

The logo is a white hexagonal shape with a green border and four small white hexagons at the corners. Inside, the text 'AI FOR SCIENCE TOWN HALL' is written in green, bold, sans-serif font.

AI FOR SCIENCE TOWN HALL

AI for Science

Barbara Helland, Advanced Scientific
Computing Research

With help from Jeff Nichols, ORNL; Rick
Stevens, ANL and Kathy Yelick, LBNL



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Development and Application of AI Critical For All Government Agencies

- Executive Order on AI

Policy Statement: Artificial Intelligence (AI) promises to drive growth of the United States economy, enhance our economic and national security, and improve our quality of life.

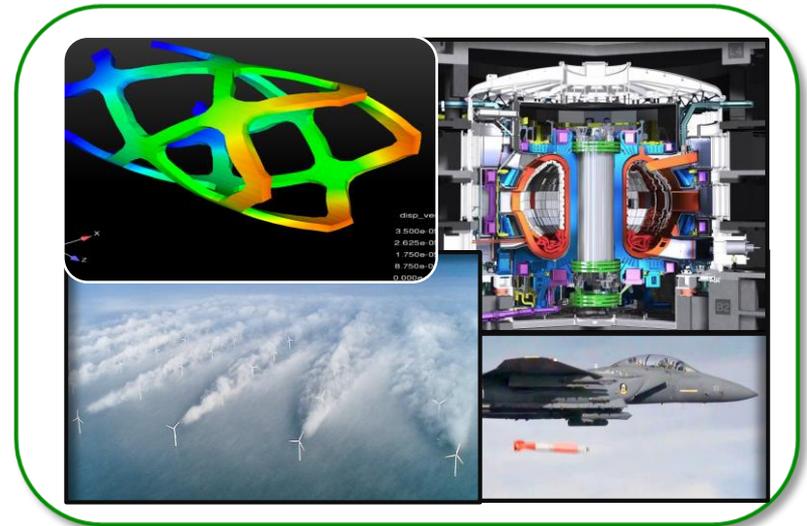
... leadership requires a concerted effort to promote advancements in technology and innovation, while protecting American technology, economic and national security, civil liberties, privacy, and American values and enhancing international and industry collaboration with foreign partners and allies.

- Supported by multiple agency strategies and programs



DOE builds on historical missions and touches all areas

- The U.S. AI strategy includes
 - 1. Long-term investment in research**
 2. Effective methods for human-AI collaboration
 3. Address ethical, legal and social implications
 - 4. Ensure the safety and security of AI Systems**
 - 5. Develop shared datasets and environments**
 6. Standards and benchmarks
 7. Understand the AI workforce
 - 8. Expand public-private partnerships**
- DOE will play a key role in AI for science and engineering
 - AI Technology office
 - Mission-driven development and application of AI/ML, i.e., innovation in Science, Energy and National security
 - Build on its HPC mission
 - Large-scale scientific data for research
 - Talent development



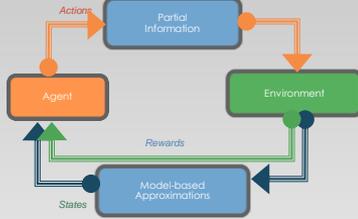
DOE research challenges touch all areas of AI

Data



- Experimental design
- Data curation and validation
- Compressed sensing
- Facilities operation and control

Learning



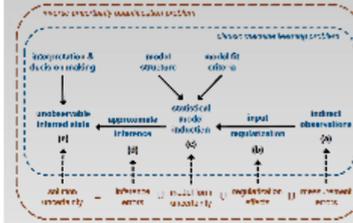
- Physics informed
- Reinforcement learning
- Adversarial networks
- Representation learning and multi-modal data
- “Foundational math” of learning

Scalability



- Algorithms, complexity and convergence
- Levels of parallelization
- Mixed precision arithmetic
- Communication
- Implementations on accelerated-node hardware

Assurance



- Uncertainty quantification
- Explainability and interpretability
- Validation and verification
- Causal inference

Workflow



- Edge computing
- Compression
- Online learning
- Federated learning
- Infrastructure
- Augmented intelligence
- Human-computer interface

How can AI help?

Data Analytics

Classification

Regression

Clustering

Dimensionality
Reduction

Inverse problems

Model
reconstruction

Parameter
estimation

Denoising

Surrogate models

Approximate
expensive
simulations

Approximate
experiments

Fill in missing
models in
simulations

.Design and control

Optimize design of
experiments

Control
instruments

Navigate state
spaces

Learn from sparse
rewards

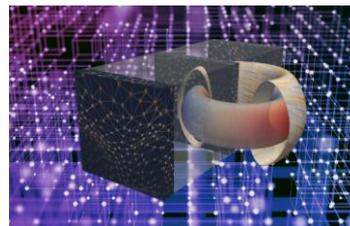
ASCR's Role in AI

Things we can do in Science with AI now



Researchers at Argonne National Laboratory are working on optimization models that use machine learning, a form of artificial intelligence, to simulate the electric system and the severity of various problems. In a region with 1,000 electric power assets, an outage of just three assets can produce nearly a billion scenarios of potential failure. Making models even more robust will give grid operators stronger guidance that can inform more reliable planning and operations for contingent events such as storms, equipment malfunctions and big fluctuations in renewable energy generation. <https://www.anl.gov/article/artificial-intelligence-can-make-the-us-electric-grid-smarter>

