Physics-Aware, Full-Stack Quantum Software Optimizations

Fred Chong

Seymour Goodman Professor, University of Chicago
Chief Scientist for Quantum Software, Infleqtion

Lead PI, the EPiQC Project, an NSF Expedition in Computing
NSF 1730449/1730082/1729369/1832377/1730088
NSF Phy-232580/OMA-2016136
DOE DE-SC0020289/0020331/QNEXT
ARO W911NF-23-1-0077; Wellcome-Leap Q4Bio
Disclosure: FC is an advisor to QCI

With Co-Pis: Ken Brown, Ike Chuang, Diana Franklin, Danielle Harlow, Aram Harrow, Andrew Houck, Margaret Martonosi, Robert Rand, John Reppy, David Schuster, Peter Shor

And Collaborators: Peter Love, Zach Manchester, Alex Pearson, Moin Qureshi, Samantha Riesenfeld

Why Computer Architecture?

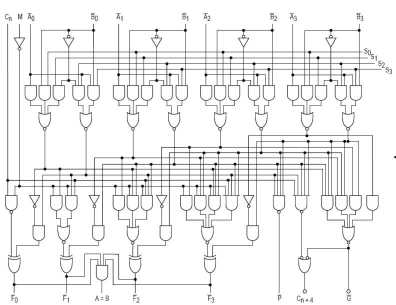
PL

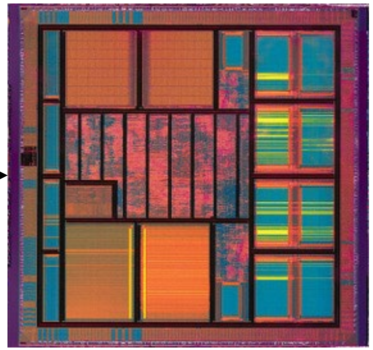
Compiler

Assembly

Assembler

Binary





Why Quantum Architecture?

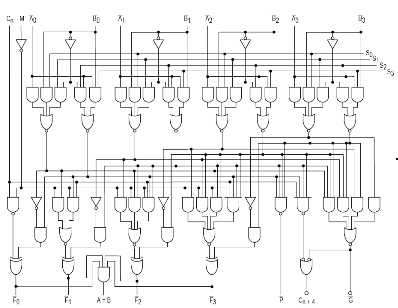
PL

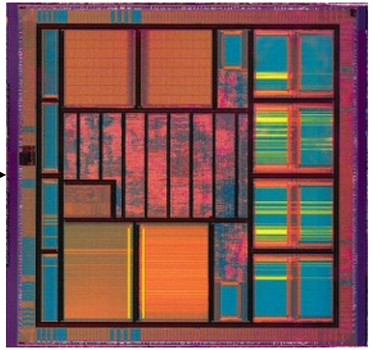
Compiler

Assembly

Assembler

Binary





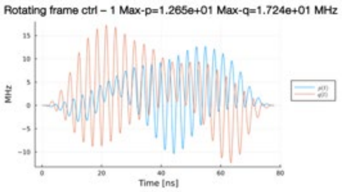
QPL

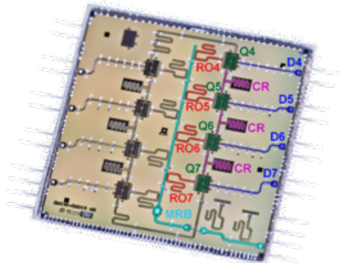
Compiler

QASM

Assembler

Waveforms





Quantum Ecosystem

NATIONAL QUANTUM INITIATIVE
THE FEDERAL SOURCE AND GATEWAY TO QUANTUM R&D ACROSS THE U.S. GOVERNMENT

NSF

DEPARTMENT OF ENERGY · UNITED STATES OF AMERICA

STAQ | SOFTWARE-TAILORED ARCHITECTURES for QUANTUM CODESIGN

Q-NEXT
NEXT GENERATION QUANTUM SCIENCE AND ENGINEERING

QACTI

MACH-|Q⟩

HQAN

SMART

NSF QuBBE

CHICAGO QUANTUM EXCHANGE

leap
Quantum for Bio
BIO

IBM

Infleqtion

Why Quantum Computing?

- Fundamentally change what is computable
  - The only means to potentially scale computation exponentially with the number of devices
- Solve currently intractable problems in chemistry, simulation, and optimization
  - Could lead to new nanoscale materials, better photovoltaics, better nitrogen fixation, and more

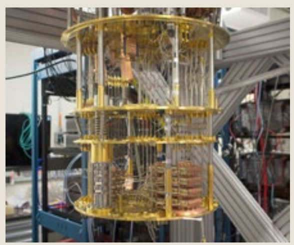


- A new industry and scaling curve to accelerate key applications
  - Not a full replacement for Moore's Law, but perhaps helps in key domains
- Lead to more insights in classical computing
  - Previous insights in chemistry, physics and cryptography
  - Challenge classical algorithms to compete w/ quantum algorithms

Why Now?

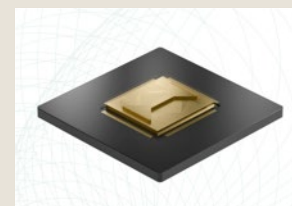
Now is a privileged time in the history of science and technology, as we are witnessing the opening of the NISQ era (where NISQ = noisy intermediate-scale quantum).

– John Preskill, Caltech



IBM
133 superconductor qubits

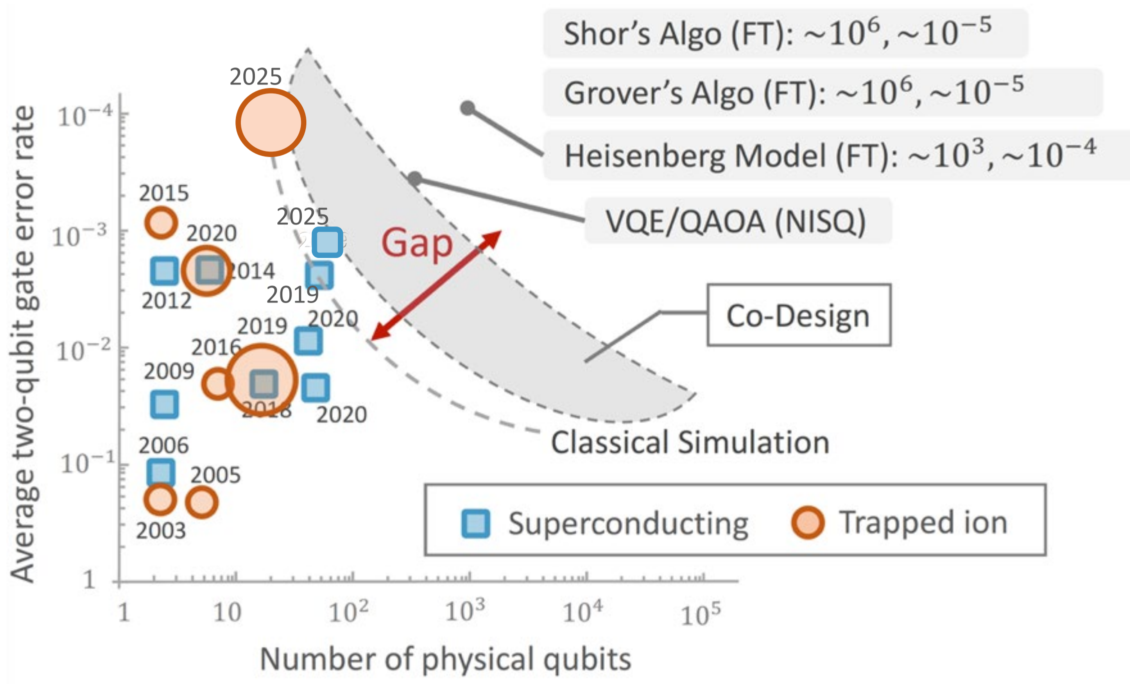
Infleqtion
100+ neutral atom qubits



Quantinuum
56 atomic ion qubits

The EPiQC Goal

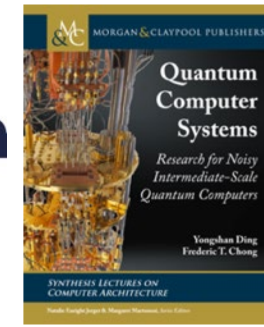
Co-design algorithms, software, and hardware to close the gap between algorithms and devices by 100-1000X, accelerating QC by 10-20 years.



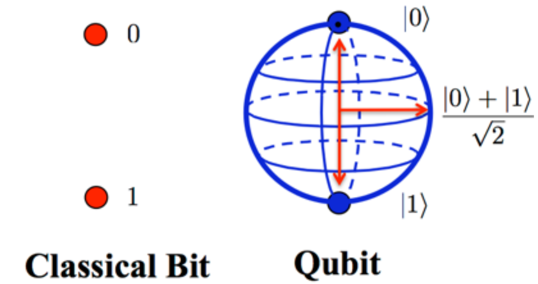
*Size of data point indicates connectivity; larger means denser connectivity.

EPiQC NSF Expedition (2018-2024)

- Many optimizations, each 2-10X, up to 10000X
- 150+ papers, 10 best paper awards
- 21 PhDs -> 7 faculty
- 1 startup
- 1 textbook, 5 EdX courses
- Techniques integrated into IBM QISKit, Google Cirq, Intel Quantum Compiler, Rigetti Pyquil, CQC TKET, ORNL XACC and QCOR



Quantum Bits (qubit)



