Exascale Science on Frontier

Bronson Messer
Director of Science
Leadership Computing Facility
Oak Ridge National Laboratory

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

Outline

- Frontier as a unique scientific instrument
- A look at a handful of recent scientific results and how they take advantage of exascale computing
- How to increase computational capability in the future? (Waiting for Moore's Law to do miracles isn't a plan.)
- Questions

Frontier

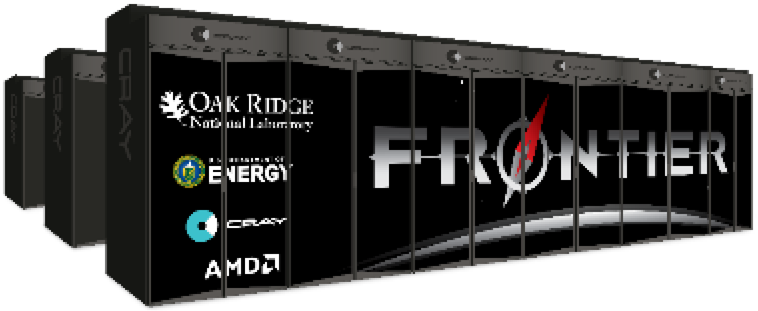

FRONTIER
OAK RIDGE National Laboratory
U.S. DEPARTMENT OF ENERGY
Hewlett Packard Enterprise
AMD


FRONTIER
CRAY


CRAY


Oak Ridge National Laboratory
Computer Rooms
FRONTIER
OAK RIDGE
National Laboratory
U.S. DEPARTMENT OF
ENERGY
Hewlett Packard
Enterprise
AMD
18

Frontier Overview



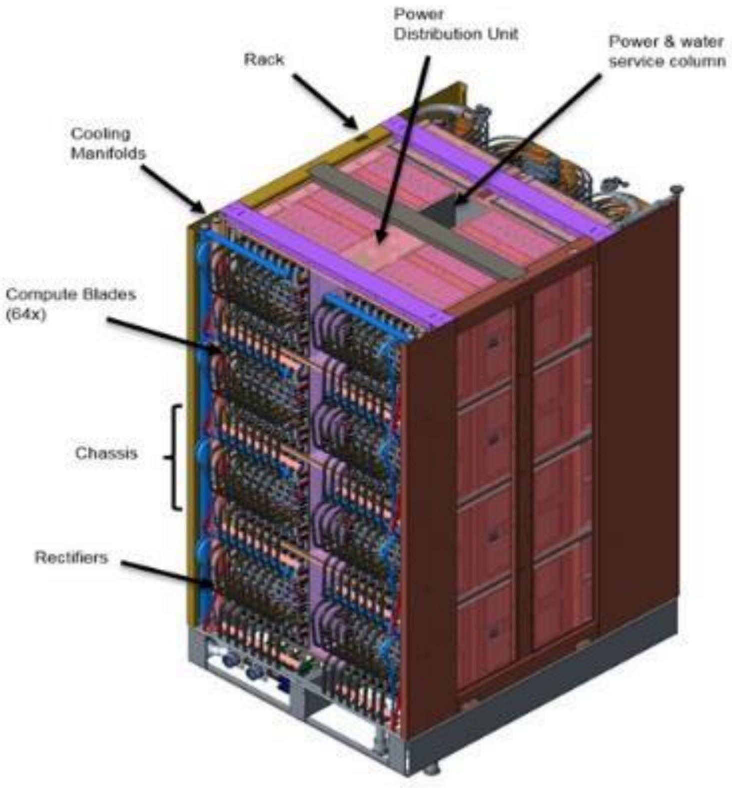
System

- 2 EF Peak DP FLOPS
- 74 compute racks
- 29 MW Power Consumption
- 9,408 nodes
- 9.2 PB memory
  (4.6 PB HBM, 4.6 PB DDR4)
- Cray Slingshot network with dragonfly topology
- 37 PB Node Local Storage
- 716 PB Center-wide storage
- 4000 ft^2 foot print

Built by HPE

Olympus rack

- 128 AMD nodes
- 8,000 lbs
- Supports 400 KW



Powered by AMD

AMD node

- 1 AMD "Trento" CPU
- 4 AMD MI250X GPUs
- 512 GiB DDR4 memory on CPU
- 512 GiB HBM2e total per node
  (128 GiB HBM per GPU)
- Coherent memory across the node
- 4 TB NVM
- GPUs & CPU fully connected with AMD Infinity Fabric
- 4 Cassini NICs, 100 GB/s network BW

Compute blade

- 2 AMD nodes

All water cooled, even DIMMS and NICs

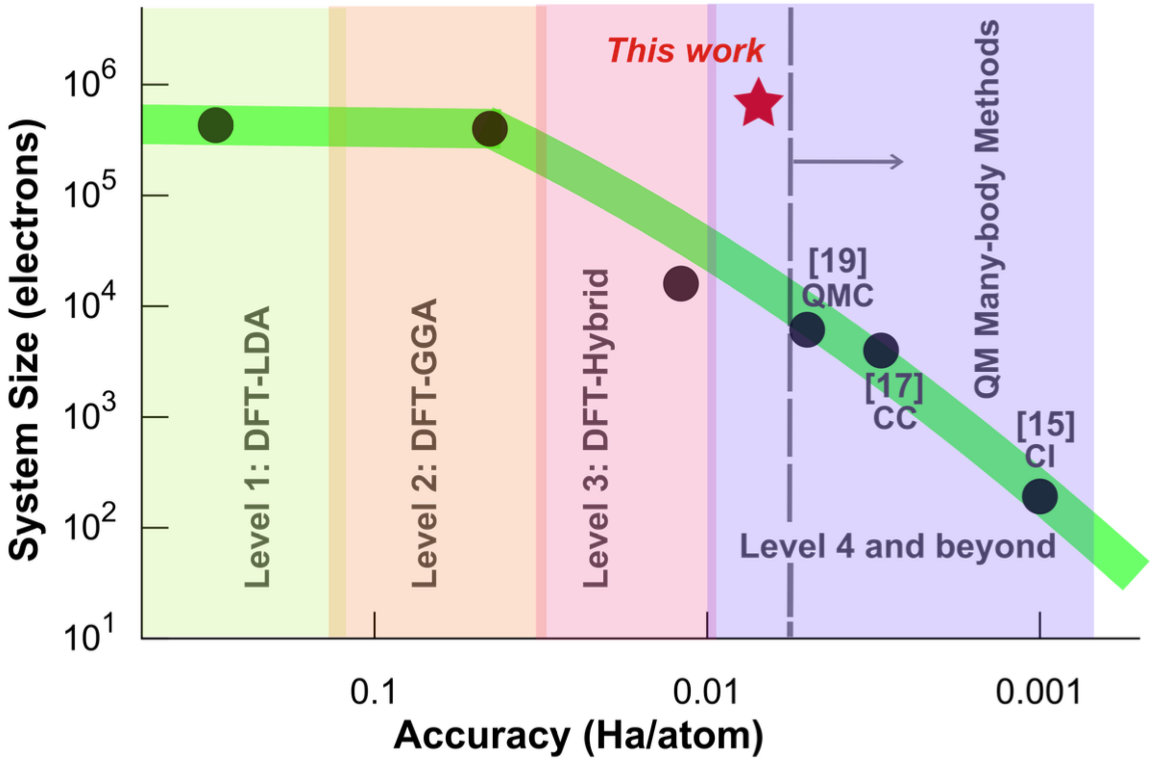
So, what do you do with all that computing power, memory, storage, etc.?

