AI for Science Town Halls

ASCAC
September 23, 2019

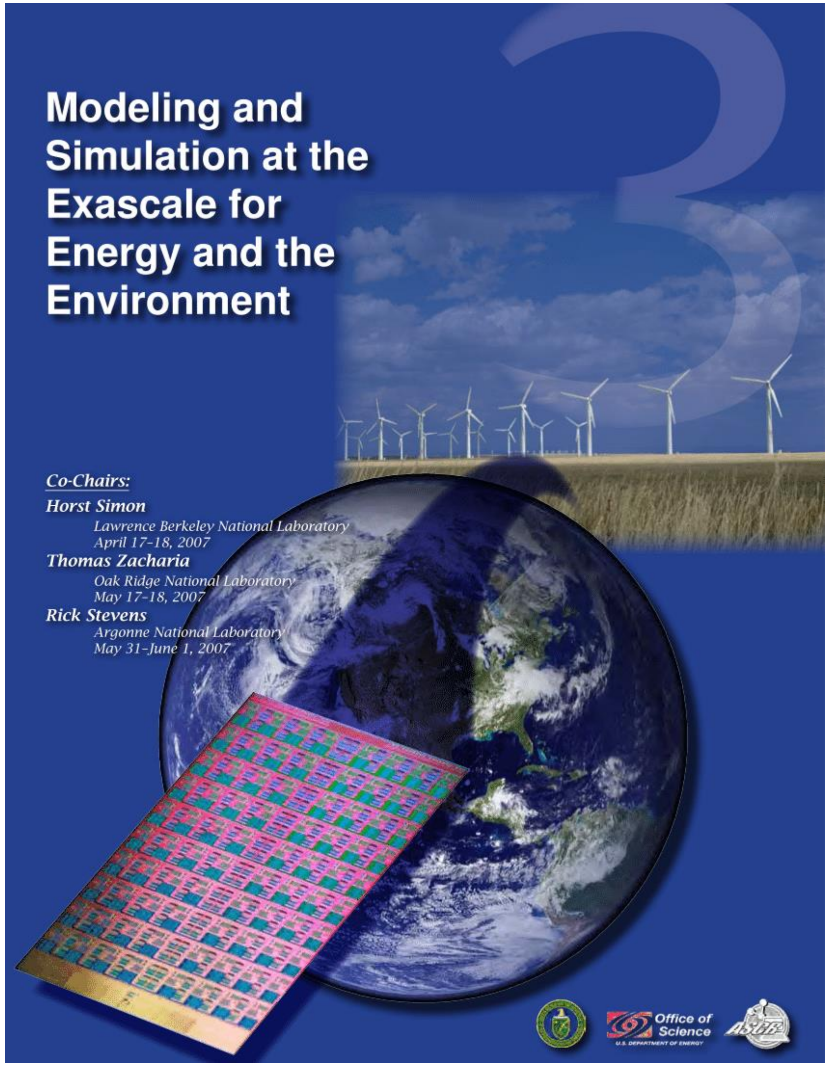
Chicago, Oak Ridge, Berkeley, and Washington, D.C.

Co-chairs: Jeff Nichols, Rick Stevens, Kathy Yelick

Note: many of the following slides are taken or use content from the three preceding townhalls, thanks to Rick Stevens, David Womble and Kathy Yelick

Why Are We Here?

- Exascale Town Halls in 2007
- Led to many other workshops (>10)
- ASCAC engagement
- NSCI
- And the Exascale Computing Initiative
  - ECP: Exascale Computing Project
  - Exascale systems
  - Application efforts across DOE
- Ideas for the next big thing, complementing exascale
- 4 Town Halls organized by the Labs
  - 4th is in DC, October 22-23



AI for Science: Integration of Modeling and Simulation, Data Analytics and Learning

- AI is transforming our “regular life” world
- AI has tremendous potential to accelerate scientific discovery
- AI Complements our Exascale Plans
  - The emerging platforms at the LCF and NERSC will be excellent platforms for machine learning, in particular deep learning training
  - The coupling of AI and HPC is a huge opportunity for DOE
  - Many uses of AI couple to experiments in ways that traditional modeling and simulation do not
  - The DOE experimental community could become major users of the DOE HPC facilities
  - Future systems directions will be impacted by AI use cases

AI is disruptive.