- 1 qubit probabilistically represents 2 states

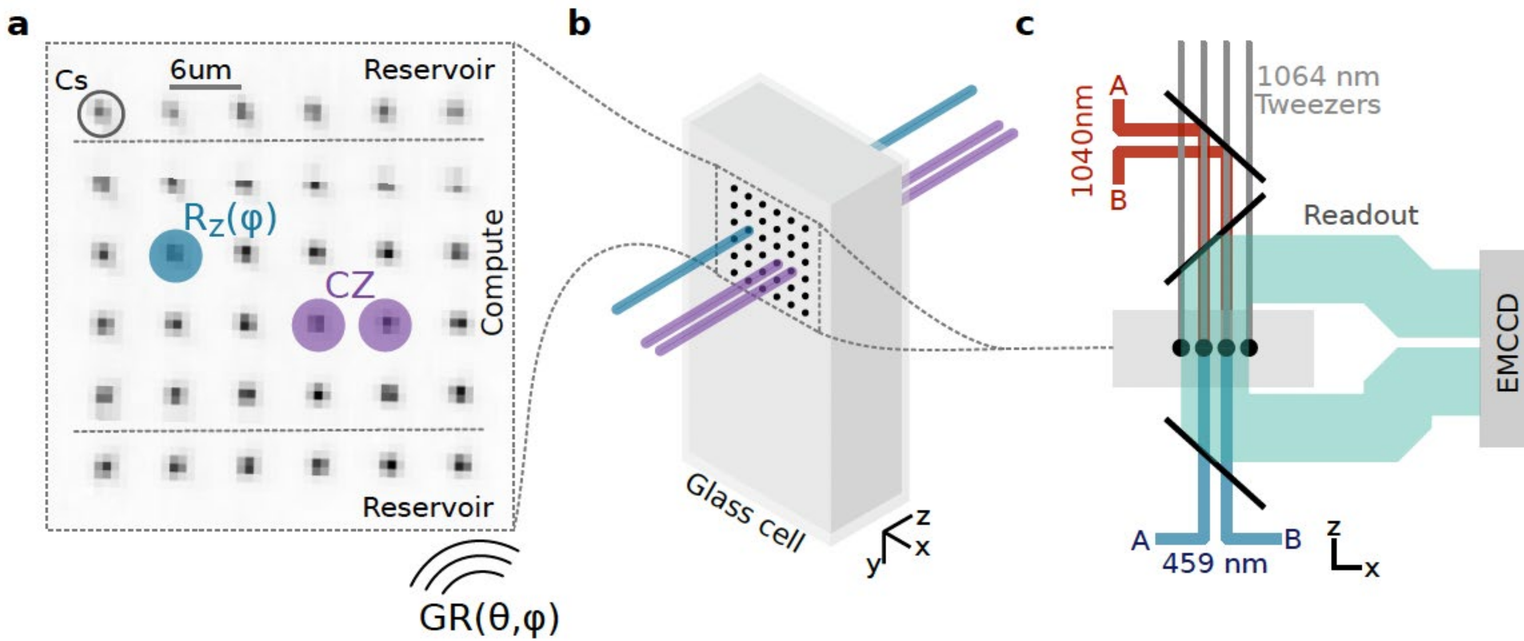
$$|a> = C_0|0> + C_1|1>$$

- Every additional qubit doubles # states

$$|ab> = C_{00}|00> + C_{01}|01> + C_{10}|10> + C_{11}|11>$$

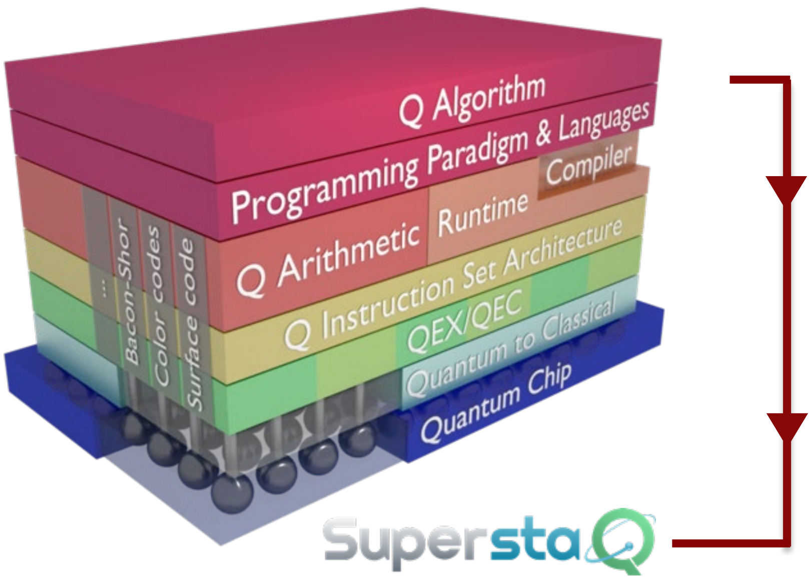
- "Parallelism" on an exponential number of states
  - But measurement collapses qubits to single classical values
  - Noise in computation and measurement

Neutral Atom Quantum Computer



[Radnaev+ 2024]
(Infleqtion)

Quantum Software: Please break abstraction layers!

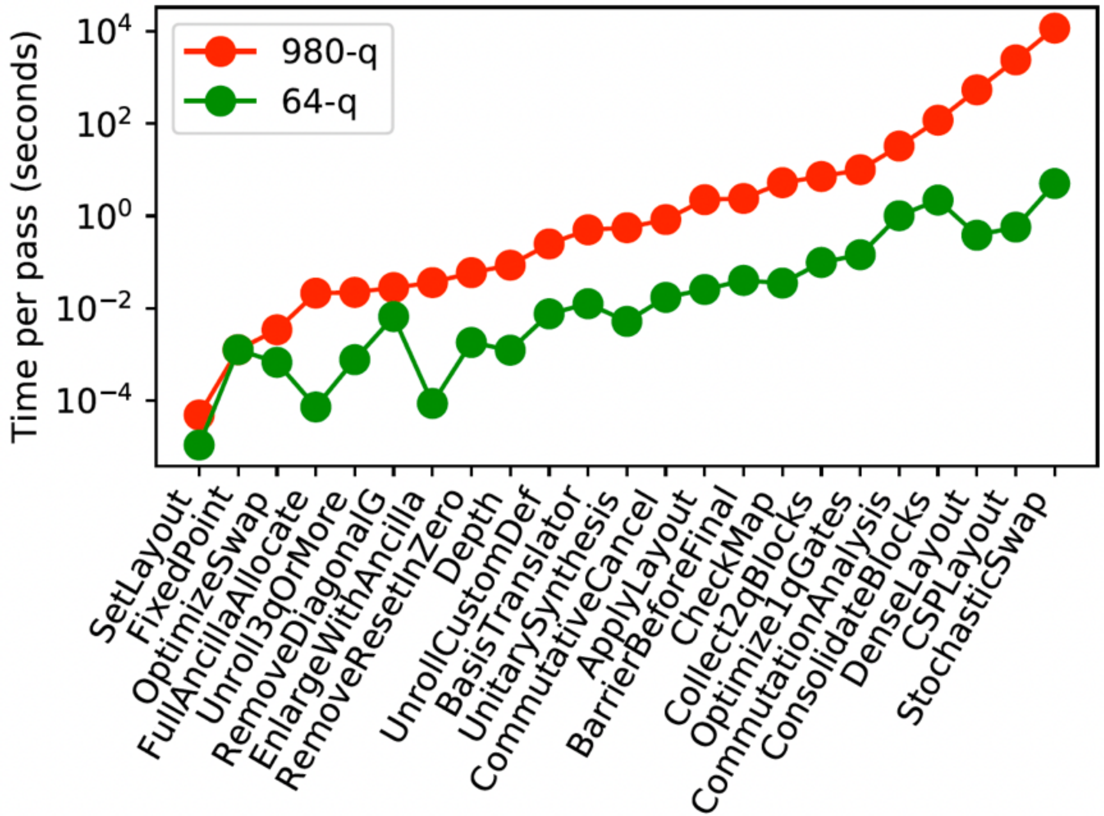


X. Fu et al. (2017)
arXiv:1708.07677

Shortcut

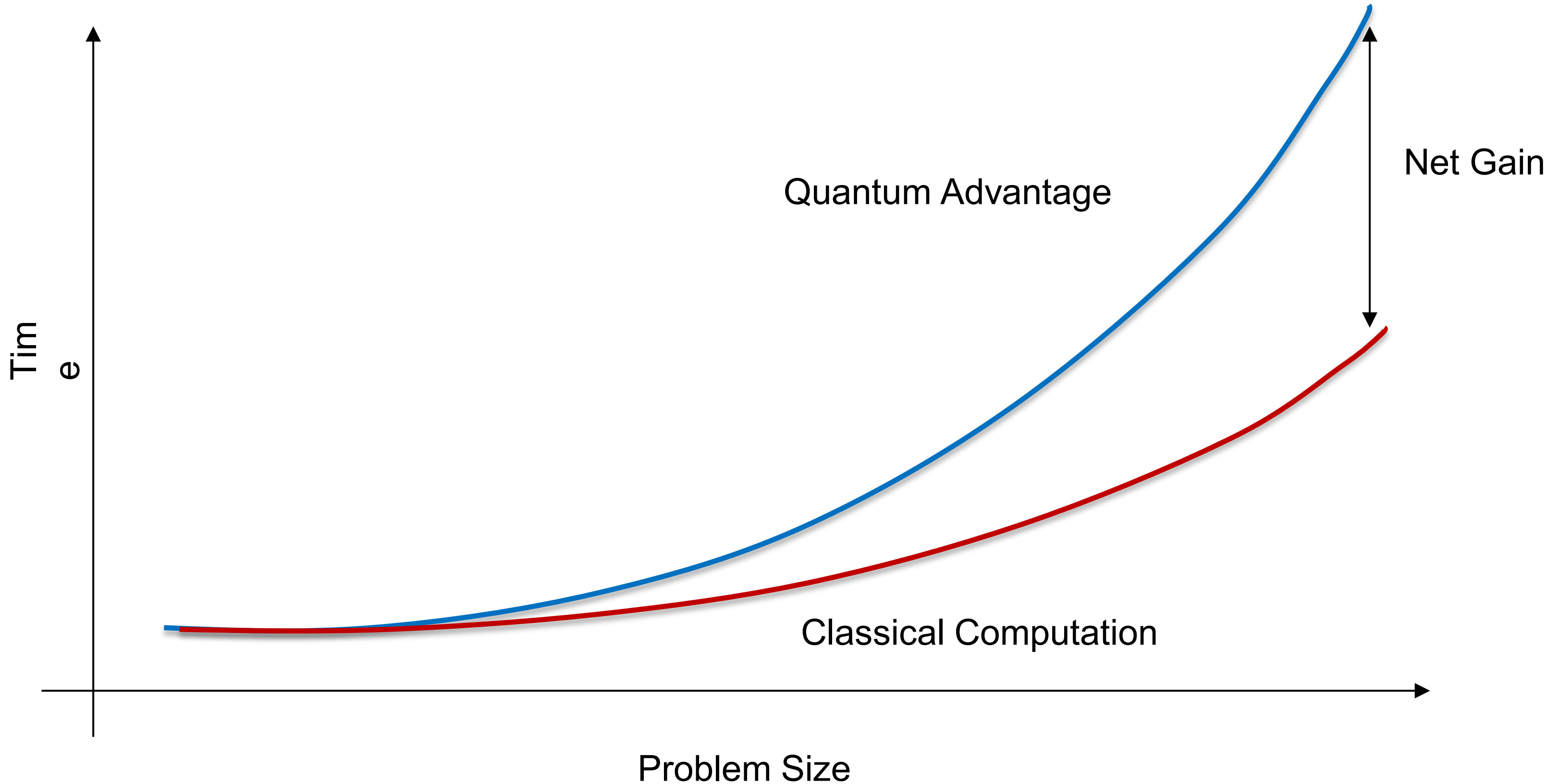
- Stack: rigid layers + interfaces
- Benefits
  - Taming complexity
- **Problems**
  - **Lost opportunities for optimization**
  - **QC stack + layers change**
- We should compile to hardware primitives. Physics first.

Scalability vs Deep Optimization



Gokul Subramanian Ravi, Kaitlin N. Smith, Pranav Gokhale, Frederic T. Chong:
Quantum Computing in the Cloud: Analyzing job and machine characteristics. IISWC 2021: 39-50

Hybrid Quantum-Classical Computing: Contest of Exponentials



The Secret Menu of Quantum Hardware


TACO BELL
CHIPOTLE
MEXICAN GRILL
Enchirito
Buffalo Chicken
Quesarito

MENU

Item	Price
1. NOT	0.1% error
2. RZ(θ)	0.01% error
3. CNOT	Market price

No substitutions permitted.

SECRET HARDWARE PRIMITIVE MENU

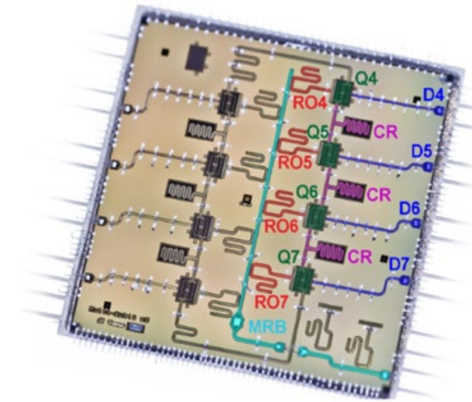
4. CR(θ)	0.4% error
5. Parametric $e^{-i\,\theta\,ZZ}$	~1.0% * θ/π
6. |1> -> |2> qutrit transition	0.3% error

Additional options available—ask your friendly physicist.

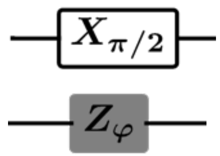
1. Swap Gates and ECA

- *Optimized SWAP networks with equivalent circuit averaging for QAOA*, Akel Hashim , Rich Rines, Victory Omole, Ravi K. Naik, John Mark Kreikebaum, David I. Santiago, Frederic T. Chong, Irfan Siddiqi, and Pranav Gokhale. Physical Review Research 4, 033028

Advanced Quantum Testbed (AQT)



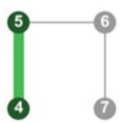
Single-qubit gates:



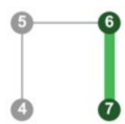
(Rabi, 30ns)

(virtual)

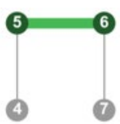
Entangling gates:



$CZ_{4,5}$: 200ns
$CS^\dagger_{4,5}$: 150ns

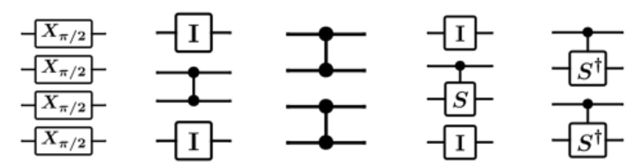


$CZ_{6,7}$: 200ns
$CS^\dagger_{6,7}$: 150ns



$CZ_{5,6}$: 200ns
$CS_{5,6}$: 150ns

Parallel gate performance:

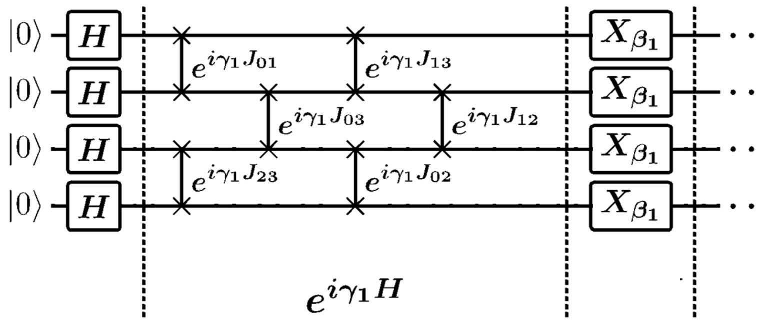


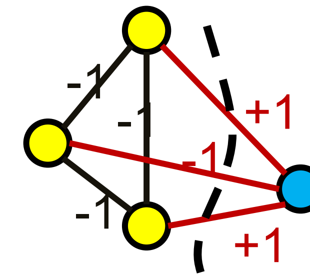
gate time	30ns	200ns	200ns	150ns	150ns
ϵ_T (10^{-2})	0.19(8)	1.09(9)	2.3(1)	0.68(9)	1.07(9)

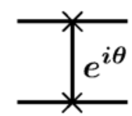
Photo: A. Hashim, et al., *Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor*, arXiv:2010.00215 (2020).