- Short answer: Everything!
- No other scientific facility has as much variety in the number and kind of scientific and technical problems that can be attacked.
- But some things are obvious candidates:
  - Problems that require resolution of a vast range of scales in space, time, etc.
  - Problems where a lot of physics is going on at every spatial point (i.e. combustion chemistry, radiation transport, etc.)
  - Problems where we need to perform many high-fidelity simulations to understand how things change with input parameters

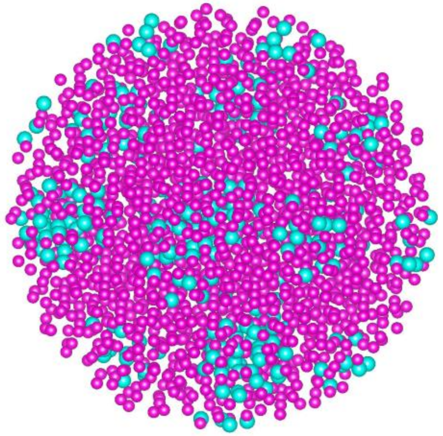
2023 Gordon Bell Prize

Simulations used exascale computing to approach quantum accuracy

- A University of Michigan-led team of eight scientists won the Association for Computing Machinery's 2023 Gordon Bell Prize on November 16 for their study that used the Frontier supercomputer to run one of the largest simulations of an alloy ever and achieve near-quantum accuracy.
- The team used density functional theory to simulate design and discovery of new materials. The model achieved the fastest sustained performance recorded for materials simulation: 659.7 petaflops using Frontier.



The Gordon Bell Prize team accepting their award at SC23. Credit: SC23

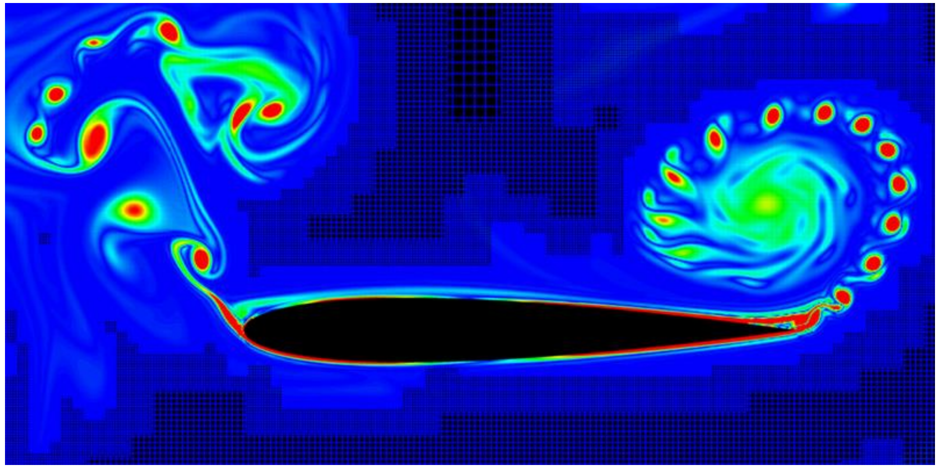


Researchers used Frontier to simulate a magnesium system of nearly 75,000 atoms. Credit: Vikram Gavini/ Univ. of Michigan

An ill-posed question, but…

- **What is the "killer app" for exascale computing?**
- I don't like this question. A lot of the value of unique supercomputing facilities is the ability to impact a huge variety of scientific problems, as pointed out earlier.
- But people ask it…
- Maybe there's not a killer app, but there is a ubiquitous physical problem that requires:
  - More memory (e.g. resolution)
  - Faster compute speed
  - Inclusion in multiphysics simulations…

Understanding turbulence



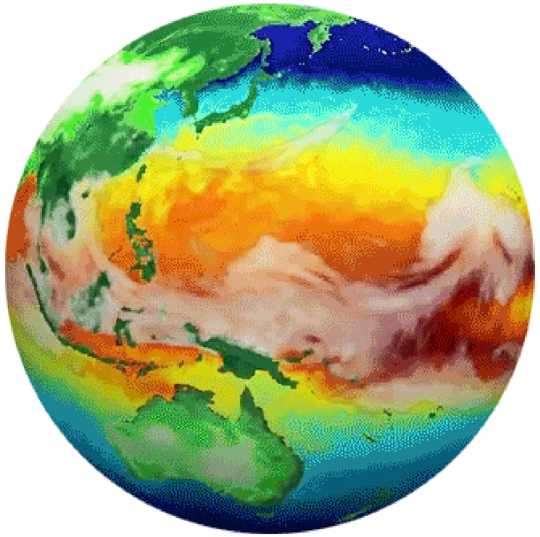
- "The last great unsolved problem in classical physics" (One of the 7 Millennium Problems)
- Werner Heisenberg assuredly never said: "When I meet God, I'm going to ask him two questions: Why relativity? And why turbulence? I really believe he'll have an answer for the first."
- We remain far away from being able to resolve turbulent physics from the largest scales where it is generated (even in terrestrial settings) to the molecular dissipation scale.
- But, there are many places where turbulence is important where other physics arrests the impact of the turbulent cascade before it gets to the smallest scales.

2023 Gordon Bell Special Prize for Climate Modeling

- A 19-member team of scientists from across the national laboratory complex won the Association for Computing Machinery's 2023 Gordon Bell Special Prize for Climate Modeling on November 16 for developing a model that uses Frontier to simulate decades' worth of cloud formations.

- The research team used Frontier to achieve record speeds in modeling worldwide cloud formations in 3D. The computational power of Frontier shrinks the work of years into days to bring detailed estimates of the long-range consequences of climate change and extreme weather within reach.

- On Frontier, it obtained a record setting performance of 1.26 simulated years per day (SYPD) for a realistic cloud resolving simulation (using 8192 Frontier nodes).



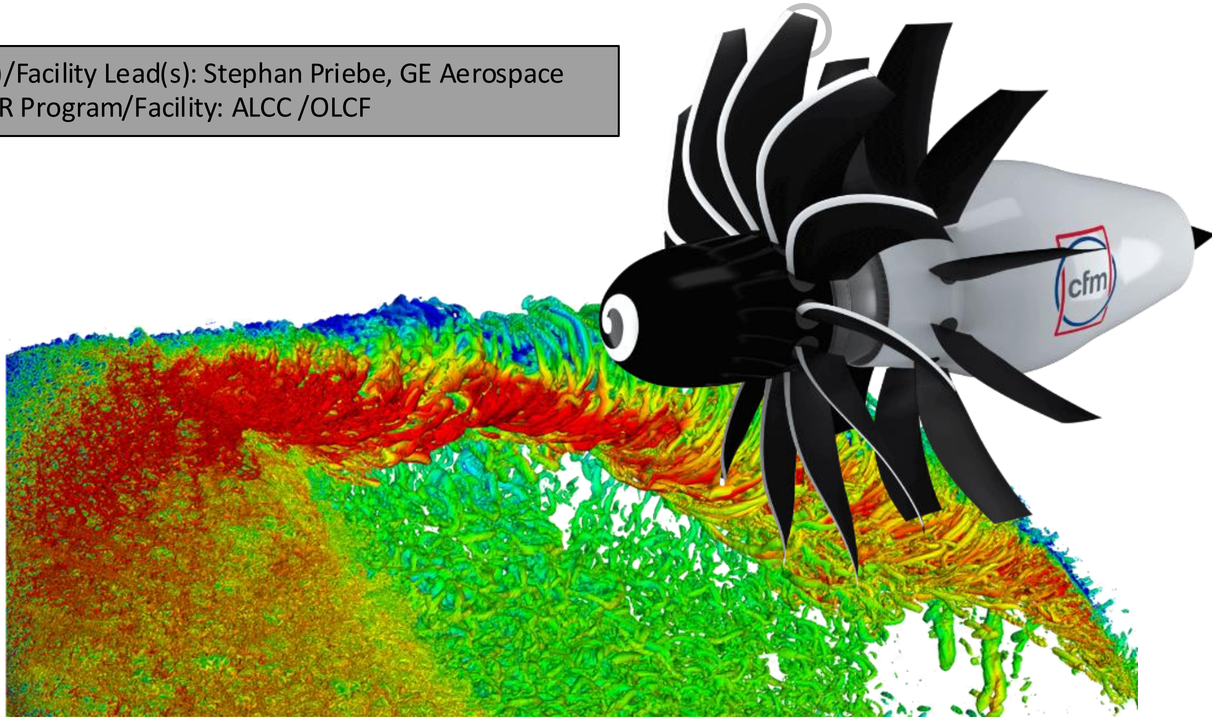
The Gordon Bell Special Prize team accepting their award at SC23. Credit: SC23

