SAND2015-8719C

Parallel scaling analysis for ALEGRA on the Excalibur system

John Niederhaus and Richard Drake
Sandia National Laboratories

DoD/SNL Workshop – Poster Session
Thursday, October 15, 2015

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

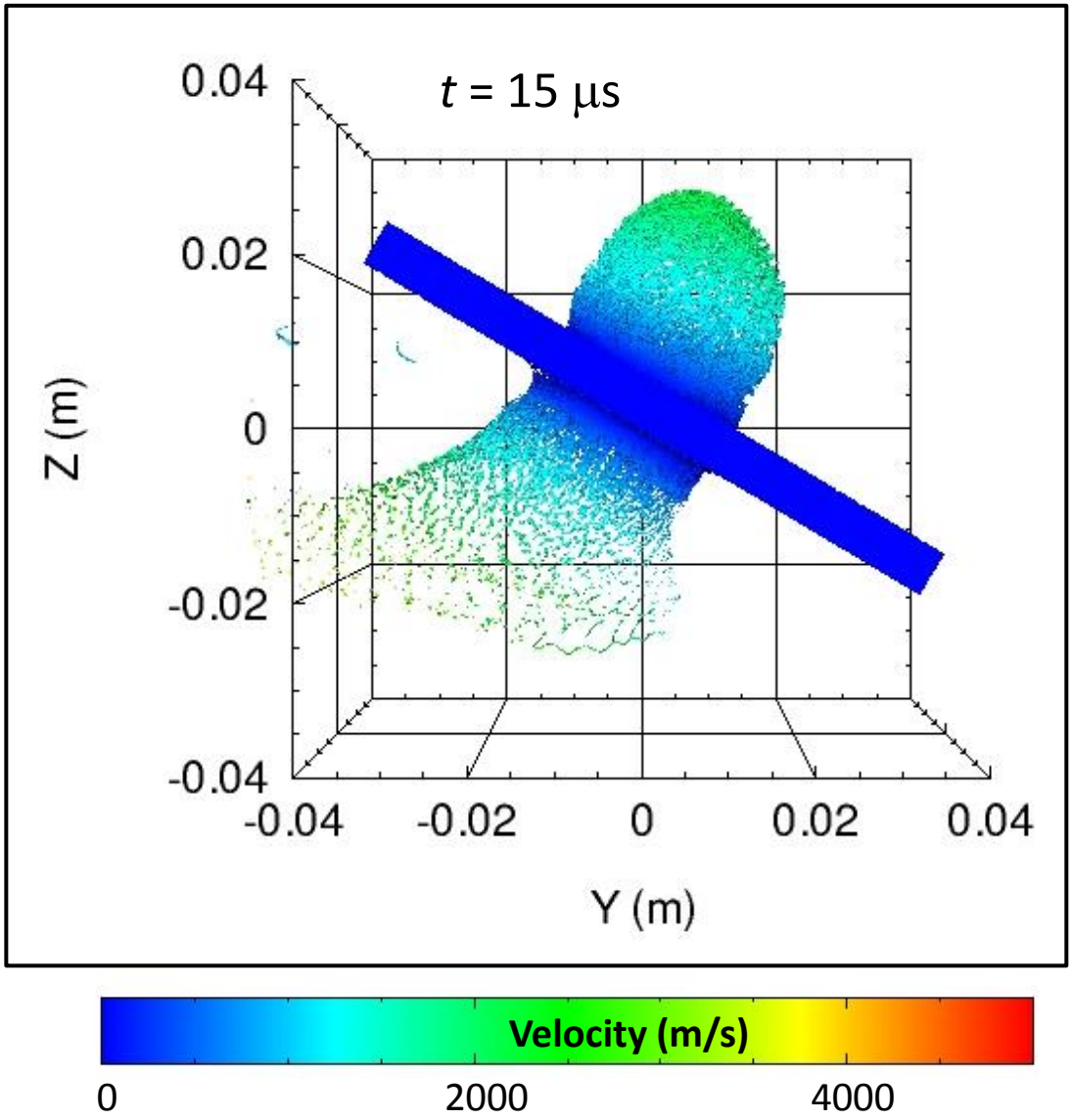
U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

