OAK RIDGE National Laboratory

A Decade of Design To Reach Exascale for ModSim

Al Geist
Leadership Computing Facility
Oak Ridge National Laboratory

ModSim'22 Workshop
Seattle WA
August 10-12, 2022

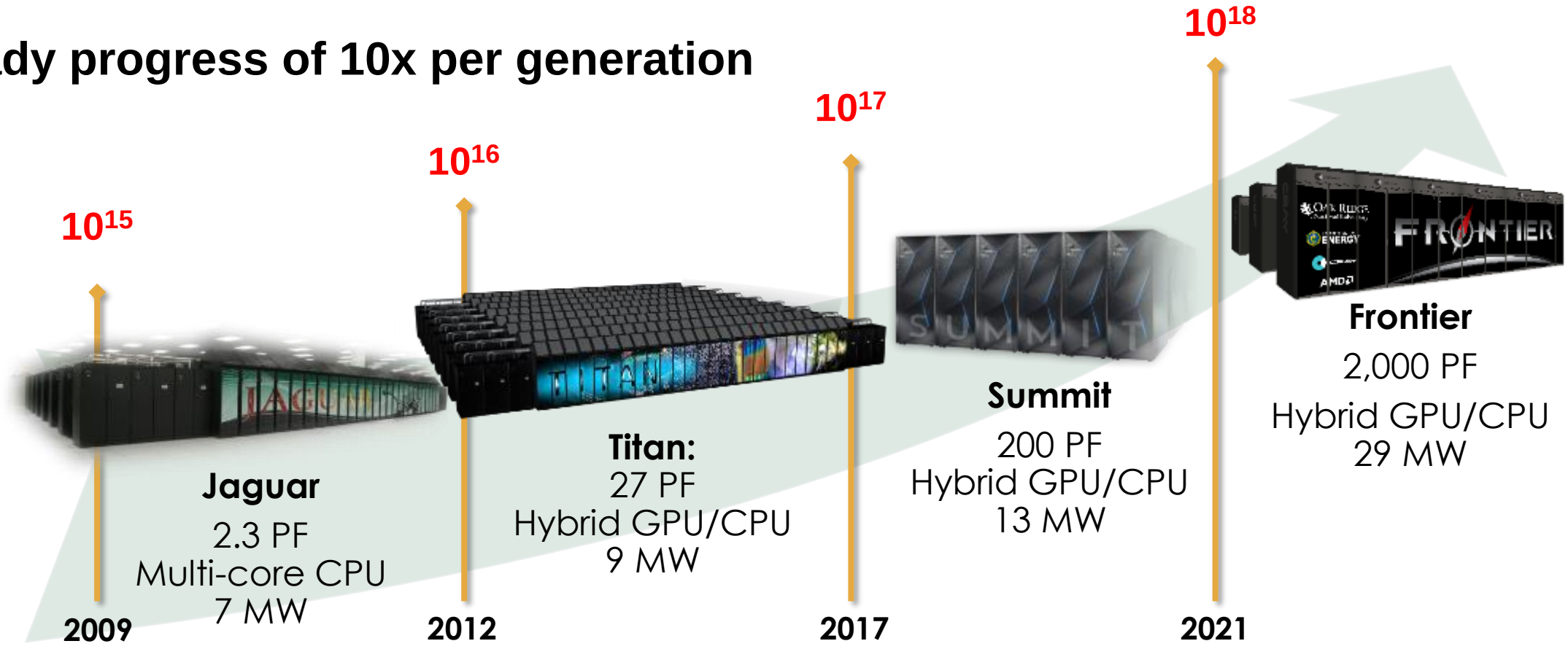
ORNL is managed by UT-Battelle, LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

Oak Ridge National Laboratory's Journey from Petascale to Exascale

Mission: Providing world-class computational resources and specialized services for the most computationally intensive global challenges	Vision: Deliver transforming discoveries in energy technologies, materials, biology, environment, health, etc.

