Optimizing Communications and I/O on Aurora for Application Performance

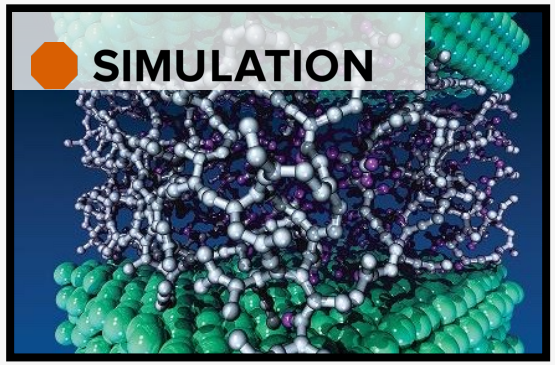
IXPUG ISC 2024 Workshop
May 16, 2024

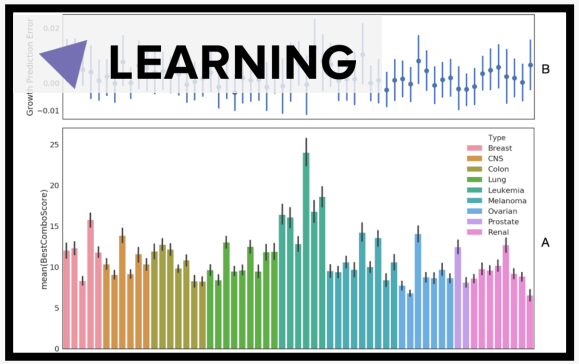
Kevin Harms and Kalyan Kumaran
Argonne Leadership Computing Facility

Argonne Leadership Computing Facility

The Argonne Leadership Computing Facility provides world-class computing resources to the scientific community.

- Users pursue scientific challenges
- In-house experts to help maximize results
- Resources fully dedicated to open science

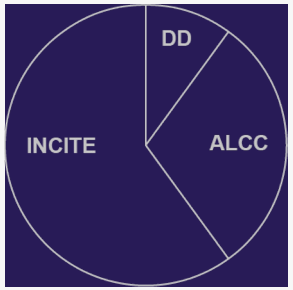




Architecture supports three types of computing

- Large-scale Simulation (PDEs, traditional HPC)
- Data Intensive Applications (scalable science pipelines)
- Deep Learning and Emerging Science AI (training and inferencing)

ALCF offers different pipelines for different applications



#2 on the Top500 @ 1.01 EF
#1 on the IO500

Intel GPU
Intel® Data Center GPU Max Series

Intel Xeon Processor
4th Gen Intel XEON Max Series CPU with High Bandwidth Memory

Platform
HPE Cray-Ex

Racks - 166
Nodes - 10,624
CPUs - 21,248
GPUs – 63,744

Interconnect
HPE Slingshot 11
Dragonfly topology with adaptive routing
Network Switch:
25.6 Tb/s per switch (64 200 Gb/s ports)
Links with 25 GB/s per direction

Platform
HPE Cray-EX

Peak FP Performance
≧ 2 Exaflops DP

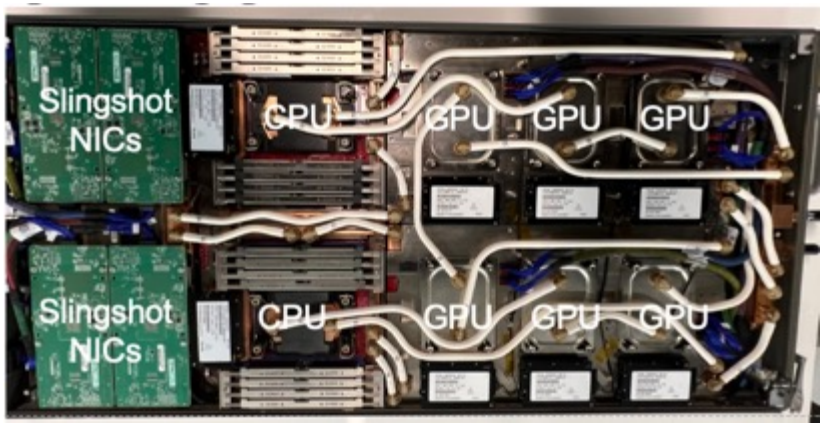
Memory
10.9PB of DDR @ 5.95 PB/s
1.36PB of CPU HBM @ 30.5 PB/s
8.16PB of GPU HBM @ 208.9 PB/s

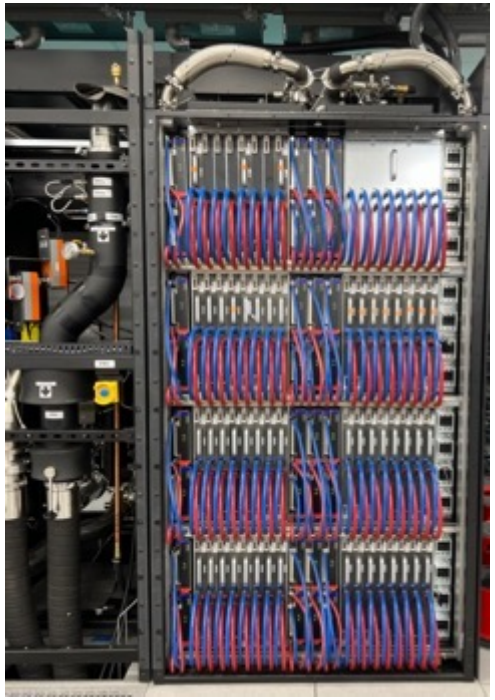
Network
2.12 PB/s Peak Injection BW
0.69 PB/s Peak Bisection BW

Storage
230PB DAOS Capacity
31 TB/s DAOS Bandwidth

Current Status of Aurora

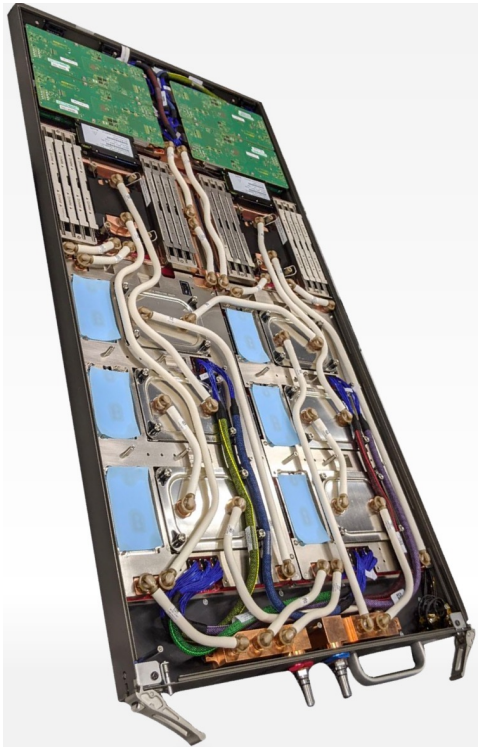
- All of the Aurora hardware is installed at ANL
- Aurora deployment and testing is underway
- Early users have access to the system and are running applications at limited scale