It won't replace the scientist, but scientists who use AI will replace those who don't.*

*Adapted from a Microsoft report, "The Future Computed"

To harness the disruptive potential of AI to improve science, we need to be leaders in AI and computational science, as well as the application of AI.

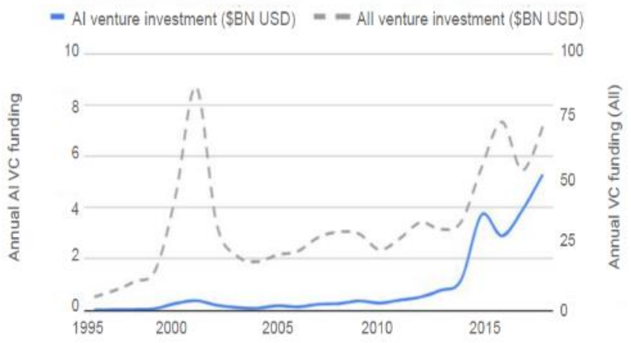
The Government/International Landscape

- Approximately 35 countries have AI strategies
  - They generally share the view that the country or business that uses its data best will be the most competitive.
  - China gets the most attention, and the competition is “asymmetric”
- The U.S. has an AI strategy that 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

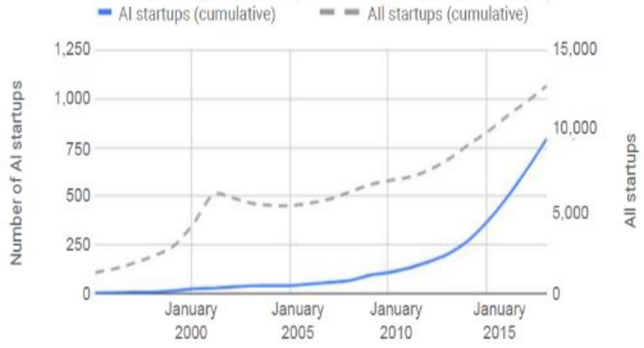
The Business Landscape

- Business
  - Will either deliver AI or use AI
  - The “Big 9” dominate, but don’t discount traditional business
  - $7.4B in start-up investments in 488 deals in 2019/Q2
  - $803M in “AI for cybersecurity” VC in last six months
- Barriers to insertion
  - Understanding: 37% of executive feel their employees understand the importance of data
  - Management needs a “BS meter”
  - Trust:
    - 49% of U.S. consumers would trust AI-generated advice for retail,
    - 38% would trust AI-generated advice for hospitality, while only
    - 20% would trust AI-generated advice for healthcare and
    - 19% for financial services
  - Example: 33% of US healthcare professionals have implemented AI into their practice, compared to a 15-country average of 46%.

Annual VC funding of AI startups (U.S., 1995 – 2017)
Source: Sand Hill Econometrics



AI startups (U.S., January 1995 – January 2018)
Source: Sand Hill Econometrics



Note: The majority of the AI startups above develop AI systems. A minority use AI as an integral part of business, but do not develop the systems themselves. See appendix for more details.

