

Preparing for Scalable Quantum Computing at NERSC



Katie Klymko NERSC nersc.gov/quantum

Summary and Outline of the Talk

NERSC is developing quantum applications and algorithms, building benchmarking tools, and evaluating the infrastructure needed to integrate quantum computing into NERSC's scientific workload.

<u>Outline:</u>

- Introduction to NERSC
- Quantum workload analysis
- NERSC's quantum team and strategy
- Pathfinding efforts for quantum: developing benchmarking tools
- NERSC's quantum user program
- Developing tools with the Berkeley Lab quantum environment





Office of

Science

NERSC is the Mission HPC Facility for DOE Office of Science Research



Office of Science

Largest funder of physical science research in the U.S.



Biological and Environmental Research



Computing



Basic Energy Sciences



High Energy Physics



Nuclear Physics



Fusion Energy, Plasma Physics





Office of Science



The Doudna Supercomputer

- The *Doudna* supercomputer will have thousands of NVIDIA Rubin GPUs and provide more than ten times the performance of *Perlmutter*, NERSC's current flagship supercomputer.
- Doudna will be built using cutting-edge technology:
 - Dell Integrated Rack Scalable Systems and PowerEdge servers
 - Dell ORv3 direct liquid-cooled server technology
 - NVIDIA Vera-Rubin CPU-GPU platform
 - High-speed NVIDIA Quantum-X800 InfiniBand networking platform
- Doudna will be a key tool for research across the DOE Office of Science mission space, powering scientific discovery in areas like fusion energy, advanced materials design, fundamental physics, biomolecular design, quantum computing, and more.
- Doudna is designed to accelerate scientific research by enabling the seamless incorporation of AI, simulations, and data analysis for the near-real-time processing of data from experimental facilities like telescopes and fusion energy research facilities.

5



May 29, 2025

DOE Announces New Supercomputer Powered by Dell and NVIDIA to Speed Scientific Discovery



Named in honor of Jennifer Doudna, the Berkeley Lab-based biochemist who was awarded the 2020 Nobel Prize for Chemistry in recognition of her work on the gene-editing technology CRISPR.

Office of





Quantum Team at NERSC











NERSC Workload Analysis Drives Quantum Computing Requirements



Quantum mechanical problem
 Quantum algorithms proposed

>50% of cycles

non-Quantum mechanical problem
 Quantum algorithms proposed

20% of cycles

What is **not** on the pie chart? Growth areas for quantum technologies!







Quantum Strategy Reflect's NERSC Mission



Pathfinding:

Understand how quantum computing will complement future NERSC workloads and enable new science applications & modes of discovery



Early User & Vendor Exploration:

Deep engagements with vendors and users at small scale to explore capabilities, requirements & best practices for the broad user community **Broad User Impact:**

Develop users and the workforce for new quantum capabilities through training, tutorials, and access to classical and quantum technologies

U.S. DEPARTMENT

of ENERGY

Office of

Science









Quantum @ NERSC Pathfinding and Early User & Vendor Exploration Activities

Vendor Collaboration



Application Development

Benchmarking

Resource Evaluation

IQuEra> R&D Partnership



Simulation of Quantum Systems



Xanadu NESAP Collaboration: Quantum machine learning workflows with PennyLane on Perlmutter





Quantum Computing Access @ NERSC (QCAN)

2022 QIS @ Perlmutter

👌 Quantum Information Science @



30+ Teams and Publications

2024 Neutral Atoms (QuEra>

J.S. DEPARTMENT OF

ENERG

2025 Superconducting

Qubits



Office of

Science





BERKELEY LAB



Inspired by Lin Lin









Office of Science

How Can We Effectively Benchmark New Quantum Algorithms and Hardware?

