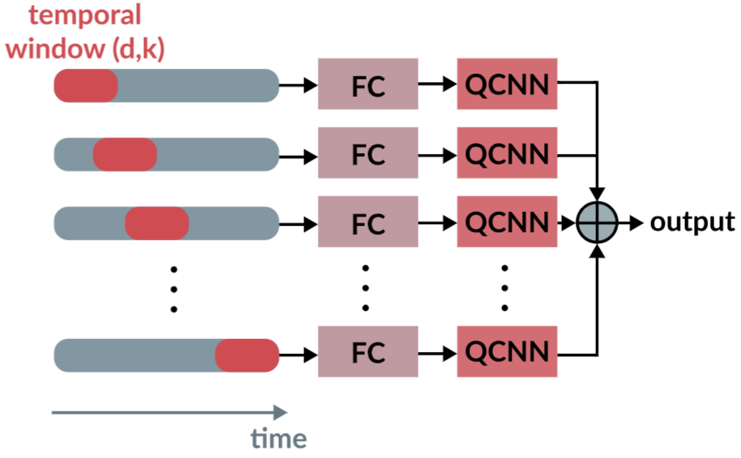
Quantum RL



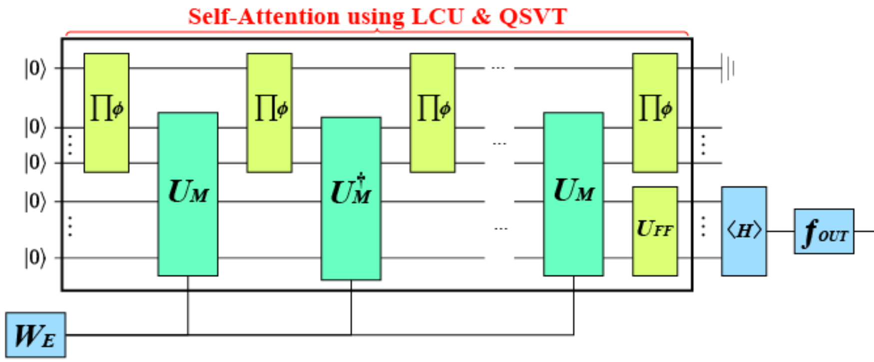
Quantum Super-resolution



Quantum Temporal Convolution



Quantum Transformers



Challenges in QML

- Noise, Decoherence, Quantum Error (Cerezo et al., 2022)

- Hardware Constraints (Cerezo et al., 2022; Chen et al., 2023)
  - Limited Number of Qubits
  - Limited Connectivity

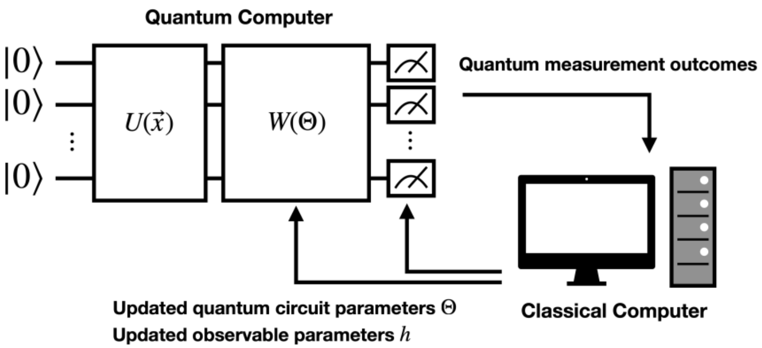
- Circuit Optimization
  - Barren Plateaus (McClean et al., 2018; Ortiz-Marrero et al., 2021; Holmes et al., 2022)
  - Overfitting (Caro et al., 2021; Kobayashi et al., 2022)

Limited Scalability, Difficulties in Trainability and Generalizability

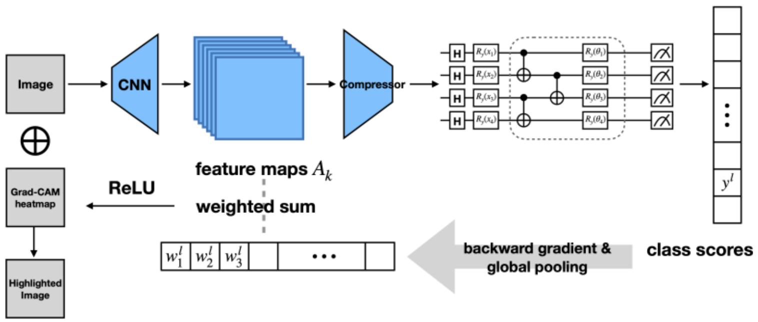
➔Limited Practical Usage

AI for Quantum Overview

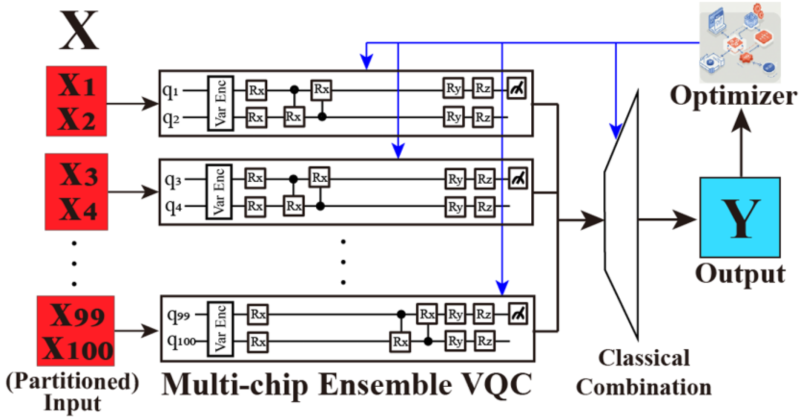
Learning to Measure



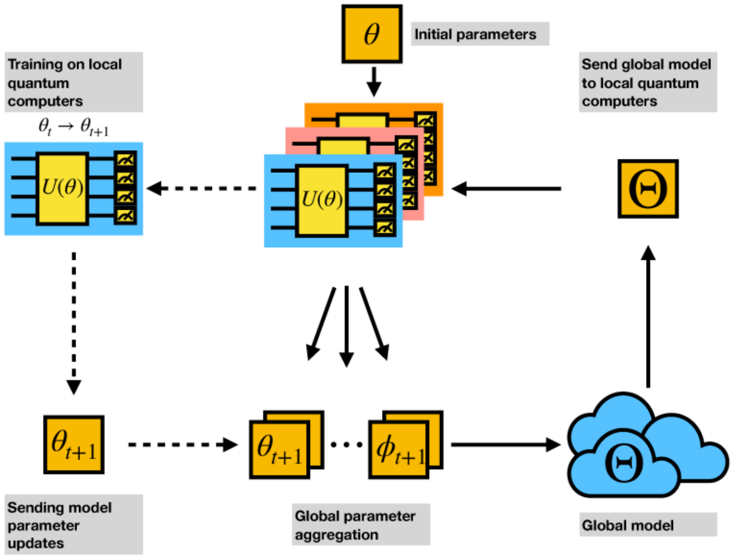
Quantum xAI (QGrad-Cam)



