Advanced Scientific Computing Research (ASCR) March 2026 SCAC Meeting

Hal Finkel, Ph.D.

Acting Associate Director

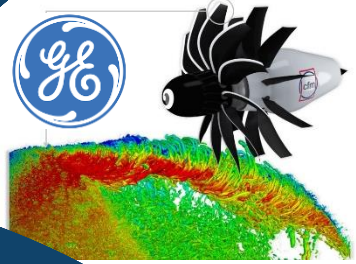
High-Performance Computing in the Office of Science Today

Enabling Scientific Breakthroughs

Pushing Innovation in HPC

DOE (ASCR + NNSA/ASC) is the only agency to deliver on the first Trump administration's strategic Exascale goals.

Discovery of Tetraneutron



Innovative Airplane Engine Design

Research

- Laying the foundations
- Making the previously impossible, possible
- Following critical technology trends
- Prepping for disruption

Facilities

- First of a kind systems
- Each facility is an innovation engine
- Derisking for industry
- Developing user communities

Quantum Computing

World's first exascale supercomputer

ASCR Pioneered…

- ➢ World's **<u>first</u>** exascale computers
- ➢ World leading HPC software ecosystem
- ➢ Wide-adoption of HPC for science & industry
- ➢ Quantum testbeds for science
- ➢ Foundations for AI revolution

Not so secret ingredient:
Strategic partnerships with

- ➢ Computing vendors
- ➢ Industry users
- ➢ Science communities
- ➢ NNSA

ASCR's End-to-End Approach for Frontier AI

Transforming Discovery Science & Infrastructure by AI

RESEARCH

- Co-design to improve computational efficiency by >100X
- Digital twins, new software, and data tools
- R&D in applied math & CS w/ highly curated data
- Edge AI & robotics, knowledge extraction

- Hardware Innovation
- Breakthrough Tools for AI
- Deep Understanding of AI Models
- AI-Driven High-Precision Science R&D

Vendor Partners | AI Companies, Industry Users, DOE, & Interagency Partners

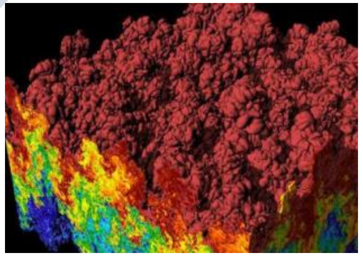
- Testbeds for computing & automated labs
- Prototypes for ultra efficient data centers
- World's most capable LCFs to train models
- IRI, ESnet, AI-enabled real-time control

ADVANCED TECHNOLOGIES & FACILITIES

Building Science & Infrastructure of AI

Next generation facilities

Automated labs

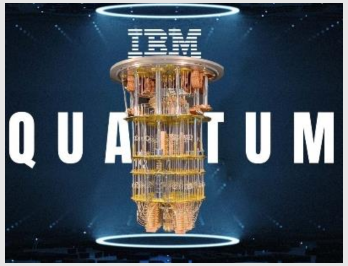


AI surrogate models

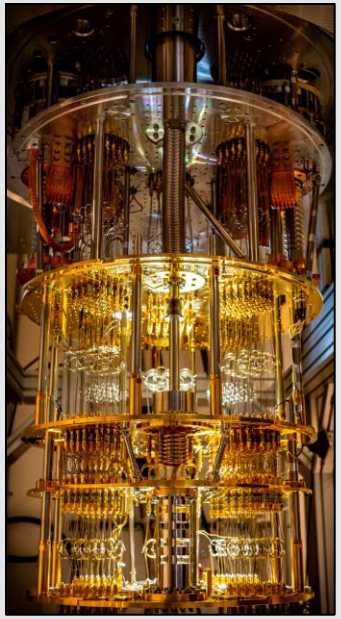
ASCR Invests in a Multi-Pronged Portfolio in QIS

Fundamental Science	National QIS Research Centers	Quantum Tech Testbeds
Continued investments in core basic research for quantum algorithms, quantum computer science, and quantum networking.	5 Centers address major cross-cutting challenges in broad ranging topics including computing, communications, and sensing.	• Exploration of innovative quantum error correction architectures and approaches for integrated and distributed quantum computing

Cloud Access to Quantum Computers - Expanding access to commercial platforms through the Quantum Computing User Program (QCUP) at OLCF and QIS @NERSC (QCAN)



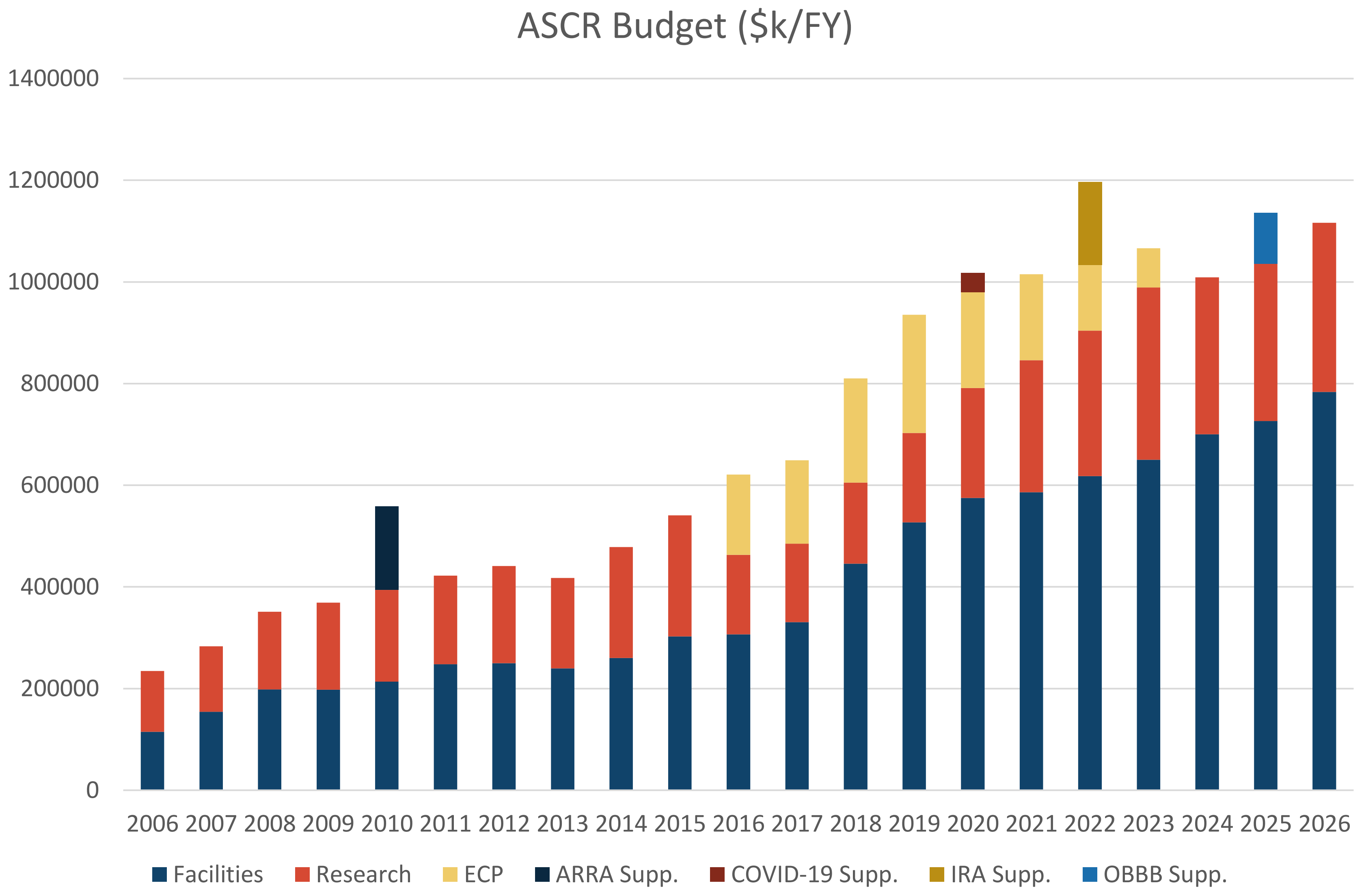
QuEra>
Computing Inc.

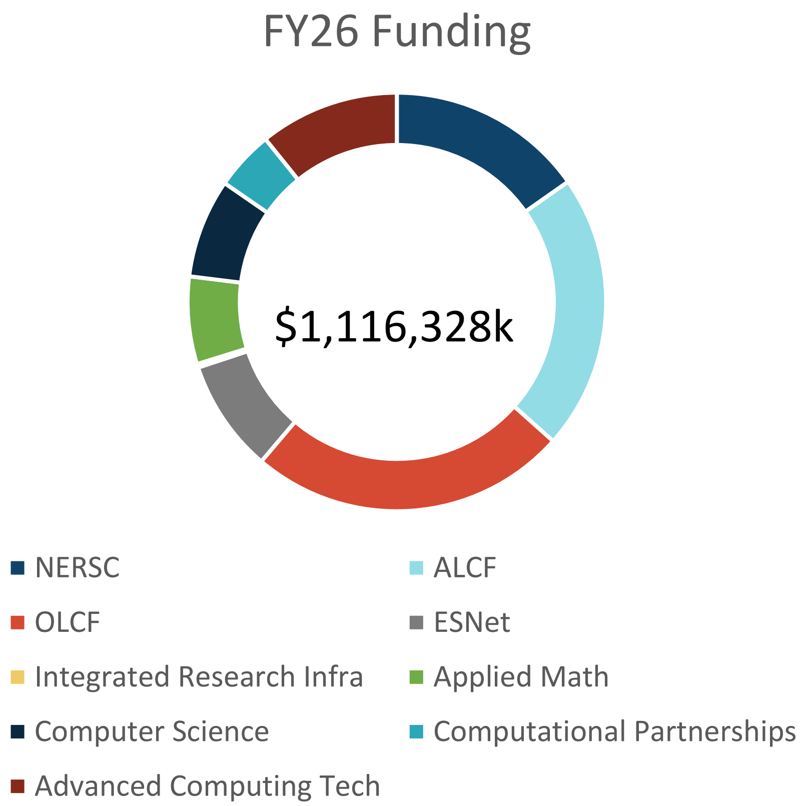


ASCR's quantum computing testbeds are open to the user community.

The first simulation of an atomic nucleus on the quantum cloud.

ASCR Budget History





ASCR's Research Portfolio in FY 2026

Targeted FY 2026 competitions for Inverse Methods, Quantum Error Correction, Correctness for Scientific Computing (with NSF), and Genesis Mission Challenges.

Awards for early-career researchers and EPSCoR institutions are spread across the portfolio.

- Computer Science
  - Advanced AI
    - AI-Driven Software Development
    - Hardware Architectures and Neuromorphic Computing
    - Transformational AI Models Consortium
  - Genesis Mission Challenges
  - Computational Physical Systems / Analog Computing
  - Next Generation (Exascale+) Software Stack
    - Exascale+ Software Ecosystem
    - Distributed Resilient Systems
    - Correctness for Scientific Computing Systems (CS2) (NSF)
    - Core Modeling Software Techniques
  - HPC System / Architecture Modeling and Co-Design
  - Advanced Networking and Wireless
  - Quantum Networking
  - Quantum Computing
    - Quantum Software Stack
    - Quantum Error Correction and Topological Approaches
    - Quantum Utility / Algorithms
    - Quantum System Benchmarking
  - Competitive Portfolios

- Applied Mathematics
  - Advanced AI / Scientific Machine Learning
    - Foundation Models and Training
    - AI Uncertainty Quantification
    - Federated and Privacy-Preserving AI
    - Physics-Informed Machine Learning
    - Core Methods and Decision Support
  - Genesis Mission Challenges
  - Randomized Algorithms
  - Inverse Methods
  - Data Reduction
  - Bridge2AI Collaborations (NIH)
  - Mathematical Multifaceted Integrated Capability Centers (MMICCs)
  - Competitive Portfolios

- Computational Partnerships
  - Scientific Discovery through Advanced Computing (SciDAC)
    - SciDAC Partnerships (BES, BER, HEP, NP, FES; NE, OE)
    - SciDAC Institutes
  - Genesis Mission Challenges
  - Scientific Computing for User Facilities (BES)
  - Partnerships for Biology
    - Rapid AI for Cancer Ecosystems (NIH)
    - Partnerships for Biology Pilots (BER)
  - Collaborative Research in Computational Neuroscience (CRCNS) (NSF)

- Advanced Computing Technologies
  - The American Science Cloud
  - Centers
    - Microelectronics Science Research Centers
    - Quantum Science Research Centers
  - Computational Science Graduate Fellowship (CSGF)
  - Testbeds
    - Quantum Testbeds
    - Neuromorphic and Microelectronics / Memory Computing Testbeds
    - Robotics Testbed Pilots

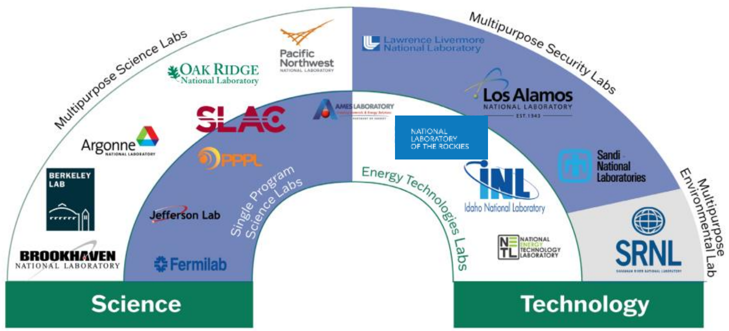
Key:
Pilots
NSF Partnerships
NIH Partnerships
Genesis Mission

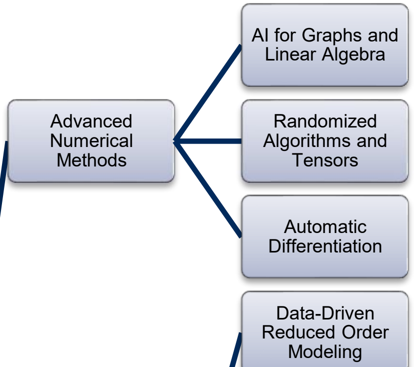
- ASCR research currently has around 400 active projects.
- Some projects are large (e.g., centers) and many are relatively small (with many projects having coordinated awards to multiple institutions).

