

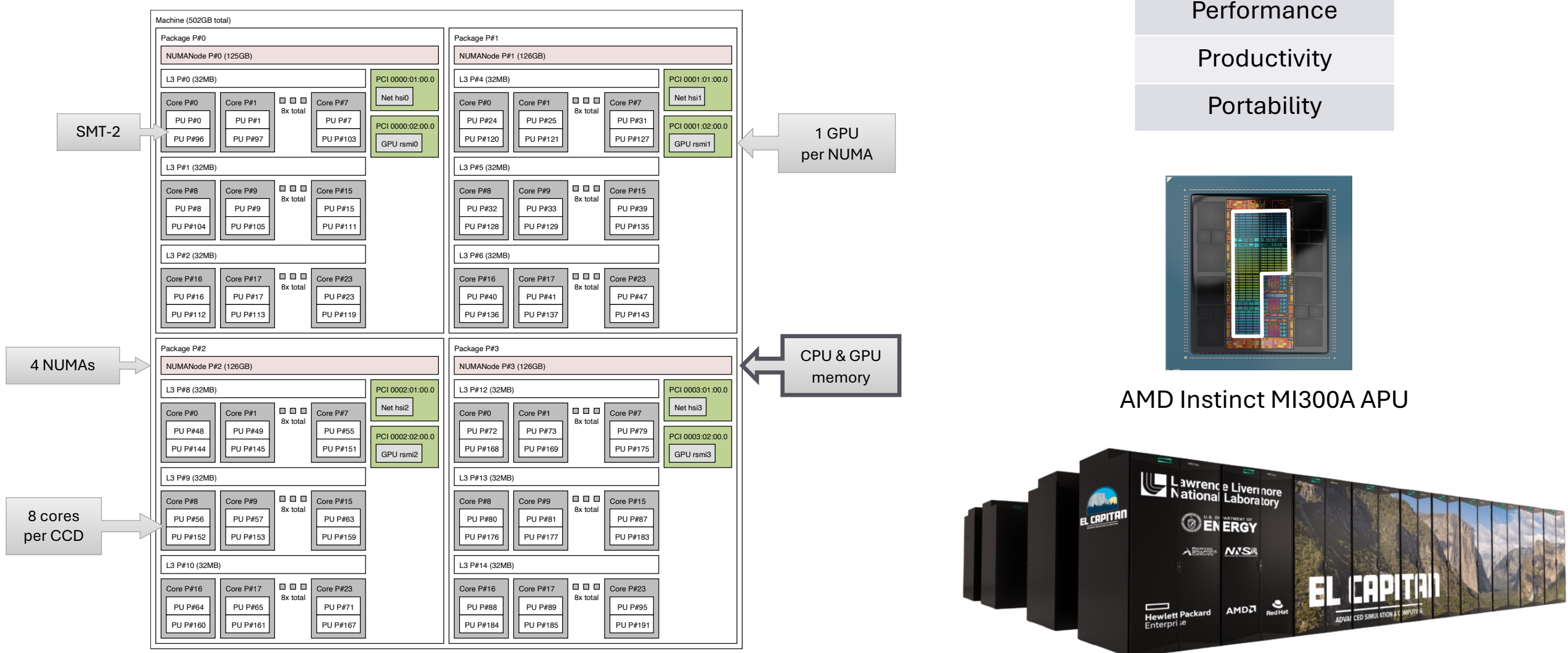
mpibind: Mapping Applications to Complex Computer Architectures

Edgar A. León (LLNL)

The complexity of modern computer systems presents significant challenges for computational scientists. The open-source mpibind library increases user productivity by mapping parallel applications to the underlying hardware efficiently, portably, and without user intervention. Project funded by NNSA's ASC program since 2014.