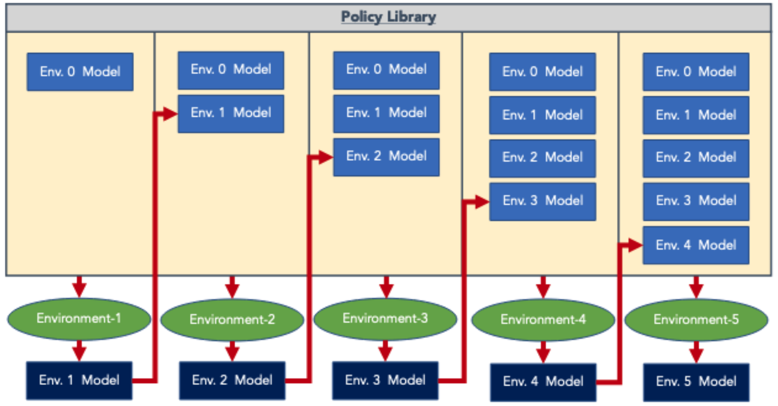
Quantum Multichip Learning



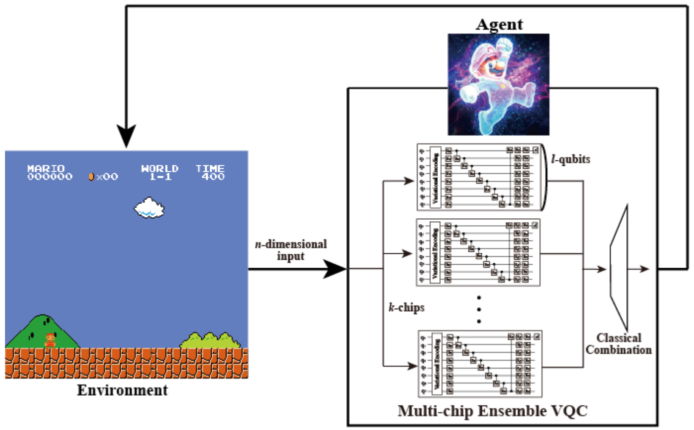
Quantum Federated Learning



Quantum Architecture Search with Continual Learning

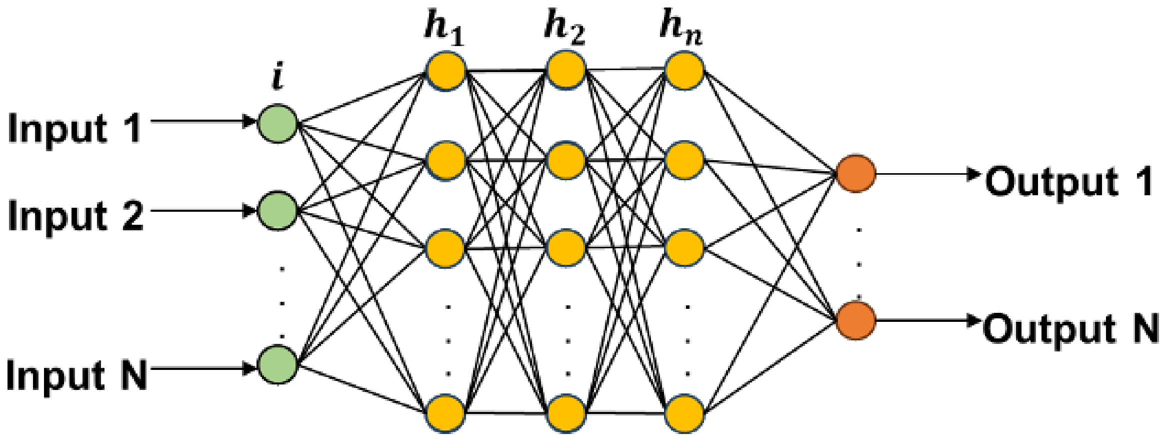


Quantum Multichip RL

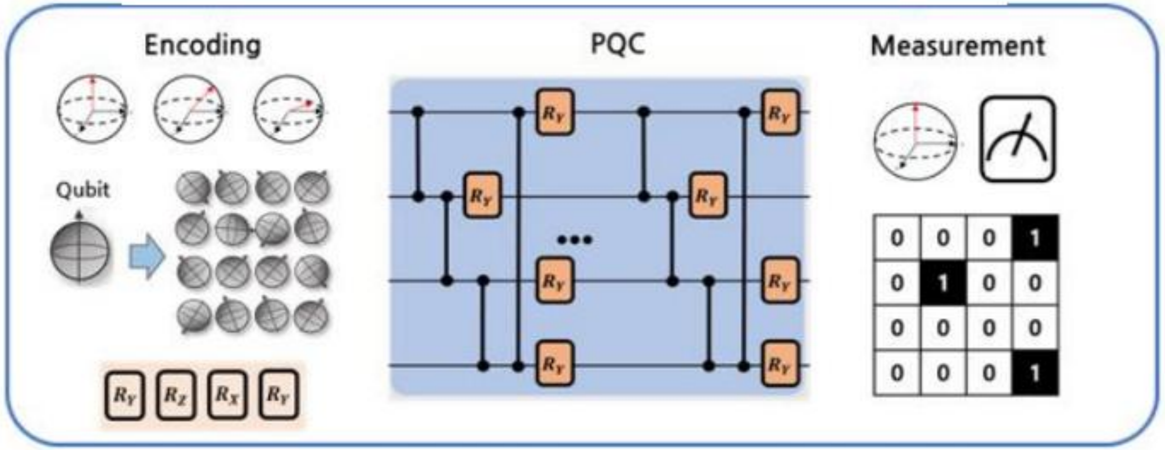


Quantum Computing

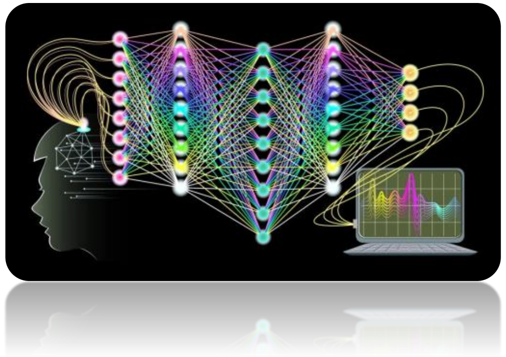
Classical Neural Network



Quantum Neural Network (QNN)



Variational Quantum Circuits (VQC)



In 1-qubit case: classical data

$$\mathcal{D} = \{(x_j, y_j)\} \qquad x_j \in \mathbb{R}^n, y_j \in \mathbb{R}$$

input label

$$\begin{tikzcd} \mathcal{H} \arrow[r, "U(\theta)"] & \mathcal{H} \arrow[d, "<\cdot|Q|\cdot>"] \\ \mathbb{R}^n \arrow[u, "V(x_j)"] \arrow[r, dashed, "f_{\mathrm{VQC}}"] & \mathbb{R}^m \end{tikzcd}$$

Step 1: Encode classical data $V : \mathbb{R}^n \to \mathcal{U}(\mathcal{H})$

$$|\psi_{x_j}\rangle := V(x_j)\,|\psi_0\rangle$$

(inject classical info → quantum state)

Step 2: Transform quantum states by a unitary $U(\theta) = e^{i\theta\sigma}$

$$|\psi_{x_j}\rangle \mapsto \boxed{U(\theta)}\,|\psi_{x_j}\rangle$$

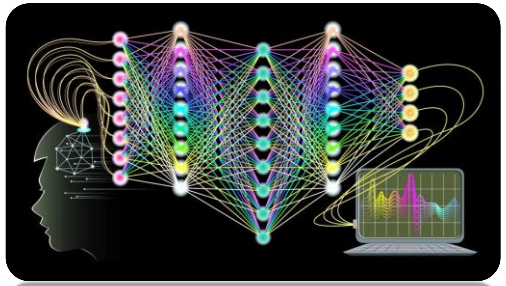
θ: parameters to be determined

(Classical network)

$$x \mapsto \boxed{Wx + b}$$

(W, b): network weights

Variational Quantum Circuits (VQC)



Step 3: Measurement: choose a Hermitian matrix Q

$$\langle Q \rangle_v := \langle v | Q v \rangle \in \mathbb{R}$$

(extract quantum info $|v\rangle$ → classical number)

In total:

$$\langle Q \rangle := \langle \psi_0 | \, V^\dagger(x_j) \boxed{U^\dagger(\theta)} Q \boxed{U(\theta)} V(x_j) \, | \psi_0 \rangle \in \mathbb{R} \quad \Longleftrightarrow \quad x \mapsto \boxed{Wx + b}$$

(VQC)

(Classical network)