SAND2015-xxxx PE

The 3D oblique impact problem

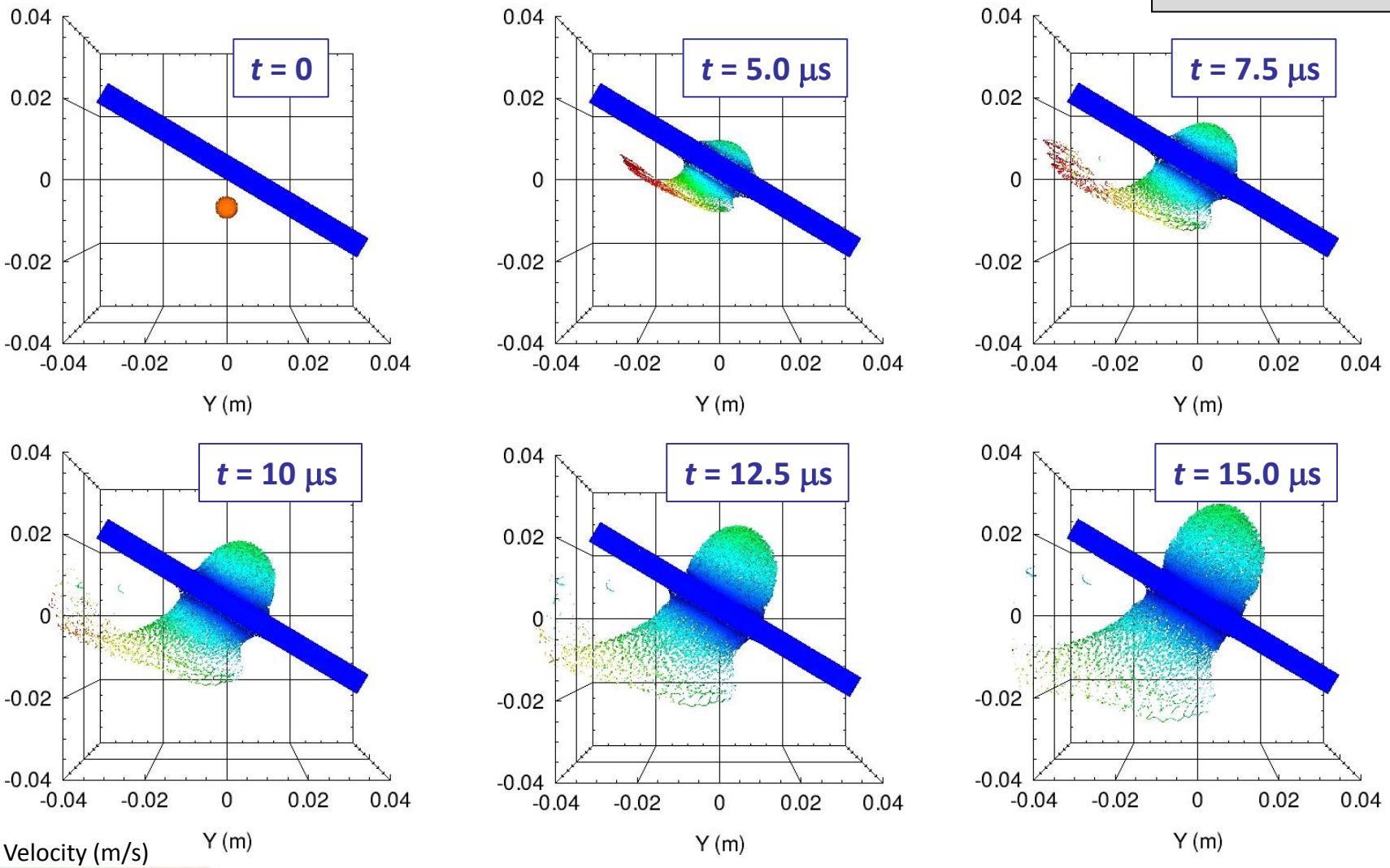
- Hypervelocity impact:
  - Cu sphere vs steel plate.
  - 30.8 degrees obliquity.
  - 4.52 km/s impact speed.
- Experimental basis: Grady and Kipp, *Int. J. Impact Engng*, 1994.
- Weak scaling study:
  - Spatial mesh refinement.
  - Increase core count to maintain ~10K elem/core.



