The Changing Landscape for Scientific Computing at NSF and DOE

Mike Norman
Director, SDSC
Distinguished Professor of Physics
Tale of 2 Agencies: Different priorities and strategies

• **DOE**
  - **Goal:** International leadership via deployed exascale systems
  - **Vision:** Converged analytics, AI, and simulation at extreme scale
  - **Strategy:** HPC systems deployed through lab procurements
  - **Strategy:** Massive investment in exascale software teams (ECP)

• **NSF**
  - **Goal:** Broaden access to advanced computing for an increasingly diverse set of academic researchers
  - **Vision:** Innovative HPC systems integrated into a national cyberinfrastructure including cloud and campus resources
  - **Strategy:** Cost-effective leadership and capacity systems deployed by winning proposer
  - **Strategy:** Proportionate investment in diverse PI-led software projects (CSSI)
DOE/NNSA exascale roadmap

Pre-Exascale Systems

<table>
<thead>
<tr>
<th>Year</th>
<th>System</th>
<th>Architecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>Titan</td>
<td>ORNL Cray/Nvidia K20 Open</td>
</tr>
<tr>
<td>2016</td>
<td>Theta</td>
<td>Argonne Cray/Intel Xeon/KNL Unclassified</td>
</tr>
<tr>
<td>2018</td>
<td>Mira</td>
<td>Argonne IBM BG/Q Open</td>
</tr>
<tr>
<td>2020</td>
<td>Sequoia</td>
<td>LLNL IBM BG/Q Secure</td>
</tr>
<tr>
<td></td>
<td>Trinity</td>
<td>LLNL Cray/Intel Xeon/KNL Secure</td>
</tr>
<tr>
<td></td>
<td>Sierra</td>
<td>ORNL IBM/NVidia P9/Volta Secure</td>
</tr>
</tbody>
</table>

Exascale Systems

<table>
<thead>
<tr>
<th>Year</th>
<th>System</th>
<th>Architecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>2021</td>
<td>Crossroads</td>
<td>LBNL Cray/AMD/Nvidia Open</td>
</tr>
<tr>
<td>2021-2022</td>
<td>ORNL TBD Open</td>
<td>ORNL TBD Open</td>
</tr>
<tr>
<td>2023</td>
<td>El Capitan</td>
<td>LBNL IBM BG/Q Secure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LLNL TBD Secure</td>
</tr>
</tbody>
</table>

Intel CPUs & GPUs
AMD CPUs & GPUs

2021 est.
2023 est.
NERSC Perlmutter

- Cray Shasta platform
- Slingshot interconnect
- AMD CPUs
- Nvidia GPUs
- Diverse MPI+X programming models supported
- All-flash PFS
ANL Aurora

- Cray Shasta platform
- Slingshot interconnect
- Intel CPUs & GPUs
- Other Intel XPU metrics?
- Diverse MPI+X programming models supported
- Unified node-level programming via One API
DELIVERING AI COMPUTE FROM EDGE TO CLOUD
FROM CPU TO XPU - ONE SIZE DOES NOT FIT ALL

SCALAR
Intel® Xeon®
Scalable Processor Family

VECTOR
Intel®
Discrete Graphics

SPATIAL
Intel®
FPGA

MATRIX
Intel® Nervana™ NNP
Intel® Movidius™ Myriad™
Intel® Mobileye® EyeQ®

ONEAPI UNIFIED DEVELOPER FRAMEWORK
ORNL Frontier

- Cray Shasta platform
- Slingshot interconnect
- AMD CPUs and GPUs
- Diverse MPI+X programming models supported
- Unified node-level programming via ROCm
- $600M / 40 MW
ROCM Tutorials

ROCM comes with a set of tutorials designed to help you understand how you can use and extend the platform.

- **hipCAFFE** How use Caffe on ROCM
- **Vector-Add** Vector Add example using the HIP Programing Language
- **Mini N-body** This sample demonstrates the use of the HIP API for a mini n-body problem.
- **Assembly Sample** The Art of AMDGCN Assembly: How to Bend the Machine to Your Will. This tutorial demonstrates GCN assembly with ROCm application development.
- **Optimizing Dispatches** ROCm With Rapid Harmony: Optimizing HSA Dispatch: This tutorial shows how to optimize HSA dispatch performance for ROCm application development.
- **CLOC Offline Compiler** ROCm With Harmony: Combining OpenCL Kernels, HCC and HSA in a Single Program. This tutorial demonstrates how to compile OpenCL kernels using the CL offline compiler (CLOC) and integrate them with HCC C++ compiled ROCm applications.

