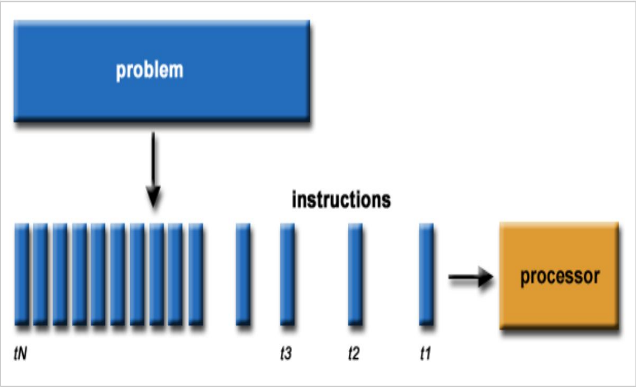
Why do HPC

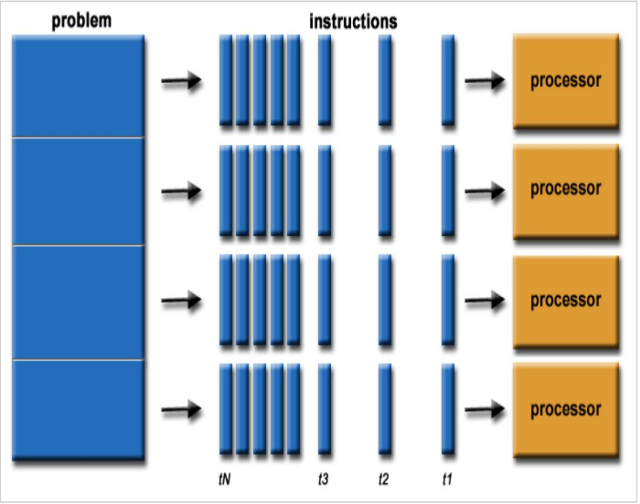
John Holmen

Michael Sandoval

What is High Performance Computing?

- High Performance Computing (HPC) is about solving the world’s largest engineering and science problems with supercomputers.
- That means doing work efficiently in parallel



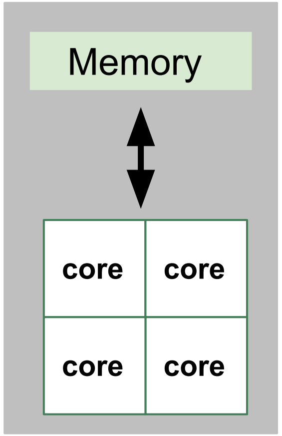


Oak Ridge National Laboratory

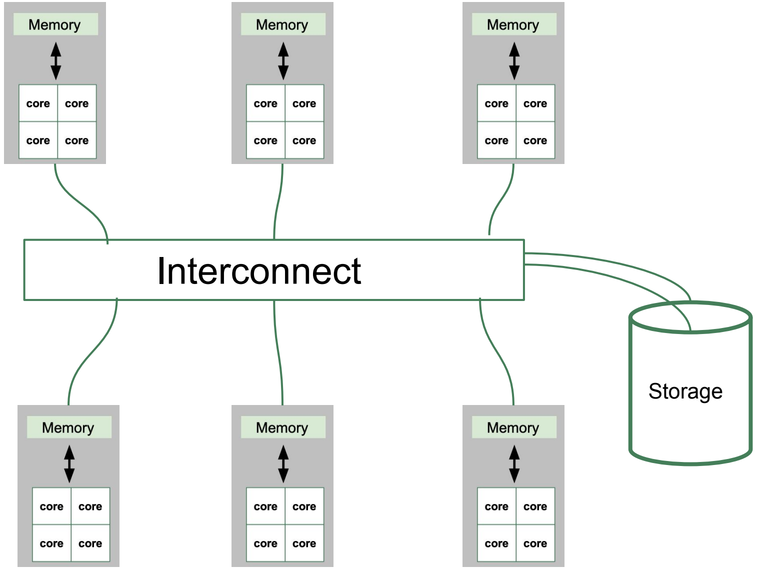
Images courtesy of : https://hpc.llnl.gov/training/tutorials/introduction-parallel-computing-tutoria
l

What is High Performance Computing?

