Hardware and Integration

Katie Antypas
ECP, Hardware and Integration Director

ASCAC Meeting
July 29, 2021

EXASCALE COMPUTING PROJECT

ECP's Technical Focus Areas

Providing the necessary components to meet national goals

Performant mission and science applications at scale

Aggressive RD&D project	Mission apps; integrated S/W stack	Deployment to DOE HPC Facilities	Hardware technology advances

Application Development (AD)

Develop and enhance the predictive capability of applications critical to DOE

24 applications
National security, energy, Earth systems, economic security, materials, data

6 Co-Design Centers
Machine learning, graph analytics, mesh refinement, PDE discretization, particles, online data analytics

Software Technology (ST)

Deliver expanded and vertically integrated software stack to achieve full potential of exascale computing

71 unique software products
spanning programming models and run times,
math libraries,
data and visualization

Hardware and Integration (HI)

Integrated delivery of ECP products on targeted systems at leading DOE HPC facilities

6 US HPC vendors
focused on exascale node and system design; application integration and software deployment to Facilities

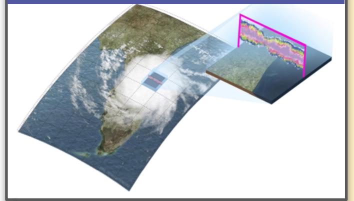
ECP **Hardware and Integration (HI)**

A capable exascale computing ecosystem made possible by “integrating” applications, software, and hardware innovations with training, outreach and allocation management with deep partnerships with DOE Facilities.

HI enables the “last mile.”

Hardware Innovations and Evaluation

Application Integration



Software Deployment

Training Outreach and Allocations

ANL
Intel/Cray

LLNL
Cray

ORNL
Cray/AMD

EXASCALE COMPUTING PROJECT

HI leadership team : Accomplished technical leaders with Facility experience

Katie Antypas, HI Director
15 years experiencing supporting HPC users and deploying HPC systems (LBNL)

Bronis de Supinski, PathForward
5 years as the CTO for the Livermore Computing facility (LLNL)

Scott Pakin, HW Evaluation
17 years in performance analysis and SW development at the ACES Facility (LANL)

Scott Parker, Application Integration at Facilities
13+ years experience working on performance optimization for scientific applications (ALCF)

Susan Coghlan, HI Deputy Director
30 years experience acquiring, deploying, managing extreme scale systems at DOE Facilities (Argonne)

Ryan Adamson, Software Deployment at Facilities
12 years of systems and security administration OLCF HPC Core Operations Group Lead (ORNL)

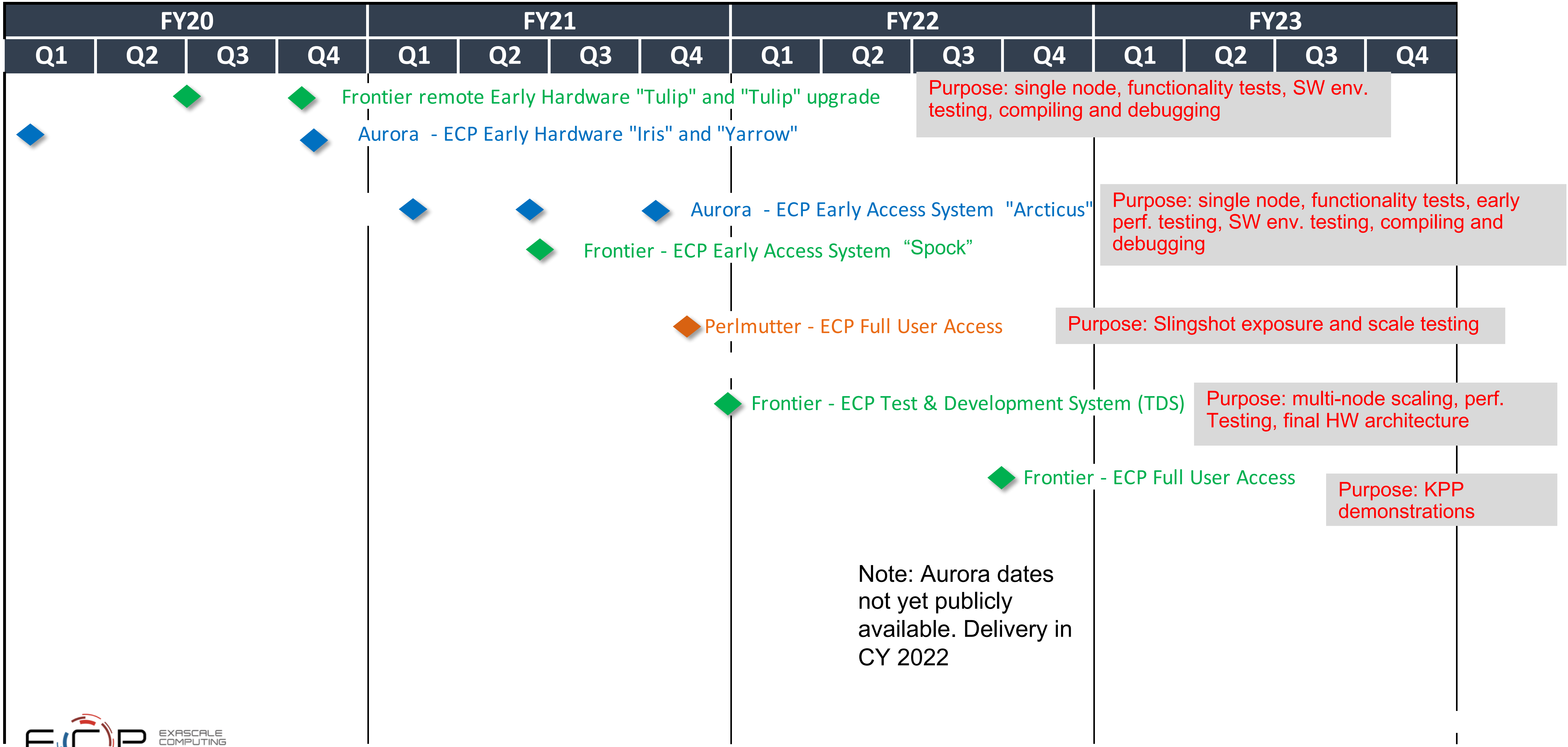
Haritha Siddabathuni Som, Facility Resource Utilization
14 years in field and manager of the ALCF User Experience Team (ANL)

