SAND2018-5894C

Exceptional service in the national interest

Sandia National Laboratories

Adapting Multiscale Solid Mechanics Codes for Next-Generation Computing Platforms

David Littlewood

Brian Lester

Michael Tupek

June 7th, 2018

U.S. DEPARTMENT OF ENERGY

National Nuclear Security Administration

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and 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.

The Supercomputing Landscape is Changing

Sandia National Laboratories

- What is the current definition of a “next generation” supercomputer?
  - On-node accelerators
    - Intel Knights Landing (KNL), NVidia GPU, etc.
  - Enables increased parallelism, e.g., MPI + X
  - Flops are cheap, memory management is critical
- DOE/NNSA Advanced Technology Systems
  - Trinity ATS-1, LANL
    - Intel Xeon (Haswell) & Intel Xeon Phi (KNL)
  - Sierra ATS-2, LLNL
    - IBM POWER9 CPUs & NVidia GPUs
  - Future ATS platforms …

ATS-1 Trinity
http://www.lanl.gov/projects/trinity/

ATS-2 Sierra
https://asc.llnl.gov

National Nuclear Security Administration

Advanced Simulation and Computing

Predicting, with confidence, the behavior of nuclear weapons through comprehensive, science-based simulations.

What is the Impact on Theoretical and Applied Mechanics?

Sandia National Laboratories

Significant opportunity and significant risk

- *Opportunity*: Run existing codes faster and on larger meshes
- *Opportunity*: New modeling approaches that were previously intractable
- *Risk*: Existing codes may not run well (or at all) on next-gen platforms

Do Next-Gen Platforms (NGP) Affect Constitutive Model Developers?

- YES, if you want to run massively parallel simulations on the most powerful computing platforms
- MAYBE, if current trends in on-node accelerators trickle down to the consumer market (modest clusters, workstations, laptops …)

What is the Impact on Theoretical and Applied Mechanics?

Sandia National Laboratories

Main concerns in moving to next-gen hardware

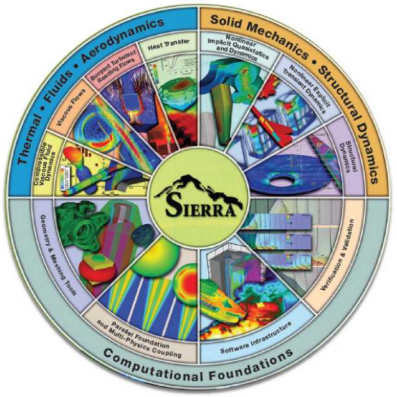
- Increased parallelism
  - `parallel_for` construct for simultaneous evaluation of the constitutive model on a large number of material points
  - Thread safety, compatibility with `parallel_for` programming models
- Data structures that are compatible with on-node accelerators
  - Memory access patterns strongly affect performance
  - For GPUs, data must be copied between host memory and device memory
- Increased software complexity
  - Greater reliance on computer science / software engineering subject matter experts

Ongoing Work in Sandia's Simulation Codes

Sandia National Laboratories

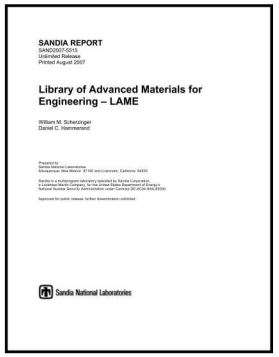
Sierra/SolidMechanics Analysis Code

- DOE/NNSA Advanced Simulation and Computing (ASC) Code
- Core capabilities
  - Explicit and implicit dynamics, implicit quasi-statics
  - Contact
  - Failure modeling



Library of Advanced Materials for Engineering (LAME)

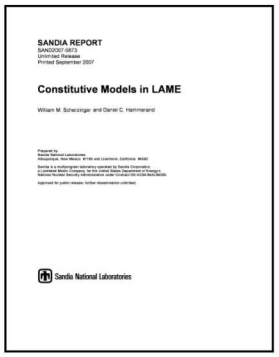
- 80+ material models
- Material models for solid elements, shells, cohesive zones, ...
- Legacy code issues (C++, Fortran)


SANDIA REPORT
SAND2007-5015
Unlimited Release
Printed August 2007

Library of Advanced Materials for Engineering – LAME

William M. Scherzinger
Daniel C. Hammerand

Sandia National Laboratories


SANDIA REPORT
SAND2007-5673
Unlimited Release
Printed September 2007

Constitutive Models in LAME

William M. Scherzinger and Daniel C. Hammerand

Sandia National Laboratories

National Nuclear Security Administration

ASC™

Advanced Simulation and Computing

Predicting, with confidence, the behavior of nuclear weapons through comprehensive, science-based simulations.