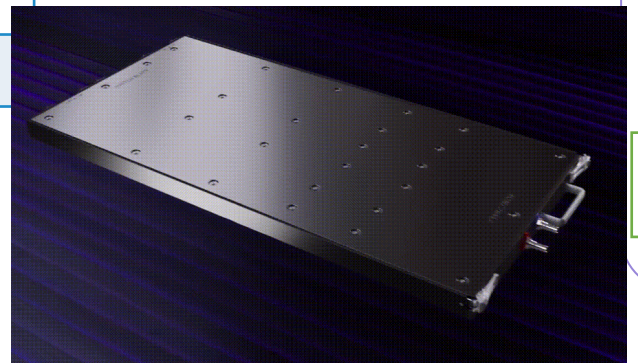


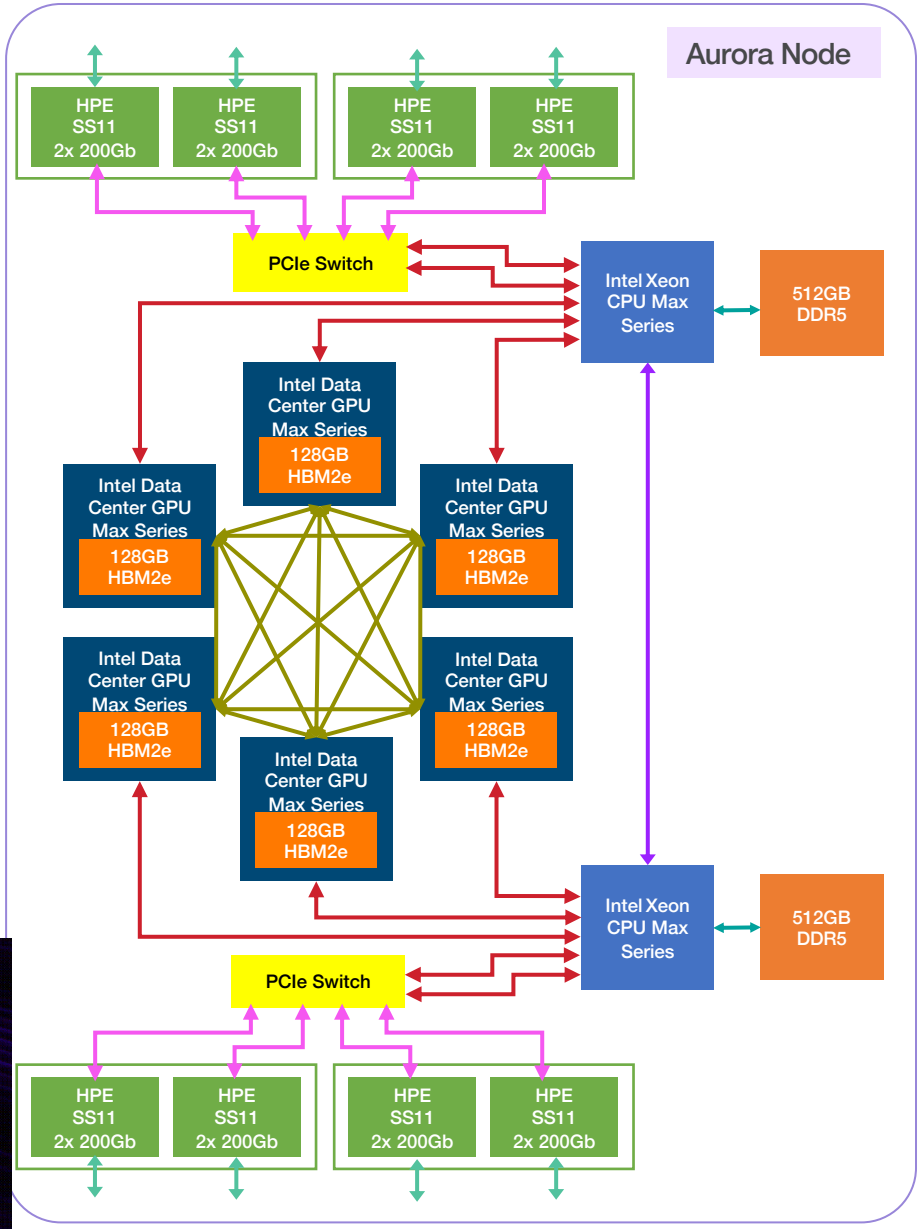


Aurora Exascale Compute Blade

NODE CHARACTERISTICS
6 GPUs - Intel Data Center GPU Max Series
2 CPUs - Intel Xeon CPU Max Series
768 GB GPU HBM Memory
19.66 TB/s Peak GPU HBM BW
128 GB CPU HBM Memory
2.87 TB/s Peak CPU HBM BW
1024 GB CPU DDR5 Memory
0.56 TB/s Peak CPU DDR5 BW
≧ 130 TF Peak Node DP FLOPS
200 GB/s Max Fabric Injection
8 NICs - HPE Slingshot 200G