Ashley Barker, Training and Productivity
8 years as a group leader of user assistance and outreach at the OLCF (ORNL)

What’s happened in Hardware and Integration in the last year?

- Early hardware is available for ECP early users at ALCF and OLCF
- Application Integration teams using early hardware to aid in Application Development (AD) teams in optimization and porting
- Added new support for Software Technologies (ST) teams through Application Integration
- Software Deployment team has converged around E4S as deployment vehicle and are building and installing ST software on early hardware
- Training and Productivity sub-project is partnering closely with facilities to put on joint training events
- Perlmutter at NERSC is identifying bugs/issues with system software and networking and serving as risk mitigation for Frontier and Aurora
- Tracking early hardware usage and plans for a new user program for application demonstration
- Hardware Evaluation is set to complete final studies on memory technologies, analytical modeling and network simulation
- PathForward sub-project has completed achieving stretch goals

ECP Teams are accessing early exascale system hardware & software



Early Access Systems

Arcticus at ALCF

HW Description –
- 17 nodes with Intel XeHP GPU (Arcticus)

SW env – SLES + Intel Aurora SDK

Access: Available to ECP members covered under appropriate NDA.

Communication and trouble shooting: Email list / slack channels, Confluence docs

Support: ALCF staff and Intel Center of Excellence staff

Notes: Systems are shared with Argonne Early Science Program. Incomplete feature support, updated frequently.

Spock at OLCF

HW Description – 12 nodes each with:
- 1x 64 core AMD EPYC CPU
- 4x MI100 GPUs w. 32 GiB HBM each
- access to OLCF home and project areas

SW env - SW - RHEL, Slurm, Cray Programming Environment (PE), AMD ROCm (HIP),

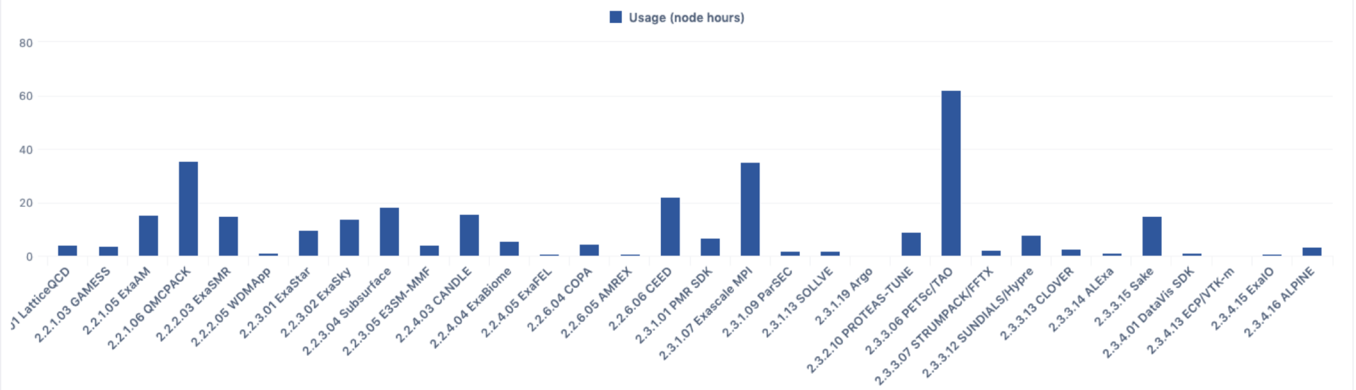
Access: Available to all ECP members *no* NDA required