 github.com/LLNL/mpibind

Scientists Face Complex Computing Architectures

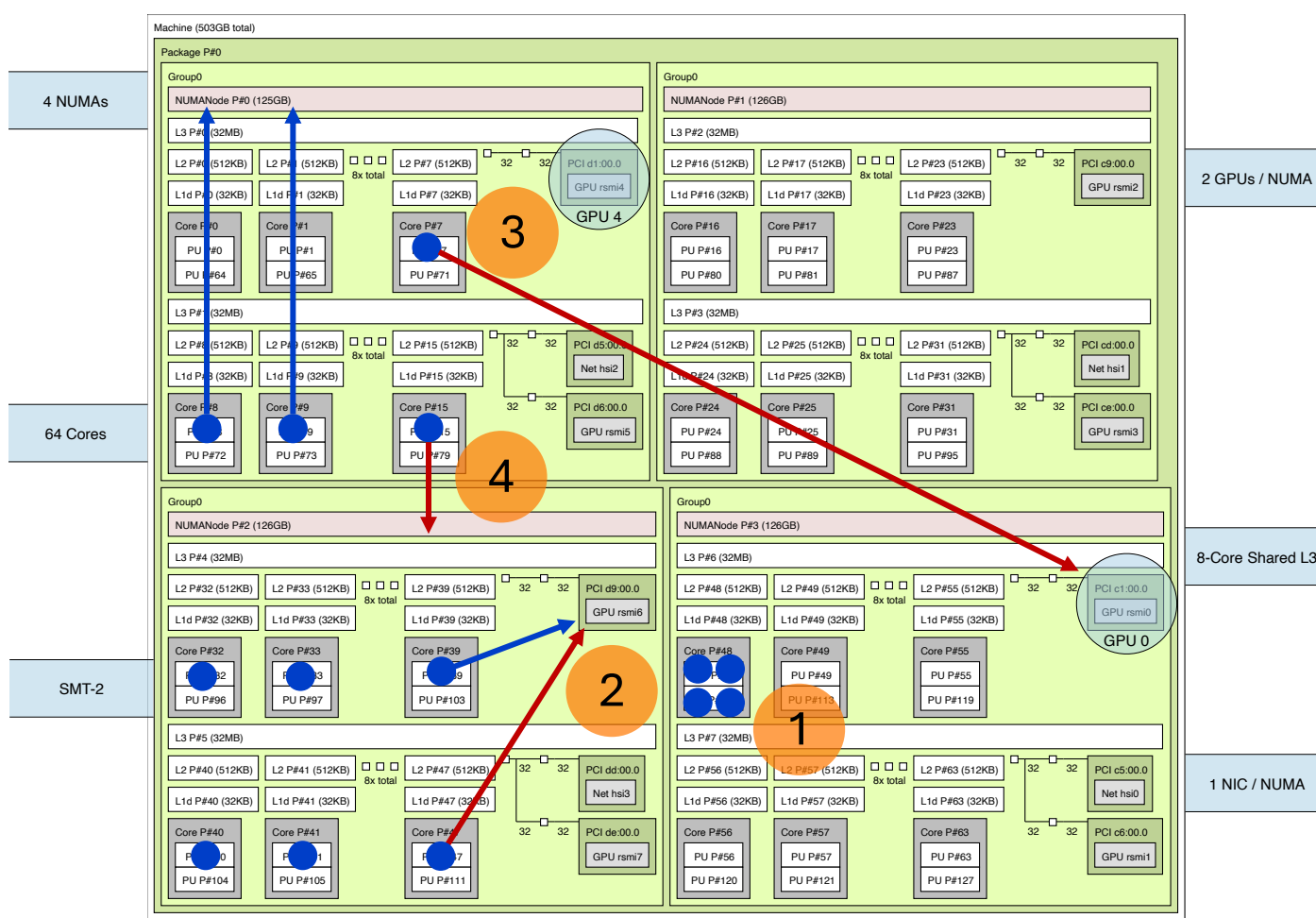


Compute-node architecture of El Capitan at LLNL.