Application: QAOA for Weighted Max-Cut

Problem: Max-cut on a fully connected graph, edge weights ∈ {-1, 1}







$$= ZZ\text{-SWAP}(\theta)$$

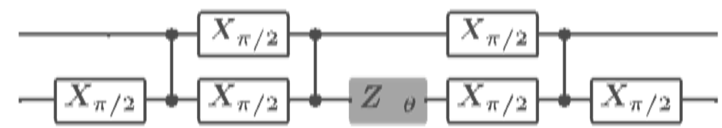
$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & e^{i\theta} & 0 \\ 0 & e^{i\theta} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

⇒ *Efficiently mapped to linear topology*

Equivalent Circuit Applications

Optimized Scheduling:

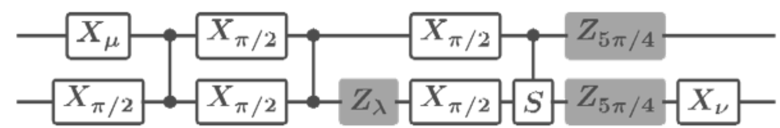
- *Iteratively select decompositions to maximize prior gate cancellation*



32 unique CZ-CZ-CZ decompositions

Equivalent Circuit Average (ECA):

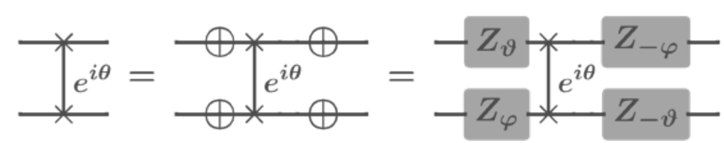
- *Randomly select decompositions to generate M logically equivalent circuits, to mitigate coherent error*



64 unique CZ-CZ-CS decompositions

ECA + Optimized scheduling:

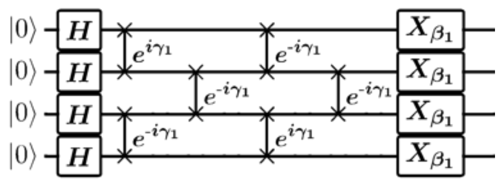
- *Only randomize over decompositions minimizing critical path depth*



2 commutation rules

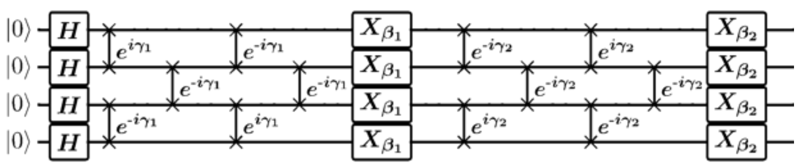
ECA + Optimized Scheduling on the AQT

QAOA, $p = 1$:

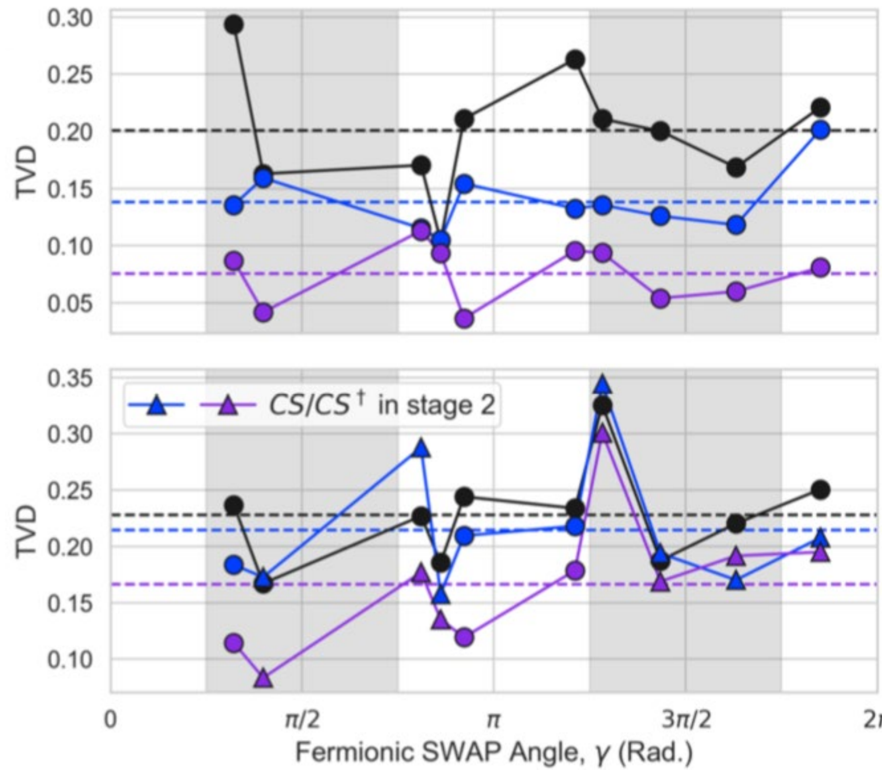


- Gate/schedule opt.: 30% error reduction
- ECA-OPT (M=20): **60% error reduction (TVD)**

QAOA, $p = 2$:



- Gate/schedule opt.: 7% error reduction
- ECA-OPT (M=20): **30% error reduction**

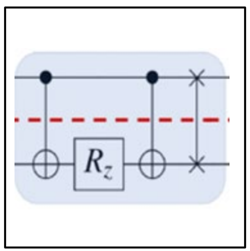


2. Direct-to-Pulse Compilation

- *Optimized Compilation of Aggregated Instructions for Realistic Quantum Computers,* Yunong Shi, Nelson Leung, Pranav Gokhale, Zane Rossi, David I. Schuster, Henry Hoffman, Frederic T. Chong, International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS '19)
- *Partial Compilation of Variational Algorithms for Noisy Intermediate-Scale Quantum Machines*, Pranav Gokhale, Yongshan Ding, Thomas Propson, Christopher Winkler, Nelson Leung, Yunong Shi, David I. Schuster, Henry Hoffmann, Frederic T. Chong, International Symposium on Microarchitecture (MICRO ’19)
- *Quantum Compilation for NISQ Algorithms with Pulse-Backed Augmented Basis Gates*, Pranav Gokhale, Ali Javadi-Abhari, Nathan Earnest, Yunong Shi, and Frederic T. Chong. The International Symposium on Microarchitecture (MICRO ‘20)

Gates vs Pulses

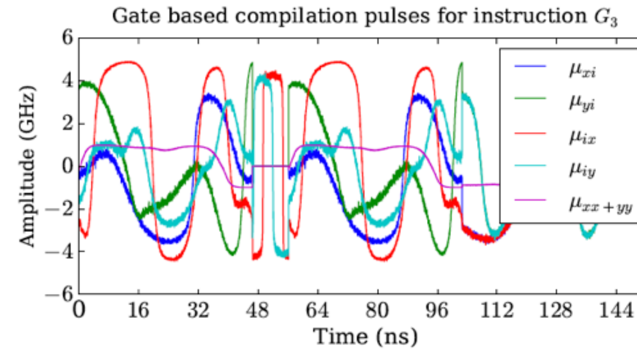
Quantum Function



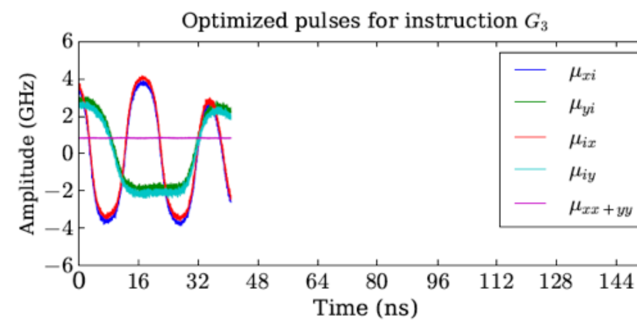
Quantum Assembly

CNOT q1, q2
RZ t1, q2
CNOT q1, q2
SWAP q1, q2

Direct to Pulse



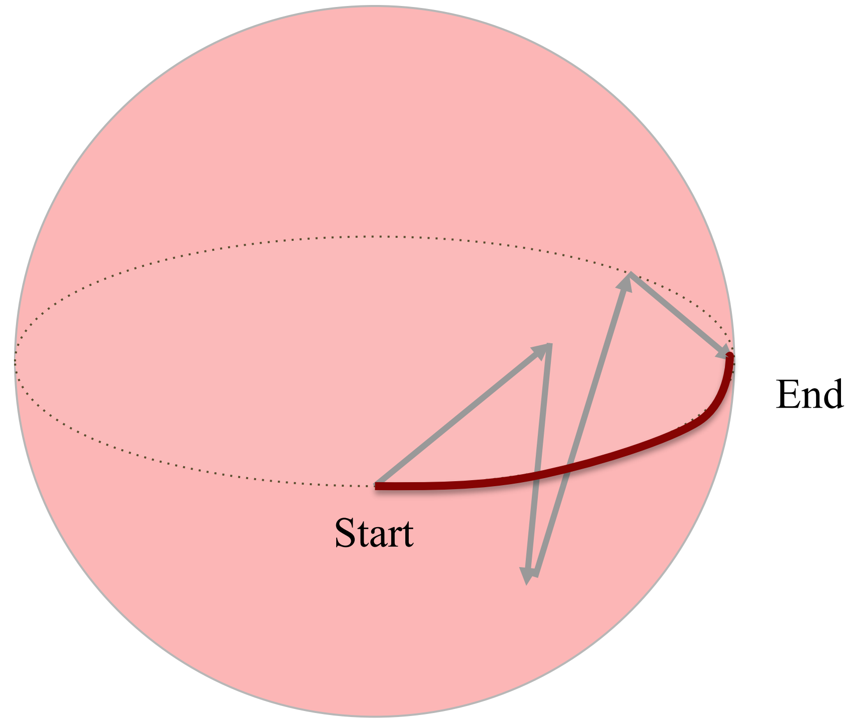
(c)