Steady progress of 10x per generation



Four Key Challenges to Reach Exascale

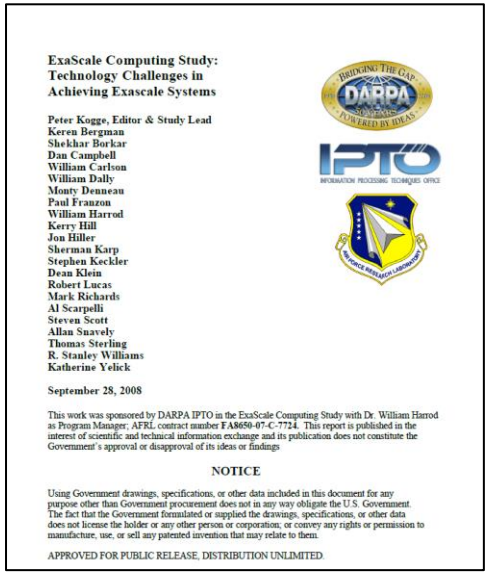
These four challenges also existed for Petascale, but were not considered show-stoppers at Petascale
In 2009 there was serious concern that Exascale Systems may not be possible

Energy Consumption: Studies in 2008 predicted that a 1 Exaflop system would consume between 150-500 MW. Vendors were given the ambitious goal of trying to get this down to 20 MW.

Reliability: Exascale computers will need to dynamically adapt to a constant stream of transient and permanent faults. Failures may happen faster than you can checkpoint a job.

Parallelism: Exascale computers will have billion-way parallelism Are there more than a handful of applications that could utilize this?

Data Movement: Memory wall continues to grow higher - Moving data from the memory into the processors and out to storage is the main bottleneck to performance.



ExaScale Computing Study:
Technology Challenges in
Achieving Exascale Systems

Peter Kogge, Editor & Study Lead
Keren Bergman
Shekhar Borkar
Dan Campbell
William Carlson
William Dally
Monty Denneau
Paul Franzon
William Harrod
Kerry Hill
Jon Hiller
Sherman Karp
Stephen Keckler
Dean Klein
Robert Lucas
Mark Richards
Al Scarpelli
Steven Scott
Allan Snavely
Thomas Sterling
R. Stanley Williams
Katherine Yelick

September 28, 2008

This work was sponsored by DARPA IPTO in the ExaScale Computing Study with Dr. William Harrod as Program Manager, AFRL contract number FA8650-07-C-7724. This report is published in the interest of scientific and technical information exchange and its publication does not constitute the Government's approval or disapproval of its ideas or findings

NOTICE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

APPROVED FOR PUBLIC RELEASE, DISTRIBUTION UNLIMITED

2008 DARPA report
Peter Kogge et al

How did ModSim help overcome these challenges for Frontier to reach an Exascale?