Communication and trouble shooting: Web documentation, help desk, slack channel

Support: OLCF Staff and HPE Center of Excellence staff

Notes: Hardening of programming environment and site specific configurations on-going

Other early hardware includes Iris and Yarrow systems at ALCF, and Birch and Tulip systems for Frontier



EXASCALE COMPUTING PROJECT

Application Integration at the Facilities Portfolio

Accelerating application readiness for the exascale architectures

Scott Parker, ALCF

Balint Joo, OLCF

Strategy: Match applications with facility readiness efforts.

Progress Assessment: Progress towards technical execution plans measured quarterly; annual external assessment.

Goal: **21** performant exascale applications that run on Aurora, Frontier

Application Development 2.2
- Chemistry and Materials Applications 2.2.1
- Energy Applications 2.2.2
- Earth and Space Science Applications 2.2.3
- Data Analytics and Optimization Applications 2.2.4
- National Security Applications 2.2.5
- Co-Design 2.2.6

Chris Knight

01 Aurora – 14 applications strongly engaged by ALCF for Aurora; others to follow as resources allow. Best practices are being developed and shared.

Stephen Nichols

02 Frontier - 12 ECP applications were selected to participate with CAAR program in July, 2019. Applications may transition in and out of the program as progress is made.

Debbie Bard

03 PreExascale – 6 ECP AD applications identified to participate in NESAP for Perlmutter with ECP funding.

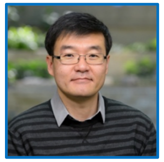
Goal: Progress towards exascale readiness develops, and NESAP-ECP apps transition to LCF facilities after NESAP.

AppInt Teams are deeply engaged and critical to the success of AD teams

Yasaman Ghadar
ALCF
EXXALT

Colleen Bert
ALCF
GAMESS

Thomas App
ALCF
QMCPACK



JaeHyuk Kw
ALCF
ExaWind

Victor Anisi
ALCF
NWChemEx

Timothy Wil
ALCF
WDMApp

Murali Eman
ALCF
CANDLE

Christopher Knight

Abhishek

What do Application Engineers do?

- 45 staff members and post-docs engage with vendor experts, AD teams, and facility performance staff
- Implement algorithms on new architectures
- Optimize applications for new hardware and software
- Implement new programming models in applications

How are teams managed?

- AppInt Engineers are embedded with facilities staff, and also partner closely with AD team
- Quarterly milestones and work plans

Paul Lin
NERSC
WDMApp

Ronnie Chatterjee
NERSC
Combustion

Michael Rowan
NERSC
WarpX

Johannes Blaschke
NERSC
ExaFEL

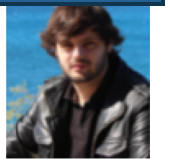
Neil Mehta
NERSC
EXXALT

Felix Wittwer
NERSC
ExaFEL

Servesh Muralidharan
ALCF
ExaFEL

Bronson Messer
OLCF
ExaSky

Rahul Gayatri
NERSC
EXXALT



Raphaël Prat
NERSC
Subsurface

Major impacts and themes from Recent AppInt Milestone Report:

- AppInt engineers have been crucial is helping AD teams gain access and make productive use of early hardware

- AppInt engineers help communicate and report issues to vendors and highlight issues raised by AD teams

- Many AppInt engineers transitioned from working on kernels and mini-apps to full applications on the early hardware

- Compiling and testing applications on early hardware has helped identify issues with the software environment, particularly with less mature areas such as Fortran and OpenMP-offload


EXASCALE COMPUTING PROJECT

ECP-U-2021-XXX

Deploy and Integrate Software Technologies and Applications to Facilities

WBS 2.4, Milestone PM-HI-1170

Scott Parker, Balint Joo, Deborah Bard, Steven Nichols, Chris Knight, and the HI Application Integration Team

Ryan Adamson, Jeff Neel, Shahzeb Siddiqui, Paul Bryant, Jay Srinivasan, Scott Atchley, and the HI Software Deployment Team

June 16, 2021

Part of this report contains information considered as:

HPE Proprietary & Confidential
Intel Corporation Proprietary
AMD Confidential – Distribution with NDA

U.S. DEPARTMENT OF ENERGY | Office of Science
NNSA

- AppInt engineers have partnered closely with programming models teams enabling Kokkos and RAJA and continue to work on porting codes to new programming environments