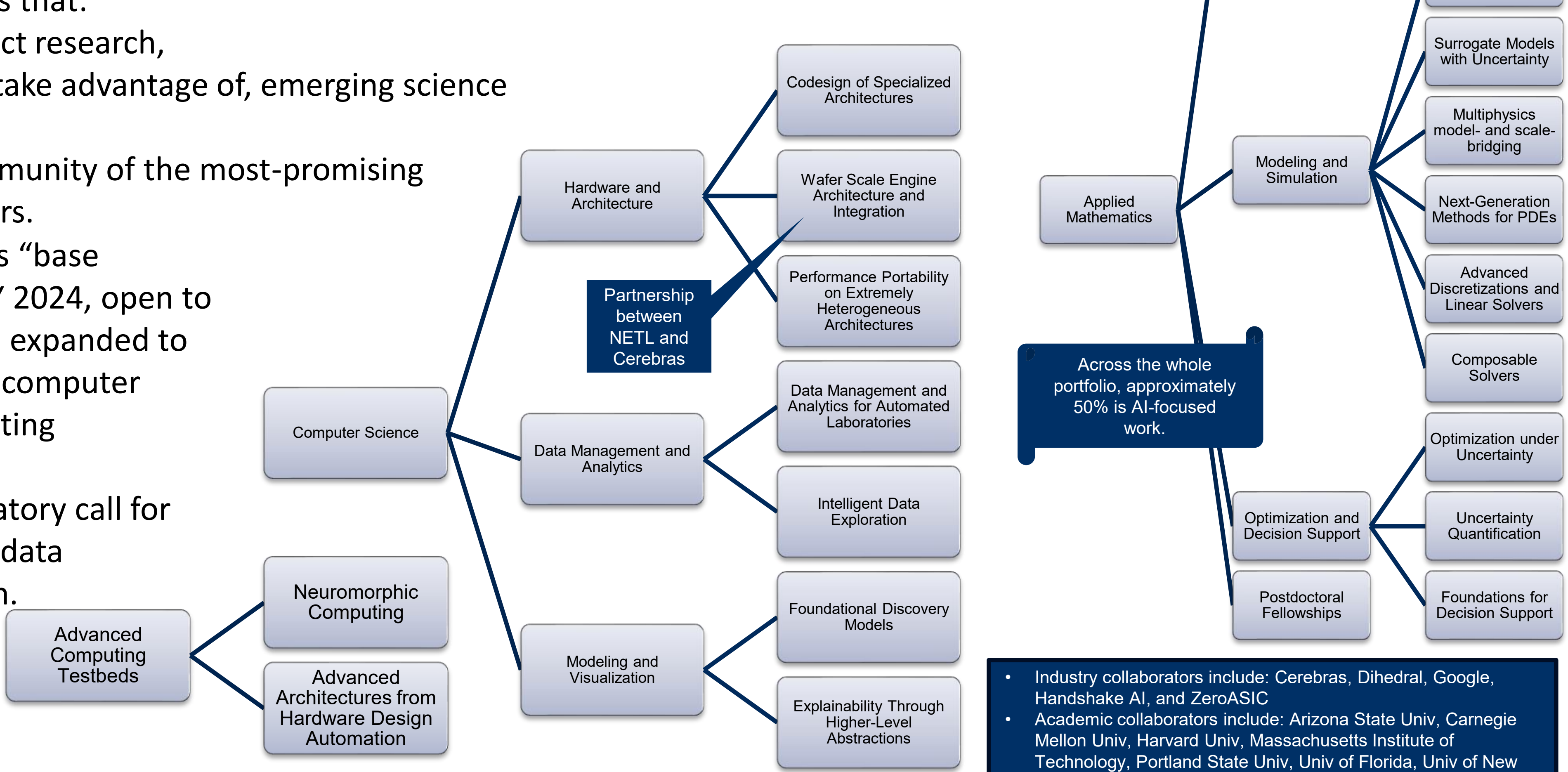
Competitive Portfolios

To ensure continued leadership in delivering on the promise of computational science, and drive innovation in efficient and versatile high-performance computing for science, ASCR seeks to invest in DOE National Laboratory-led portfolios that:

- Support long-term, high-impact research,
- Aggressively respond to, and take advantage of, emerging science and technology trends, and
- Collaborate with a broad community of the most-promising academic and industry partners.
- A full re-competition of ASCR's "base program", via solicitation in FY 2024, open to all DOE National Laboratories, expanded to include applied mathematics, computer science, and advanced computing technologies / testbeds.
- Followed by an FY 2025 laboratory call for computer science research in data management and visualization.
- ANL, BNL, LANL, LBNL, LLNL, NETL, ORNL, PNNL, SNL are all in the current portfolio.







- Industry collaborators include: Cerebras, Dihedral, Google, Handshake AI, and ZeroASIC
- Academic collaborators include: Arizona State Univ, Carnegie Mellon Univ, Harvard Univ, Massachusetts Institute of Technology, Portland State Univ, Univ of Florida, Univ of New Mexico, and Univ Of Texas at Arlington

ASCR 2026 Research Funding Opportunities

Open-Call Panels: In DE-FOA-0003600: Inverse Methods and Quantum Error Correction
Targeted Solicitations: Correctness for Scientific Computing (CS2) (NSF partnership NSF 24-571)
Cross-Cutting Solicitations: Early Career Research Program (DE-FOA-0003602) and Building EPSCoR-State/National Laboratory Partnerships (DE-FOA-0003615)
Genesis Mission RFA (DE-FOA-0003612): ASCR is planning to partner in many areas, and the following focus areas on those on which ASCR is leading:

- Discovering Quantum Algorithms with AI (Application-aware Error Correction, Computational Tools for Fault Tolerant Quantum Computational Science,
- Realizing Quantum Systems for Discovery (AI for Quantum Computing and Networking)
- Recentering Microelectronics in America (AI-Driven Architecture Design; 3D non-volatile compute-in-memory technology; Physics-Based Circuit Design, Simulation, and Emulation; Power Electronics and Communication Networks; Low-temperature Electronics for Sensors and Computation; Transform Neuromorphic Computing Connectivity, Communication, and System Hardware Integration)
- Achieving AI-Driven Autonomous Laboratories (Advanced Robotics for Dynamic Laboratory Environments, AIOps - AI for Network Operations, Accelerate the design and prototyping of neuromorphic computing circuit primitives for robotic embodied physical artificial intelligence)
- HPC Code Curation, Translation, and Development for Accelerated Scientific Discoveries (AI-Driven Code Porting and Optimization; Automated Scientific Problem-to-Code Generation; Neuro-Symbolic Agents for Code Development; Performance Prediction and Feedback Loops; Trustworthy AI for Scientific Software; Multi-Modal Data Integration for Code Intelligence; Partnerships for HPC AI Advancement)
- AI for Scientific Reasoning (Trustworthy Mathematical and Symbolic Reasoning, Hypothesis Generation from Multi-Modal Data, Composable and Modular Foundation Models)
- Cybersecurity for AI-driven Science Workflows (AI for Adversarial Robustness and Resilience, Data Provenance and Integrity Verification, Real-Time Attack Detection and Mitigation for AI Models)

ASCR Facilities provide world-leading computing, data, and networking infrastructure for extreme-scale science while advancing U.S. competitiveness

World's First Exascale Computers

Argonne Leadership Computing Facility

Oak Ridge Leadership Computing Facility

Leadership Computing Facilities (ALCF, OLCF)
Unique national HPC resources for extreme-scale applications, delivering the exascale (10^{18}) era of supercomputing

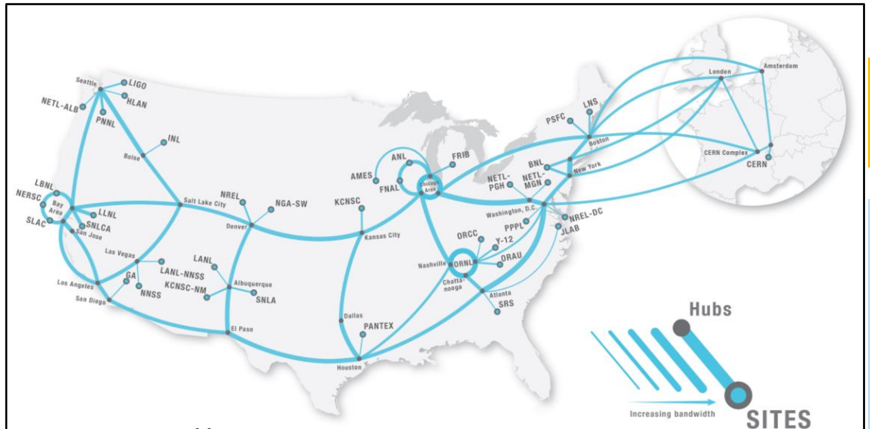
SC's Largest User Facility

NERSC at LBNL

High Performance Production Computing Facility (NERSC)
Dedicated HPC resource serving 11,000+ users annually

NNSA National Nuclear Security Administration | U.S. DEPARTMENT OF ENERGY Office of Science | ECP EXASCALE COMPUTING PROJECT

ASCR's decades long partnership with NNSA led the Nation to win the race in exascale

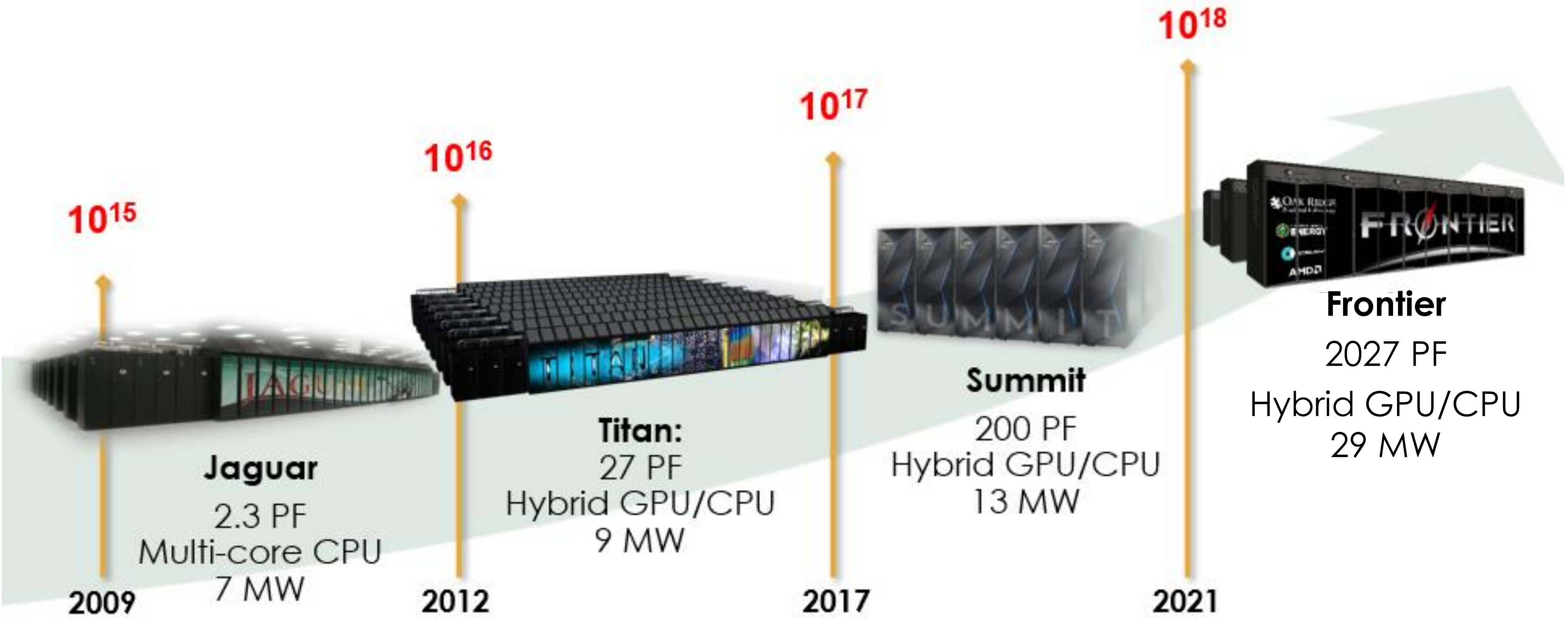


ESnet, managed by LBNL

Each ASCR Facility is an Innovation Engine

Energy Sciences Network (ESnet):
- Connects all DOE national labs and dozens of other DOE sites to 150+ global research networks, commercial cloud providers, and the internet
- Engineered for lossless transmission of huge data flows