- New algorithm papers are put on the arXiv every day, but are often not run on a standard set of examples – how useful are they?
- Hardware updates are being introduced quickly, and new technologies are being introduced, but it's often unclear how to compare them





HamLib: A Library of Hamiltonians for Benchmarking Quantum Algorithms and Hardware

Commonality between all qubit technologies and quantum computational models: **solve a computational problem through Hamiltonian evolution and encode the problem in the Hamiltonian**

- Explicitly: adiabatically or through analog quantum computation
- More subtly: the circuit model
- Different types of encodings possible





HamLib: A Library of Hamiltonians for **Benchmarking Quantum Algorithms and Hardware**

- Project led by Intel and NERSC, including Sandia, NASA, LBL, Texas A&M, Oxford
- One of the first quantum benchmarking data sets available
- Independently integrated into Pennylane software from Xanadu
- Published (Quantum 8 (2024): 1559) in December 2024, 20+ citations



Physical Sciences







Office of

Science

HamLib: A Library of Hamiltonians for Benchmarking Quantum Algorithms and Hardware

3-20 qubits For testing on early quantum hardware

10-40 qubits For running on simulators 40-1000 qubits For medium-term algorithm analysis





Want thousands of problem instances







Analogy: Classical Libraries for Benchmarking

Deep learning on images

Molecular datasets





(QM9)

syncedreview.com/2020/06/23/google-deepmind-researchers-revamp-imagenet







Savings in the Problem Preparation Pipeline



Characteristics of a Good Hamiltonian Library

- "Well-spaced" 2-, 4-, 6-, ..., 100- qubit problems
- Inclusion of "real-world" problems
- Already mapped to qubits

A few practical details:

- The dataset mainly contains qubit Hamiltonians (with minimal auxiliary data)
- Files are zipped as HDF5 and represented using represented using OpenFermion's QubitOperator class
- All provided in a Pauli representation using multiple different encodings

$$H_{\text{encoded}} = \sum_{i} c_i \bigotimes_{k} \{\sigma_{ik}\}$$

• Stored on NERSC servers https://portal.nersc.gov/cfs/m888/dcamps/hamlib/







Hamlib Stats



hamlib - 1,570,721 Hamiltonians





Hamlib Chemistry Stats









HamLib Team



Nicolas Sawaya



Daniel Marti-Dafcik



Yang Ho



Daniel Tabor



David Bernal





Pradeep Dubey



Nathan Bishop

111

Anne Matsuura





Bert de Jong





Simon Benjamin



Ojas Parekh





Norm Tubman

















Daan Camps



Library for Optimizing Quantum Circuits Simulations

Doru Thom Popovici, Harlin Li, Naoki Yoshioka, Naoto Aoki, Nobuyasu Ito, Daan Camps, Katie <u>Klymko</u>, Anastasiia Butko

- Graph Construction
 - Create a dependency graph between the gates
- Partition and Optimization
 - Partition the graph into clusters to better utilize the memory hierarchy
 - Construct a hierarchical representation
- Generation of Efficient Implementation
 - Generate efficient code C++/cuda/hip



Output: Optimized implementation

Bringing Science Solutions to the World

U.S. DEPARTMENT

ENERGY

Office of

Science



Preliminary Results: Execution Time on Perlmutter

aft indep aiskit



We start with 30 qubits on 4 GPUs

The blue plot represents the execution of the simulation without optimizations for the local computation

The red and orange plots show our optimized implementation that better map to the memory hierarchy of the GPU

- random indep giskit - random indep giskit

32

- random indep giskit

36





- gft indep giskit - gft indep giskit



20

30



34

NERSC's R&D Partnership with QuEra Computing: First DOE Lab To Partner with a Neutral Atom Vendor

out of Harvard and MIT researchers

systems for science applications

23

Array Processors")

- Quera partners with <u>NERSC</u>
- to offer expanded quantum capabilities
- New partnership helps users of the National Energy Research Scientific Computer Center solve problems in quantum dynamics, chemistry, high-energy physics, and other fields
- **Engagement themes:**

Year 1 (2023-24): Initial hardware explorations for accelerating science applications. Expertise development.

Year 2 (2024-25): Large-scale experiments for expertise and use-case demonstration. User program established.







Boston-based quantum computing startup focused on neutral atom hardware

Current hardware is Aquila, an analog quantum computer with 256 qubits in a

reconfigurable 2D geometry ("Dynamically Field-Programmable Neutral Atom

R&D collaboration has been focused on application development and

understanding the resources needed to use current and future neutral atom



Office of

Science

Aquila: QuEra's Neutral Atom Analog Computer



[2] Henriet et al, Quantum 4, 327 (2020).

