Computational Nuclear Physics: AI/ML

Amber Boehnlein

Jefferson Lab Associate Director

Computational Science
& Technology

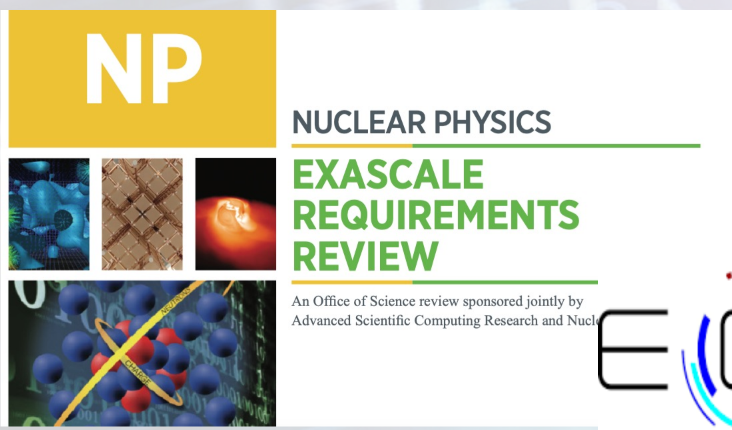
(with many thanks to Malachi Schram)

Outline

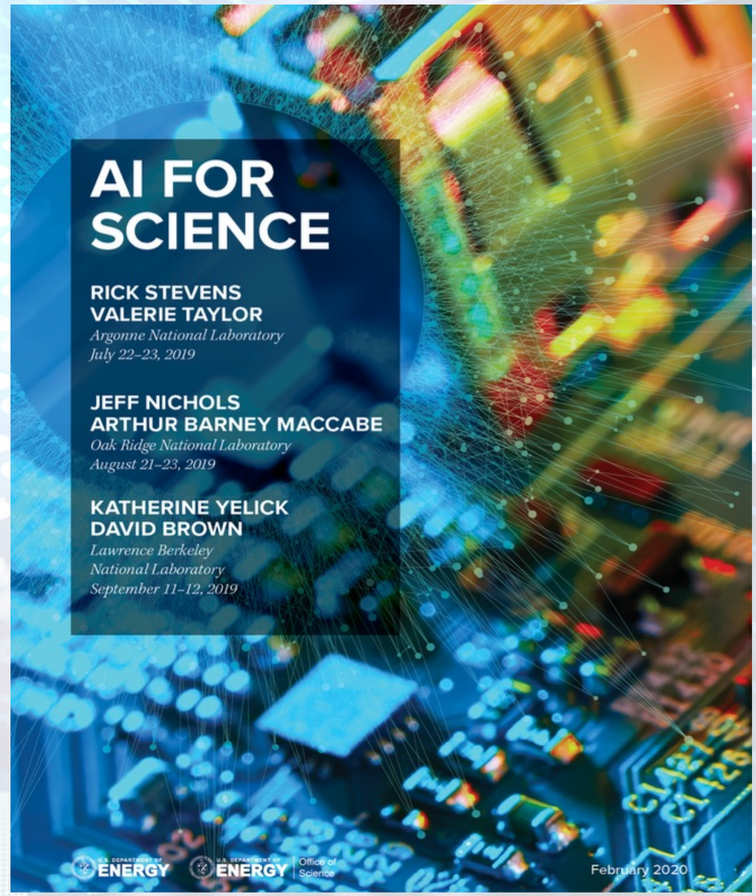
- The goals of this talk
  - Prompt dialog for drafting the white paper
  - Support for general recommendations (see Peter's talk)
- Outline
  - Landscape
  - Glimpse into the future
  - Observations