ASCR's Public-Private Partnerships Have Shaped the AI Landscape



Titan: ORNL worked with **NVIDIA** to make GPUs viable for HPC

Summit: NVIDIA added Tensor cores to make GPUs viable for AI

Frontier: ORNL worked with **AMD** on CPUs and GPUs for HPC

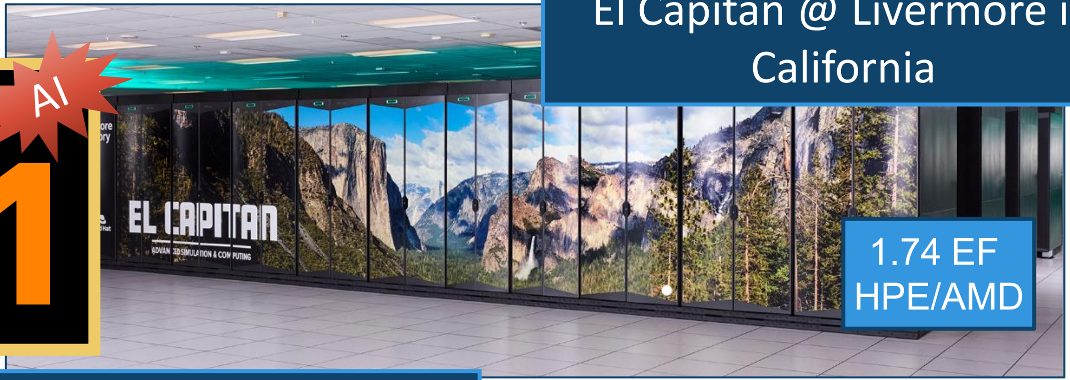
Aurora: ANL worked with **Intel** on GPUs for HPC

DOE Exascale Supercomputers Lead in Performance and Efficiency

TOP500 # 1

HPL-MxP # 1 (AI)

El Capitan @ Livermore in California



1.74 EF
HPE/AMD

Exascale Today Enables the AI of Tomorrow

Frontier @ Oak Ridge in Tennessee

1.35 EF
HPE/AMD

TOP500 # 2

HPL-MxP # 3 (AI)

GREEN 500
Honorable Mention

Aurora @ Argonne in Illinois

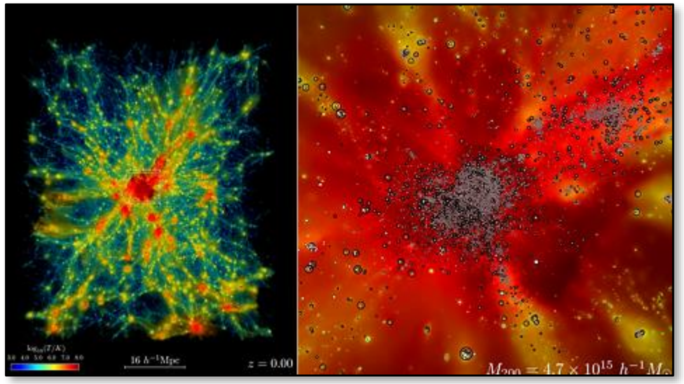
1.01 EF
Intel/HPE

TOP500 # 3

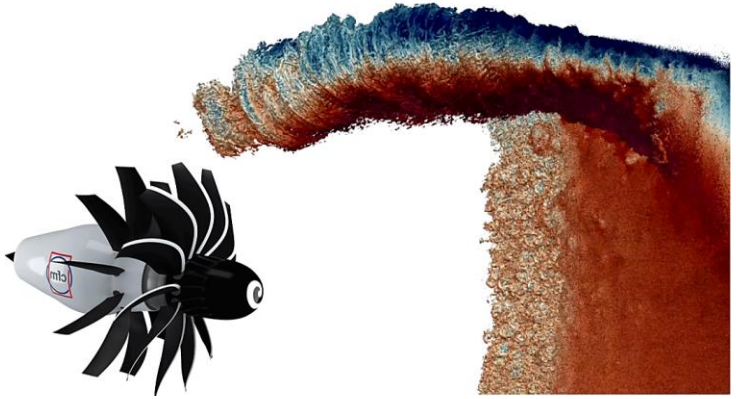
HPL-MxP # 2 (AI)

OLCF Frontier Is Delivering Impactful Exascale Science

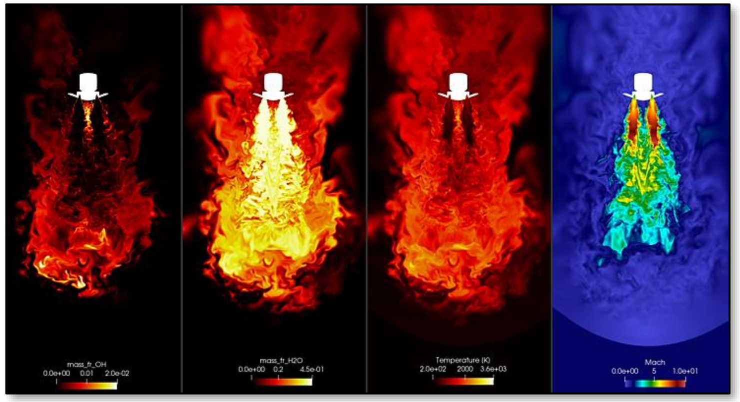
World's largest simulation of the expanding universe
Argonne National Laboratory



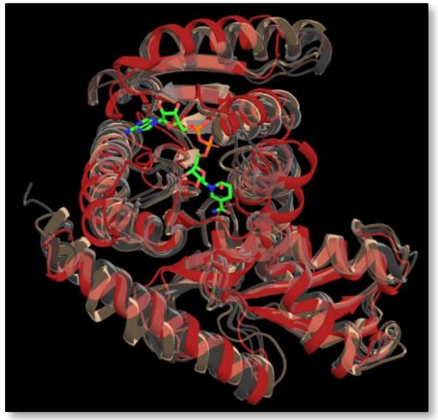
Flight-scale aero-acoustic simulations for jet engine design only possible at the exascale
GE Aerospace



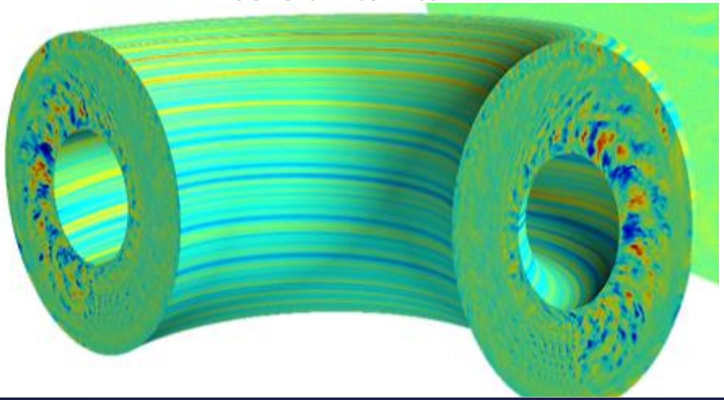
First detailed simulation of retro-propulsion human-scale Mars lander descent
NASA Langley Research Center



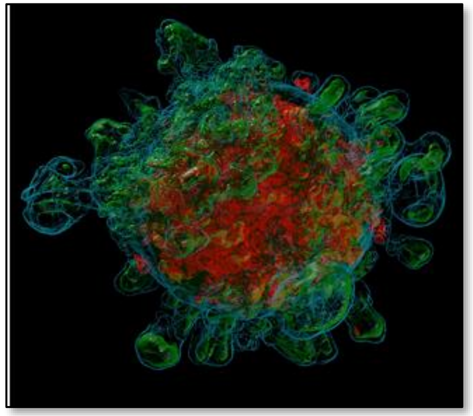
AI-training of large language model for protein design
Argonne National Laboratory



Bridging simulation and data analysis of DIII-D fusion plasma turbulence in unprecedented detail
General Atomics

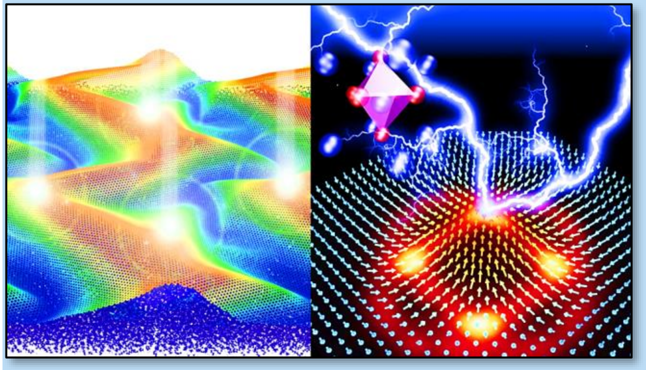


Revealing origins of heavy elements from stellar supernova
Lawrence Berkeley National Laboratory



ALCF Aurora is Delivering Impactful Exascale Science

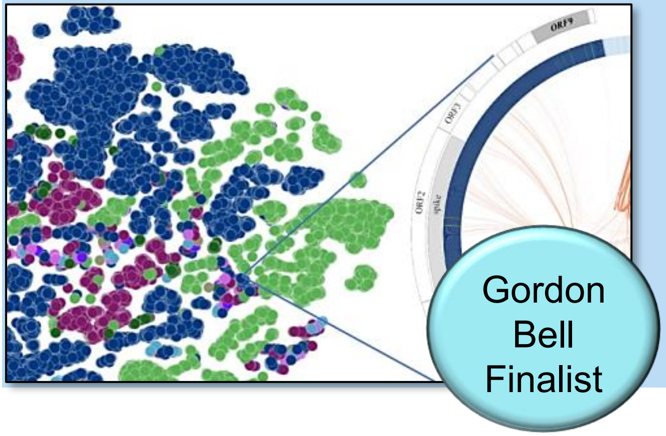
Advanced Quantum Materials



Advanced scalable manufacturing of quantum materials and ultrafast control of their emergent properties. Used AI-guided simulations in tandem with state-of-the-art x-ray, electron-beam, neutron experiments at DOE facilities.

University of Southern California

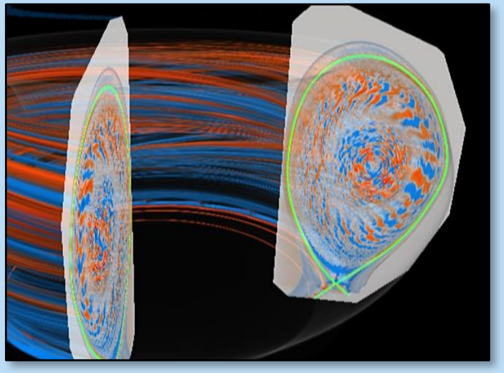
New Generation of Protein Discovery



Combined protein sequence data, experimental results, molecular simulations, and text narratives to provide detailed insights into properties with AI technologies. Scaled to large virus and bacterial genomes. Approach speeds up protein discovery for medicines, catalysts, and other applications.

Argonne National Laboratory

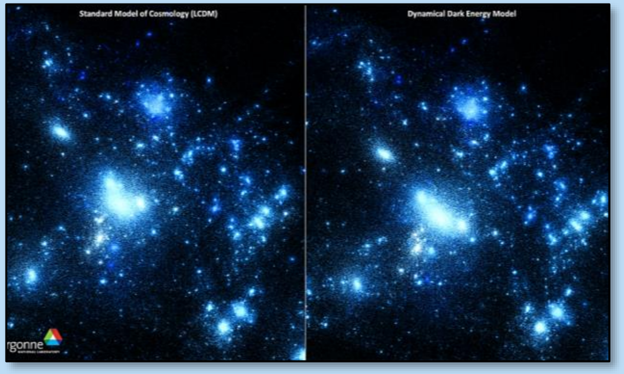
Designs for Fusion Power



Answered questions critical to the design of fusion power plants with first-principles simulations. Provided insights into issues with power exhaust, including mitigating stationary heat-flux densities and avoiding unacceptably high transient power flow to material walls

Princeton Plasma Physics Laboratory

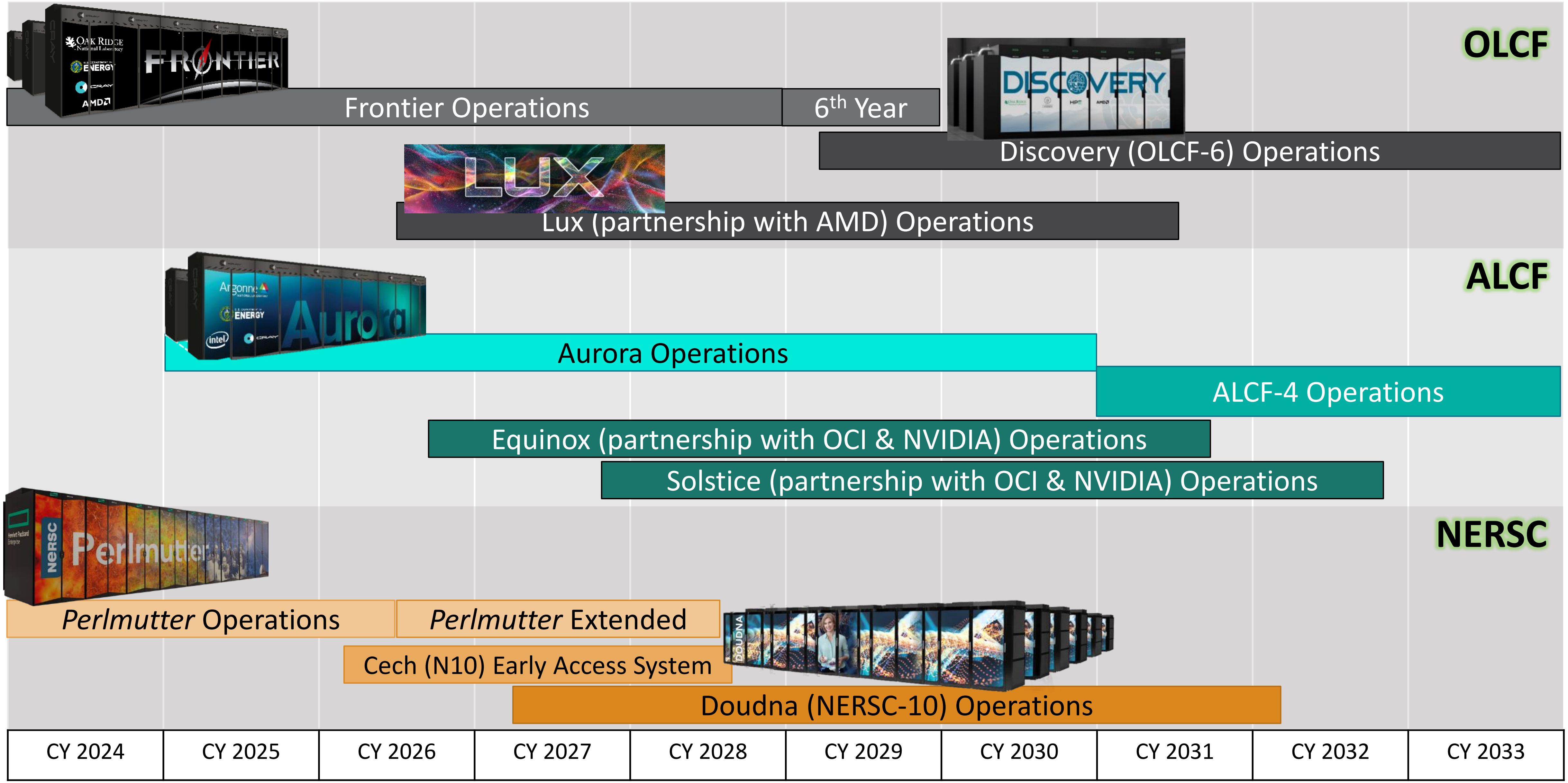
Dark Energy Discovery with DESI



Used Aurora to test ground-breaking observations from DESI that dark energy may not be constant. Simulations provide a key feedback to understand if observations are real or an artifact.