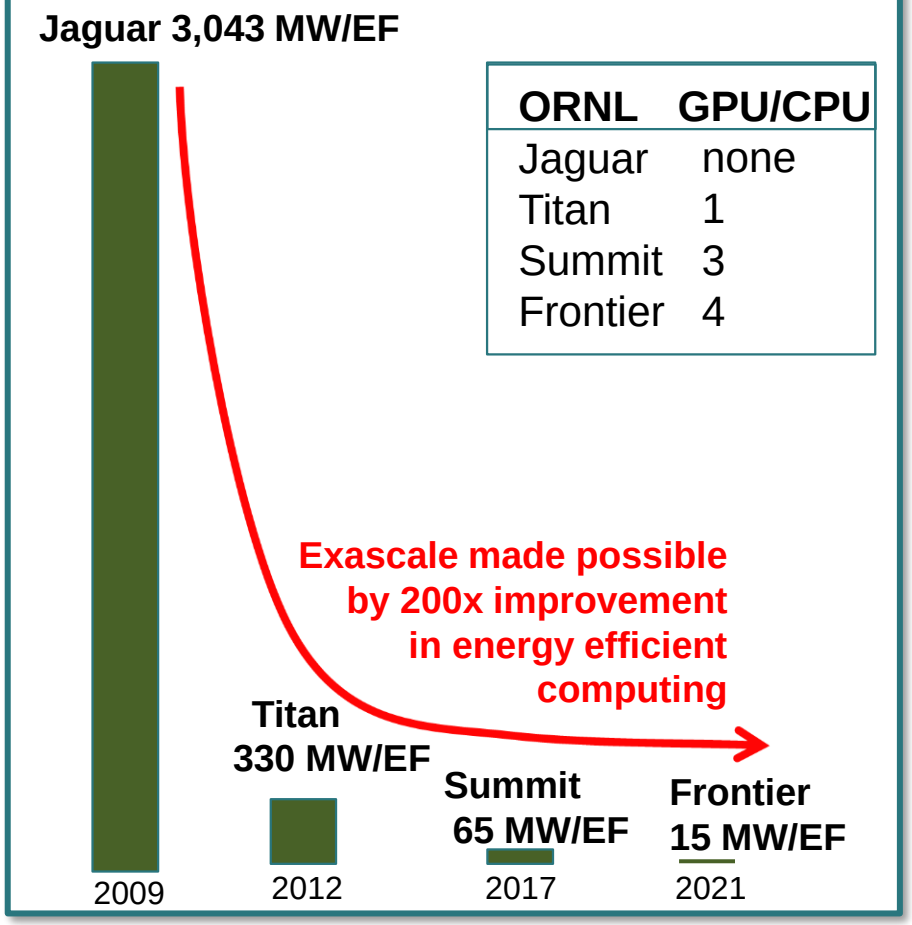
Energy Efficient Computing – Frontier achieves 14.5 MW per EF

The most serious Challenge to reaching Exascale has been energy consumption

- **ORNL pioneered GPU use in supercomputing** beginning in 2012 with Titan thru today with Frontier. Significant part of energy efficiency improvements.
- **ASCR [Fast, Design, Path] Forward vendor investments** in energy efficiency (2012-2020) further reduced the power consumption of computing chips (CPUs and GPUs)..
- **200x reduction in energy per FLOPS** from Jaguar to Frontier at ORNL
- ORNL achieves additional energy savings from using warm water cooling in Frontier (32 C).
  ORNL Data Center PUE= 1.03

Frontier first US Exascale computer

Multiple GPU per CPU drove energy efficiency



Reliability – What's important - Mean Time to Application Failure

The Normal Tactic is Avoidance Detection Containment Recovery

Typical supercomputer stays up about a week, but there are many faults during this time that the system must dynamically handle without system interrupt.

Reliability methods often conflict w/ Energy

- Reducing voltage margins saves energy but adversely affects component reliability (FIT rate)
- Silent error rates increasing with memory size (cosmic rays changing bits in DRAM)
- Redundancy consumes energy and costs $$

Applications need faster checkpoint or resilience aware algorithms

Resilience – ECC and solder have been common causes for undetected errors that cause wrong answers across big systems

- Jaguar's memory experienced 8,000 bit flips every hour from cosmic rays-- that is random data changes at rate of 350/min!
  - Single-bit errors corrected by ECC
  - Double-bit uncorrectable error averaged 1 every 24 hours
- Titan had a problem with having too much gold in the solder. Thermal stress from chips heating and cooling cracked the solder, changing resistance and corrupting data

Frontier has 4 TB NVM on node for fast checkpoint. Frontier Target for MTBAF is 6 hours.

Billion-way Parallelism

As supercomputers got larger and larger, we expected them to be more specialized and limited to just a small number of applications that can exploit their growing scale

Solution: On our journey to Exascale, we found the Summit architecture with few, large-memory, multi-GPU nodes could excel at broad range of applications including modeling and simulation, data analytics, and artificial intelligence

- Adapting to this architecture has been a journey for the ModSim community
- The GPUs hide between 1,000 and 10,000 way concurrency inside their pipelines so the users don't have to think about as much parallelism. (But it is not trivial)
- Nvidia "Tensor cores" opened the door to incorporate ML and AI into ModSim applications
- Frontier Users and system software only have to deal with 9,400 nodes not a million

Frontier Exascale computer uses and improves on Summit's successful architecture
- 5 TB of on-node memory, 4 GPU per node, Peak of >10 ExaOps (FP16)

