

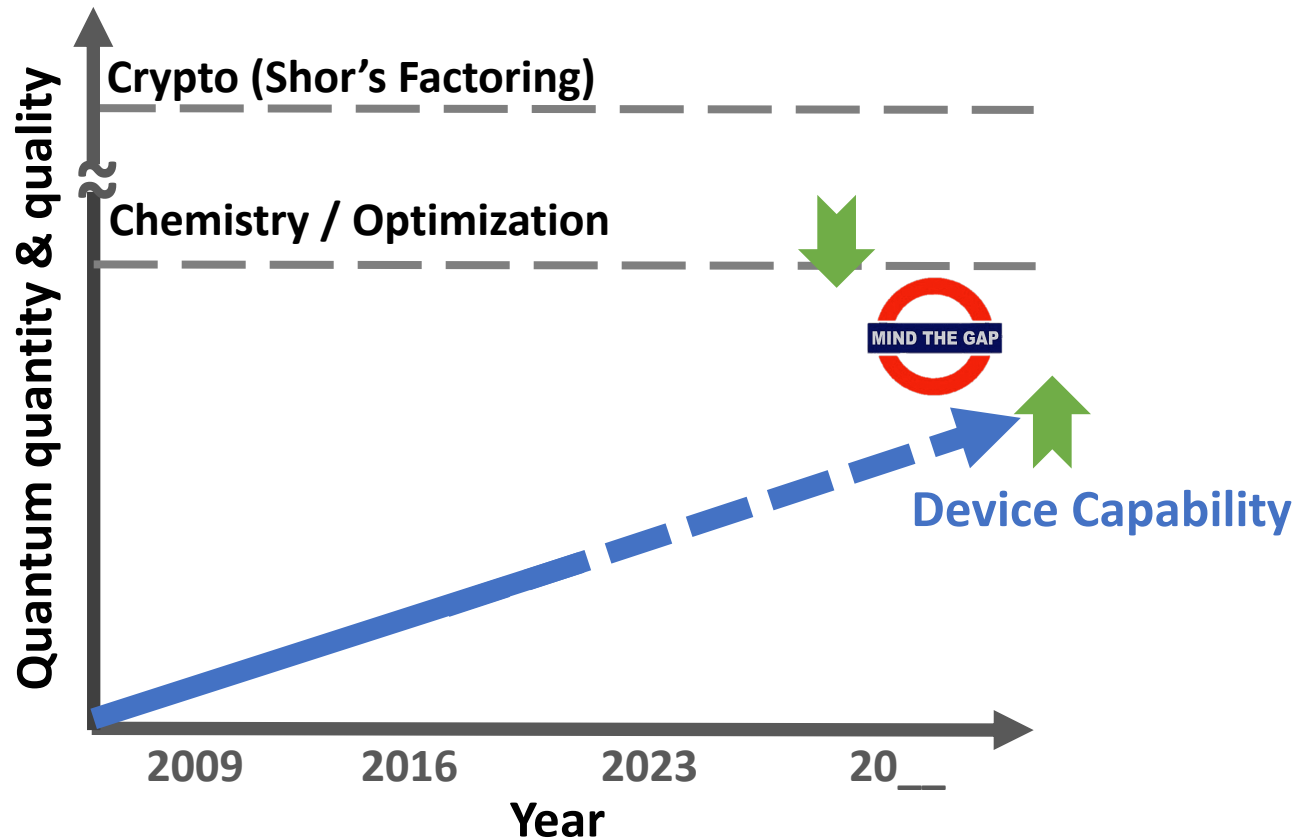


# Quantum-Classical Architectures

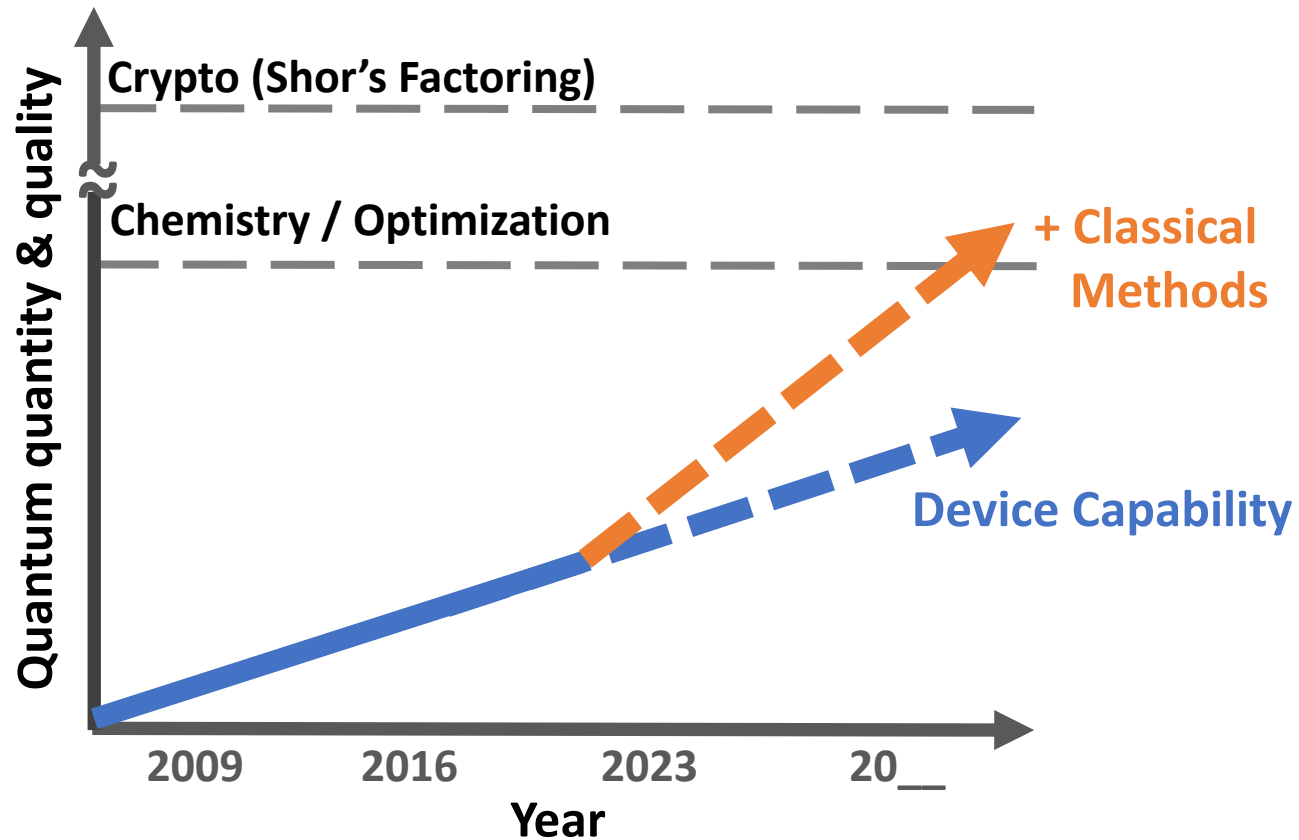
Gokul Subramanian Ravi

Assistant Professor, University of Michigan

# Wide gap between application requirements and technology capability

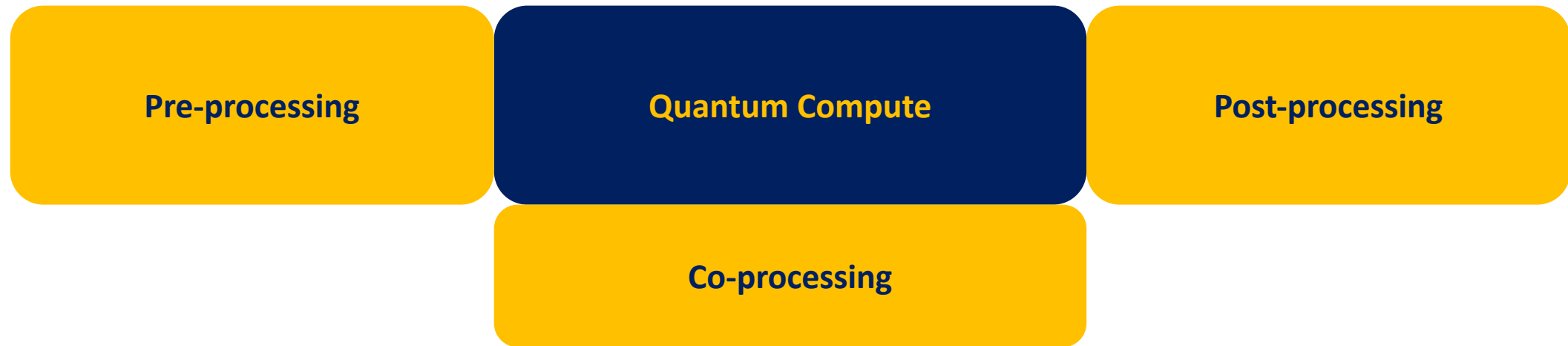


# Bridging the gap: Hybrid quantum-classical computing approaches

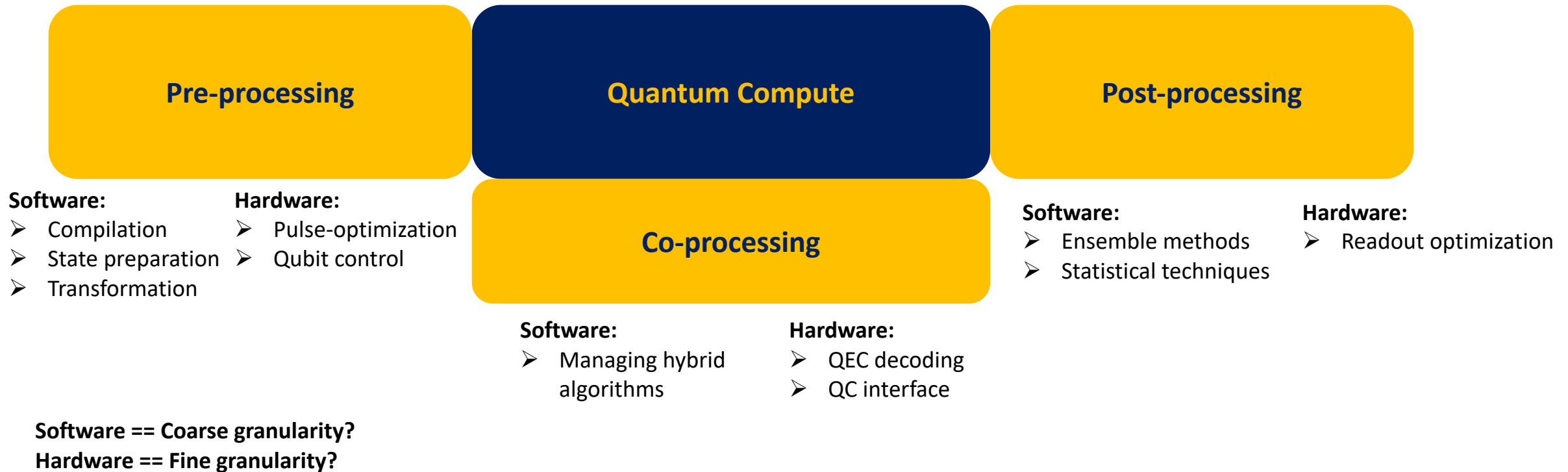


1. PL and Compilation
2. Computer Architecture
3. Feedback-based Optimization
4. High performance computing
5. Cryogenic hardware design
6. Classical simulation
7. Multi-chip / distributed computing
8. Cloud resource management