Argonne National Laboratory

ASCR Announced HPC Systems and Projects Timeline



ASCR Advanced Computing Technologies (ACT) Bridges the Gaps

Bridging the computing technologies readiness gap

Research
New computing approaches & ideas

Pathfinding Testbeds
Rapid technology evaluation for suitability, performance, & scaling

Advanced Prototypes
Developing and testing of technology prototypes ready for scaling

Technology Scaling Testbeds
Bringing advanced prototypes to facilities scale for evaluation

Facilities
Computing technology solutions ready for use in production

ACT →

ASCR ACT: Microelectronics

Vision: Anticipate future scientific computing needs leveraging advanced and emerging microelectronics technologies.

Goal: Accelerate scientific innovation by investing in R&D, prototyping, scaling, and transitioning advanced microelectronics technologies that increase experimental throughput and efficiency.

Investments include:

- **Public-Private Partnership** (BNL-GreatSky): ***Superconducting Optoelectronic Computing***
  - Protype a Josephson Junction-based AI accelerator chip that delivers ~100X compute-power performance over state-of-the-art.
- **Public-Private Partnership** (PNNL-Micron): ***Advanced Memory to Support AI for Science***
  - Developing an ecosystem of memory-centric technologies designed to overcome memory capacity, bandwidth, and latency constrains in scientific computing AI-based applications.
- **Testbed** (DUKE-Lead Collaboration): ***Neuromorphic Computing Testbed***
  - Develop & Prototype emerging analog neuromorphic computing circuit primitives; the building blocks of neuromorphic computing (indistinguishable from biology).
- **DOE SC Microelectronic Centers:** *Network of 16 projects led by 10 national laboratories*
  - Transforming computing efficiency, developing microelectronics for extreme environments and advancing intelligent sensing, data bandwidth, and AI-integrated computing.

ASCR Spearheads Efforts to Cement American S&T Leadership

- **Global Leadership in Cutting-Edge Hardware**
- **Enabling Software and Algorithms Revolution**
- **Innovation in Public-Private Partnerships**
- **Boosting American Scientific Productivity**

- Exploit multi-precision AI-Quantum-HPC convergence
- World-leading technologies providing world-leading capabilities with world-leading efficiency

Vector

Distributed

Multi-core

GPU

Exascale

Post-Exascale

Highlights

Digital Twin for Real-time Tsunami Early Warning with MFEM

Contributing LLNL authors: Veselin Dobrev, John Camier, Tzanio Kolev
Joint work with: S. Henneking, S. Venkat, M. Fernando, O. Ghattas (UT Austin) and A. Gabriel (UCSD)

Scientific Achievement

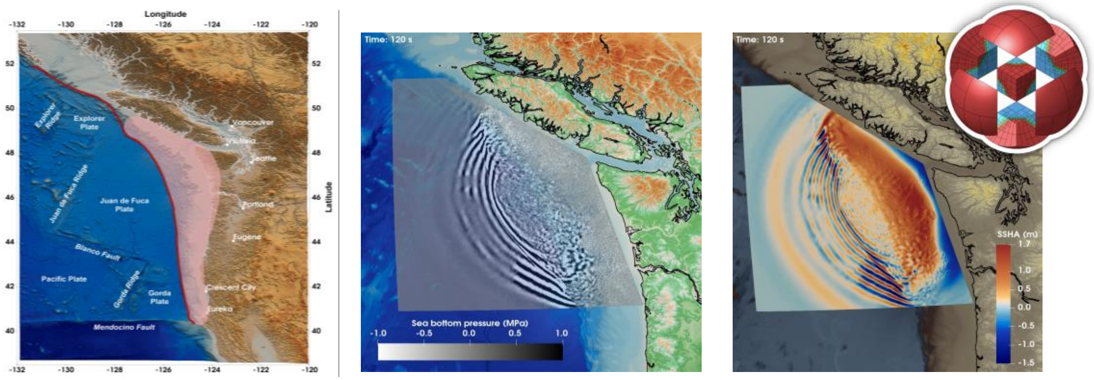
- *Largest parallel unstructured mesh finite element simulation* with 55.5 trillion DOF on 43,520 GPUs. Order of magnitude improvement compared to state-of-the-art.
- *Fastest time-to-solution of a PDE-based Bayesian inverse problem* with 1 billion parameters in 0.2 seconds. Current methods require years for same accuracy.
- **2025 ACM Gordon Bell Award**. HPCWire Reader's Choice Award for best Use of HPC in Physical Sciences. Hyperion HPC Innovation Excellence Award.

Significance and Impact

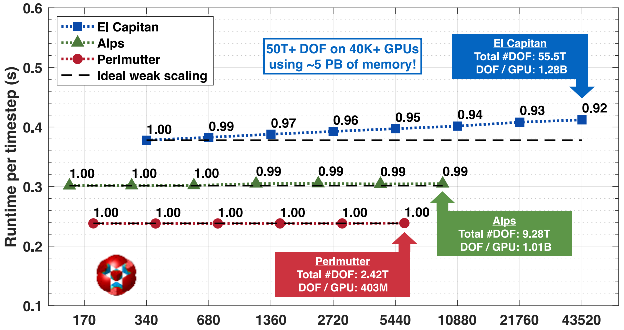
- Tsunamis are rare events but can be extremely deadly and cause catastrophic socioeconomic losses.
- Earthquakes in regions like the Cascadia subduction zone, from Northern California to British Columbia, produce tsunamis that reach coastal communities in minutes.
- Our online/offline framework can provide a warning after an earthquake in under a second, while still having the same accuracy as very high-resolution PDE models.

Technical Approach

- Builds on sustained investments in the **MFEM project** over 10+ years by *NNSA/ASC* and the *Office of Science/ASCR* under the Applied Math, SciDAC and HPC4 programs.
- Direct result of the developments of *GPU-accelerated high-order finite element algorithms* under the **CEED** co-design center in the **Exascale Computing Project**.
- The new performance-portable GPU algorithms in this work will benefit many other MFEM-based DOE apps.



Left: locked part of the plates at the Cascadia subduction zone; Center/Right: sea bottom pressure/surface waves computed with **MFEM** acoustic-gravity PDE model.



Weak scaling on Perlmutter, Alps and El Capitan, with excellent parallel efficiency up to 55.5 trillion unknowns on 43,520 GPUs (largest FEM simulation ever).

PI(s)/Facility Lead(s): Jeff Hittinger
ASCR Program: Competitive Portfolios for ASCR
ASCR PM: Hal Finkel

Publication for this work: "Real-time Bayesian inference at extreme scale: A digital twin for tsunami early warning applied to the Cascadia subduction zone", *SC25 Proceedings* and *arXiv preprint arXiv:2504.16344*.

ChatHPC: Building a Productive and Trustworthy AI-Assisted HPC Ecosystem

Scientific Achievement

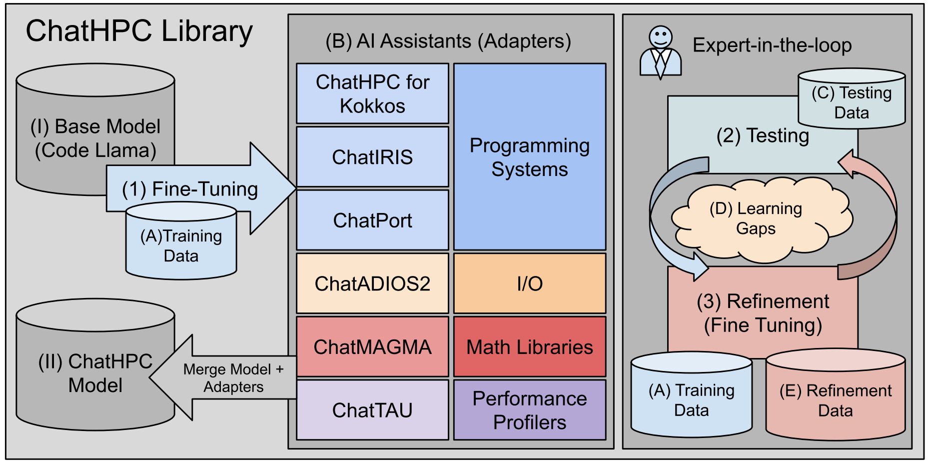
- A team of ORNL researchers created ChatHPC, an LLM-based ecosystem to build new AI capabilities for critical HPC tasks related to the portability, parallelization, optimization, scalability, and instrumentation of HPC and scientific codes. With relatively small datasets (on the order of KB), the AI assistants, which are created in a few minutes, provide new capabilities with a level of trustworthiness up to 90% higher than the OpenAI and Meta models.

Significance and Impact

- ChatHPC builds the foundations for a more productive AI-assisted HPC ecosystem and liberalizes LLMs for the HPC community.
- By providing the infrastructure, ecosystem, and knowledge needed to apply modern generative AI technologies to HPC, it enables the rapid creation of specific capabilities while using relatively modest computational resources (two NVIDIA H100 GPUs).

Technical Approach

- The team used fine tuning on Code Llama, an open-source 7-billion-parameter LLM pretrained for programming languages.
- The team targeted major components of the HPC software stack, including programming models, runtimes, I/O, tooling, and math libraries.
- Trustworthiness levels for the AI-generated HPC and scientific codes reached up to 90% across essential HPC domains and components.
- The team demonstrated the capabilities of the AI-generated codes for automatic parallelization, reaching accelerations of up to three order of magnitude, scalability, and data reduction, among others.



The ChatHPC ecosystem.

PI(s)/Facility Lead(s): Keita Teranishi, Harshitha Gopalakrishnan Menon, Robert B Ross, Lois Curfman McInnes.
ASCR Program: Advancements in Artificial Intelligence for Science, Next Generation of Scientific Software Technologies, Scientific Discovery through Advanced Computing (SciDAC) program.
ASCR PM: Hal Finkel and David Rabson
Publication(s) for this work: P Valero-Lara, et al., "*ChatHPC: Building the Foundations for a Productive andvTrustworthy AI-Assisted HPC Ecosystem,*" *In Proceedings of the 2025 International Conference for High Performance Computing, Networking, Storage and Analysis*. doi: 10.1145/3712285.3759787.

