



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Advanced Scientific Computing Research at the Department of Energy

NIH Pre-Meeting for 2019 ML-MSM

Integrating Machine Learning with Multiscale Modeling for
Biomedical, Biological, & Behavioral Systems

October 8, 2019

Steven Lee, PhD

Office of Advanced Scientific Computing Research
Program Manager for Applied Mathematics
Research Division Director (Acting)

Steven.Lee@science.doe.gov

Outline of NIH Pre-Meeting Webinar

Objective: Identify perspectives, challenges, and opportunities of integrating machine learning with multiscale modeling (ML-MSM) in biomedical, biological, and behavioral systems

- Methods guided by fundamental principles of mathematics and physics
- Four approaches: ODEs, PDEs, theory-driven, data-driven
- Context: Digital Twins, Human Safety

Scientific Machine Learning & Artificial Intelligence

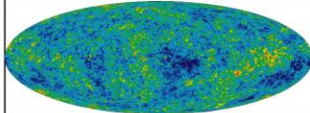
Scientific progress will be driven by

- Massive Data: sensors, simulations, networks
- Predictive Models & Adaptive Algorithms
- Heterogeneous High-Performance Computing

Trend: Human-AI collaborations will transform the way science is done.

Exemplars of Scientific Achievement

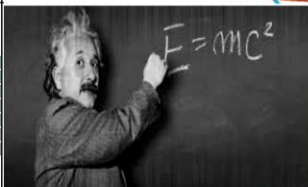
Cosmic Microwave Background



DNA Structure



Periodic Table of the Elements



Human-AI insights enabled via scientific method, experimentation, & AI reinforcement learning.

BASIC RESEARCH NEEDS FOR Scientific Machine Learning
Core Technologies for Artificial Intelligence

Prepared for Sub-Committee on Energy Advanced Scientific Computing Research

U.S. DEPARTMENT OF ENERGY

DOE Office of Science Programs

Advanced Scientific Computing Research

- Delivering world leading computational and networking capabilities to extend the frontiers of science and technology

Basic Energy Sciences

- Understanding, predicting, and ultimately controlling matter and energy flow at the electronic, atomic, and molecular levels

Biological and Environmental Research

- Understanding complex biological, climatic, and environmental systems

Fusion Energy Sciences

- Building the scientific foundations for a fusion energy source

High Energy Physics

- Understanding how the universe works at its most fundamental level

Nuclear Physics

- Discovering, exploring, and understanding all forms of nuclear matter



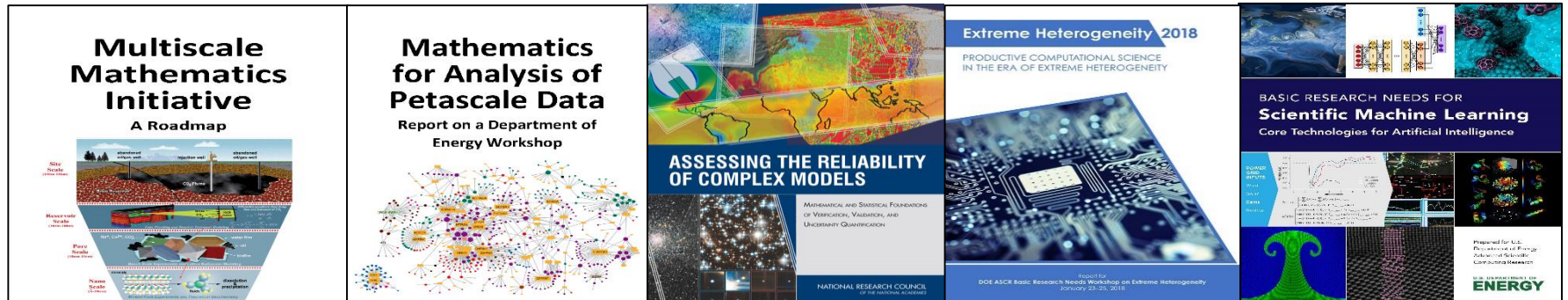
DOE Applied Mathematics program

Office of Advanced Scientific Computing Research (ASCR)