A team of scientists from Princeton Plasma Physics Laboratory (PPPL) and Princeton University is working with a Harvard graduate student to applying deep learning to forecast sudden disruptions that can halt fusion reactions and damage the doughnut-shaped tokamaks, like ITER, that house the reactions. <https://www.pppl.gov/news/2019/04/artificial-intelligence-accelerates-efforts-develop-clean-virtually-limitless-fusion>

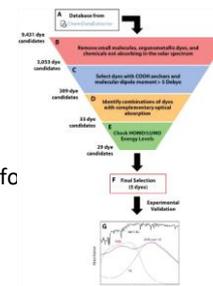


Depiction of fusion research on a doughnut-shaped tokamak enhanced by artificial intelligence. (Eliot Feibush/PPPL and Julian Kates-Harbeck/Harvard University)



Researchers at the U.S. Argonne National Laboratory and the University of Cambridge in England have developed a novel "design to device" approach to identify promising materials for dye-sensitized solar cells (DSSCs) that can be manufactured with low-cost, scalable techniques. Using a combination of simulation, data mining and machine learning the team was able to pinpoint five high-performing, low-cost dye materials from a pool of nearly 10,000 candidates for fabrication and device testing.

<https://www.sciencedaily.com/releases/2019/03/190305092837.htm>

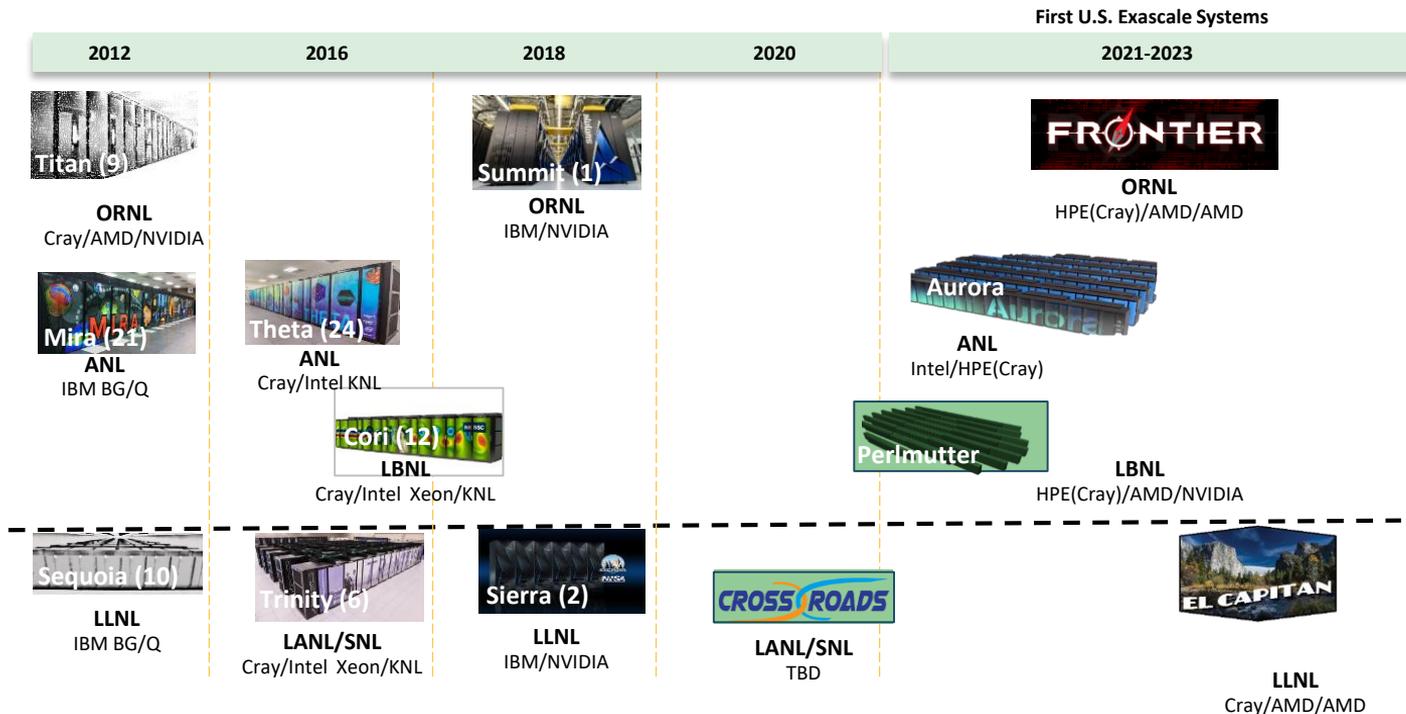


Pre Exascale: Summit Excels Across Simulation, Analytics, AI



- Data analytics – CoMet bioinformatics application for comparative genomics. Used to find sets of genes that are related to a trait or disease in a population. Exploits cuBLAS and Volta tensor cores to solve this problem 5 orders of magnitude faster than previous state-of-art code.
 - **Has achieved 2.3 ExaOps** mixed precision (FP₁₆-FP₃₂) on Summit
- Deep Learning – global climate simulations use a half-precision version of the DeepLabv3+ neural network to learn to detecting extreme weather patterns in the output
 - **Has achieved a sustained throughput of 1.0 ExaOps (FP₁₆)** on Summit
- Nonlinear dynamic low-order unstructured finite-element solver accelerated using mixed precision (FP₁₆ thru FP₆₄) and AI generated preconditioner. Answer in FP₆₄
 - **Has achieved 25.3 fold speedup** on Japan earthquake – city structures simulation
- Half-dozen Early Science codes are reporting >25x speedup on Summit vs Titan