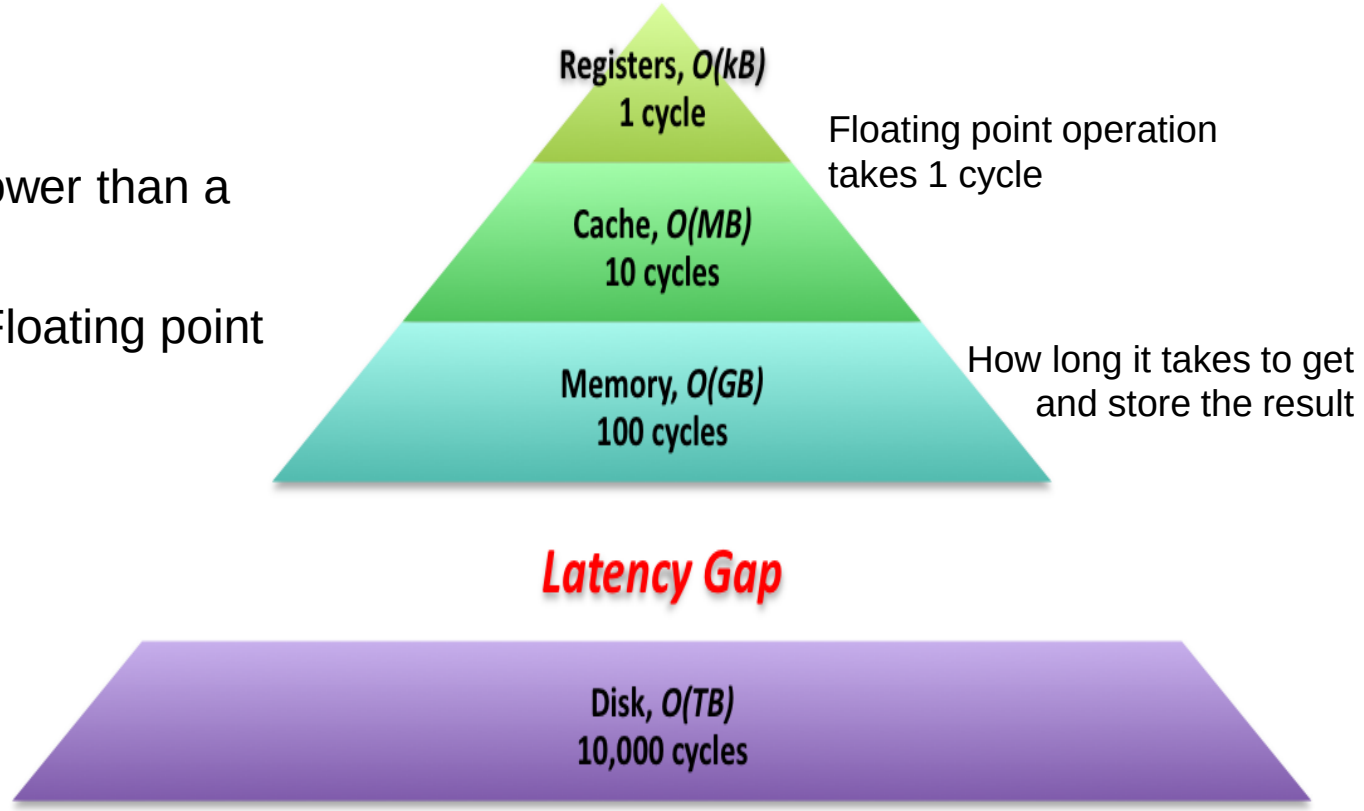
Data Movement Dominates Time and Energy

Memory Wall Challenges

- Time to move a byte is orders of magnitude slower than a Floating point operation
- Energy to move a byte is much higher than a Floating point operation

Needs for Exascale

- Data Locality
- Communication avoiding algorithms
- Explicitly managed memory hierarchies
- Faster memory