Adapting Constitutive Models for Next-Gen Hardware

Sandia National Laboratories

Design considerations

- Hide as much software engineering complexity as possible
- Strive for compatibility with existing material model API
- Strive for future-proof design

Software engineering strategy

- Store material properties and state data in user-friendly, flexible data containers that are amenable to both CPU and GPU architectures
  - FullTensor `def_grad(K_F_XX)`
  - SymTensor `stress(K_S_XX)`
- Data containers are treated by material model developers as simple arrays, but are in fact views into hardware-specific storage classes
- Move parallelism (loop over material points) to a high-level interface
  - Material models implement `get_stress()` for a single material point
  - Material model is evaluated for many material points using a `parallel_for` construct that is hidden from material model developers

Sandia National Laboratories

Are there any possible complications?

- Fortran will not be well supported
  - Convert to C++
- Material model code must be thread safe
- Degree to which performance is improved is strongly dependent on the algorithm and source code for a given material model
- Material models that use function pointers are a challenge
- Good performance on GPUs in particular requires a large workset such that the GPU is fully saturated with work
- End-to-end code performance requires that analysis code as a whole be full compatibility with next-generation hardware
  - Modified material model is necessary but not sufficient for improved performance

Preliminary Results

- Definition of NGP material model interface
- Creation of NGP data containers
- NGP versions of Neo-Hookean and J2-plasticity with linear hardening
- Unit tests that exercise material models on conventional and next-gen systems
  - Call material models on a large number of material points
  - Assess correctness and performance
- Initial performance benchmarking on select hardware platforms
  - Intel Haswell
  - Intel Knights Landing (KNL)
  - Intel Broadwell
  - NVidia Kepler GPU

Sandia National Laboratories

Operated for the U.S. Department of Energy by National Technology and Engineering Solutions of Sandia, LLC.

Albuquerque, New Mexico 87185

date: October 12, 2017

to: Distribution

from: David Littlewood, Org. 1444, MS-1322
Michael Tupek, Org. 1542, MS-0845

subject: Adapting Material Models for Improved Performance on Next-Generation Hardware

1 Overview

Sandia researchers in computational solid mechanics and high-performance computing have successfully modified an initial set of material models in the Library of Advanced Materials for Engineering (LAME) [3, 2] for improved performance of the *Sierra/SolidMechanics* [4] analysis code on next-generation computing platforms. This is critical because, in many cases, evaluation of the material model constitutes a major fraction of the overall run time of a mechanics simulation. Elements of the improved performance demonstrated here are utilization of the *Kokkos* library for performance portability across disparate hardware architectures and the use of Single Instruction, Multiple Data (SIMD) vectorization. Successful use of these extensions within LAME required two major changes to the material model code: 1) material properties and state data must be stored in data containers with layouts that are amenable to disparate compute architectures, and 2) parallelism for the evaluation of multiple material points must be moved from the material model itself to a higher-level software interface. Separating out the implementation of the material model from the particulars of both the data layout and the loop parallelism strategy results in immediate performance benefits across different compute architectures. Initial testing of constitutive models modified for performance on Next-Generation Platform (NGP) architectures shows significant performance gains on Intel's Knights Landing (KNL) and Nvidia GPUs relative to more traditional multi-core x86 hardware.

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and 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.

Unlimited Release SAND2017-11100 O.

[Memorandum SAND2017-11100 O]

Performance Results for the Neo-Hookean Model

Sandia National Laboratories

- Full utilization of a single compute node on conventional and next-gen hardware platforms
- `get_stress()` called on ~1M material points divided into 2K worksets
- Speed-up is given relative to serial execution on Intel Haswell architecture

Table 1: Performance comparison for a single node at full utilization: Speed-up of the neo-Hookean model relative to serial Haswell.