Step 4: Choose loss function

$$\mathcal{L}(\theta; \mathcal{D}) = \sum_{j=1}^{N} \| \langle Q \rangle (x_j; \theta) - y_j \|^2$$

to determine best θ

$$\begin{tikzcd} \mathcal{H} \arrow[r, "U(\theta)"] & \mathcal{H} \arrow[d, "<\cdot|Q|\cdot>"] \\ \mathbb{R}^n \arrow[u, "V(x_j)"] \arrow[r, dashed, "f_{\mathrm{VQC}}"] & \mathbb{R}^m \end{tikzcd}$$

Brookhaven National Laboratory

Quantum Learning to Measure Quantum Neural Networks

- **Challenge**: Existing QML (VQC) models rely on fixed measurement observables (e.g., Fig. 1 with Pauli matrices), limiting flexibility and task-specific optimization.
- **Approach**: Learnable, parameterized observables Q for VQCs (Fig. 2).

$$\langle Q \rangle := \langle \psi_0 | V^\dagger(x_j) \boxed{U^\dagger(\theta)} \boxed{Q} \boxed{U(\theta)} V(x_j) |\psi_0\rangle \in \mathbb{R}$$

- **Results**: Demonstrated QNN (make moon, 4 qubit, 2 layers)

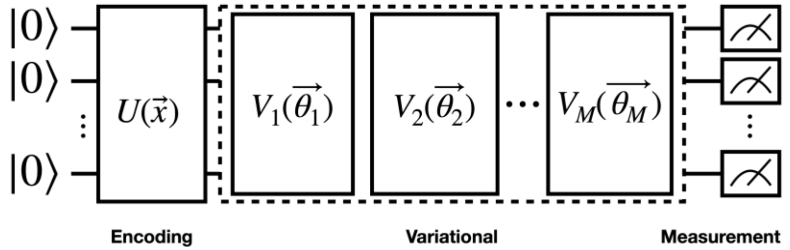


Fig. 1: Conventional VQC with fix (predefined) observables

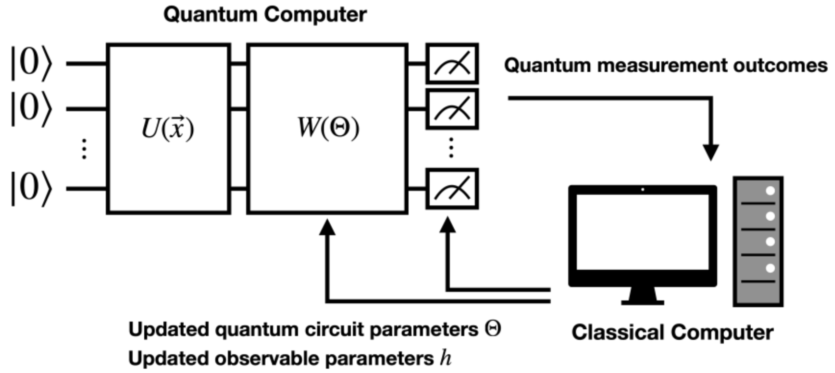
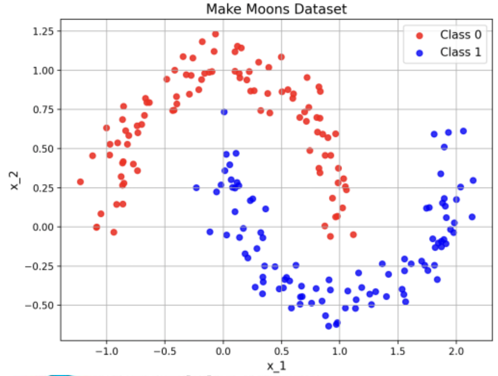
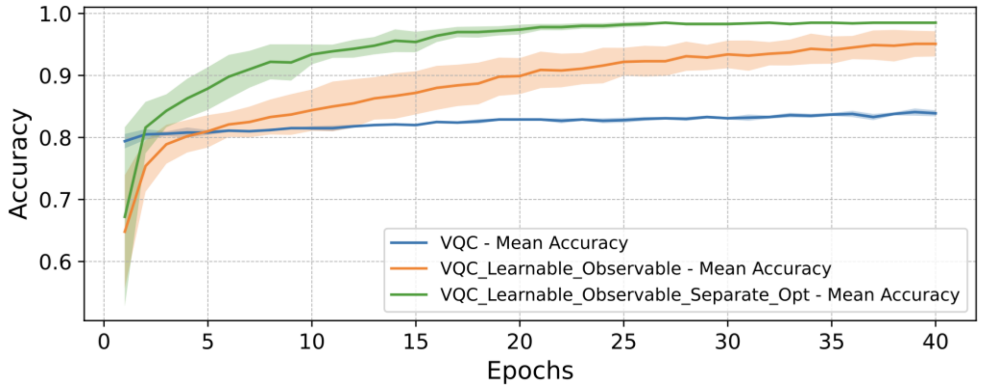


Fig. 2: VQC with learnable observables (Hermitians)

$$\begin{tikzcd}
\mathcal{H} \arrow[r, "\boxed{U(\theta)}"] & \mathcal{H} \arrow[d, "\boxed{<\cdot|Q|\cdot>}"] \\
\mathbb{R}^n \arrow[u, "V(x_j)"] \arrow[r, dashed, "f_{\mathrm{VQC}}"] & \mathbb{R}^m
\end{tikzcd}$$





Brookhaven National Laboratory

Samuel Yen-Chi Chen, Huan-Hsin Tseng, Hsin-Yi Lin, Shinjae Yoo, "Learning to Measure Quantum Neural Networks", arXiv:2501.05663 (2025)

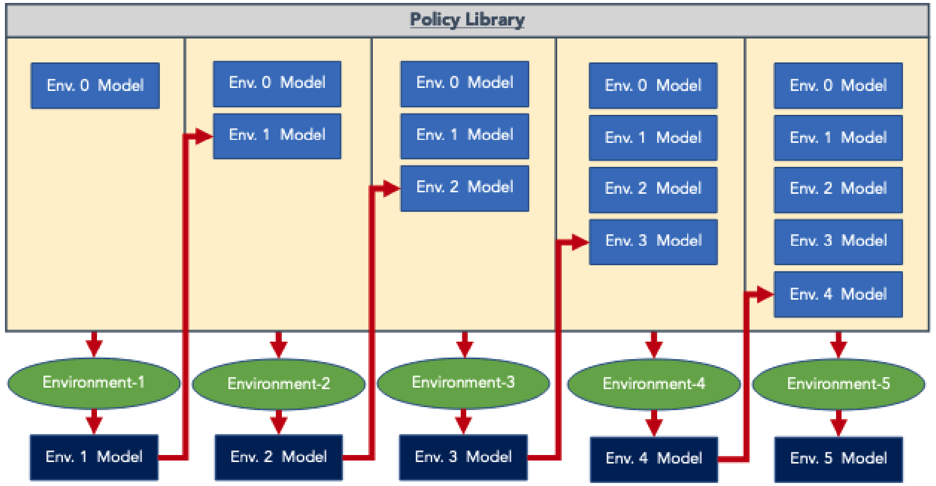
Quantum Architecture Search via Continual Reinforcement Learning

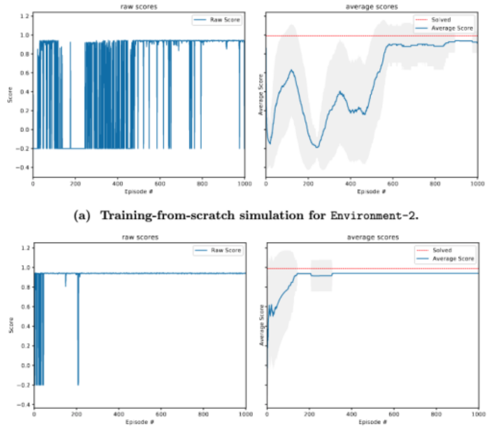
Motivation: Existing quantum architecture search schemes assume some prior knowledge of quantum circuits, are sampling from a set of potential circuits, and cannot automatically reuse previously learned policy.