White House Executive Order on AI

EXECUTIVE ORDERS

Executive Order on Maintaining American Leadership in Artificial Intelligence

INFRASTRUCTURE & TECHNOLOGY

Issued on: **February 11, 2019**

★ ★ ★

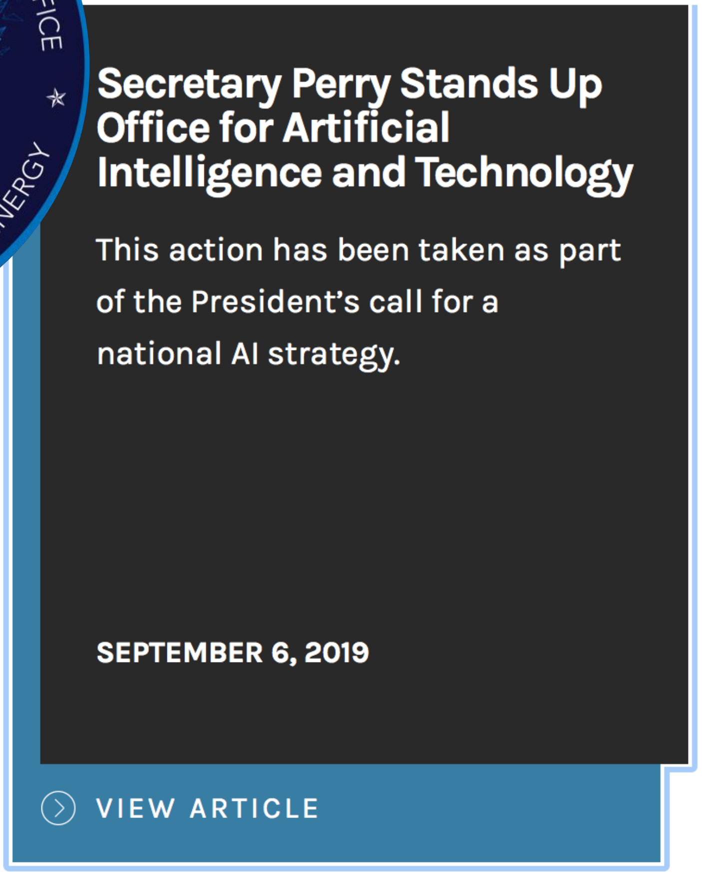
By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. Policy and Principles. Artificial Intelligence (AI) promises to drive growth of the United States economy, enhance our economic and national security, and improve our quality of life. The United States is the world leader in AI research and development (R&D) and deployment. Continued American leadership in AI is of paramount importance to maintaining the economic and national security of the United States and to shaping the global evolution of AI in a manner consistent with our Nation's values, policies, and priorities. The Federal Government plays an important role in facilitating AI R&D, promoting the trust of the American people in the development and deployment of AI-related technologies, training a workforce capable of using AI in their occupations, and protecting the American AI technology base from attempted acquisition by strategic competitors and adversarial nations. Maintaining American leadership in AI 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. It is the policy of the United States Government to sustain and enhance the scientific, technological, and economic leadership position of the United States in AI R&D and deployment through a coordinated Federal Government strategy,

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.

DOE's Artificial Intelligence and Technology Office



Vision:

Transform DOE into a world-leading AI enterprise by accelerating the research, development, delivery, and adoption of AI.

Mission:

The Artificial Intelligence and Technology Office (AITO), the Department of Energy's center for Artificial Intelligence, will **accelerate the delivery** of AI-enabled capabilities, **scale** the department-wide development and impact of AI, and **synchronize** AI activities to advance the agency's core missions, **expand partnerships**, and support American AI Leadership.

AI for Science Town Halls

The Integration of modeling and simulation, data analytics and learning

- ~1.5 days to capture ideas, problems, requirements and challenges for an AI for Science initiative
- Each townhall
  - 1 plenary, 3 keynotes, half-day breakouts on domains, half-day breakouts on crosscuts
  - All breakouts were consistent, with slight tailoring to accommodate what we learned and local influences
- What problems could be attacked?
- What data, simulations, and experiments do we need?
- What kind of methods, software and math do we need?
- What kind of computer architectures and infrastructure do we need?

Goal is to Look 10+ Years Out

- We are looking for transformational ideas
- What could be the impact of a sustained push on AI in some problem domain?
  - Building superhuman capabilities in science
- What scale
  - Big Problems, Big Pushes, Big Data, Big Systems?
  - Fine grain innovation, many thousands of small teams?
- Coupling to experiments, simulations, user and computing facilities?
- What does "scientific production" look like in this space?