Department of Energy (DOE) Roadmap to Exascale Systems



SciML Priority Research Needs January 2018

SciML
Foundational
Research
Themes

Domain-aware – leverages and respects scientific domain knowledge – physics principles and symmetries, constraints, uncertainties and 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

SciML
Capabilities

Data-Intensive – scientific inference and data analysis – ML methods for multimodal data, in situ data analysis and to optimally guide data acquisition

Adaptive Simulations – ML hybrid algorithms and models for better scientific computing tools– ML-enabled adaptive algorithms, parameter tuning and multiscale surrogate models

Controlling Complex Systems – automated decision support, adaptivity, resilience, control – exploration of decision space with ML, ML-based resource management and control, optimal decisions for complex systems

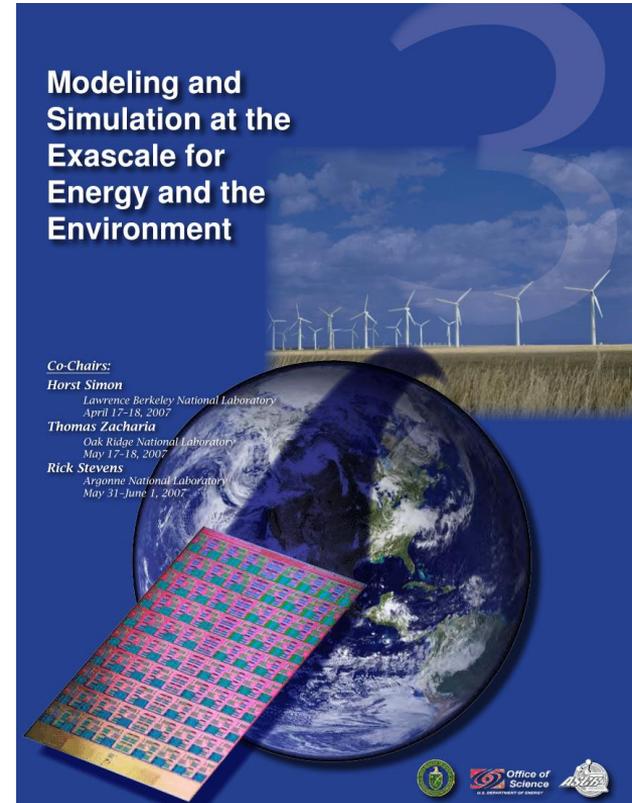


Purpose: Define priority research directions for applied mathematics in scientific machine learning (ML). Identify the challenges and opportunities for increasing the rigor, robustness, and reliability of ML for DOE missions.

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

AI for Science Town Halls: Learning from Exascale

- Modeled after Exascale Town Halls in 2007
- Collecting community input on the opportunities and challenges facing the scientific community in the era of convergence of High Performance Computing (HPC) and artificial intelligence (AI) technologies and data
- Engage the DOE science community in a series of broad and open discussions about
 - opportunities that can be realized by advancing and accelerating the development of AI capabilities specifically for science and science use cases,
 - opportunities from the DOE Office of Science as well as selected topics from energy and technology domains and will include approaches combining experiments, traditional modeling and simulation, and machine/deep learning
 - opportunities that include the DOE science user facilities. Led to many other workshops (>10)