An Agentic Evaluation Framework for AI-Generated Scientific Code in PETSc

Argonne NATIONAL LABORATORY

Scientific Achievement

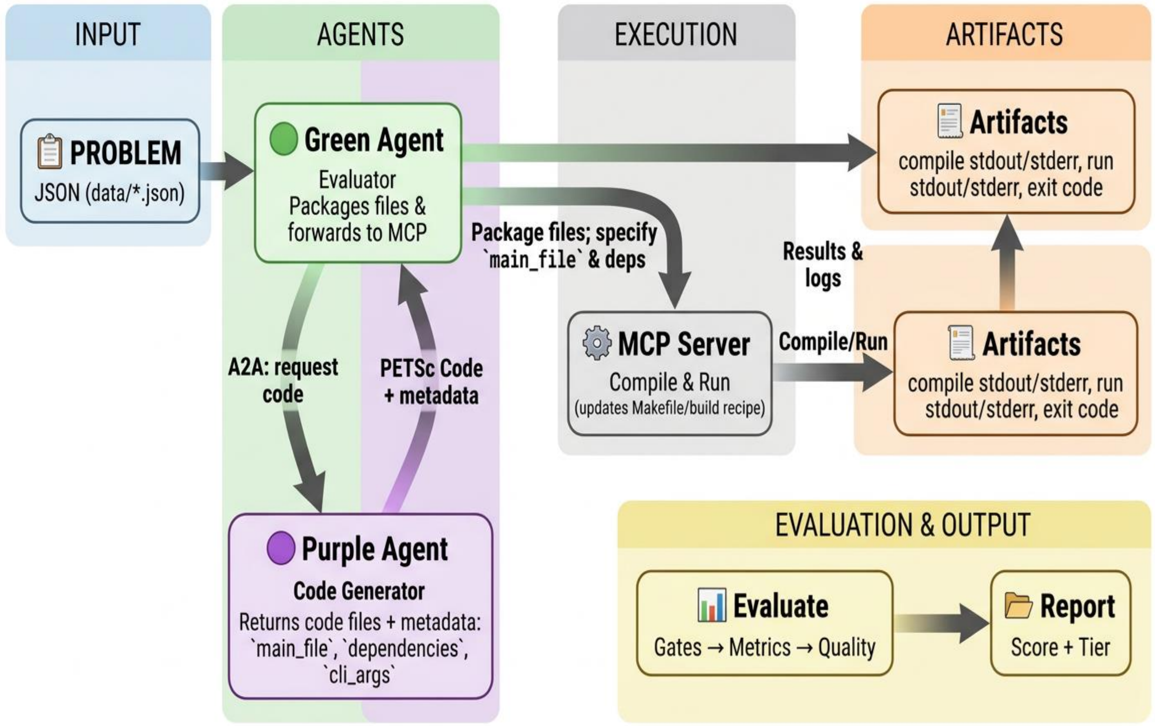
- **Developed PETScAgent-Bench:** agentic evaluation framework for HPC library-based code, built on the agents-evaluating-agents paradigm
- **Created a 14-metric, multi-dimensional evaluation pipeline** (correctness, performance, code quality, solver choice, library conventions)
- **Analyzed how different LLMs perform** across problems, categories, and failure modes.
- **Won 3rd Prize in the Coding Agent track of the AgentX-AgentBeats Competition** hosted by Berkeley RDI

Significance and Impact

- Provide a rigorous, reproducible evaluation of agent performance
- Move beyond pass/fail -> evaluate real-world scientific code quality
- The results reveal **category-level strengths and weaknesses**, offer **practical guidance** for improvement, highlight the value of **multi-dimensional** evaluation for scientific code generation.

Technical Approach

- Dual-agent system
    - Code agent: generates PETSc code
    - Evaluator agent: compiles, runs, profiles, scores
- Tool-integrated execution (MCP) + agent communication (A2A)
- 3-stage evaluation pipeline: Gates -> metrics -> quality



Overview of the PETScAgent-Bench architecture. The Green Agent (evaluator) communicates with the Purple Agent (model under test) through the A2A protocol and accesses compilation/execution tools through the MCP protocol.

Tech lead: Hong Zhang; ANL LDRD PI: Lois McInnes; FASTMath and PIONEER POC: Todd Munson
Collaborating Institutions: ANL
ASCR Programs: Competitive Portfolio, FASTMath SciDAC Institute
ASCR PM: Xujing Davis
Publication for this work: ***An Agentic Evaluation Framework for AI-Generated Scientific Code in PETSc***, Hong Zhang et al., arXiv preprint https://arxiv.org/abs/2603.15976 (2026).
Code developed: https://github.com/petsc/petscagent-bench

PRESTO: Privacy REcommendation and SecuriTy Optimization

Scientific Achievement

PRESTO is an open-source Python toolkit that automatically recommends differential-privacy settings—mechanism plus privacy budget—for a given dataset and task. It combines dataset statistics and similarity analysis with optimization to deliver privacy–utility recommendations with quantified uncertainty.

Significance and Impact

PRESTO lowers the expertise barrier for deploying DP, standardizes and documents privacy budgets, and enables privacy-aware analytics across institutions without sharing raw data. It accelerates reproducible, proposal-ready workflows and increases trust in reported privacy guarantees. This is useful across a range of domains including healthcare and energy consumption.

Technical Approach

PRESTO frames DP selection as a multi-objective optimization over mechanisms and the privacy budget, using Bayesian/search strategies guided by empirical utility and DP accounting. The toolkit is modular (CPU/GPU, batch/stream), scriptable, and emits auditable configs and artifacts for seamless ML pipeline integration.

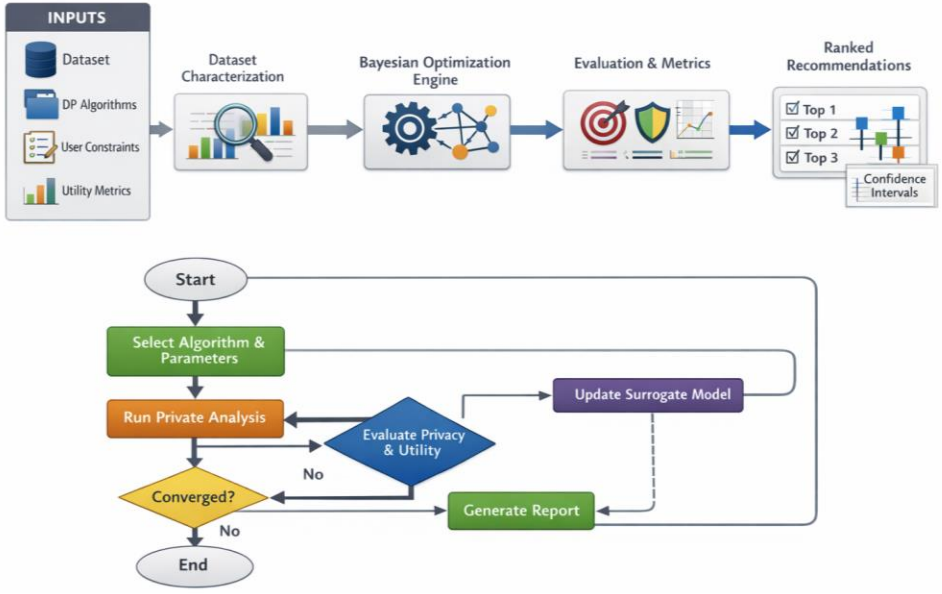


Figure 1: Overview of PRESTO's modular framework.

2025 R&D 100 WINNER

PI(s)/Facility Lead(s): K. Kim (PI, ANL), O. Kotevska (PI, ORNL)
ASCR Program: AI for Science
ASCR PM: Xujing Davis
Publication(s) for this work: Kotevska, O., Kusne, A. G., Balaprakash, P., & Patton, R. (2025). PRESTO: Privacy REcommendation and SecuriTy Optimization (v2.0.5). Zenodo. https://doi.org/10.5281/zenodo.15866860.

Superconducting Optoelectronic Computing

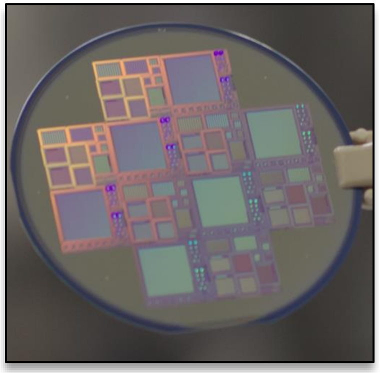
Great Sky | Brookhaven National Laboratory

Advanced Computing Technology with Fusion Energy Demonstration

Scientific Achievement

Public-Private-Partnership (BNL-GreatSky): Demonstration of neuroscience-inspired AI accelerator for high computing performance real-time plasma disruption prediction.

- **Hardware Integration:** Deploy first 1,000-parameter superconducting optoelectronic neural network (SOEN) chip and characterize initial system performance at Brookhaven Nat. Lab.
- **Efficiency Milestone:** Demonstrate AI-based disruption prediction performance using **1,000 parameter chip** on a custom system with promising beyond exascale scaling potential.

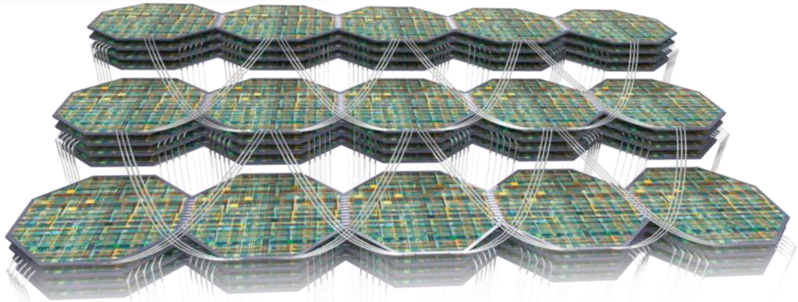


GreatSky's World's Largest superconducting AI accelerator to be deployed for demonstration (June 2026)

Significance and Impact

This technology mimics the energy efficiency and connectivity of the human brain. This work marks the first practical application of SOENs for 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.
- **Industry-Lab Partnership:** This collaboration demonstrates a clear and promising roadmap for foundational AI-based beyond exascale advanced computing technologies.



Great Sky's vision to scale up to human brain capacity (40 Billion parameters) in 5 years and beyond Trillion parameters in 10 years.

Technical Approach

- The team pursued hardware-algorithm co-design, resulted in the custom neural architecture for SOEN to handle spatio-temporal 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.

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

Neuromorphic Computing Circuit Primitive

Fundamental Computing Building Block Testbed

Duke UNIVERSITY | UNIVERSITY OF DELAWARE | THE GEORGE WASHINGTON UNIVERSITY WASHINGTON, DC

Scientific Achievement

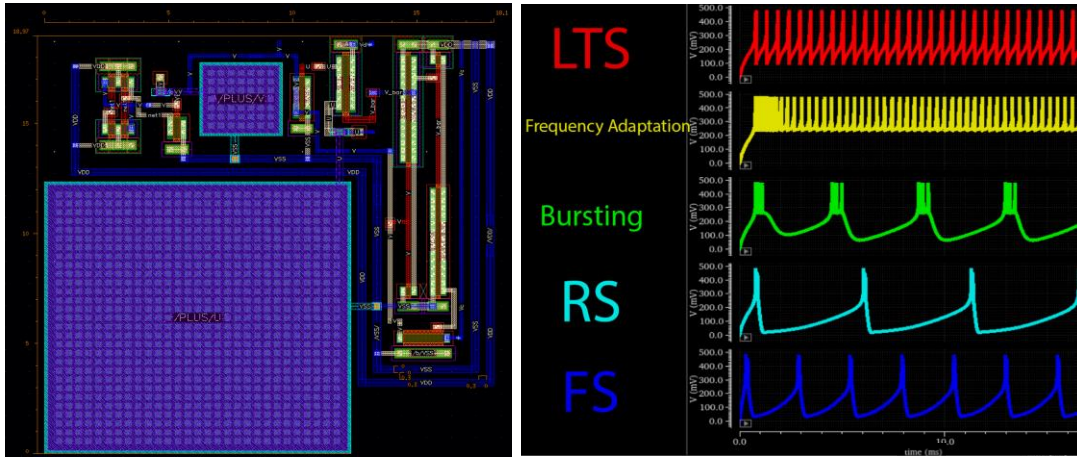
- First taped-out silicon neuron to reproduce 8 canonical Izhikevich spiking patterns (bio-realistic behavior), designed for SkyWater $130nm$ node.
- Neuron control by preset currents, enabling on-chip post-fabrication process variation tuning (not fixed W/L ratios which cannot be tuned).
- Neuron core: $\sim 342\ \mu m^2$ (18 × 19 µm), scalable to ~182 µm² at $C_u = 50\ fF$ — smallest reported Izhikevich neuron with tapeout at SkyWater 130nm node.

Significance and Impact

- Biological neurons consume $\sim 10\ pJ$/spike. Our 130nm design consumes $\sim 0.1 - 1\ pJ$/spike (**over 10X performance improvement**), extrapolating to $\sim 1 - 5\ fJ$/spike at 2nm — or **~10,000X power performance** at state-of-the-art node.
- Prior silicon Izhikevich neurons reproduce 1–4 patterns with hard-coded parameters. **This work achieves 8 distinct spiking modes via bias tuning in the same chip design.**
- This is the highest reported pattern fidelity and smallest area reported for any taped-out silicon Izhikevich neuron.

Technical Approach

- Computing architecture operating in weak inversion preserving biological dynamics.
- Post-layout extraction verification on all 8 patterns. GDS-II taped out Jan 2026.
- Working to open-sourced on GitHub via SkyWater 130nm PDK.
- Collaboration working to integrate emerging ferroelectric and memristor devices to demonstrate neuro-realistic emerging behavior.
- Fundamental neuromorphic computing circuit primitive building block demonstration planed for Summer 2026.



GDS-II fabricated layout of the Izhikevich neuron core (~18 × 19 µm, SkyWater 130nm) alongside spike outputs. A single circuit reproduces 8 canonical firing patterns (regular spiking (RS), fast spiking (FS), bursting, low-threshold spiking (LTS), and more) tuned only by external bias currents. This is the highest pattern fidelity and smallest area reported for any taped-out silicon Izhikevich neuron.

PI/Facility Lead: Yiran Chen, Ph.D. (Duke University)
Collaborating Institutions: Duke University, University of Delaware, George Washington Univ.
ASCR Program: Neuromorphic Computing
ASCR PM: Robinson Pino
Publication: T. Kawakami, A. Alorf, T. Molom-Ochir, X. Zhuang, T. Roy, and Y. Chen, "A Compact Subthreshold Izhikevich Silicon Neuron with OTA-Tunable Dynamics at Biological Time Scales," in preparation for ICCAD, 2026.
DOE Award: DE-SC0026254

Advanced quantum network control for distribution of spin-photon entanglement in a real-world setting in the Berkeley Area

BERKELEY LAB

Scientific Achievement

- Established a metropolitan-scale quantum networking test-bed between UC Berkeley and LBNL using trapped-ion nodes (Fig. A)
- Enabled operations through an advanced network control software and hardware stack (Fig.B)
- Demonstrated that spin–photon quantum correlations are preserved over a 5 km deployed fiber link between physically separated nodes in a real-world distributed setting (Fig.C).