Frontier enables first-ever insights into product-scale open fan aeroacoustic performance under realistic conditions

The Science
To evaluate the potential efficiency delivered by an open fan architecture while simultaneously reducing noise levels, designers need to understand airflow behavior around the blades of the engine, including the complex physics of turbulence, mixing, and potential flow for separation. GE Engineers used Frontier to perform first-ever 3D Large Eddy Simulations at realistic flight-scale conditions with unprecedented detail toward revealing breakthrough insights into improving aerodynamic and acoustic performance in next generation aircraft engine technology. The high-fidelity simulations performed on Frontier provide unique insights to better understand important factors in novel propulsion design, such as how complex turbulent flow physics measured in smaller scale prototype rig tests scale up to flight performance of the full-scale engine. The prior state-of-the-art was limited to a reduced scale by the available computational resources (as computational costs grow exponentially with size). Those simplifications introduce uncertainty when translating engineering knowledge and learnings to the "real thing."

The Impact
The project lends crucial support to pursuit of reduction and eventual elimination of fossil fuel use for aircraft propulsion and achieve sustainable energy in flight. With its partner Safran, GE announced in 2021 the Revolutionary Innovation for Sustainable Engines (RISE) technology demonstration program to deliver over 20% lower fuel consumption and CO_2 emissions compared to today's most efficient aircraft engines. Fluid dynamics simulations on Frontier are enabling GE to extend learnings from prior research to realistic flight conditions and discover new ways of controlling turbulence, improve fan performance, and more efficiently guide future physical testing. The study also offers a rich high-fidelity dataset from which machine learning algorithms can train reduced order models useful in exploring design trade-offs.

PI(s)/Facility Lead(s): Stephan Priebe, GE Aerospace
ASCR Program/Facility: ALCC /OLCF



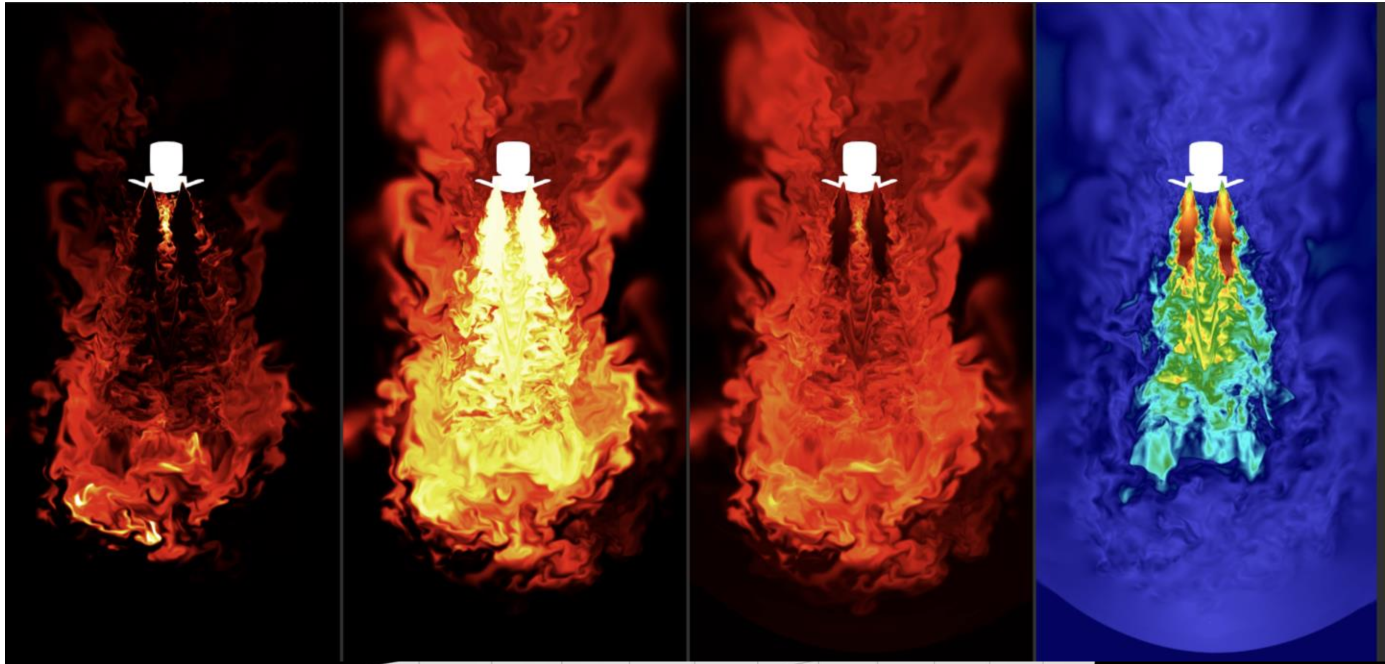
The RISE engine featuring open fan blades (top). GE visualization of turbulent flow in the tip region of an open fan blade only made possible using the Frontier supercomputer (bottom). Credit: CFM, GE Research.

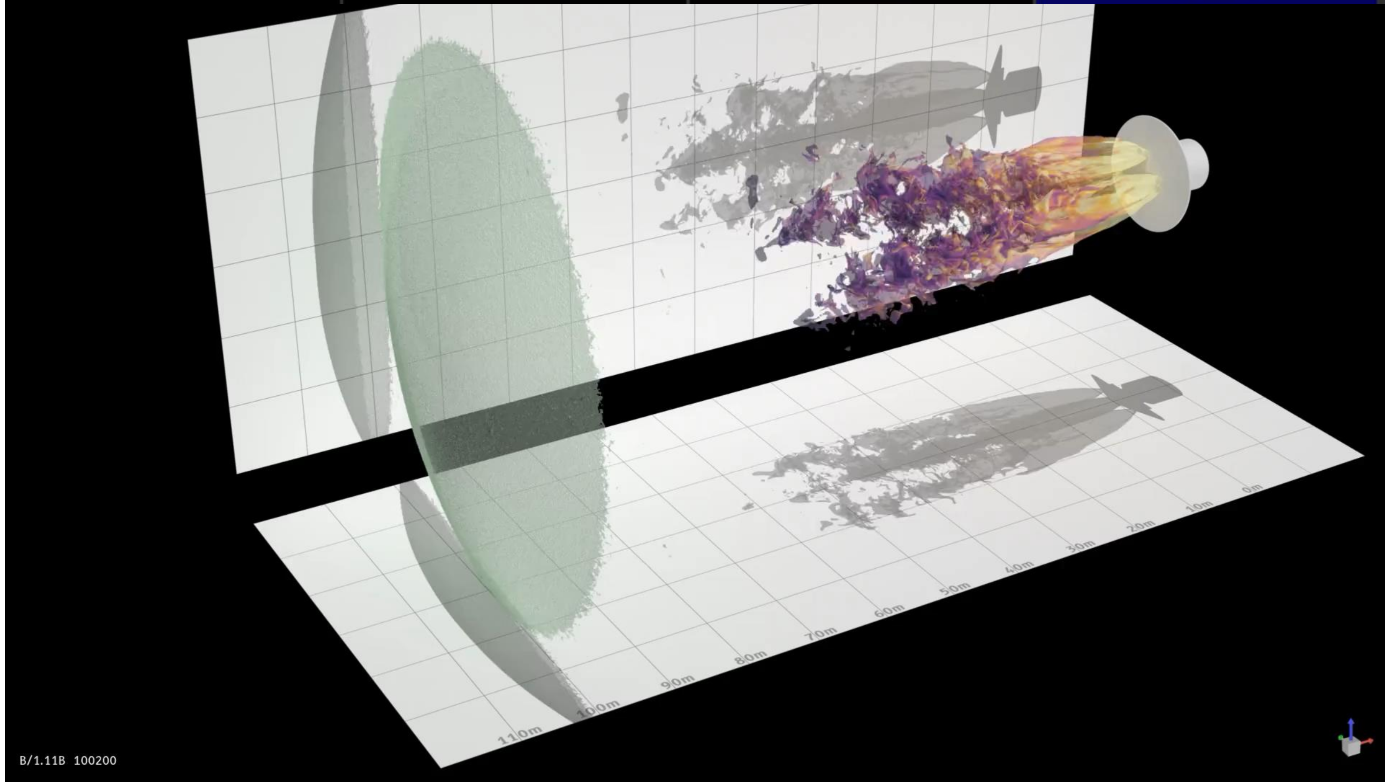
"Frontier has unlocked the ability for us to numerically fly our fans during the design phase, years in advance of actual flight testing, with a level of detail previously unattainable." —Dr. Stephan Priebe, Senior Engineer in Computational Fluid Dynamics & Methods, GE Research