© 2016 AMD Corporation [Disclaimer and Legal Information](#)
Upshot: DOE is “all in” on dense GPU clusters

- 3 systems, 3 different kinds of GPUs
- Application porting is a major issue
- Open platforms Perlmutter, Aurora and Frontier require different node-level programming approaches (CUDA, One API, ROCm), creating consternation amongst application teams
  - Only CUDA is mature
- $1.8B being spent on ECP, which is an ambitious software readiness program
  - Big effort in “performance portability” SW: Kokkos, RAJA, hipify
  - Big effort in math libraries too
  - All exascale software will be open source and available via SDKs
Tale of 2 Agencies: Different priorities and strategies

• **DOE**
  • **Goal**: International leadership via deployed exascale systems
  • **Vision**: Converged analytics, AI, and simulation at extreme scale
  • **Strategy**: HPC systems deployed through lab procurements
  • **Strategy**: Massive investment in exascale software teams (ECP)

• **NSF**
  • **Goal**: Broaden access to advanced computing for an increasingly diverse set of academic researchers
  • **Vision**: Innovative HPC systems integrated into a national cyberinfrastructure including cloud and campus resources
  • **Strategy**: Cost-effective leadership and capacity systems deployed by winning proposer
  • **Strategy**: Proportionate investment in diverse PI-led software projects (CSSI)
NSF OAC’s philosophy

Conclusion

- Science and society are being transformed by compute and data – an agile cyberinfrastructure ecosystem is essential
- Rapidly changing application requirements; resource and technology landscapes
  - Our cyberinfrastructure ecosystem must efficiently evolve
- Forward-looking approach to cyberinfrastructure ecosystem, aimed at transforming science
Leadership Computing, part of a Balanced Portfolio

- **Leadership Class Systems**
  - Highly specialized instruments
  - Extreme-scale capabilities

- **Innovative Systems**
  - Support small-medium scale computations, data analytics
  - Emerging and specialized systems

- **Campus Systems**
  - Resources managed, operated by single institutions
  - Serve local, regional (and sometimes national/international) needs

- **Distributed, Federated Systems**
  - Integrate distributed systems (e.g., multicampus, services)
  - Support distributed workflows, exploit data localities
Planning for the Future: OAC-Funded Computing Ecosystem

Community-driven planning and actions

System Hardware and Software Overview

- “Vanilla” 8000 node Dell/EMC Intel CPU cluster
- 56 core dual socket Cascade Lake nodes
- HDR Infiniband fabric
- CPU applications “compile and go”
- Standalone mini-Summit cluster (IBM P9/Nvidia V100)
- Shared PFS
EXPANSE
COMPUTING WITHOUT BOUNDARIES
5 PETAFLOP/S HPC and DATA RESOURCE

HPC RESOURCE
13 Scalable Compute Units
728 Standard Compute Nodes
52 GPU Nodes: 208 GPUs
4 Large Memory Nodes

DATA CENTRIC ARCHITECTURE
12PB Perf. Storage: 140GB/s, 200k IOPS
Fast I/O Node-Local NVMe Storage
7PB Ceph Object Storage
High-Performance R&E Networking

REMOTE CI INTEGRATION

LONG-TAIL SCIENCE
Multi-Messenger Astronomy
Genomics
Earth Science
Social Science

INNOVATIVE OPERATIONS
Composable Systems
High-Throughput Computing
Science Gateways
Interactive Computing
Containerized Computing
Cloud Bursting
Closing thoughts/suggestions

- **SCEC can build on a diversity of compute and data resources available from NSF and DOE**
  - High end simulations can be accommodated via PRAC and INCITE programs
  - Some architectures will require substantial code development to exploit (Aurora, Frontier), others no so much (Frontera)
  - Need to chose your HPC path forward taking software development into account
  - Other NSF resources (Expanse, Bridges 2) may be useful for ensemble runs, data analysis, and data sharing

- **SCEC should strongly advocate for its needs in multiple venues**
  - LCCF Phase 2 requirements gathering (next 12 months)
  - SCEC should advocate NSF OAC for something like PetaApps program for coming leadership computing systems
RELENTLESS INNOVATION CONTINUES

Transistor efficiency (Perf / W)

- 2x scaling
- Planned intra-node optimizations
- 4x Reduction in design rules
- EUV
- Next-gen Foveros & EMIB packaging


- 14
- 14+
- 14++
- 10
- 10+
- 10++
- 7
- 7+
- 7++
- 1272
- 1274
- 1276
INCREASING THE PACE OF INNOVATION


INTEL® XEON® PROCESSOR E5 V3 HASWELL
INTEL® XEON® PROCESSOR E5 V4 BROADWELL
INTEL® XEON® SCALABLE PROCESSOR SKYLAKE
2ND GEN INTEL® XEON® SCALABLE PROCESSOR CASCADE LAKE
COOPER LAKE & ICE LAKE
SAPPHIRE RAPIDS
NEXT GEN

DRIVING LEADERSHIP WORKLOAD PERFORMANCE

5 TO 7 QUARTER CADENCE
MOVING TO
4 TO 5 QUARTER CADENCE

intel 2019 INVESTOR MEETING