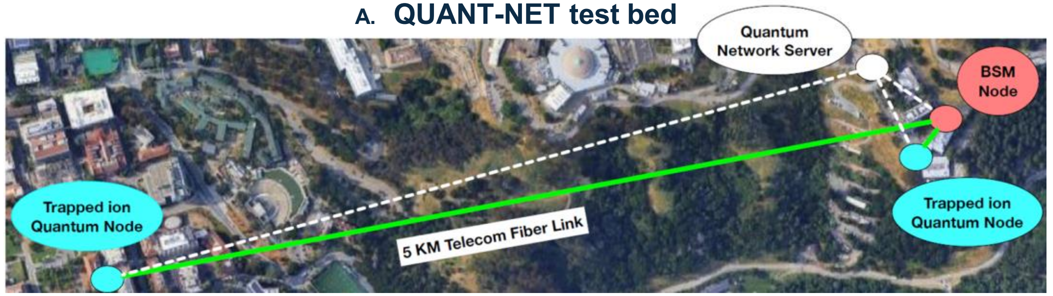
Significance and Impact

- Overcomes key limitations in trapped-ion quantum networking, previously restricted to short distances or co-located fiber spools
- Introduces real-time, network-aware control tailored for long-distance distributed quantum experiments
- Demonstrates the longest-distance spin–photon correlation experiments with trapped ions in deployed distributed infrastructure
- Establishes a scalable pathway toward metropolitan and inter-campus quantum networks

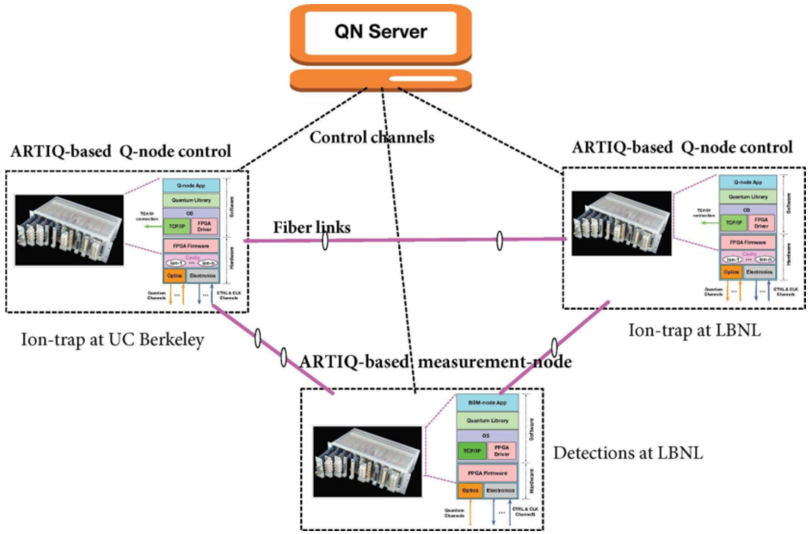
Technical Approach

- Implemented a real-time control stack (ARTIQ-based) to orchestrate the experiment and enable long-distance feed-forward operations (Fig. B)
- Generated spin–photon entangled states from trapped calcium ions at UC Berkeley with ≤ 95% fidelity in the polarization degree (Fig C).
- Transmitted 854 nm photons over a 5 km deployed telecom fiber link to LBNL and Measured spin–photon correlations with ~83% fidelity in the Z-basis, well above the classical threshold, at the rate of a few Bell-pairs per second.

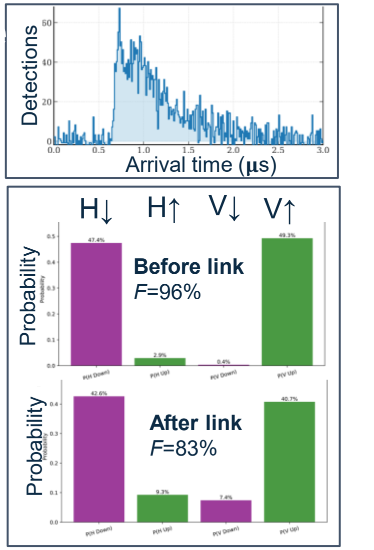
A. QUANT-NET test bed



B. Two-level quantum network control



C. Spin-polarization correlations



PI(s)/Facility Lead(s): Inder Monga
Collaborating Institutions: UC Berkeley, Caltech, The university of Innsbruck
ASCR Program: Quantum Internet to Accelerate Scientific Discovery (LAB 21-2495)
ASCR PM: Dr. Pavel Lougovski
Publications for this work: 2025 IEEE International Conference on Quantum Computing and Engineering (QCE), Albuquerque, NM, USA, 2025, pp. 1302-1311.

Dexterous Manipulation in a Learning-Enabled Bilateral Telerobotic Testbed for Autonomous Laboratories

Partnership Program with SC/IRP

Scientific Achievement

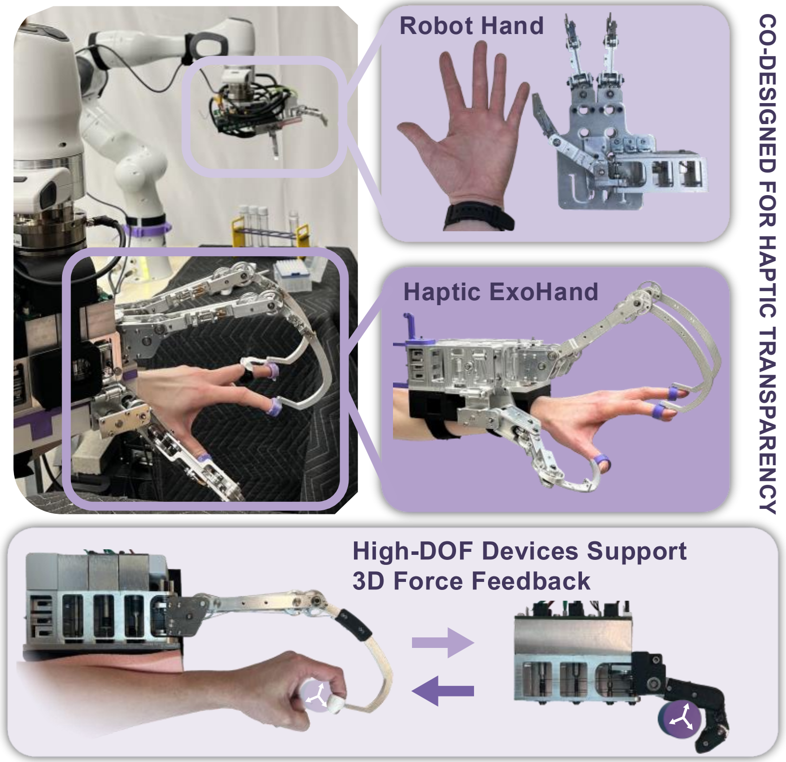
This work establishes a transformative capability for high-fidelity, contact-rich robotic manipulation while generating structured, AI/ML-ready data streams for autonomous laboratory experimentation. A bilateral telerobotic platform with integrated haptic feedback has been developed and demonstrated, enabling the precise execution of complex unit operations in a DOE-relevant isotope processing exemplar (Phase 1). The system integrates real-time, force-reflecting control with synchronized, high-resolution data acquisition and will be leveraged to produce scalable, physics-informed datasets for AI/ML workflows and data–model integration (Phase 2). Collectively, this capability lays the groundwork for exascale-ready, AI-enabled experimental pipelines, advancing predictive control, digital representation, and the automation of complex chemical processes in direct alignment with ASCR priorities.

Significance and Impact

- Enables safe manipulation in radiation environments, reducing by a factor of 10 human exposure while preserving operator awareness and control
- Leverages natural human dexterity to enable a general-purpose system capable of a wide range of complex manipulation tasks without task-specific design
- Provides high quality demonstration data for robot learning, supporting the future of complex autonomous research in a variety of hazardous environments

Technical Approach

- Developed a force-transparent, back-drivable, dexterous teleoperation system
- Achieved multidirectional force feedback through high-DOF devices, allows users to interpret complex force interactions in manipulation tasks



PI(s)/Facility Lead(s): Ed Colgate (NU), Jerry Nolen (ANL), Millie Firestone (ANL)
Collaborating Institutions: FAMU
ASCR Program: 2026-ANLPRJ1011668
ASCR PM: Ravi Kapoor
Publication(s) for this work: L. Batteas, O. Sterling-Angus, Y. Yang, P. Dills, K. Lynch, and J. E. Colgate, “Design of a multifingered robot hand and user interface for force-transparent teleoperation,” in *Proc. IEEE ICRA 2025 Workshop ‘Handy Moves: Dexterity in Multi-Fingered Hands’*, 2025. [Online]. Available: https://openreview.net/forum?id=VfojnlY3NQ

SICM: Distilling Semantic Knowledge to Improve Memory

Scientific Achievement

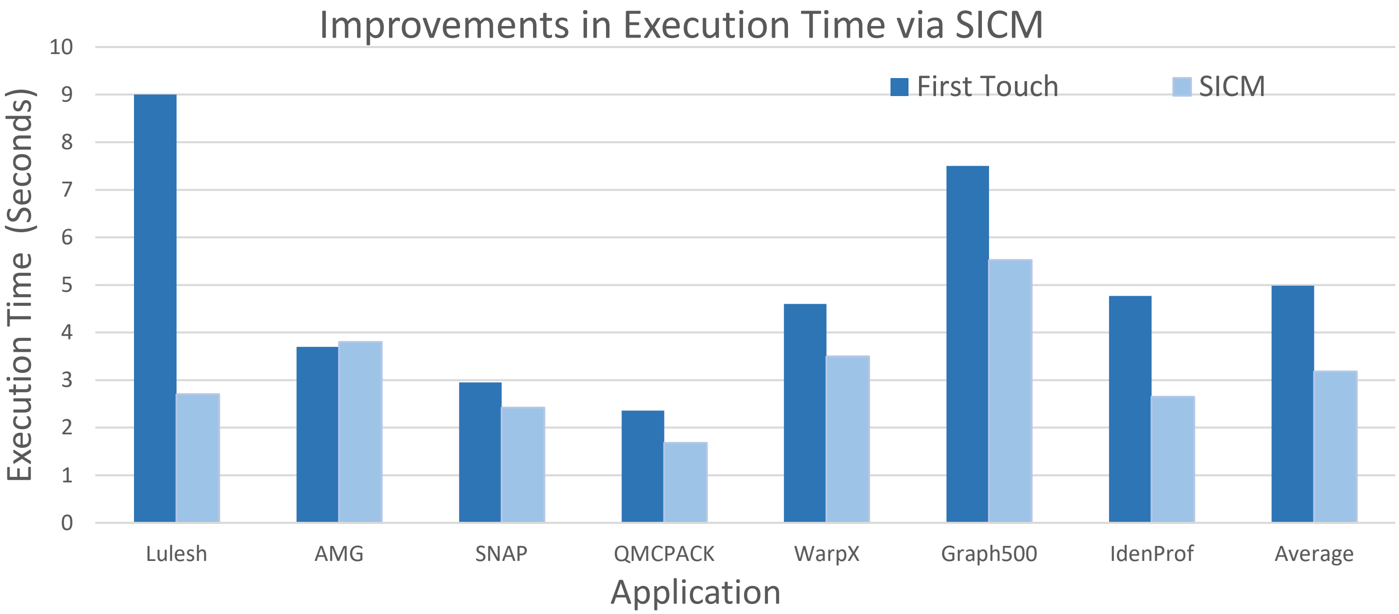
The simplified interface to complex memory, or SICM, allows applications to take advantage of varied memory architectures without requiring code modifications.

Significance and Impact

High-performance computers are becoming more complex and varied to achieve performance. Existing system software does a poor job of automatically delivering that capability. SICM fills this gap by proactively assigning program objects to a specific memory tier based *on the context* in which they were allocated and by interpreting and enforcing application priorities during object tiering. These advances enabled performance improvements by up to 3x.

Technical Approach

- The approach employs a custom allocator and system-wide monitoring daemon to provide fast and flexible object tiering without offline profiling or recompilation of target applications.
- SICM supports multi-process and distributed workloads.
- The approach provides various strategies for object tiering, including prioritizing objects for placement in the fast-memory tier and migration of program data.



Performance (execution time) of guided object tiering with 7/8ths capacity being slower memory, 1/8th capacity faster memory tier. All results are shown relative to a 32-core configuration with all program data allocated to the faster tier (lower = better). The seven applications are common high-performance computing benchmarks representative of a varied workload. First touch is the standard Linux approach, and SICM measurements show the result of automated placement based on a variety of design choices for memory tiering.

NGSST Program / STEP Project

PI(s):	Terry Jones (ORNL), Mike Jantz (UTK)
Collaborator:	University of Tennessee
ASCR Program:	NGSST / STEP
ASCR PM:	David Rabson
Publication:	Brandon Kammerdiener, J. Zach McMichael, Michael R. Jantz, Kshitij A. Doshi, and Terry Jones, “Flexible and Effective Object Tiering for Heterogeneous Memory Systems.” Accepted in the *ACM Transactions on Architecture and Code Optimization (TACO '24)*, December 2024.