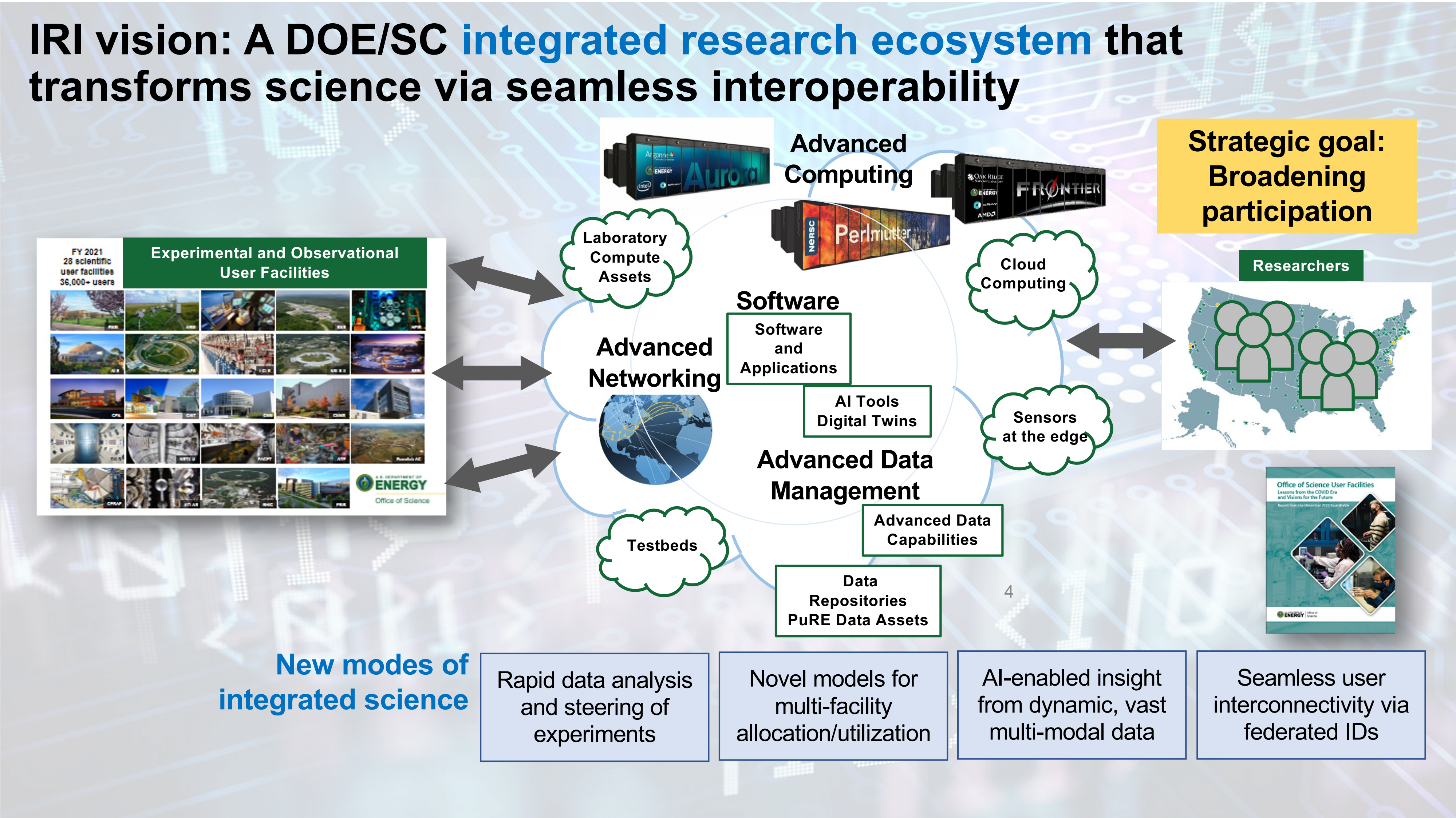
SC-Wide Activites

- The Exascale Computing Project
  - Requirements Gathering – 2015/2016
- AI for Science Townhalls – 2019
- Integrated Research Infrastructure Architectural Blueprint Activity- 2022
- AI@DOE workshop 2022
- AI For Science and Security 2022



