Markus Diefenthaler (Jefferson Lab)

Update from AI Roundtable for HEP and NP

Artificial Intelligence (AI) Initiative in the Office of Science: Round Tables

- **Purpose of Round Tables**: *"Bring together scientific expertise across the research programs in the Office of Science to identify and prioritize aspirational opportunities for* ***addressing the most significant scientific challenges*** *aligned with DOE-SC* ***through AI-enabled transformations****."*

- Six Round Tables, held between October 28 – November 8:
  - Biosciences and Environmental Sciences
  - **High Energy and Nuclear Physics, October 30–31**
  - Materials and Chemical Sciences
  - Fabrication Science
  - Fundamental Energy Research
  - User Facility Science and Operations

- **Deliverable of the Round Tables**: *"The roundtables are* ***charged to develop a set of Priority Research Opportunities (PROs)*** *consistent with the missions of the Department, in* ***which scientific impact*** *will be* ***uniquely enabled*** *and/or* ***significantly accelerated*** *by the* ***coordinated development of AI tools and methods,*** *and to highlight the path to pursue these scientific questions in the context of the DOE-SC programs."*

High Energy and Nuclear Physics Round Table

Including accelerator and detector technologies.

- **Dates and Location**: October 30–31, 2024, Rockville, MD, event website.
- **Co-chairs**: Markus Diefenthaler (Jefferson Lab), Jennifer Ngadiuba (Fermilab), Daniel Ratner (SLAC)
- **Program Managers**: Eliane Lessner (BES, Accelerator & Detector R&D), Jeremy Love (HEP), Sharon Stephenson (NP)
- **Goal: PRO for each theme**:

AI for Theory and Inverse Problems: AI tools and algorithms that facilitate theoretical predictions, calculations, and understanding, and data analysis methods that map experimental results onto parameters of interest. AI analysis that supports both theoretical discovery and deeper understanding of experimental results, including accelerating and empowering joint experimental-theoretical analysis.

AI for Experimental Precision: Applications of AI that lead to improved experimental results either through improved efficiency, precision, or sensitivity of HEP and NP experiments. AI tools and workflows that would improve experimental results via uncertainty quantification, improved identification or classification, data processing or reconstruction, and improved final calibration.

Resource-aware AI: Development of AI models, techniques, and datasets the facilitate development of AI for applications in computationally resource constrained environments such as ASICs & FPGAs on the edge. Additionally, AI-enabled co-design; Digital Twins; and methods to establish trustworthiness, reliability, and robustness of AI-based control systems should be considered. AI-assisted autonomous experimentation and control of experimental systems, including accelerators and detectors is in scope, while hardware systems and operations of these systems are considered in the Facilities Roundtable.

Participants and Assignment for first Breakout Sessions

Alex Scheinker
Farah Fahim
Kazuhiro Terao
Aobo Li
Zein-Eddine Meziani
Nicolas Schunck
Pietro Musumeci
Salman Habib
Jan Strube
Jesse Thaler
Nobuo Sato
Tia Miceli
David Lawrence

Jin Huang
Taritree Wongjirad
Dean Lee
Peter Ostroumov
Andrea Delgado
Julia Gonski
Michael Wagman
Torri Jeske
Nathan Cook
Paolo Califiura
Angelo Dragone
Michelle Kuchera
Silviu-Marian Udrescu

Participants from BNL:
- Georg Hoffstaetter
- Jin Huang

Observer from BNL:
- Torre Wenaus

Survey Prior to the Round Table

DOE Office of Science AI Roundtables: High Energy Physics and Nuclear Physics, including topics on general development related to Instrumentation and Accelerator Technology

Recognizing the potential of AI for the Nation's scientific discovery, energy research and national security future, DOE-Office of Science (SC) is commissioning a set of Scientific Roundtables to identify and prioritize the scientific challenges with the highest potential for impact through applications of AI. The Roundtables will identify Priority Research Opportunities (PROs) for using evolving AI capabilities. This request seeks input on the applications of AI and enabling capabilities to aid discovery science in High Energy and Nuclear Physics, including topics related to Instrumentation and Accelerator technologies of interest to all SC.