Applied Math program develops the mathematical & scientific computing foundations to accelerate the pace of scientific discoveries

Portfolio in FY19: \$30M/year for ~50 projects at Labs, universities, non-profits
 Scientific enabling technologies are being built on ASCR developments in:

Core Applied Math	Optimization, Linear algebra, Uncertainty Quantification (UQ), Differential equations, Machine Learning (ML), Meshes, Multigrid, Reduced order models
Scientific Software/ Libraries	High performance software codes (PETSc, Trilinos, SUNDIALS), Automatic differentiation, Parallel-in-time integrators, Meshes, Tensors, & more
Math Centers	Science at user facilities, Power Grid, Additive Manufacturing, Materials Design
Workshops	Multiscale Math, Petascale Data, UQ, Extreme Heterogeneity, Scientific ML



Preparing for Scientific Machine Learning and AI: Pre-Workshop Report (January 2018)

Executive Summary

1. Introduction

1. Scope
2. Definitions

2. Motivation and impact

1. Use cases and examples
2. Need for robust scientific machine learning
3. Need for interpretable scientific machine learning

3. Scientific computing and machine learning

1. Supervised machine learning methods
2. Unsupervised machine learning methods
3. Reinforcement machine learning methods
4. Other machine learning methods

4. Computational foundations for scientific machine learning

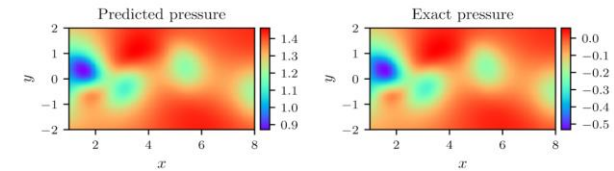
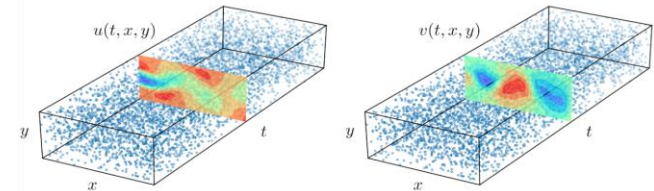
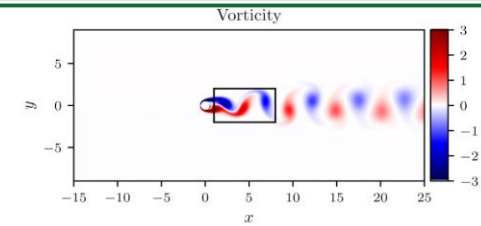
1. Rigorous analysis
2. Model reduction and multi-fidelity modeling
3. Computational complexity
4. Optimization
5. Statistics and uncertainty quantification

PRD1: Domain-Aware Scientific Machine Learning

Leveraging Scientific Domain Knowledge

Key Points: How can domain knowledge be effectively incorporated into Scientific ML methods?

- Established domain models based on physical mechanism & scientific knowledge
- Scientific ML offers significant opportunity to complement traditional domain models
- Domain knowledge: physical principles, symmetries, constraints, computational predictions, uncertainties, etc
- Potential to improve accuracy, interpretability, & defensibility while reducing data requirements & accelerating training process



Correct PDE	$u_t + (uu_x + vv_y) = -p_x + 0.01(u_{xx} + v_{yy})$ $v_t + (uv_x + vv_y) = -p_y + 0.01(v_{xx} + v_{yy})$
Identified PDE (clean data)	$u_t + 0.999(uu_x + vv_y) = -p_x + 0.01047(u_{xx} + v_{yy})$ $v_t + 0.999(uv_x + vv_y) = -p_y + 0.01047(v_{xx} + v_{yy})$
Identified PDE (1% noise)	$u_t + 0.998(uu_x + vv_y) = -p_x + 0.01057(u_{xx} + v_{yy})$ $v_t + 0.998(uv_x + vv_y) = -p_y + 0.01057(v_{xx} + v_{yy})$