Some technical specifications:

- Rb atom
- Maximum qubits: 256
- Maximum duration: 4-20 microseconds
- Field of View: (75-100, 76) microns
- Minimum spacing: 2-4 microns



[1] Aquila v1.0 (2023) https://arxiv.org/abs/2306.11727







Hamiltonian and Rydberg Interactions

$$H = \frac{\Omega(t)}{2} \sum_{j} \hat{\sigma}_{x,j} - \sum_{j} \Delta_{j}(t) \hat{n}_{j} + \sum_{j < k} V_{jk} \hat{n}_{j} \hat{n}_{k}$$
Rydberg interaction
$$V_{jk} = \frac{C_{6}}{r_{jk}^{6}} : \text{van der Waals}$$

$$\implies \text{Entangle atoms}$$

$$\hat{n}_{j} = |1\rangle\langle 1|_{j} : \text{Number operator}$$

$$\Delta_{j} = \Delta_{\text{glob}} + \Delta_{\text{loc},j} : \text{Detuning field}$$

$$\hat{1} \qquad \hat{1}$$

$$\text{Global Local}$$

$$H = \frac{\Omega(t)}{2} \sum_{j} \hat{\sigma}_{x,j} - \sum_{j} \Delta_{j}(t) \hat{n}_{j} + \sum_{j < k} V_{jk} \hat{n}_{j} \hat{n}_{k}$$

$$Rydberg blockade :$$

$$\frac{\varphi_{j}}{\varphi_{j}} = \frac{\varphi_{j}}{\varphi_{j}} + \frac{\varphi_{j}}{\varphi_{j}} +$$

Rabi drive

$$\hat{\sigma}_{x,j} = |0\rangle\langle 1|_j + |1\rangle\langle 0|_j$$
: Pauli-X



 Ω : Rabi frequency





· · 10

 $V(r > r_B) = 0$

 x_j Distance (µm)

Coarsening Development with QuEra



2024 IEEE International Conference on Quantum Computing and Engineering (QCE)

Quantum dynamics for cosmology and high energy physics

- *Physical Review B* 110 (15), 155103 - *Physical Review B* 110 (15), 155114





Bringing Science Solutions to the World



Probing the Phase Diagram of the Lieb Lattice



In condensed matter physics and materials science, computing the ground state phase diagram tells us how important materials behave under different conditions.





Bringing Science Solutions to the V

Comparison of Phases Between Aquila and Perlmutter



Comparison of Phases Between Aquila and Perlmutter



Quantum Computing Access @ NERSC (QCAN)

Quantum Information Science on Perlmutter

Perlmutter

through access to

NERSC's classical

10+ teams selected per

• Started January 2022

year on a rolling basis

resources

• 30+ publications

• Early user exploration

🙀 Quantum Information Science @

Atoms Access

Neutral

- through QuEra Computing
- Close collaborations between NERSC, science teams, and hardware vendor
- Up to 25h of QC usage per selected team
- 20 teams applied, 3 selected
- Started October 2024

Superconducting

Qubits Access through <u>IBM</u>



- Broader access to quantum hardware for a subset of NERSC users (up to 15 projects)
- Support from IBM team
- IBM Quantum Network: community of users
- Started January 2025

We are adapting access based on advancements in quantum hardware, algorithms and software development, and increased understanding of users' needs.









QIS@Perlmutter: Science Highlights 2022-2025

Perlmutter GPU hours dedicated for research in areas that include quantum simulation, error mitigation, chemistry, materials science, condensed matter physics, and optimization: accelerating quantum simulations at supercomputing scale.

Users have access to NVIDIA tools like CUDA-Q. **30+ Project Teams, 30+ Publications**





Hybrid Transformer Architecture Smaldone et al, *arXiv*:2502.19214

Office of

Science

Generation



of ENERGY

PI: Marwa Farag (NVIDIA)



QCAN QuEra Project Awardees

Dr. Huan-Hsin Tseng



Maximum Independent Set Problems

The MIS problem is **NP-hard** and has **broad applicability**, including in high-energy and condensed-matter physics. This project proposes to develop and implement an **improved MIS algorithm** on the Aquila neutral atom quantum computer.