While participation in the roundtable series is by invitation only, input from the broader community is sought. You are encouraged to provide your input by responding to the workshop information survey. All perspectives are invited, whether DOE laboratory, industrial, or academic in origin; scientific application, technology, or workforce in nature; near-term, mid-term, or long-term in outlook. We encourage you to forward this request to experts working to apply AI to research into High Energy Physics and Nuclear Physics including topics on general development related to Instrumentation and Accelerator technologies.

Please give your responses to the questions below, thank you for your time. **We appreciate responses by Oct 28, 2024.**

We have received 45 responses from both the roundtable participants and the wider community.

The survey highlights **strong community support for leveraging AI in HEP and NP**, including accelerator and detector technologies.

There is **consensus on the transformative potential of AI**, **provided** that **dedicated resources**, **cross-institutional collaboration**, and **skilled personnel** are in place.

AI Opportunities From the Survey

AI for Theory and Inverse Problems:

Theory and Computational Physics: AI applications in theoretical physics aim to simplify complex calculations, perform symbolic regression, and model many-body interactions. Generative models are used for theory-informed simulations and to produce synthetic data for model testing. AI can empower joint experimental-theoretical analysis.

Simulation and Modeling: AI-driven models help simulate complex physical systems, reducing computational demands and enabling faster simulations. This includes high-dimensional space simulations, phase-space unfolding, and detector response modeling. Generative models are utilized for inference, data quality monitoring, and creating synthetic datasets.

Linked to Experimental Measurements and Precision: High-precision measurements require high-precision simulations / models.

AI for Experimental Precision

Experimental Measurements and Precision: AI enhances precision in measurements, a lot of responses related to data processing and analysis. AI is being used to process and analyze large volumes of data efficiently, specifically fast turnaround to the analyzer. Applications include anomaly detection, real-time calibration, and low-level to high-level data reduction.

Resource-aware AI :

Hardware Optimization and Design: AI is applied in designing detectors and accelerator systems by optimizing operational parameters / performance and exploring hardware co-design.

Operational Efficiency and Diagnostics: AI applications are targeted at improving operational efficiency, such as predictive maintenance, diagnostics for particle accelerators, and automation of legacy systems. This includes the use of large language models for synthesizing operational knowledge and documentation.

Automation and Control: AI aids in the autonomous control and calibration of accelerator and experimental setups. ML approaches and digital twins are used for real-time tuning, optimization, and predictive maintenance.

Principal Challenges and Recommendations From the Survey

Challenges

- **Funding and Institutional Support**:
  Need for dedicated funding streams and institutional prioritization of AI projects.
- **Workforce and Expertise**:
  Shortage of skilled personnel with AI and physics knowledge; challenges in talent retention and career stability.
- **Technical and Modeling Issues**:
  Difficulties in accurate modeling, reliability, and interpretability; strict accuracy demands in HEP and NP experiments.
- **Community Adoption**:
  Hesitancy due to trustworthiness and accuracy concerns; need for rigorous testing and validation.

Recommendations

- **Working Groups**:
  Establish cross-institutional groups to share knowledge and standardize AI methodologies.
- **Data Infrastructure**:
  Invest in workflow management systems for AI use cases, including data and analysis preservation and metadata.
- **Interdisciplinary Training**:
  Develop programs to build foundational skills in AI and HEP and NP.
- **Collaboration Opportunities**:
  Enhance partnerships with labs and tech companies to accelerate AI development and support.