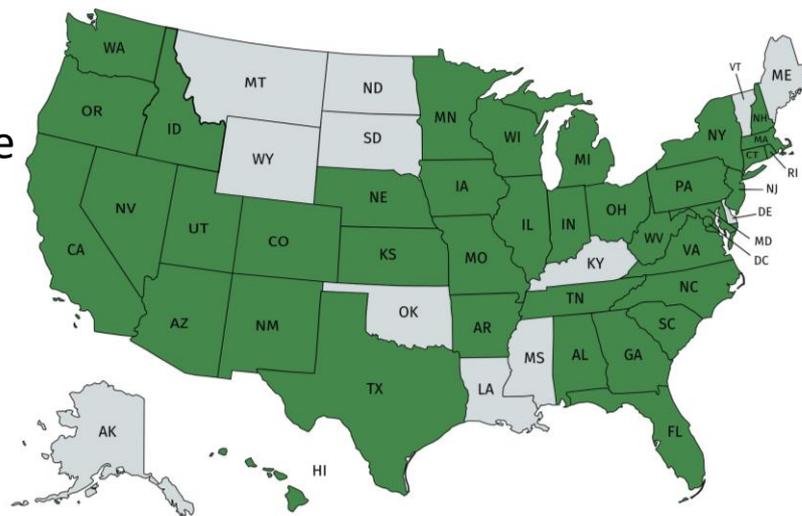


The AI for Science Town Halls so far

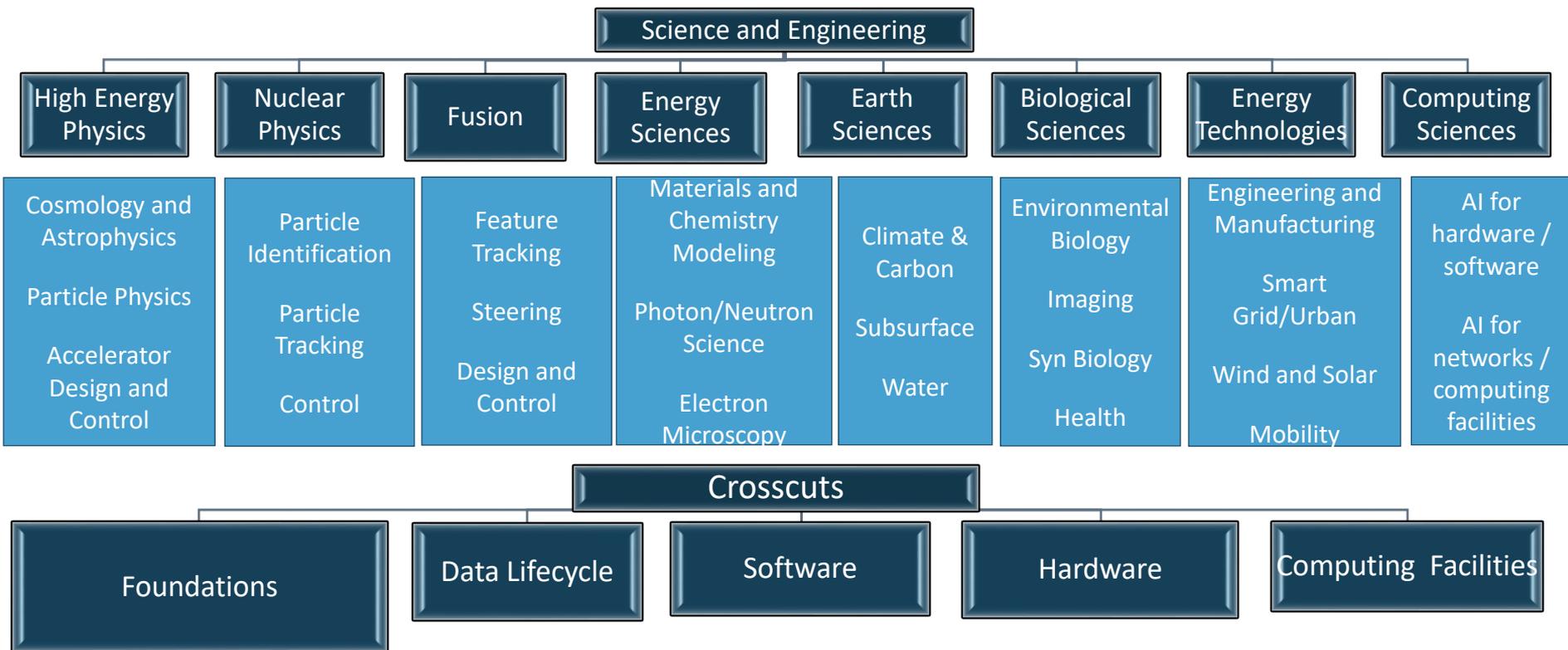
- Over 1000 registrations across 4 Town Halls

ANL	July 22-23	357	
ORNL	Aug 20-21	330	
LBNL	Sept 11-12	349	+100 online
DC	Oct 22-23	273	+ ?
Totals	1309	1309	

- All 17 DOE National Laboratories
- 39 Companies from large and small
- Over 90 different universities
- 6 DOE/SC Offices + EERE and NNSA



Breakouts and Subtopics



AI for Science Vision: 2020 \Rightarrow 2030

- AI will enable us to attack **new problems**
- AI becomes **equal partners** to modeling and simulation and data analysis
- AI will **enable experimentalists to harness the power of Exascale** computing
- AI will power **automated laboratories** and change the nature of experimental science
- AI will need **new computing architectures, new software environments, new policies** and create **new user communities** and new ways of dissemination
- AI **will improve how DOE laboratories operate** and how work is done

In Ten Years...

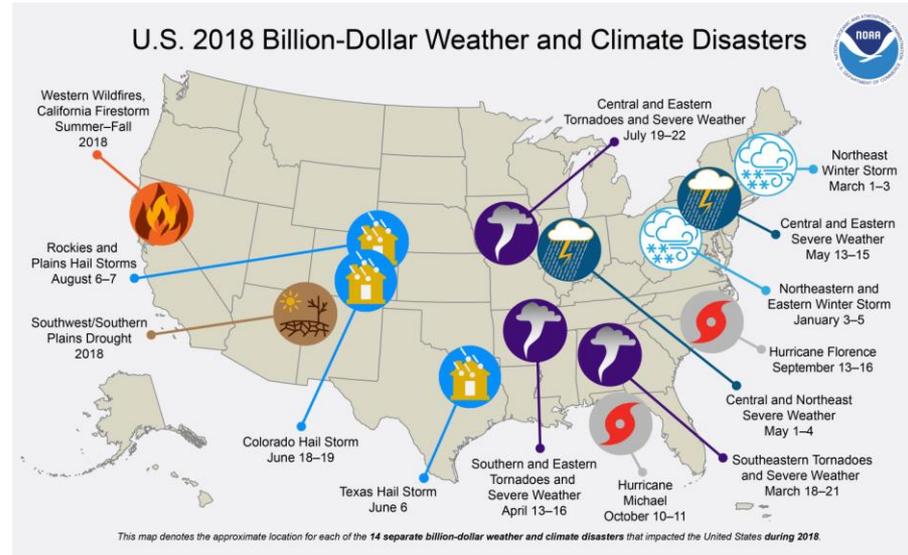
- **Learned Models Begin to Replace Data**
 - queryable, portable, pluggable, chainable, secure
- **Experimental Discovery Processes Dramatically Refactored**
 - models replace experiments, experiments improve models
- **Many Questions Pursued Semi-Autonomously at Scale**
 - searching for materials, molecules and pathways, new physics
- **Simulation and AI Approaches Merge**
 - deep integration of ML, numerical simulation and UQ
- **Theory Becomes Data for Next Generation AI**
 - AI begins to contribute to advancing theory
- **AI Becomes Common Part of Scientific Laboratory Activities**
 - Infuses scientific, engineering and operations

Earth and Environmental Sciences: State of the Art

A predictive understanding of the Earth system is crucial for utilizing its energy and water resources while mitigating costly environmental hazards.