EXASCALE COMPUTING PROJECT





IRI vision: A DOE/SC integrated research ecosystem that transforms science via seamless interoperability
Advanced Computing
Strategic goal: Broadening participation
FY 2021
28 scientific user facilities
36,000+ users
Experimental and Observational User Facilities
Laboratory Compute Assets
Cloud Computing
Researchers
Software
Software and Applications
AI Tools Digital Twins
Advanced Networking
Sensors at the edge
Advanced Data Management
Advanced Data Capabilities
Testbeds
Data Repositories PuRE Data Assets
4
New modes of integrated science
Rapid data analysis and steering of experiments
Novel models for multi-facility allocation/utilization
AI-enabled insight from dynamic, vast multi-modal data
Seamless user interconnectivity via federated IDs

IRI Patterns and Practice Areas

IRI Science Patterns	Description	Facilitator(s)
Time-Sensitive Patterns	Workflows that have time critical/sensitive requirements, e.g., experiment steering, near real-time event detection, deadline scheduling to avoid falling behind.	Chin Guok
Data-Integration -Intensive Patterns	Analysis of combined data from multiple sources - can include data from multiple sites, experiments and/or simulations. Tracking metadata and provenance for reproducible science. Interactive analysis of data, possibly at scale.	Nicholas Schwarz
Long-term Campaign Patterns	Sustained access to resources at scale over a longer time needed to accomplish a well defined objective. Robustness, reproducibility and reliability are important to accomplish. Likely to involve significant logistical planning. Examples include sustained simulation production and large data (re)processing for collaborative use.	Tom Uram & David Cowley

IRI Practice Area	Description	Facilitators?
Resource Co-Operations	Allocations/provisioning of multiple heterogeneous resources must be aligned in time and planned in advance to enable integrated workflows. IRI requires new levels of cooperation, collaboration, co-scheduling, and joint planning - across facilities and across DOE programs.	Arjun Shankar & Johannes Blaschke
Cybersecurity and Federated Access	Users require a distributed research Infrastructure with seamless access and consistent services while Office of Science Lab cyber personnel operate an environment with cybersecurity requirements and policies set at the Federal level. Operators of user facilities also have different missions, and thus different requirements, across the lab complex. Balancing these constraints can also lead to sources of impedance.	Amber Boehnlein & Eric Lancon
User Experience	Understanding users' needs and experiences is critical to technologists' ability to develop effective IRI solutions. This group will engage on approaches for enabling users: requirements gathering, user-centric (co)-design, liaising approaches, and related topics. (This topic has implications for all other Practice areas).	Kjiersten Fagnan
Workflows, Interfaces & Automation	Assembling system components to support IRI science cases systematically in the form of end-to-end pipelines. Users should be able to manage these overlays and middlewares effectively across facilities.	Debbie Bard & John MacAuley
Scientific Data Lifecycle	Users need to manage their data across facilities from creation (along with meta-data), staging, movement, storage, dissemination, curation, archiving, publishing, etc. Technologists need to understand the requirements across different communities to develop solutions appropriate for an integrated research infrastructure (IRI).	Jini Ramprakash & Eric Lancon
Portable / Scalable Solutions	Users and technologists need to move/translate their efforts across heterogeneous facilities (be portable) as well as go from smaller to larger resources (be scalable).	Shane Canon

AI/ML is not a separate practice area due to the cross cutting nature

NP: Artificial Intelligence and Machine Learning

- The Nuclear Physics Community is very active in AI/ML
- Active application of ML to NP problems
  - Funded through Lab Directed Research and Development
  - NSF –AI for Fundamental Interactions
  - Competitive funding (SciDAC-5)
  - As part of Research or Ops
- Community activities
  - Schools & Hackathons
  - Workshops
    - AI For Nuclear Physics Workshop Mar. 4-6, 2020
    - AI4EIC – Sept 2021 and Sept 2022