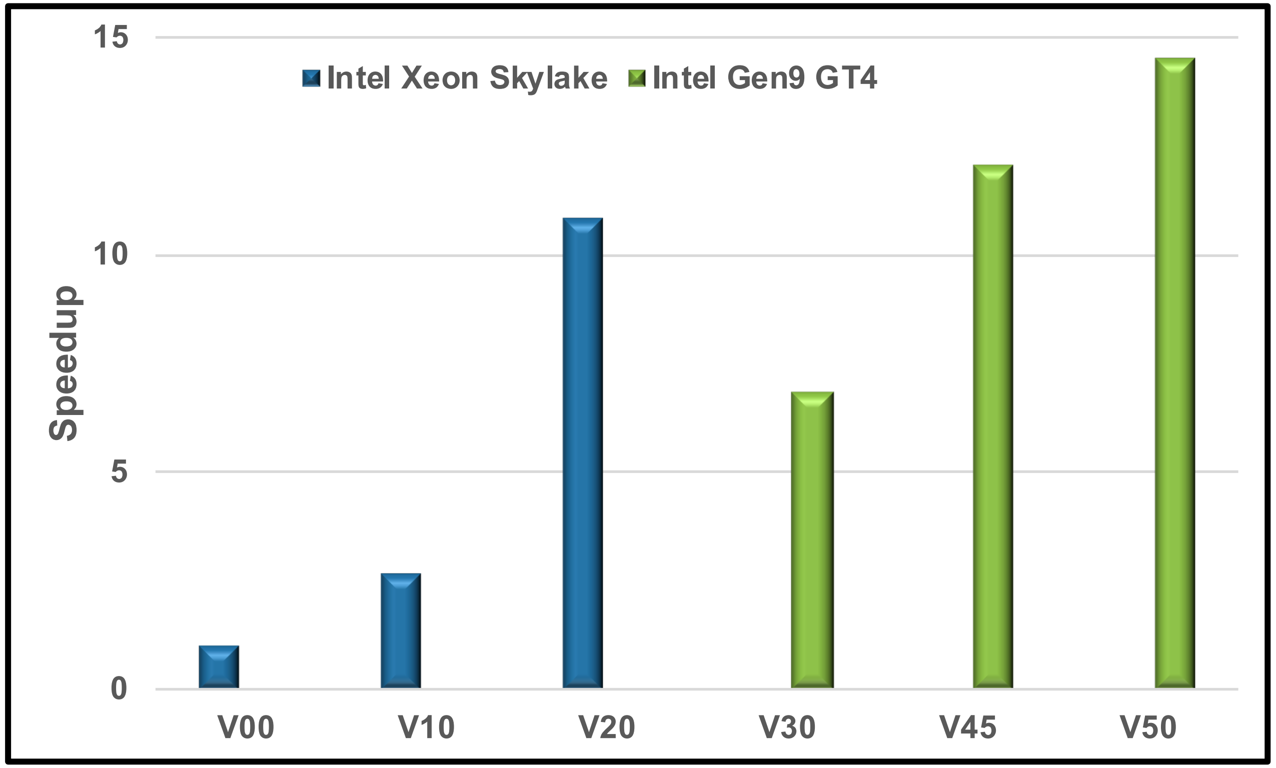
- AppInt engineers have been instrumental in interpreting performance results on early hardware, conducting roofline analyses, and collaborating with AD teams using vendor and non-vendor tools, and leading to significant performance gains.

Current report is NDA, however team is looking into options for a public version next year

GAMESS

Colleen Bertoni, ALCF

- ***Goal:*** Enable quantum chemistry on extremely large systems of interest in catalysis and energy research.
- ***Programming models***: Linear algebra libraries, CUDA, plans for HIP/DPC++ OpenMP
- Key physics module: RI-MP2 electron correlation method kernel
  - MPI/OpenMP threading for CPU with OpenMP offload for part of the RI-MP2 code for GPU



Porting RI-MP2 mini-app to Intel GPUs with OpenMP offload

Series of progressive optimizations, including OpenMP threading (V10), porting to MKL (V20), offloading to GPU (V30), restructuring loops (V45), and enabling concurrent CPU+GPU computation (V50)