Dr. Bert de Jong





Spin Transport in 2D Lattices

Information Scrambling in Negatively Curved Spaces

32

Spintronics uses spin transport in solid-state devices for efficient information processing and storage. This project proposes to carry out an experiment of spin transport on Aquila.

Prof. Lex Kemper





Models of quantum gravity that exhibit AdS/CFT correspondence, are **notoriously difficult to classically simulate.** The results will elucidate information scrambling and dynamics in curved spaces









NERSC IBM Quantum Hub Opened in 2025

LBL employees and selected external users will get access to access multiple IBM quantum processor systems, all with 100+ qubits, and new systems as they come online





Credit: Christopher Tirrell from IBM





Utility-scale Quantum Processor from IBM



Credit: Christopher Tirrell from IBM







Office of Science

NERSC Benefits from LBNL's Quantum Research Community & Interacts with Broader DOE Efforts

Berkeley Lab-led Quantum Centers, Testbeds, and Facilities

QUANTUM SYSTEMS ACCELERATOR Catalyzing the Quantum Ecosystem		AQT	QUANT-NET
National QIS Research	Extensive toolset for the	DOE Testbed that provides	DOE Testbed that will
Center to catalyze national	exploration of new	users with full-stack access	enable quantum
leadership for co-design of	materials through synthesis,	to an advanced quantum	entanglement distribution
algorithms, quantum	fabrication,	computation system based	between Berkeley Lab and
devices, and engineering	characterization, and	on superconducting	UC Berkeley with optical
solutions.	simulation.	circuits.	fiber.

DOE Quantum Programs

Oak Ridge Leadership Computing Facility Quantum Computing User Program



Quantum Performance Laboratory NERSC Quantum events have participants from PNNL, LLNL, ORNL, SNL, and BNL







Office of

Science

Full Stack Superconducting Quantum Computing Platforms





aqt.lbl.gov

Cryogenic Platform

Bluefors dilution fridge with 160 RF lines, operating at 10 mK. Modular and extensible cryopackaging, developed in partnership with Bleximo, mitigates crosstalk, provides control and readout for 128 qubits, and can accommodate multiple chips.



Superconducting Quantum Processors

Current chip is an 8-qubit transmon ring design with high-coherence qubits (T_1 and T_2 >100 µs). A novel 8-qubit QPU with arbitrary dynamically reconfigurable (up to all-to-all) qubit-qubit connectivity will soon be available.





Software Stack

Open programming stack and flexible user interface, with demonstrated tailoring and mitigation of coherent errors to improve algorithmic performance. Can support non-standard user software needs.

Commercial and Custom Controls

Cutting-edge controls equipment capable of fast feedback and on-the-fly state detection, which can accommodate custom user needs. Both commercial Zurich Instrument solutions and an in-house custom modular solution called QubiC are available.

QubiC: Open-source FPGA-based BERKELEY LABQUANTUM Bit Control System FFFFFF Bringing Science Solutions to the World



QubiC 1.0

- Started in 2018 by Gang Huang, based on VC707.
- Auxiliary RF components, firmware, software, calibration protocols published & open-sourced.
- Low-cost DACs-ADCs





- Based on ZCU216 RFSOC
- •10 GSPS allows direct generation up to ~10 GHz in 2nd Nyquist Zone
- AC- and DC-coupled AFEs
- Adopted by multiple groups
- •Used for various co-design projects
- QubiCSV: data management and visualization



- Novel Architecture on FPGA to enable Mid-Circuit Measurement and Feedforward
- Flexible Function Proc enables arbitrary logic



Multi-Platform Sync



- 2x VC707 and 2x ZCU216 synchronized thru ref clk
 - 24-ch 16-bit DAC @1GSPS
 - 32-ch 14-bit DAC @ 8 GSPS
- RMS jitter < 10ps
- Next step: decision distribution