Time evolution of the simulation

Sandia National Laboratories

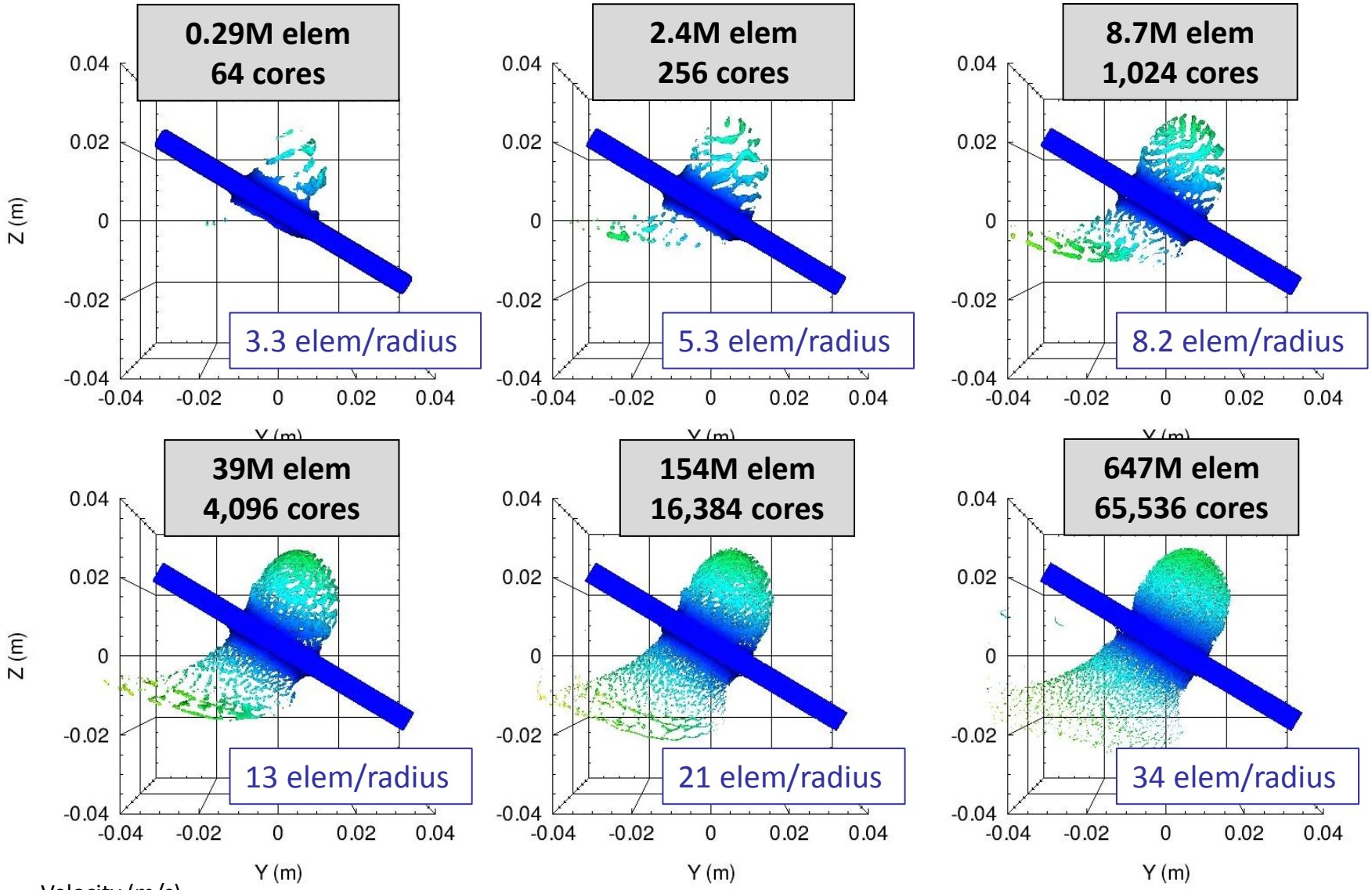
647M elem
65,536 cores



Maximum memory use per core in this calculation: 2.8 GB.