Configuration	Platform	Speed-up
Haswell: 32 cores	ascic	25.1
Haswell: 32 cores + 4-wide SIMD	ascic	88.5
Broadwell: 32 cores	ascic	40.0
Broadwell: 32 cores + 4-wide SIMD	ascic	124
KNL: 64 cores + 4x hyperthreads	mutrino	42.7
KNL: 64 cores + 2x hyperthreads[1] + 8-wide SIMD	mutrino	177
Kepler: Nvidia GPU	ascicgpu	162

Performance Results for the J2-Plasticity Model

Sandia National Laboratories

- Full utilization of a single compute node on conventional and next-gen hardware platforms
- `get_stress()` called on ~1M material points divided into 2K worksets
- Speed-up is given relative to serial execution on Intel Haswell architecture

Table 2: Performance comparison for a single node at full utilization: Speed-up of the J2-plasticity model relative to serial Haswell.

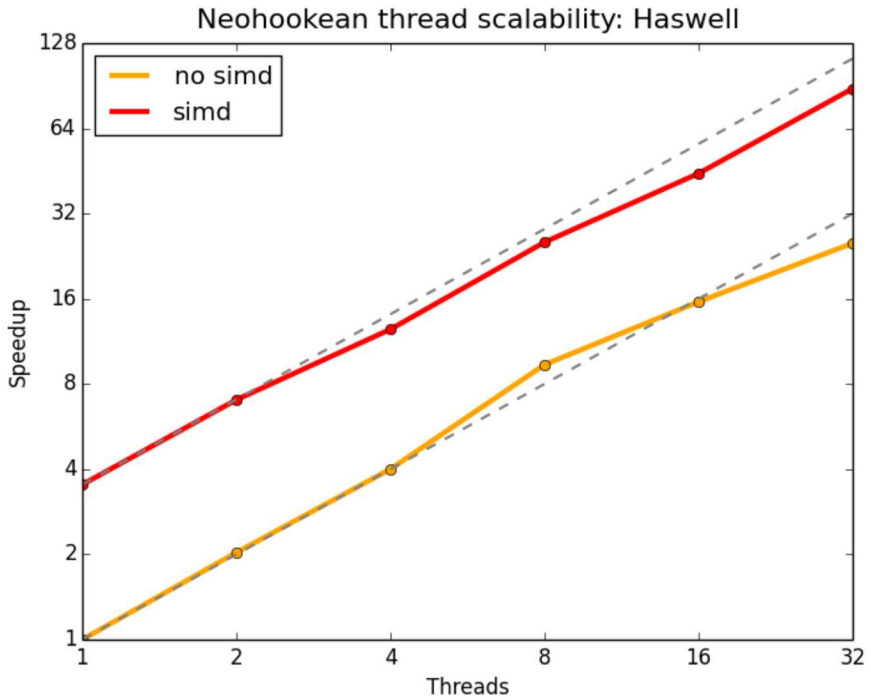
Configuration	Platform	Speed-up
Haswell: 32 cores	ascic	24.4
Haswell: 32 cores + 4-wide SIMD	ascic	42.2
Broadwell: 32 cores	ascic	34.9
Broadwell: 32 cores + 4-wide SIMD	ascic	53.8
KNL: 64 cores + 4x hyperthreads	mutrino	32.2
KNL: 64 cores + 4x hyperthreads + 8-wide SIMD	mutrino	76.4
Kepler: Nvidia GPU	ascicgpu	71.3