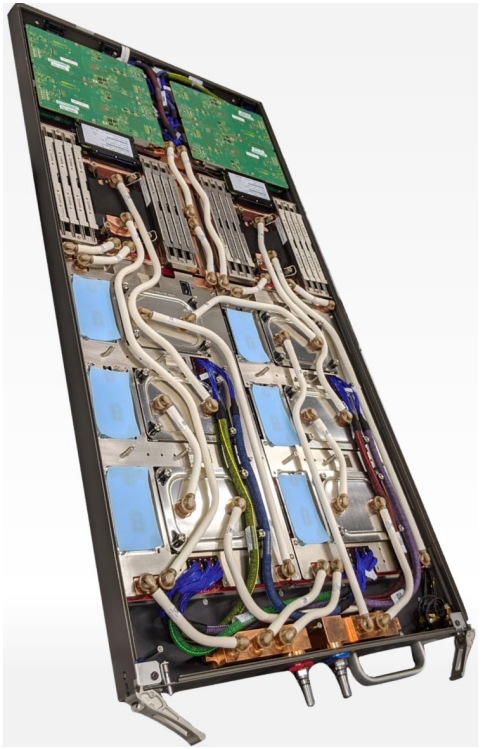


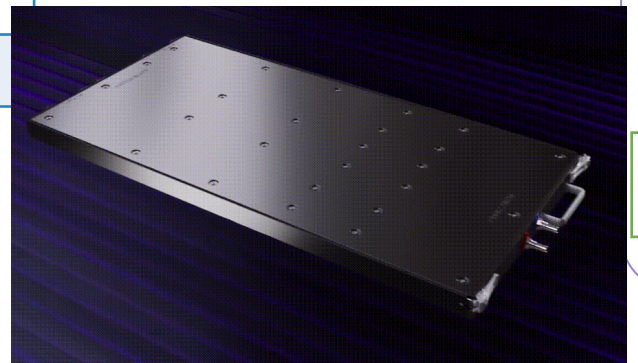


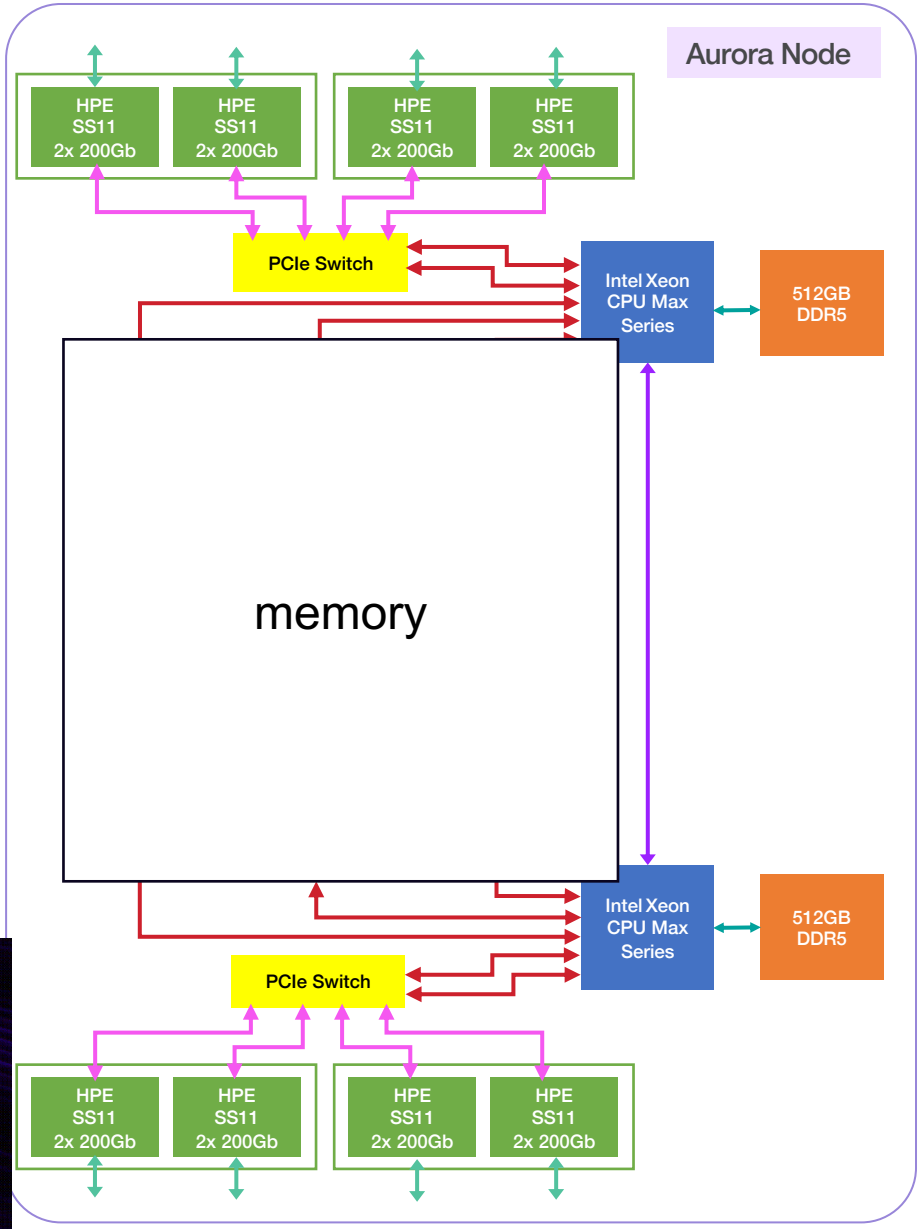


Aurora Exascale Compute Blade

NODE CHARACTERISTICS
6 GPUs - Intel Data Center GPU Max Series
2 CPUs - Intel Xeon CPU Max Series
768 GB GPU HBM Memory
19.66 TB/s Peak GPU HBM BW
128 GB CPU HBM Memory
2.87 TB/s Peak CPU HBM BW
1024 GB CPU DDR5 Memory
0.56 TB/s Peak CPU DDR5 BW
≧ 130 TF Peak Node DP FLOPS
200 GB/s Max Fabric Injection
8 NICs - HPE Slingshot 200G

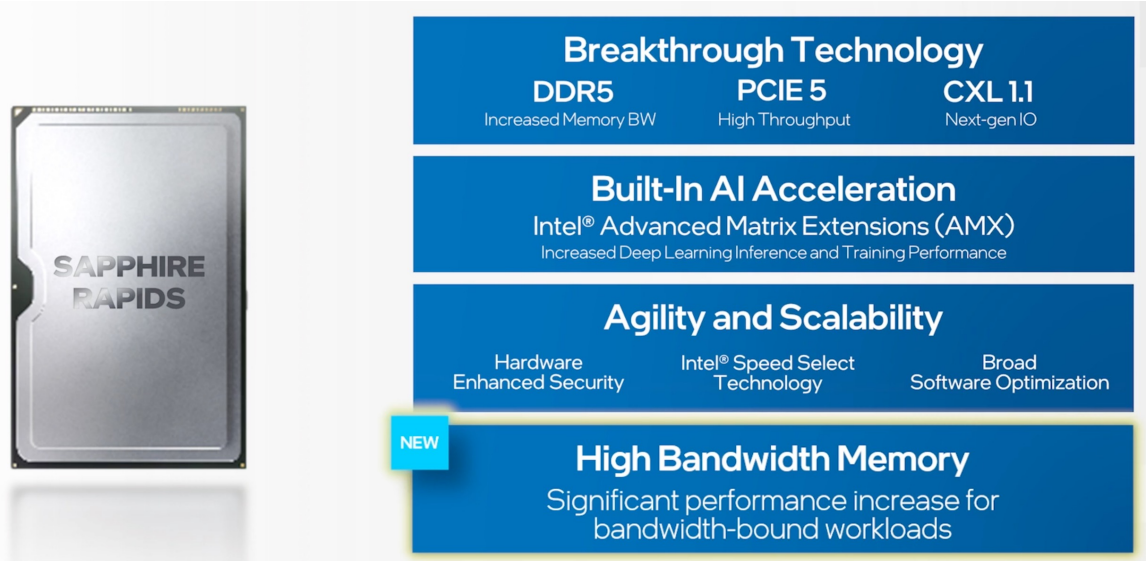






4th Gen Intel® Xeon Max Series CPU with HBM (Sapphire Rapids)

XEON DESCRIPTION	
Cores	52
Vector Extension	AVX-512
Threads (#)	2
Total HBM Memory (GB)	64
Peak HBM Memory BW (TB/s)	1.43
Total DDR5 4400 Memory (GB)	512
Peak DDR5 4400 Memory BW (TB/s)	0.28



HPE Slingshot Interconnect

Consistent, Repeatable Application Performance

- Advanced congestion control
- Fine grained adaptive routing
- Very low average and tail latency

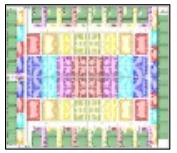
Extremely Scalable RDMA Performance

- Connectionless protocol
- Fine grained flow control
- MPI HW tag matching & progress engine
- Dragonfly topology – 3 switch hops (typical)