Agenda

Wednesday, October 30

8:15 a.m.	Linda Horton - Welcome
8:30 a.m.	Round Table Charge and Organization
9:00 a.m.	**Breakout Sessions:** Scientific Challenges / Opportunities, in the context of the three themes of the charge. 5-6 groups, pre-assigned. Each group has a scribe who captures the discussion in a Google doc, and reports on it later.
10:15 a.m.	**Coffee Break** During the coffee break, we can create a list out of the 5-6 lists and start grouping them.
10:45 a.m.	**Reports from the Breakout Sessions**
11:00 a.m.	**Group Discussion:** Identify most compelling science opportunities, based on the common list.
12:00 p.m.	**Lunch Break**
1:00 p.m.	**Breakout Sessions:** Develop quad charts for compelling science opportunities
2:30 p.m.	**Coffee Break**
3:00 p.m.	**Present quad charts**
3:30 p.m.	**Group Discussion**: Initial Identification of PROs
5:00 p.m.	**Discussion: Report Outline**
5:30 p.m.	**Adjourn**

Thursday, October 31

8:30 a.m.	**Group Session**: Finalize PROs
9:30 a.m.	**Breakout:** Writing Session
10:30 a.m.	**Coffee Break**
10:30 a.m.	**Breakout:** Writing Session
11:30 a.m.	**Summary and Next steps**
12:00 p.m.	**Adjourn**

Three main steps:

1. **Discuss scientific challenges and converge to common challenges.**
2. **Develop quad charts for the common challenges and converge on three PROs.**
3. **Finalize quad charts / PROs and discuss report outline and writing assignment.**

Outcome of the Group Discussion

Reminder: Three Themes

Theme 1: Inverse Problems	Theme 2: AI for Experimental Precision	Theme 3: Resource-aware AI

Five Key Scientific Challenges / PRO Topics

PRO Topic 1: Inverse problems **(Theme 1)**

PRO Topic 2: Harnessing the power of existing datasets, including combining data from various generations and types of facilities, experiments, and simulations **(Themes 1, 2, 3)**

PRO Topic 3: Determining the highest priority for experiments, analyses, or simulations **(Themes 1, 2)**

PRO Topic 4: Pushing the frontiers of accelerators, experiments, and computability **(Themes 2, 3)**

PRO Topic 5: Understanding high-dimensional complex systems using low-dimensional or sparse data, including pattern recognition **(Themes 1, 2, 3)**.

Quad Charts

- **High Energy and Nuclear Physics**
  PRO Topic yyy

Scientific challenge	Summary of research direction(s) with AI
Where reasonable, participants can optionally draw from recent DOE-SC workshop reports that identify most significant scientific challenges relevant to this PRO. For example, material and chemical sciences roundtable participants could consider scientific challenges described in recent BRN reports https://science.osti.gov/bes/Community-Resources/Reports **Challenge to solve ("Need")**	Example of research directions that would significantly help to address the identified scientific challenge and could potentially be greatly advanced through AI **Research needed for solutions ("Approach")**
Potential impact (to science and DOE mission?)	**Capabilities & Near-Term Pathways**
Why is this important? ("Benefits")	Key needed capabilities or barriers/actions that must be tackled to advance the identified research directions. Could include, for example, needed new experimental tools, compute capabilities and methods. **Are there technical or infrastructure barriers?**

High Energy and Nuclear Physics
PRO Topic 1: AI for Inverse Problems and Inference

Scientific challenge

Reconstructing the nature of the Universe on microscopic and macroscopic scales—and the intrinsic properties of matter that emerge from its fundamental laws—is a grand challenge in high-energy and nuclear physics.

To address this challenge, one has to efficiently solve **high-dimensional inverse problems** to optimize experimental performance and relate large-scale measured data to **physical observables**, which then have to be compared to precision theory calculations to draw **robust inferences** on the underlying principles, parameters, and dynamics, and in the presence of well-characterized uncertainties and well-specified assumptions.

Summary of research direction(s) with AI

- AI-enabled **end-to-end simulation-based inference**:
  - Developing AI-enhanced data processing systems
  - High-dimensional deconvolution (aka unfolding) of experimental data
  - Generative models for accelerating Monte Carlo simulations, including parametrized systematics
  - Parameter inference to underlying theories/models
- Building comprehensive **domain-specific foundational models** to forecast the impact of new measurements and calculations
- **Uncertainty quantification** for inverse model solutions
- Building **model-based control systems** based on validated solutions to inverse problems.

Potential impact (to science and DOE mission?)

