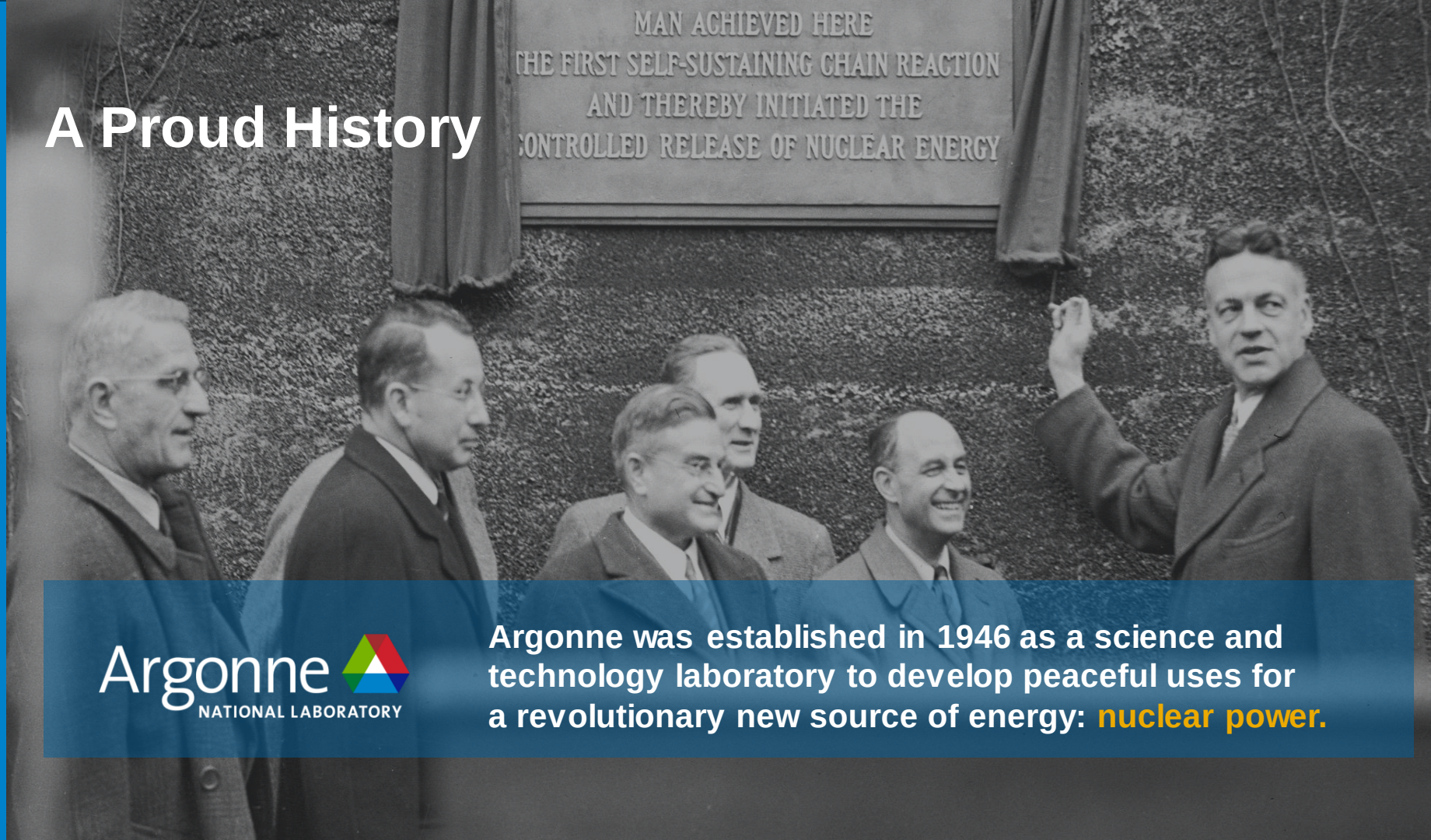
Argonne National Laboratory: Computing at Argonne

Foundation Models for the Electric Grid Workshop
February 12, 2025

A Proud History



Argonne NATIONAL LABORATORY

Argonne was established in 1946 as a science and technology laboratory to develop peaceful uses for a revolutionary new source of energy: nuclear power.