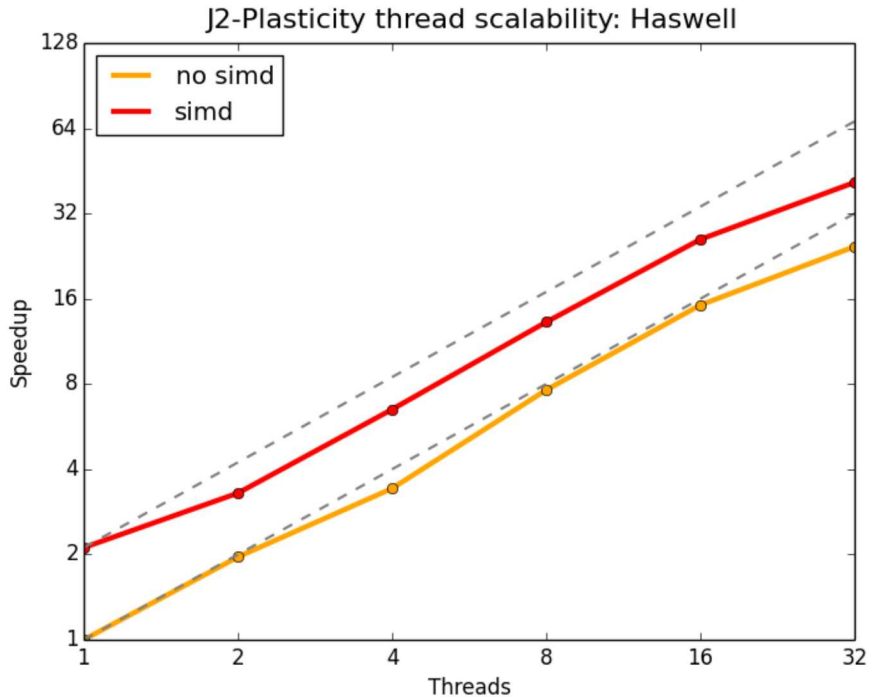
Performance Results: Thread scalability on Intel Haswell

Sandia National Laboratories

Unit test results for Neo-Hookean and J2-plasticity models

- `get_stress()` called on ~1M material points divided into 2K worksets
- Speed-up is given relative to serial execution on Intel Haswell architecture
- Results show that NGP material models scale well on traditional hardware
- Results demonstrate effectiveness of SIMD vectorization