Mapping Applications to the Machine Is Hard

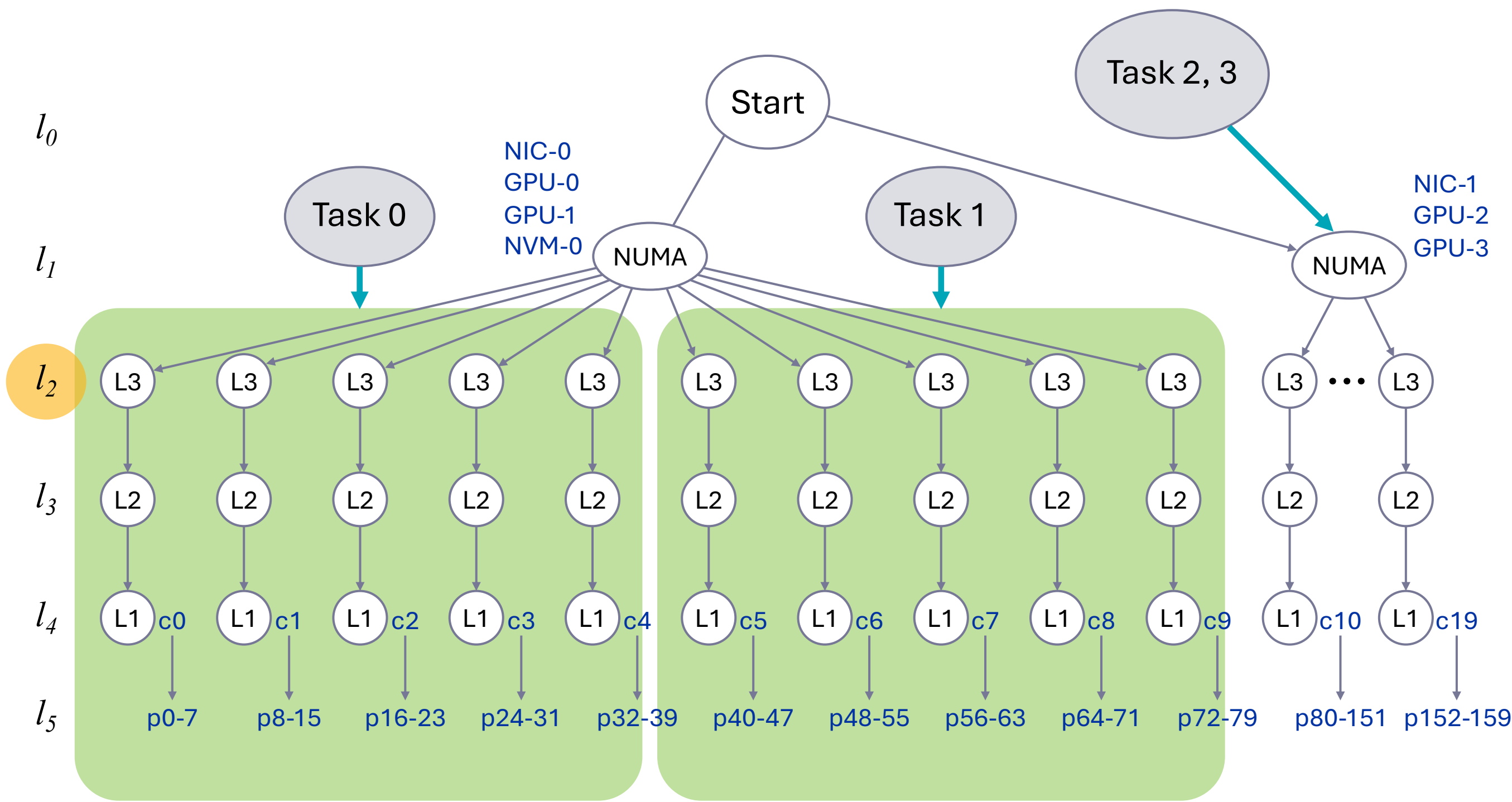
What could go wrong?

1. Multiple threads running on a single core
2. Multiple GPU kernels sharing a GPU
3. Processes launching kernels on non-local GPUs
4. Threads accessing remote memory, etc.