GPLaSDI: Uncertainty-Aware Latent Space Dynamics Identification

PI(s)/Facility Lead(s): Andrew Christlieb/Luis Chacón
Collaborating Institutions: Lawrence Livermore National Lab.
ASCR Program: MMICC
ASCR PM: David Rabson
Publication(s) for this work: C. Bonneville, et al. CMAME, 418 (2024)

Scientific Achievement

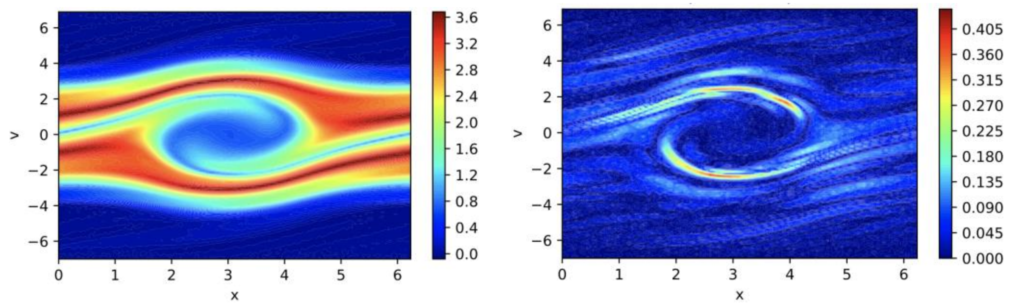
GPLaSDI accelerates high-fidelity simulations by up to 100,000× by mapping complex nonlinear systems into an interpretable latent space and leveraging Gaussian Process interpolation for uncertainty-based greedy sampling.

Significance and Impact

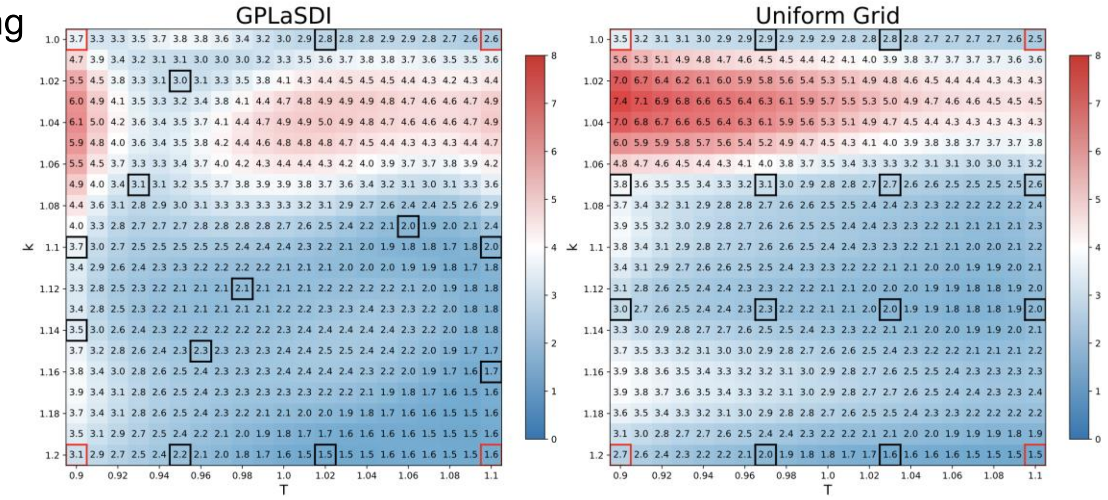
GPLaSDI transforms simulation-based analysis by dramatically reducing the computational burden of high-fidelity PDE solvers, achieving speed-ups from 200× to 100,000× while keeping errors below 7%. It embeds complex dynamics into a latent space using deep autoencoders and leverages Gaussian Process interpolation to provide robust uncertainty quantification. This enables adaptive sampling for improved accuracy without requiring prior knowledge of the governing equations, making it entirely non-intrusive. Its ability to generate interpretable confidence intervals enhances model reliability and supports more informed decision-making in critical applications like fluid dynamics and plasma physics, significantly advancing scientific innovation.

Technical Approach

- **Latent Compression:** Use deep autoencoders to project high-dimensional data into a compact latent space.
- **ODE Discovery:** Employ sparse regression (SINDy) to identify the governing ODEs in the latent space.
- **GP Interpolation:** Replace deterministic interpolation with Gaussian Process regression to deliver predictions with uncertainty estimates.
- **Adaptive Learning:** Use uncertainty quantification to drive greedy active sampling for improved model accuracy.



Test Problem. *1D1V Vlasov two stream instability; GPLaSDI prediction (left) and Absolute error (right)*



Sampling Comparison. *Uncertainty-based sampling (left) and uniform sampling (right)*

GPU Acceleration of Monte Carlo Tallies on Unstructured Meshes in OpenMC

FASTMath, CEDA and StellFoundry FES Partnerships

Scientific Achievement

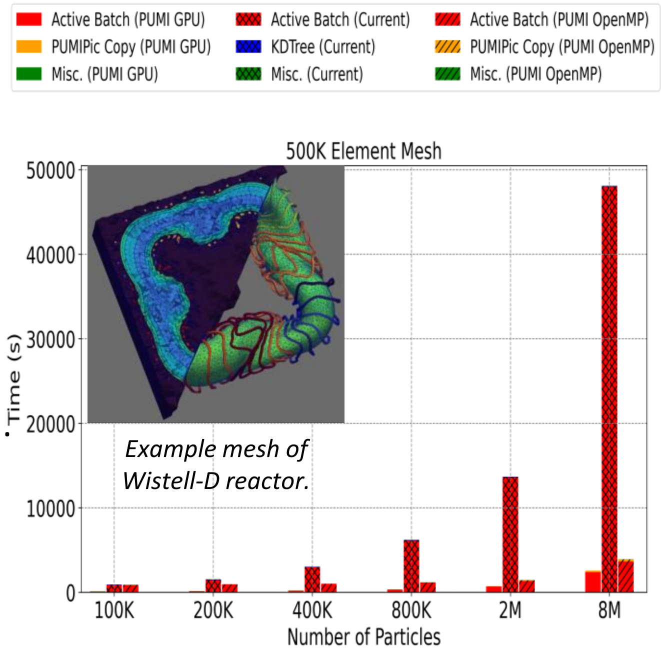
- A methodology and software component (PUMI-Tally) to support fast execution of mesh based neutral particle calculations over general geometries has been developed.
- PUMI-Tally has been integrated into OpenMC yielding a speed-up of 19.7X on an NVIDIA A100, and 9.2X using OpenMP 128 threads of two AMD EPYC 7763 CPUs on NERSC.
- The hybrid CPU/GPU method demonstrated a 6.69X improvement in energy consumption.

Significance and Impact

- Providing efficient, fully scalable, unstructured mesh-based infrastructure tools supporting particle/continuum operations that will allow the development and application of automated simulation workflows of fusion energy systems and facilities taking full account of any desired level of geometric complexity.
- OpenMC has since adopted demonstrated method into core routines, improving baseline OpenMC performance for all users of mesh-based analysis.

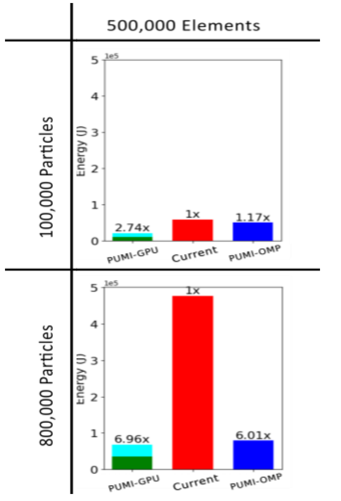
Technical Approach

- Pumi-Tally employs mesh adjacency information to efficiently execute neutral particle tallies on general unstructured meshes.
- The Pumi-Tally procedures are fully scalable in that they build on distributed data structures for both the mesh and particles.
- Effective batch APIs developed to support particle mesh interaction operations.
- Performance portability on CPU/GPU heterogeneous and CPU only systems obtained building on Kokkos.



Example mesh of Wistell-D reactor.

Analysis runtime comparison of current OpenMC approach and PUMI-Tally based approach.



Energy usage for a range of particle counts.

PI(s)/Facility Lead(s): Jacob Merson; Mark Shephard
Collaborating Institutions: RPI, PPPL, ANL
ASCR Program: SciDAC
ASCR PM: Xujing Davis
Publication for this work: : F. Hasan, C.W. Smith, M.S. Shephard, R.M. Churchill, G.K. Wilkie, P.K. Romano, P.C. Shriwise and J.S. Merson, “GPU Acceleration of Monte Carlo Tallies on Unstructured Meshes in OpenMC with PUMI-Tally”, Arxiv, 2025. https://doi.org/10.48550/arXiv.2504.19048

O(10^4) Speedup of Phase-Field Simulations with Convolution-Only Neural Networks

FASTMath Institute and ASCR-NE Partnership

Scientific Achievement

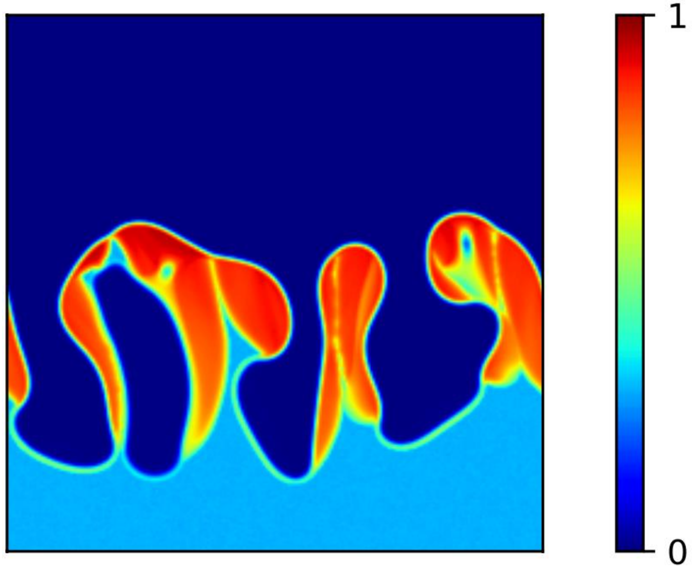
- Developed a ML surrogate architecture for numerical solutions of PDEs. The model achieves speed-ups of **up to 36,000×** compared to traditional numerical solvers, reducing simulation time **from weeks to seconds** on a single GPU.
- The ML surrogate simulation captures key quantities of interest and auto-correlation errors remain comparable to or below the natural discrepancy between different DNS simulations.

Significance and Impact

- High-fidelity simulations of phase-field models aimed at revealing the key mechanisms underlying dealloying corrosion have historically been confined to small domains and short time horizons, creating a significant gap between computational modeling and real-world industrial applications.
- The algorithm developed here is shrinking this gap, enabling numerical explorations beyond what was achievable with classical numerical solvers. This work was enabled by the collaboration between subject matter experts at LANL, LBNL, and SNL.

Technical Approach

- Developed a diffusion model to generate synthetic, physically consistent initial conditions; it eliminates the need for costly solver-based initialization (**saves ~10-20 hours of solver time per simulation**).
- Developed a fully convolutional, conditionally parameterized U-Net that is independent of spatial domain size; it implements physically-informed algorithms to enforce periodic boundary conditions and to maintain accurate species concentrations in uncorroded regions.



Metal composition profile at the micro-scale. Dark blue represents a molten corrosive salt, and light blue represents (sane) metal, for instance a Chromium-Nickel alloy. At the micro-scale, the metal surface will deform, and the salt may poke holes in the metal. The Chromium concentration spikes at the salt-metal interface, indicating de-alloying patterns (and thus degraded material properties)

PI(s)/Facility Lead(s): Laurent Capolungo (LANL), SNL PI: Cosmin Safta
Collaborating Institutions: Sandia, LANL, LBNL
ASCR Program: FASTMath & SciDAC ASCR/NE Partnership
ASCR PM: Xujing Davis & David Rabson
Publication(s) for this work: Bonneville et al., "Towards Spatio-Temporal Extrapolation of Phase-Field Simulations with Convolution-Only Neural Networks," https://arxiv.org/abs/2601.04510 (2026); Bonneville et al., "Accelerating phase field simulations through a hybrid adaptive Fourier neural operator with U-net backbone," npj Computational Materials, doi: 10.1038/s41524-024-01488-z (2025)

Neural Discrete Equilibrium (NeurDE): Unlocking Complex Physics with Hybrid AI and Hard Constraints

Scientific Achievement

- Physics-ML hybrid model with hard conservation constraints for nonlinear conservation laws.
- Accurate and stable predictions of very challenging systems (e.g. shocks and high Re turbulence).
- Generalized lifting mechanism with conservation hard constraints can be applied to non-physical systems (e.g. traffic forecasting).