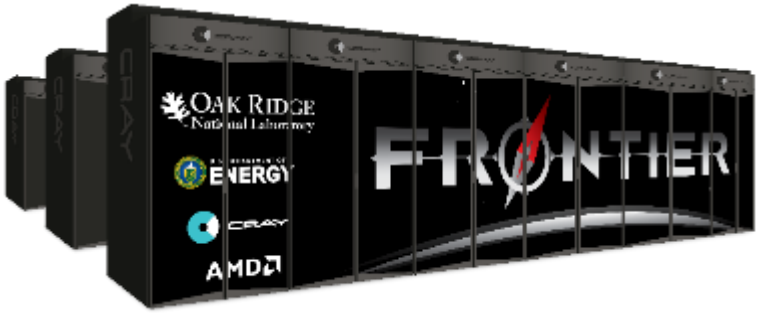
Frontier having High Bandwidth memory soldered onto the GPU increases BW an order of magnitude and GPUs are well suited for latency hiding. Frontier kicks the can down the road.

In the end, Exascale did not Require Exotic Technology, Architecture, or new Programming Paradigms. It was Incremental Steps - not a Giant Leap.

System	Jaguar (2009)	Titan (2012)	Summit (2017)	Frontier (2021)
Peak	2.3 PF	27 PF	200 PF	2,000 PF
# nodes	18,688	18,688	4,608	9,408
Node	1 AMD CPU	1 AMD Opteron CPU 1 NVIDIA Kepler GPU	2 IBM POWER9™ CPUs 6 NVIDIA Volta GPUs	1 AMD Trento CPU 4 AMD MI250X GPUs
Memory	0.3 PB DDR2	0.6 PB DDR3 + 0.1 PB GDDR	2.4 PB DDR4 + 0.4 HBM + 7.4 PB On-node storage	4.6 PB DDR4 + 4.6 PB HBM2e + 36 PB On-node storage with 66 TB/s Read 62 TB/s Write
On-node interconnect	NA	PCI Gen2 No coherence across the node	NVIDIA NVLINK Coherent memory across the node	AMD Infinity Fabric Coherent memory across the node
System Interconnect	Cray SeaStar 2.0 GB/s	Cray Gemini network 6.4 GB/s	Mellanox Dual-port EDR IB 25 GB/s	Four-port Slingshot network 100 GB/s
Topology	3D Torus	3D Torus	Non-blocking Fat Tree	Dragonfly
Storage	15 PB, 0.2 TB/s Lustre	32 PB, 1 TB/s, Lustre	250 PB, 2.5 TB/s, IBM Spectrum Scale with GPFS	695 PB HDD+11 PB Flash Performance Tier, 9.4 TB/s and 10 PB Metadata Flash. Lustre
Power	7 MW	9 MW	13 MW	29 MW

