Exascale computing challenges and approaches

Ian Karlin

Lawrence Livermore National Laboratory

Lawrence Livermore National Laboratory

August 19, 2020

This work was performed under the auspices of the U.S. Department of Energy by Argonne National Laboratory under Contract DE-AC02-06-CH11357, Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, Los Alamos National Laboratory under Contract DE-AC5206NA25396, and Oak Ridge National Laboratory under Contract DE-AC05-00OR22725.

Rob Neely's slide from 2014: Sequoia is on the Path to Exascale

- Many of the new challenges presented by Sequoia will be directly transferrable to future ATS platforms
  - Trinity, Sierra, …
- Many expected challenges can be deferred *for now*
- Sequoia may be difficult, but it could be a lot worse!

Exascale Challenge	Sequoia
Extreme MPI scaling	X
Low memory capacity per core	X
Low memory bandwidth	X
Simple "in-order" cores	X
Fine-grained threading	X
SIMD	X
Heterogeneity	
Non-uniform memory	
Multi-level memory hierarchies	
Burst Buffers	
Hierarchical interconnects	
OS noise	
Application-driven resilience	
Application-driven power management	

My update today:
Sequoia was the next step on the Path to Exascale and Sierra

- Many of the new challenges presented by Sequoia and then Sierra will be directly transferrable to future ATS platforms
  - Crossroads, El Cap …
- Sierra has been a challenge, but without Sequoia lessons learned and prep it might have been intractable.

Exascale Challenge	Sequoia	Sierra
Extreme MPI scaling	X	
Low memory capacity per core	X	GPU
Low memory bandwidth	X	X
Simple “in-order” cores	X	X
Fine-grained threading	X	X
SIMD	X	SIMT
Heterogeneity		X
Non-uniform memory		X
Multi-level memory hierarchies		X
Burst Buffers		X
~~Application-driven resilience~~*		
~~Application-driven power management~~		
Performance portability		X
Complex workflows		Start

*OS noise and hierarchical interconnects also removed because challenges are being solved by vendors

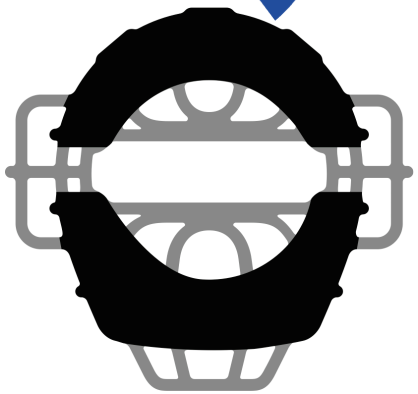
Exascale is an evolution from pre-exascale machines with some challenges mitigated and others growing in complexity.

As you well know exascale systems are heterogenous and GPU accelerated.

The accelerators themselves are diverse and most exascale codes still need to run on CPU only machines.

Performance portability is one of the largest challenges and NNSA has various performance portability approaches

Kokkos



Umpire

Fortran and C codes from all Labs

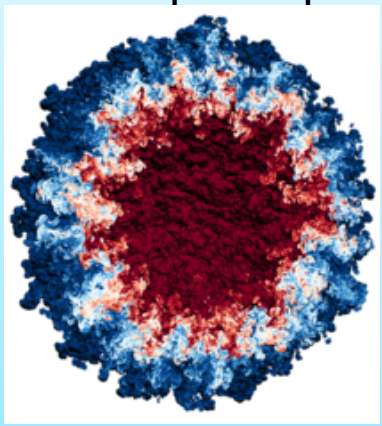
Some codes with a few hot loops program to the metal

HIP

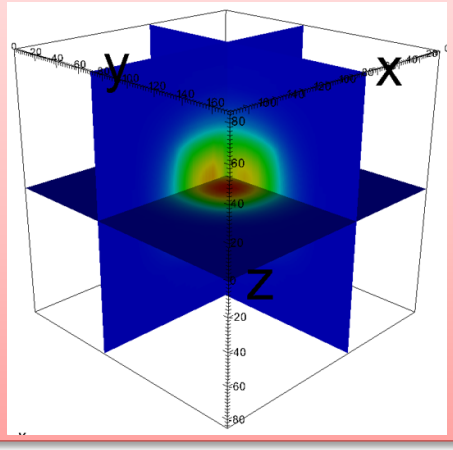
In addition, to addressing performance portability concerns these tools address other challenges, such as, managing the memory hierarchy.