QubiC: Open-source FPGA-based U.S. DEPARTMENT OF BERKELEY LABQUANTUM Bit Control System Office of Science Bringing Science Solutions to the World **Distributed Processor Multi-Platform Sync QubiC 2.0** QubiC 1.0 REF CLK Function Proc or Results Bus LCDAC LCDAC LCDAC VC707 DO DAC CDAC LCDAC CLK104 C707 **ZCU216** CLK104 Started in 2018 by Ga ZCU216 Huang, based on VC7 Auxiliary RF compone firmware, software, 2707 and 2x ZCU216 calibration protocols hronized thru ref clk Get Involved Open source repos published & open-so https://gitlab.com/LBL-QubiC https://go.lbl.gov/qubicsignup |-ch16-bit DAC @1GSPS Low-cost DACs-ADCs • AC- and DC-coupled AFEs arbitrary logic • 32-ch 14-bit DAC @ 8 GSPS • Adopted by multiple groups • RMS jitter < 10ps •Used for various co-design Next step: decision distribution projects QubiCSV: data management U.S. DEPARTMENT Office of and visualization of ENERGY Science



QubiC: Hardware Efficient Randomized Compiling

(a)



Bringing Science Solutions to the World

The Problem:

- Ouantum Processors suffer from both coherent and stochastic noise.
- Coherent noise is much more problematic than stochastic noise.
- Thankfully, there is Randomized Compiling to help us convert coherent noise to stochastic noise.
- Unfortunately, Randomized Compiling is very expensive in terms of classical steps (compilation and upload). often saved for final runs.



The Solution: Bypass Compilation

- 1. Randomly select twirling Pauli on FPGA from {I, X, Y, Z}
- Look up the "undoing Pauli" through a 2-qubit gate in a 4x4 element table:

 $P_i, P_j \rightarrow P'_i, P'_j$

3. Single-qubit gate combination around U3: Given: $U3 = Z(\phi_2) X_{90} Z(\phi_1) X_{90} Z(\phi_0)$ We find: $P_i U3P'_{i-1} = Z(\phi'_2) X_{90} Z(\phi'_1) X_{90} Z(\phi'_0)$ where ϕ'_n is given by a 64-element map.

39

Experimental Results





QubiC: Hardware Efficient



Hardware-Efficient Randomized Compiling

<u>The</u>

- Quantum F from both stochastic
- Coherent r problemati noise.
- Thankfully Randomize help us cor noise to ste
- Unfortunal Compiling in terms of compilatio often save



Neelay Fruitwala,^{1,*} Akel Hashim,^{2,3,*} Abhi D. Rajagopala,³ Yilun Xu,¹ Jordan Hines,² Ravi K. Naik,³ Irfan Siddiqi,^{2,3,4} Katherine Klymko,⁵ Gang Huang,^{1,†} and Kasra Nowrouzi^{3,‡}

 ¹Accelerator Technology and Applied Physics Division, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA
 ²Department of Physics, University of California at Berkeley, Berkeley, CA 94720, USA
 ³Applied Math and Computational Research Division, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA
 ⁴Materials Sciences Division, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA
 ⁵National Energy Research Scientific Computing Center, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA
 (Dated: June 21, 2024)

Randomized compiling (RC) is an efficient method for tailoring arbitrary Markovian errors into stochastic Pauli channels. However, the standard procedure for implementing the protocol in software comes with a large experimental overhead — namely, it scales linearly in the number of desired randomizations, each of which must be generated and measured independently. In this work, we introduce a hardware-efficient algorithm for performing RC on a cycle-by-cycle basis on the lowest level of our FPGA-based control hardware during the execution of a circuit. Importantly, this algorithm performs a different randomization per shot with zero runtime overhead beyond measuring a circuit without RC. We implement our algorithm using the QubiC control hardware, where we demonstrate significant reduction in the overall runtime of circuits implemented with RC, as well as a significantly lower variance in measured observables.



-compile

d circuit

d data

27

nt/server

2.00

1.50

1.25

0.75

0.25



QubiC: Parametrized Circuit Execution



Bringing Science Solutions to the World

The Problem:

- Many commonly-deployed protocols for benchmarking and mitigation (e.g. RB, CB, RC, etc) consist of 100s-1000s of circuits.
- The classical steps in running these circuits often take the majority of total run time.
 <u>Memory</u> and <u>upload</u> are particularly problematic.
- This makes it <u>expensive</u> to characterize QPUs and mitigate noise/errors.