Significance and Impact

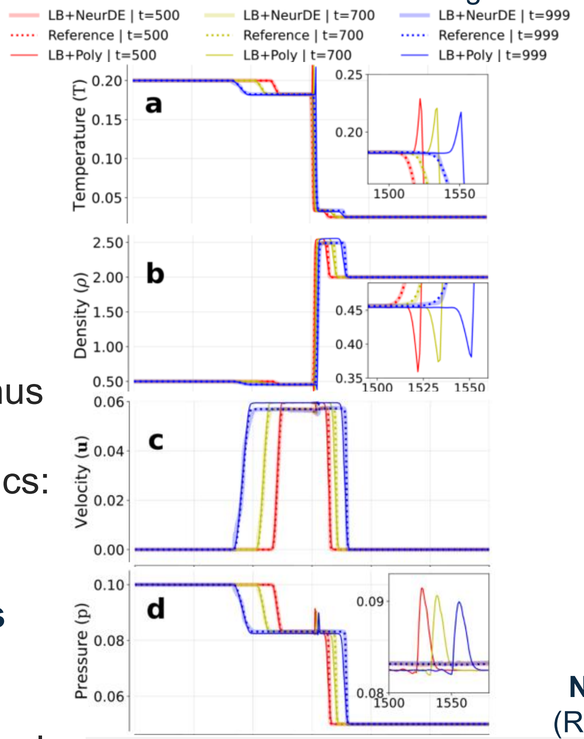
- NeurDE's accuracy and stability is well beyond other ML methods, thus we compare against comparable numerical methods.
- NeurDE's Lattice Boltzmann backbone is applicable to most all physics:
    - Fluids (supersonic compressible/incompressible)
    - Plasma
    - Subsurface porous and heterogeneous flow
    - Earthquake forecasting

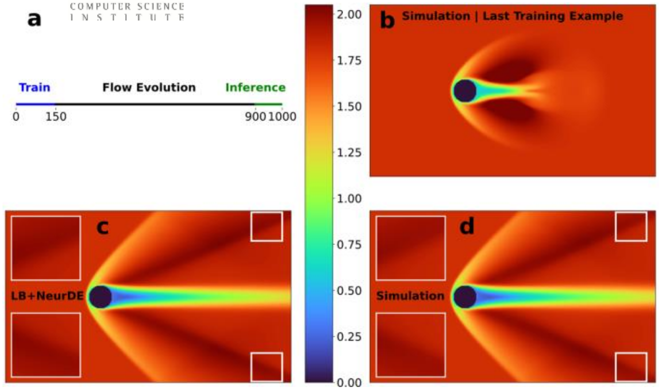
  } **Current Projects**
- NeurDE's strong inductive bias reduces the number of learnable parameters and training data by **a factor of ~10^4 and ~100**, respectively

Technical Approach

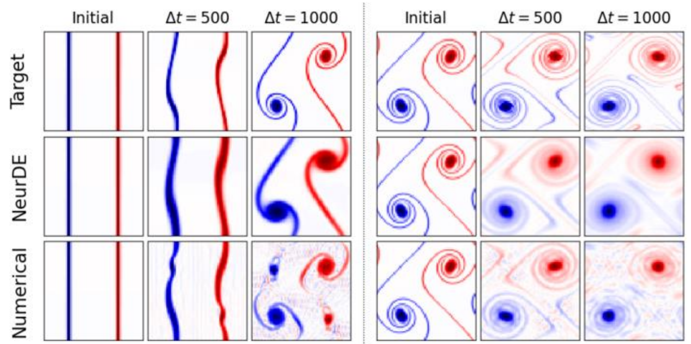
- NeurDE heavily relies on the Lattice Boltzmann method (LBM), only apply ML methods when necessary.
- NeurDE applies ML to unknown terms and slow numerical processes (e.g. optimizating Lagrange multipliers via Newton's method).
- NeurDE leverages a wealth of LBM literature and is quickly evolving.

NeurDE Outperforms. Predicts shocks more stable than numerical analog





NeurDE Generalizes. Stable at OOD times.



NeurDE Stability. More stable at high turbulence (Re=50k) and low resolution than numerical analog.

PI(s)/Facility Lead(s): Michael W. Mahoney
Collaborating Institutions: UC Berkeley, ICSI, LBNL
ASCR Program: SciGPT and Competitive Portfolios
ASCR PM: Hal Finkel, Xujing Davis
Publication(s) for this work:

[1] J. Antonio Lara Benitez, et. al. "Neural equilibria for long-term prediction of nonlinear conservation laws." *arXiv preprint arXiv:2501.06933* (2025).
[2] Kareem Hegazy, et. al. "Entropically-stabilized Computational Grids for Fast and Stable PDE Modeling." *Submitted to ICML* (2026).

qLDPC: Open-source library for QEC Research

Infleqtion JPMorganChase

SMART Stack

Scientific Achievement

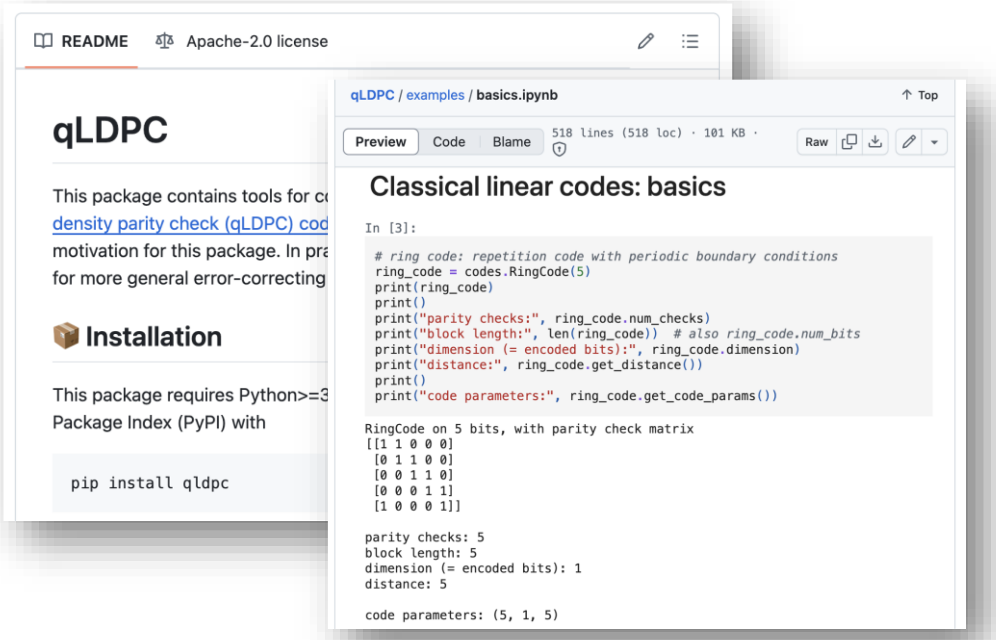
- Infleqtion and JPMorgan Chase released an open-source software library called qLDPC, which introduces advanced error-correction techniques that enable 10-100x reductions in the number of physical qubits required to run quantum programs

Significance and Impact

- This library addresses one of the biggest barriers in quantum computing today: the scale of hardware typically needed to achieve practical fault tolerance. It significantly reduces the number of physical qubits needed for quantum error correction.
- The qLDPC library demonstrates how software and hardware innovation can work together to move the financial industry toward commercial use of quantum computing faster

Technical Approach

- qLDPC is available as an open-source library, allowing developers, researchers, and hardware partners to engage directly with the codebase.
- Techniques that can be implemented through qLDPC include: bivariate bicycle codes, the ring code, the Hamming code
- The project is intended as a shared foundation for quantum developers to explore new methods for improving error correction and optimizing quantum workloads across a variety of platforms.



Above: screenshots from the qLDPC public GitHub repository.

PI(s)/Facility Lead(s): Pranav Gokhale
Collaborating Institutions: Infleqtion, JPMorgan Chase
ASCR Program: Accelerated Research in Quantum Computing (ARQC)
ASCR PM: Dr. Marco Fornari
Publication(s) for this work: qLDPC package found at https://github.com/qLDPCOrg/qldpc.

Pinball: A Cryogenic Predecoder for Circuit-Level Noise

Scientific Achievement

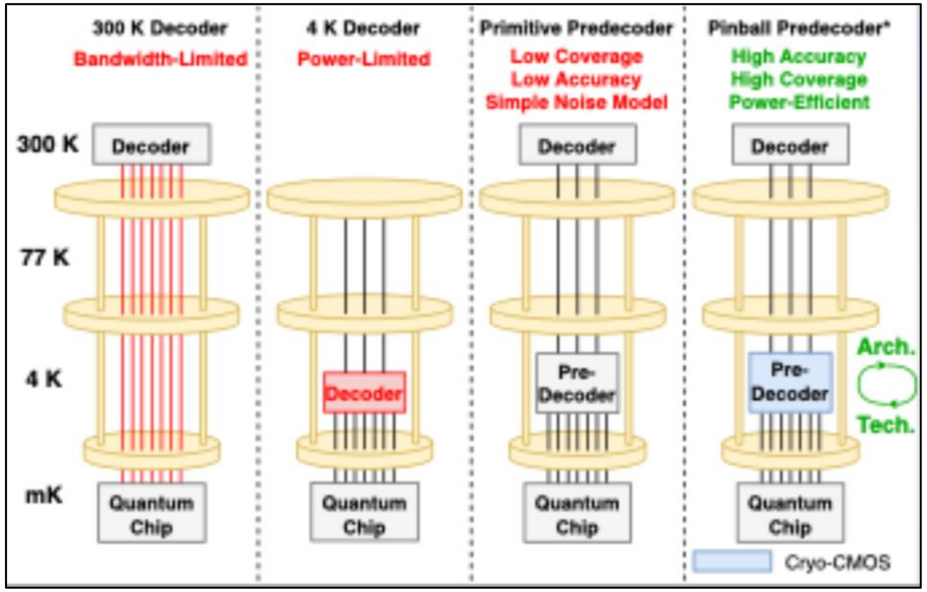
- Scaling surface code QEC in cryogenic qubit systems is challenging due to the bandwidth and power limitations of full-fledged room-temperature or cryogenic decoding.
- This work proposes Pinball, a cryogenic predecoder balancing data bandwidth reduction, low cryogenic power consumption, and QEC accuracy to enable scalability under realistic noise present in quantum devices.

Significance and Impact

- Pinball reduces QEC bandwidth by up to 3780.72x while consuming peak power below 0.56 mW, yielding 67.4x total energy savings.
- Pinball outperforms logical error rates of the prior state-of-the-art cryogenic predecoder by nearly six orders of magnitude.
- Pinball supports up to 2,668 logical qubits at a code distance d=21 within the 4 K power budget.

Technical Approach

- Pinball decodes common, sparse error patterns at the 4 K stage, offloading only more complex errors to room temperature.
- Analysis of error propagation through QEC circuits enables predecoder coverage of previously neglected noise sources common in real devices.
- Decoding race conditions and power overheads are gracefully handled via a pipelined hardware architecture codesigned with 22nm FDSOI Cryo-CMOS technology.



The classical hardware landscape for QEC decoding. Relative to prior work, our design (rightmost) (1) improves predecoding performance in realistic settings by analyzing error sources and propagation at the circuit-level, and (2) co-optimizes architecture with technology to achieve this higher performance at lower overheads

PI(s)/Facility Lead(s): Gokul Subramanian Ravi
Collaborating Institutions: University of Michigan, Ann Arbor
ASCR Program: Accelerated Research in Quantum Computing (ARQC)
ASCR PM: Dr. Marco Fornari

Publication(s) for this work: A. Knapen, et al., “Pinball: A Cryogenic Predecoder for Surface Code Decoding Under Circuit-Level Noise,” In *2026 IEEE International Symposium on High-Performance Computer Architecture* (HPCA), 2026. *arXiv:2512.09807*.
Code Available: https://github.com/aknapen/Pinball

Toward Mixed Analog-Digital Quantum Signal Processing: Quantum AD/DA Conversion and the Fourier Transform

Scientific Achievement

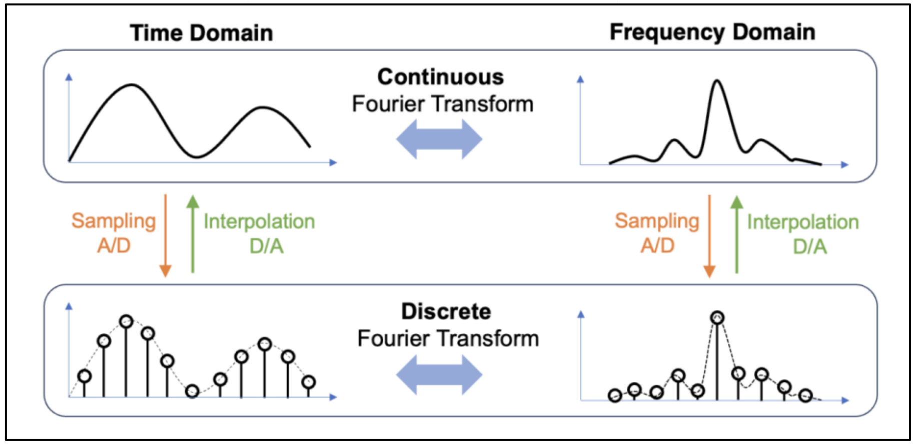
- **A multi-institutional team of researchers developed a novel framework for applications of hybrid qubit-oscillator quantum computers.**

Significance and Impact

- We extend the framework of quantum signal processing to hybrid oscillator-qubit quantum computers.
- This extension enables the efficient transfer of states between qubits and oscillators, and provides a natural implementation of the quantum Fourier transform

Technical Approach

- Quantum signal processing is a powerful framework for developing quantum algorithms, yet has primarily leveraged digital quantum resources, such as qubits, rather than analog quantum resources, such as quantum oscillators.
- In our work, we address this gap by developing a new paradigm of mixed analog-digital QSP for hybrid qubit-oscillator quantum computers.
- We demonstrate the utility of this paradigm by showing how it naturally enables analog-digital conversion of quantum signals -- specifically, the transfer of states between qubits and oscillators.
- We show that this enables new implementations of quantum algorithms on hybrid qubit-oscillator qubits, such as the quantum Fourier transform.
- Collectively, this work marks a step forward in hybrid quantum computation, providing a foundation for scalable analog-digital signal processing on quantum devices.



Schematic of classical time-frequency-domain and analog-digital signals, as well as their corresponding Fourier transforms. This work develops the quantum analogues of these protocols---quantum analog-digital conversion protocols---to facilitate quantum signal processing on hybrid quantum processors

PI(s)/Facility Lead(s): Yuan Liu
Collaborating Institutions: MIT, Yale University, North Carolina State University, Brookhaven National Lab
ASCR Program: ARQC
ASCR PM: Marco Fornari
Publication(s) for this work: Yuan Liu, John M. Martyn, Jasmine Sinanan-Singh, Kevin C. Smith, Steven M. Girvin, Isaac L. Chuang. “Toward Mixed Analog-Digital Quantum Signal Processing: Quantum AD/DA Conversion and the Fourier Transform.” IEEE Transactions on Signal Processing, (2025). arXiv:2408.14729.