Significant overachievement (ALCC)

Flying on Mars

- A vehicle carrying astronauts and all their life-support systems will weigh 20 to 50 tons or more.
- Mars' thin atmosphere — about 100 times less dense than Earth's — won't support a parachute landing for such a large craft.
- Answer: retropropulsion
- But: Steady flight in the Martian atmosphere means understanding induced turbulence *in that atmosphere*.
- "The team focused on the ultimate simulation they had hoped for years earlier: a truly autonomous, closed-loop test flight leveraging the world's most powerful supercomputing system."





New Clues to Improving Fusion Confinement

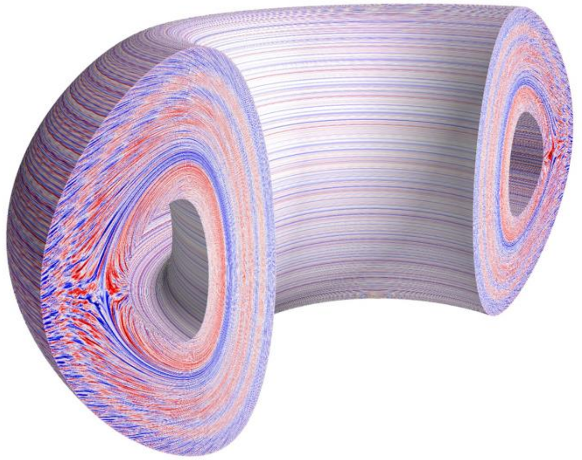
Using ORNL’s Frontier exascale supercomputer, researchers discover an unexpected feature of multi-scale turbulence in the edge of a fusion reactor’s plasma

The Science

Scientists have developed sophisticated techniques to improve the confinement of superhot plasma within tokamak fusion reactors. Confinement in tokamaks is limited by the small but steady heat loss caused by turbulence. One important method to improve tokamak confinement is a byproduct of the commonly used neutral beam injection, or NBI, system. NBI uses an intense particle beam to heat the plasma to 150 million degrees Celsius for fusion to occur. In addition to heating the plasma, this beam also causes the plasma to rotate around the tokamak’s chamber. Previous studies had indicated that this rotation would reduce the turbulence caused by ions. It was also theorized that the rotation would not affect the electrons because they are so light and fast. The results of a new study on Frontier contradict that theory.

The Impact

The researchers found that with the interaction between ions and electrons, the rotation can increase the overall level of turbulence and subsequently reduce the quality of confinement. While the ions’ turbulence is favorably reduced, the total turbulence can increase, which is unfavorable for a reactor. The team’s findings could help to optimize the design of future tokamaks, including ITER, the international collaboration in France to build the world’s largest tokamak.



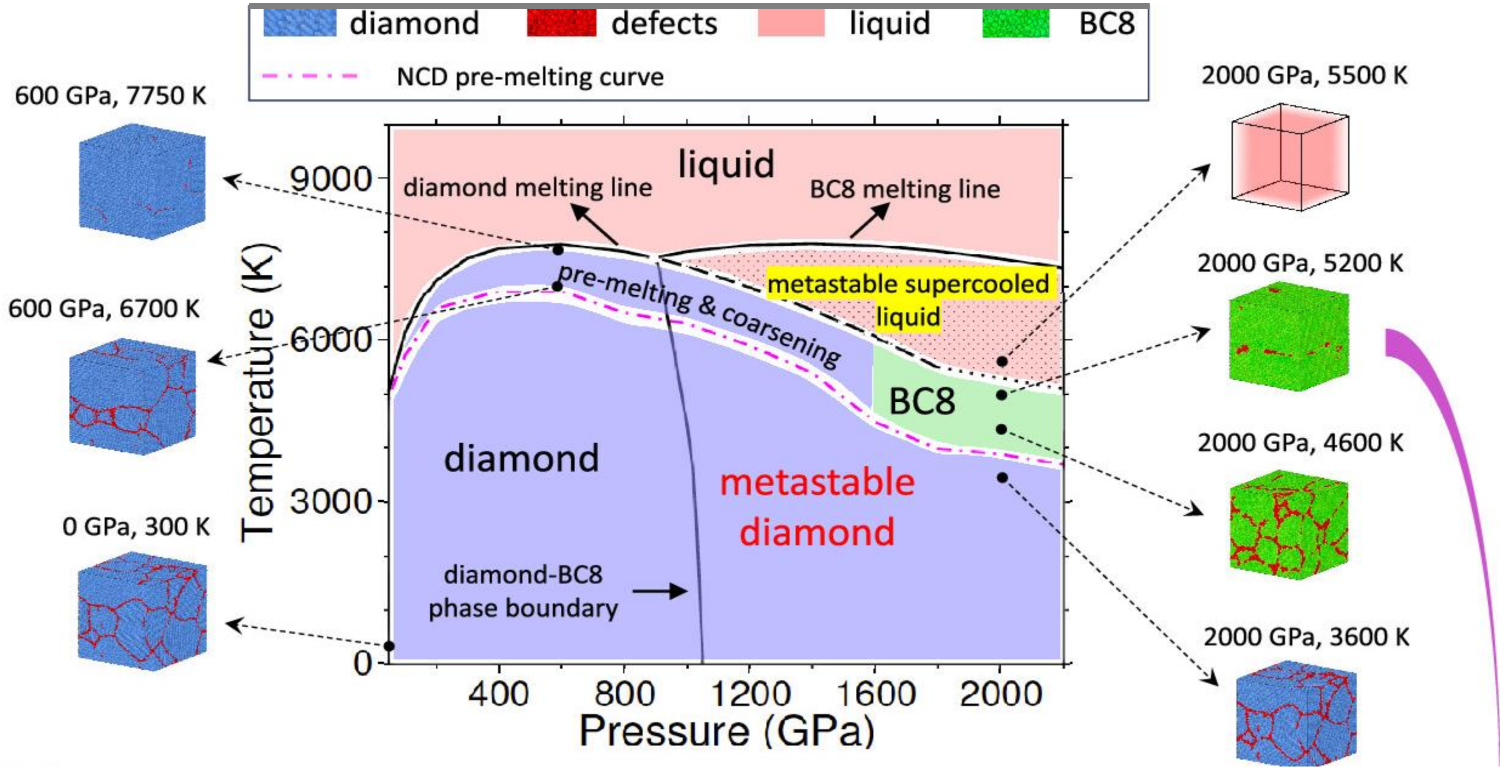
Researchers used the CGYRO gyrokinetic code to create a multiscale simulation of plasma temperature fluctuations driven by turbulence. The model was based on data from DOE’s DIII-D National Fusion Facility for an “ITER baseline” tokamak scenario. Image: Emily Belli, General Atomics