# Quantum-classical research directions



# Quantum-classical research directions



# Quantum-classical research directions

Pre-processing

Quantum Compute

## Software:

➤ Compilation

➤ State preparation

➤ Transformation

## Hardware:

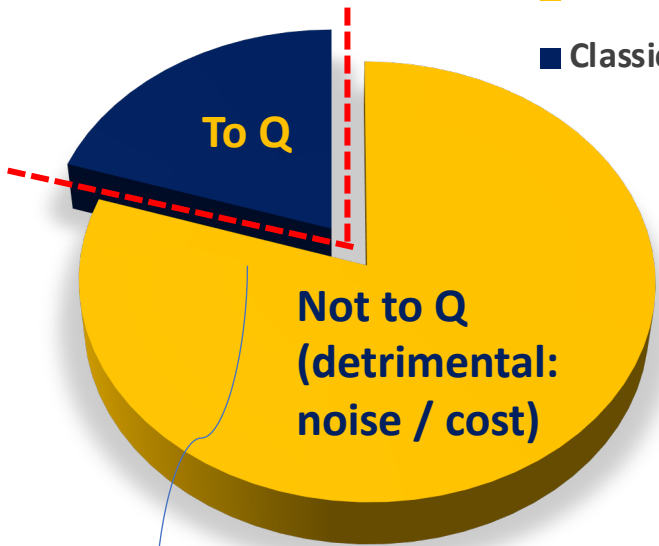
➤ Pulse-optimization

➤ Qubit control

# Classical state preparation / initialization

Application fraction

- Classically tractable
- Classically intractable

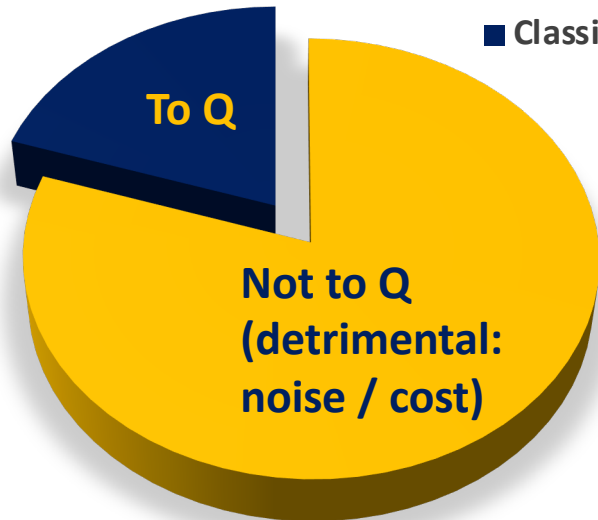


Not a clear boundary

# Classical state preparation / initialization

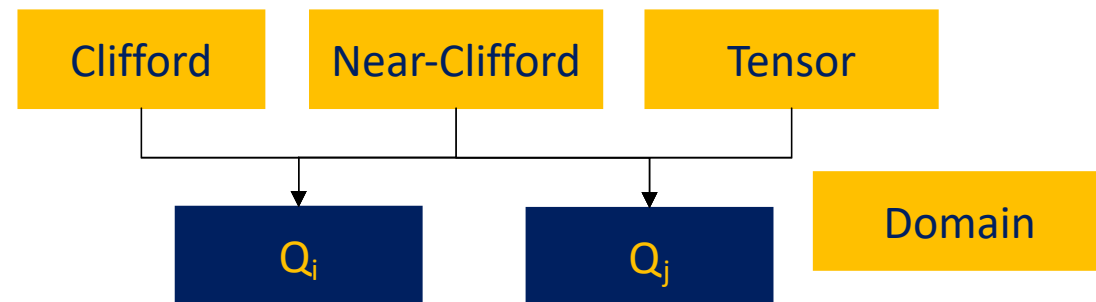
Application fraction

- Classically tractable
- Classically intractable



*Given a classical compute budget, maximize application-tailored processing to reduce the computational load on the quantum device.*

- **Clifford simulation:** scales polynomially, low overheads, 1000s of qubits in seconds, can be run on a laptop, but limited capability.
- **Near-Clifford simulation:** cost and capability both scale exponentially in proportion to *non-cliffordness*, might require HPC.
- **Tensor networks:** cost and capability both scale exponentially in proportion to *entanglement*
- **Domain-specific insights and tools**

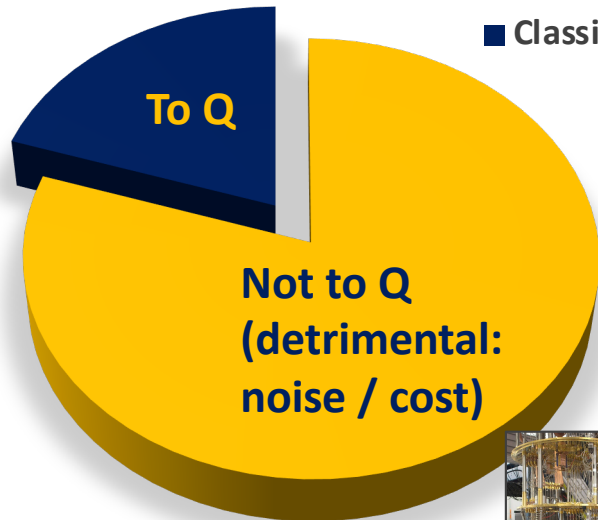




# Classical state preparation / initialization

## Application fraction

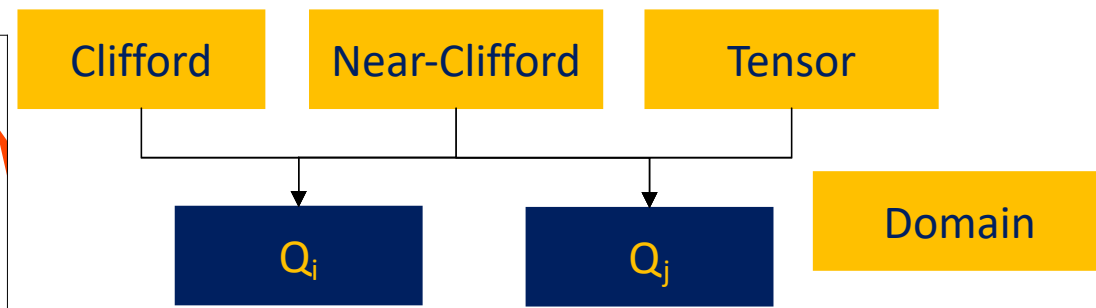
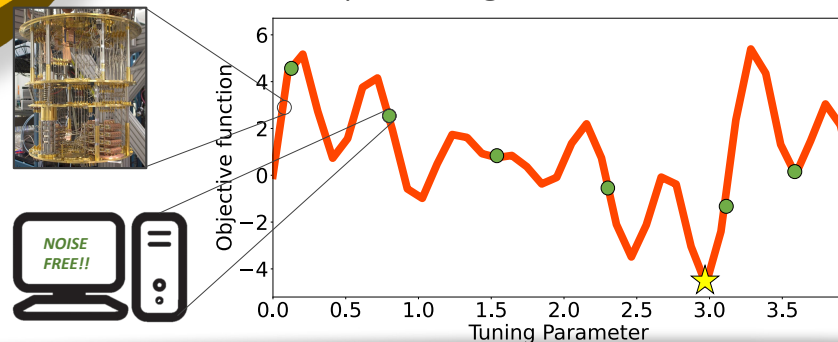
- Classically tractable
- Classically intractable



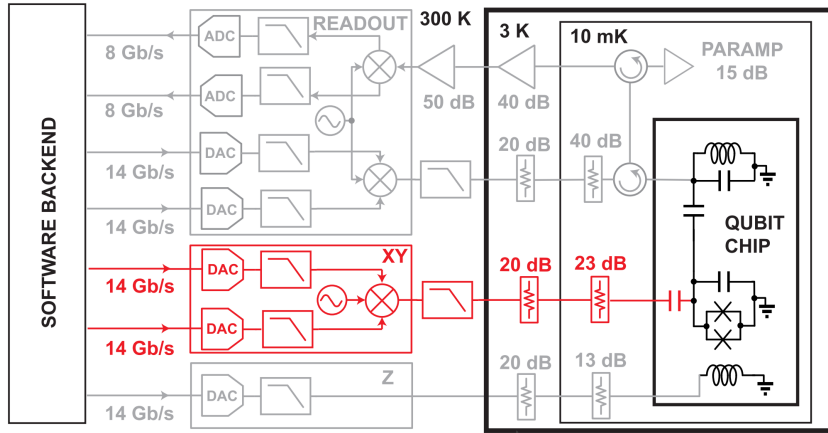
**Given a classical compute budget, maximize application-tailored processing to reduce the computational load on the quantum device.**

- **Clifford simulation:** scales polynomially, low overheads, 1000s of qubits in seconds, can be run on a laptop, but limited capability.
- **Near-Clifford simulation:** cost and capability both scale exponentially in proportion to *non-cliffordness*, might require HPC.
- **Tensor networks:** cost and capability both scale exponentially in proportion to *entanglement*
- **Domain-specific insights and tools**

CAFQA: A classical simulation bootstrap for variational quantum algorithms. ASPLOS 2023

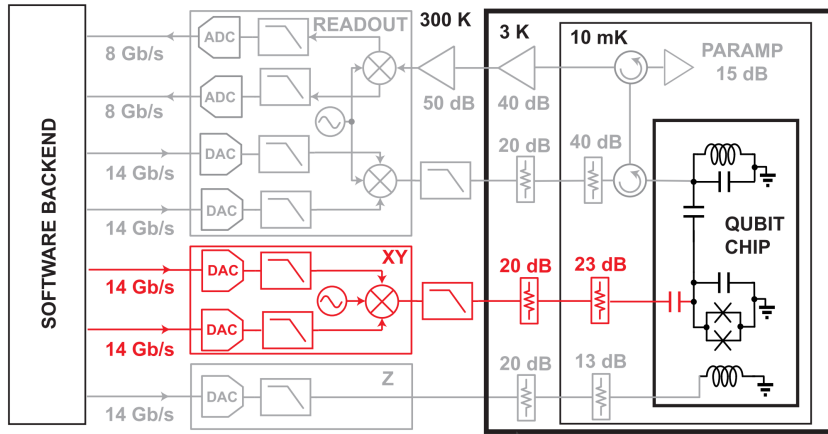


# Qubit control

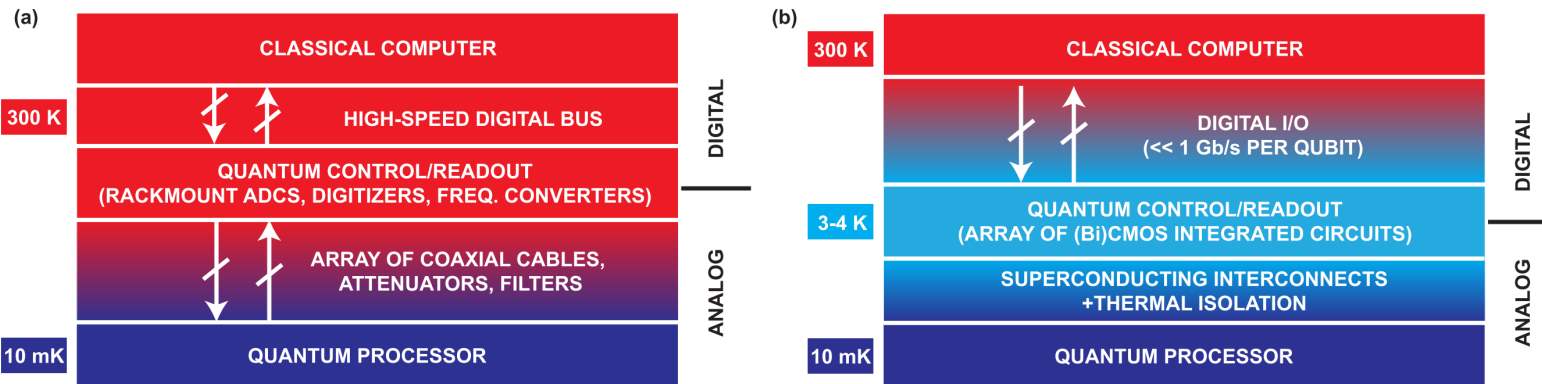


Design and Characterization of a 28-nm Bulk-CMOS Cryogenic Quantum Controller Dissipating Less Than 2 mW at 3 K (Google). 2019