AI approaches are playing useful roles in

- Geophysical characterization and change detection
- Data assimilation and model–data integration
- Data-driven and physics-informed machine learning
- Surrogate modeling



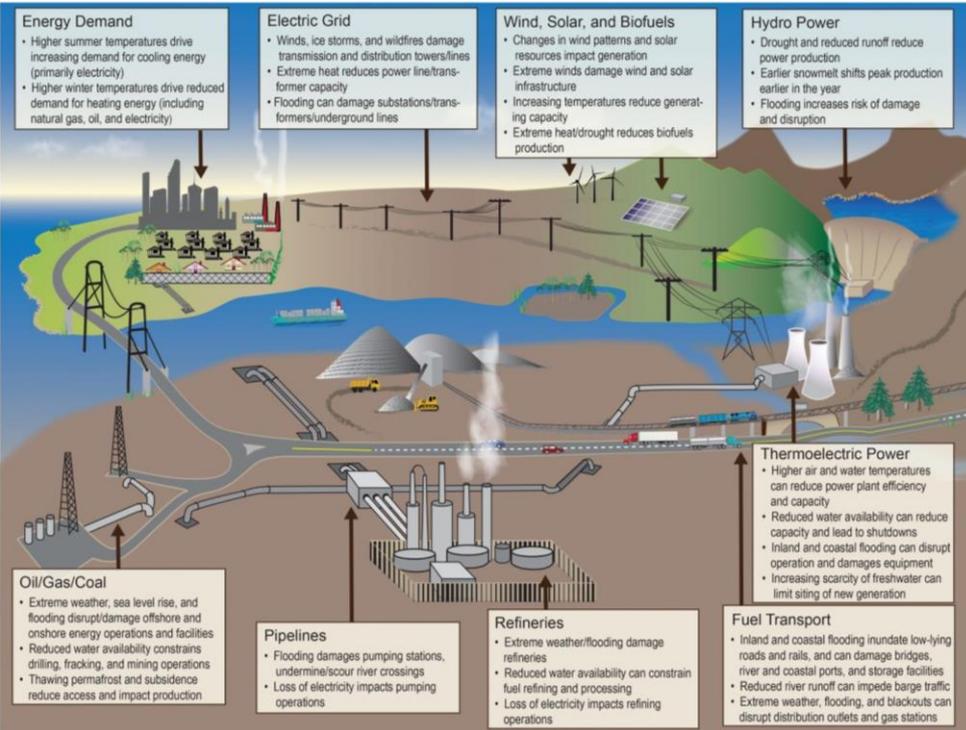
Earth and Environmental Sciences: State of the Art

A variety of AI methods (e.g., data mining, artificial neural networks, and random forests) are beginning to be used for

- Producing weather forecasts
- Environmental data gap filling
- Satellite remote sensing and geophysical image analysis
- Process parameter estimation and uncertainty quantification
- Spatiotemporal pattern discovery
- Physics-constrained simulation



Grand Challenge #1



Project environmental risk and develop resiliency in a changing environment

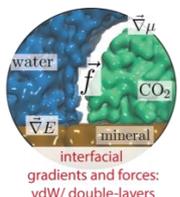
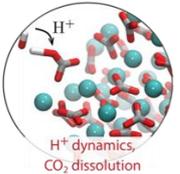
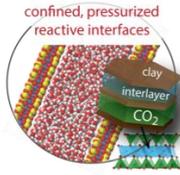
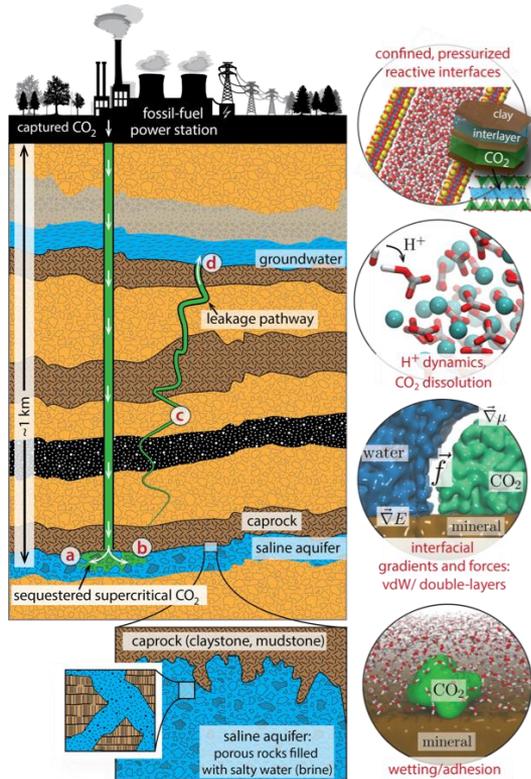
- Increasing frequency of weather extremes and changing environment pose risks to energy infrastructure and the built environment
- Sparse observations and inadequate model fidelity limit the ability to identify vulnerability, mitigate risks, and respond to disasters