PI/Facility Lead(s): Emily Belli, General Atomics
Oher team members: Jeff Candy, General Atomics; Igor Sfiligoi, UC San Diego
ASCR Program/Facility: OLCF/Frontier
ASCR PM: Christine Chalk
Funding: INCITE
Publication(s) for this work: E. A. Belli, J. Candy, and I. Sfiligoi. “Flow-shear destabilization of multiscale electron turbulence,” *Plasma Physics and Controlled Fusion* (2024): 10.1088/1361-6587/ad2c28

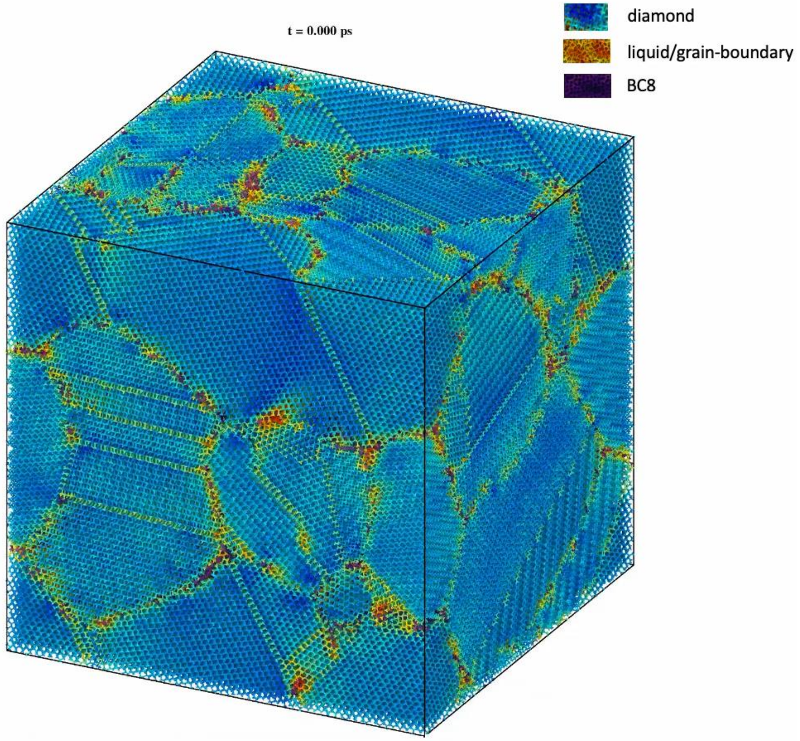
Frontier Simulations Could Help Build a Better Diamond

Supercomputer study guides effort to create otherworldly BC8 on Earth

Classic interatomic potentials unable to accurately traverse this phase diagram at the requisite accuracy. To achieve the required level of accuracy, the team trained an ML interatomic model by using an extensive cache of quantum mechanical data on various states of carbon, including BC8.



K. Nguyen-Cong et al, "Extreme Metastability of Diamond and its Transformation to BC8 Post-Diamond Phase of Carbon"



PI/Facility Lead(s): Ivan Oleynik, Univ. South Florida
ASCR Program/Facility: OLCF/Frontier
ASCR PM: Christine Chalk
Funding: ALCC
Publication(s) for this work: Kien Nguyen-Cong, et al., *The Journal of Physical Chemistry Letters* 2024 15 (4), 1152-1160

LAMMPS on 8000 Frontier nodes

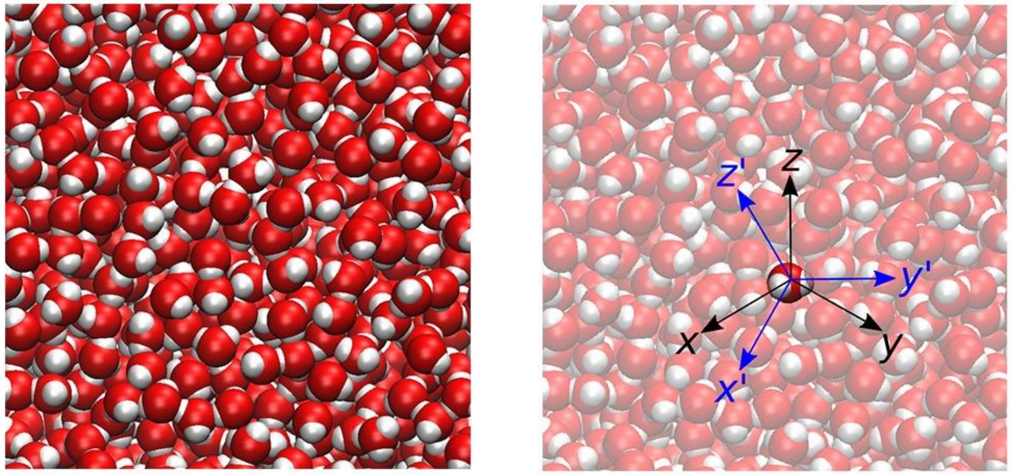
Study reveals flaw in long-accepted approximation used in water simulations

The Science

Computational scientists at DOE's ORNL have published a study that questions a long-accepted factor in simulating the molecular dynamics of water: the 2 femtosecond (one quadrillionth of a second) time step. According to the team's findings, using anything greater than a 0.5 femtosecond time step can introduce errors in both the dynamics and thermodynamics when simulating water using a rigid-body description. One of the tenets of statistical mechanics is that, if a system is in equilibrium, then the temperatures associated with its translational motion (movement along a line) and rotational motion should be the same. If those two temperatures differ, then the simulation is not in equilibrium.

The Impact

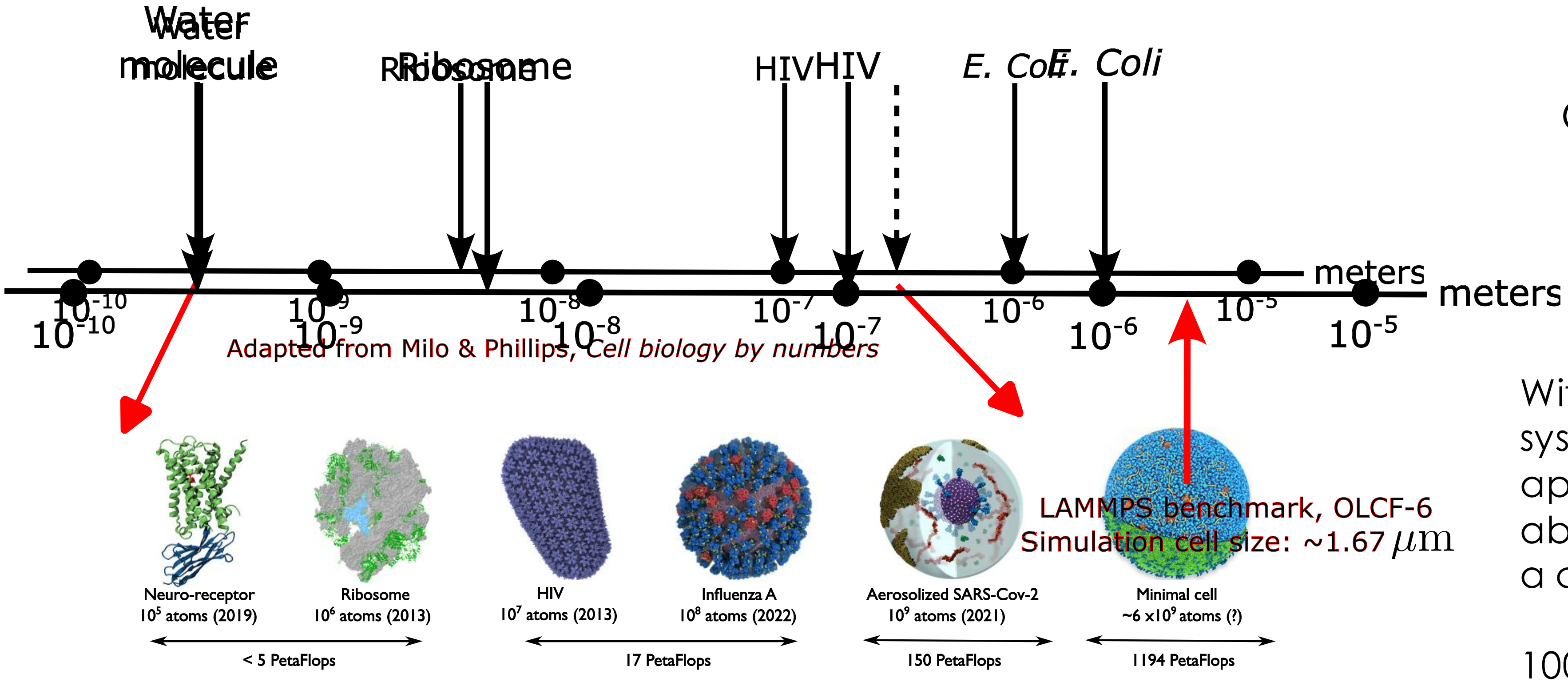
The 2-femtosecond time step has been accepted as a standard in water simulations for almost 50 years. This does mean that code performance must be raised to achieve the same physical times in the same wallclock time.