Compute-node architecture of Frontier at ORNL.

mpibind Follows the Memory Hierarchy to Map Applications Efficiently



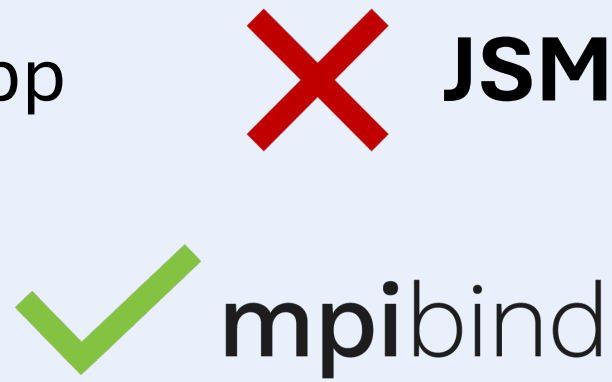
mpibind Provides a Simple Interface and Requires Minimal User Input

8 MPI processes on a CORAL node (2 NUMAs, 40 cores, 4 GPUs)

`jsrun -a 2 -c 10 -g 1 -r 4 -d packed -b packed:5 ./app` ❌ JSM

`srun -n8 -mpibind=on ./app`

`flux run -n8 -o mpibind=on ./app`



Required	Number of tasks
Optional	Number of threads
	Greedy (true or false)
	GPU optimized (true or false)
	SMT (1 to num. HW threads per core)
	Restricted topology (Mem or CPU)

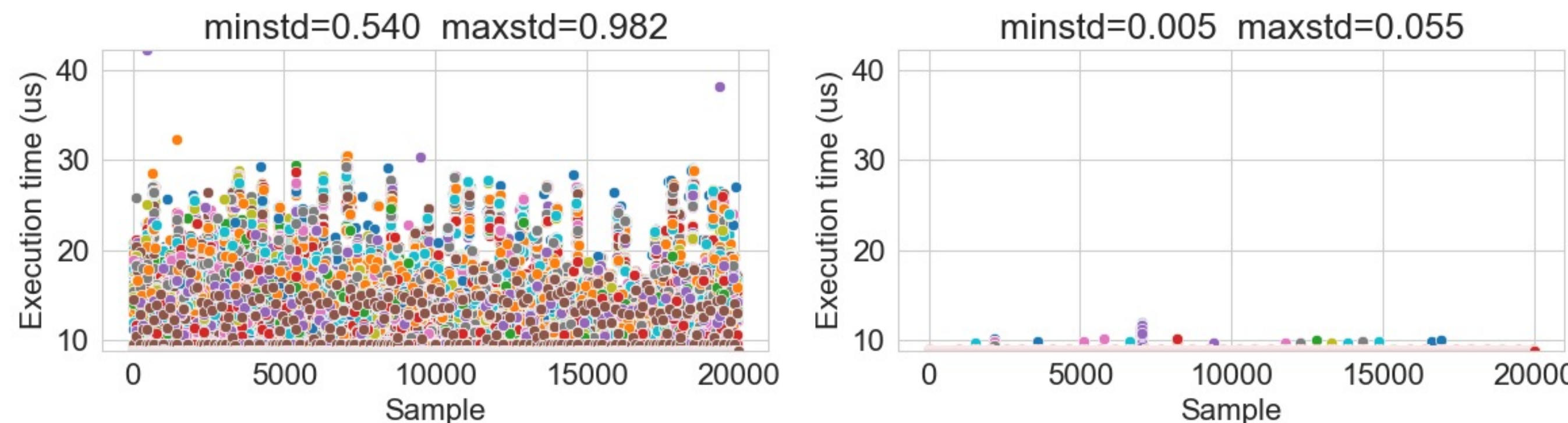
mpibind Provides Portability

- Mapping algorithm relies on abstract memory-compute tree
- Algorithm is separate from applying affinity

Portable Mapping Algorithm	Slurm and Flux Plugins
Given application tasks: (1) Assign CPUs (2) Assign GPUs (3) Provide thread mapping	Use job info as input to mpibind
	Get mpibind mapping
	Bind tasks
	Set environment variables

mpibind Is a Critical Component for Mitigating System Noise

Parallel FWQ on El Capitan: Before (left) and after (right) noise mitigation



mpibind Helps Scientists Make Better Use of Modern Supercomputers

Performance	• Leverages local memory and local devices • Helps mitigate system noise • Maximizes the cache available to each worker
Productivity and Portability	• Provides a simple interface • Requires minimal user input • Uses same algorithm across architectures, MPI libraries, and resource managers

mpibind increases scientists' productivity across computer architectures, MPI libraries, and resource managers