Current One-of-a-Kind Facilities

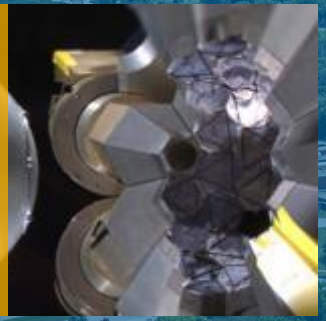
Enabling science from nanoscale to exascale

Advanced Photon Source

APS

Argonne Tandem Linear Accelerator System

ATLAS



Argonne Leadership Computing Facility

ALCF

Center for Nanoscale Materials

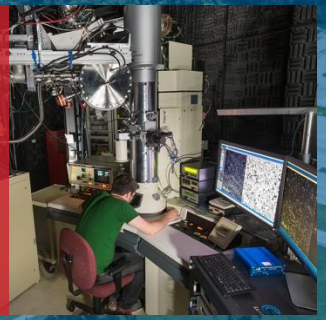
CNM

Atmospheric Radiation Measurement – The Southern Great Plains and third ARM mobile facility

ARM

Intermediate Voltage Electron Microscope

IVEM



ALCF Systems Evolution

5.7 TF

IBM BG/L
2004

557 TF

Intrepid
IBM BG/P
2007

10 PF

Mira
IBM BG/Q
2012

11.7 PF

Theta
Intel-Cray XC40
2017

15.6 PF

ThetaGPU
NVIDIA
DGX A100
2020

44 PF

Crux
HPE-AMD
+
Polaris
HPE-NVIDIA
2021

>1 EF

Aurora
Intel-HPE
2024

JLSE (2013)

AI Testbed (2020)

Edge Testbed (2021)