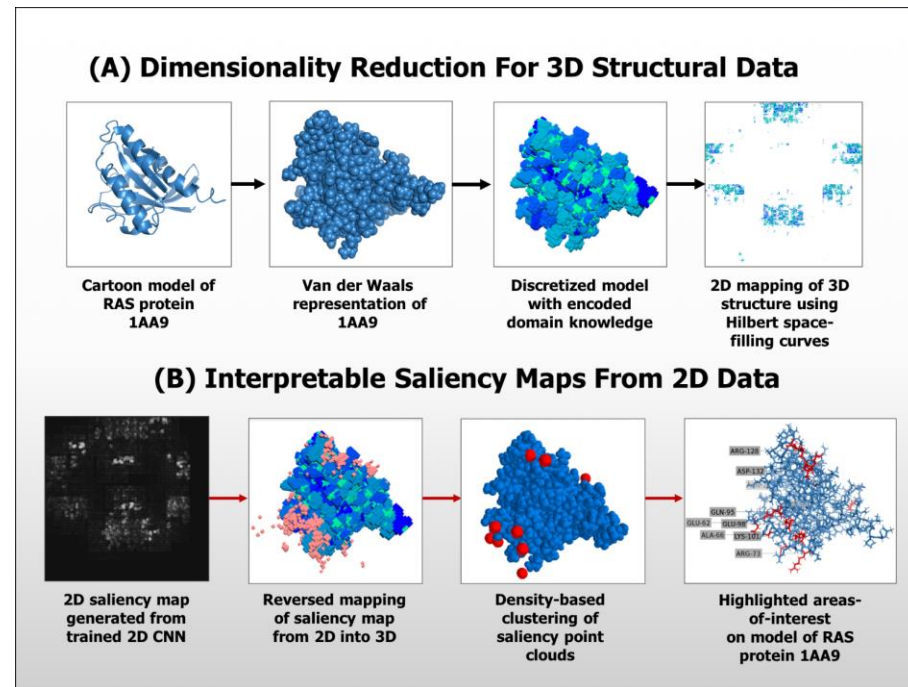
This example illustrates the capabilities obtained by incorporating domain knowledge into a deep neural network. Given scattered and noisy data components of an incompressible fluid flow in the wake of a cylinder, we can employ a physics-informed neural network that is constrained by the Navier-Stokes equation in order to identify unknown parameters, reconstruct a velocity field that is guaranteed to be incompressible and satisfy any boundary conditions, as well as recover the entire pressure field. Figure from: Raissi et al.

PRD2: Interpretable Scientific Machine Learning

Explainable and Understandable Results

Key Points: How to balance the use of increasingly complex ML models with the need for users to understand conclusions & derive insights?

- Physical understanding has been the bedrock of modeling
- User confidence linked to the conviction that model accounts for domain knowledge (variables, parameters, physical laws, etc.)
- Need exploration & visualization approaches for “debugging” complex machine learning models
- Need metrics to quantify model differences



High-level data pipeline overview for dimensionality reduction of 3D protein structures (A) and interpretation of saliency maps from trained CNN model (B). Saliency maps generated from CNN models can then be clustered to identify areas along the 3D structure that are regions that highly influence the output of the CNN model. From these salient regions, specific residues can be identified that fall in close proximity to the salient regions.

