SAND2022-4433D

Sandia National Laboratories

SNL ATDM* Software Ecosystem Operating Systems and On-Node Runtime

Stephen Olivier (PI), Ron Brightwell (PM), Matthew Dosanjh, Kurt Ferreira, Scott Levy, Kevin Pedretti, Andrew Younge

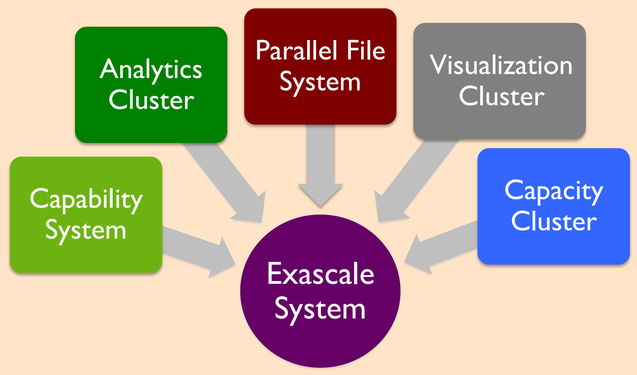
Key Thrusts

- **Containers and virtualization** technologies for productivity, portability, and performance
- **Application characterization** to understand MPI usage and response to system noise
- **Runtime systems** for on-node tasking, centered on the SNL Qthreads multithreading library
- **Standards body participation**, encompassing work in the MPI Forum and the OpenMP Language Committee
- **Preparing for future platforms**, especially NNSA ASC Advanced Technology Systems (ATS) and Vanguard Advanced Architecture Prototype Systems (AAPS)

Interactions

- **Applications:** Representative applications incorporating simulation, analytics, visualization, and/or tool components
- **Resource Management:** OS/runtime mechanisms, policies, and interfaces to enable more effective resource management
- **Testbeds:** Novel platforms able to test custom OS/R stacks

The growing complexity of applications and platforms requires a flexible system software stack, not a one-size-fits all solution.



Highlight: Containers

- **Adoption of Containers in HPC** is limited but growing, as frameworks like Singularity/Apptainer, Shifter, PodMan, and CharlieCloud mature and evolve.
- **Containerizing full mission DOE apps** such as NALU and SPARC and running at scale (2048 nodes / 114,688 containers) provided insights into app performance and viable usage models.
- **Diverse stakeholders** – vendors, facilities, dev ops teams, and app developers – have collaborated with us on solutions for container use cases.

Recent Publications/Presentations:

R. Priedhorsky et al., "Minimizing privilege for building HPC containers", *SC21*, Nov. 2021. IEEE.

S. Levy and K.B. Ferreira, "An Initial Examination of the Effect of Container Resource Constraints on Application Perturbation", *2021 RADR* workshop at *IPDPS*, June 2021.

Highlight: Standards Work

- **OpenMP Language Committee** contributions include work on complex memory support, interoperability, C++ support, and task parallelism.
- **OpenMP 5.2** specification released in November 2021, one year after 5.1.
- **MPI Forum** contributions include work on MPI sessions and partitioned communication for use with threading.
- **MPI 4.0** released in June 2021, six years after MPI 3.1.

Recent Publications/Presentations:

R.E. Grant et al., "A Portable Implementation of Partitioned Point-to-Point Communication Primitives", *EuroMPI 2020*, Nov. 2020. Springer.

S. L. Olivier, "Evaluating the Efficiency of OpenMP Tasking for Unbalanced Computation on Diverse CPU Architectures", 2020 *Intl. Workshop on OpenMP*, Sept. 2020, Springer.

Highlight: MPI Analysis

- **MPI usage in applications** is not well understood, despite its crucial impact on application performance and design of future HPC interconnects.
- **Our extension of the LogOPSim** network simulator analyzes the behavior of MPI message matching with metrics such as queue depth.
- **Using traces of real applications**, we evaluate their MPI usage, optimize where possible, and influence vendor hardware and software design decisions.

Recent Publications/Presentations:

K.B. Ferreira and S. Levy, "Evaluating MPI resource usage summary statistics", *Parallel Computing*, Vol. 108, Dec. 2021. Elsevier.

K. B. Ferreira et al., "Hardware MPI message matching: Insights into MPI matching behavior to inform design", *Concurrency and Computation. Practice and Experience*, 32 (3), Feb 2020. Wiley.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.
SAND No. SAND2022-XXXX

* Advanced Technology Development and Mitigation (ATDM) is a subprogram of the NNSA Advanced Simulation and Computing (ASC) program. Through ATDM, the NNSA supports the development of open-source software technologies that are both important to the success of national security applications and externally impactful to the rest of the Exascale Computing Project and the broader community.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.