ALCF AI Testbed

https://www.alcf.anl.gov/alcf-ai-testbed

Cerebras CS-2

SambaNova DataScale SN30

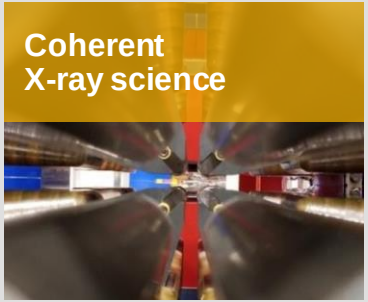
Graphcore Bow Pod64

Habana Gaudi1

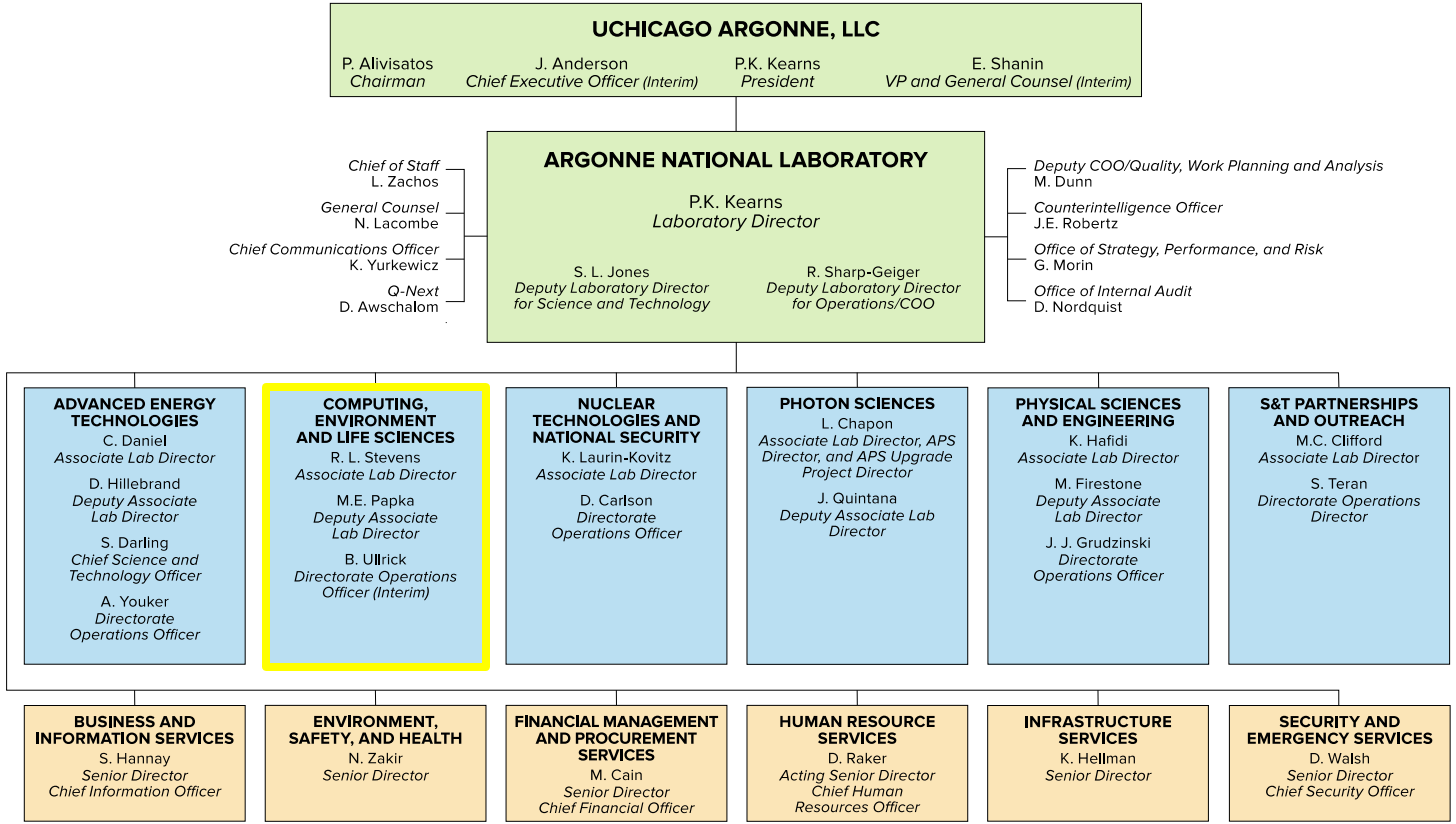
GroqRack

- Infrastructure of next-generation machines with AI hardware accelerators
- Provide a platform to evaluate usability and performance of AI4S applications
- Understand how to integrate AI systems with supercomputers to accelerate science