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

AI for Science Town Halls – Argonne

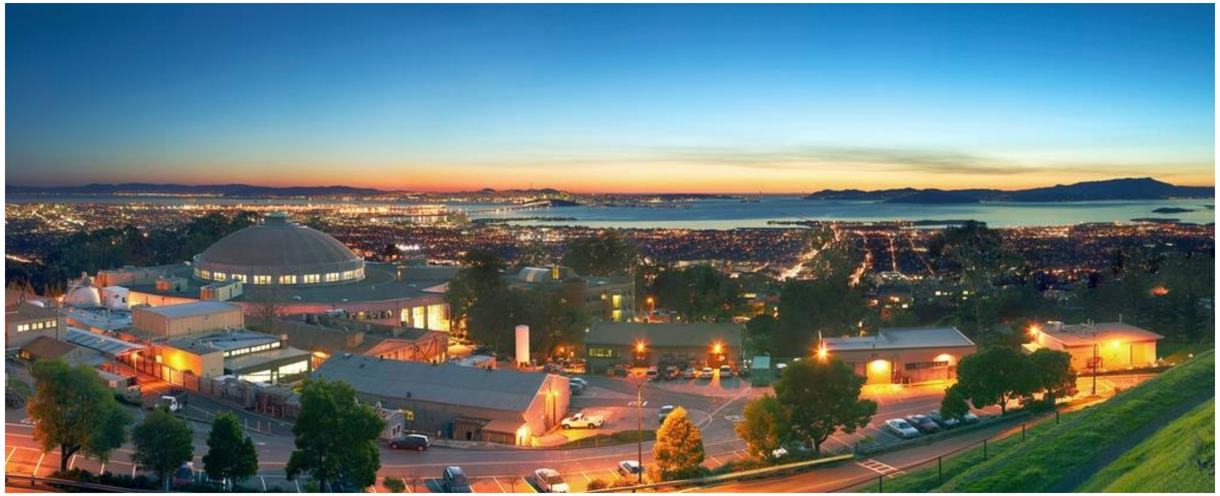
- 1st of 4 Town Halls
- July 22-23
- Held at Advanced Photon Source
- 357 participants
- Introductory remarks by Congressman Bill Foster

AI for Science Town Halls – ORNL

- August 20-21
  - ORNL Conference Center
- 319 participants
- Opening remarks by Steve Binkley
- Keynote: “AI for Science Opportunities” - ORNL AI Program Director David Womble

AI for Science Town Halls – Berkeley

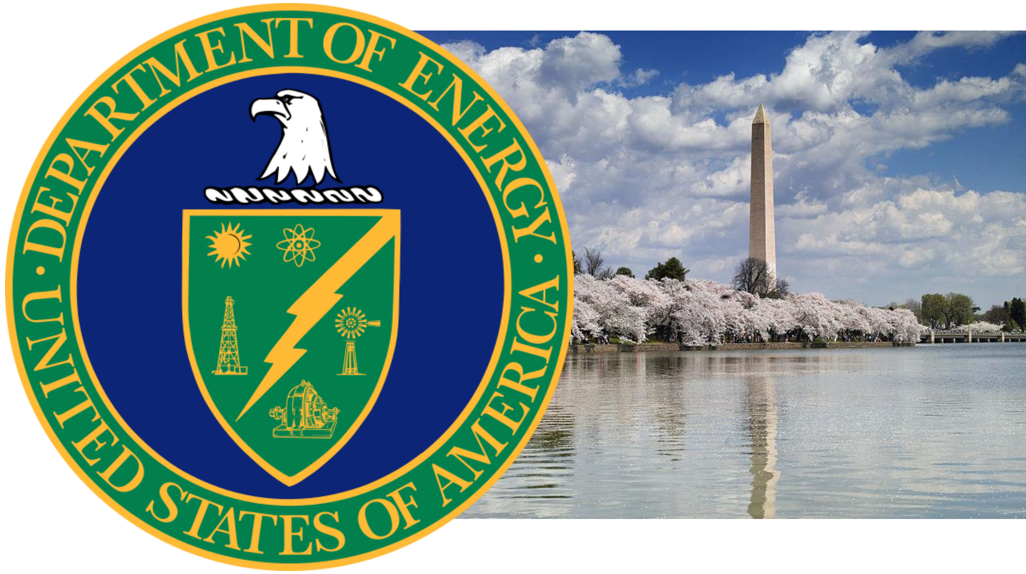
- September 11-12
- Opening remarks by Barb Helland
- 358 attendees (in person)
  - 121 virtual