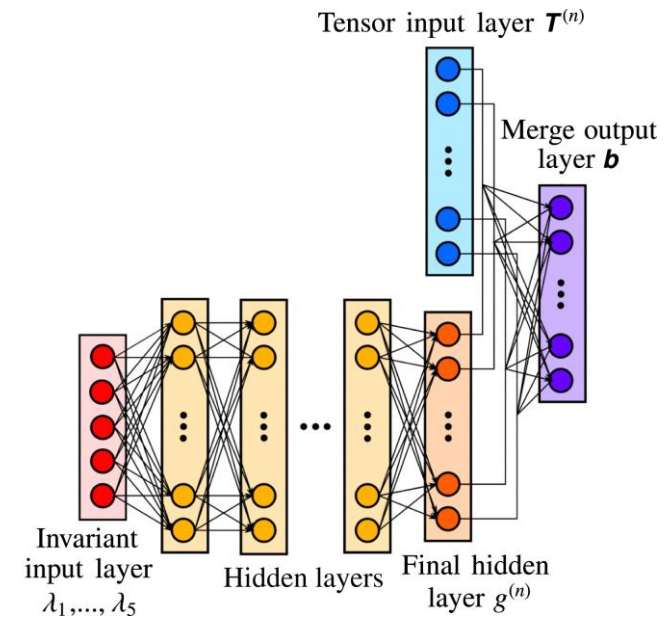
Image credit: Rafael Zamora-Resendiz and Silvia Crivelli, LBNL.

PRD3: Robust Scientific Machine Learning

Stable, Well-Posed, and Efficient Formulations

Key Points: How can computationally efficient Scientific ML methods be developed and implemented to ensure outcomes are not unduly sensitive to perturbations in training data and model selection?

- Scientific ML methods need to establish the properties of robustness & reliability
- Integration of protocols for verification & validation are in their infancy
- Progress will require research proving that developed methods and implementations are stable and well-posed



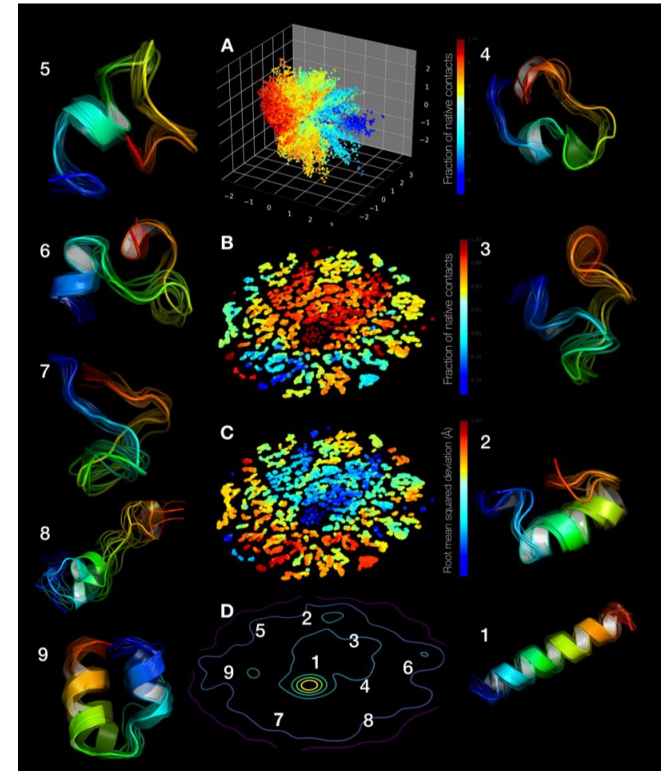
In the context of Reynolds averaged incompressible turbulence modeling, a neural network has been used in an eddy viscosity turbulence closure model. From physical arguments, the model needs to satisfy rotational invariance, ensuring that the physics of the flow is independent of the orientation of the coordinate frame of the observer. A special network architecture, a tensor basis neural network (TBNN), embeds rotational invariance by construction. Without this guarantee, the NN model evaluated on identical flows with the axes defined in different directions could yield different predictions.
Image credit: SNL.

PRD4: Data-Intensive Scientific Machine Learning

Automated Scientific Inference and Data Analysis

Key Points: What novel approaches can be developed for reliably finding signals, patterns or structure within high-dimensional, noisy, uncertain input data?