Approach: Offer a continual reinforcement learning (DRL) agent to generate desired quantum circuits without encoded physics knowledge that can reduce training episodes with previous learning knowledge.

Results: The DRL agent can generate quantum circuits under the effects of noise and can learn quickly when device noise patterns change.

Impact: Results suggest the possibilities of building ML models to rewrite and update quantum AI.





(a) Training-from-scratch simulation for Environment-2.

(b) Policy reuse simulation result for Environment-2.
Starting Policy Library: from Scratch - Environment-0, Policy Reuse - Environment-1.

Brookhaven National Laboratory

Ye, E., & Chen, S. Y. C. (2021). Quantum Architecture Search via Continual Reinforcement Learning. arXiv preprint arXiv:2112.05779.

Quantum xAI

Motivation:

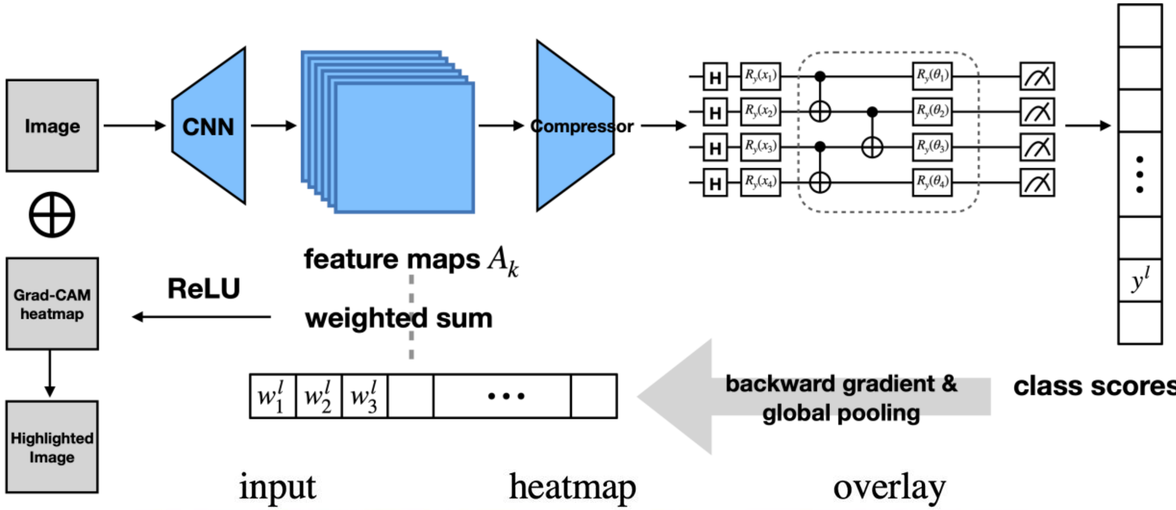
- AI is widely applied on various co-design and operation activities with the specified objective functions
- Next generation detector / accelerator / reactor design and operation
- Disruptive energy efficient computing is seriously required

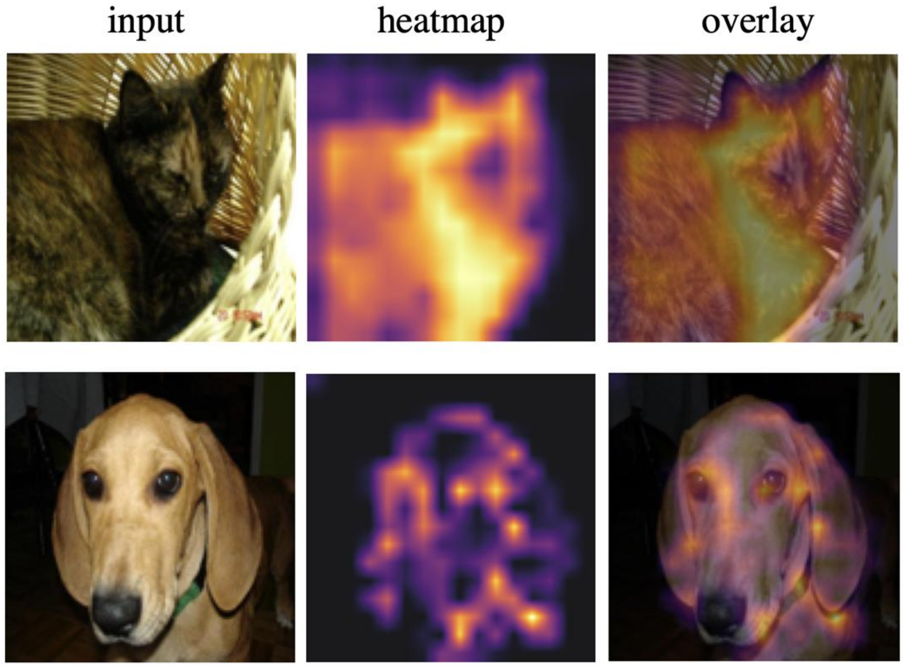
Challenges:

- Design and operation **competing objective** (energy vs accuracies vs noise)
- Algorithmic behavior understanding given the device condition

Related Work:

- QGrad-Cam for QML (IBM 100+ qubit test awards) [1]
- Causal Analysis for understanding entanglement, expressibility, etc. [2]





1. Lin, Hsin-Yi, Huan-Hsin Tseng, Samuel Yen-Chi Chen, and Shinjae Yoo. "Quantum Gradient Class Activation Map for Model Interpretability." In *2024 IEEE Workshop on Signal Processing Systems (SiPS)*, pp. 165-170. IEEE, 2024.
2. Park, Junghoon, Samuel Yen-Chi Chen, Shinjae Yoo, Huan-Hsin Tseng, and Wells Fargo. "Over the Quantum Rainbow: Explaining Hybrid Quantum Reinforcement Learning.“, QCE 2024

Brookhaven National Laboratory

HPCwire
Since 1987 - Covering the Fastest Computers in the World and the People Who Run Them

- Home
- Topics
- Sectors
- Exascale
- Specials
- Resource Library
- **Podcast**
- Events
- Job Bank
- About

Rigetti Demonstrates Industry's Largest Multi-Chip Quantum Computer and Halves Two-Qubit Gate Error Rate

July 16, 2025

BERKELEY, Calif., July 16, 2025 — Rigetti Computing, Inc., a pioneer in full-stack quantum-classical computing, today announced that it has achieved its mid-year performance milestone of 99.5% median two-qubit gate* fidelity on its modular 36-qubit system, a 2x reduction in median two-qubit gate error ra[te] from Rigetti's previous best results on its 84-qubit single chip Ankaa-3 syste[m]

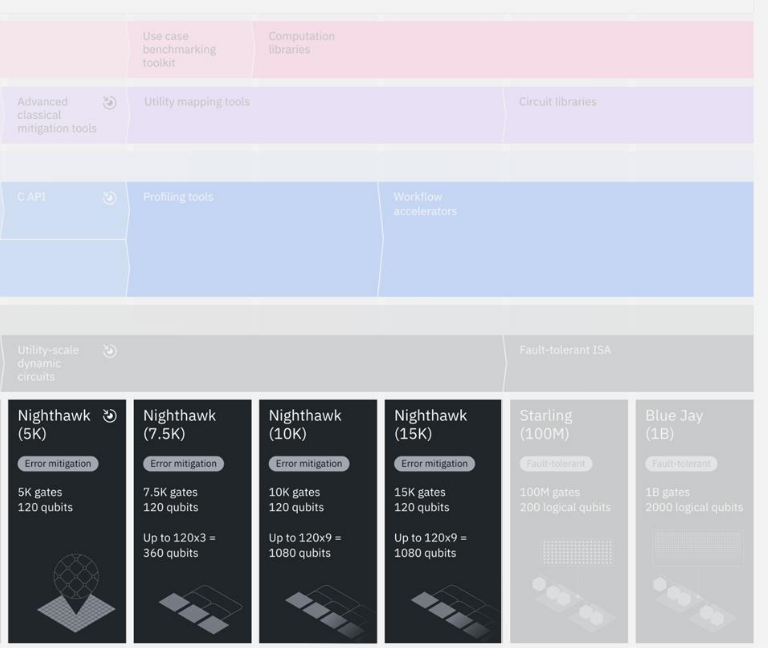
Composed of four 9-qubit chips ("chiplets") tiled together, the 36-qubit system is based on Rigetti's proprietary modular chip technology and unlocks the Company's path to building a 100+ qubit chiplet-based system.