- Supercomputers are made of compute servers, networks, and storage servers, that are all specially optimized to do operations in parallel.



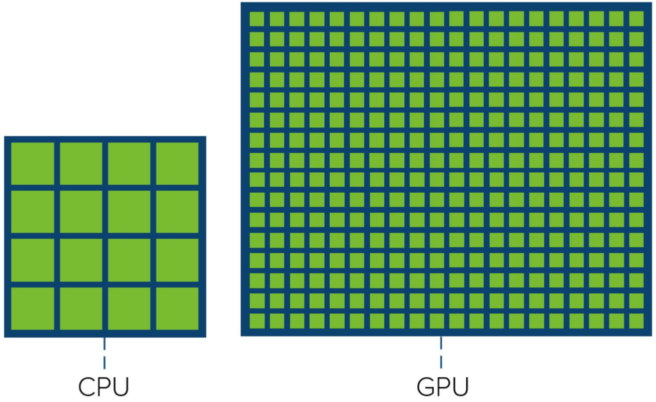
Node



Cluster

Processors

- Modern HPC nodes combine Central Processing Units CPUS with accelerators such as Graphical Processing Units GPUs.
  - CPUs generally have several compute cores that can run 10s of tasks at a time
  - GPUS have streaming multiprocessor that can run 1000s of tasks at a time.



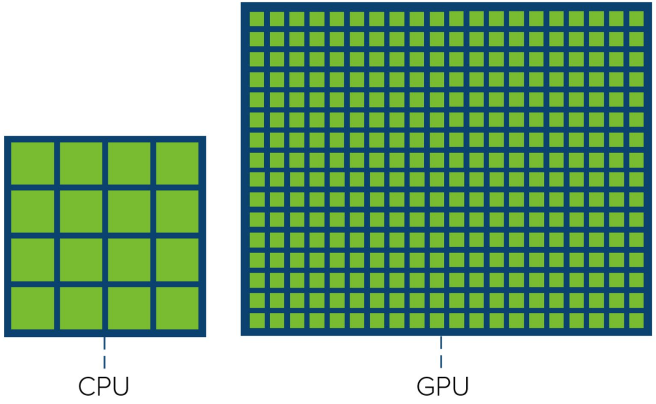
Images courtesy of : https://blogs.nvidia.com/blog/2009/12/16/whats-the-difference-between-a-cpu-and-a-gpu/

Processors

Much of an HPC programmers work is figuring out how to efficiently use all of the different levels of parallelism on a compute node and within a group of nodes.

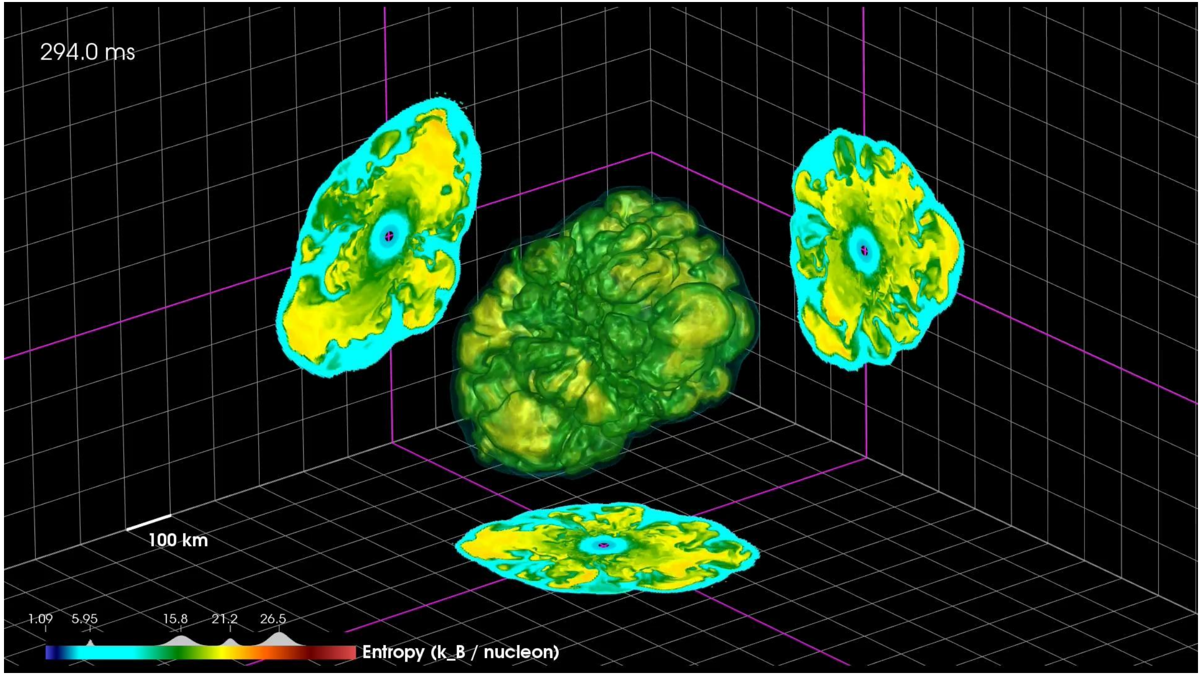
- Generally work is setup on the CPU and then moved to the GPU for processing.
- CPUs and GPUs have separate memory so the programmer must plan for the data movement between them.