- Scientific ML methods require the development of improved methods for statistical learning in high-dimensional Scientific ML systems with noisy and complex data
- Need approaches required to identify structure in complex high-dimensional data
- Scientific ML requires efficient sampling in high-dimensional parametric and model spaces



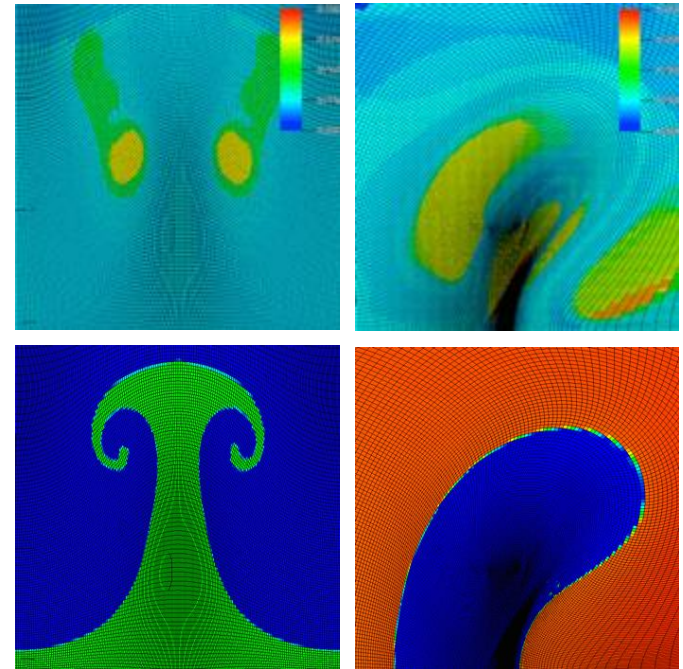
ML techniques reveal Fs-peptide folding events from long time-scale molecular dynamics simulations. A low dimensional embedding of the simulation events reveal transitions from fully unfolded states (blue) to fully folded states (red). A two dimensional embedding using t-test stochastic neighborhood embedding shows the presence of near native states (labeled state 1) versus partially unfolded (2-7) and fully unfolded states (8-9) in the picture. Image Credit: Arvind Ramanathan, ORNL.

PRD5: Machine Learning-Enhanced Models and Simulations

Predictive Scientific Computing

Key Points: What is the role and potential advantages of ML-embedded approaches in computational model and algorithm development?

- Combination of scientific computing with learned adaptivity for more efficient simulations
- ML for in-situ parameter tuning
- ML for sub-grid physics models
- Progress will require the development of new methods to quantify tradeoffs and optimally manage the interplay between traditional and ML models and implementations



The arbitrary Lagrangian-Eulerian (ALE) method is used in a variety of engineering and scientific applications for enabling multi-physics simulations. Unfortunately, the ALE method can suffer from simulation failures, such as mesh tangling, that require users to adjust parameters throughout a simulation just to reach completion. A supervised ML framework for predicting conditions leading to ALE simulation failures was developed and integrated into a production ALE code for modeling high energy density physics.

Image credit: M. Jiang, LLNL.

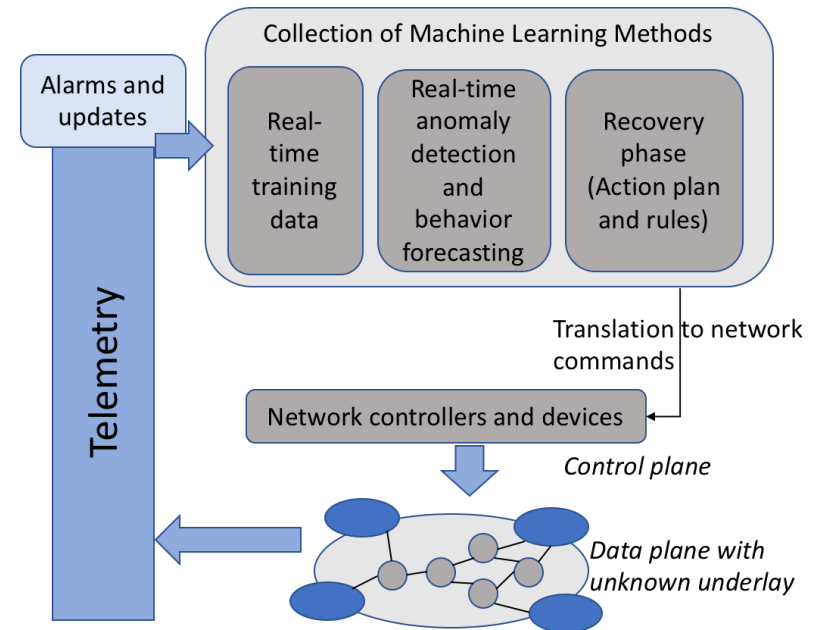


PRD6: Intelligent Automation and Decision Support

Management and Control of Complex Processes and Systems

Key Points: What are the challenges in managing the interplay between automation & human decision-making?