# Qubit control

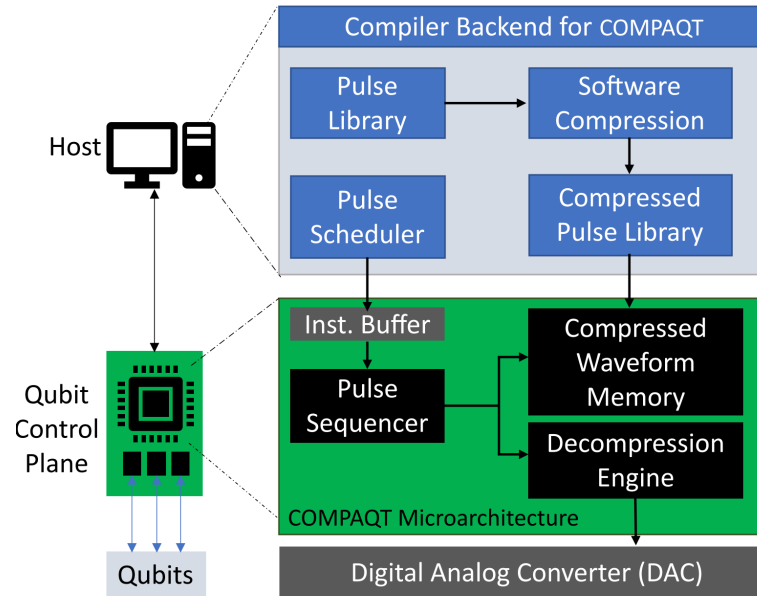
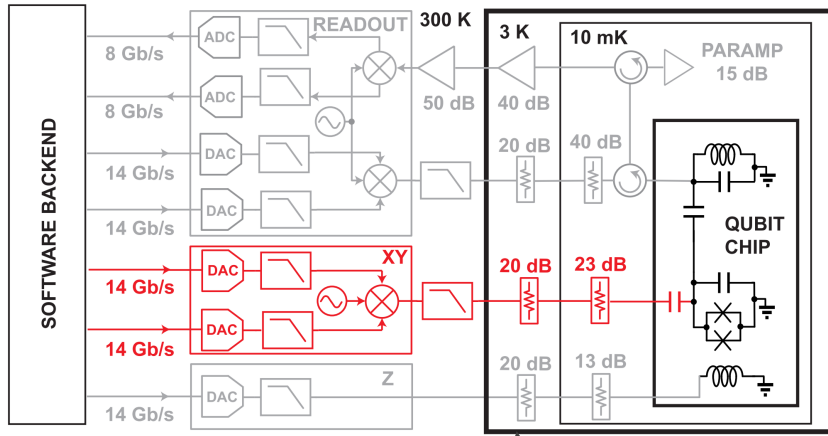


Design and Characterization of a 28-nm Bulk-CMOS Cryogenic Quantum Controller Dissipating Less Than 2 mW at 3 K (Google). 2019

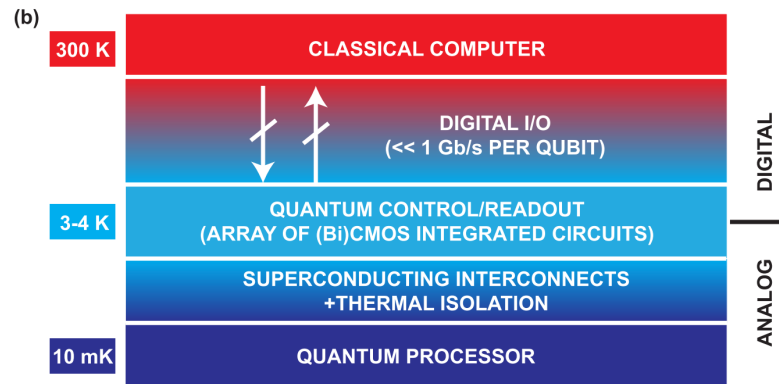
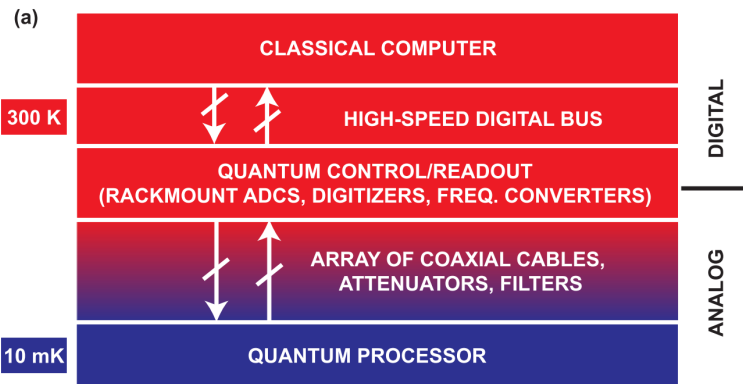


# Qubit control

COMPAQT: Compressed Waveform Memory Architecture for Scalable Qubit Control. MICRO 2022

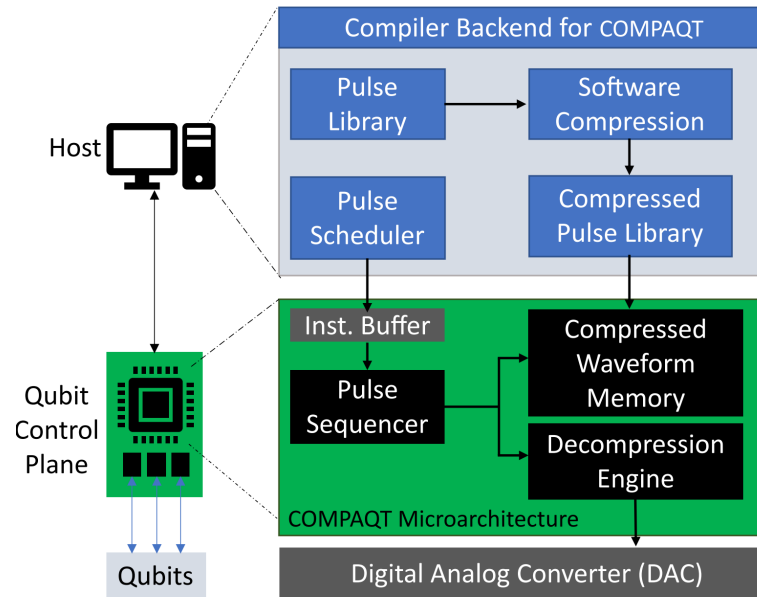
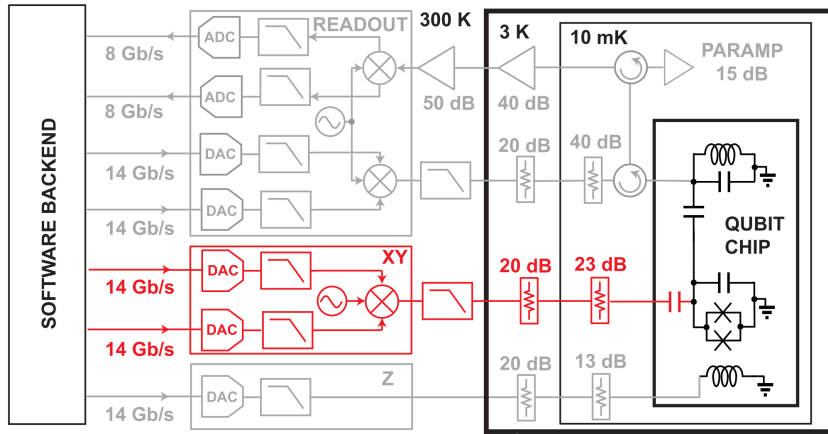


Design and Characterization of a 28-nm Bulk-CMOS Cryogenic Quantum Controller Dissipating Less Than 2 mW at 3 K (Google). 2019

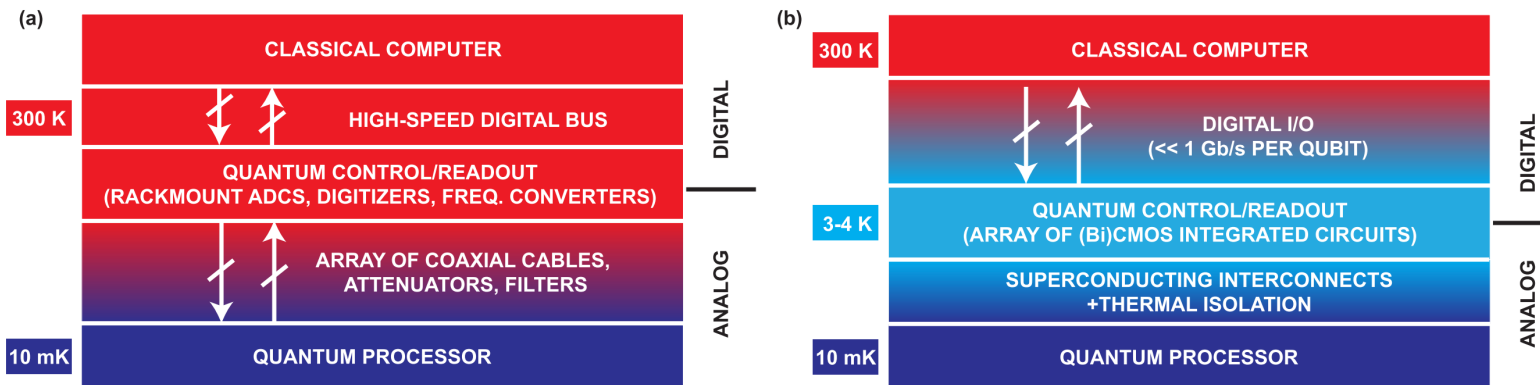


# Qubit control

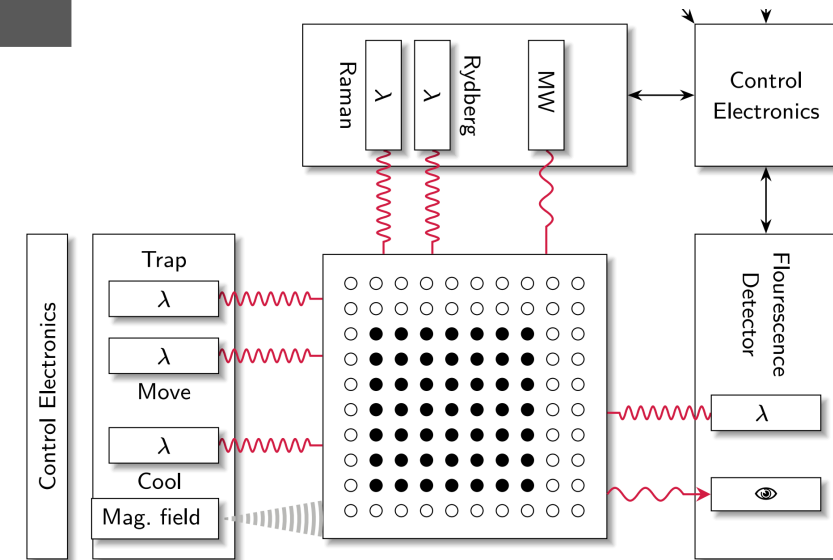
COMPAQT: Compressed Waveform Memory Architecture for Scalable Qubit Control. MICRO 2022



Design and Characterization of a 28-nm Bulk-CMOS Cryogenic Quantum Controller Dissipating Less Than 2 mW at 3 K (Google). 2019