We will cover several of the main tools and methods for utilizing the different levels of parallelism and their associated storage within an HPC cluster.



Why do HPC?

HPC allows simulations of nature to be built with enough detail that they can make accurate predictions of natural phenomena, so you can test things that are impossible to test physically or perhaps to expensive:



Movie Courtesy of Michael Sandoval

More HPC Applications / Examples

Reaching New Heights in Weather Forecasting's Exascale Future

ECMWF and ORNL researchers use the power of Summit to simulate the Earth's atmosphere for a full season at 1-square-kilometer grid-spacing

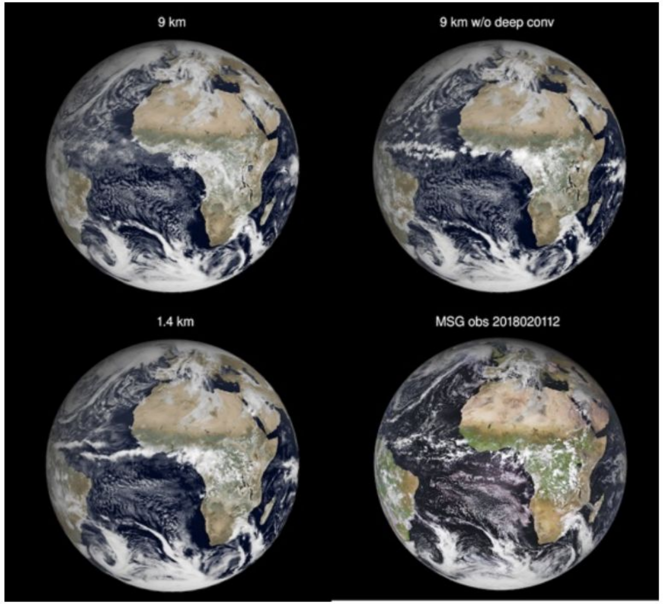
The Science

Using Summit, a team of researchers from ECMWF and ORNL achieved a computational first: a global simulation of the Earth's atmosphere at a 1-square-kilometer average grid-spacing for a full 4-month season. Completed in June, the milestone marks a big improvement in resolution for the "European Model," which currently operates at 9-kilometer grid-spacing for routine weather forecast operations. It also serves as the first step in an effort to create multi-season atmospheric simulations at high resolution, pointing toward the future of weather forecasting—one powered by exascale supercomputers.

The Impact

The team has made the simulation's data available to the international science community. By eliminating some of the fundamental modelling assumptions prevalent in conventional simulations, the high-resolution data may help to improve model simulations at coarser resolutions.

PI(s)/Facility Lead(s): Nils Wedi, ECMWF
ASCR Program/Facility: INCITE/OLCF
ASCR PM: Christine Chalk
Publication(s) related to this work: Wedi, N. P., et al. A baseline for global weather and climate simulations at 1 km resolution. *Journal of Advances in Modeling Earth Systems*, 12 (2020), e2020MS002192. doi: 10.1029/2020MS002192



These simulated satellite images of Earth show the improvement in resolution of the ECMWF Integrated Forecasting System from 9-kilometer grid-spacing with parametrized deep convection (top left), to 9-kilometer grid-spacing (top right), and 1-kilometer grid-spacing (bottom left). On the bottom right is a Meteosat Second Generation satellite image at the same verifying time. Image courtesy ECMWF.