- Outer-Loop applications include optimization, uncertainty quantification, inverse problems, data assimilation, & control.
- New mathematically & scientifically justified methods to guide data acquisition and ensure data quality and adequacy.
- Scientific ML methods for improving system resilience or responsiveness.



Exascale applications are exponentially raising demands from underlying DOE networks such as traffic management, operation scale and reliability constraints. Networks are the backbone to complex science workflows ensuring data is delivered securely and on-time for important compute to happen. In order to intelligently manage multiple network paths, various tasks such as pre-computation and prediction are needed to be done in near-real-time. ML provides a collection of algorithms that can add autonomy and assist in decision making to support key facility goals, without increased device costs and inefficiency. In particular, ML can be used to predict potential anomalies in current traffic patterns and raise alerts before network faults develop. Image credit: Prabhat, LBNL.

Scientific AI/Machine Learning: Priority Research Needs

Scientific Machine Learning: Foundations

Domain-Aware: Leverages & respects scientific domain knowledge. Physics principles, symmetries, constraints, uncertainties & structure-exploiting models

Interpretable: Explainable and understandable results. Model selection, exploiting structure in high-dimensional data, use of uncertainty quantification with machine learning

Robust: Stable, well-posed & reliable formulations. Probabilistic modeling in ML, quantifying well-posedness, reliable hyperparameter estimation

Data-Intensive Scientific ML: Scientific inference & data analysis. ML methods for multimodal data, in situ data analysis & optimally guide data acquisition

Machine Learning-Enhanced Simulations: ML hybrid algorithms & models for predictive scientific computing. ML-enabled adaptive algorithms, parameter tuning & multiscale surrogate models

Intelligent Automation and Decision Support: Adaptivity, automation, resilience, control. Exploration of decision space with ML, ML-based resource management, optimal decisions for complex systems

Scientific Machine Learning: Capabilities



January 2019

Advances in 6 Priority Research Directions (PRDs) are needed to develop the next generation of machine learning methods and artificial intelligence capabilities.



U.S. DEPARTMENT OF
ENERGY

Office of
Science

<https://www.osti.gov/biblio/1478744>

National Artificial Intelligence Research and Development Strategic Plan

AI R&D Strategies

1. Make Long-Term Investments in Research
2. **Develop Effective Methods for Human-AI Collaboration**
3. Understand and Address the Ethical, Legal, and Societal Implications of AI
4. Ensure the Safety and Security of AI Systems
5. Develop Shared Public Datasets and Environments for AI Training and Testing
6. Measure and Evaluate AI Technologies through Standards and Benchmarks
7. **Better Understand the National R&D Workforce Needs**
8. Expand Public-Private Partnerships to Accelerate Advances in AI



THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN: 2019 UPDATE

A Report by the
SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE
of the
NATIONAL SCIENCE & TECHNOLOGY COUNCIL