The 2023 P5 and NSAC-LRP reports recommend high-priority experiments and scientific areas involving the **acquisition of large data sets** and the **development of high-fidelity simulations**, which will benefit from AI-enabled solutions to inference and inverse problems, for example:

- Cosmic frontier: dark energy and dark matter physics, nature of primordial fluctuations
- Neutrino physics: cross-section modeling, event reconstruction, rare processes
- Energy frontier: Higgs physics, searches for new physics
- Accelerator R&D: Performance improvements based on virtual diagnostics and machine parameter inference
- Hadronic physics: Hadron structure (quark and gluon 3D imaging of nucleons and nuclei) and hadron spectroscopy.
- Theory: accelerating ab initio simulations in lattice QCD, nuclear many-body physics, and collider event generators

Capabilities & Near-Term Pathways

- Workforce: cross-disciplinary training and **career opportunities** in AI for physics (and physics for AI)
- Curation and long-term data storage
- Modernizing legacy hardware/software tools
- Common frameworks for simulatiogen tools
- Integration of AI inverse models into control loops
- Accessibility to AI-ready high performance computing systems

High Energy and Nuclear Physics
PRO Topic 2: Discovery with high-dimensional, multimodal data

Scientific challenge

Maximize the scientific insight we can extract from **unique datasets** in HEP and NP.

HEP/NP data are produced by diverse, complex detector systems that generate **high-dimensional** data with **multiple modalities** arising both within individual detectors and across different experiments. Much of this data is unique, as the experiments that generated it may be difficult or even impossible to repeat. AI has advanced how we analyze data for current experiments and data modalities. To maximize **insights beyond traditional methods**, we must enable AI to maximally learn from diverse information from across experiments – **past, present, and future**.

Summary of research direction(s) with AI

- Interoperable foundation models that combine multimodal data to latent representations that may capture multiple scales (global, sensor level, physical object level)
- Anomaly detection methods for automated discovery
- AI that incorporates physics information, structure and symmetry to encode irregular and sparse data for physics inference
- Techniques such as transfer learning to address domain shifts among different datasets

Potential impact (to science and DOE mission?)

- Physics discoveries and high precision measurement in line with P5 scientific priorities and NSAC LRP
- New insights from combinations that are lost from individual channels, datasets, and experiments
- Maximize past, present, and future investments from the DOE and other facilities
- Push forward core AI methodology for the further benefit of science and society through HEP and NP's uniquely structured, large and well-annotated datasets/tasks.

Capabilities & Near-Term Pathways

- Scalable approach to distributed training and inference from laptops/desktops to HPC e.g. (thousands of GPUs).
- building AI-ready data and software - creating frameworks for interacting with and preserving diverse HEP/NP datasets, incorporating FAIR principles in the process.
- Joint interdisciplinary AI research projects addressing unique HEP/NP challenges
- training and career paths for researchers at the intersection of physics and AI ("data physicists")

High Energy and Nuclear Physics
PRO 4: “Codesign for scientific infrastructure and enabling technology” (1, 2, 3)

Scientific challenge

- Enabling precision understanding and control of scientific instruments
  - Autonomous control and optimization of large scale scientific instrumentation, including essential subsystems (e.g. lasers, LLRF systems)
  - High-precision operation in novel parametric spaces
- Enabling more data-intensive experiments
  - Real-time high-fidelity physics-relevant information extraction & validation
  - Prediction and interpretation of unique data signatures for low-event-rate experiments
  - Improve physics information density in storage and reduce bias (e.g. via full streaming DAQ)
- Enabling theoretical calculations beyond current limits

Summary of research direction(s) with AI

- Co-design of novel software/hardware for controls, resource-constrained systems
- Software and hardware infrastructure to support digital twins
- MLOps framework for development, deployment, monitoring, history of ML models
- AI-driven design, optimization, and self-calibration, including differentiable programming
- Real-time energy-efficient distributed AI-heterogeneous computing at the edge
  - Intelligent data reduction and feature extraction
- Accelerating state-of-the-art codes, including MCMC, fast emulators, uncertainty quantification, and theory-experiment collaboration, as well as ML for scientific computing
- Trust-worthiness and explainability

Potential impact
(to science and DOE mission?)