Grand Challenge #1

- New tools are needed to accelerate projection of weather extremes and impacts on energy infrastructure
- Building resiliency to address evolving risks will benefit from integration of smart sensing systems, built-for-purpose models, ensemble forecasts to quantify uncertainty, and dynamic decision support systems for critical infrastructure



Grand Challenge #2

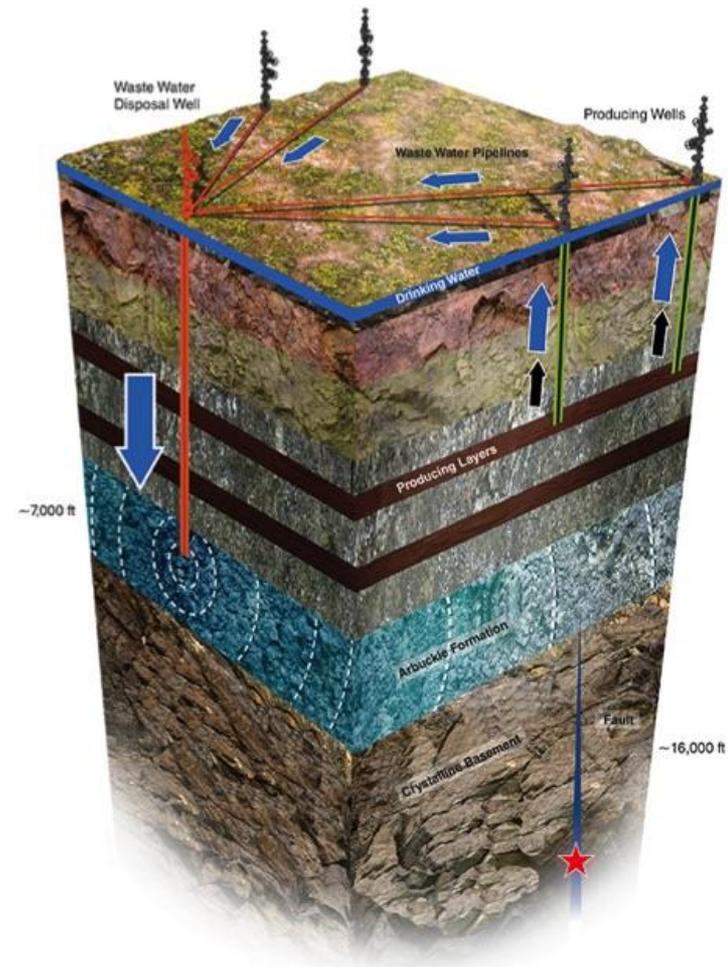


Characterize and modify subsurface conditions for responsible energy production, CO₂ storage, and contaminant remediation

- National energy security and transition to renewable energy resources relies on utilization of subsurface reservoirs for energy production, carbon storage, and spent nuclear fuel storage
- Subsurface data are uncertain, disparate, diverse, sparse, and affected by scaling issues
- Subsurface process models are incomplete, uncertain, and frequently unreliable for prediction

Grand Challenge #2

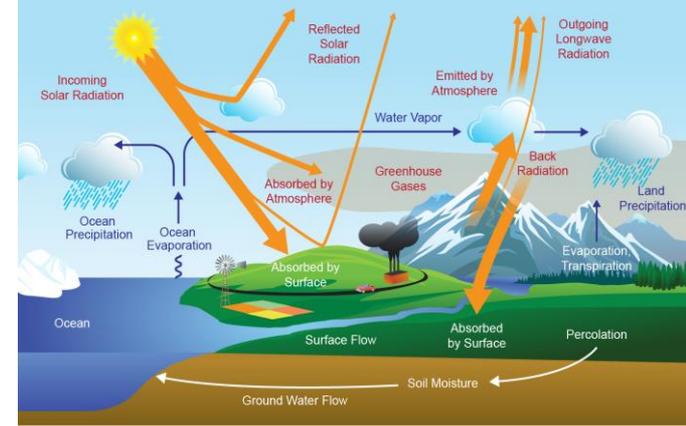
- We need to substantially increase hydrocarbon extraction efficiency, discover and exploit hidden geothermal resources, reduce induced seismicity and other impacts, improve geologic CO₂ storage, and predict long-term fate and transport of contaminants
- Mitigating risks requires improved subsurface characterization and assimilation of real-time data streams into predictive models of geological and ecological processes



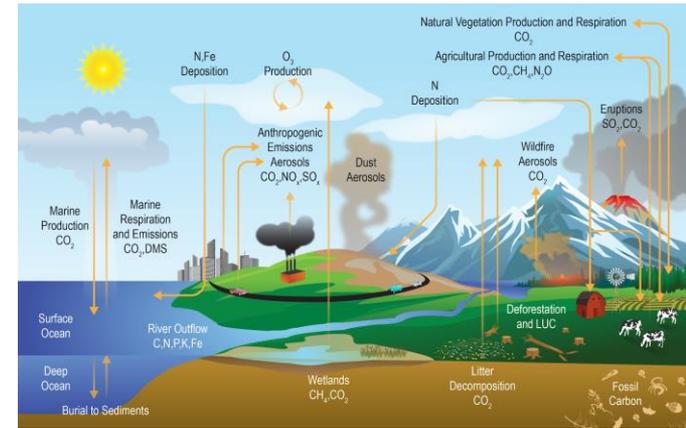
Grand Challenge #3

Develop a predictive understanding of the Earth system under a changing environment