Results from excalibur

Effect of mesh refinement



t = 15 μs

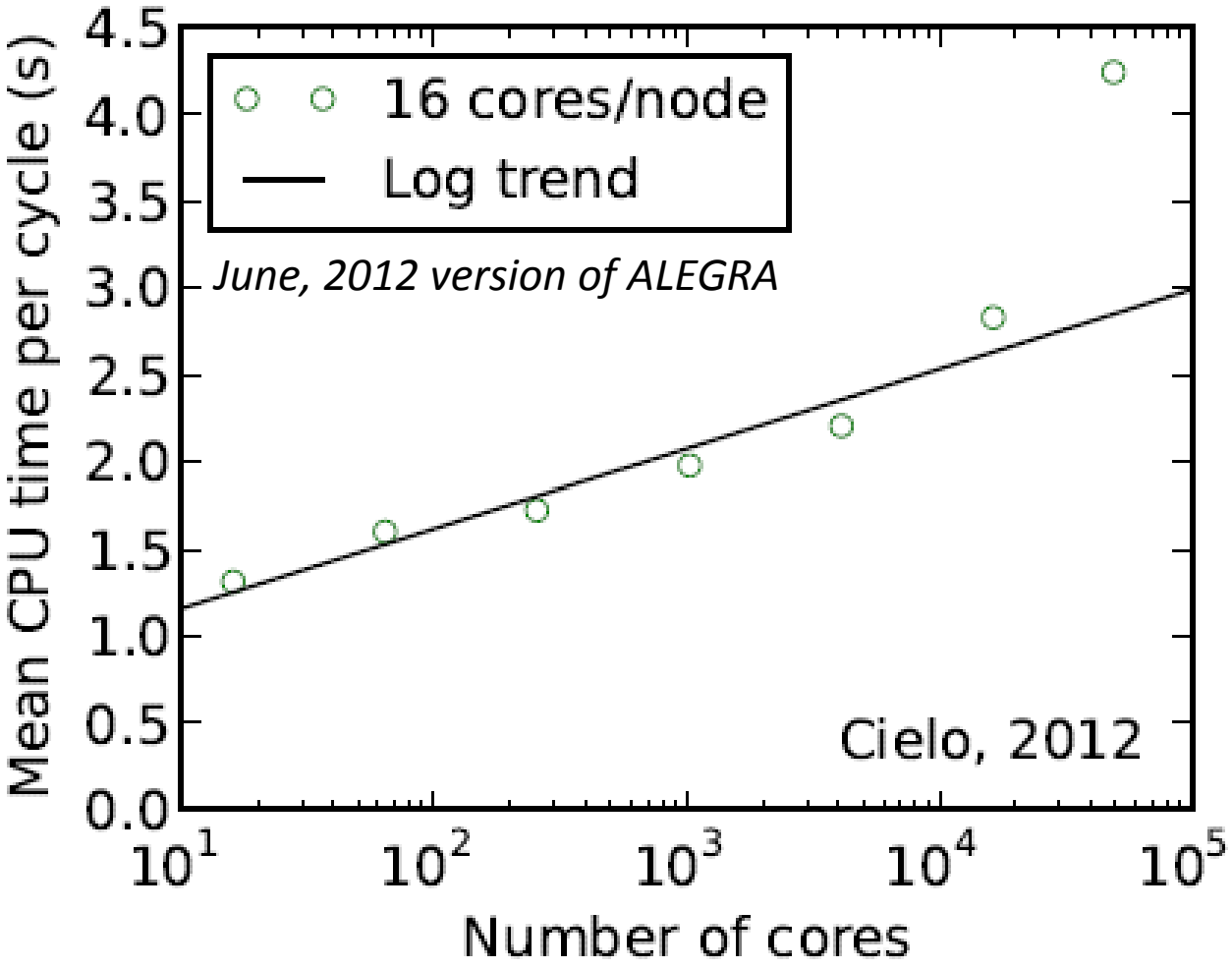
Results from excalibur

Comparison of two Cray systems

	Cielo	*Excalibur*
System	Cray XE6	Cray XC40
Online date	November, 2011	April, 2015
Core type	AMD Opteron	Intel Xeon E5
Core speed	2.4 GHz	2.3 GHz