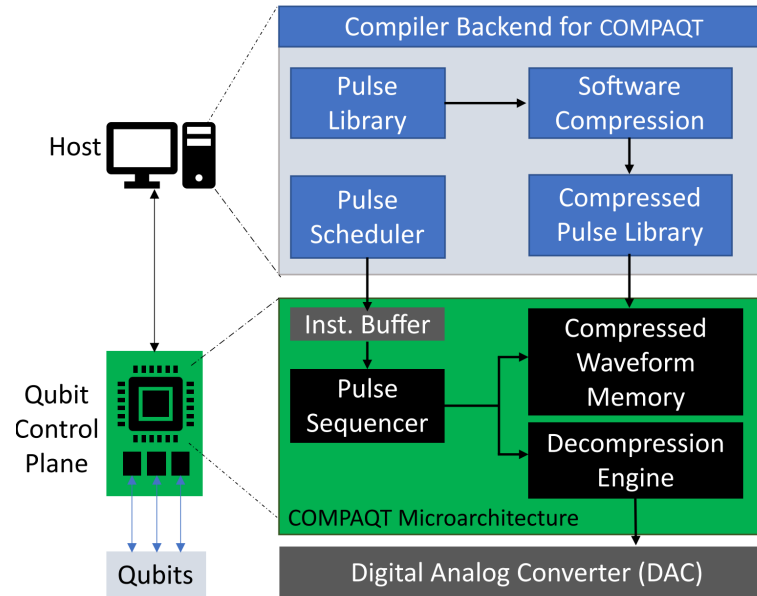
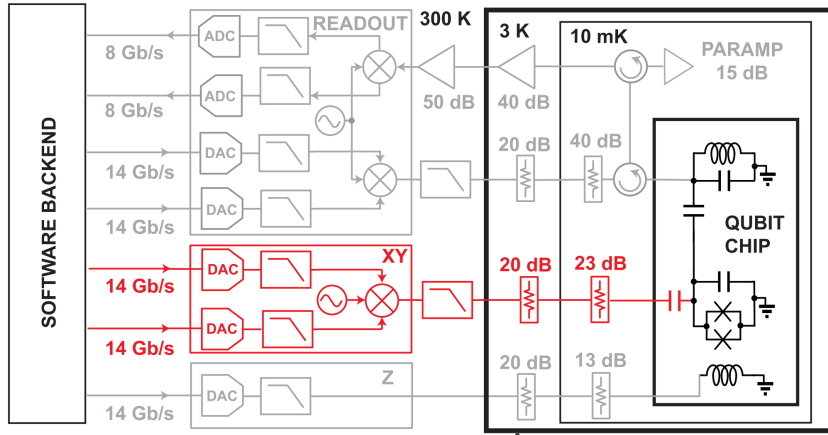


Neutral Atom Quantum Computing Hardware: Performance and End-User Perspective. 2023



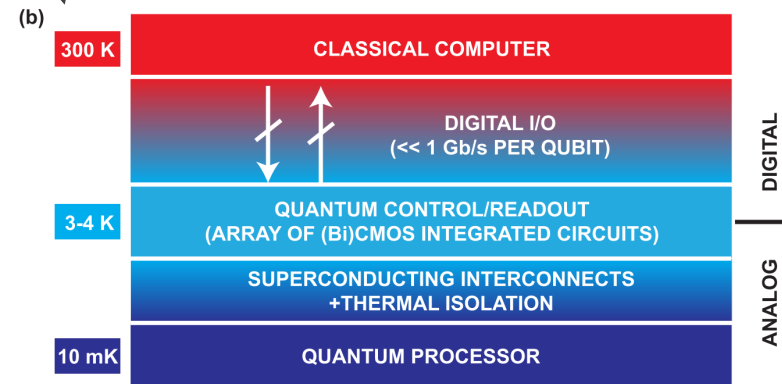
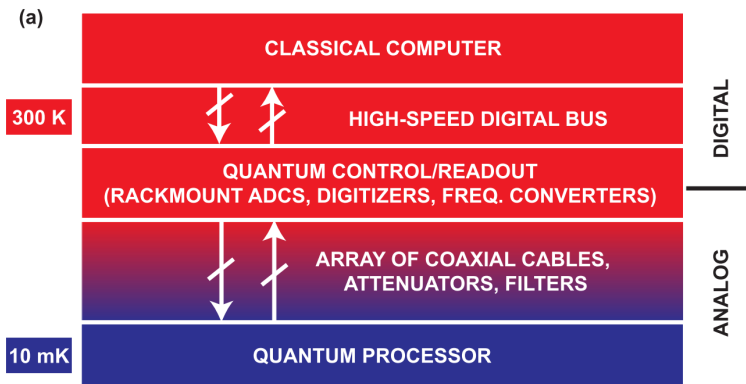
# Qubit control

COMPAQT: Compressed Waveform Memory Architecture for Scalable Qubit Control. MICRO 2022

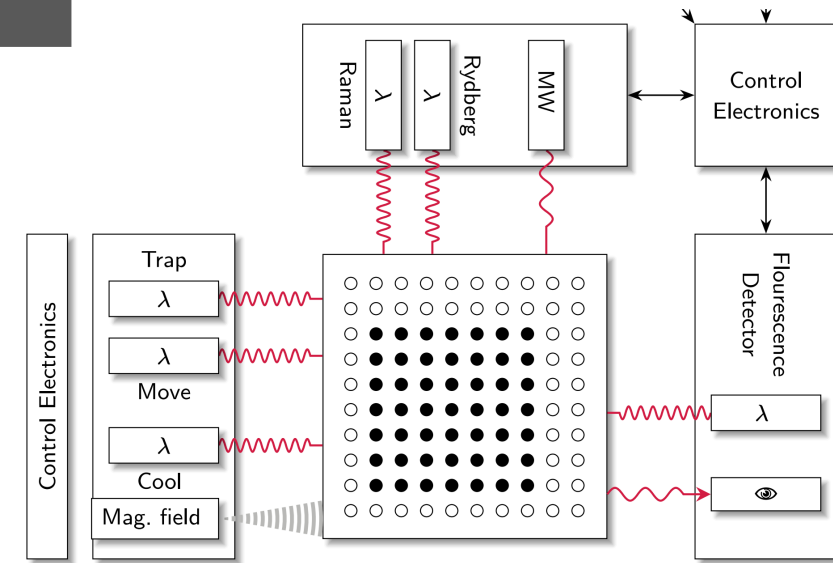


*Software and hardware methods to reduce control overheads as we scale to millions of qubits.*

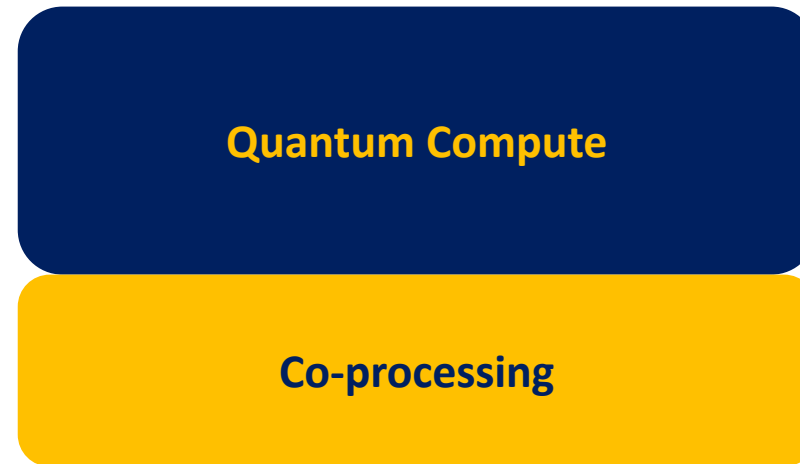
Design and Characterization of a 28-nm Bulk-CMOS Cryogenic Quantum Controller Dissipating Less Than 2 mW at 3 K (Google). 2019



Neutral Atom Quantum Computing Hardware: Performance and End-User Perspective. 2023



# Quantum-classical research directions



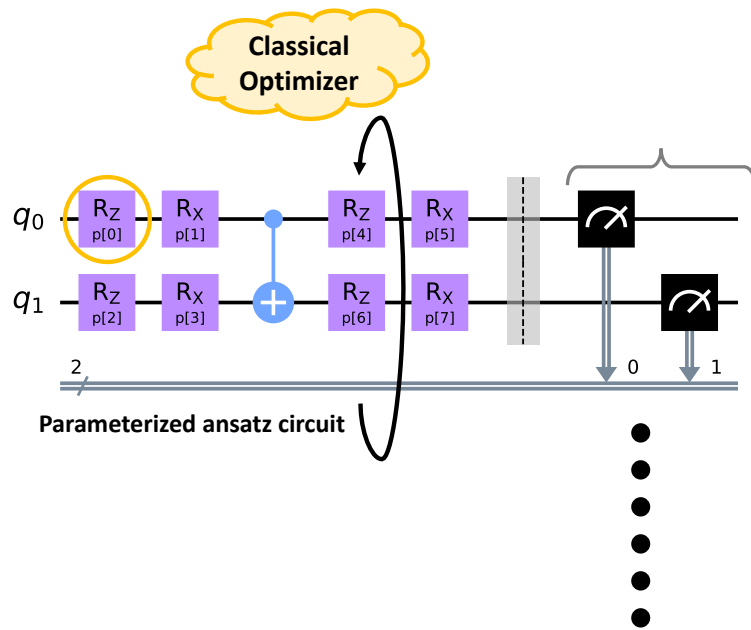
## Software:

- Managing hybrid algorithms

## Hardware:

- QEC decoding
- QC interface

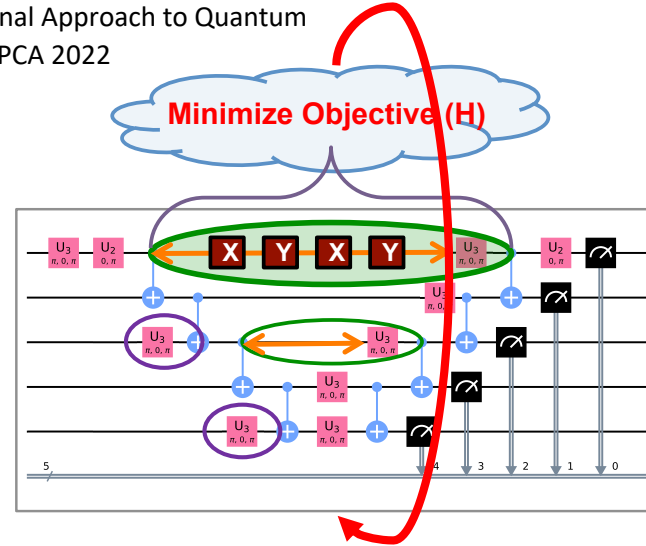
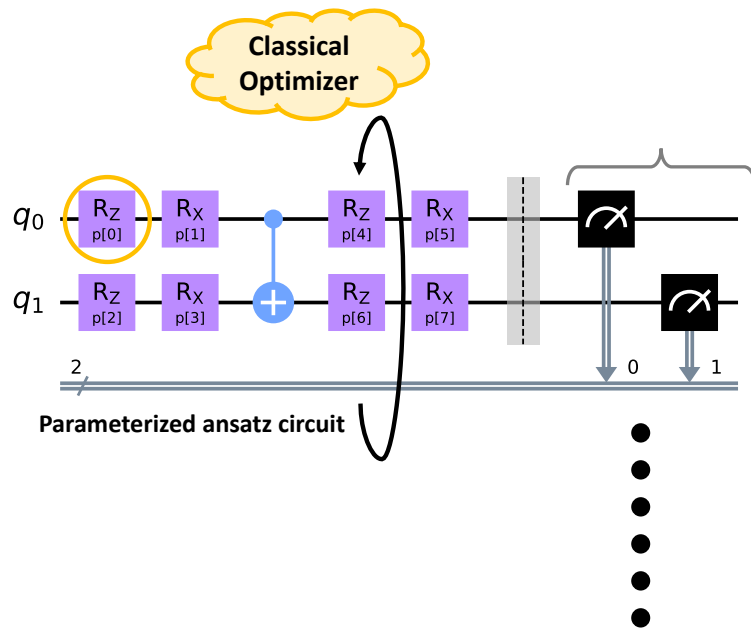
# Managing hybrid quantum algorithms