- To advance the nation's energy and infrastructure security, a foundational scientific understanding of complex and dynamic hydrological, biological, and geochemical processes and their interactions is required (across atmosphere, ocean, land, ice)
- Knowledge must be incorporated into Earth system models to project future climate conditions for various scenarios of population, socioeconomics, and energy production and use

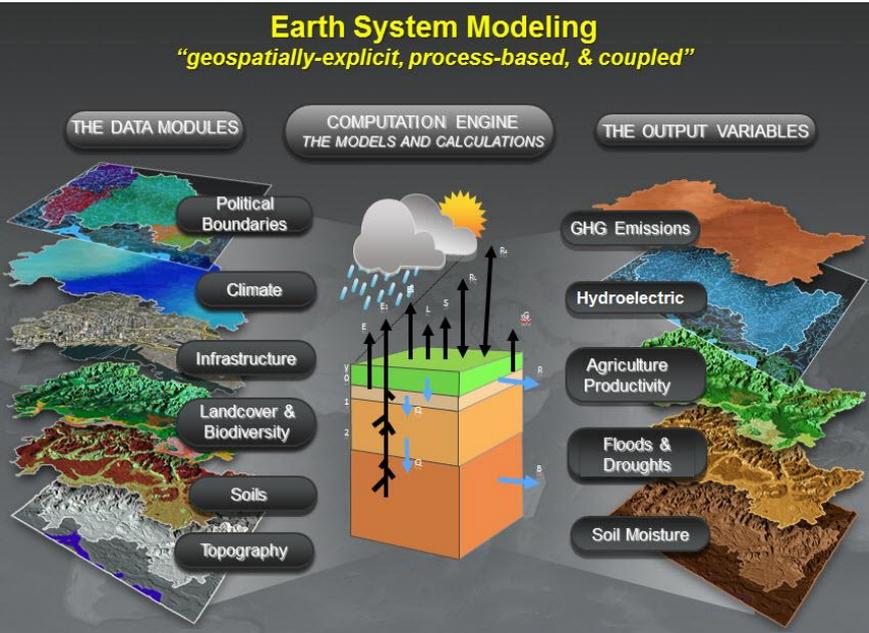


Energy & Water Cycles



Carbon & Biogeochemical Cycles

Grand Challenge #3

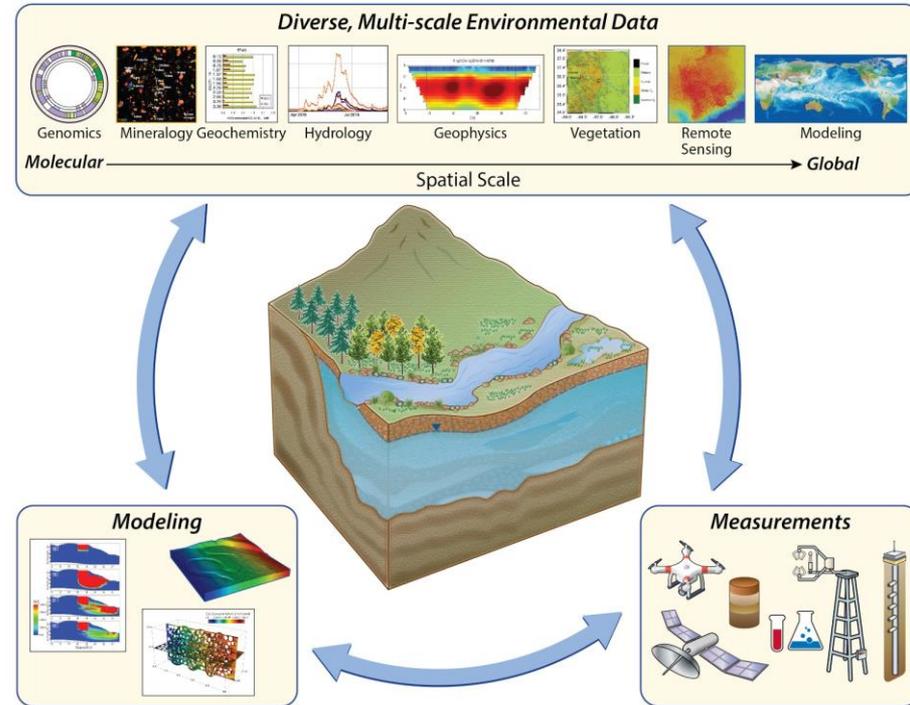


- Accurate predictions are needed to quantify changes in atmospheric and ocean circulation and weather extremes, to close the carbon cycle, and to understand responses and feedbacks of human, terrestrial, and marine ecosystems to environmental change
- Advances in genomics and bioscience data need to be leveraged to provide detailed understanding of plant–microbial interactions and their adaptations and feedbacks to the changing environment

Grand Challenge #4

Ensure global water security under a changing environment

- Water resources are critical for energy production, human health, food security, and economic prosperity
- Water availability and water quality are impacted by environmental change, weather extremes, and disturbances such as wildfire and land use change