Oak Ridge National Laboratory | Leadership Computing Facility   ECMWF   NVIDIA

GE Spins up Supercomputer Models to Zero in on Energy Loss in Turbines

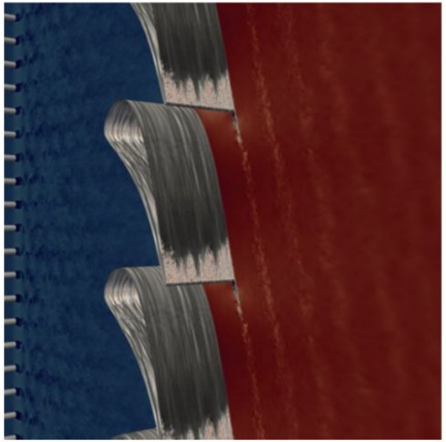
A team at GE Aviation and the University of Melbourne is studying turbulent flows on the Summit supercomputer for better engines

The Science

High-pressure turbines are vital components of gas turbines used to propel jet engines. The more efficient these jet engines are, the better they are for the aircraft industry and their customers. But these large, dynamic systems are difficult to study via experiments and physical testing. A team led by scientists at General Electric (GE) Aviation and the University of Melbourne used the Summit supercomputer to run for the first time real-engine cases capturing the largest eddies, or circular fluid movements, down to those that were tens of microns away from the turbine blade surface. From the simulations, the researchers determined which regions near a turbine blade experience a greater loss of energy. For the case with the highest Mach number, which describes the flow's velocity compared with the speed of sound, they discovered an extra loss of energy resulting from strong shock waves, or violent changes in pressure, that interact with the edge and wake of the flow to cause a massive amount of turbulence.

The Impact

More accurate prediction of real-engine conditions will lead to more efficient engines that consume less fuel and other positive derivative effects. A 1 percent reduction in fuel consumption across a fleet of engines is equal to about 1 billion dollars a year in fuel cost savings. Reduced fuel consumption also translates into reduced emissions—a 1 percent reduction in fuel burn reduces CO_2 emissions by roughly 1.5 percent.



A row of upstream bars produces highly turbulent flow that gets accelerated through a high-pressure turbine blade row and interacts with the blade surface, causing significant temperature variations. Image Credit: Richard Sandberg, University of Melbourne

PI(s)/Facility Lead(s): Richard Sandberg, Univ. Of Melbourne; Sriram Shankaran, GE Aviation
ASCR Program/Facility: INCITE/OLCF
ASCR PM: Christine Chalk
Publication(s) for this work: Y. Zhao and R. D. Sandberg, "High-Fidelity Simulations of a High-Pressure Turbine Vane," *Journal of Turbomachinery* 143, no. 9 (2021).

Y. Zhao and R. D. Sandberg, "Using a New Entropy Loss Analysis," *Journal of Turbomachinery* 142, no. 8 (2020): 081008

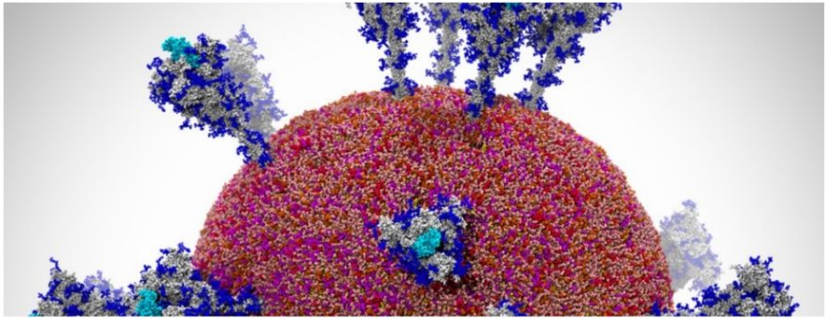
OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY   THE UNIVERSITY OF MELBOURNE   GE Aviation   U.S. DEPARTMENT OF ENERGY INCITE LEADERSHIP COMPUTING

Gordon Bell Special Prize-Winning Team Reveals AI Workflow for Molecular Systems in the Era of COVID-19 (PI: Rommie Amaro, UCSD)