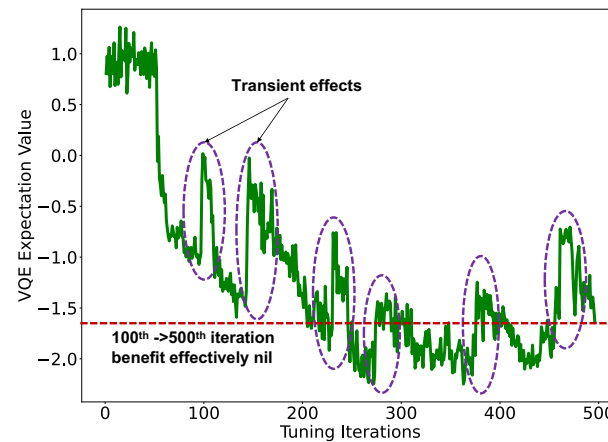
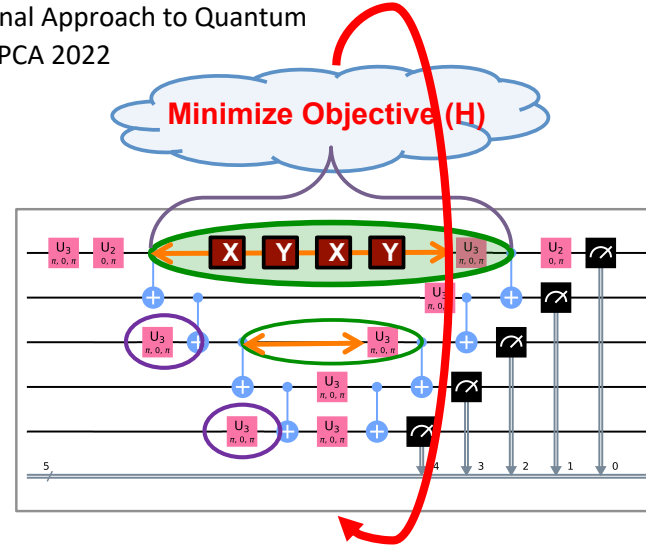
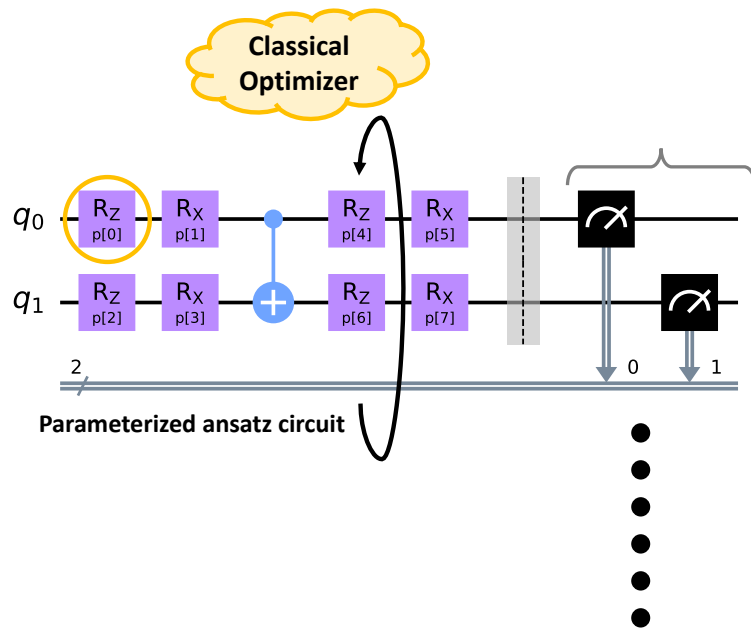
# Managing hybrid quantum algorithms

VAQEM: A Variational Approach to Quantum Error Mitigation. HPCA 2022



# Managing hybrid quantum algorithms

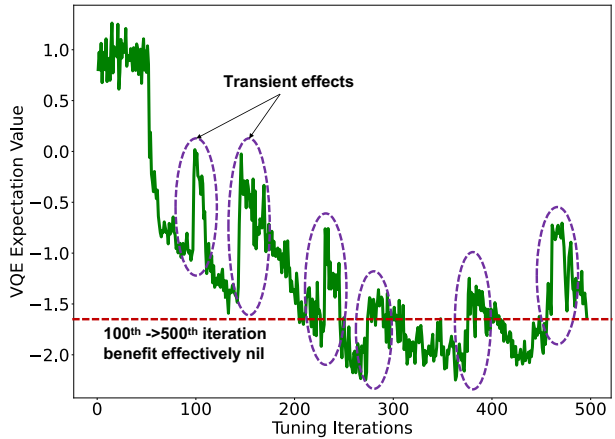
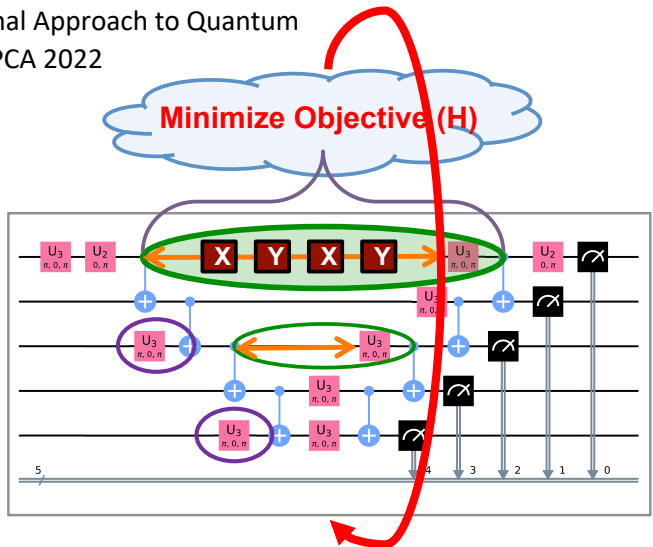
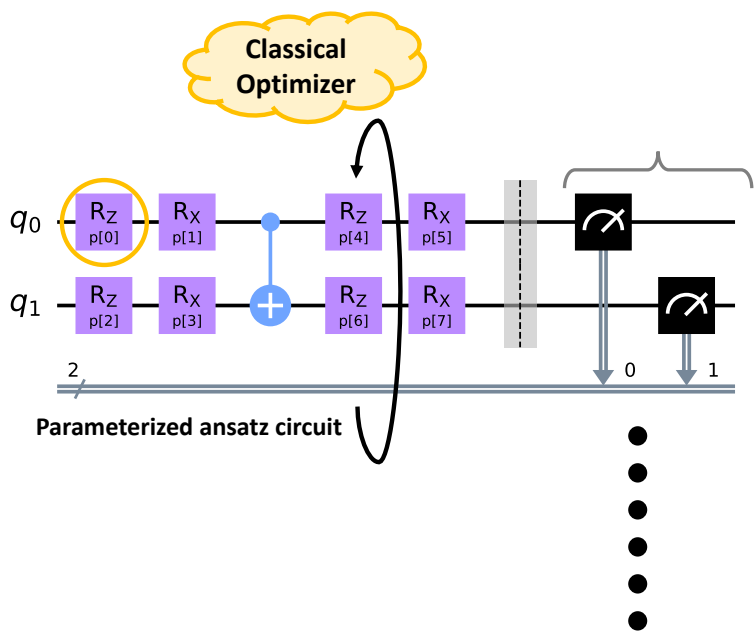
VAQEM: A Variational Approach to Quantum Error Mitigation. HPCA 2022



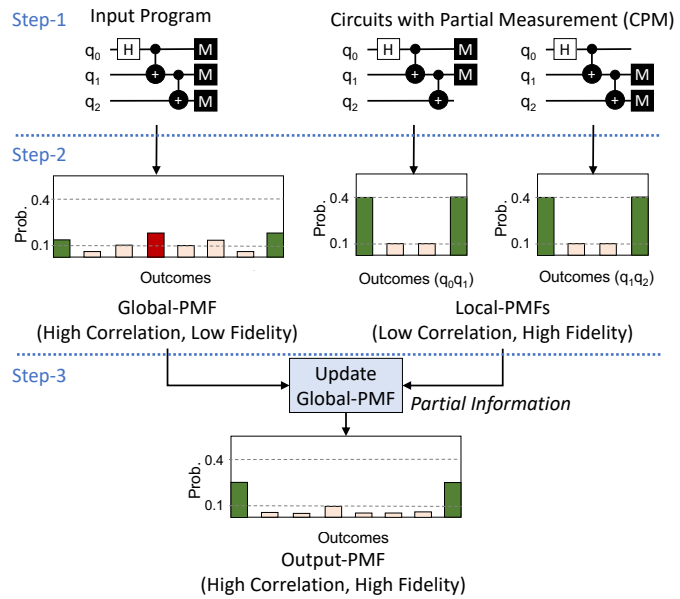
Navigating the Dynamic Noise Landscape of Variational Algorithms with QISMET. ASPLOS 2023

# Managing hybrid quantum algorithms

VAQEM: A Variational Approach to Quantum Error Mitigation. HPCA 2022



VarSaw: Application-tailored Measurement Error Mitigation for Variational Quantum Algorithms. ASPLOS 2024

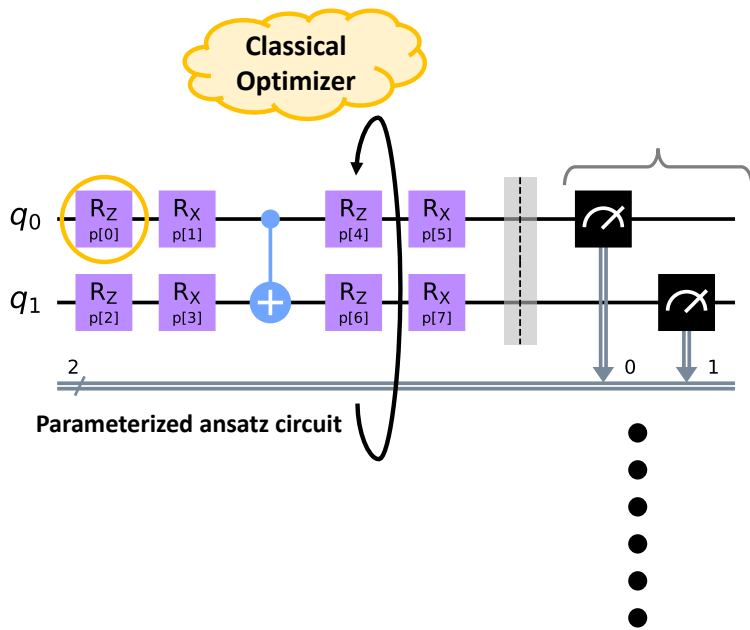


Navigating the Dynamic Noise Landscape of Variational Algorithms with QISMET. ASPLOS 2023

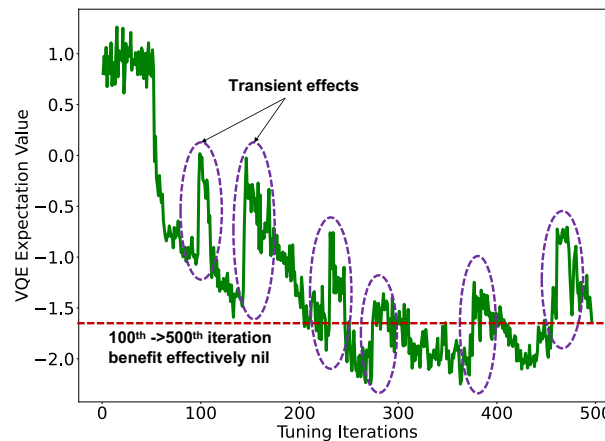
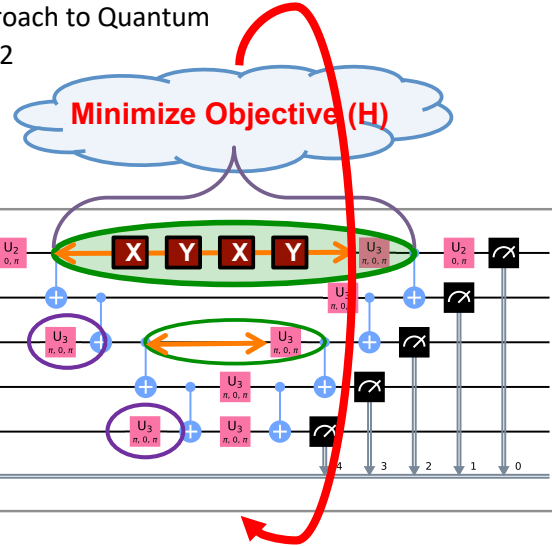