JUNE 2019



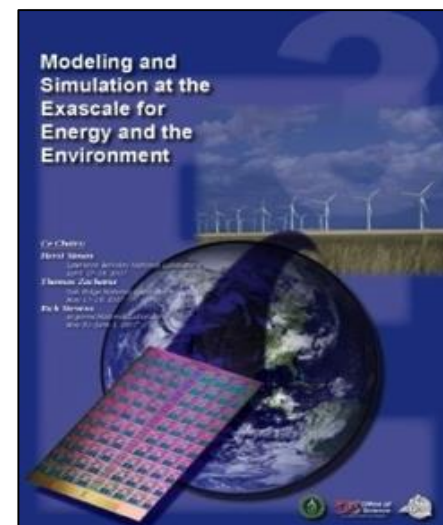
U.S. DEPARTMENT OF
ENERGY

Office of
Science

AI for Science Townhalls

**Organized by Argonne, Oak Ridge and Berkeley
with participation from all the DOE laboratories...**

- Four “Townhalls” aimed at getting input from the DOE community on opportunities and requirements for the next 5-10 years in computing with a focus on convergence between HPC, data and AI
- July (Argonne), August (Oak Ridge), September (Berkeley), October 22-23 (Washington DC)
- Modeled after the 2007 Townhalls that launched the Exascale Computing Initiative
- Each meeting covers roughly the same ground, geographically distributed to enable local participation
- Applications in science, energy and technology
- Software, math and methods, hardware, data management, computing facilities, infrastructure, integration with experimental facilities, etc.
- Expect 200-300 people per meeting
- Output will be a report to guide strategic planning at Labs and DOE



DOE Scientific Machine Learning and AI

Lens of Applied Mathematics & Scientific Computing

Capability Themes	Relevant Funding Announcements since 2005
Data-Intensive Scientific Machine Learning	2009 – 2012: Mathematics for Analysis of Petascale Data 2009 – 2012: Joint Mathematics Computer Science Institute 2012 – 2015: Resilient Extreme-Scale Solvers 2013 – 2016: DOE Data-Centric Science at Scale
Machine Learning-Enhanced Scientific Simulations	2005 – 2008: Multiscale Mathematics Research and Education 2008 – 2011: Multiscale Mathematics for Complex Systems 2013 – 2016: Uncertainty Quantification (UQ) for Extreme-Scale Science 2019 – 2021: UQ for Scientific Machine Learning & Artificial Intelligence
Intelligent Automation and Decision Support for Complex Systems	2009 – 2012: Mathematics for Complex, Interconnected Systems 2010 – 2013: Uncertainty Quantification (UQ) for Complex Systems 2012 – 2017: Mathematical Multifaceted Integrated Capability Centers I 2017 – 2022: Mathematical Multifaceted Integrated Capability Centers II



DOE Applied Math program has laid the groundwork to harness Machine Learning and Artificial Intelligence for scientific purposes

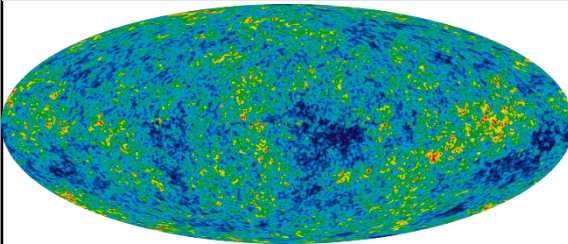
Scientific progress will be driven by

- Massive Data: sensors, simulations, networks
- Predictive Models & Adaptive Algorithms
- Heterogeneous High-Performance Computing
- Scientific Machine Learning & AI

Trend: Human-AI collaborations will transform the way science is done.

Exemplars of Scientific Achievement

Cosmic Microwave Background



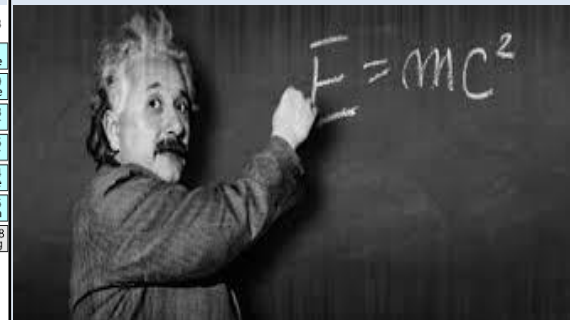
DNA Structure



Periodic Table of the Elements

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 H																	2 He
2	3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba	57 La	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	89 Ac	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Nh	114 Fl	115 Mc	116 Lv	117 Ts	118 Og
	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu				
	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr				

Special Relativity



Human-AI insights enabled via scientific method, experimentation, & AI reinforcement learning.