Rigetti plans to launch its 36-qubit system on August 15, and remains on track to release its 100+ qubit chiplet-based

QUANTUM COMPUTING

Multi-Chip Frameworks for Scalable and Robust Quantum Machine Learning

May 14, 2025
BY QUANTUM NEWS



1. More experiments with logical qubits
2. More specialized hardware/software (as opposed to universal quantum computing)
3. More networking noisy intermediate-scale quantum (NISQ) devices together
4. More layers of software abstraction
5. More workforce development tools
6. Improved and novel physical qubits

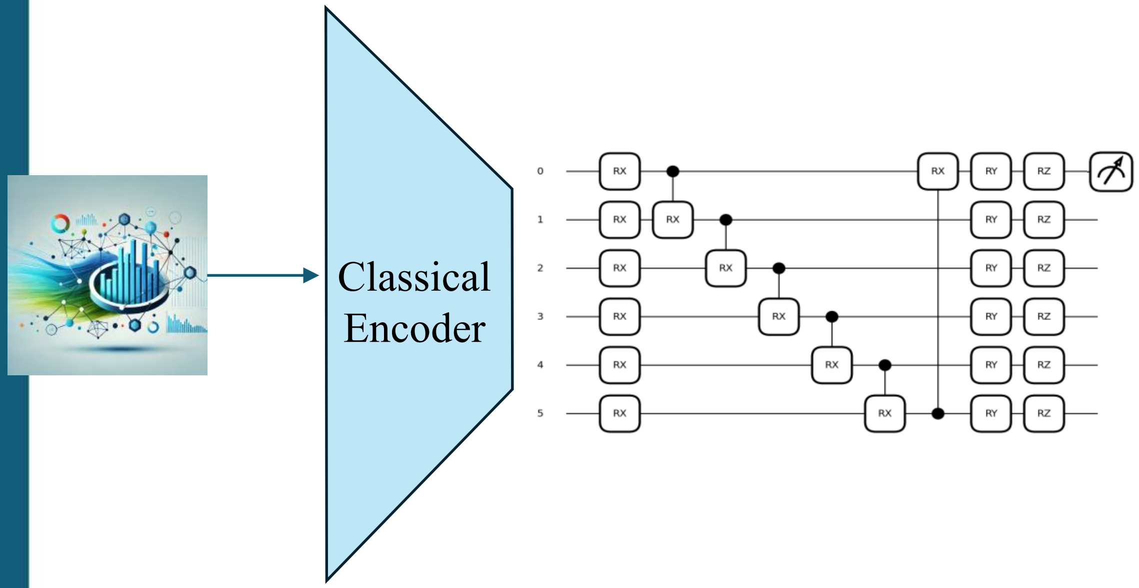
Quantum computing's six most important trends for 2025

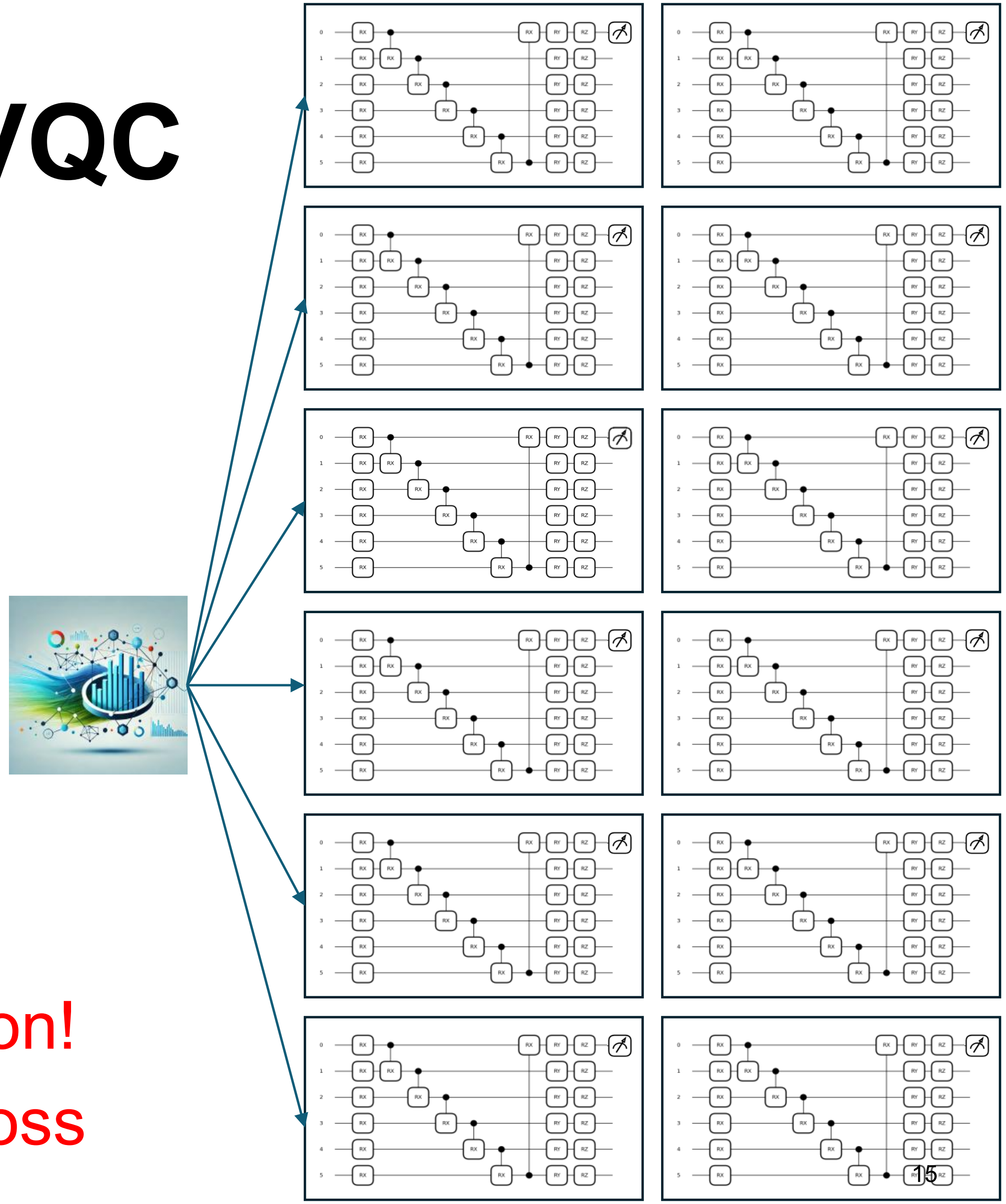
February 04, 2025

Brookhaven National Laboratory

https://quantumzeitgeist.com/multi-chip-frameworks-for-scalable-and-robust-quantum-machine-learning/

Multi-Chip Ensemble VQC





No Dimension Reduction!