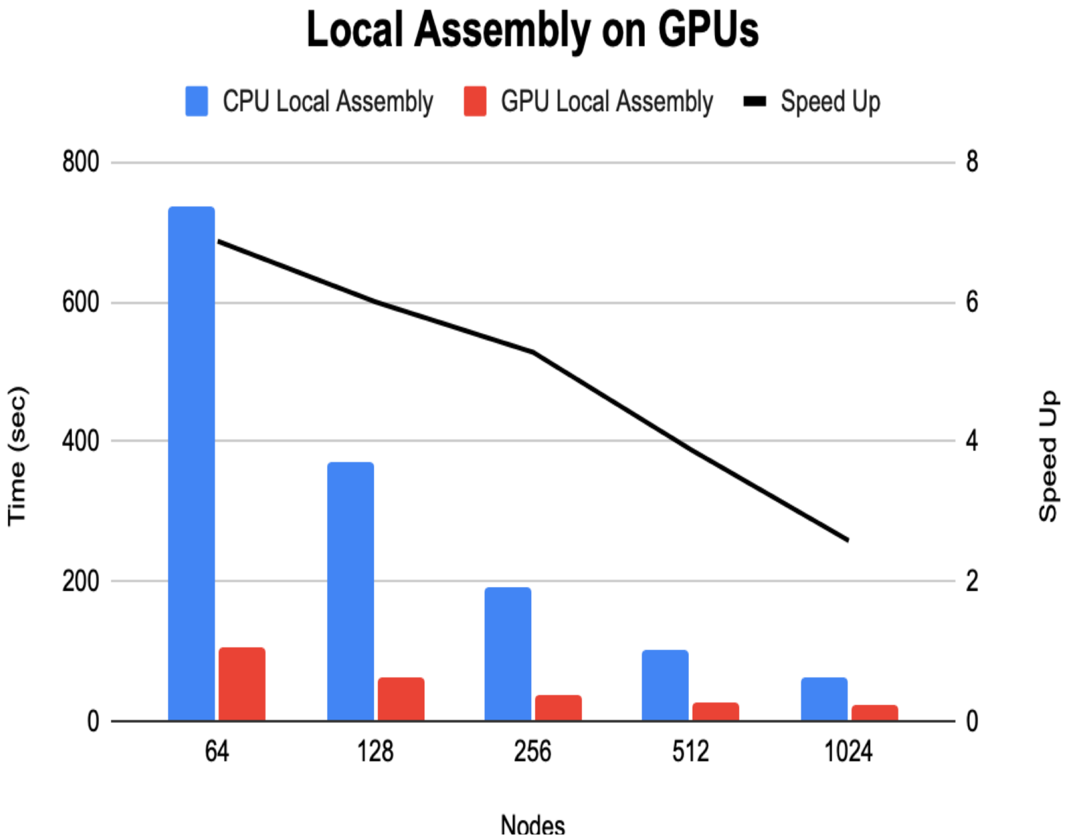
Gamess PI: Mark Gordon, Iowa State University

ExaBiome

Muaaz Gul Awan, Jonathan Madsen,
NERSC

GPU Local Assembly Module for MetaHipMer:

- Graph traversal implemented using hashtables which induce a random memory access pattern that is not suitable for GPUs. Low level CUDA intrinsics used for performant implementation.
- **Local assembly module is 7x faster with GPUs when using 64 Summit nodes.** Performance becomes more communication dominant at higher number of nodes.
- **Integrating this in MetaHipMer pipeline improved performance by upto 42% at 64 nodes**.
- **Local assembly portion has now been reduced from 34% to 6.3% of total MetaHipMer runtime**.
- Accepted to SC21 and **nominated for best paper award**
- Able to process a 16TB metagenome assembly, largest to date



Feedback from ExaBiome:

- Performance improvement on GPUs required low level intrinsics making code somewhat NVIDIA specific.
- Team is working to port code to SYCL and HIP; some differences in intrinsics
- ExaBiome codes rely greatly on Integer performance and for that reason instruction roofline is used frequently for performance deep dives. However, the performance metrics required for constructing instruction roofline are not yet available through Intel and ROCM profilers. We are working closely with vendors on this.

ExaBiome PI: Kathy Yelick, UC Berkeley

NWChemEx

Dmitry Liakh,
and Elvis Maradzike OLCF Post-Doc

NWChemEx is a new redesigned and improved version of the NWChem code capable of performing various quantum-chemistry simulations on heterogeneous exascale HPC platforms

Goal: Larger molecular systems can be calculated **faster** using **coupled-cluster** methods, allowing new quantum chemistry applications

Key challenges:

- Efficient utilization of GPU accelerators
- Optimization of the intra-node data transfers and inter-node communication
- Efficient implementation of the reduced-scaling coupled-cluster algorithms (task granularity challenge)
- Interoperability between all necessary libraries and runtimes

Accomplishments:

- NWChemEx introduces a **portability layer** called Tensor Algebra for Many-body Methods (**TAMM**).
- **TAMM** offloads tensor operations to appropriate processing **backends**.
- **OLCF TAL-SH library** has been integrated with TAMM as a processing backend for NVIDIA and AMD GPU.
- The **CCSD** module of NWChemEx has been fully integrated with the TAL-SH backend on Summit (NVIDIA GPU), demonstrating a **speed-up of 40X** compared to CPU-only NWChem on Titan.