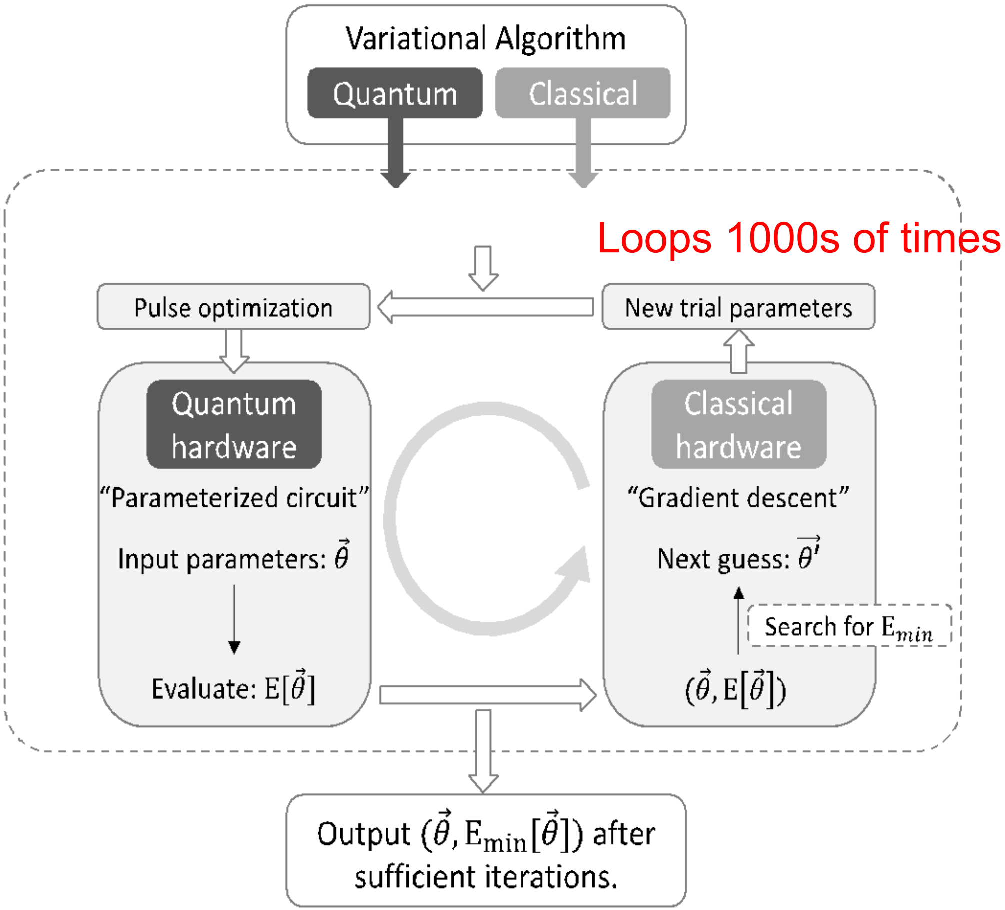
(d)

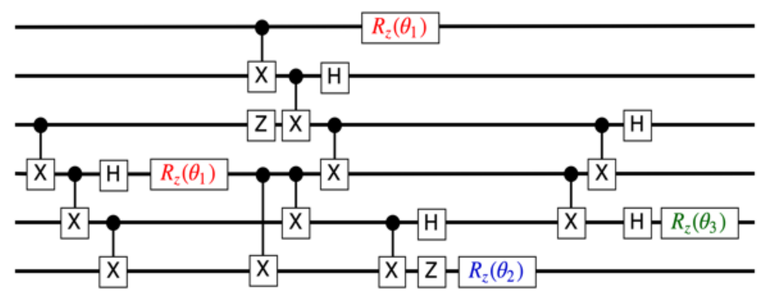
Direct-to-Pulse Results

- **2X to 10X faster**
- But it can take hours to compile a program before we can run it
- This is a problem for an important class of algorithms that alternates between classical and quantum computing

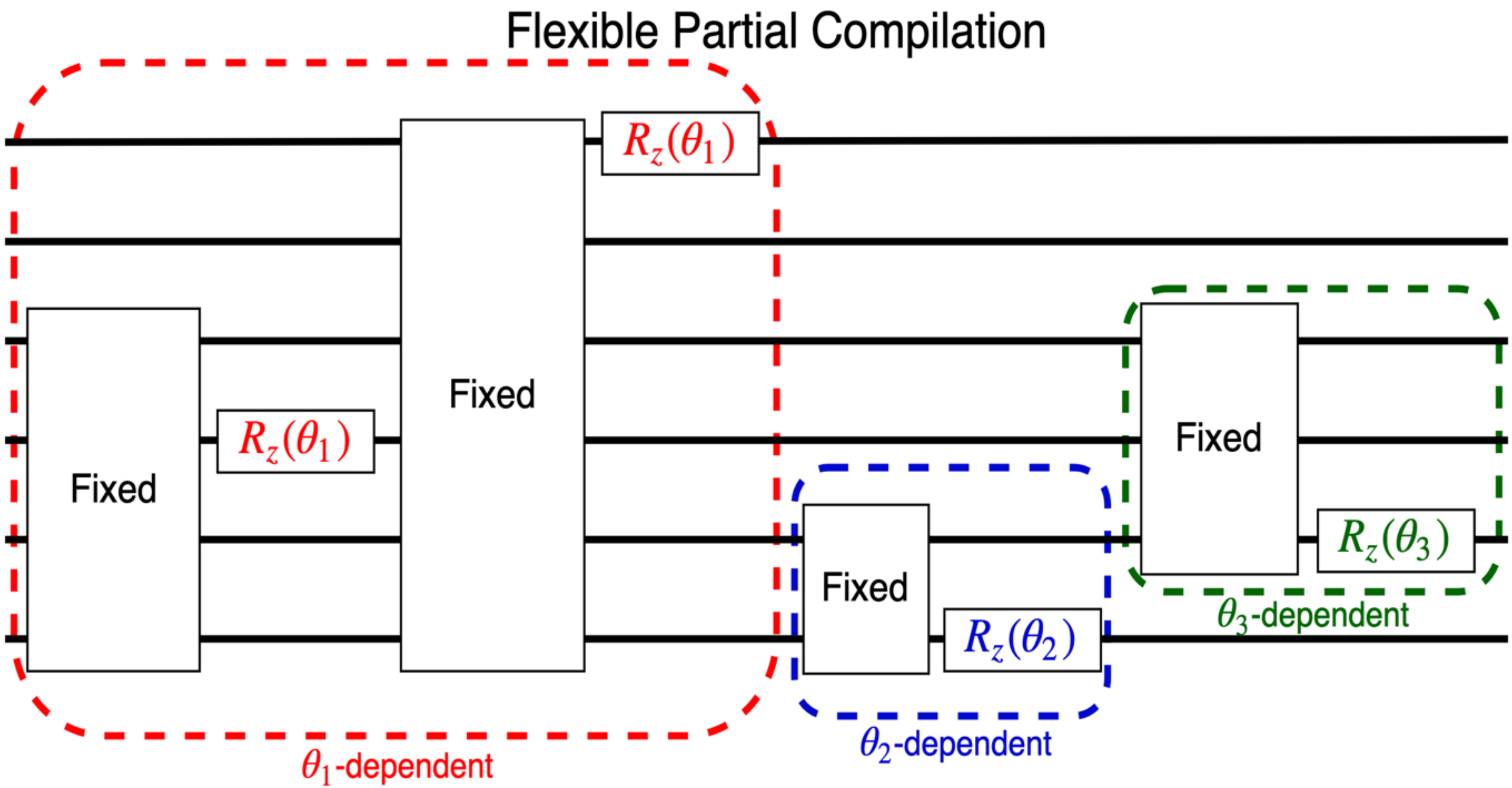


Variational Quantum Algorithms





Flexible Partial Compilation



Partial Compilation Results

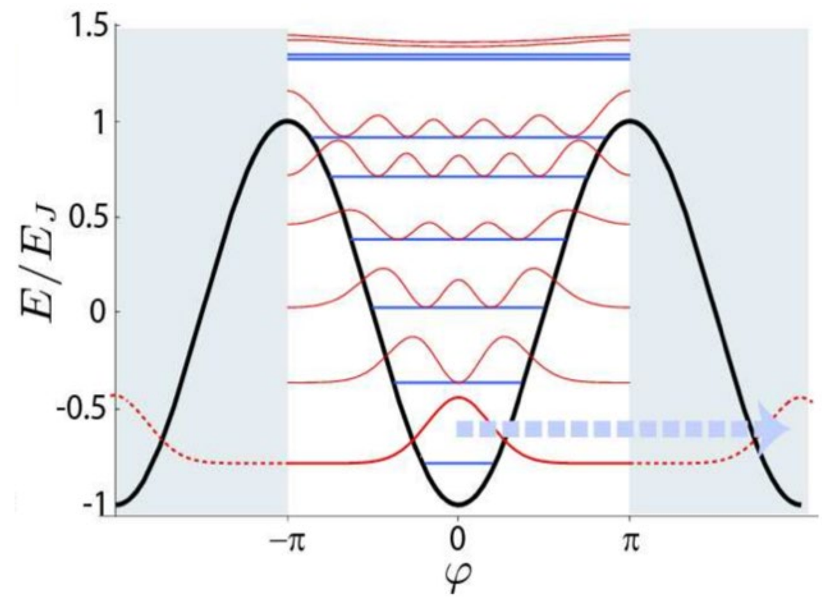
- **2x** pulse speedups
- **10-80x** faster compilation than previous method
- **2 patents** pending
- *The key was to break the abstraction of machine instructions and target pulses*

3. Qutrits instead of Ancilla

- *Asymptotic Improvements to Quantum Circuits via Qutrits*, Pranav Gokhale, Jonathan Baker, Casey Duckering, Natalie Brown, Ken Brown, and Frederic T. Chong. International Symposium on Computer Architecture (ISCA ‘19) (**QIP Best Poster, 3 of 480**)
- *Efficient Quantum Circuit Decompositions via Intermediate Qudits*, Jonathan M. Baker, Casey Duckering, Frederic T. Chong. International Symposium on Multi-Valued Logic (ISMVL’20)
- *Extending the Frontier of Quantum Computers with Qutrits,* P. Gokhale, J.M. Baker, C. Duckering, N.C. Brown, K.R. Brown, F.T.Chong. **IEEE Micro Top Picks in Computer Architecture** (2020)
- *Improved Quantum Circuits via Intermediate Qutrits*, Jonathan Baker, Pranav Gokhale, Casey Duckering, Natalie Brown, Ken Brown, and Frederic T. Chong. ACM Transactions on Quantum Computing (2020).

Qutrits versus Qubits

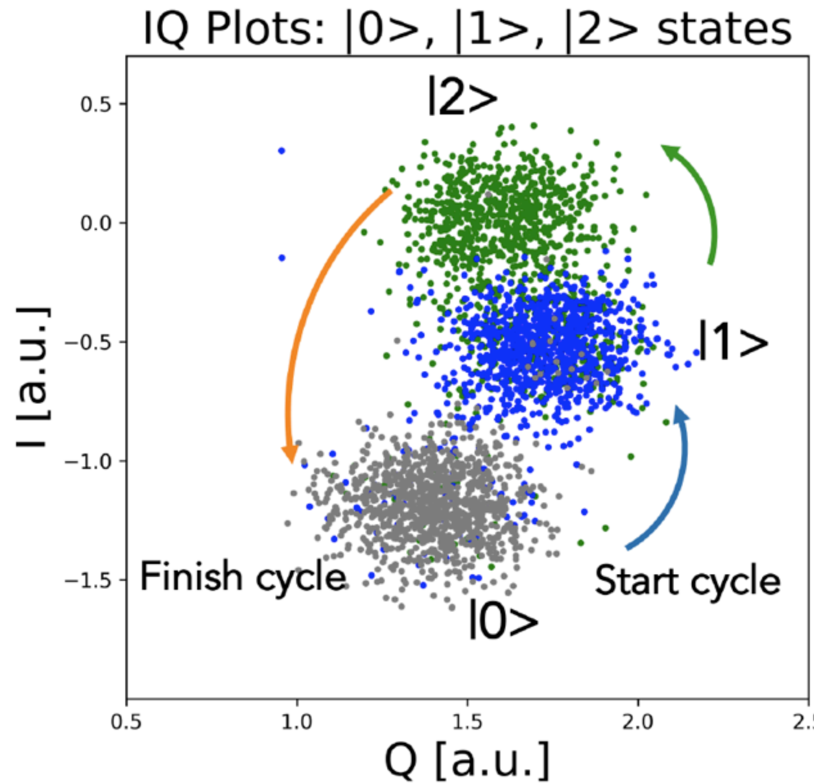
- Store 3 values instead of 2 in each hardware device
- 3-level logic is not new, but makes more sense for quantum devices
- Especially useful for programs that need some extra quantum bits to be more efficient (some temporary space)



[Koch 07]

Qutrit Results

- Fewer devices needed
  - Up to 70X reduction for some programs
- A lot of interest from hardware platforms
  - IBM OpenPulse experiment
- Also won the "Top Picks" best papers for 2019 award
- *The key was to break the binary abstraction*



[Gokhale+ Micro20]