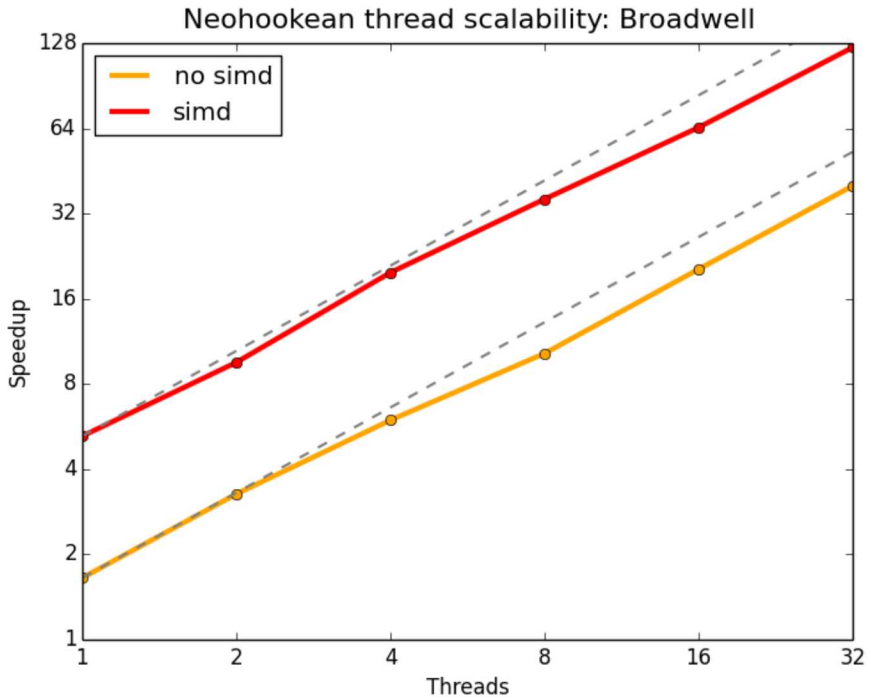


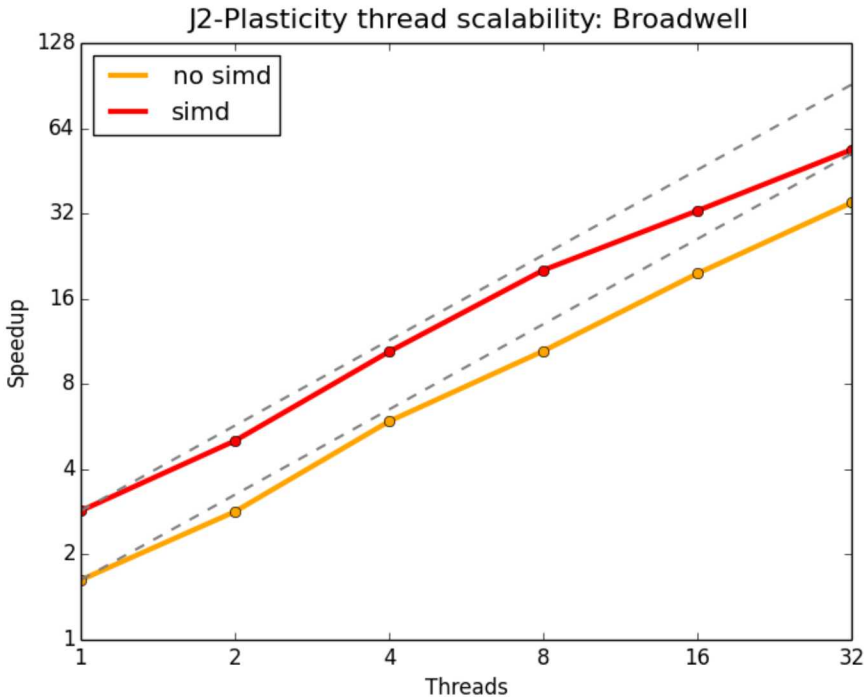
Performance Results: Thread scalability on Intel Broadwell

Sandia National Laboratories

Unit test results for Neo-Hookean and J2-plasticity models

- `get_stress()` called on ~1M material points divided into 2K worksets
- Speed-up is given relative to serial execution on Intel Haswell architecture
- Results show that NGP material models scale well on traditional hardware
- Results demonstrate effectiveness of SIMD vectorization