Oak Ridge National Laboratory

AI for Science Town Halls – Washington, D.C.

- Opening remarks by Chris Fall
- October 22-23
  - Renaissance Hotel, downtown
- Theme: What Can We Do in AI for Science to Move the Country Forward?
- Visionary Keynote
  - "Energy & Investments: The role of AI in changing how business works" - Claire Curry, Bloomberg New Energy Finance



Agenda

Day 1	Day 2
$\frac{X}{AI}$	$\frac{AI}{X}$
Science and engineering breakthroughs that critically use AI at scale	**Methods, computing/networking/data facilities, software, and hardware for AI**

Science Breakthroughs

What: is the challenge problem?

Why: is this important to science, society, etc.?

How: what kind of AI is critical and why?

Scale: what is the data size/rate, compute cost, etc.?

Timeframe: is this a 3,5, or 10-year goal?

Crosscut Challenge Highlight

What problem are you solving?

Why is this important?

Which applications need it?

Why DOE? How does this fit into DOE expertise / facilities / team science?

Cosmology and Astrophysics Breakout

AI for Science, Berkeley Town Hall

Universe: The Movie
Brought to you by AI

Problem: reconstructing the past from the Big Bang until today and predicting the future of our visible Universe, from the largest scales down to our own galaxy, using all existing data (galaxy positions, stellar mass, velocities, dark matter maps, gas distribution, tSZ, kSZ, X-ray).

What is dark energy? What is its density evolution in time? What is the nature of dark matter? Did inflation happen?

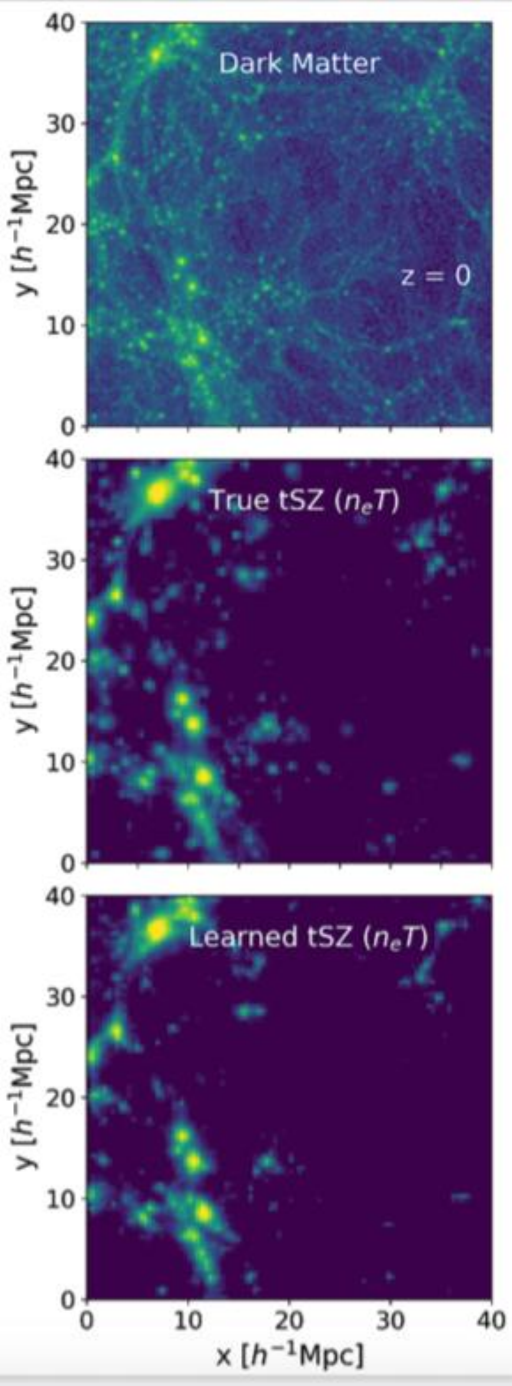
Impact: providing tightest possible constraints on fundamental physics questions as stated above by solving optimal inference problem. Equivalent to several DESI+LSST.