4. Ququart Gates and Compilation

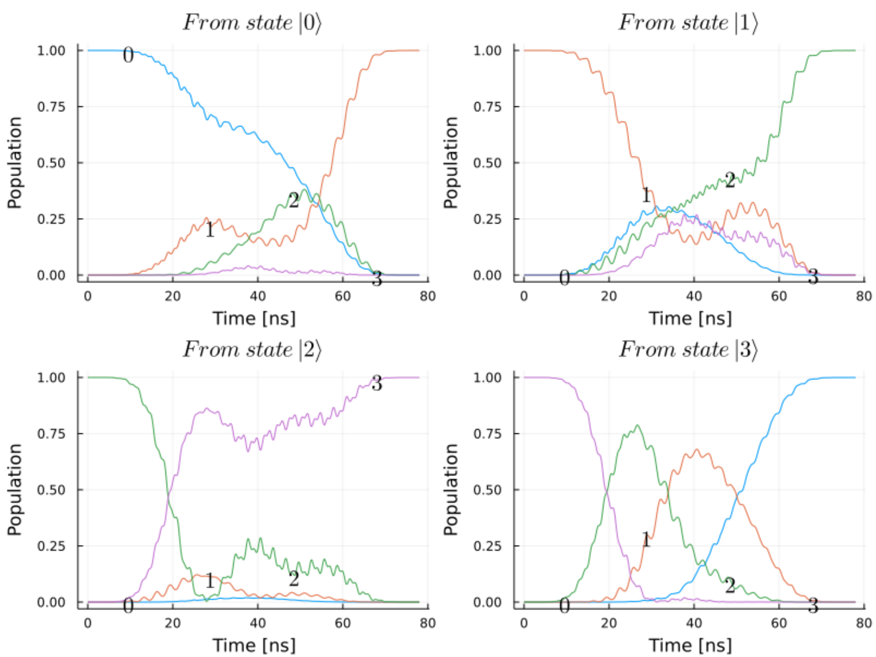
- *Time-Efficient Qudit Gates through Incremental Pulse Re-seeding*, Lennart Maximilian Seifert, Jason Chadwick, Andrew Litteken, Frederic T. Chong and Jonathan M. Baker. IEEE International Conference on Quantum Computing and Engineering, 2022.
- *Qompress: Efficient Compilation for Ququarts Exploiting Partial and Mixed Radix Operations for Communication Reduction*, A. Litteken, L. Seifert, J. Chadwick, N. Nottingham, F. Chong, and J. Baker. International Symposium on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2023.
- *Dancing the Quantum Waltz: Compiling Three-Qubit Gates on Four-Level Architectures*, Andrew Litteken, Lennart Maximilian Seifert, Jason Chadwick, Natalia Nottingham, Tanay, Ziqian Li, David Schuster, Frederic T. Chong, and Jonathan M Baker. International Symposium on Computer Architecture (ISCA), 2023.

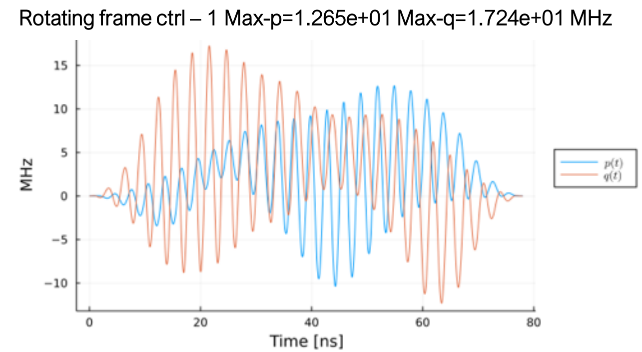
Motivation

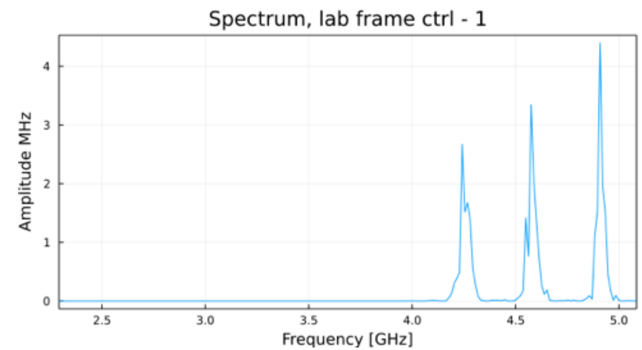
- Ququarts would give more general compression/decompression and fast qubit operations inside a ququart
- Theory predicts quadratic scaling of gate duration with radix
- Maybe pulse implementation would be better…

Quantum optimal control for qudits

- Example: QOC for X_4, $F = 99.9\%$

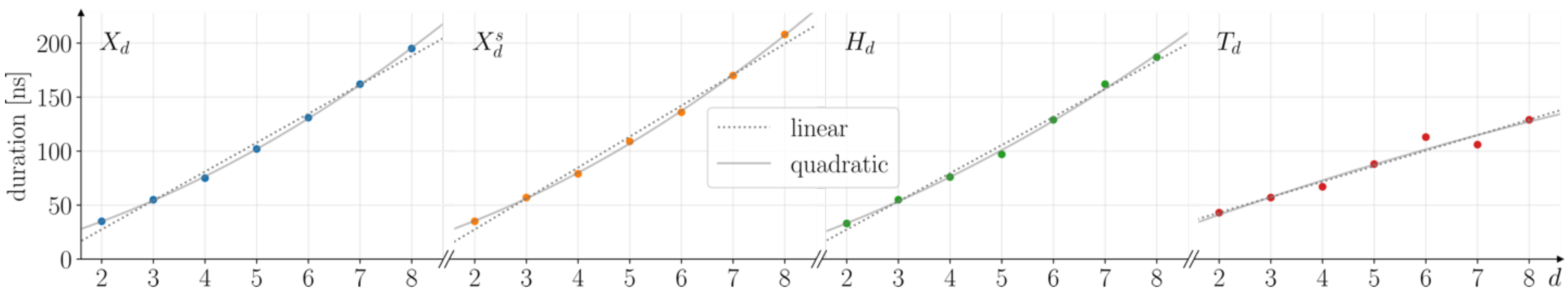






Pulse duration minimization using IPR

- Time-optimized pulse duration scaling: Single-qudit gates
  - Observe near-linear scaling up to $d = 8$ qudits

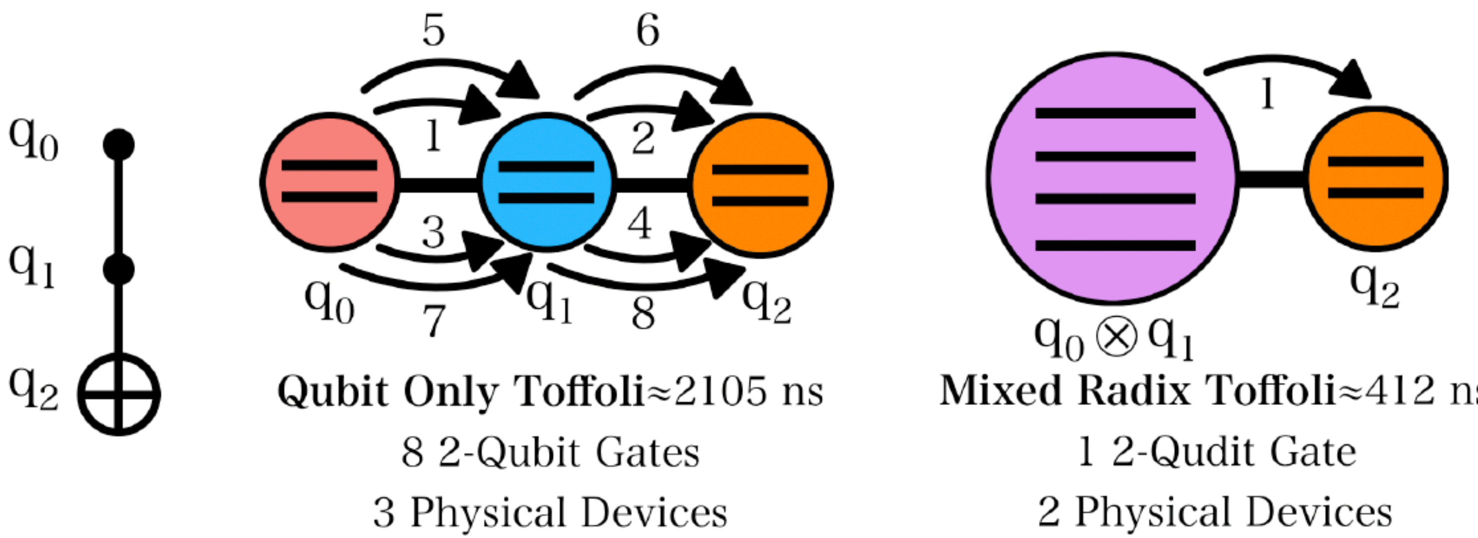


Qudit Conclusion

- Near-linear scaling in practical qudit regime $d \leq 8$
- *Enables upcoming compiler optimizations allowing 2X device savings with comparable fidelity*

Qudit Conclusion

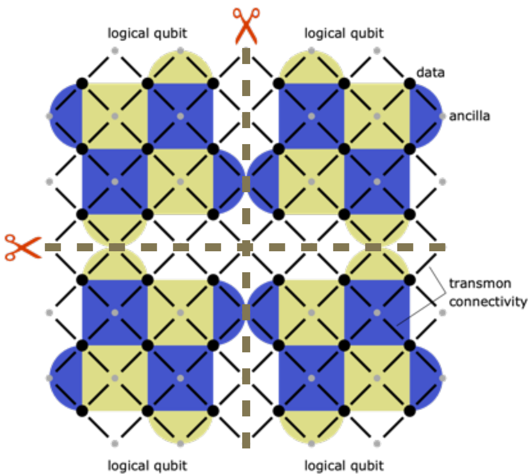
- Near-linear scaling in practical qudit regime $d \leq 8$
- *Enables compiler optimizations allowing 2X device savings with comparable fidelity (Qompress ASPLOS23)*
- *Enables ququart-qubit Toffoli (Quantum Waltz ISCA23)*

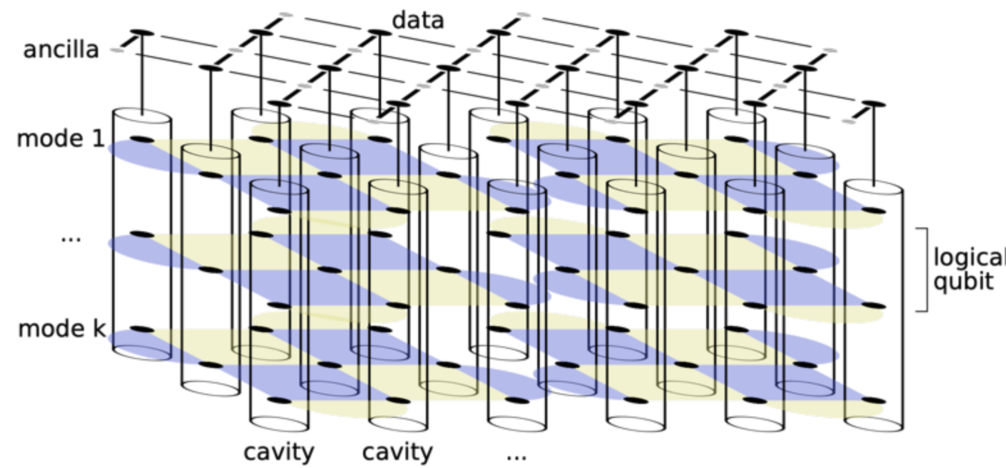


5. Virtualized Logical Qubits

- “Virtualized Logical Qubits: A 2.5D Architecture for Error-Corrected Quantum Computing,” Casey Duckering, Jonathan Baker, David Schuster, Frederic T. Chong. Micro 2020. **Micro Top Pick 2021.**