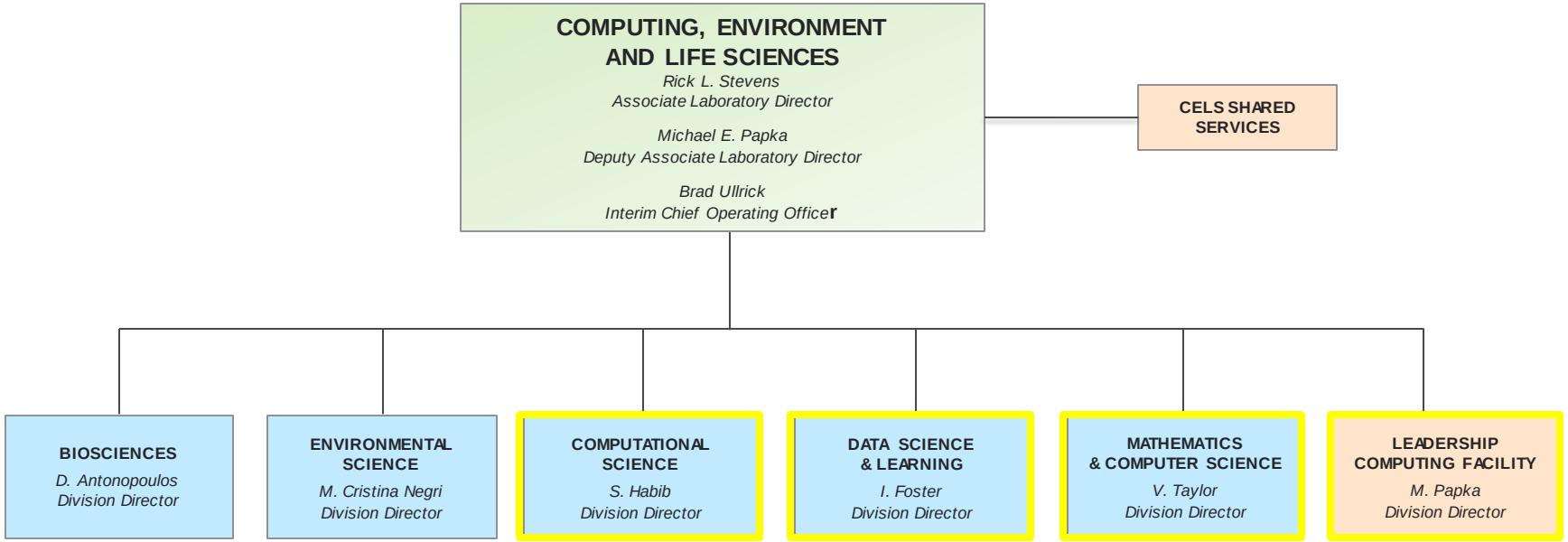
Current Strategic Initiatives



Argonne Organizational Chart



Computing, Environment and Life Sciences



Computing@Argonne

Computing@Argonne Leadership Team

Rick Stevens
ALD, CELS

Valerie Taylor
DD, MCS

Ian Foster
DD, DSL

Salman Habib
DD, CPS

Michael Papka
Deputy ALD, CELS
DD, ALCF

Sven Leyffer
DDD, MCS

Rob Ross
DDD, MCS

Rajeev Thakur
DDD, DSL

Timothy Williams
DDD, CPS

Jini Ramprakash
DDD, ALCF

Paul Hovland
Strategic Lead for
Research Partnerships

Computing Strategy: Strategic Areas (EAZQ)

Delivering on **exascale science**
(near term investments)

Creating world leading **AI** for
Science, Energy, Security
(10 year investments)

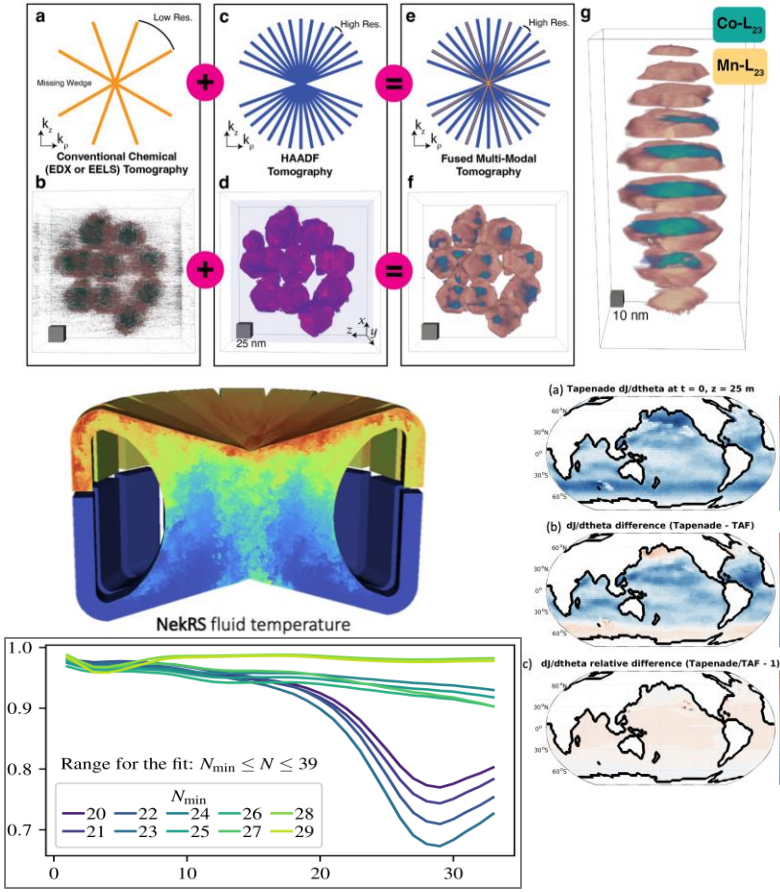
Pathways to **Zettascale**
(10-20 year investments)

Quantum computing and QIS
(20+ year investments)

Applied Mathematics, Numerical Software, and Statistics

Priority Research Directions

- **Complex DOE Applications [EAZQ] –** Effectively integrating differential and algebraic equations, complex systems, experimental data, and uncertainties (UQ) toward SciML/AI
- **Scalable Algorithms & Libraries [EAZQ] –** Solving problems modeled by ordinary/partial differential equations; exploiting spatiotemporal correlations for predictions & insights from experimental data toward energy efficiency
- **Optimization for Decision [EAZQ] –** Models, theory, and algorithms for optimal design, decision, control, inverse problems of complex systems under UQ toward digital twins
- **Math Foundations of NextGen Systems [EAZQ] –** Automatic differentiation, optimization for explainable AI/ML; control, compilers, & design of quantum systems/algorithms