Universe: The Movie
Brought to you by AI

What kind of AI and why? Conventional methods (such as 2 point correlations) are missing information encoded in the data. To optimally extract information while maintaining robustness we need AI combined with statistical methods and HPC simulations. Generative models and discriminative AI models are crucial in solving the problem. Example: AI learned tSZ map can be created 1000 times faster.

Scale: cosmological surveys measure on scales of (10Gpc)^3 and resolve scales of 1Mpc, thus having 10^12 dynamic range and O(100) PB datasets. To solve it we need to combine AI based surrogate models with N-body and hydrodynamic cosmology simulations. To explore the posterior parameter space we will need to run O(100) exascale-class simulations.

Timeframe: 10 year effort to get full dynamical range.



Numerical aspects of learning (improving robustness and stability)

Co-leads:

Sandeep Madireddy (Argonne)

Clayton Webster (ORNL)

Stefan Wild (Argonne)

Writer: Rachel Harken (ORNL)

4 Crosscut Opportunities

- Robust ML and AI approaches to increase trust
- Advanced deep network architecture design
- Characterizing the loss landscape of science-informed ML
- Exploiting variable precision for performance

- **What scientific grand challenges could these address?**
  - Robustness is an essential ingredient for all AI models that augment/model science
  - Particularly important for mission-critical and high-risk scenarios
  - Important theoretical guarantees could increase and justify adoption in complex scientific applications

Robust ML and AI Approaches to Increase Trust

- **What is this?**
  - Build AI models that are accurate and stable, where the stability is with respect to variability/perturbations in the data
  - Develop strategies for achieving this robustness through the choice of architecture, initialization and optimization

- **What is unique to/important for science?**
  - The data, information content and noise distributions are unique compared to industry applications
  - Consequence and risks can be much higher when deployed in practice
  - Computational and numerical capabilities in the DOE ecosystem could be used to improve robustness
  - The diversity of the application space provides a richer testbed for developing and testing robustness approaches

New capabilities imagined

- 3-5 years
  - A suite of databases and challenge problems common and transferable between science applications that can be used to test and improve the robustness in the AI models
  - Novel architecture formulations (e.g., Implicit networks)
  - Domain informed representation learning (input space representations)
  - Robustness metrics be a commonplace in architecture choice
- 10-15 years
  - Rigorous theoretical basis involving approximation theory for domain-informed architecture choice/design
  - Probabilistic approaches to deep learning and uncertainty quantification

Advanced Deep Network Architecture Design

What is this?

- Design innovative deep networks to resolve physical and engineering systems comprised of multiple complex phenomena.
- This requires capabilities that go beyond black-box tools developed by industry that lack important properties, e.g., stability, robustness.

What math needs to be done to enable future advances in science?

- How to design continuous ML models given the desired physical and/or analytical properties?
- How to design proper discretization schemes to enforce the desired numerical properties?
- How to integrate ML models into well-established simulators and accelerate the solvers?

What scientific grand challenges could this address?