Virtualized Logical Qubits





Virtualize logical qubits by storing them in memory layers

Physical 2D address, virtual 2D+mode index

Load to apply error correction and to compute

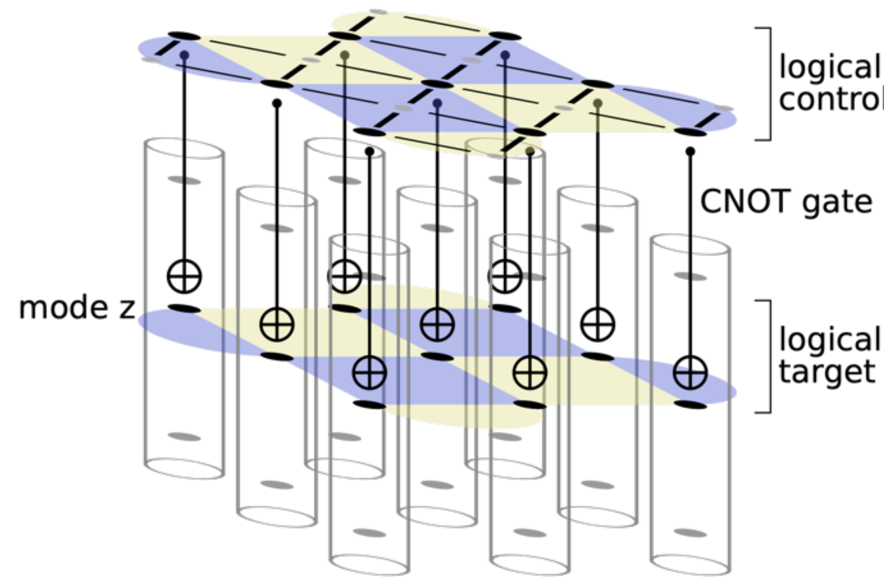
Transversal CNOT

Not possible in 2D

6x faster

No measurements

Verified with process tomography



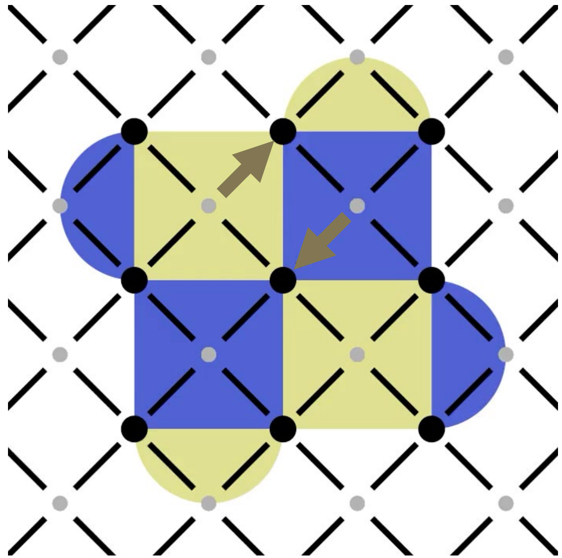
Compact Version

Use each transmon for both data and ancilla qubits

Same hardware connectivity

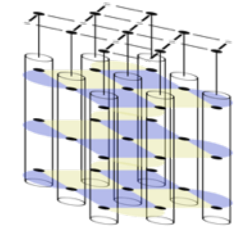
Saves 2x transmons

Slower and requires more memory accesses



With adjusted coordinates

VLQ Summary



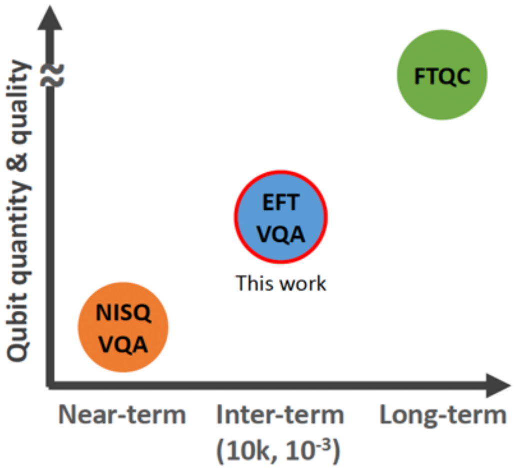
- We virtualize logical qubits with memory separate but local to computation
- 10x reduction in transmon qubits and control hardware
- Minimum proof of concept for 10 logical qubits requires only 11 transmons and 9 cavities.
- *The key was to go beyond 2 dimensions and match the computation to the architecture*

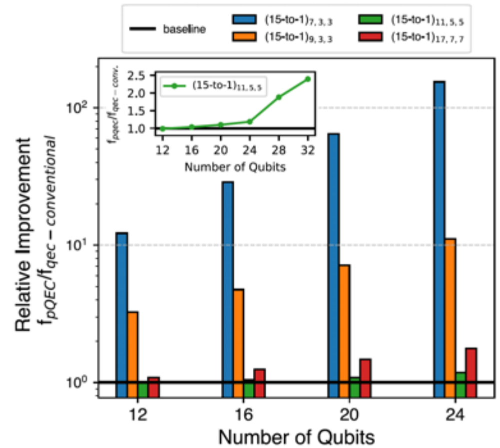
6. Interleaved Logical Qubits in Atom Arrays

3X faster using transversal gates

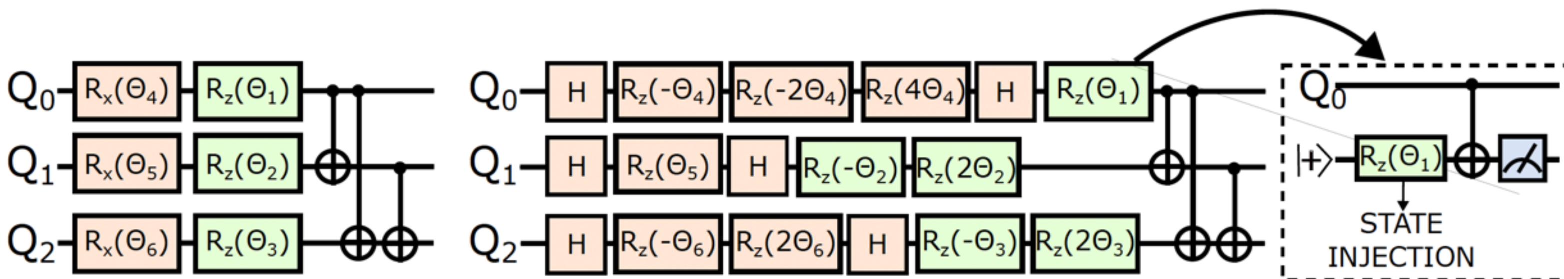
J. Viszlai, S. F. Lin, S. Dangwal, C. Bradley, V. Ramesh, J. M. Baker, H. Bernien, and F. T. Chong, “Interleaved Logical Qubits in Atom Arrays,” presented at the 2025 IEEE International Symposium on High-Performance Computer Architecture (HPCA), March. 2025.

7. Partial Error Correction





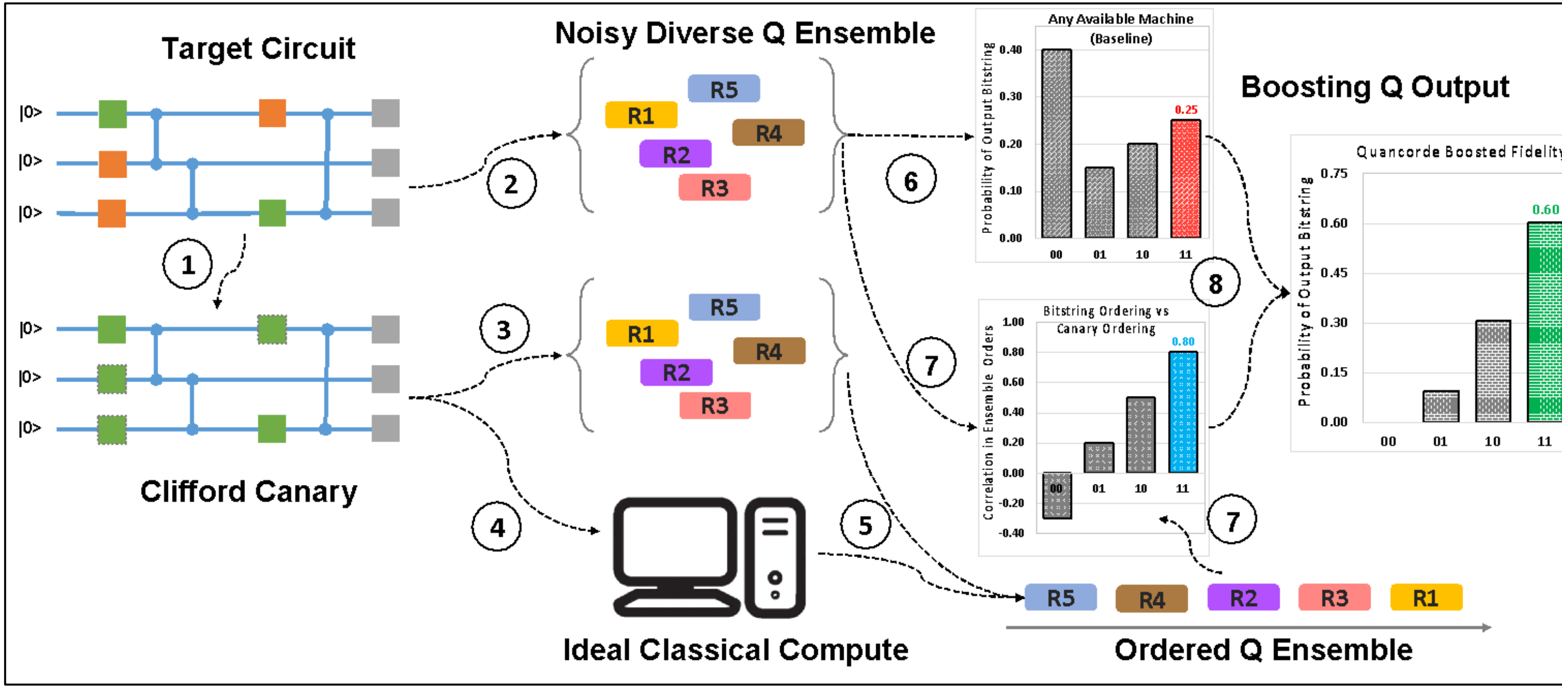
>25X higher fidelity with physical rotations



"Variational Quantum Algorithms in the Era of Early Fault Tolerance," S. Dangwal, S. Vittal, L. Seifert, F. Chong, and G. Ravi. International Symposium on Computer Architecture (ISCA 2025)

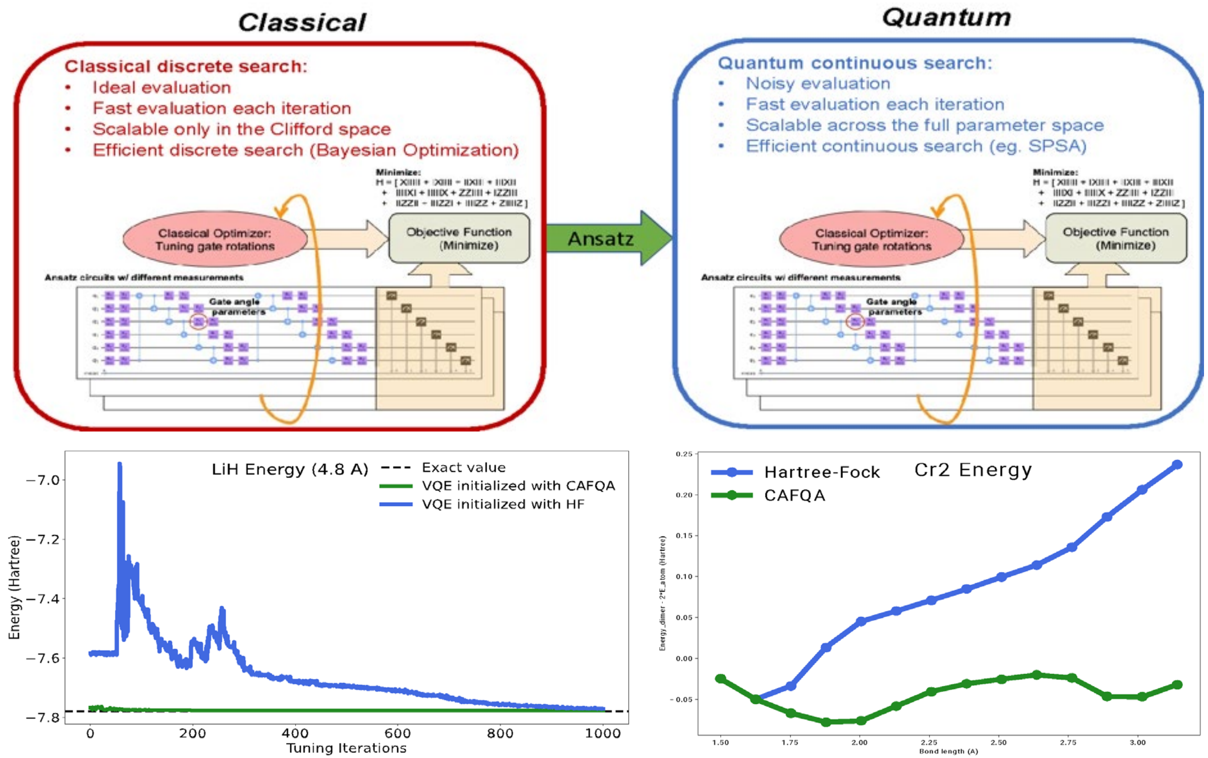
Hybrid Quantum-Classical Computations

8. Quancorde: Boosting fidelity with Quantum Canary Ordered Diverse Ensembles



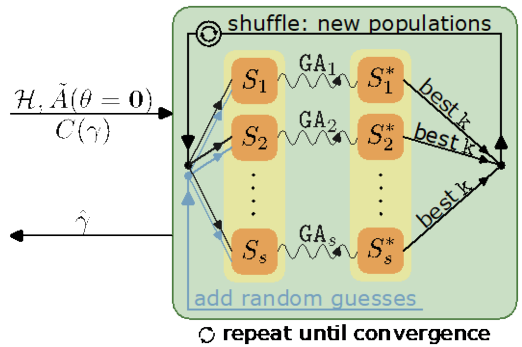
Gokul Subramanian Ravi, Jonathan Baker, Kaitlin Smith, Nathan Earnest, Ali Javadi-Abhari, Frederic Chong. ICRC, December 2022 (Selected Highlight Paper)