GPU-based computing is leading to large performance increases on diverse applications

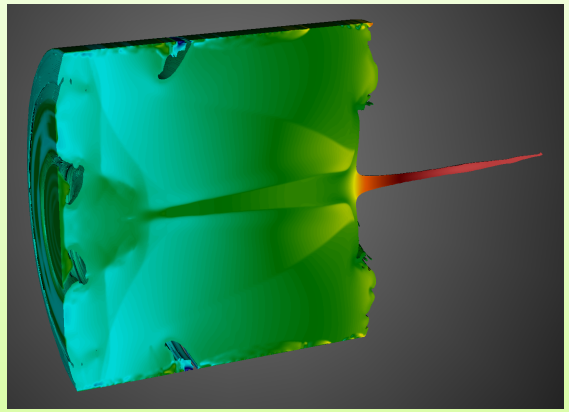
Ares, RT Mixing
13x speedup



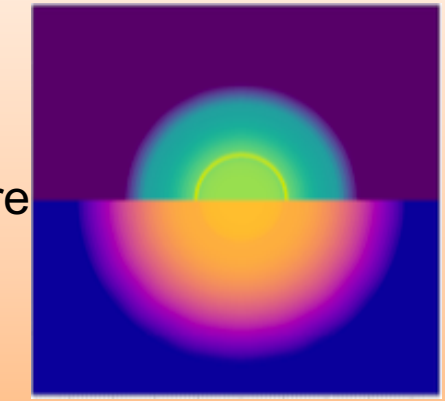
Ardra, Reactor Safety
16x speedup



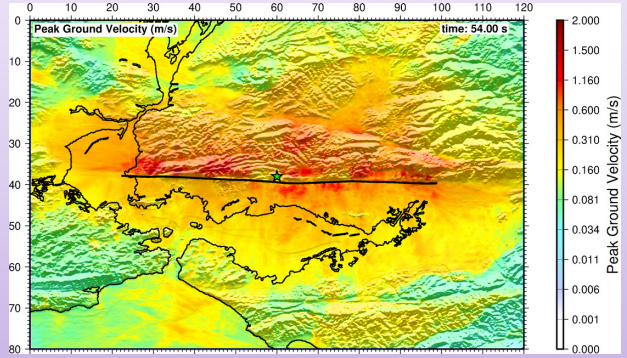
ALE3D, Shaped Charge
8x speedup



Kull/Teton
Radiating Sphere
7x speedup



SW4, Hayward Fault, 28x speedup



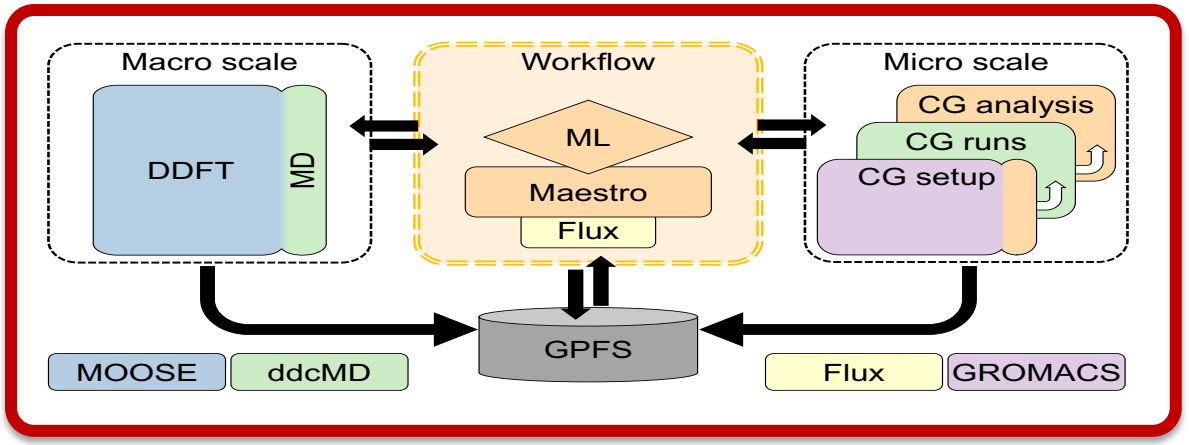
However, it's not all sunshine, lollipops, and rainbows on GPUs. Exascale will involve solving or mitigating these challenges.

- MPI performance concerns
  - Kernel launch overheads cause long times packing buffers
  - When to send from the CPU vs. GPU
- Some apps/algorithms aren't there yet
  - Monte Carlo
  - Multi-grid setup phase
- Tool suites don't handle all of our use cases
  - Tools do not provide all the debugging and performance information we desire
- Interoperability of parallel models (OMP, CUDA, …), compliers, memory handling, across libraries can be difficult
- Strong scaling is limited by need for massive parallelism to feed a GPU

Many of your projects are looking at some of these issues.