Native Ethernet

- Native IP – no encapsulation
- High-scale bandwidth integration to campus

HPE Slingshot Switches - 64 ports @ 200 Gbps



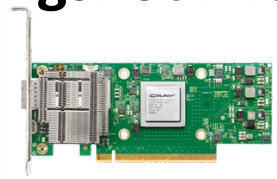
HPE Switch ASIC

Rack switches

100% DLC Switches

HPE Slingshot NICs - 200 Gbps

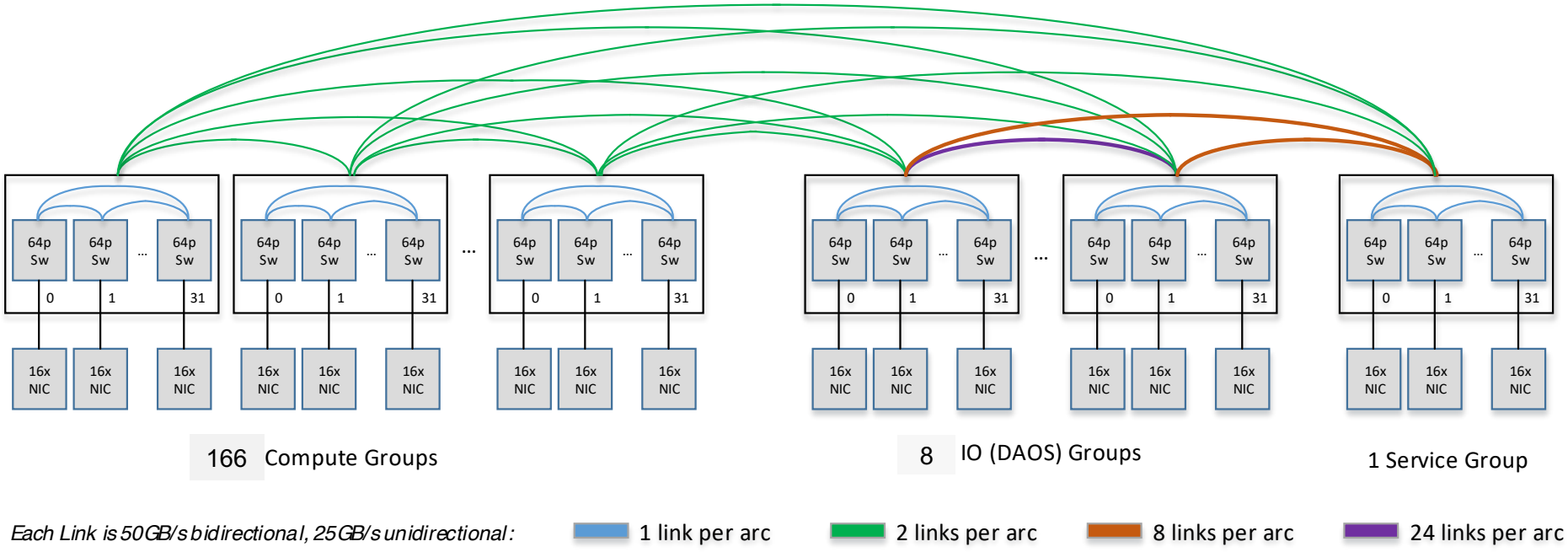
HPE NIC ASIC



PCIe Adapters

100% DLC NIC Mezz

Fabric



- 1-D Dragonfly Topology - 175 total groups (166 compute + 8 IO + 1 Service),
- All the global links are optical, all the local links in compute groups are electrical
- 2 global links between any two compute groups
- 24 links between any two IO groups, 8 links between the Service group and each IO group

MPI Benchmark Results

- Bisection
  - —9759 nodes
  - —8 ranks per node or 1 rank per NIC
  - —78,072 Slingshot endpoints
  - —1 MiB message size
  - —Aggregate BW - 357036.51 GB/sec
- MPI_All2all
  - —9758 nodes
  - —8 ranks per node
  - —64K message size
  - —Aggregate BW - 141683.50 GB/sec
- OSU Pairs
  - —9758 nodes
  - —8 ranks per node
  - —4K message size
  - —Aggregate messages/s – 59,826,805,718.19

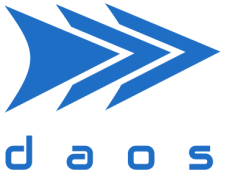
MPI node configuration

- Aurora has 8 NICs per node and 2 CPU sockets per node
  - MPICH selects NIC to use based on affinity and NUMA distance
  - Distribute ranks between both sockets to use all NICs
  - Set CPU binding to get best performance

- Libfabric and MPICH have many tuneables
  - These are often set to conservative values to reduce resource usage in default modes
  - Setting values higher may improve performance for some/many workloads
  - https://docs.alcf.anl.gov/aurora/known-issues/
  - `export FI_CXI_DEFAULT_CQ_SIZE=131072`
  - `export FI_CXI_OVFLOW_BUF_SIZE=8388608`
  - `export FI_CXI_CQ_FILL_PERCENT=50`

- Multi-NIC striping
  - When using few ranks than NICs, multi-NIC striping breaks messages and sends data across multiple NICs from a single process
  - https://www.mpich.org/static/docs/slides/2021-sc-bof/Intel.pdf
  - `export MPIR_CVAR_CH4_OFI_ENABLE_MULTI_NIC_STRIPING=1` (or 0 to turn off)