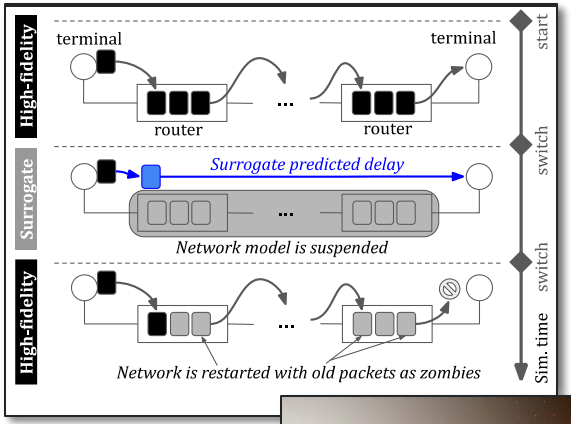


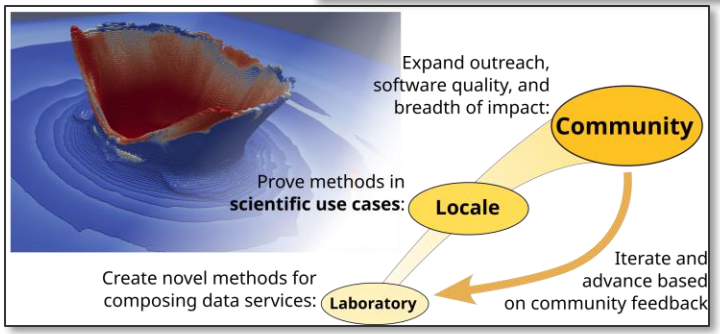
Top: Multi-modal chemistry tomography. Bottom: NekRS fusion simulation at exascale; AD of ocean-circulation model; QAOA advantage

Computer Science

Priority Research Directions

- **Transforming the Fundamentals of Computing [EAZQ] –** reimagining codesign; energy-efficient architectures and systems software; exploiting smart network and storage devices
- **Enhancing Scientific Programming and Trustworthiness in Computing [EAZQ] –** the science of scientific software; ensuring correctness in scientific computing; program synthesis for scientific computing; AI/LLM for scientific software development; FAIR data and workflows
- **Innovations in How We Represent Data [EAZQ] –** multivariate functional and AI-derived representations of data; privacy-preserving federated learning; error-bounded lossy compression of scientific data
- **Accelerating Science from Exascale to the Edge [EAZQ] –** science at the edge, in the cloud, and in the computing continuum

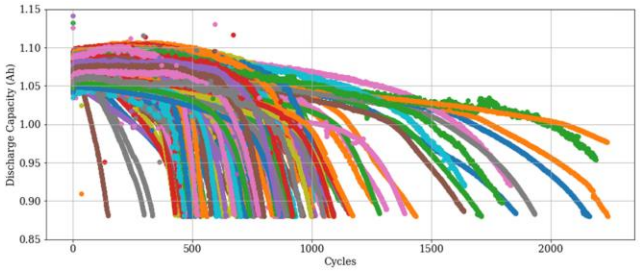




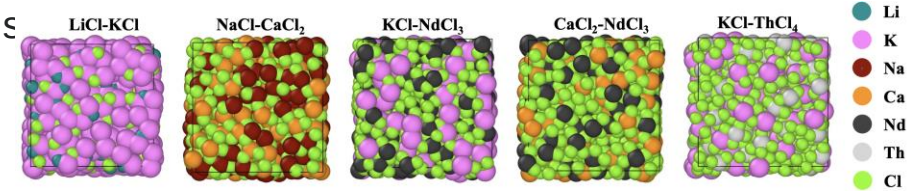
AI for Science

Priority Research Directions

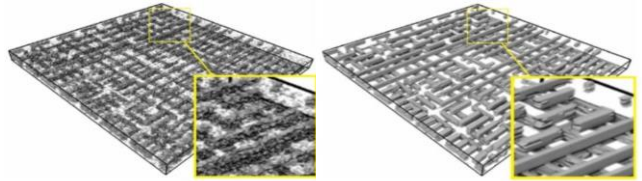
- **Pioneering new applications [EAZQ] –** Effectively incorporating scientific domain knowledge within AI methods across a broad spectrum of DOE SC problems
- **Scientific AI software systems [EAZQ] –** Making AI operational in science with composable services for AI at all scales
- **Robust foundations for AI [EAZQ] –** Understanding the capabilities and limits of current methods and developing new methods for robust, quantifiable, interpretable AI
- **Exploring the hardware basis for scientific AI [EAZQ] –** Understanding the interplay between evolving, emerging accelerators, and scientific AI performance