Minimize Information Loss

Brookhaven National Laboratory

Advantages of Multi-Chip Ensemble

Improved Scalability

→Multi-Chip Ensemble can encode & process high dimensional data with minimal dimension reduction

Improved Trainability & Generalizability

→Multi-Chip Ensembles have reduced entanglement & expressibility

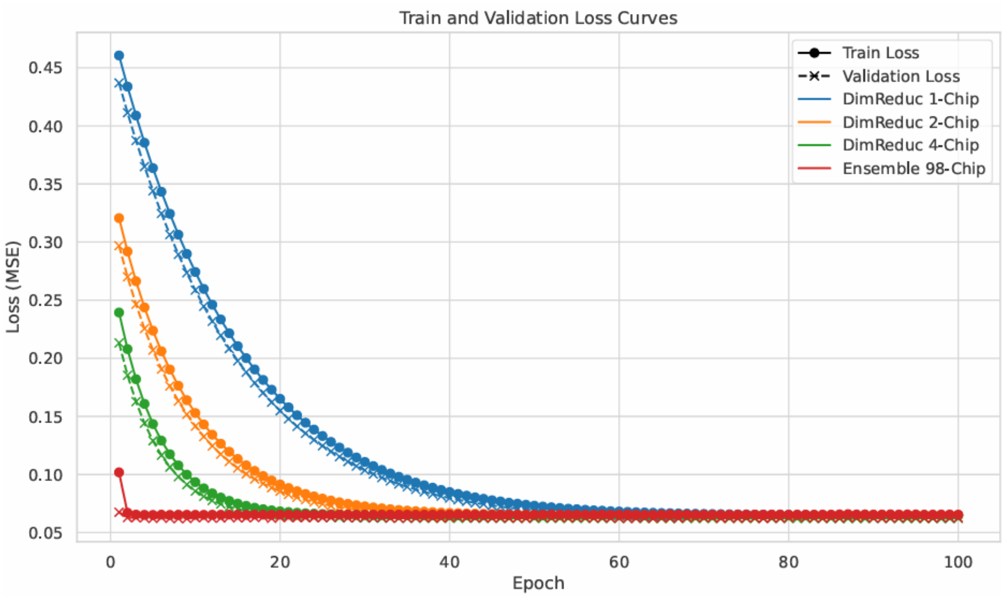
→This reduces the risk of barren plateaus & overfitting caused by excessive levels of entanglement & expressibility

Improved Noise Resistance

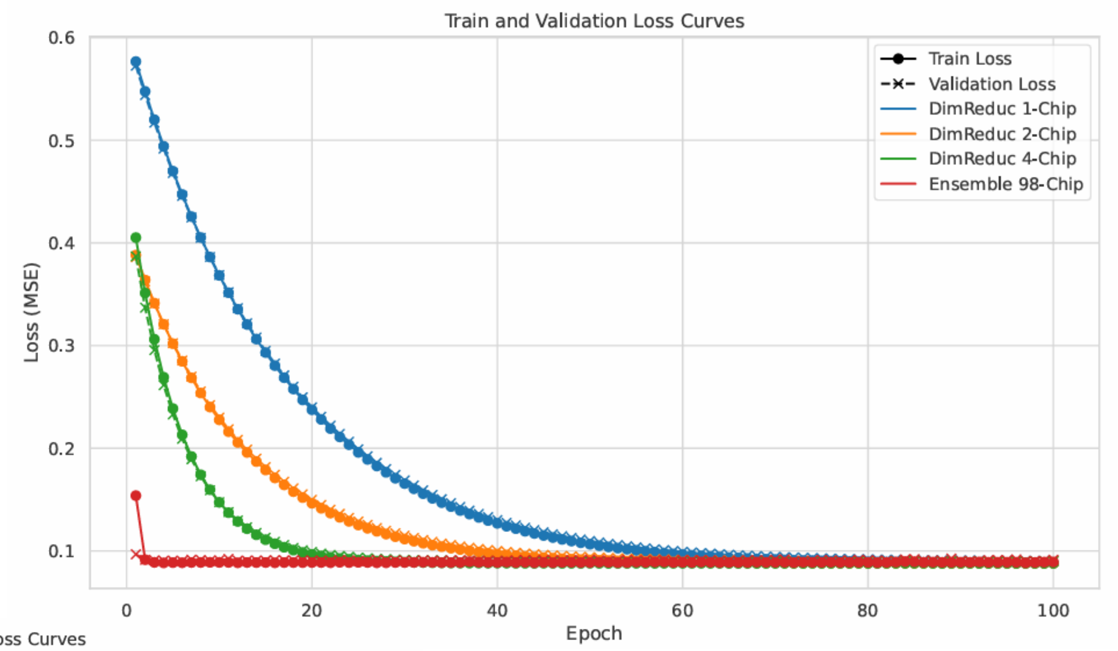
→By combining measurements of multiple VQCs, multi-chip ensembles naturally mitigates quantum errors

Brookhaven National Laboratory

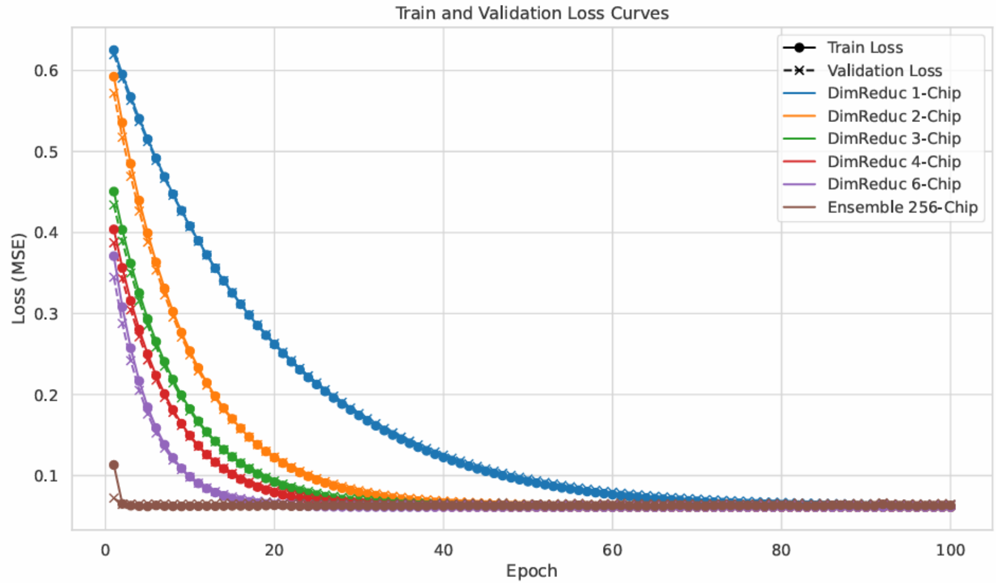
Model Performance



MNIST



Fashion MNIST



CIFAR-10

Brookhaven National Laboratory

Superconducting optoelectronic networks (SOENs)

Optical input

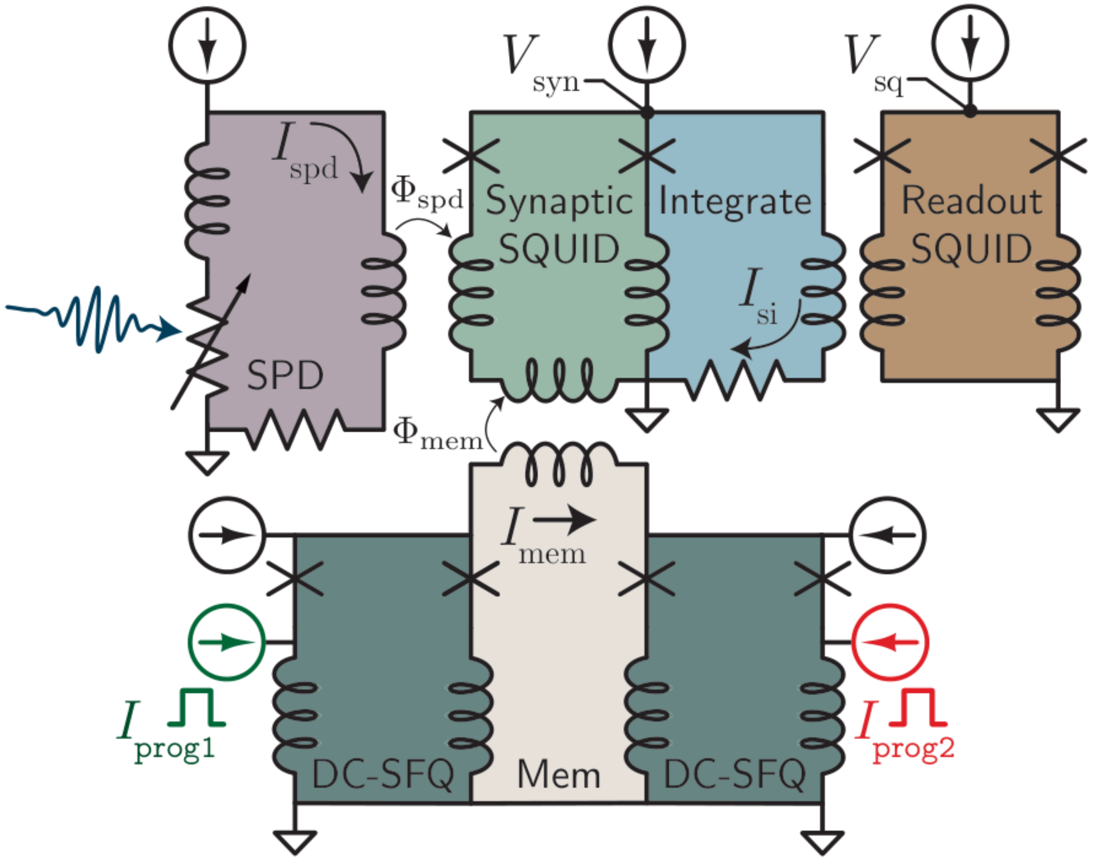
Single-photon detectors receive signals.