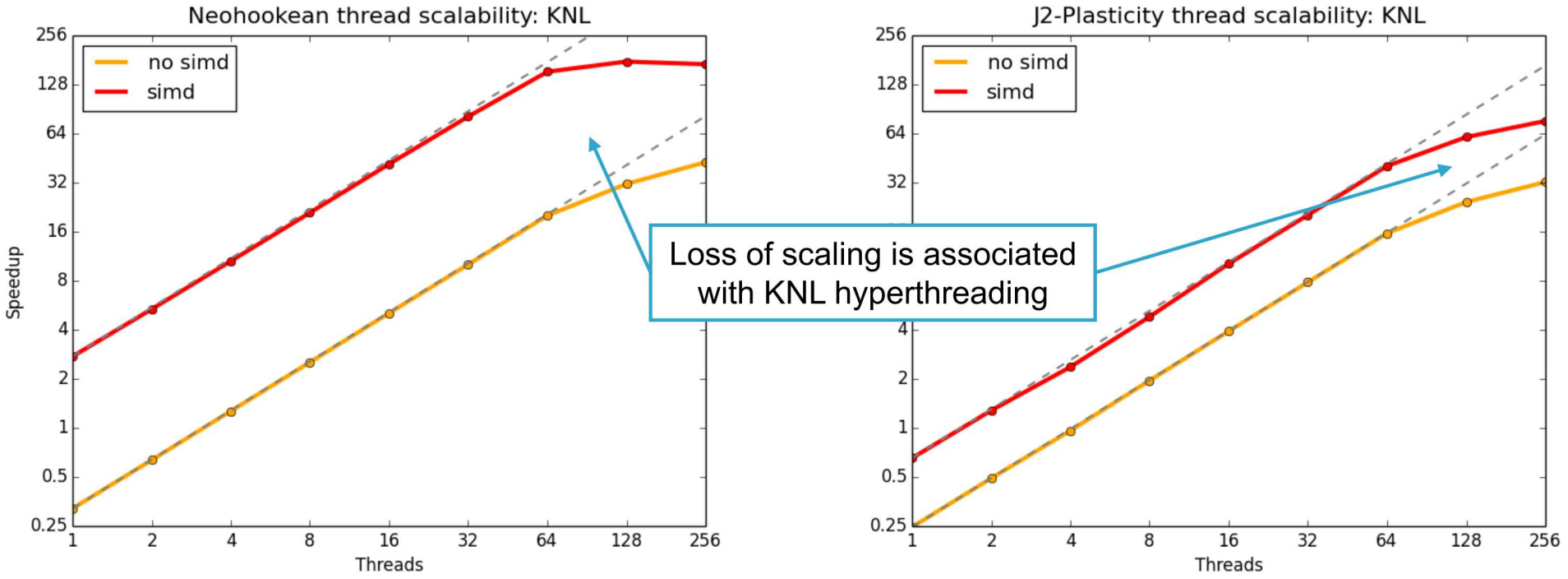


Performance Results: Thread scalability on Intel Knights Landing

Sandia National Laboratories

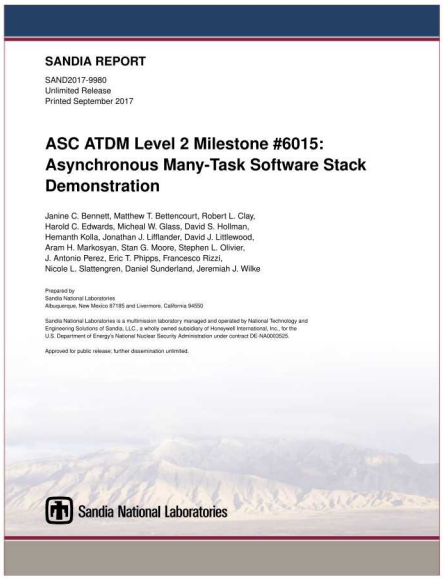
Unit test results for Neo-Hookean and J2-plasticity models

- `get_stress()` called on ~1M material points divided into 2K worksets
- Speed-up is given relative to serial execution on Intel Haswell architecture
- Test executed on 1 KNL CPU with 64 cores, 8-wide SIMD, and up to 4 hyperthreads per core



Moving from Single Scale to Multiscale

- Utilize a suite of software tools to take full advantage of next-generation computing platforms
  - Kokkos for performance portability
  - DARMA for asynchronous many-task scheduling
  - Qthreads for high-performance multi-threading
  - MPI + X via standard MPI, Kokkos, Qthreads
- Vision: provide stem-to-stern Kokkos compatibility
  - Enable bona fide performance evaluation of NGP material models
- Introduce contact algorithm
  - Exercises critical load balance features
  - Highly relevant to Sandia missions


SANDIA REPORT

SAND2017-9980
Unlimited Release
Printed September 2017

ASC ATDM Level 2 Milestone #6015: Asynchronous Many-Task Software Stack Demonstration

Janine C. Bennett, Matthew T. Bettencourt, Robert L. Clay, Harold C. Edwards, Micheal W. Glass, David S. Hollman, Hemanth Kolla, Jonathan J. Lifflander, David J. Littlewood, Aram H. Markosyan, Stan G. Moore, Stephen L. Olivier, J. Antonio Perez, Eric T. Phipps, Francesco Rizzi, Nicole L. Slattengren, Daniel Sunderland, Jeremiah J. Wilke

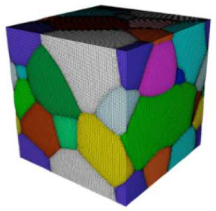
Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and 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.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