Battery capacity fade curve prediction

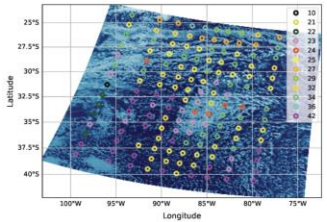


Machine-learning potentials for combinatorial molten salt mixtures



Joint ptycho-tomography with deep generative priors

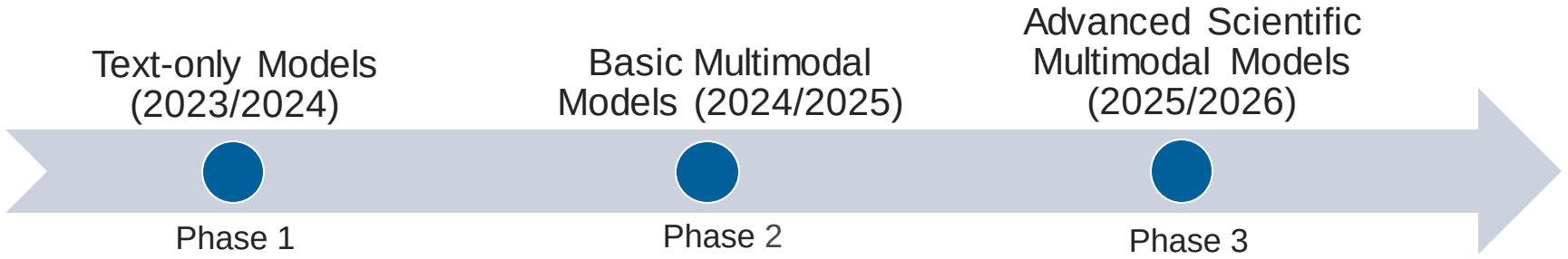
Unsupervised cloud classification



AuroraGPT: A Foundation Model for Open Science

Priority Research Directions

- **General-purpose scientific foundation model [EAZQ]**: Trained on general corpora, scientific texts, science data
- **Explore pathways [EAZQ]** toward a "Scientific Assistant"
- **Multimodal [EAZQ]**: Images, tables, equations, proofs, time series, graphs, fields, etc.
- **Exascale Computing for AI [EAZQ]:** Exploit exascale systems to speed up model training and inference