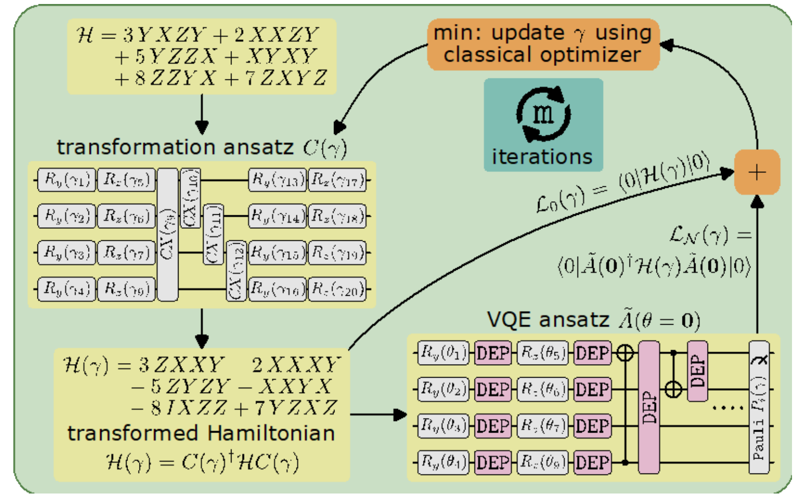
9. CAFQA: A classical simulation bootstrap for variational quantum algorithms

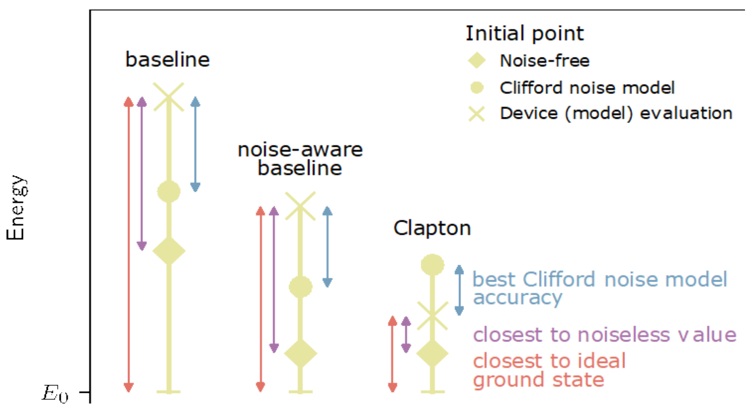


Gokul Subramanian Ravi, Pranav Gokhale, Yi Ding, William M. Kirby, Kaitlin N. Smith, Jonathan M. Baker, Peter J. Love, Henry Hoffmann, Kenneth R. Brown, Frederic T. Chong. ASPLOS, March 2023.

10. Clapton: Clifford-Assisted Problem Transformation for Error Mitigation in Variational Quantum Algorithms (VQAs)







L. M. Seifert, S. Dangwal, F. T. Chong, and G. S. Ravi, “Clapton: Clifford-Assisted Problem Transformation for Error Mitigation in Variational Quantum Algorithms,” ASPLOS’25.

Inflection

THE UNIVERSITY OF CHICAGO

Massachusetts Institute of Technology

leap + BIO

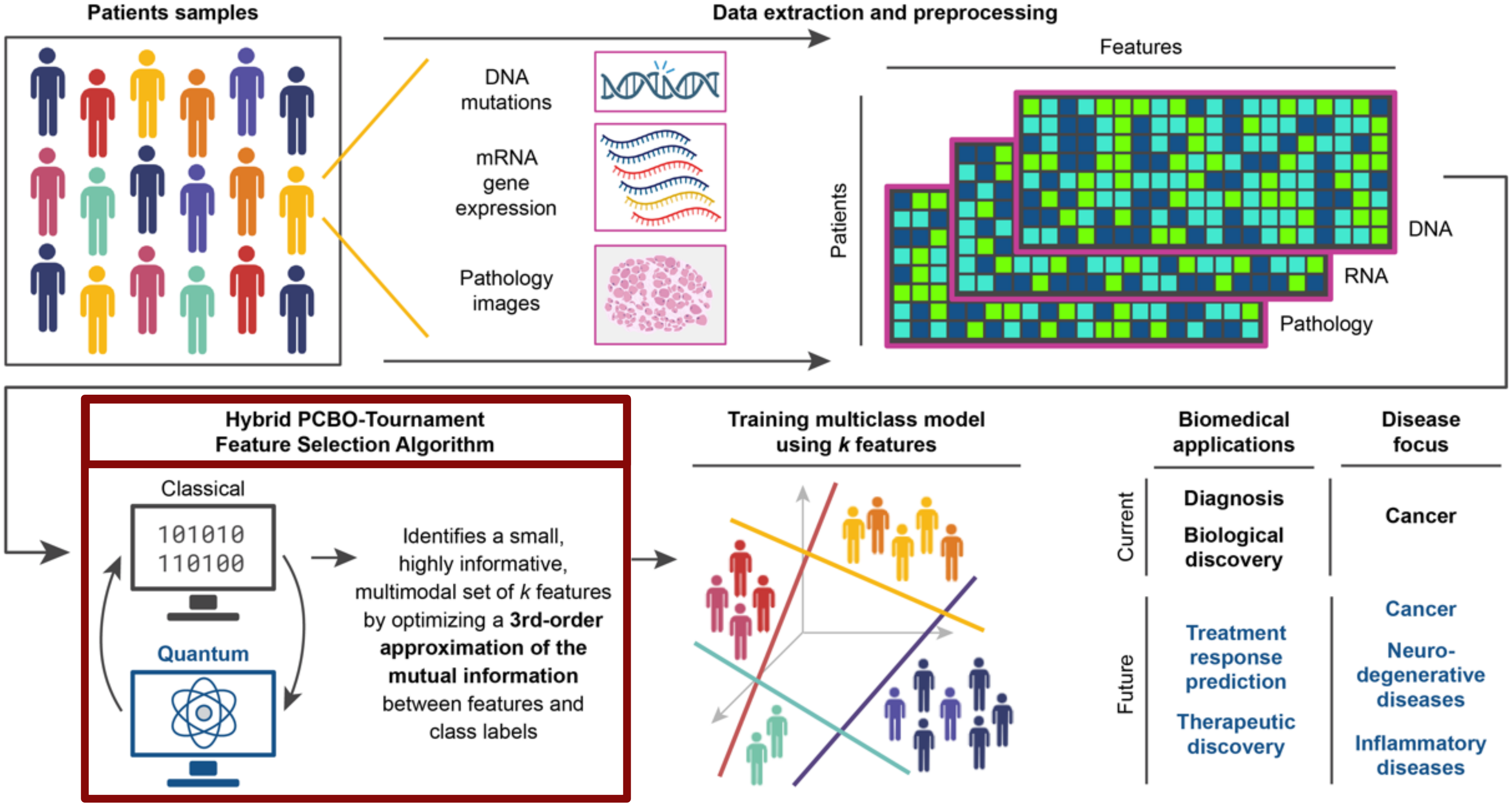
Quantum Biomarker Algorithms for Multimodal Cancer Data

Fred Chong and Teague Tomesh, Infleqtion

Samantha Riesenfeld and Alex Pearson, UChicago

Aram Harrow, MIT

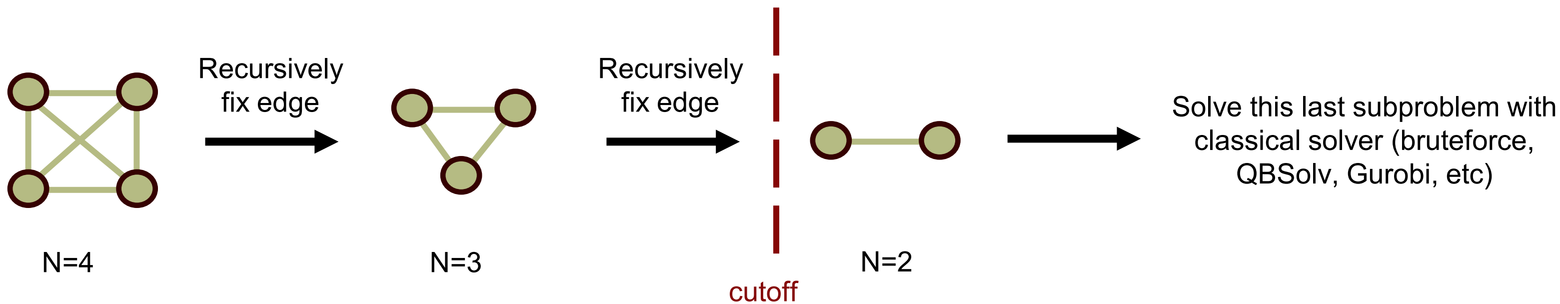
Multimodal Feature Selection



Teague

Hyper-RQAOA

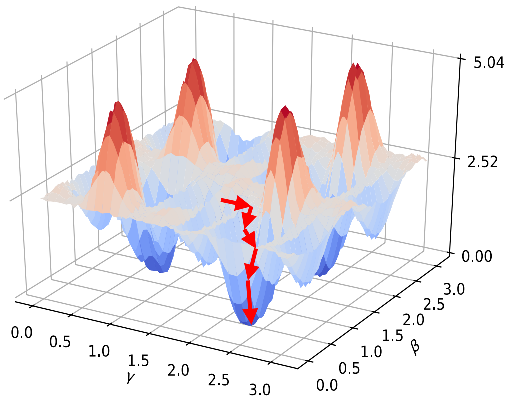
$$H_{\mathrm{PCBO}} = \sum_{l} \alpha_l P_l = \sum_{i} \underbrace{C(i) Z_i}_{\text{“}1^{\text{st}}\text{-order edge”}} + \sum_{i<j} \underbrace{C(i,j) Z_i Z_j}_{\text{“}2^{\text{nd}}\text{-order edge”}} + \sum_{i<j<k} \underbrace{C(i,j,k) Z_i Z_j Z_k}_{\text{“}3^{\text{rd}}\text{-order edge”}}$$