Superconducting dynamics

Josephson junctions and loops provide nonlinear integration.

Programmable response

Persistent-current memory sets synaptic weight.

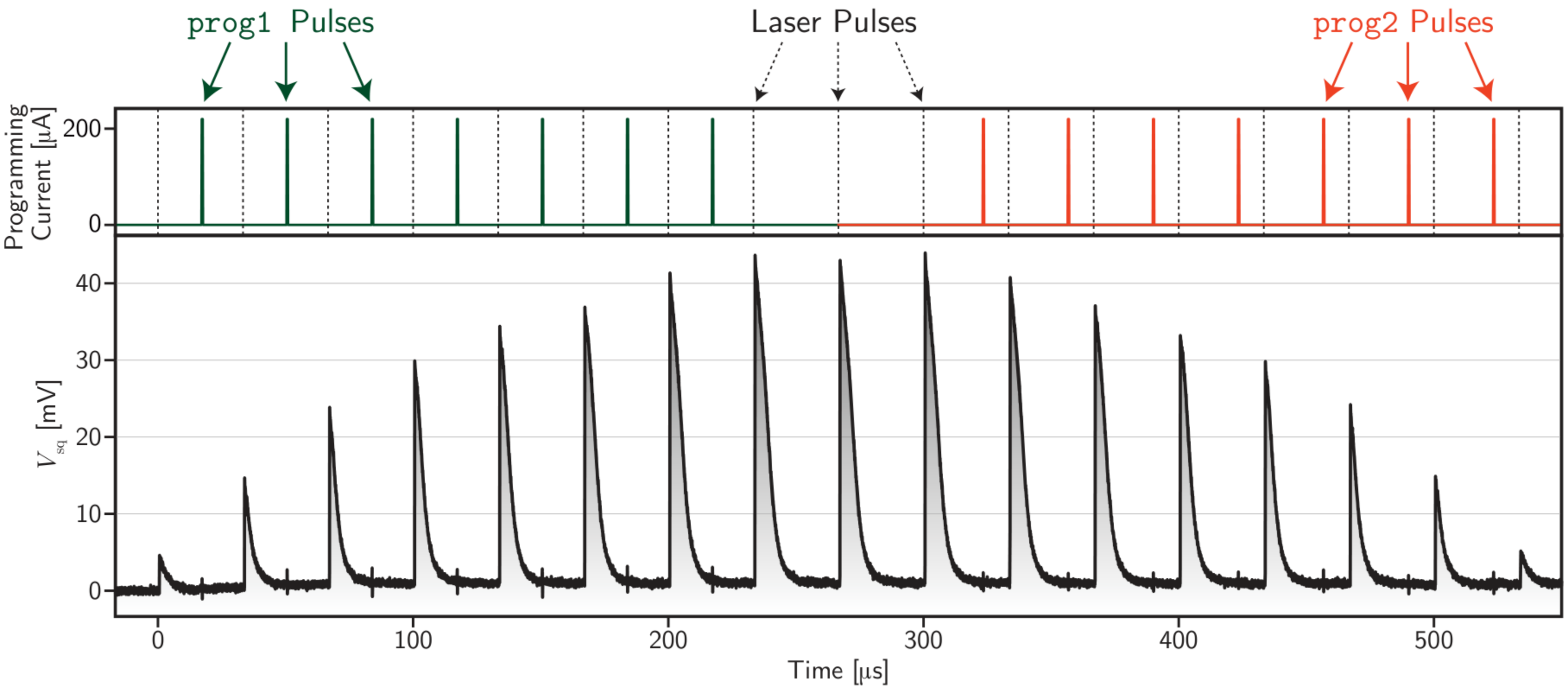


Brookhaven National Laboratory

Device architecture: Primavera et al., ICONS 2024. Integrated optical scaling remains a development step.

Memory pulses tune the response to optical inputs

Measured programmable synapse at 2.3 K



Optical probe pulses and local superconducting memory. Primavera et al. (2024).

Superconducting Optoelectronic Networks (SOENs)

Advanced Computing Technology with Fusion Energy Demonstration

DIII-D NATIONAL FUSION FACILITY SAN DIEGO · PPPL · GREAT SKY · Brookhaven National Laboratory

Scientific Achievement

Demonstration of neuroscience-inspired AI accelerator for high-throughput inference applied to real-time plasma disruption prediction.

- **Hardware Demonstration:** World's largest analog **superconducting neural network with 1,008 parameters**. The chip has been demonstrated at Great Sky in Boulder, CO, and a full system will be deployed to Brookhaven Nat. Lab. in Dec. 2026 for further qualification.
- **Efficiency Milestone:** Demonstrate AI-based disruption prediction performance with a custom hardware platform with potential for scaling beyond 1T parameters

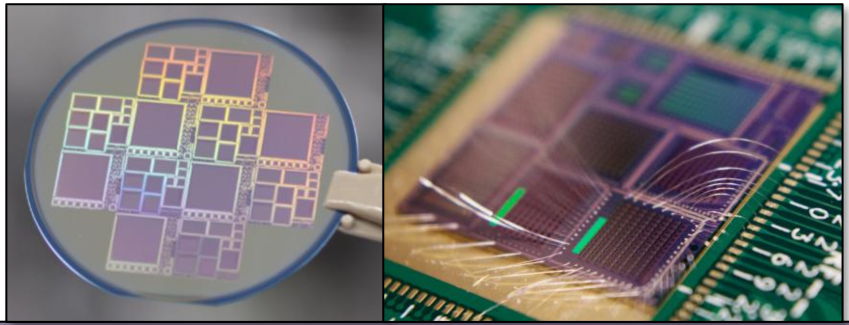
Significance and Impact

This technology captures the processing, connectivity, and efficiency of biological brains. This work marks the first application of SOENs to a critical challenge in fusion energy.

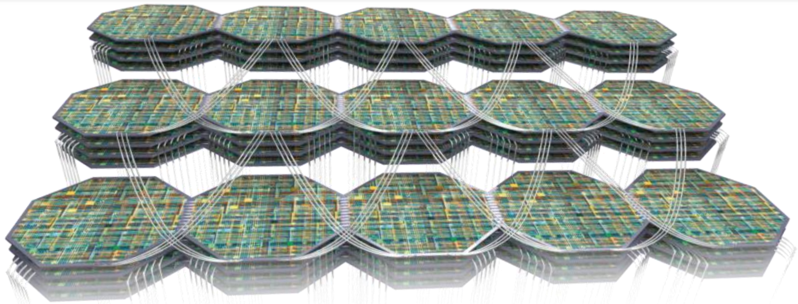
- **Overcoming the AI Power Gap:** By utilizing superconducting circuits and single-photon optical links, the system targets a transition away from the gigawatt-scale data centers required by traditional GPUs toward ultra-low-power alternatives.
- ***Public-Private-Partnership (BNL-Great Sky):*** This collaboration demonstrates a clear and promising roadmap for foundational AI going beyond exascale computing technologies.