Frontier System

HPE Cray EX 235a -- Also used in LUMI and Adastra Systems



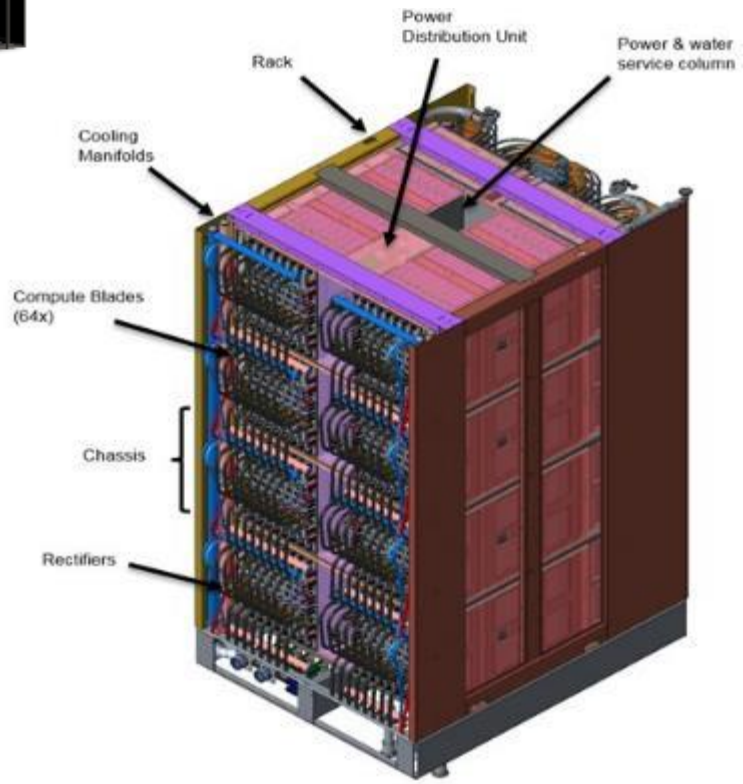
System

- 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

Frontier TDS

Olympus rack

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



Energy Efficient Node

AMD extraordinary engineering

- 1 AMD "Trento" CPU (optimized Milan)
- 4 AMD MI250X GPUs

GPU has extensive power management that can rapidly turn off unused resources and vary power across the GPU at a very fine level.
CPU has the same ability

- 512 GiB DDR4 memory on CPU
- 512 GiB HBM2e total per node
- 4 Cassini NICs connected to the 4 GPUs

Compute blade

- 2 AMD nodes

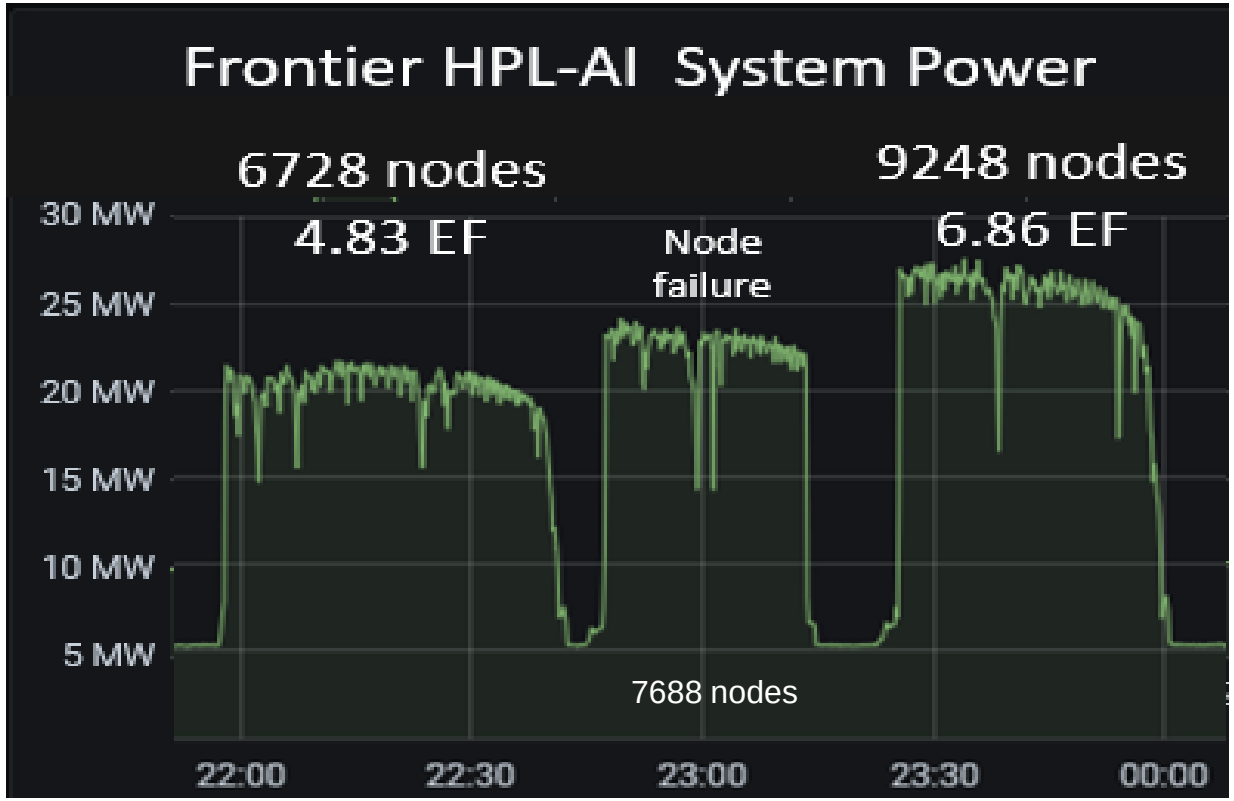
All water cooled, even DIMMS and NICs

Future ModSim: Understand how power fluctuations and harmonics affect resilience and reliability of the system components