Research by team at ANL and UC San Diego leads to a novel understanding of SARS-CoV-2 and a new method for studying disease

The Science

Imaging techniques, such as X-ray imaging and cryogenic electron microscopy, can provide snapshots of viruses such as SARS-CoV-2, but these fall short of capturing the dynamic movements of viral proteins. Computer simulations can help scientists capture the movements of these structures virtually. Now, a team led by Rommie Amaro at the University of California San Diego and Arvind Ramanathan at ANL have built a first-of-its-kind workflow based on AI and have run it on OLCF's Summit supercomputer to simulate the virus's spike protein in numerous environments, including within the SARS-CoV-2 viral envelope comprising 305 million atoms—the most comprehensive simulation of the virus performed to date. The accomplishment has earned the ACM Gordon Bell Special Prize for HPC-Based COVID-19 Research.



A snapshot of a visualization of the SARS-CoV-2 viral envelope comprising 305 million atoms. Image Credit: Rommie Amaro, UC San Diego; Arvind Ramanathan, ANL

The Impact

The team was able to successfully scale NAMD to 24,576 of Summit's NVIDIA V100 GPUs. The results of the team's initial runs on Summit have led to discoveries of one of the mechanisms that the virus uses to evade detection as well as a characterization of interactions between the spike protein and the protein that the virus takes advantage of in human cells to gain entry—the ACE2 receptor. The team is now integrating their scientific code, NAMD, into their workflow pipeline to fully automate the transition from simulation to AI for data processing without gaps.

PI(s)/Facility Lead(s): Rommie Amaro
ASCR Program/Facility: INCITE / OLCF
ASCR PM: Christine Chalk
Publication(s) for this work: Lorenzo Casalino, Abigail Dommer, Zied Gaieb, Emilia P. Barros, Terra Sztain, Surl-Hee Ahn, Anda Trifan, Alexander Brace, Anthony Bogetti, Heng Ma, Hyungro Lee, Matteo Turilli, Syma Khalid, Lillian Chong, Carlos Simmerling, David J. Hardy, Julio D. C. Maia, James C. Phillips, Thorsten Kurth, Abraham Stern, Lei Huang, John McCalpin, Mahidhar Tatineni, Tom Gibbs, John E. Stone, Shantenu Jha, Arvind Ramanathan, and Rommie E. Amaro. "AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics." *International Journal of High-Performance Computing Applications, SC20*(2020).

What Background do I need to do HPC?

You need to be able to teach yourself new techniques and ideas and then be persistent in using them.

What Background do I need to do HPC?

There are many paths to HPC:

- Domain scientist, mathematician, engineer, or a computer scientist depending on what you want to do.
- Academic Careers:
  - Generally, you get an advanced degree and use HPC for your thesis or post-education work.
- Technical Careers:
  - Get a degree that focuses on the programming environment, languages, systems engineering, hardware or algorithms used in computing.

What Background do I need to do HPC?

HPC programers today have:

- Experience using compiled languages like C, C++, Fortran
- Experience with interpreted languages like python
- Experience with advanced maths such as differential equations, Linear Algebra and statistics
- Experience with MPI, OpenMP, OpenACC, Cuda and other tools used for parallel and GPU computing
- Experience with ML/AI
- Quantum computing is on the horizon

You need to be able to teach yourself new techniques and ideas and then be persistent in using them.

What do OLCF Computational Scientists and Engineers do?

HPC Career; NCCS Organization example:

- Advanced Technologies
  - Analytics and AI at Scale, Data Lifecycle and Scalable Workflows, Technology Integration
- Operations
  - Platforms, Software Services Development, System Acceptance and User Environment, User Access, Outreach, Communication, User Assistance
- Science Engagement
  - Chemistry and Materials, Life Sciences and Engineering, Nuclear Particles and Astrophysics, Algorithms and Performance Analysis
- HPC Systems
  - Clusters, Cybersecurity and Information Engineering, Infrastructure and Networking, Infrastructure Operations, Scalable Systems, Storage and Archive
- Advanced Computing Methods and Quantum Information Science
  - research and development related to data systems, data analytics, geospatial sciences, modeling and simulation, discrete computing, quantum sciences, and cyber security