For NNSA Exascale is a way point not an end goal

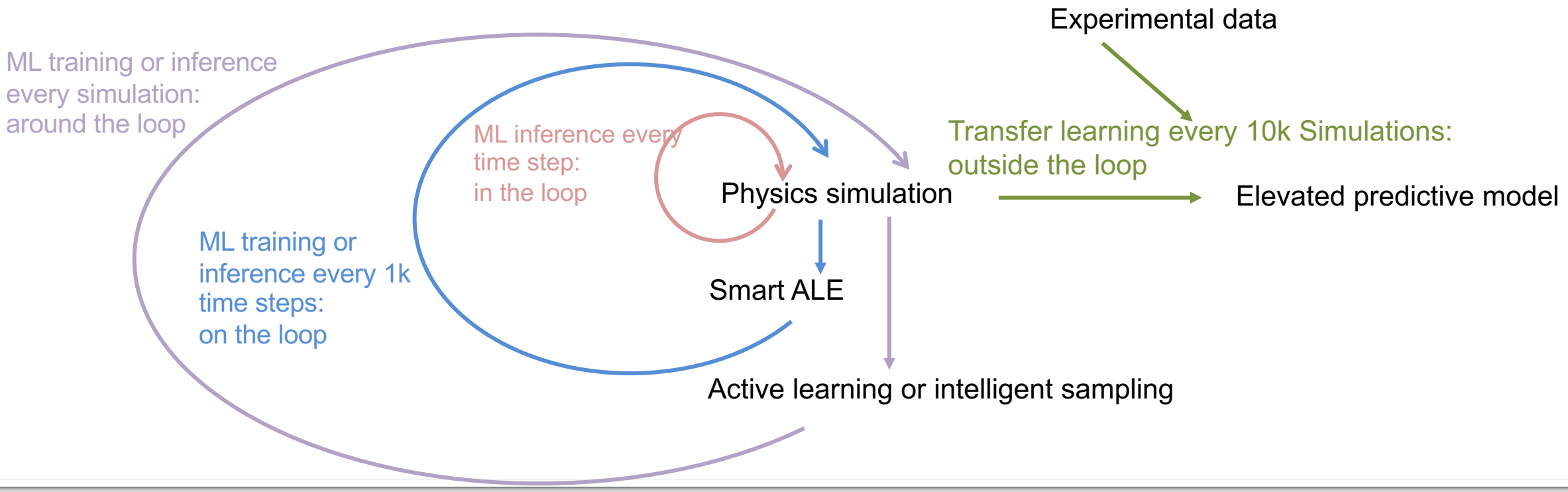
- Exascale will enable us to solve problems we could not have tackled before, however:
- We will continue to have mission drivers needing more compute
- Technology is driving us to find other solutions to address this insatiable demand
- Machine learning is one promising approach to enable us to do more with our current machines



Complex workflows involving ML are being applied to cancer, COVID-19 and NIF simulations today and being investigated and started to be picked up by all labs for ASC workloads.

Cognitive simulation is one approach to continued advancement: Wrap simulation in multiple layers of ML inference and training

- High-precision scientific **simulation**
- Frequent ML **training**
- Potentially very high-frequency **inference**



Exascale is not the large transition point that happened at pre-exascale systems, however, there are many opportunities to innovation

- Many of the challenges people identified a decade ago have held true e.g.:
  - Memory bandwidth challenges
  - Lots of threads
- Some new challenges have been identified e.g.:
  - Performance portability
  - GPU specific challenges
- Some challenges did not pan out:
  - Application resilience and power control beyond checkpoint restart.
- Complex workflows and their integration with ML models will continue to grow in importance.

I'm here to help make sure we share the lessons we have learned getting ready for Sierra and help connect you to key ASC tools that might help you run well on Lassen and uEICap

My official role in PSAAP

- Understand your computing needs and concerns and help plan machine purchases to get you the cycles you need
  - Leverage other procurements
  - Budget constrained
  - Work with CRT to understand usage
- Work with CRT to provide trainings/seminar series
- I will leverage other responsibilities at LLNL including:
  - Benchmarking lead for our procurements
  - El Cap COE deputy

I'm here to help make sure we share the lessons we have learned getting ready for Sierra and help connect you to key ASC tools that might help you run well on Lassen and uEICap

Lawrence Livermore
National Laboratory