Summary of Frontier's #1 HPL-AI run

- 6.86 Exaflop/s
- 9248 nodes
- ~25 MW av. Power
- Run time 38 min.

When the job launched, the data center cooling and power infrastructure had to handle a 20 MW surge in under 5 seconds.



ModSim needs to start considering affects of huge power swings and power oscillations

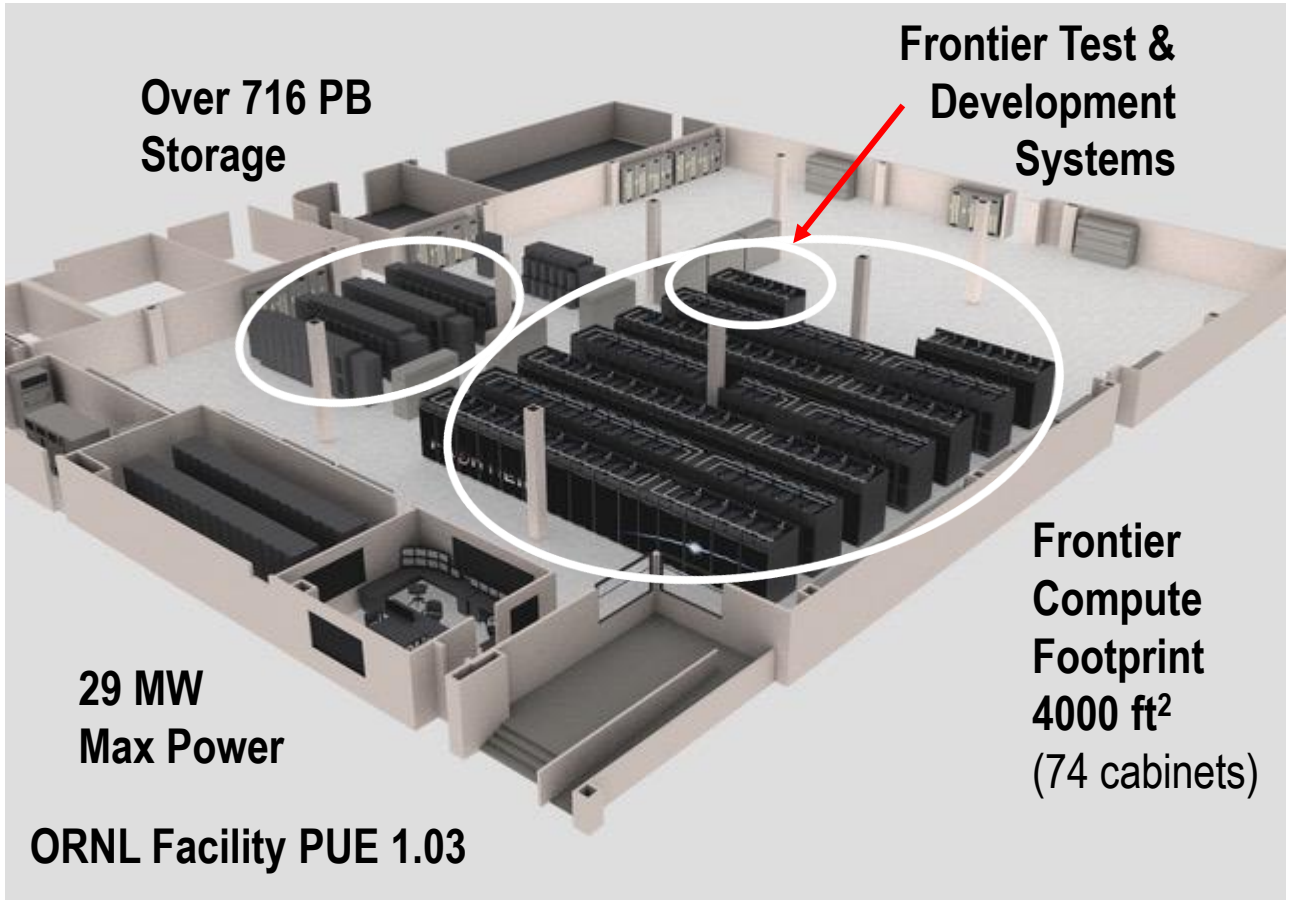
Fronter TDS Ranked #1 Green500 62 Gflops/Watt (Nov.'21 #1 was 39 GF/W)
Frontier Ranked #2 Green500 52 Gflops/Watt