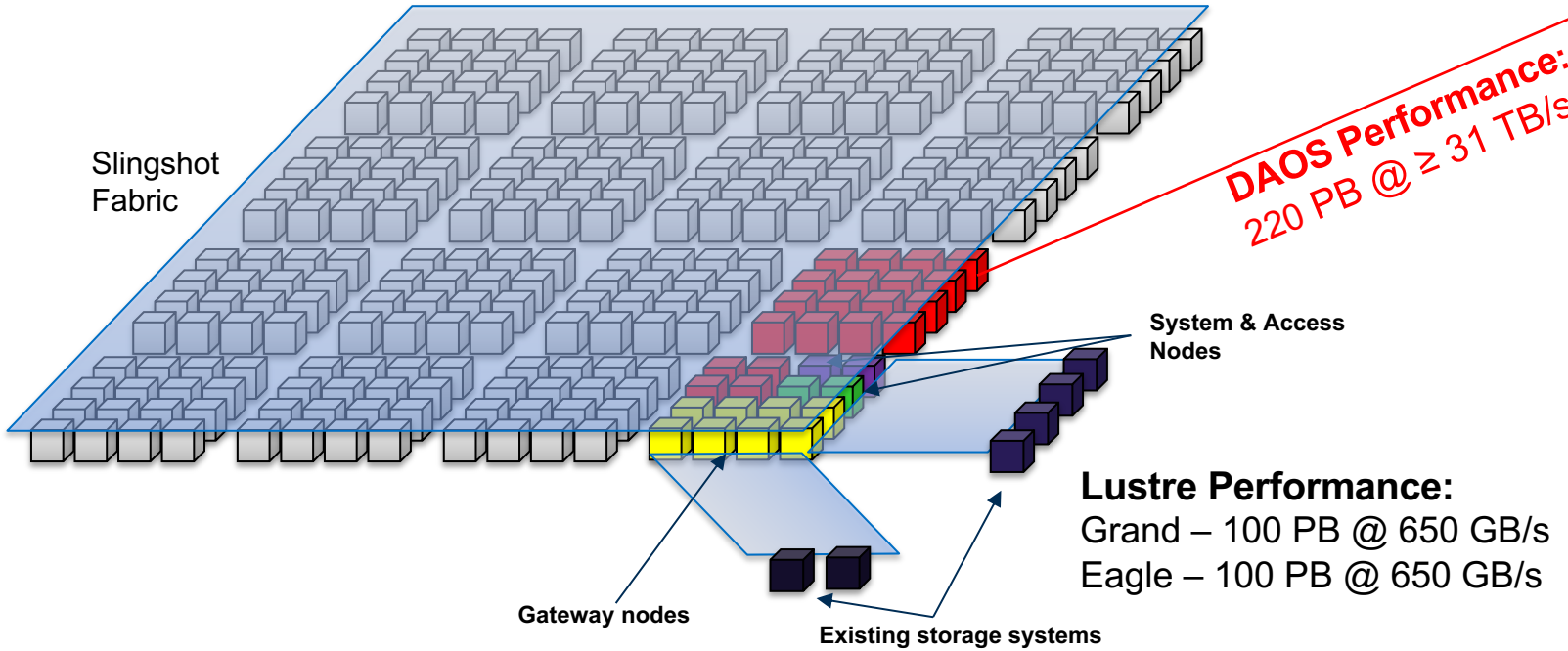
Aurora Storage Systems



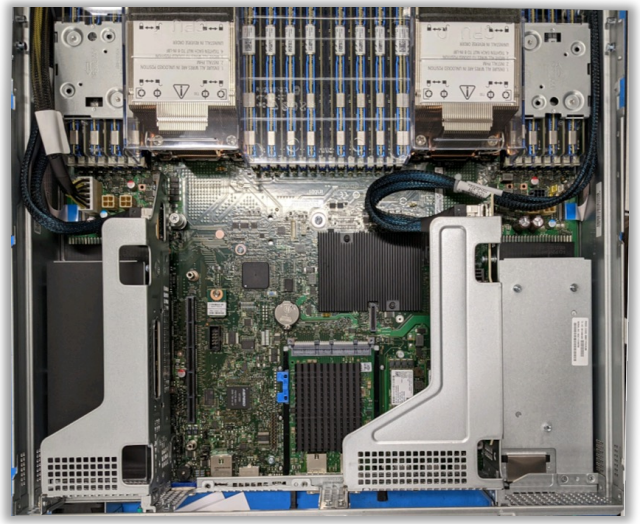
- Aurora will have multiple storage systems:
  - DAOS: provides Aurora's main "platform" high performance storage system
  - Lustre: will utilize existing ALCF systems (Grand, Eagle) for center-wide data access and data sharing
- DAOS is a novel file system architecture the provides:
  - A flexible storage API that enables new I/O paradigms
  - Compatibility with existing I/O models such as POSIX, MPI-IO and HDF5
  - Open source storage solution

System	Capacity	Performance
Aurora DAOS	220 PB @ EC16+2 ▪ 250 PB NVMe ▪ 8 PB Optane PMEM	31 TB/s Read & Write
Eagle	100 PB @ RAID6 ▪ 8480 HDD ▪ 40 Lustre MDT	> 650 GB/s Read & Write
Grand	100 PB @ RAID6 ▪ 8480 HDD ▪ 40 Lustre MDT	> 650 GB/s Read & Write

Aurora Storage Architecture



DAOS Performance:
220 PB @ ≥ 31 TB/s



Lustre Performance:
Grand – 100 PB @ 650 GB/s
Eagle – 100 PB @ 650 GB/s

The Aurora open-source storage strategy strongly favors cooperation:

- DAOS: object storage system for in-fabric high-performance platform storage (the first of its kind on a DOE leadership system!)
- Lustre: parallel file systems for facility-wide access and data sharing

Namespace integration will make it easier for users to manage data.

- 1024 DAOS server nodes, each with:
  - 16 x 512GB persistent memory
  - 16 x 15.3TB NVMe drives
  - 2 x HPE Slingshot NICs
  - Dual CPU with 512 GB RAM

IO-500 results (SC'23)

- RF0
  - [RESULT] ior-easy-write 21029.853648 GiB/s
  - [RESULT] ior-hard-write 10500.042182 GiB/s
  - [RESULT] ior-easy-read 11890.364024 GiB/s
  - [RESULT] ior-hard-read 6348.025367 GiB/s

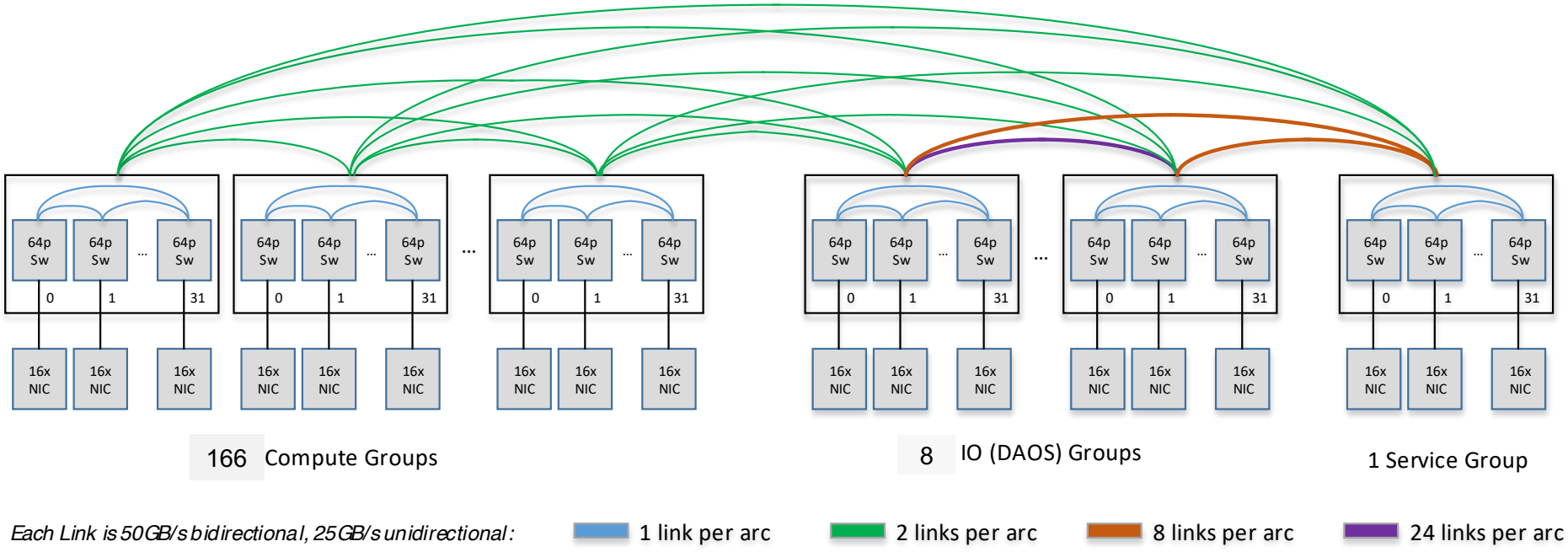
- RF1
  - [RESULT] ior-easy-write 20693.627959 GiB/s
  - [RESULT] ior-hard-write 4216.340094 GiB/s
  - [RESULT] ior-easy-read 12122.872205 GiB/s
  - [RESULT] ior-hard-read 9706.548821 GiB/s

- RF1 configuration
  - EC_16P1GX ior easy
  - RP_2GX ior hard
  - RP_2G1 for mdtest files
  - RP_2GX for mdtest hard dir
  - RP_2G32 for mdtest easy dir