- Noisy/adversarial data from experiments/simulation
- High-dimensional input parameters
- Irregular data geometry

New capabilities imagined

- 3-5 years
  - Advance closure models in CFD
  - New pre-conditioners for systems derived in physics and engineering applications
- 10-15 years
  - New deep network architectures that can accommodate unstructured temporal/spatial meshes.
  - Transfer the advanced architectures into the state-of-the-art capabilities in applications

Characterizing the Loss Landscape of Science-informed ML

- **What is this?**
    - Understanding properties of critical points and the landscape around them using visualization and analysis
    - Understand how the loss landscape will affect performance of training algorithms
    - Characterization of the effects of regularizations on loss landscape

- **What is unique to/important for science?**
    - Characterization of the loss landscape in order to expedite training of ML models by informing choice of ML model design (hyperparameter optimization, initializations)
    - Understanding how regularizations encourage certain properties for critical points and landscapes will equip us with new guides to improve the interpretability of neural networks via regularizations.
- **New capabilities imagined (**3-5 years)
    - Tackle short horizon bias: Mathematical theory that can bound the number of critical points for a given loss function
    - Improved understanding of loss landscape for more complicated ML models of practical interests, e.g., GANs, Physics-informed ML

Ø Training parameterized models on near-term quantum devices
Ø Multiplicity, bifurcation and other critical behavior regularly encountered in chemically reacting systems

Exploiting Variable Precision for Performance

- **What is this?**
  - Performance from faster arithmetic, data movement, data storage

- **Challenges**
  - Numerical implication... reevaluate numerical analysis
  - Can it be automated?
  - Issues with different vendor “standards” for half precision
  - Can we verify we have meaningful results?
  - Redefining what accuracy means

- **What is unique to/important for science?**
  - Bigger problems faster
  - Maintaining accuracy while enhancing throughput for scientific exp

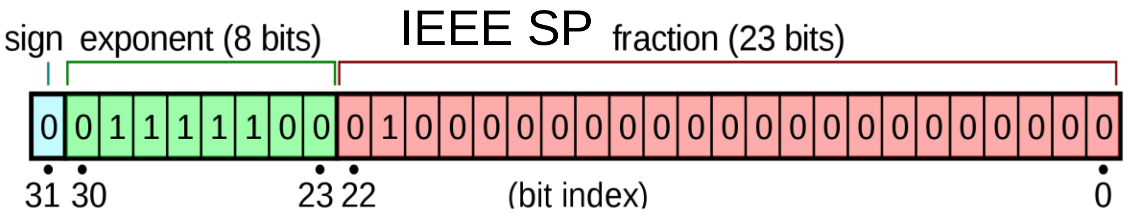
New capabilities imagined

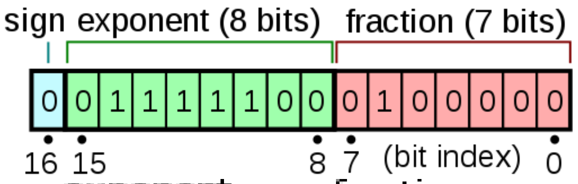
- LA Library for half and mixed precision
- Fast initial guesses and preconditioners
- Precompute minimum precision to train models
- Precision-informed neural architecture search

What scientific grand challenges could this address?

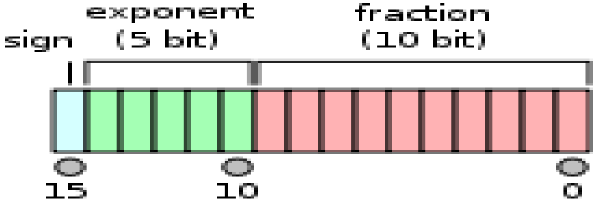
- Make edge computing a reality for science
- Higher fidelity models for same cost

Type	Size	Range	$u = 2^{-t}$
half	16 bits	$10^{\pm 5}$	$2^{-11} \approx 4.9 \times 10^{-4}$
single	32 bits	$10^{\pm 38}$	$2^{-24} \approx 6.0 \times 10^{-8}$
double	64 bits	$10^{\pm 308}$	$2^{-53} \approx 1.1 \times 10^{-16}$
quadruple	128 bits	$10^{\pm 4932}$	$2^{-113} \approx 9.6 \times 10^{-35}$