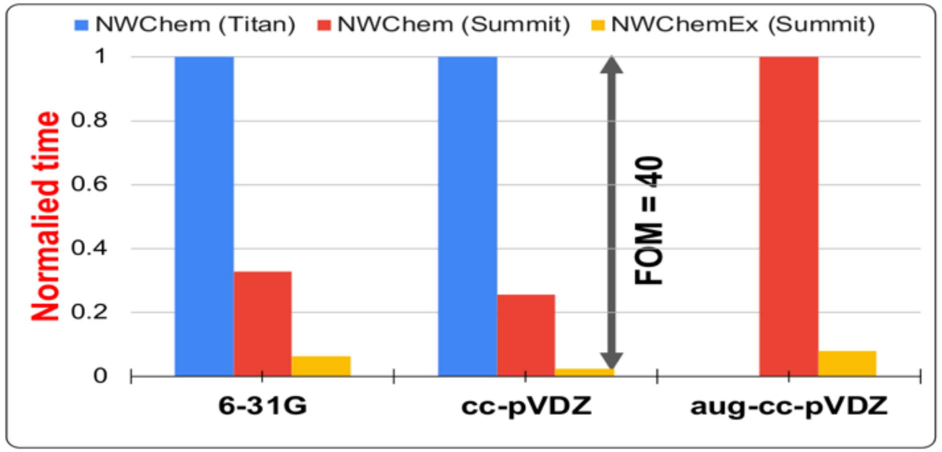
~39,000 basis functions

Basis set	# basis functions
6-31G	424
cc-pVDZ	737
aug-cc-pVDZ	1243

Coupled cluster singles and doubles

Basis	Platform	Nodes	NWChem	NWChemEx
6-31G	Titan Summit	406 100	8.5 min 2.8 min	-- 0.53 min
cc-pVDZ	Titan Summit	890 220	51 min 13 min	-- 1.3 min
aug-cc-pVDZ	Summit	256	74 min	5.8 min

Note: Timing for single iteration



TAMM based coupled-cluster single and doubles (CCSD) achieves **FOM of 40**

EXASCALE COMPUTING PROJECT

NWChemEx PI: Theresa Windus, Ames Lab

We are increasing support for ST teams through Application Integration

- ECP funds vendor support for engaging with ST teams
  - Vendor Center of Excellence support at ALCF for Kokkos, Raja, PETSc/TAO, STRUMPACK, SuperLU, Hypre, SLATE, Trillinos, HeFFTE, VTK-m
  - Vendor Center of Excellence support at OLCF: Kokkos, Exa-PAPI++, VTK-m
- In the last quarter we've also added support for:
  - Math library engagement
  - HDF5 and I/O support
  - Extending the scope and impact of Kokkos
  - Deeper engagement support support for ST teams at the OLCF
  - Increasing vendor support at OLCF

Tom Papatheodore, OLCF

Paul Lin and Rahul Gayatri, NERSC

The Software Deployment team’s role is to package and integrate Software Technology products on the exascale systems

PMR Core (17)	Compilers and Support (7)	Tools and Technology (11)	xSDK (16)	Visualization Analysis and Reduction (9)	Data mgmt, I/O Services, Checkpoint restart (12)	Ecosystem/E4S at-large (12)
QUO	openarc	TAU	hypre	ParaView	SCR	mpiFileUtils
Papyrus	Kitsune	HPCToolkit	FleSCI	Catalyst	FAODEL	TriBITS
SICM	LLVM	Dyninst Binary Tools	MFEM	VTK-m	ROMIO	MarFS
Legion	CHiLL autotuning comp	Gotcha	Kokkoskernels	SZ	Mercury (Mochi suite)	GUFI
Kokkos (support)	LLVM openMP comp	Caliper	Trilinos	zfp	HDF5	Intel GEOPM
RAJA	OpenMP V & V	PAPI	SUNDIALS	VisIt	Parallel netCDF	BEE
CHAI	Flang/LLVM Fortran comp	Program Database Toolkit	PETSc/TAO	ASCENT	ADIOS	FSEFI
PaRSEC*		Search (random forests)	libEnsemble	Cinema	Darshan	Kitten Lightweight Kernel
DARMA		Siboka	STRUMPACK	ROVER	UnifyCR	COOLR
GASNet-EX		C2C	SuperLU		VeloC	NRM
Qthreads		Sonar	ForTrilinos		IOSS	ArgoContainers
BOLT			SLATE		HXHIM	Spack
UPC++			MAGMA			
MPICH			DTK			
Open MPI			Tasmanian			
Umpire			TuckerMPI			
AML						

Legend
- PMR
- Tools
- Math Libraries
- Data and Vis
- Ecosystems and delivery

In Software Deployment we have come a long way in the last 18 months

Dec 2019 Review

- Just transitioned L3 leadership to Ryan Adamson (ORNL)
- Opportunities for integrating spack, SDK, and E4S software ecosystems had not yet been explored fully
- Demonstrated early instances of continuous integration infrastructure on production systems
- Limited testing of ST packages on production systems
- Had a strong vision, but lacked a simple, easily understood plan by stakeholders