# Managing hybrid quantum algorithms

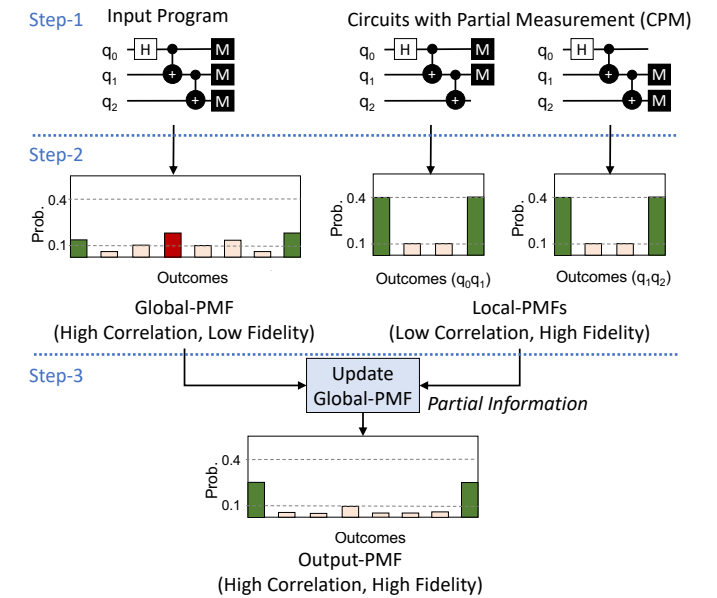
VAQEM: A Variational Approach to Quantum Error Mitigation. HPCA 2022



**Increased tuning parameters and additional features makes optimizer and related classical processing efficiency critical. Non-trivial HW/SW resources.**

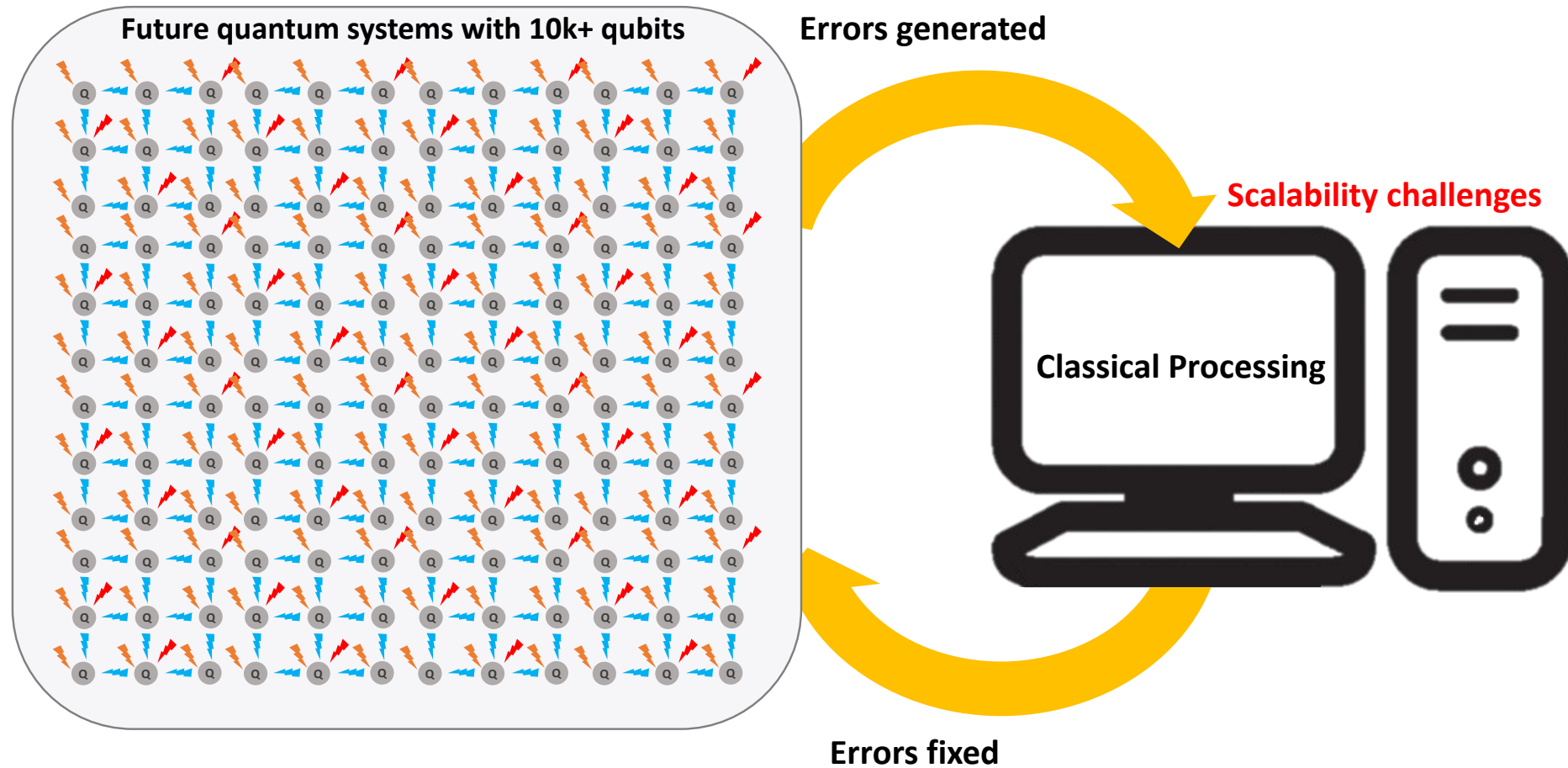


VarSaw: Application-tailored Measurement Error Mitigation for Variational Quantum Algorithms. ASPLOS 2024

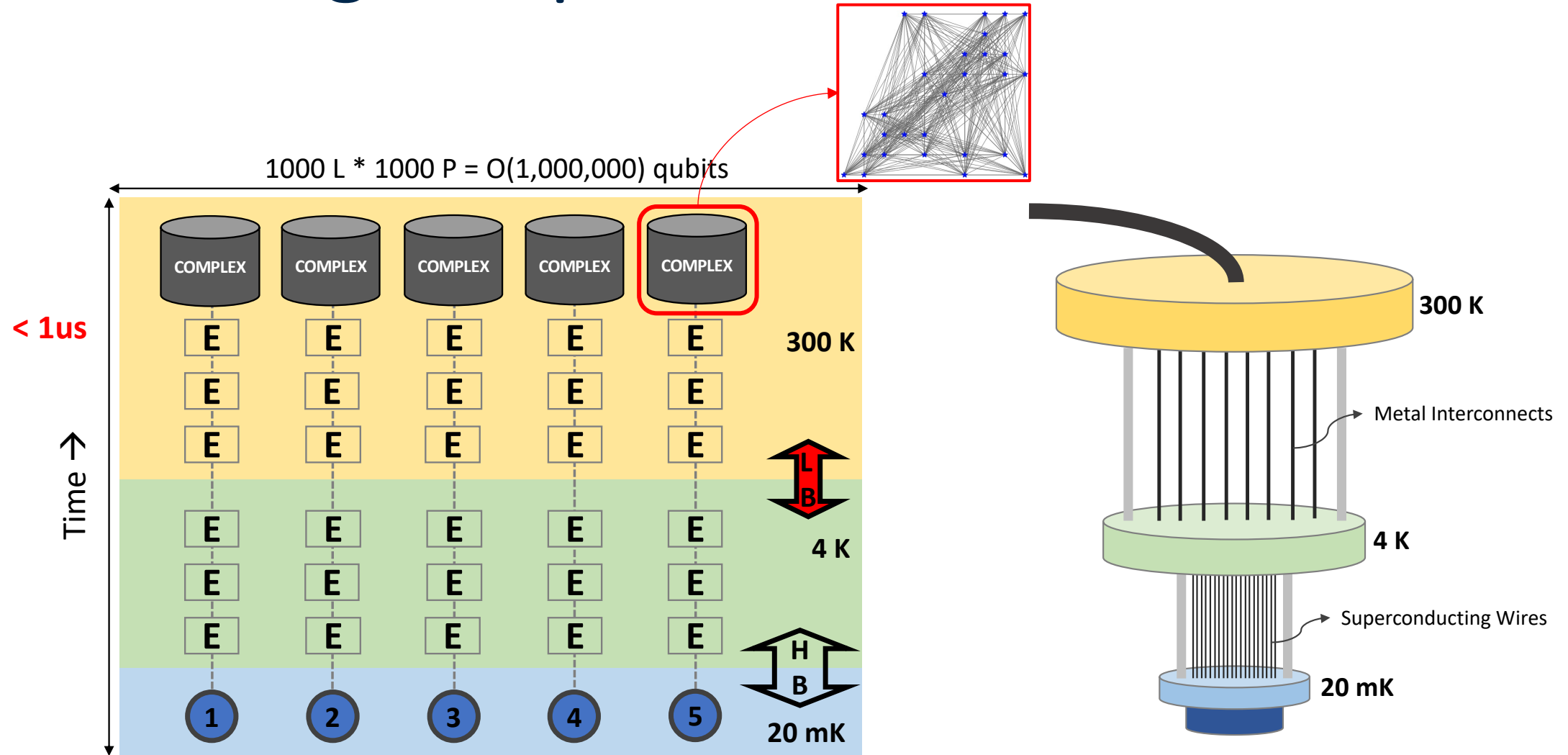


Navigating the Dynamic Noise Landscape of Variational Algorithms with QISMET. ASPLOS 2023