Cores per node	16	32
Memory type	DDR3-1333 MHz	DDR4-2133 MHz
Memory per node	32 GB	128 GB
Total compute cores	143,104	99,136
Interconnect type	Gemini 3D Torus	Aries/Dragonfly
Physical location (owner):	*LANL (DOE)*	*ARL (DoD)*

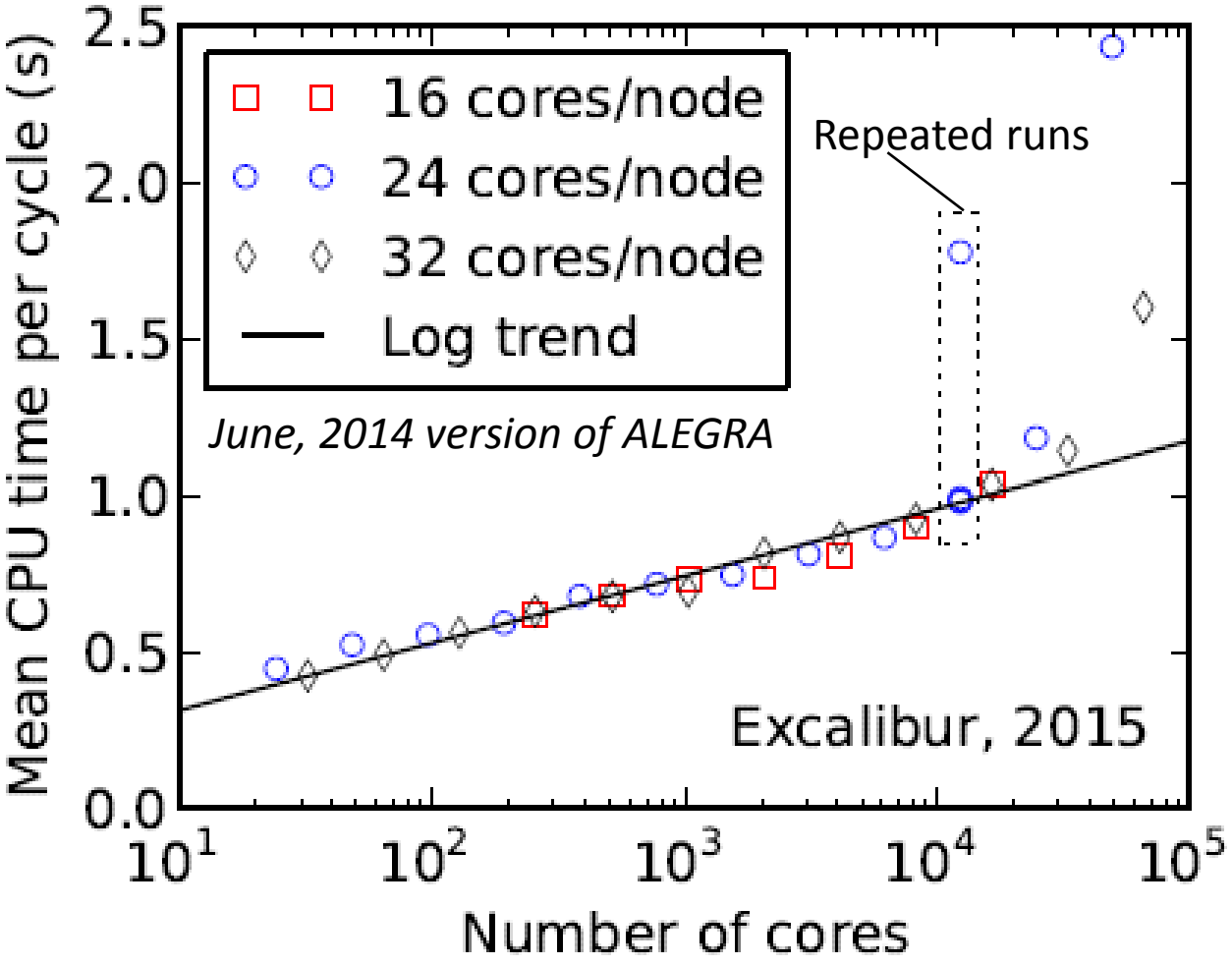
These turned out to be the critical differences.