18 months later

- Broke apart Software Deployment area into Software Integration (testing and deployment) and Continuous Integration
- Made decision to leverage the E4S software stack and packaging ***significantly simplifying*** testing, deployment and continuous integration plan
- E4S build pipelines ***with CI*** deployed at each facility on production systems (including GPU architectures)
- Software testing on early access hardware has begun
- Site-Local CI deployed on all facility managed early hardware delivered

	Q3 FY21
E4S Testsuite Tests	61/67
ST Spack Tests	42/67
ST Spack Tests Targeting Exascale Features	25/67
ST products ready for installation on early hardware	16/67

EXASCALE COMPUTING PROJECT

Delivering an Open, Hierarchical software Ecosystem

Levels of Integration | **Product** | **Source and Delivery**

ECP ST Open Product Integration Architecture

- Build all SDKs
- Build complete stack
- Containerize binaries

E4S

Source: ECP E4S team; Non-ECP Products (all dependencies)
Delivery: spack install e4s; containers; CI Testing

- Group similar products
- Make interoperable
- Assure policy compliant
- Include external products

SDKs

Source: SDK teams; Non-ECP teams (policy compliant, spackified)
Delivery: Apps directly; spack install sdk; future: vendor/facility

- Standard workflow
- Existed before ECP

ST Products

Source: ECP L4 teams; Non-ECP Developers; Standards Groups
Delivery: Apps directly; spack; vendor stack; facility stack