AI4NP WINTER SCHOOL

Theory Alliance FACILITY FOR RARE ISOTOPE BEAMS

FRIB-TA Summer School: Machine Learning Applied to Nuclear Physics

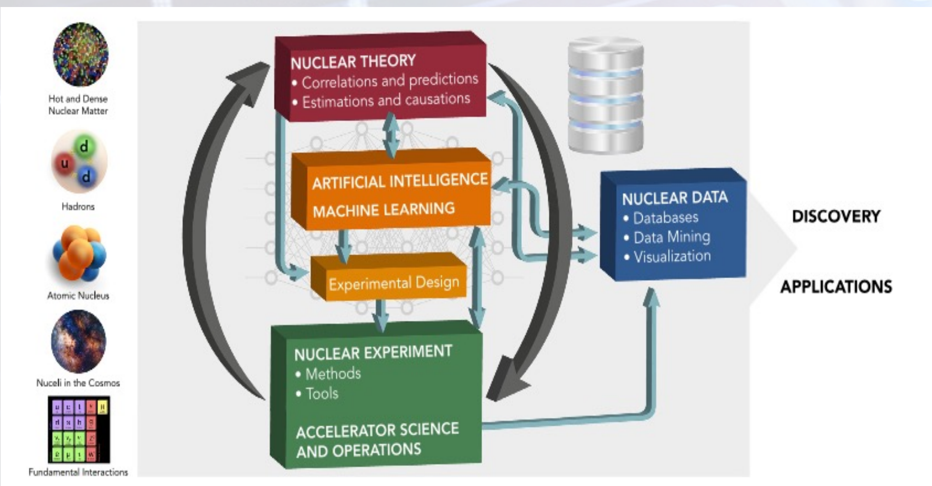
Jefferson Lab

Condensed Recommendations from the 2020 AI4NP Workshop Report

- **Comprehensive Data Management**
  - In order to verify and validate new ML models, a full AI/ML workflow framework is required (including provenance and artifact curation).
  - This will also require dataset curation to enable reproduceable and compare results.
  - Ties back to the IRI ABA
- **Computing resources for AI/ML**
  - Most computing resources for AI/ML workflows are currently available at select institutions.
  - New hardware is being deployed to provide access to GPUs for AI/ML theory efforts.
- **Uncertainty quantification**
  - There is an active community working on new technologies to better quantify uncertainties from the model predictions.
  - Integration of these techniques is being explored in NP theory.
- **Workforce development**
  - The NP community is increasing its efforts to develop a AI/ML knowledgeable workforce
  - JLab Data Science Department that has started to work directly with the NP theory community.

Review of Modern Physics: ML for Nuclear Physics

- Documented the lifecycle
  - Ties to IRI ABA
- Documented the state of practice as of 2021
  - Emulation of experimental data
  - Parameter optimization and extraction
    - Inverse problems



CONTENTS

Jefferson Lab

AI for Science and Security

- Three topical AI for Science & Security workshops recently held to discuss a long-term vision:
  1. June – AI Surrogates, AI for Complex Systems
  2. July – Properties and Inverse Design, Foundation Models
  3. August – Autonomous Discovery, AI for Programming
- Interesting themes:
  - Reinforcement learning for complex controls
  - ML-based surrogate models and digital twins
  - Uncertainty quantification (UQ), verification & validation (V&V), and guaranties
  - Building in physical constraints and relationships into the ML models
  - Integrating ML-based function surrogates with UQ & guaranties into code and simulations
  - Combined data and AI models management (ties to IRI ABA)
  - Trustworthiness, robustness, and explainability