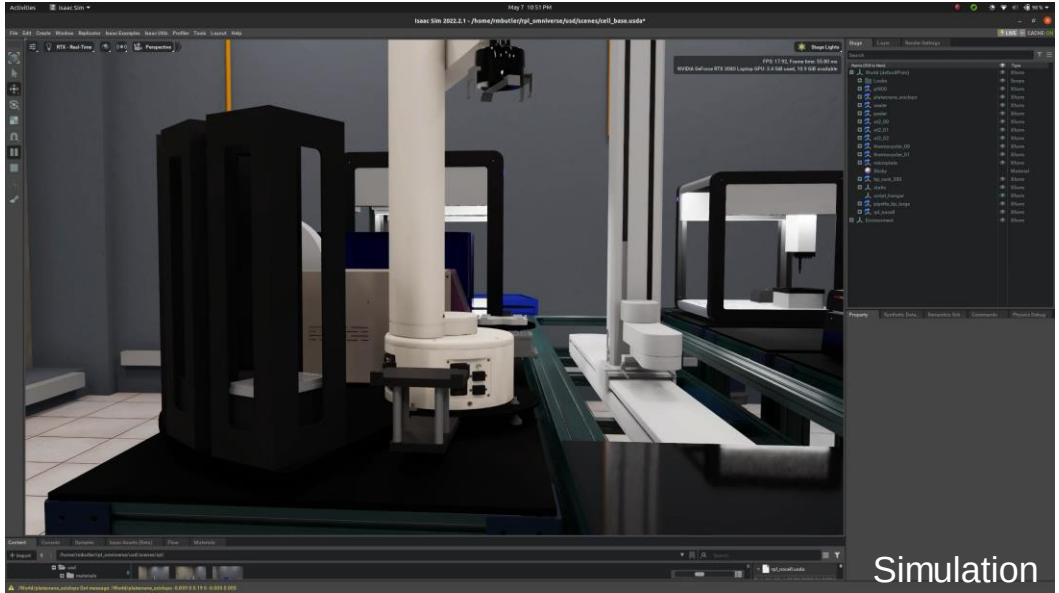
Trillion Parameter Consortium

A*STAR
AI Singapore
AIST
Allen Institute For AI
Amazon Web Services, Inc. (AWS)
AMD
Argonne National Laboratory
Barcelona Supercomputing Center
Brookhaven National Laboratory
CalTech
CEA
CSCS
Cerebras Systems
CINECA
CSC – IT Center for Science
CSIRO
Deep Forest Sciences
ETH Zürich
Fermilab National Accelerator Lab
Flinders University
Fujitsu Limited
Groq
Harvard University
HPE
Indiana University
INESC TEC
Inria
Intel
Jülich Supercomputing Center
Kotoba Technologies, Inc.
LAION
Lawrence Berkeley National Laboratory
Lawrence Livermore National Laboratory
Leibniz Supercomputing Centre
Los Alamos National Laboratory
Microsoft Research
National Center for Supercomputing Applications
National Energy Technology Laboratory
National Renewable Energy Laboratory
National Supercomputing Centre, Singapore
NCI Australia
New Zealand eScience Infrastructure
Northwestern University
NVIDIA
Oak Ridge National Laboratory
Pacific Northwest National Laboratory
Pawsey Institute
Princeton Plasma Physics Laboratory
Princeton University
RIKEN
Rutgers University
SambaNova
Sandia National Laboratories
Seoul National University
SLAC National Accelerator Laboratory
Sony Research
Stanford University
STFC Rutherford Appleton Laboratory, UKRI
Stonybrook University
SURF
Texas Advanced Computing Center
Thomas Jefferson National Accelerator Facility
Together AI
Tokyo Institute of Technology
Université de Montréal
University of Buffalo
University of California San Diego / SDSC
University of Chicago
University of Delaware
University of Illinois Chicago
University of Illinois Urbana-Champaign
University of Michigan
University of New South Wales
University of Southern California / ISI
University of Tokyo
University of Toronto / Acceleration Consortium
University of Utah
University of Virginia
University of Washington

Autonomous Discovery at Argonne

LDRD-funded projects engage scientists from CELS (ALCF, BIO, DSL, MCS) and PSE (CNM, MSD) to advance autonomous discovery

In our Rapid Prototyping Lab, we are developing and applying a modular hardware and software architecture, and engaging many students



DOE Recognitions

Lois McInnes
2024 SC Distinguished Scientist Fellow

Rob Ross
2020 Earnest Orlando Lawrence Award

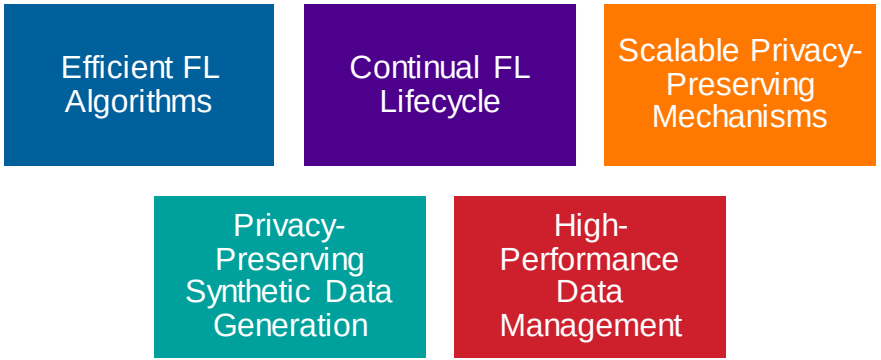
Ian Foster
2019 SC Distinguished Scientist Fellow

DOE AI FOR SCIENCE PROJECT

Privacy-Preserving Federated Learning for Science: Building Sustainable and Trustworthy Foundation Models

- **Project Lead:** Argonne (PI: Kibaek Kim)
- **Collaborating Institutions:** Brookhaven, Oak Ridge, Arizona State U., Rutgers U.
- **Objective:** Develop federated learning (FL) algorithms and AI foundation models that are sustainable, efficient, and privacy-preserving.
- **Use Cases:** Electric grid operations, scientific experiments (e.g., X-ray image science)
- **Deployment:** Running on DOE supercomputers and cloud systems for large-scale experiments.