Frontier Test & Development systems(TDS) arrived early to allow staff to prepare codes while Frontier was being delivered and stabilized. Ran HPL on TDS first.

May 2, 2022 HPL run on 128 node TDS system achieved 19.2 PF using av. 309 KW
#29 TOP500 list
#1 Green500 achieving 62 Gflop/W

May 16, 2022 HPL-AI was run on 9,248 nodes of Frontier achieved 6.86 Exaflops using av. 25 MW

May 27, 2022 HPL run on 9,248 nodes of Frontier achieved 1.102 EF using av. 21.1 MW
#1 TOP500 list
#2 Green500 achieving over 52 Gflop/W



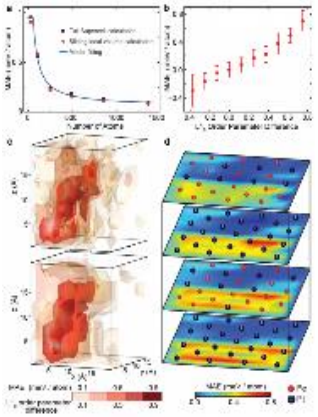
Why is TDS GF/W higher? HPL gets higher efficiency at 128 nodes verses 9,248 nodes

Initial CAAR Early Science Results on Crusher ~10x Summit

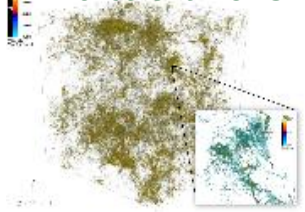
ModSim challenge: Diverse range of algorithms and we get new proposals every year

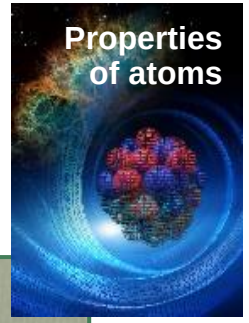
Science Area	CAAR App	Recent Results on Crusher
Advanced materials	**LSMS**	MI250 getting up to 10x speedup over Summit V100
Turbulent Flows	**GESTS**	Crusher MI250 achieves 12x speedup over Summit V100
Porus Media	**LBPM**	Crusher MI250 slightly faster than Summit V100.
Plasma Physics	**PIConGPU**	Seeing 2.5x – 5x speedup over Summit
Atomic nucleus	**NuCCOR**	Crusher MI250 performance gains of up to 8x over Summit V100
Bioinformatics	**CoMet**	Has been run on Frontier up to 8,000 nodes
Astrophysics	**Cholla**	Total of 15x speedup = Crusher HW getting additional 3x over Summit + 5x from SW

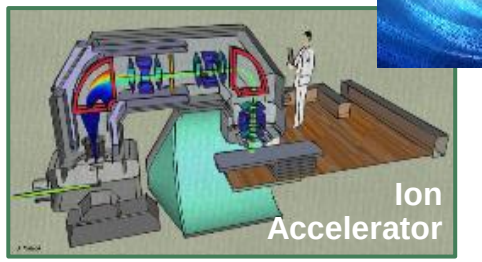
Magnetic anisotropy in FePt nanoparticle



Turbulent Flows







Astrophysics hydrodynamics

It is an exciting decade ahead for ModSim
Questions?