# Decoding for quantum error correction

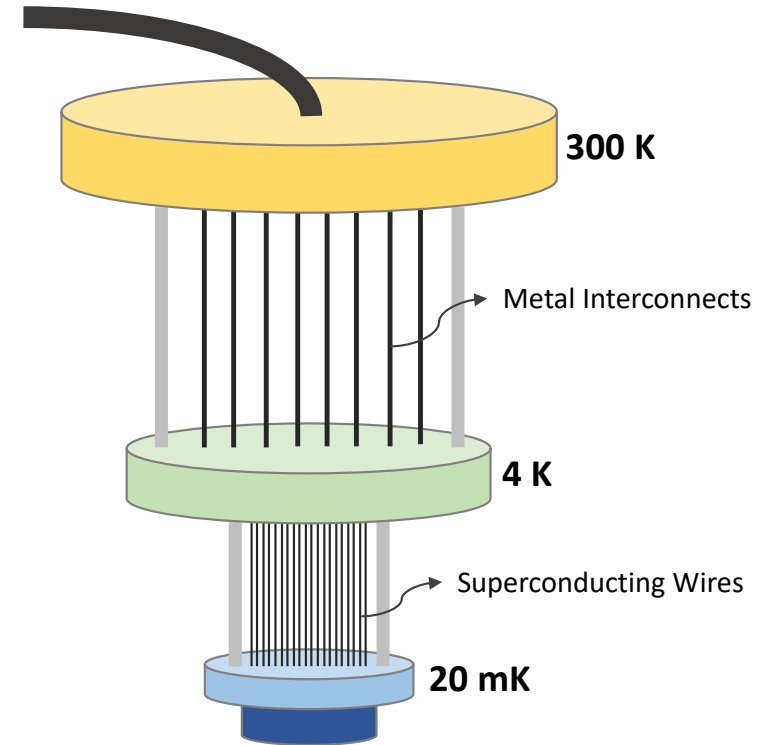
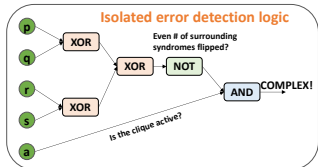
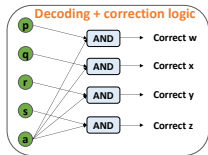
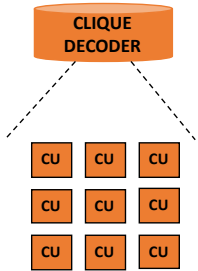
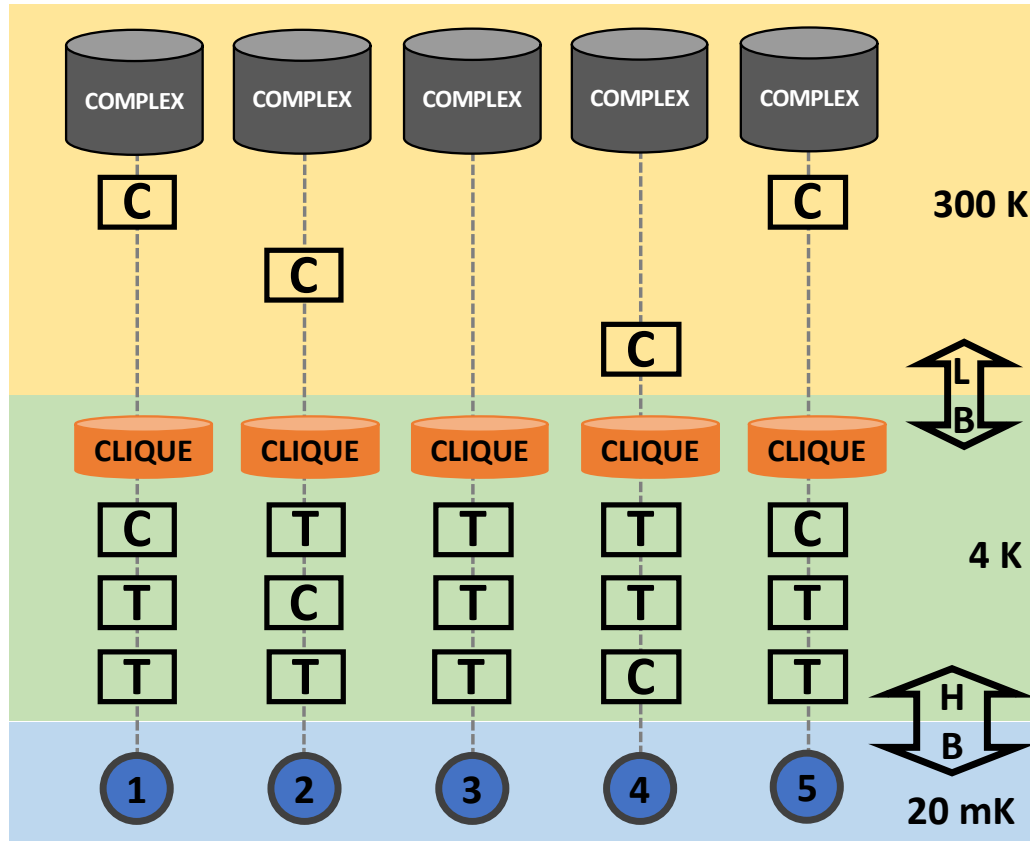


# Decoding for quantum error correction

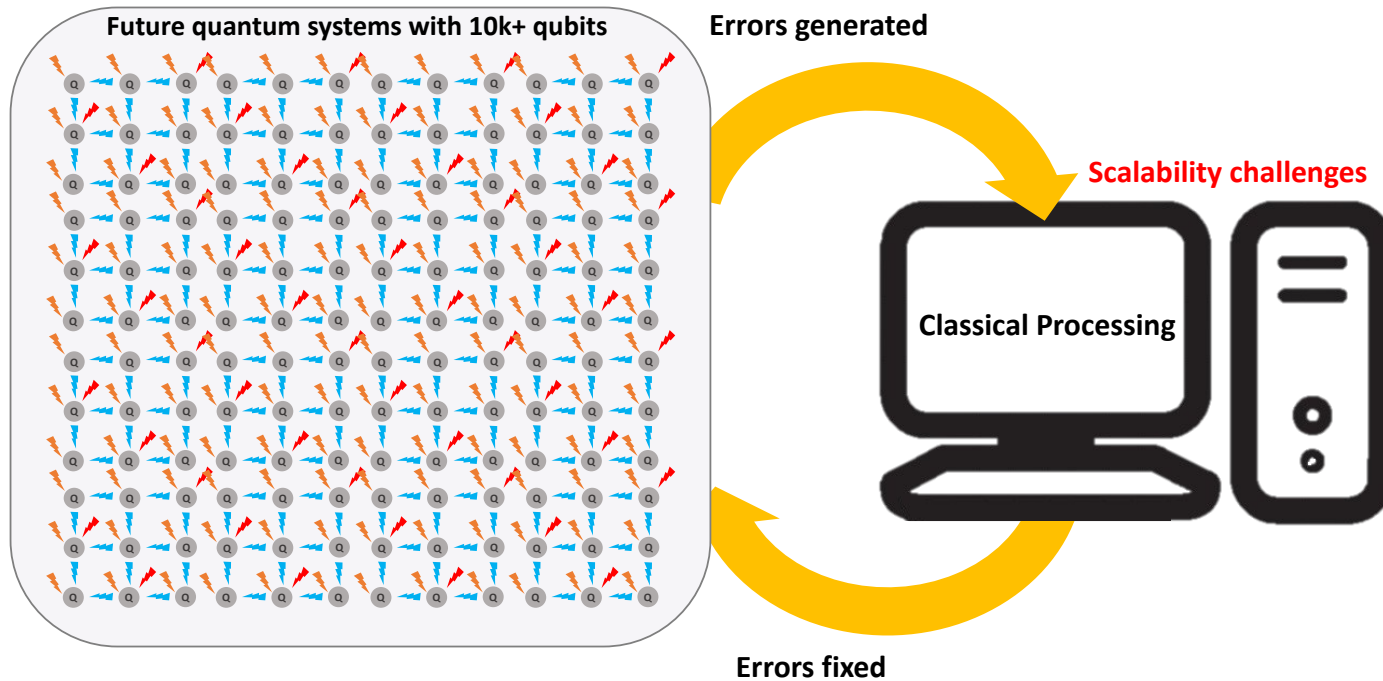


# Decoding for quantum error correction

Clique: Better Than Worst-Case Decoding for Quantum Error Correction. ASPLOS 2023



# Decoding for quantum error correction



## **Opportunities:**

- *Efficient decoders for exotic codes.*
- *Latency reduction in decoding.*
- *Bandwidth reduction at quantum-classical interface.*
- *ASICs/HPC/GPUs?*



# Quantum-classical research directions

Quantum Compute

Post-processing

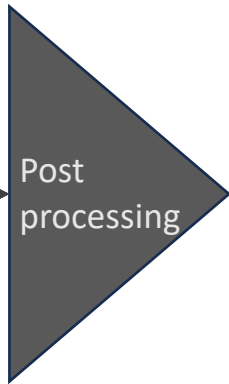
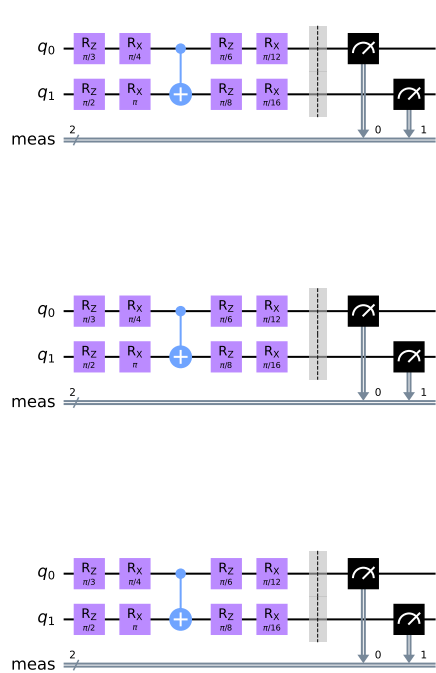
**Software:**

- Ensemble methods
- Statistical techniques

**Hardware:**

- Readout optimization

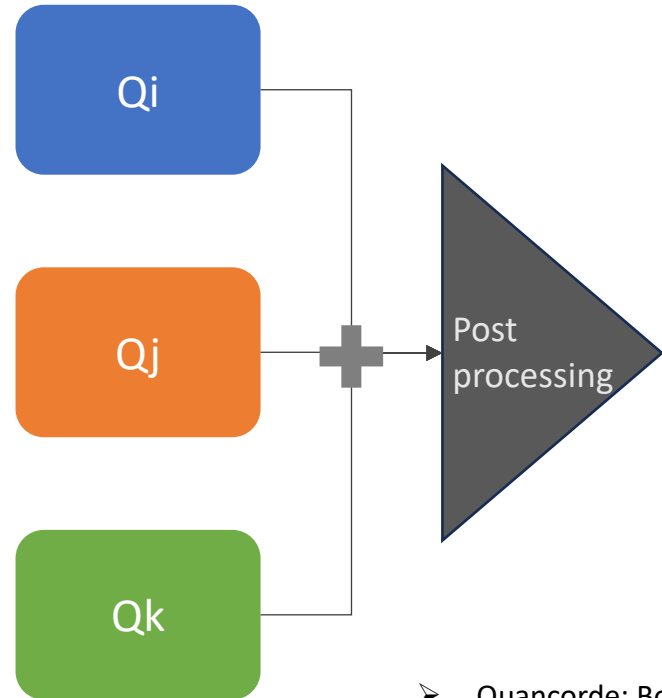
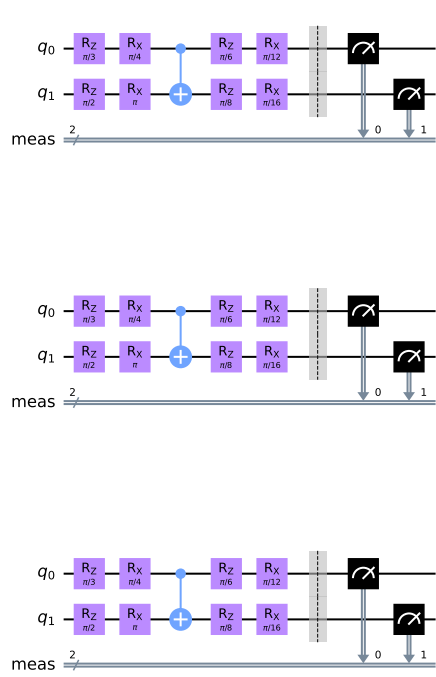
# Resource ensemble methods



Boosting fidelity / functionality beyond the capability of any single device.

- Quancorde: Boosting fidelity with Quantum Canary Ordered Diverse Ensembles. ICRC 2023.
- EQC : Ensembled Quantum Computing for Variational Quantum Algorithms. ISCA 2022.
- CutQC: Using Small Quantum Computers for Large Quantum Circuit Evaluations. ASPLOS 2021.
- Ensemble of Diverse Mappings. MICRO 2019.
- Zero noise extrapolation (many papers).
- Probabilistic error cancellation. Nature Physics 2023.

# Resource ensemble methods



## Opportunities:

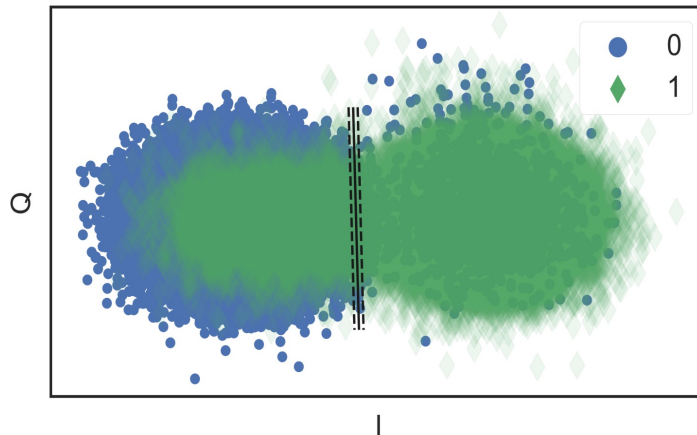
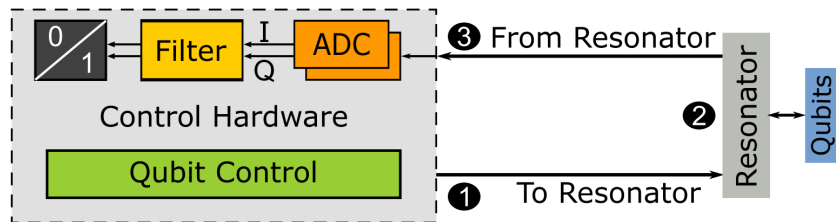
- Reducing classical post-processing overheads.
- Reducing quantum execution overheads.
- (Classically simulable) noise modeling.
- Full-stack resource management framework to manage ensemble methods.