	RC	СВ	RB
# Circuits	1540	6000	736
# Unique Circuits	77	21	32
Classical Time (%)	94	68	43
NERS	C		

The Solution: RIP & Stitch!

- Note that in all these cases, only a small minority of the total set of circuits are structurally unique! The rest are only off by gate parameters (e.g. phase).
- •On the host computer:
 - Read the circuits
 - dentify unique circuits
 - Peel off the parameters
- Compile and upload <u>only</u> the unique circuits, plus a table of parameters
- At run-time on FPGA: <u>Stitch</u> the parameters back in







The Proble

- Many commonly-dep protocols for benchr and mitigation (e.g. F RC, etc) consist of 100s-1000s of circu
- The classical steps in these circuits often 1 majority of total run <u>Memory</u> and <u>upload</u> particularly problem.
- This makes it <u>expen</u>: characterize QPUs a mitigate noise/error:

	RC
# Circuits	1540
# Unique Circuits	77
Classical Time (%)	94
NERS	C

QubiC: Parametrized Circuit





Hardware-Assisted Parameterized Circuit Execution

Abhi D. Rajagopala,¹ Akel Hashim,^{1,2} Neelay Fruitwala,³ Gang Huang,³ Yilun Xu,³ Jordan Hines,⁴ Irfan Siddiqi,^{2,1,5} Katherine Klymko,⁶ and Kasra Nowrouzi¹

 ¹Applied Math and Computational Research Division, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA
 ²Quantum Nanoelectronics Laboratory, Department of Physics, University of California at Berkeley, Berkeley, CA 94720, USA
 ³Accelerator Technology and Applied Physics Division, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA
 ⁴Department of Physics, University of California at Berkeley, Berkeley, CA 94720, USA
 ⁵Materials Sciences Division, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA
 ⁶National Energy Research Scientific Computing Center, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA
 (Dated: August 30, 2024)

Standard compilers for quantum circuits decompose arbitrary single-qubit gates into a sequence of physical $X_{\pi/2}$ pulses and virtual-*Z* phase gates. Consequently, many circuit classes implement different logic operations but have an equivalent structure of physical pulses that only differ by changes in virtual phases. When many structurally equivalent circuits need to be measured, generating sequences and measuring each circuit is unnecessary and cumbersome since compiling and loading sequences onto classical control hardware is a primary bottleneck in quantum circuit execution. In this work, we developed a hardware-assisted protocol for measuring parameterized circuits on our FPGA-based control hardware, QubiC. This protocol relies on a hardware-software co-design technique in which software identifies structural equivalency in circuits and compiles unique circuits to reduce the overall waveform compilation time. The hardware architecture then performs real-time stitching of the parameters to the circuit before measuring on quantum hardware. This work demonstrates that this protocol speeds up the total execution time for many quantum circuits and algorithms.

Keywords: Quantum Computing, Parameterized Circuits, FPGA

Research to Address the Classical Requirements for Future Quantum Workloads

- NERSC Team awarded best paper at ISC 2024 for "Evaluation of the classical hardware requirements for large-scale quantum computations"
- Paper explores the network latency, network bandwidth between the quantum computer and the HPC system, and compute requirements for applications in physics and chemistry





10

Above plot shows estimated FLOPs required for real-time decoding as a function of number of logical qubits (x-axis) and depth (y-axis). Red markers show resources required for different large-scale problems from condensed matter physics and quantum chemistry. All problems require less than 1 petaflop of compute.

2000

DEPARTMENT

1500

250

Office of

Science



Summary

- The NERSC quantum team is engaging in strategic pathfinding/R&D efforts and collaborations to prepare for future NERSC systems.
- NERSC has a quantum user program designed for both scientific impact and breadth of users which will inform future requirements.
- Understanding the future role of quantum computing in the HPC workload and developing the infrastructure to serve users is a community-wide activity which will require collaboration between NERSC staff, vendors, researchers, users, and the DOE community at-large.





Acknowledgements









QuEra Computing



Berkeley Lab

