This schematic shows the rotational motion of a specific water molecule. At equilibrium, on average, the energy associated with rotational motion should equal the energy associated with the translation of the molecule as a whole. Credit: Dilip Asthagiri/ORNL, U.S. Dept. of Energy

PI/Facility Lead(s): Dilip Asthagiri and Tom Beck, ORNL
ASCR Program/Facility: OLCF/Summit
ASCR PM: Christine Chalk
Funding: OLCF Staff Program 2023
Publication(s) for this work: D. Asthagiri and T. Beck, "MD Simulation of Water Using a Rigid Body Description Requires a Small Time Step to Ensure Equipartition," *Journal of Chemical Theory and Computation* (2024): https://doi.org/10.1021/acs.jctc.3c01153.

Study reveals flaw in long-accepted approximation used in water simulations



Beck, Carloni, Asthagiri, All-atom simulations in the exascale era, JCTC **20**, 1777 (2024).

OLCF-6 LAMMPS Benchmark

With 90% memory footprint, largest system we could simulate has approximately 4.66E11 atoms or about 155 billion water molecules in a cube of side ≈ 1.6 μm

100 steps (at 0.5 fs timestep!) of the benchmark on 9261 nodes required about 16 minutes : ≈7.5 minutes for execution and the rest for start up (replicating the system, etc.)

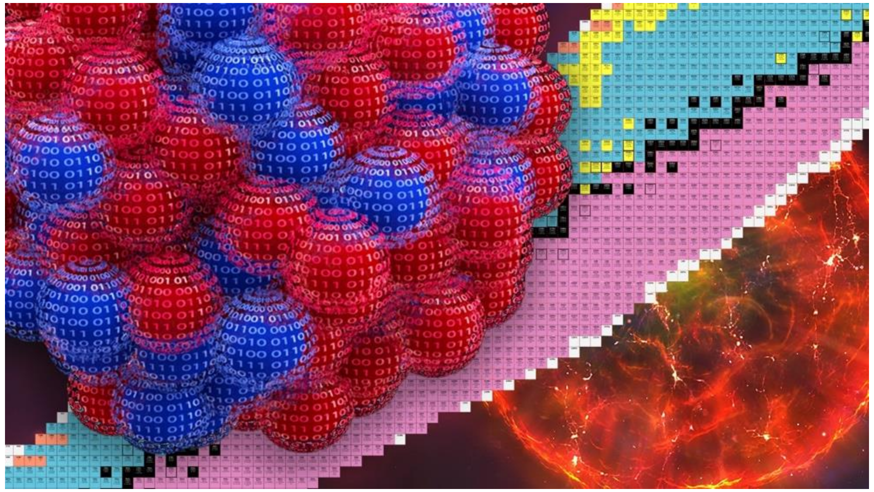
Frontier Simulations Provide New Insights Into Calcium-48's Magnetic Properties

The Science

Calcium-48 is an important isotope in a variety of settings. Ithe nucleus is composed of 20 protons and 28 neutrons — a combination that scientists call "doubly magic." Magic numbers — such as 20 and 28 — are specific numbers of protons or neutrons that provide stability by forming a closed shell within the nucleus. The strong binding and simple structure of calcium-48 also makes it an interesting test subject for studying the strong and weak nuclear forces that hold nuclei together or break them apart.

The Impact

Nuclear physicists at Oak Ridge National Laboratory used Frontier to calculate the magnetic properties of calcium-48's atomic nucleus. Their findings will not only provide a better understanding of how magnetism manifests inside other nuclei but will also help to resolve a decade-old disagreement between experiments that drew different conclusions about calcium-48's magnetic behavior. Additionally, the research could provide new insights into the subatomic interactions that happen inside supernovae.



The Frontier supercomputer simulated magnetic responses inside calcium-48, depicted by red and blue spheres. Insights into the nucleus's fundamental forces could shed light on supernova dynamics. Credit: ORNL, U.S. Dept. of Energy

PI/Facility Lead(s): Gaute Hagen, ORNL
ASCR Program/Facility: OLCF/Frontier
ASCR PM: Christine Chalk
Funding: INCITE
Publication(s) for this work: B. Acharya, B. S. Hu, S. Bacca, G. Hagen, P. Navrátil, and T. Papenbrock. "Magnetic Dipole Transition in ^{48}Ca," *Physical Review Letters* (2024): 10.1103/PhysRevLett.132.232504

Oak Ridge National Laboratory | Leadership Computing Facility

The impact of the ^{48}Ca result on supernovae

- The matrix elements describing the magnetic dipole response also appear in the calculation of the electron capture/β-decay rates.

- The effects on supernova shock propagation from differences of this size have been known…for a long time…

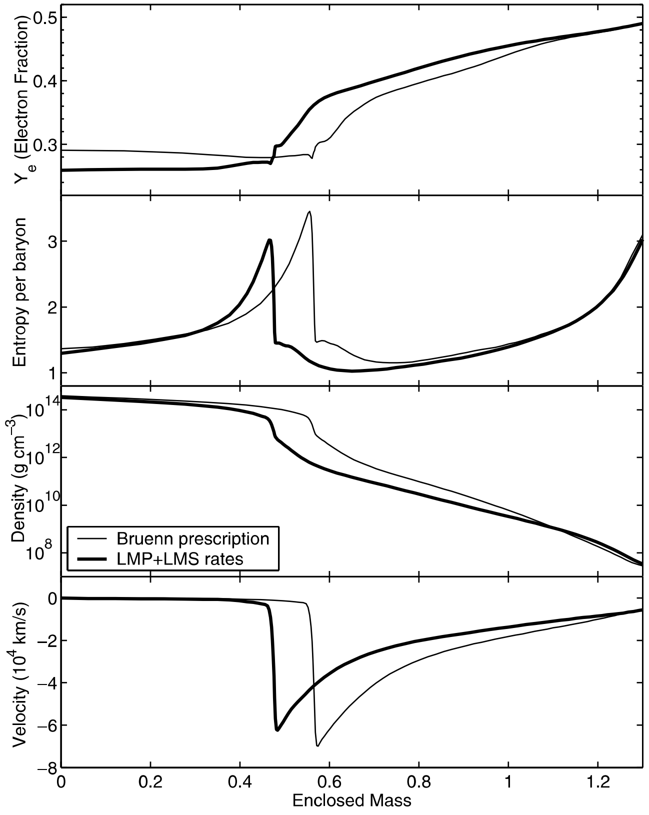


FIG. 1. The electron fraction, entropy, density, and velocity as functions of the enclosed mass at the beginning of bounce for a 15 $M_\odot$ model. The thin line is a simulation using the Bruenn parametrization, while the thick line is for a simulation using the LMP and hybrid reaction rate sets.