ECP ST Individual Products

Hardware Evaluation: Study Excess Data Movement

Objective

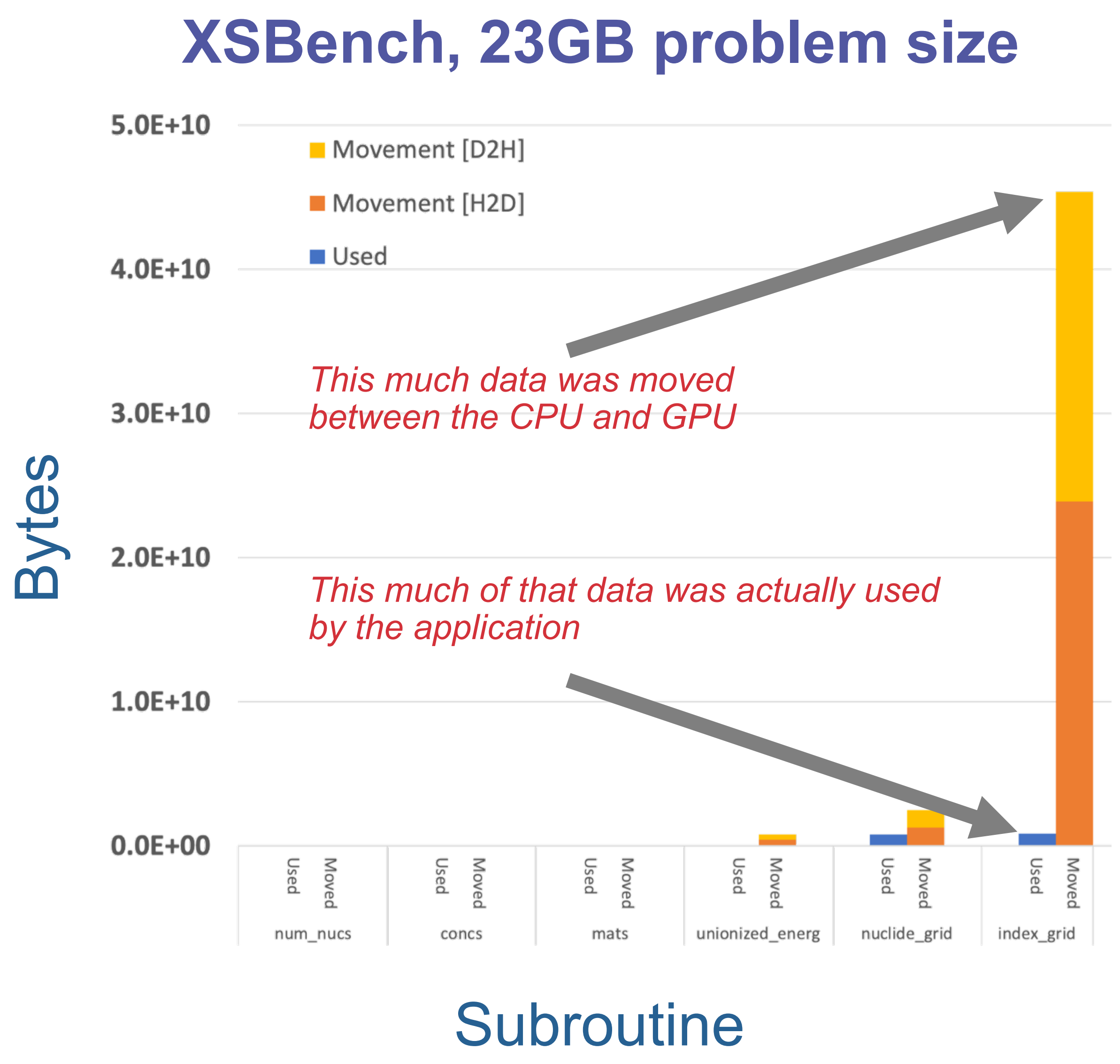
- Identify excess CPU/GPU data movement

Data Movement Insights

- Irregular accesses lead to excess inbound and outbound data movement at both the L1 and L2 GPU cache levels.
- Use of both Unified and managed memory may lead to higher eviction of Unified memory pages

Impact

- Inform Facilities, AD, and ST of excess data movement causes in CPU/GPU interactions
- New and enhanced tools relate excess data movement to application data structures and access patterns
- Identified potential ways to reduce excess data movement

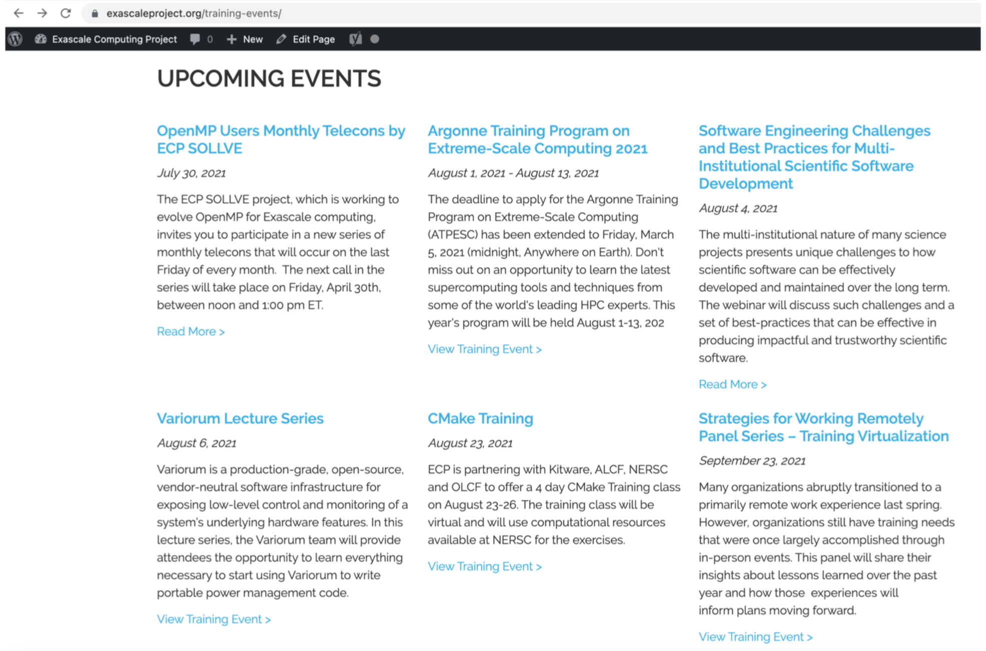


Credit: Ivy Peng, Gwen Voskuilen, David Boehme, Shirley Moore, Maya Gokhale

Selected ECP Training Events

- **Training at the 2021 ECP Annual Meeting**
  - Offered 29 distinct tutorials on a wide variety of topics to ECP community
  - Tutorials were led by members of ECP, DOE Facilities, and/or vendor partners
  - A total of 2,050 people attended the 29 tutorials
  - The two highest attended tutorials were the *Intel COE Tutorial on SYCL and DPC++* (174 attendees) and the *Frontier Tutorial* (169 attendees)
  - Select tutorials were made available on the ECP External Training Website

- **ECP Training Advisory Group**
  - Strong training partnership between ECP and the Labs
  - The TAG has worked together to host a number of training events relevant to ECP and Facility users
    - Upcoming CMake training is an example of close collaboration between ALCF, ECP, NERSC, and OLCF
    - The training is open to users across those organizations and helping to address a high-impact training need and enhance portability across the systems



Summary

- Hardware and Integration is playing a key role in preparing applications and software going forward through strong partnerships with facilities
- Challenges and mitigating risk going forward
  - Deploying applications on the test systems and full exascale systems
    - Expect to be part of system hardening and ECP applications will expose system issues
    - Partner closely with facilities to report bugs and issues as early as possible
    - Application Integration engineers will play crucial role in mitigating risk, responding software and hardware issues and serving as the conduit between facilities and ECP
  - Packaging and deploying software on test and full exascale systems
    - Hardening environment between Spack and newer programming models