Google TPU: bfloat16
largest fl pt number $O(10^{38})$



IEEE float16

largest fl pt number
65,504

Domain Breakouts

Argonne	Oak Ridge	Berkeley	DC
Materials, Chemistry, and Nanoscience	Materials, Chemistry, and Nanoscience	Materials, Chemistry, and Nanoscience	Materials, Chemistry, and Nanoscience
		Materials Synthesis and Chemistry	
Environment, Climate and Earth Science	Environment, Climate and Earth Science	Climate and Carbon	Earth Systems
		Subsurface and Geoscience	
		Water	
Biology and Life Science	Biology and Life Science	Synthetic Biology	Biology and Life Sciences
		Health	
Fundamental Physics	Fundamental Physics	Cosmology and Astrophysics	Fundamental Physics
		Particle Physics	
		Accelerator Science	

Domain Breakouts

Argonne	Oak Ridge	Berkeley	DC
Engineering and Technology	Advanced Manufacturing	Engineering and Manufacturing	Engineering Manufacturing
Energy (wind, solar, fossil, etc.)	Transportation and Mobility	Transportation / Mobility	Smart Energy Infrastructure
		Urban	
	Energy Generation & Distribution	SmartGrid	
		AI Networking and Computing Facilities	AI for Computer Science
		AI for Computer Hardware and Software	
[Energy (wind, solar, fossil, etc.)]	Fusion	Fusion	Fusion

Crosscuts Breakouts

<table>
<tr><th>Argonne</th><th>Oak Ridge</th><th>Berkeley</th><th>DC</th></tr>
<tr><td rowspan="3">Optimization /UQ /Statistics</td><td rowspan="2">Numerical Aspects of Learning</td><td>Performance Optimization of Deep Learning</td><td rowspan="8">Machine Learning Foundations and Open Problems</td></tr>
<tr><td>Foundations and Challenges of Deep Learning</td></tr>
<tr><td>Model Applicability and Characterization</td><td>Opportunities and Foundations of Traditional Machine Learning</td></tr>
<tr><td rowspan="5">Convergence of Simulation and Data Methods</td><td>Decision Support</td><td>Reinforcement/Streaming learning for Decision support and control</td></tr>
<tr><td rowspan="4">Science Informed Learning</td><td>ML for science problems with limited data</td></tr>
<tr><td>Science-informed learning</td></tr>
<tr><td>Uncertainty Quantification</td></tr>
<tr><td>Use of AI with Simulation</td></tr>
</table>

Crosscut Breakouts

Argonne	Oak Ridge	Berkeley	DC
Software Environments and Software Research	Software Environments and Software Research	Software Environments and Research	Software Environments and Software Research
Data Infrastructure and Life Cycle	Data Infrastructure and Life Cycle	Data Life Cycle	Data Life Cycle & Infrastructure
Hardware and Architecture	Hardware and Architecture	Hardware Technology	Hardware Architectures
Imaging and Scientific User Facilities	Data Collection, Reduction, Analysis, and Imaging for Scientific User Facilities	Light Sources Electron Microscopy Imaging	Support for AI for Experimental Facilities
[Facilities Integration]	*[Facilities Integration and AI Ecosystem]*	*[Facilities Infrastructure and Integration; the AI Ecosystem]*	Support for AI at the Edge
Facilities Integration	Facilities Integration and AI Ecosystem	Facilities Infrastructure and Integration; the AI Ecosystem Cybersecurity and Privacy	Facilities Integration and AI Ecosystem

What Happens Next?

- Domains/Crosscuts (7 + 7) working to create summary documents (6-page) and presentations (12 slide) based on the prior 3 townhall artifacts
- These summary documents are presented, vetted and tweaked at the DC townhall
- Summary documents are finalized and used to create townhall report to ASCR
- Next steps are based on the need for BRN for each of the relevant offices; SC, applied, NNSA, etc.
- Assessment of final steps need to be determined by long term goals of ASCR/SC/DOE: program versus project, etc.