- Enhanced measurement precision and discovery potential of fundamental physics
  - In line with P5 scientific priorities and NSAC LRP
- Synergy with US microelectronics and fabrication (academia & industry)
- Tech transfer potential (energy-efficient AI, global leadership and competitiveness in AI operational capability)
- Improved reliability, safety, control, and productivity of scientific infrastructure.
- Reusable, shareable simulations/surrogate models for designing new autonomous scientific instruments.
- New theoretical breakthroughs
- Sandbox for training and development of DOE workforce

Capabilities & Near-Term Pathways

- Various-scale projects (prototypes) demonstrating tangible scientific capability enhancement
- Accelerator digital twin framework inter-lab collaboration + industry (pilot for general digital twins for complex systems with dedicated facility access)
- High-performance, high-throughput, and energy-efficient AI computing resources, capability requirements vary by use-case.
- Streamlined support and access of compute & storage resources
- Co-simulation frameworks (AI/ML – Hardware (ASIC, FPGA))
- Access to Deep Sub-Micron Semiconductor technologies
- Legal framework for hardware design collaborations (joint NDAs, CAD access)
- Human capital: pipeline of engineering & cross-disciplinary expertise/training of current staff

High Energy and Nuclear Physics
PRO Topic 5: *Virtual views of high dimensional complex systems (1-3)*

Scientific challenge

Many high energy and nuclear physics facilities, accelerators, detectors, laser systems, and experiments involve highly complex high-dimensional dynamic systems that are not directly observable.

An existing challenge is how to create AI/ML models for physically consistent virtual views of such systems based on low-dimensional and sparse non-invasive measurements.

Summary of research direction(s) with AI

1. Functional mappings such as neural operators or invertible normalizing flows.
2. Development of adaptive physics-constrained generative diffusion models for dynamic systems.
3. Low-dimensional latent embeddings for generative models of dynamic systems.
4. Digital twins as virtual constructs that mimic the structure, context, and behavior of physical systems.
5. Pattern recognition on sparse data, including self-supervised methods.
6. Multimodal and multiscale latent representations for the reconstruction of physics observables.

Potential impact (to science and DOE mission?)

1. Enable automatic/autonomous control of high-dimensional complex dynamic systems, such as the 6D phase space of charged particle beams in DOE HENP facilities, accelerators, high power lasers, detectors, and experiments.
2. Accelerator R&D: Performance improvements based on virtual diagnostics and machine parameter inference
3. Improve understanding of design and operation of future complex systems and experiments.
4. Enable high-fidelity prediction of complex systems from low-fidelity simulations.

2023 DOE AI for Science, Energy, and Security report

Capabilities & Near-Term Pathways

1. Need for a multi-scale approach to distributed training and inference from laptops to servers and HPC.
2. Resources in support of open data (curation, storage, AI-ready, indexed and maintained, data engineering).
3. Workforce training and career opportunities for state-of-the-art physics-informed generative AI methods.
4. Modernizing legacy hardware/software tools
5. Integration of AI inverse models into control loops
6. Production and validation of large synthetic datasets.
7. Large scale modeling and simulations of physical systems.

Summary and Next Steps

- **Developed quad charts for four PROs**:

PRO Topic 1: Inverse problems **(Theme 1)**

PRO Topic 2: Harnessing the power of existing datasets, including combining data from various generations and types of facilities, experiments, and simulations **(Themes 1, 2, 3)**

PRO Topic 4: Pushing the frontiers of accelerators, experiments, and computability **(Themes 2, 3)**

PRO Topic 5: Understanding high-dimensional complex systems using low-dimensional or sparse data, including pattern recognition **(Themes 1, 2, 3)**.

- **Report is work in progress. Draft reports from the six Round Tables are due December 20.**
- Following the last Round Table, a virtual **meeting** will be held for DOE and the co-chairs of the individual roundtables to **identify cross-cutting opportunities** (to-be-scheduled).
- DOE SC has discussed the possibility of webinars to present the outcomes of the Round Tables.