- 300 Aurora compute nodes
  - 18 compute groups

- 642 DAOS servers
  - 8 DAOS groups

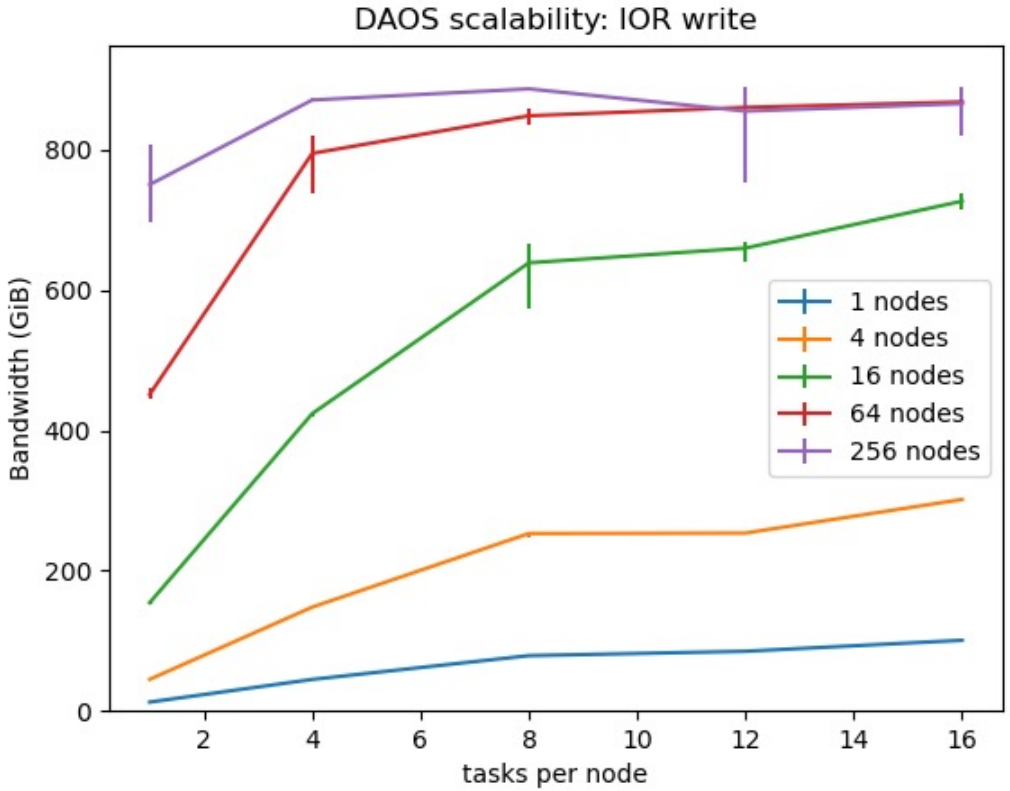
Fabric



- 1-D Dragonfly Topology - 175 total groups (166 compute + 8 IO + 1 Service),
- All the global links are optical, all the local links in compute groups are electrical
- 2 global links between any two compute groups
- 24 links between any two IO groups, 8 links between the Service group and each IO group
- Total injection bandwidth: 2.12PB/s
- Total bisection bandwidth: 0.69PB/s

IO Performance

- Larger number of client nodes increases performance
- DAOS needs reasonable number of ranks per node to achieve ideal performance
  - More ranks per node for efficient single node performance
  - At scale, even inefficient I/O configuration will still saturate
- Distribute I/O process (like MPI) over both sockets to take advantage of all NICs
  - DAOS specifically selects NICs in a round-robin fashion based on NUMA



Data from 20 DAOS node configuration
Courtesy of Rob Latham (Argonne)

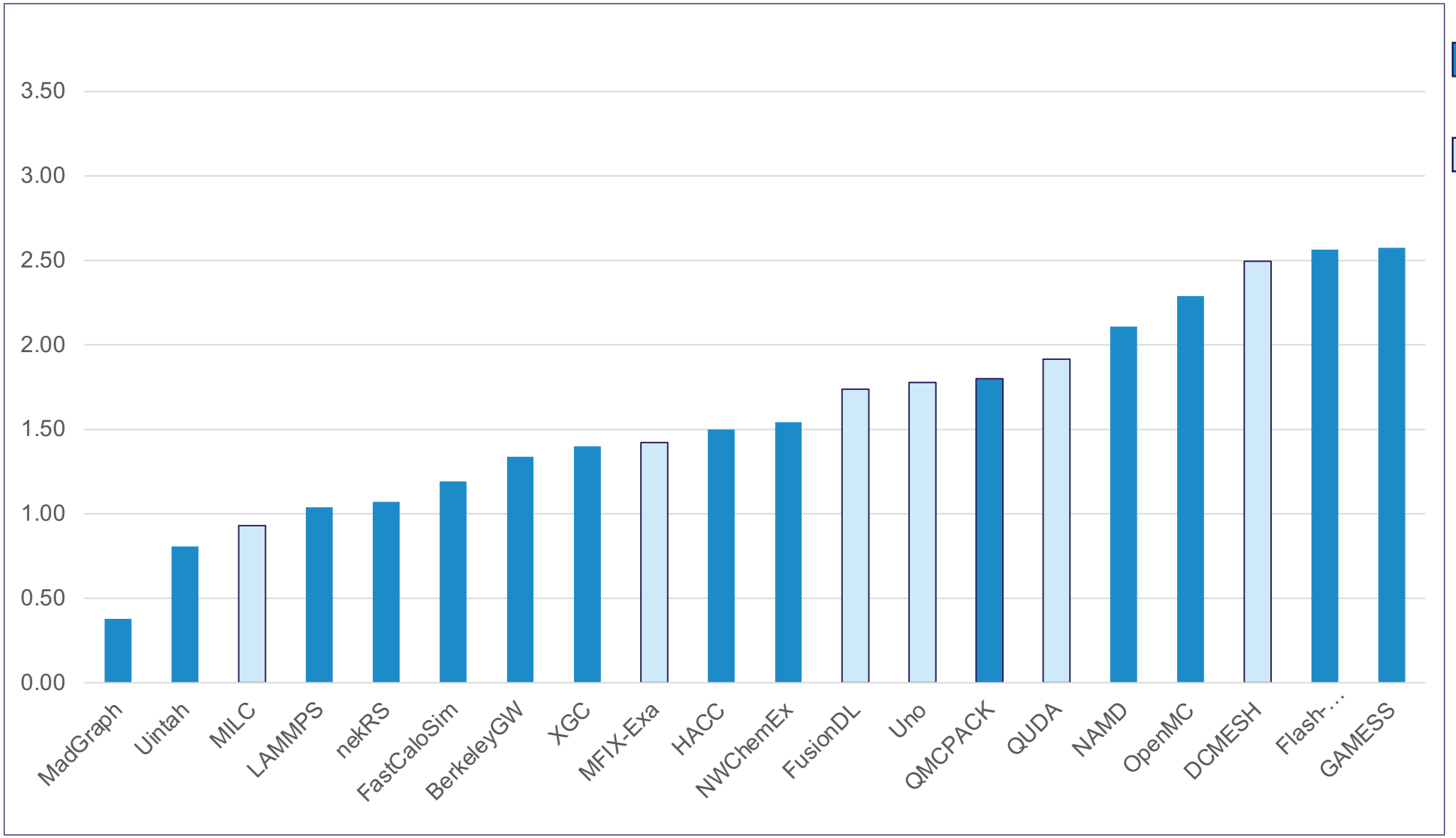
Aurora Applications

Argonne Leadership Computing Facility

Aurora Applications Overview

- Early applications for Aurora come from two programs:
  - —The Argonne Early Science Program (ESP) : 19 projects
    - 9 Simulation projects
    - 5 Learning projects
    - 5 Data projects
  - —The DOE Exascale Computing Project (ECP) : 15 projects
- Some projects contain multiple codes
  - —44 codes are targeted for Aurora as part of ESP or ECP projects
    - 3 codes are intended for use on the CPU only
- Involves effort from over 60 Argonne and Intel staff and over one hundred outside collaborators
- Almost all projects involve teams from outside Argonne