Hix, Messer, et al. (2003)

Science demands continue to grow, and the exascale has *accelerated* their growth.

- The level of fidelity that exascale computing provides has opened up a wide range of possibilities across disciplines. *This has been catalyzed by innovation and software from ECP*.

- New demands stem from
  - The need to include and accurately account for larger domains (more coupling, environmental factors, etc.)
    - e.g., fusion, human digital twins, engineering turbulent flows

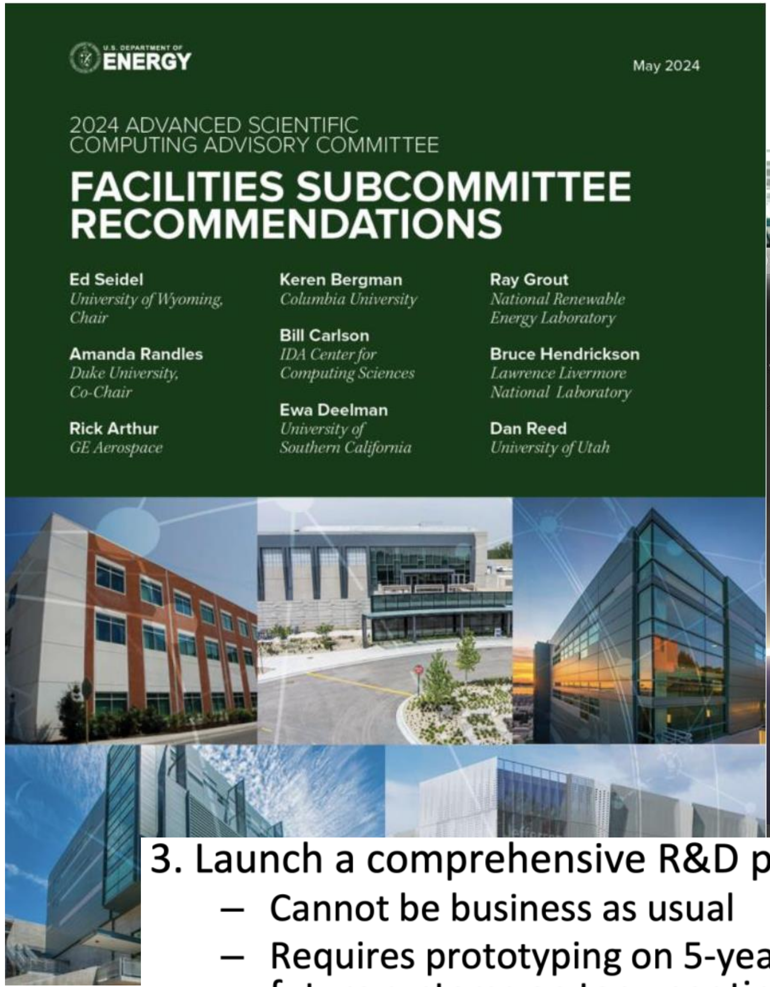
    memory and memory BW
  - The need to increase throughput for high-fidelity simulations and for experimental data analyses
    - AI help for design of experiments is essential here!

    AI/ML, WAN, and more usable FLOP/s
  - The need to relax approximations that have been/are being shown to break down as capability increases
    - sub-grid models — AI/ML
    - improving useful, but "wrong" theories (e.g., MD) — quantum computing
    - coupling in multiphysics — algorithms

- <u>Increases in capability are required to meet these demands.</u>

Vendor engagement needed to realize these requirements

- Post-Exascale HPC deployments facing significant challenges:
  - Transistor costs dramatically increasing
  - Power per SoC dramatically increasing
  - Architectural choices dominated by AI
  - Network power budgets rapidly growing as bandwidth increases


U.S. DEPARTMENT OF ENERGY

May 2024

2024 ADVANCED SCIENTIFIC COMPUTING ADVISORY COMMITTEE

FACILITIES SUBCOMMITTEE RECOMMENDATIONS

Ed Seidel, *University of Wyoming, Chair*

Amanda Randles, *Duke University, Co-Chair*

Rick Arthur, *GE Aerospace*

Keren Bergman, *Columbia University*

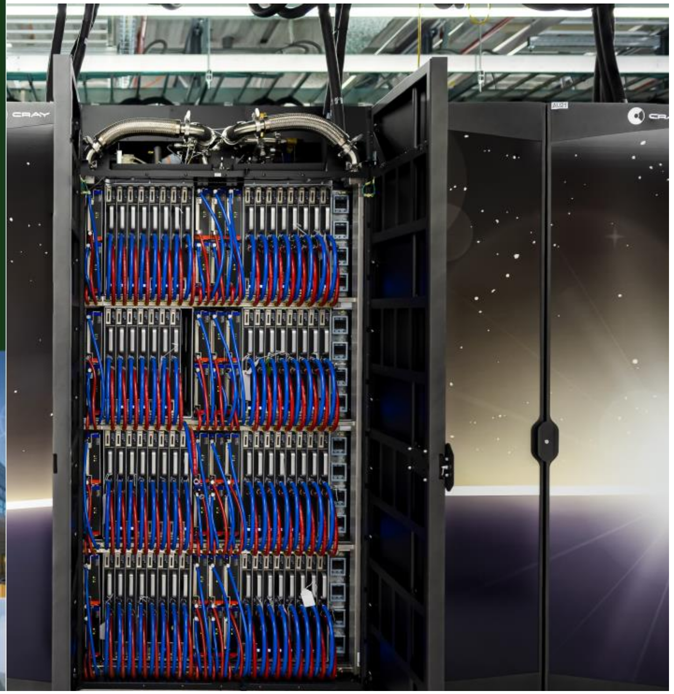
Bill Carlson, *IDA Center for Computing Sciences*

Ewa Deelman, *University of Southern California*

Ray Grout, *National Renewable Energy Laboratory*

Bruce Hendrickson, *Lawrence Livermore National Laboratory*

Dan Reed, *University of Utah*



3. Launch a comprehensive R&D program
   - Cannot be business as usual
   - Requires prototyping on 5-year timescale that informs pathways to future systems on ten-year timescale
   - DOE cannot go it alone; other agencies, vendors, science communities all needed
   - New governance model needed, both within DOE and across agencies

New Frontiers

- Based on the *-Forward vendor R&D tradition
- Reinvigorating vendor R&D programs
- ***Develop <u>perennial</u> 2 Year Programs***
- First 2-year initial program RFP released Sept 6th
- Themes
  - Hardware emphasis on energy efficiency
  - Software emphasis on sustainable software
  - Cross-cutting maturation of software-hardware ecosystems