Grand Challenge #4

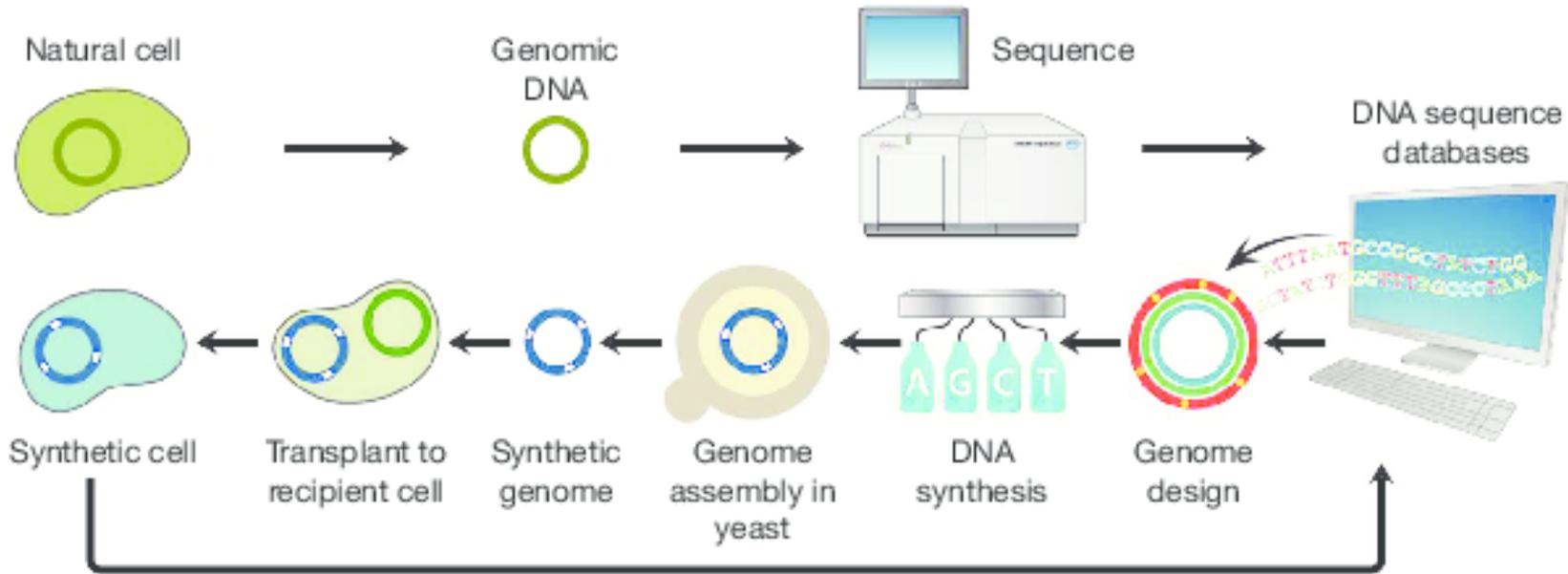


- Methods are needed to integrate disparate and diverse multi-scale data with models of watersheds, rivers, and water utility infrastructure
- Predictions of water quality and quantity require data-driven models and smart sensing systems
- Water resource management must account for changes in weather extremes, population, and economic growth

Use AI to Accelerate Synthetic Biology

- **AI to predict the relationship between Genotypes and Phenotypes**
 - Today ML can predict antibiotic resistance from genomes without culturing the organism as accurately as we can measure resistance in the lab
- **ML to predict protein function from protein sequence**
 - Today DL can predict protein structure from sequence (DeepMind, TTIC, etc.)
- **Generative models to design biosynthetic pathways**
 - Today ML can predict metabolic pathways from genomes
- **Generative models to compose collections pathways into subsystems**
- **Generative models to translate from collections of functions to a set of modules**
- **Models to wrap biological modules with regulation and signaling systems**
- **Seq2seq models to translate functional blocks into genome sequences**
- **AI to control the routine fabrication and synthesis of novel whole organisms**

Building the Database to Support BioDesign



Today we have >300,000 genomes >100,000 metagenomes

In ten years we could have > 10,000,000 genomes >1,000,000 metagenomes

Food for Models!

With a robust biodesign capability we could...

- Replace chemical factories with small safe portable biomanufacturing
- Democratize and accelerate drug development
- Produce novel food grade protein and fiber sources
- Produce biological carbon capture systems
- Produce designer polymers that are environmentally benign
- Harness bespoke biological systems for water purification
- Integrate 3d printable bioinks with biological computing and control to produce new types of smart matter

What's Next after the AI for Science Town Halls

Work with other Program Offices to build a coordinated SC AI research program similar to our Quantum program

- As a first step, on Tuesday, SC Director Fall announced forthcoming charge to ASCAC
 - to establish a subcommittee to look at the outputs from the several workshops and subcommittee reports that have identified and enumerated the scientific opportunities and some challenges from the intersection of AI/ML with data-intensive science and high performance computing activities and
 - to analyze the opportunities and challenges for the Office of Advanced Scientific Computing Research (ASCR) and the Office of Science associated with Artificial Intelligence and Machine Learning.

AI Revolution

- **Learned Models Replace Data**
- **Experimental Discovery Refactored**
- **Questions Pursued Semi-Autonomously**
- **Simulation and AI Merge**
- **Theory Becomes Data**
- **AI Laboratories**