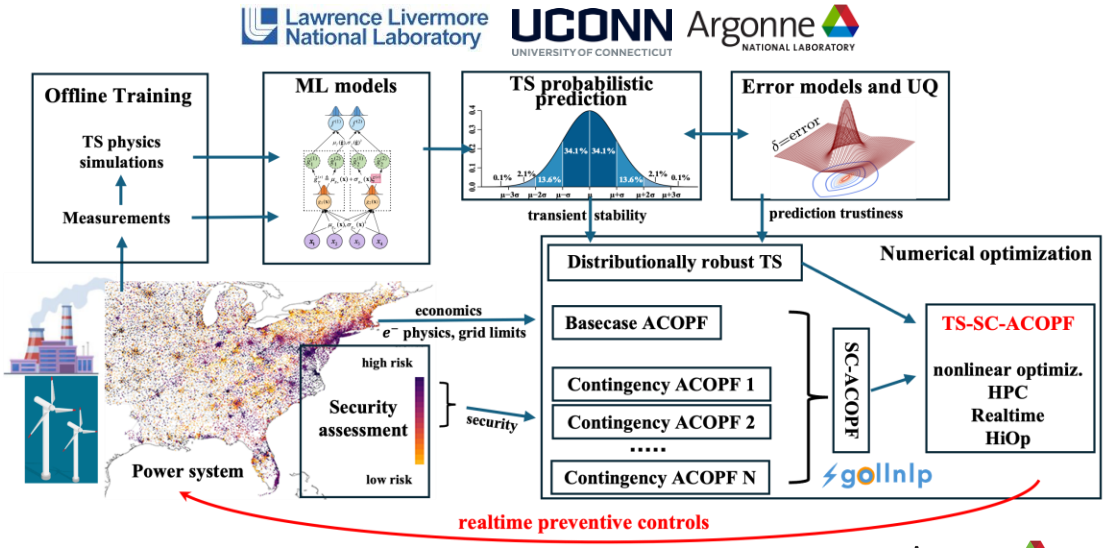
Key Research Thrusts



Scalable Learning and Optimization for Secure and Economic Grid Operations

- Address power grid operational challenges from increased weather disruptions, security threats, and the integration of distributed energy resources
- Combine physics-informed machine learning with statistical risk modeling in control loops, addressing the computational intractability of traditional high-fidelity stability simulations
  - Trustworthy ML algorithms for rapid stability risk assessment
  - Risk-averse optimization techniques for robust contingency management
  - End-to-end HPC framework on DOE LCF

- Enable real-time preventive controls that simultaneously optimize for security, economic efficiency, and resilience; transforming how operators manage power grid complexities in the face of multiple contingency scenarios



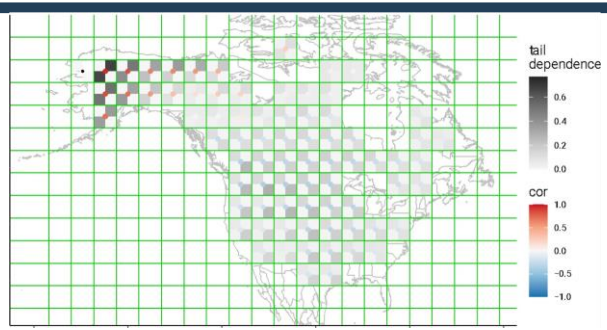
Grid Modernization Initiative -CASCDES

Project Objectives:

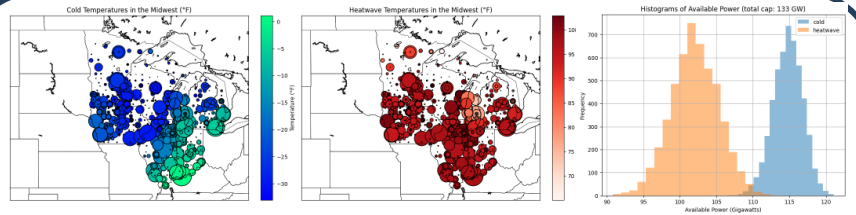
- Forecast the likelihood of wide-area extreme impacts and their impact to power market prices
- Assess the effectiveness of investment and market adaptations to mitigate this risk.
- Work with industry and stakeholders to develop tools/approaches that will become the industry standard for assessing climate risk for wide-area extremes

Potential Impact:

- Enhanced preparedness for wide-area extreme weather events on power systems improving power system resilience.
- Informed interregional network planning based on insights into the impacts of such events.
- Policy development and market adaptations to promote and enhance power system resilience and reduce outages.



Our approach captures abnormal correlations between Alaska and Illinois



Left: Extreme cold temperatures. Middle: Extreme hot temperatures. Radius is proportional to generator capacity. Right: Available power during the exreme event

THANKS!