Performance on cielo, 2012



- Full complement of cores per node: 16.
- Log trend slope: 0.199.
- At 1K cores: ~2.0 s/cycle.
- Excursion begins at 16K cores.

Performance on excalibur, 2015



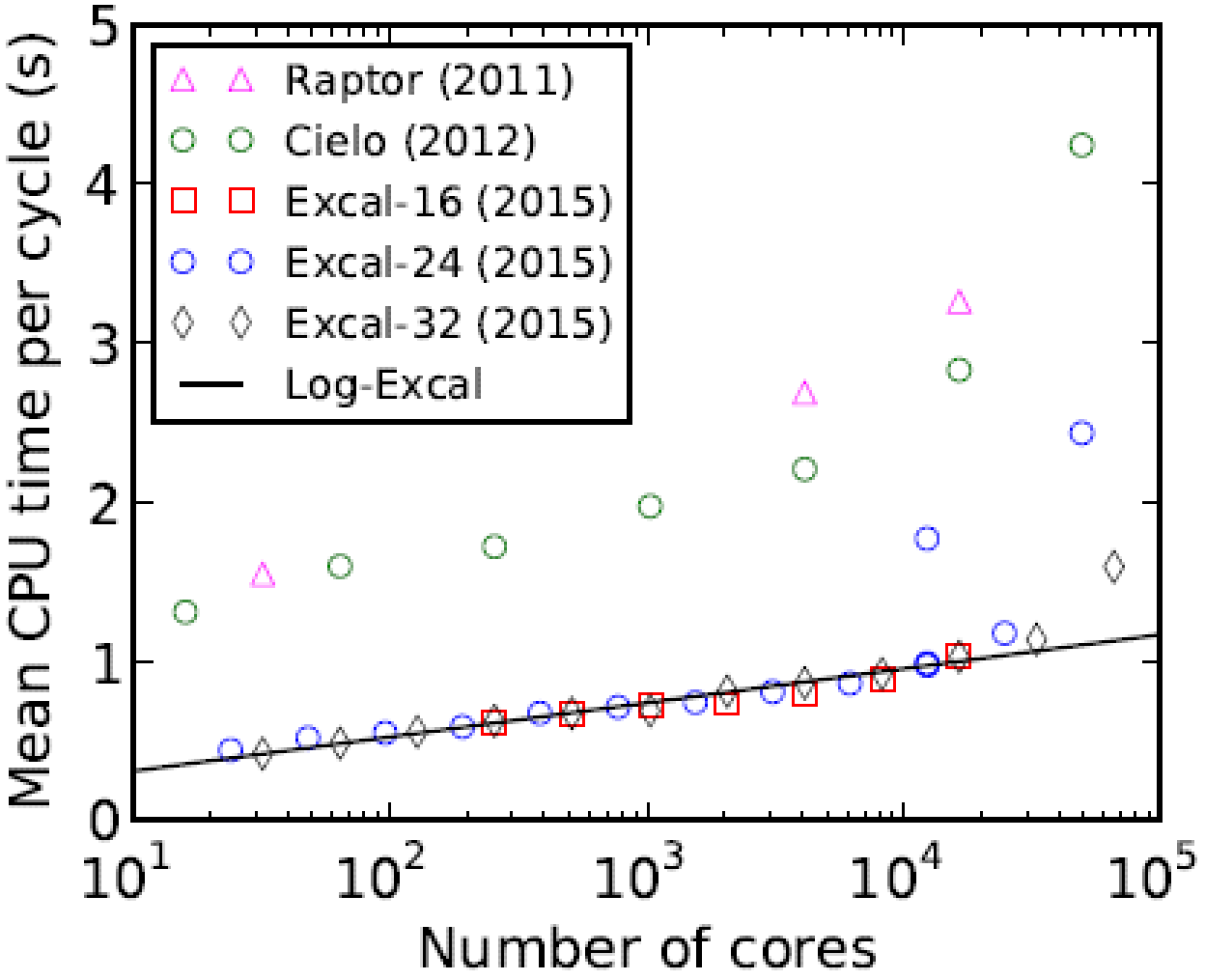
- Variable use of cores per node: 16, 24, 32.
- Log trend slope: 0.093.
- At 1K cores: ~0.75 s/cycle
- Excursion begins at 16K cores.
- Bad nodes?