New Frontiers

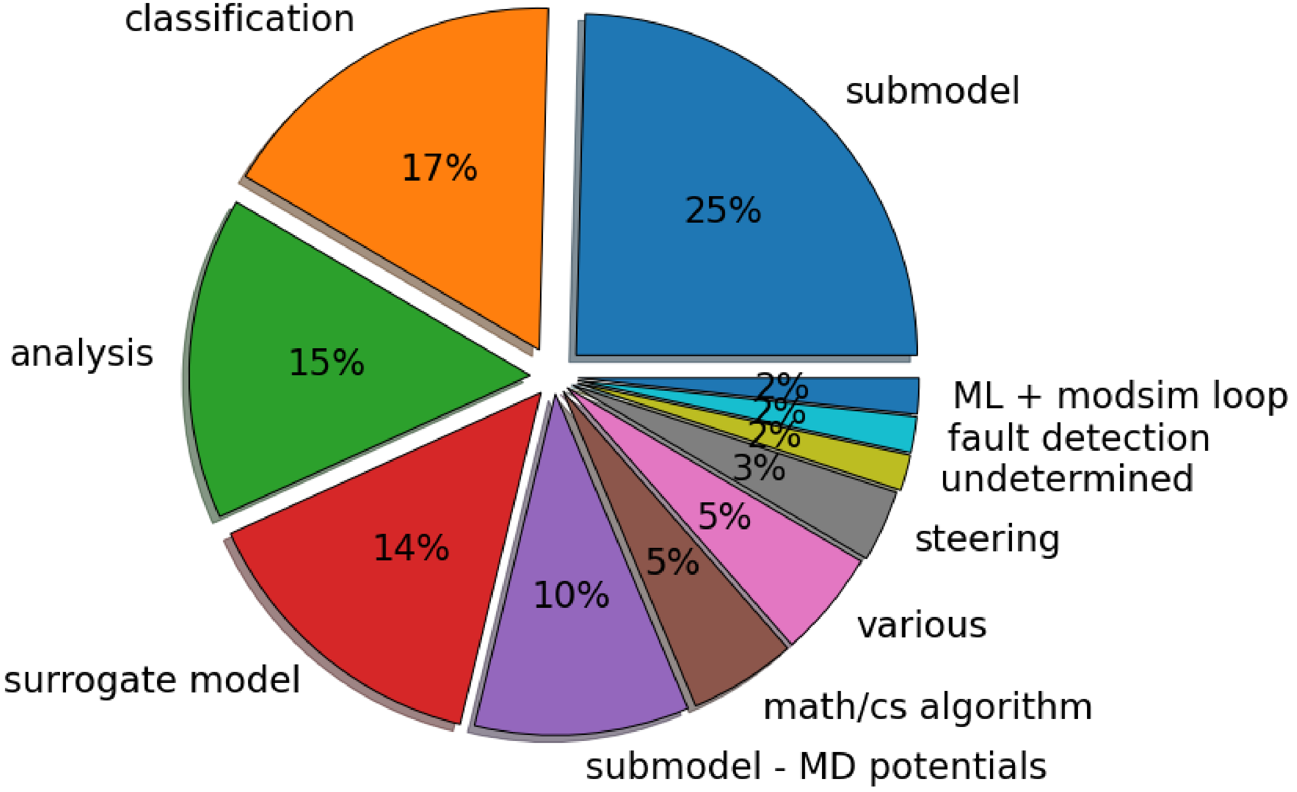
TECHNICAL REQUIREMENTS

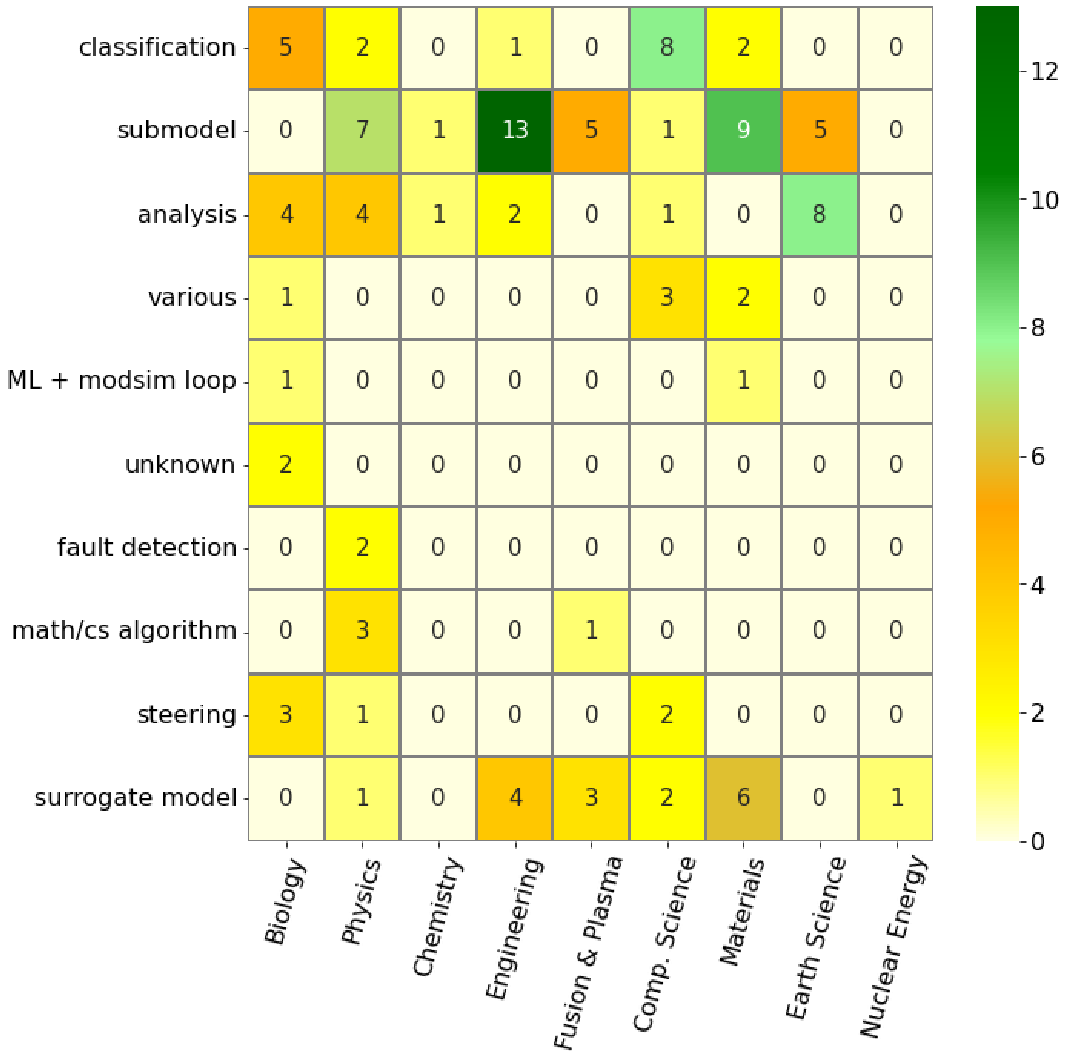
March 2024

Questions?

And a HUGE thanks to all our OLCF users, especially those whose work was highlighted here!

AI/ML are found throughout the OLCF workload today





	Biology	Physics	Chemistry	Engineering	Fusion & Plasma	Comp. Science	Materials	Earth Science	Nuclear Energy
classification	5	2	0	1	0	8	2	0	0
submodel	0	7	1	13	5	1	9	5	0
analysis	4	4	1	2	0	1	0	8	0
various	1	0	0	0	0	3	2	0	0
ML + modsim loop	1	0	0	0	0	0	1	0	0
unknown	2	0	0	0	0	0	0	0	0
fault detection	0	2	0	0	0	0	0	0	0
math/cs algorithm	0	3	0	0	1	0	0	0	0
steering	3	1	0	0	0	2	0	0	0
surrogate model	0	1	0	4	3	2	6	0	1

"Learning to Scale the Summit: AI for Science on a Leadership Supercomputer." Wayne Joubert, Bronson Messer, Philip C. Roth, Antigoni Georgiadou, Justin Lietz, Markus Eisenbach and Junqi Yin, Accepted, ExSAIS 2022: Workshop on Extreme Scaling of AI for Science (IPDPS 2022)