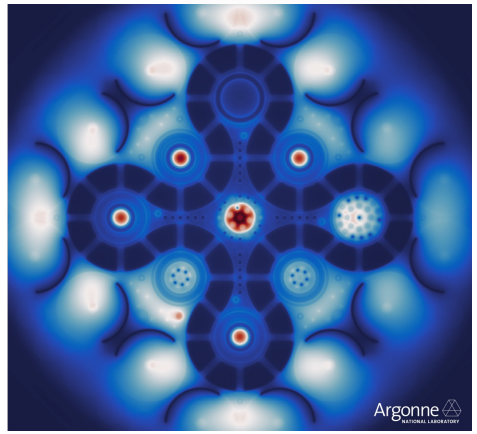
PVC Performance versus A100

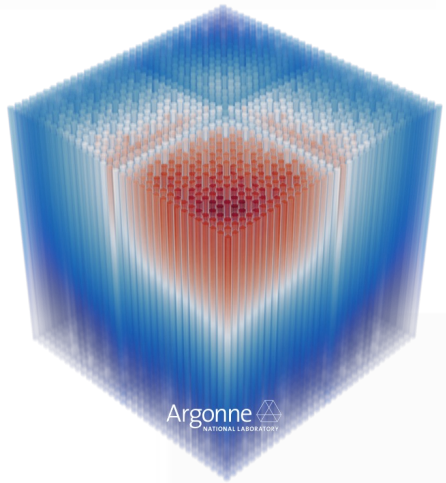


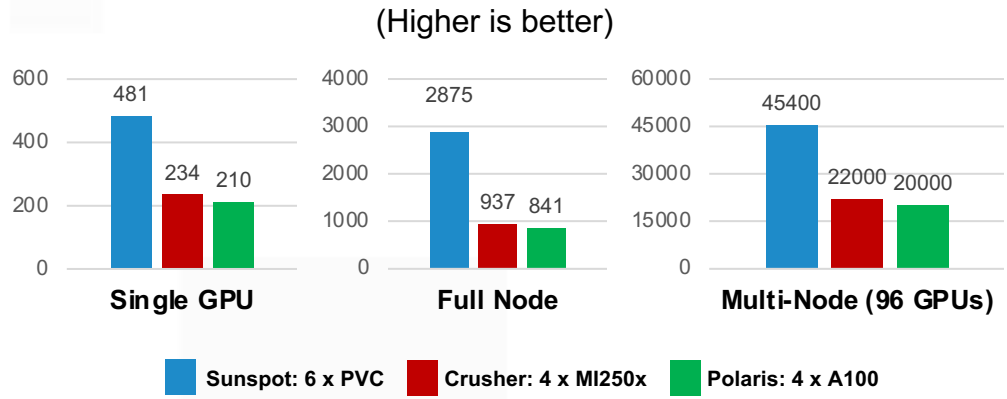
OpenMC *(courtesy of John Tramm)*

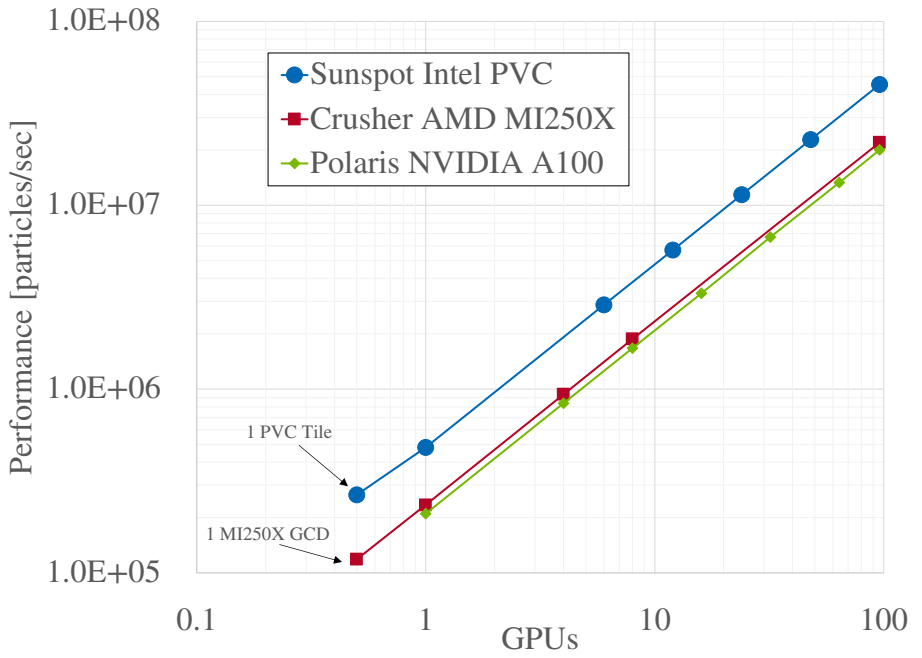
https://docs.openmc.org

- OpenMC is being developed as part of the ECP ExaSMR project (PIs: Steven Hamilton, Paul Romano)
- OpenMC is a Monte Carlo particle transport code written in C++ and the OpenMP target offloading programming model
- The project seeks to accelerate the design of small modular nuclear reactors by generating virtual reactor simulation datasets with high-fidelity, coupled physics models for reactor phenomena that are truly predictive
- The Monte Carlo method employed by OpenMC is considered the "gold standard" for high-fidelity but these methods suffer from a very high computational cost.
- The extreme performance gains OpenMC has achieved on GPUs is finally bringing within reach a much larger class of problems that historically were deemed too expensive to simulate using Monte Carlo methods.