Technical Approach

- The team pursued hardware-algorithm co-design to deliver a custom neural architecture to analyze spatiotemporal dynamics of fusion electron cyclotron emission imaging (ECEi) data.
- The platform integrates Josephson-junction superconducting circuits for high-speed, low-energy computation, and single-photon optical signaling and photonic waveguides for high-fan-out communication.



Great Sky's 1k-parameter superconducting neural network: operational, demo went live June 2026



Great Sky's vision: scale to 40-billion-parameter models in 5 years and beyond a trillion parameters in 10 years.

PI(s)/Facility Lead(s): Shinjae Yoo
Collaborating Institutions: Great Sky
ASCR Program: Emerging Computing Technologies
ASCR PM: Pavel Lougovski
Publication(s) for this work: J. Shainline, M. Churchill, S. Yoo, et al. "Superconducting Optoelectronic Computing Networks for Fusion Disruption Prediction", manuscript in preparation.

1KP Programmable Network Demo

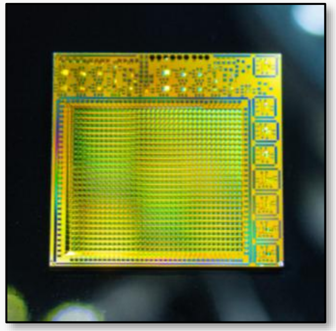
DIII-D NATIONAL FUSION FACILITY SAN DIEGO · PPPL · GREAT SKY · Brookhaven National Laboratory

The Chip

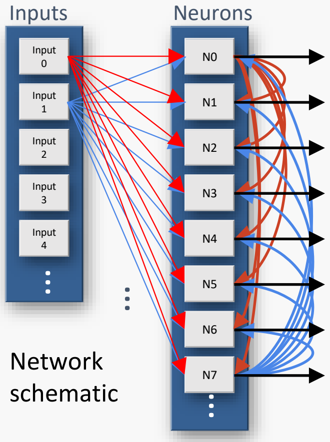
1000-parameter recurrent neural network

Programmable input-to-hidden and hidden-to-hidden weights

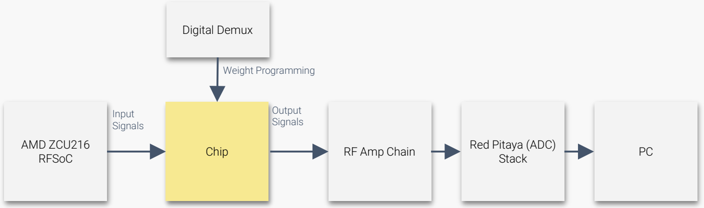
Eight input channels driving 28 neurons with all-to-all connectivity



1,000 parameter analog superconducting neural network



Network schematic

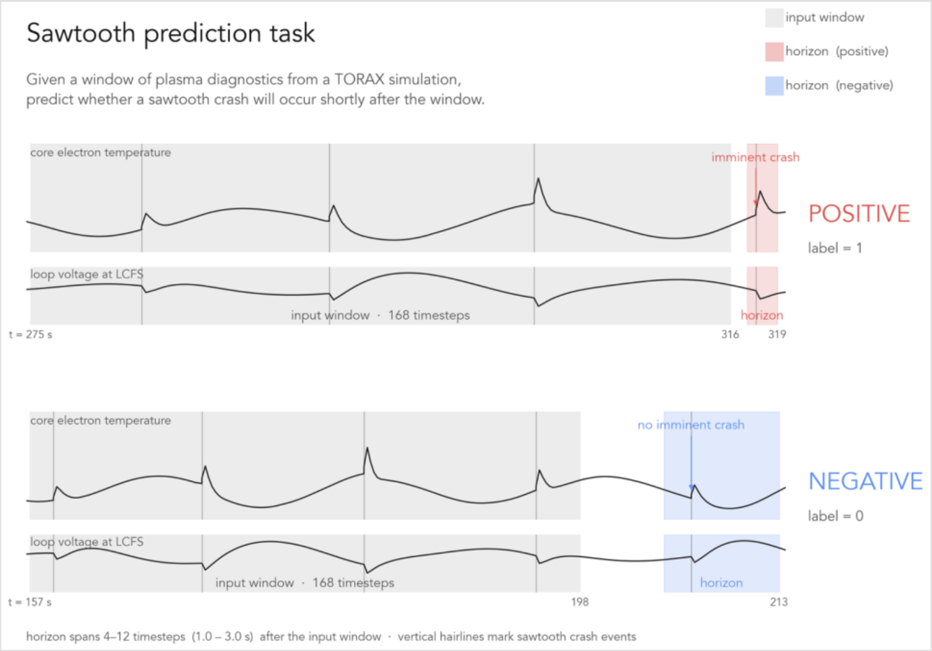


Three Datasets and Tasks

Fusion disruption prediction: Eliminating plasma disruptions is a critical issue to realize fusion power plants. Great Sky demonstrated superconducting neural networks for low-latency disruption prediction using historical DIII-D ECEI data.

Anomaly detection: Tokamak operation can involve transient phenomena such as edge-localized modes, tearing modes, and other anomalous plasma dynamics. Great Sky trained its neural network chip to identify anomalous events historical DIII-D ECEI data.

Sawtooth prediction: Sawtooth crashes are recurrent core instabilities that can affect confinement and control. Great Sky generated synthetic data using the TORAX plasma simulator and used the labeled dataset to train a model to predict imminent sawtooth crash events.



Results and Outlook

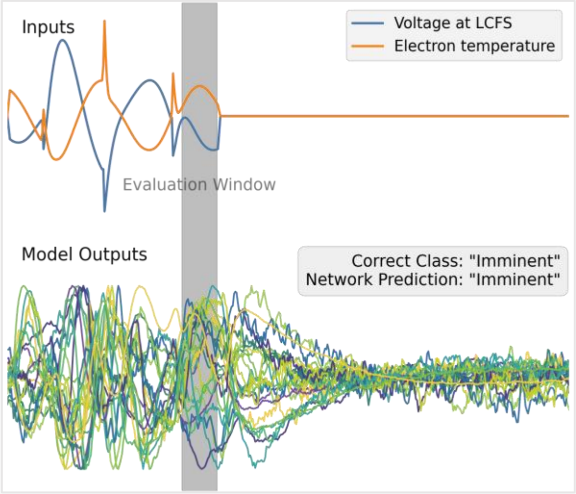
Fusion disruption prediction: A 1000-parameter model was able to achieve an F1 score of 0.79 with a lead time of 300 ms ahead of the disruption event, showing that our network efficiently integrates temporal information that boosts its performance over a multilayer-perceptron baseline.

Anomaly detection: Models were trained to distinguish low- and high-anomaly behavior without manual labeling. The physical network achieved an F1 score of 0.80, showing promise for robust anomaly detection.

Sawtooth prediction: Tested accelerator with hardware-in-the-loop training of programmable weights, achieving a final accuracy of 88%, showing that a small network with 1,000 parameters predicts sawtooth crashes.

In all cases, the chips processed the full data sequence in three microseconds or less.

Next chip: Increase parameter count 10x, providing further opportunities for fusion diagnostics and control



AI co-design connects hardware to scientific value

Quantum

Adapt circuits and measurements to physical constraints.

Neuromorphic

Train physical dynamics for fast scientific inference.

Next evidence: reproducible performance and energy at system level

Shinjae Yoo   sjyoo@bnl.gov

Backup