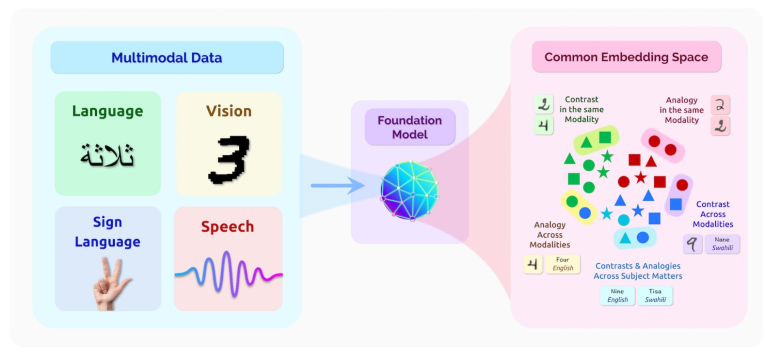
Jefferson Lab

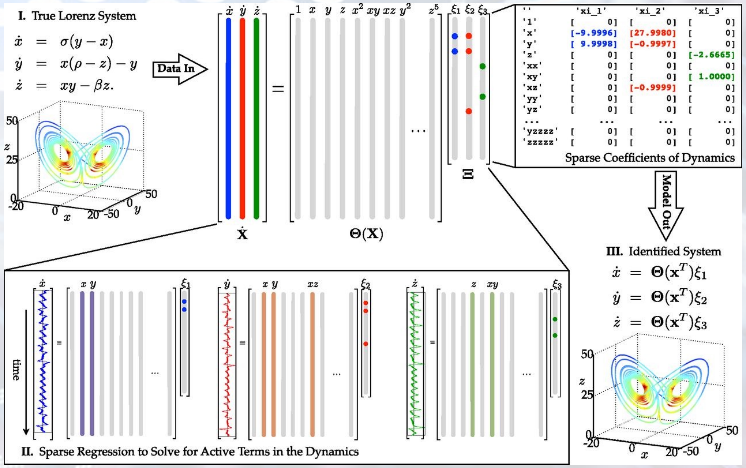
AI@DOE Executive Summary

- Top R&D priorities:
  1. Incorporate prior knowledge into AI systems
  2. Training for rare events
  3. Explainability, interpretability and understanding
  4. Automation and optimization -- self-driving labs, hypothesis generation
  5. Targeted algorithm development -- development focused on DOE missions
  6. Sustainable AI -- energy-efficient solutions (green AI)
- Enablers:
  - Underlying ecosystem to enable AI R&D -- virtual AI user facility
  - Ethics framework to guide AI R&D – identify unexpected biases

Trending now…

- **Scalable Distributed Learning**
  - In order to efficiently train over from large dataset, the need for a distributed learning computing infrastructure will likely be required. (Ties to IRI)
- **Uncertainty quantification for deep learning models**
  - Detailed study of uncertainty estimation techniques for AI/ML in NP applications as it relates to higher dimensionality and unique modalities.
- **Computing co-design for AI/ML and NP theory**
  - AI/ML for NP theory is evolving and
  - New techniques will be developed that might not perform optimally on existing hardware.
- **Foundations Models for NP theory**
  - A foundation model is any model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks.
- **Techniques to advance scientific discovery**
  - Sparse Identification of Nonlinear Dynamical Systems (SINDy) is an algorithm to discover governing dynamical equations
- **Techniques for explicit physics knowledge integration**
  - Applications of automatic differentiation through known physics equations into the ML models
  - Low energy nuclear physics examples has shown some improved results





multi-disciplinary investments needed

Jefferson Lab

Observations

- ECP project is coming to an end
  - Application Areas – sustainability needed
- IRI ABA is concluding
- Expect additional planning activities
  - Data Lifecycle
  - Co-design
  - Need to make the case for expanded access to hardware
- Funding landscape over-constrained (Recent HEP FOA: 30 proposals for 3 awards)
- A split between ML as business as usual and the application of advanced techniques
  - Need workforce development within NP AND multi-disciplinary collaborations