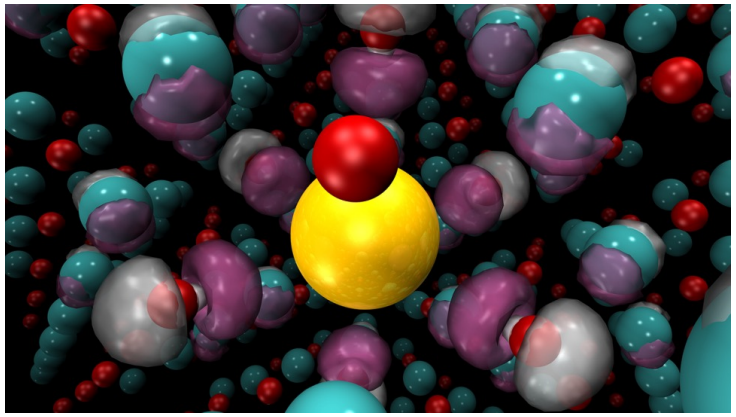


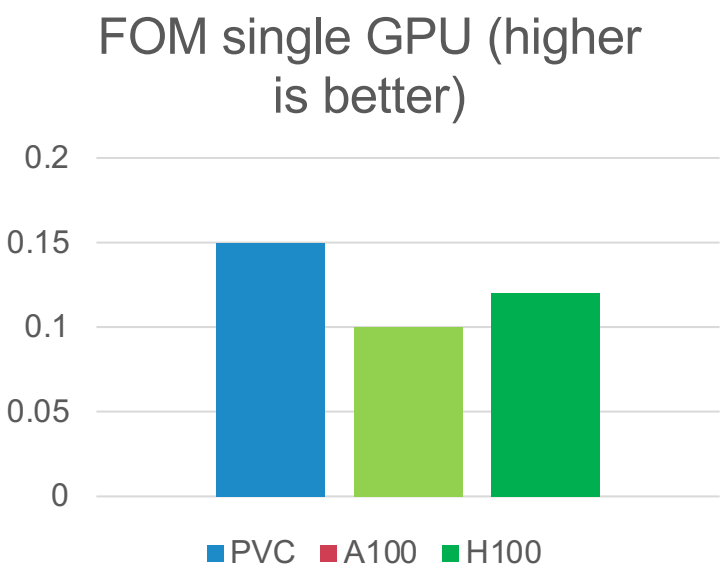


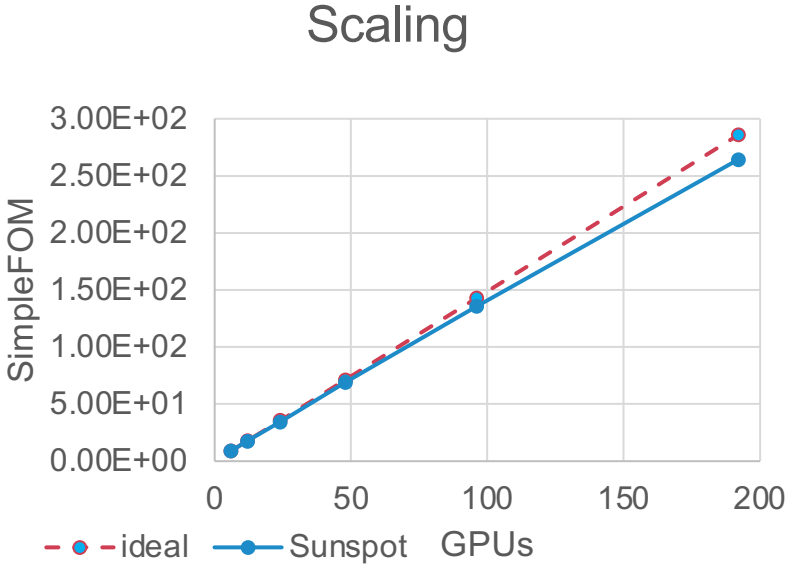
QMCPACK *(courtesy Thomas Applencourt, Ye Luo, Jeongnim Kim)*

ECP Project PI: Paul Kent

- QMCPACK, is a high-performance open-source Quantum Monte Carlo (QMC) simulation code.
- Science case: computing the quantum mechanical properties of materials with benchmark accuracy, including for energy storage and quantum materials.
- QMCPACK uses C++ and OpenMP target offload, plus wrappers (eg SYCL) around vendor optimized linear algebra.







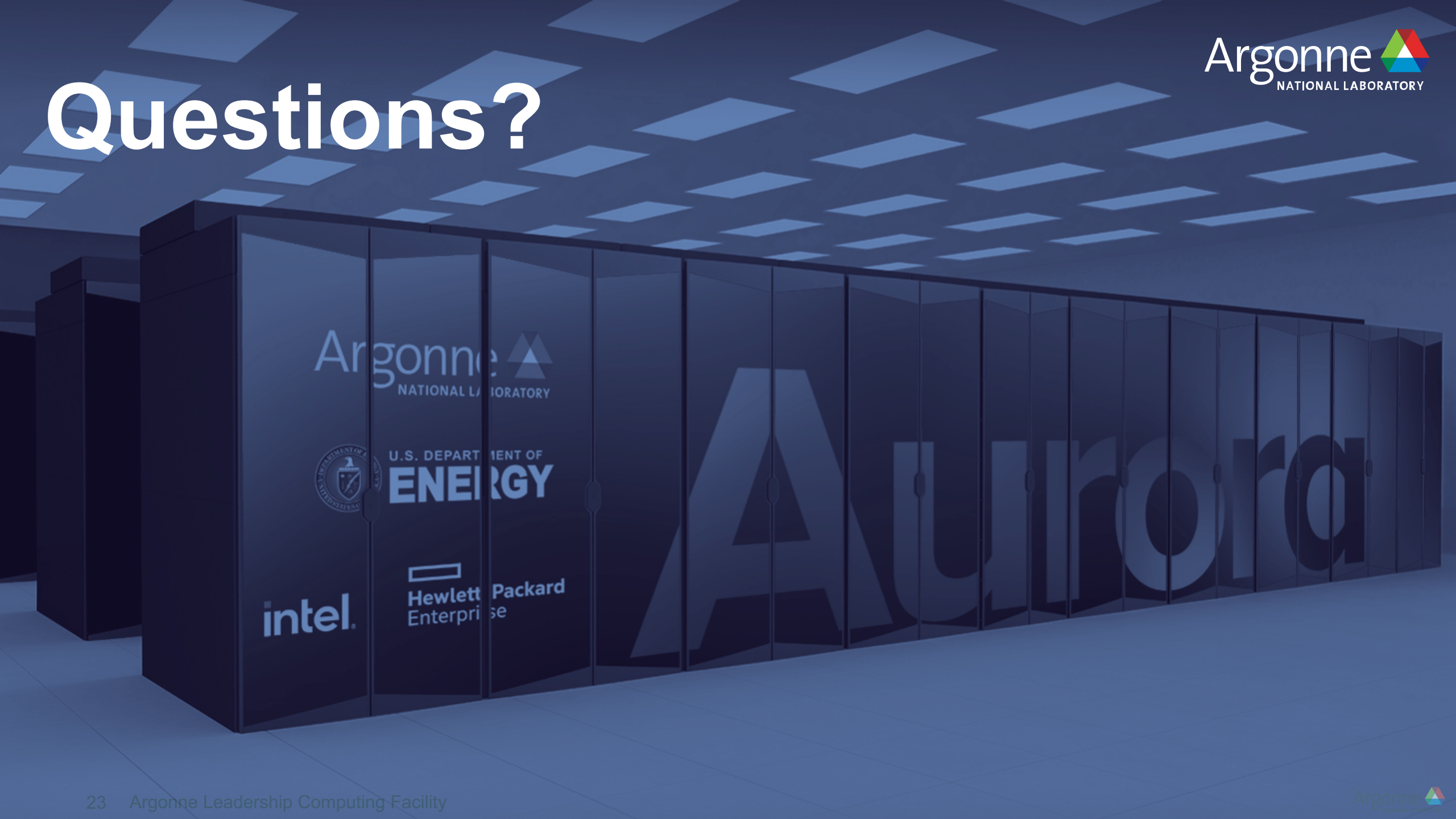
- Running `dmc-a512-e6144-DU64` problem. This simulates a supercell of nickel oxide with 6144 electrons and 512 NiO atoms total.
- Intel® Data Center GPU Max Series: 2 MPI ranks per GPU, 8 Walkers per rank, 64 GB of HBM per stack. Using Intel(R) oneAPI DPC++/C++ Compiler 2022.12.30
- A100 (40GB): 1 MPI Rank, 7 Walkers. LLVM15 compiler.
  H100: llvm/clang 17, cuda 11.8): 1 MPI Rank, 7 Walkers
- The Figure Of Merit (FOM) measure is throughput (walker moves/second). Higher is better.

Final Thoughts

- Off-node communication in the form of I/O to storage or message exchange with other nodes requires some thought related to placement of the job within the network and selection of appropriate configuration on the node for tuneables and process/thread placement

- DAOS will provide a next generation level of performance for applications
  - —Are there ways to utilize these capabilities beyond defensive I/O

Questions?


Argonne
NATIONAL LABORATORY
U.S. DEPARTMENT OF
ENERGY
intel
Hewlett Packard
Enterprise
Aurora