Boosting fidelity / functionality beyond the capability of any single device.

- Quancorde: Boosting fidelity with Quantum Canary Ordered Diverse Ensembles. ICRC 2023.
- EQC : Ensembled Quantum Computing for Variational Quantum Algorithms. ISCA 2022.
- CutQC: Using Small Quantum Computers for Large Quantum Circuit Evaluations. ASPLOS 2021.
- Ensemble of Diverse Mappings. MICRO 2019.
- Zero noise extrapolation (many papers).
- Probabilistic error cancellation. Nature Physics 2023.

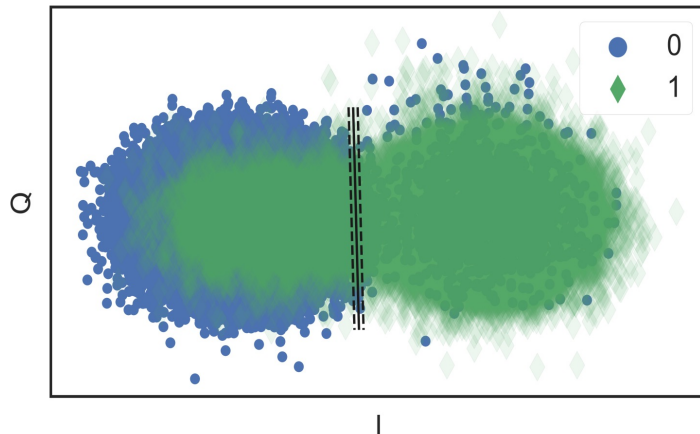
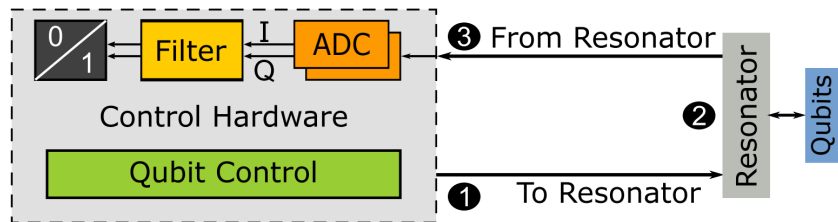
# Readout optimization

Scaling Qubit Readout with Hardware Efficient  
Machine Learning Architectures. ISCA 2023

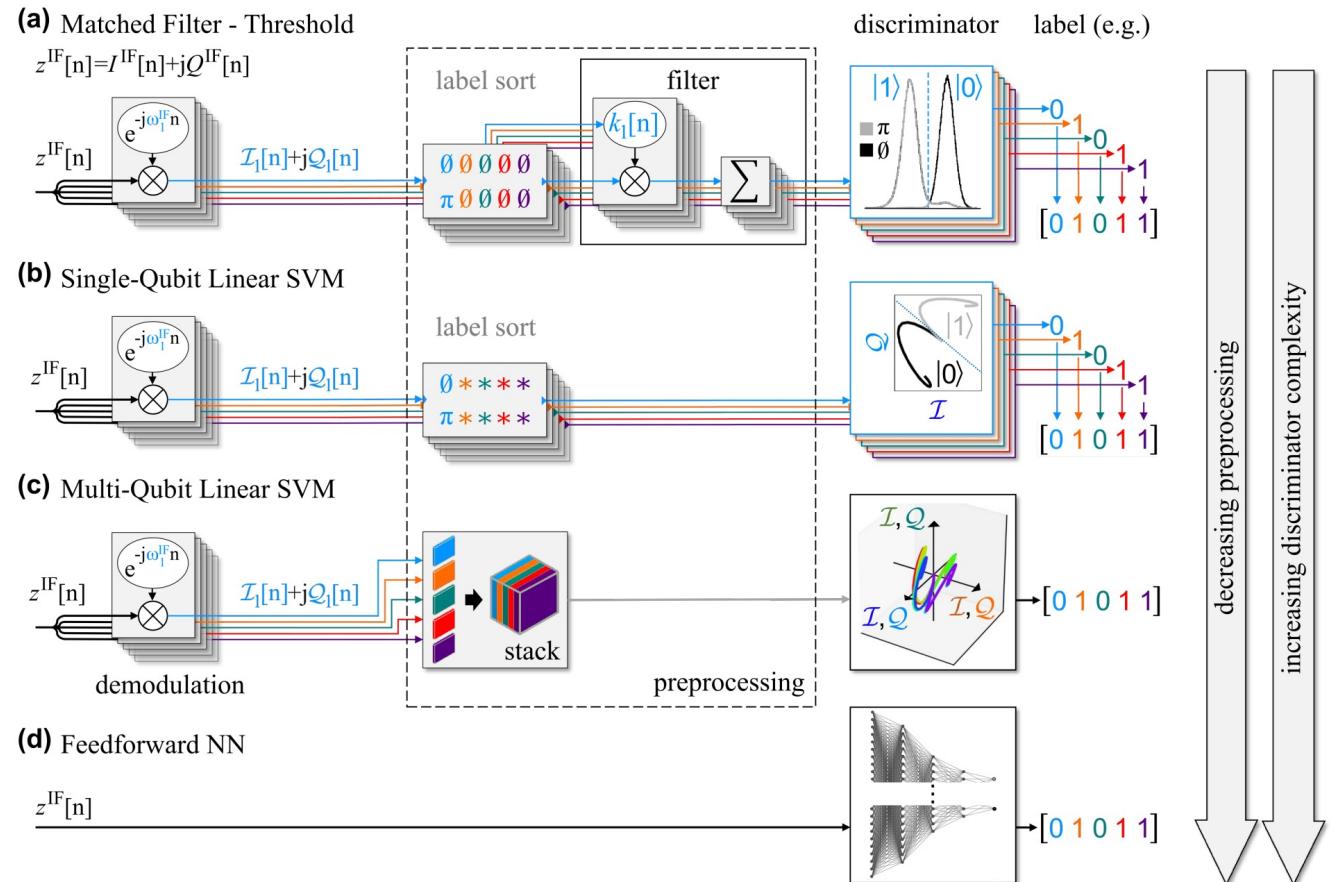


# Readout optimization

Scaling Qubit Readout with Hardware Efficient Machine Learning Architectures. ISCA 2023



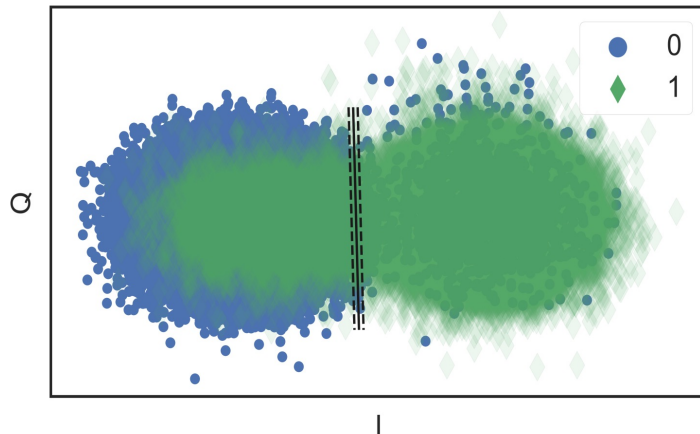
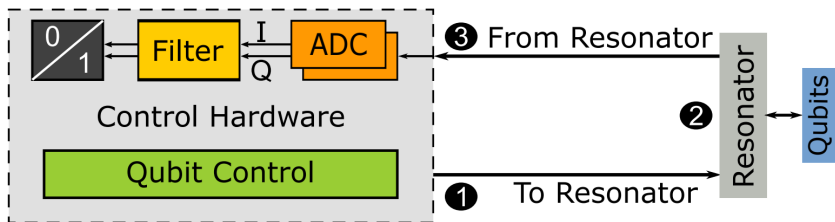
Deep Neural Network Discrimination of Multiplexed Superconducting Qubit States. 2022.



# Readout optimization

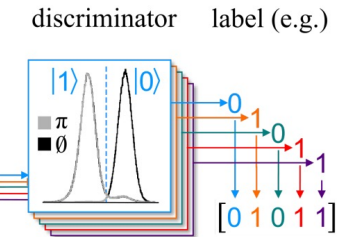
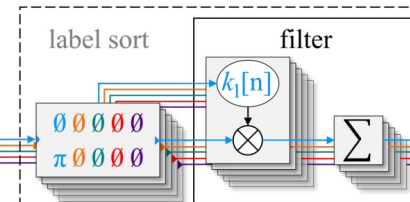
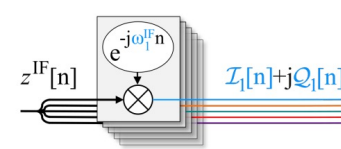
Deep Neural Network Discrimination of Multiplexed Superconducting Qubit States. 2022.

Scaling Qubit Readout with Hardware Efficient Machine Learning Architectures. ISCA 2023

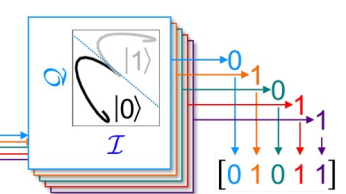
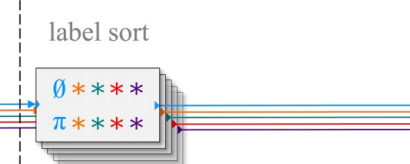
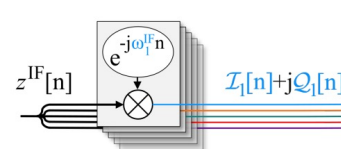


(a) Matched Filter - Threshold

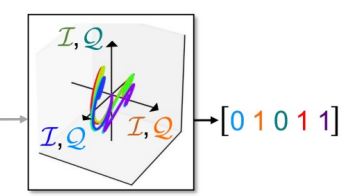
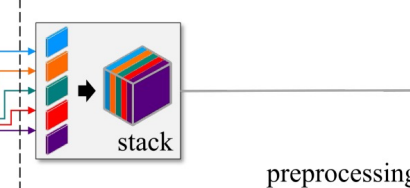
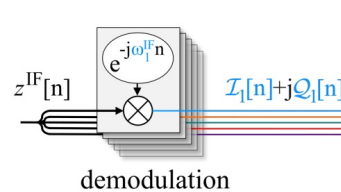
$$z^{IF}[n] = I^{IF}[n] + jQ^{IF}[n]$$



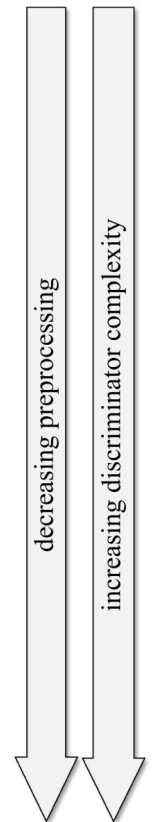
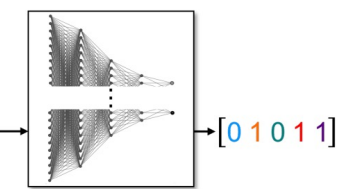
(b) Single-Qubit Linear SVM



(c) Multi-Qubit Linear SVM



(d) Feedforward NN



**Need for high accuracy, low latency, low overheads, circuit/device-awareness, etc.**

# Quantum-classical research directions