Observations on scaling studies

Sandia National Laboratories

- Simulations are a direct representation of users' experience.
  - Realistic production problem with a realistic frequency of output is used.
  - Timestep is variable based on CFL condition.
- For core counts < 10,000:
  - Reliable logarithmic scaling is observed.
  - This is expected for parallel calculations where communication is needed and "all-reduce" types of operations are involved.
- For core counts > 10,000:
  - Performance becomes much more variable from one run to the next.
  - Significant excursions (50-100%) in performance trend are apparent.
  - Decreasing the cores per node does not eliminate the issue.

Performance comparison



Raptor: Cray XE6, 16 cores/node, AFRL DSRC

Comparing 2012 and 2015 results

Sandia National Laboratories

- Significant performance improvements are observed on excalibur for N < 10,000 cores:
  - Factor of 2 improvement in scaling trend slope.
  - Factor of 3 improvement in overall throughput (CPU time/cycle).
  - Most likely attributable to improvements in memory configuration/speed.
  - Massive refactor of ALEGRA data layout in 2013-2014 also helpful!
- Speculative explanation for excursions on both machines:
  - Poor node allocation, resulting in sub-optimal communication topology. "Bursty" use of interconnect. Poor load balancing.
  - Unpredictable hardware faults – effects magnified at large core counts.
  - Known: excursions not correlated with physical features seen in solutions.

Conclusions

- Weak scaling with a 3D hypervelocity impact problem is a useful tool, despite variable timestep.
- Major performance improvement (roughly 3x) is observed on excalibur relative to previous generation of Cray systems and previous versions of ALEGRA.
- Performance variability and degradation is seen for N > 10K cores.
- ALEGRA performance for multiphysics problems is not evaluated.

Acknowledgments

Compute time on Cielo was provided by DOE's ASC Cielo Capability Computing Campaign (CCC-5) under the administration of Joel Stevenson (SNL-9326).

Compute time on Excalibur was provided by DoD's ARL DSRC during early-access testing time and subsequent dedicated time, administered by Thomas Kendall and Phillip Matthews (ARL).

Early versions of the ALEGRA simulations and test configuration were developed with the help of David Hensinger (SNL-1555).

Thanks also to Simon Hammond (1422), Ryan Grant (1423) and Paul Lin (1426) for helpful discussions.