Aram

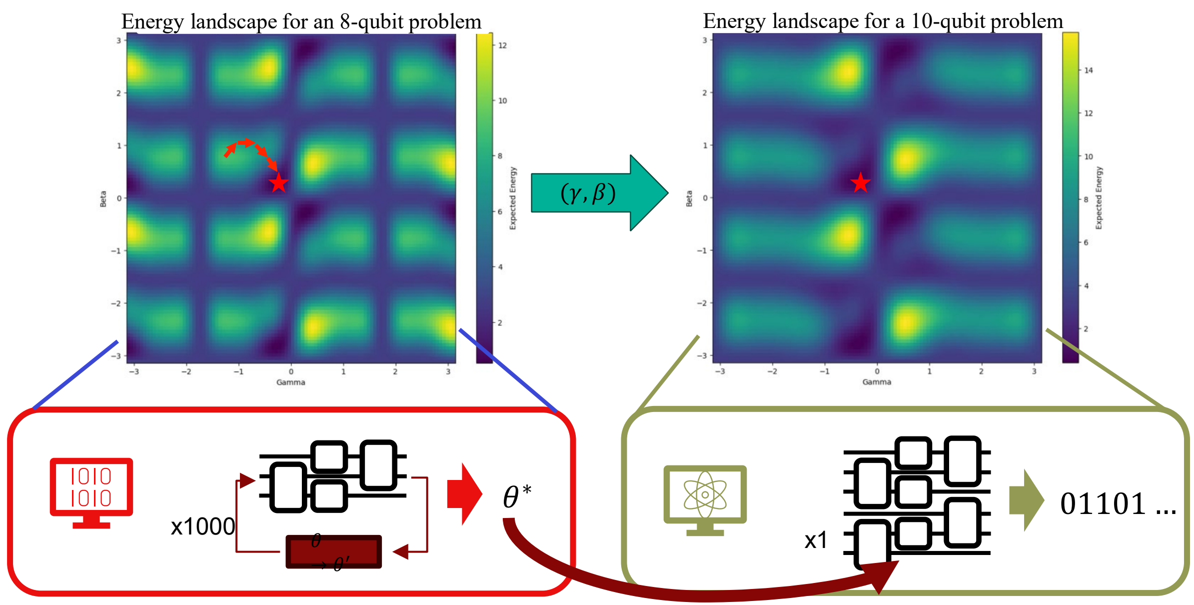
First demonstration of parameter transfer for RQAOA and real-world problems

Variational optimization is *expensive*



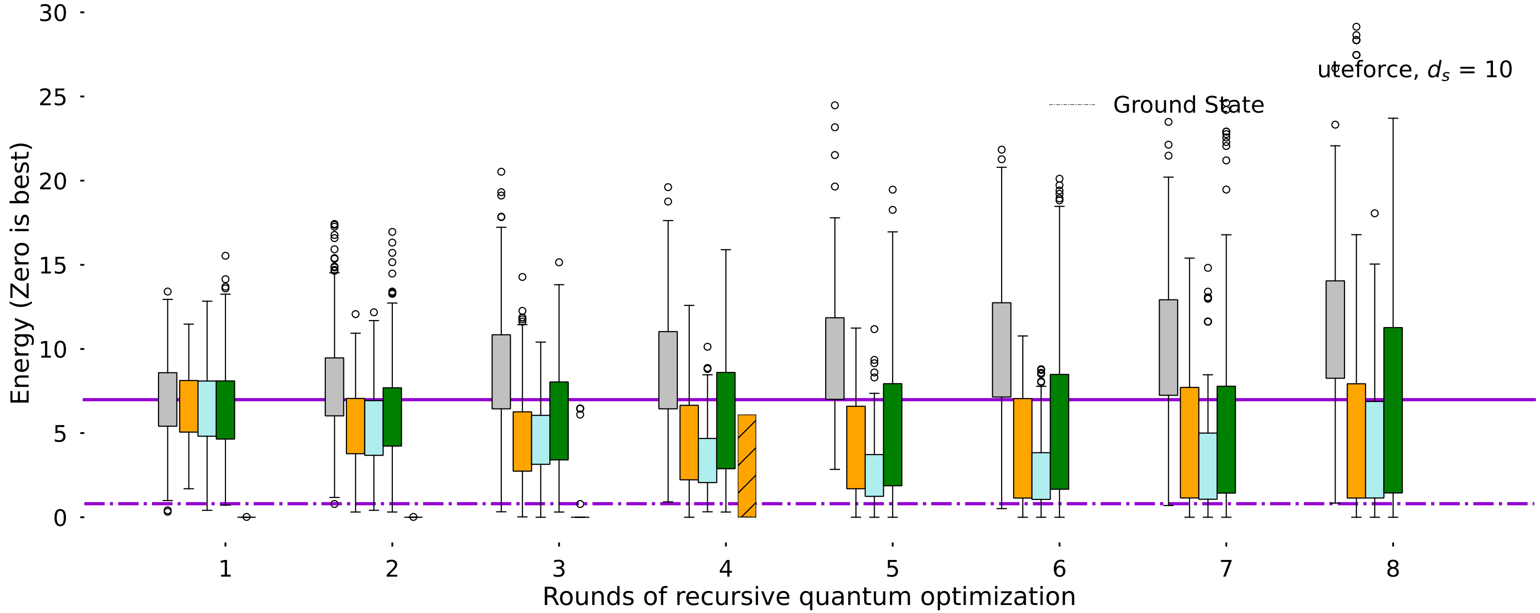
- Many quantum circuit evaluations required to traverse the loss landscape
  - Limits the scalability of our classical simulations and hardware experiments

Because QAOA exhibits the *parameter concentration*[1] property, we can *transfer* parameters between problems

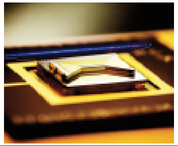
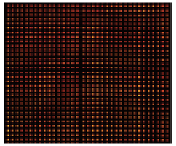
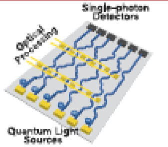
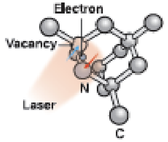
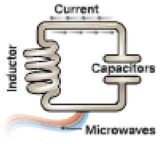
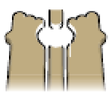
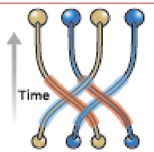


[1] Brandao, Fernando GSL, Michael Broughton, Edward Farhi, Sam Gutmann, and Hartmut Neven. "For fixed control parameters the quantum approximate optimization algorithm's objective function value concentrates for typical instances." *arXiv preprint arXiv:1812.04170* (2018).

A Hybrid Path to Quantum Advantage



- We envision this hybrid approach as a viable pathway to quantum advantage
- QBSolv struggles at 20 qubits -> Gurobi struggles at 50-100 qubits

		Qubit Coherence Time (sec)	Two-qubit Gate Fidelity	Qubits Connected	Companies	Pros	Cons
Natural Qubits							
	**Trapped Ions** Electrically charged atoms, or ions, are held in place with electric fields. Qubits are stored in electronic states. Ions are pushed with laser beams to allow the qubits to interact.	>1000	99.9%	High	IonQ, Quantinuum, AQT Oxford Ionics, Universal Quantum	Very stable. Highest achieved gate fidelities.	Slow operation. Many lasers are needed.
	**Neutral Atoms** Neutral atoms, like ions, store qubits within electronic states. Laser activates the electrons to create interaction between qubits.	1	99.5%	Very high; low individual control	Infleqtion, Atom Computing, QuEra, Pasqal, Planqc, M²	Many qubits, 2D and maybe 3D.	Hard to program and control individual qubits; prone to noise.
	**Photonics** Photonic qubits are sent through a maze of optical channels on a chip to interact. At the end of the maze, the distribution of photons is measured as output.	–	–	–	PsiQuantum, Xanadu	Linear optical gates, integrated on-chip.	Each program requires its own chip with unique optical channels. No memory.
	**Diamond Vacancies** A nitrogen atom and a vacancy add an electron to a diamond lattice. Its quantum spin state, along with those of nearby carbon nuclei, can be controlled with light.	10	99.2%	Low	Quantum Diamond Technologies, Quantum Brilliance	Can operate at room temperature.	Difficult to create high numbers of qubits, limiting compute capacity.
Synthetic Qubits							
	**Superconducting Circuits** A resistance-free current oscillates back and forth around a circuit loop. An injected microwave signal excites the current into super-position states.	0.00005	99.9%	High	Google, IBM, QCI, Rigett, Oxford Quantum Circuits	Can lay out physical circuits on chip.	Must be cooled to near absolute zero. High variability in fabrication. Lots of noise.
	**Silicon Quantum Dots** These "artificial atoms" are made by adding an electron to a small piece of pure silicon. Microwaves control the electron's quantum state.	0.03	~99%	Very Low	HRL, Intel, SQC, Oxford Quantum Ocean, DIRAQ, Quantum Motion, EeroQ	Borrows from existing semiconductor industry.	Only a few connected. Must be cooled to near absolute zero. High variability in fabrication.
	**Topological Qubits** Quasiparticles can be seen in the behavior of electrons channeled though semiconductor structures. Their braided paths can encode quantum information.	–	–	–	Microsoft	Designed to be more robust to environmental noise.	Existence not yet confirmed.

A summary of some of the leading quantum information technologies and their characteristics. Table modified from Gabriel Popkin, Quest for qubits. Science 354, 1090-1093(2016). DOI:10.1126/science.354.6316.10 90.

5 Year Update to the Next Steps in Quantum Computing Report

Unlocking True Commercial Advantage

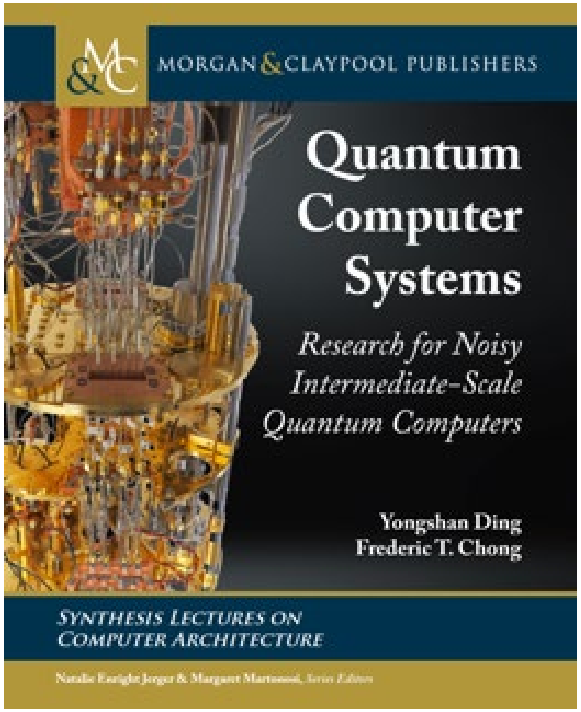
Designing data center ready solutions at scale

	2024 Error Detection	2026 Error Correction	2028 Commercial Advantage
Logical Performance	Dual-species crosstalk-free measurement	Logical circuit depth > 1,000 Universal quantum gate set	Logical circuit depth > 1 million.
Logical Qubits	2	>10	>100
Logical Operations Rate		10,000/sec	100,000/sec
Physical qubits	1,600	8,000	40,000
Fidelity 2Q (CZ):	99.50%	99.90%	99.95%
Fidelity Local 1Q:	99.90%	99.95%	99.95%
Fidelity Global 1Q:	99.99%	99.99%	99.99%
Software	Tooling (QCVV) for verifying fault-tolerant properties	Optimized compilation of fault-tolerant circuits	Exponential speedup demonstration
Enabling technologies	Optimized laser pulse modulation	Advanced photonic beam steering	Realtime atom reloading

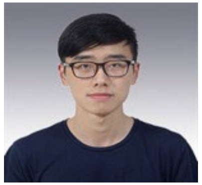
Infleqtion

Summary

- QC is at a historic time
- Physics-aware, full-stack SW can greatly accelerate progress
- Hybrid quantum-classical compute will be key
- More info:
  epiqc.cs.uchicago.edu
  infleqtion.com



EPiQC Alum



Yongshan Ding (Yale)
Quantum RAM,
Crosstalk mitigation,
Qubit reuse,
Synthesis book

Jonathan Baker (UT Austin)
Qudit circuits,
Memory architectures,
2.5D error correction,
Circuit mapping/sched

Kaitlin Smith (Northwestern)
Modular architectures,
Qudit circuits,
Information leakage

Gokul Ravi (Michigan)
Cross-layer optimization,
Hybrid quantum-classical,
Error mitigation

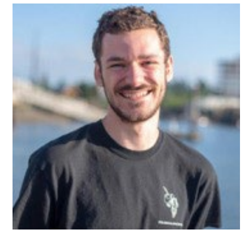
Prakash Murali (Cambridge)
Noise-aware mapping,
Design-space studies,
Resource estimation

Poulami Das (UT Austin)
Error mitigation,
QEC decoding,
Variational algorithms

Saeed Mehraban (Tufts)
Complexity theory,
NISQ computation,
Holomorphic QC

Upcoming Graduates

Siddharth Dangwal 2026
Partial QEC,
Measurement error mitigation,
Scalable noise simulation



Joshua Viszlai 2026
QEC-HW co-design,
Efficient QEC decoding,
QLDPC memories