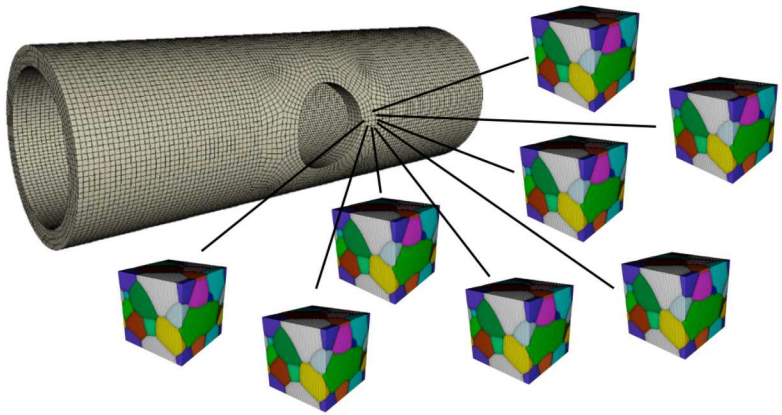
NimbleSM



Moving from Single Scale to Multiscale

- Multiscale modeling approaches that were previously intractable may become feasible on NGP platforms
- Multiscale techniques that do not require communication across compute nodes are well suited for on-node accelerators (e.g., GPU)
- Ongoing work focuses on FE^2
  - RVE solves are carried out on a single CPU + accelerator(s)
  - Amenable to asynchronous execution and load balancing

FE^2 multiscale



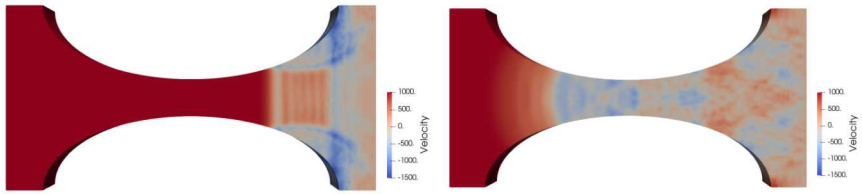
Independent RVE sub-models are associated with individual integration points in the macroscale mesh

Asynchronous Execution and Load Balancing for Multiscale

Sandia National Laboratories

Elastic wave propagation

Good load balance

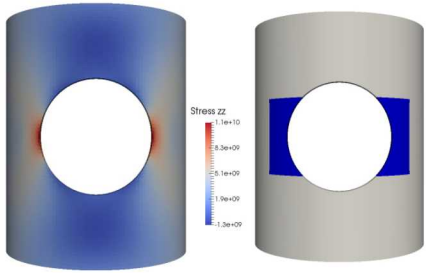


time = 5.0e-6 s

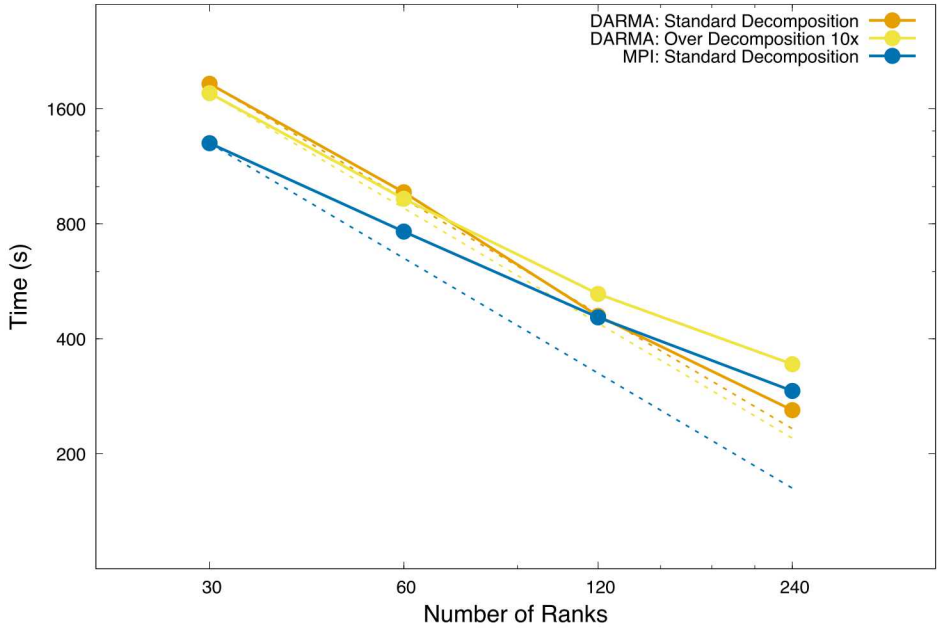
time = 15.0e-6 s

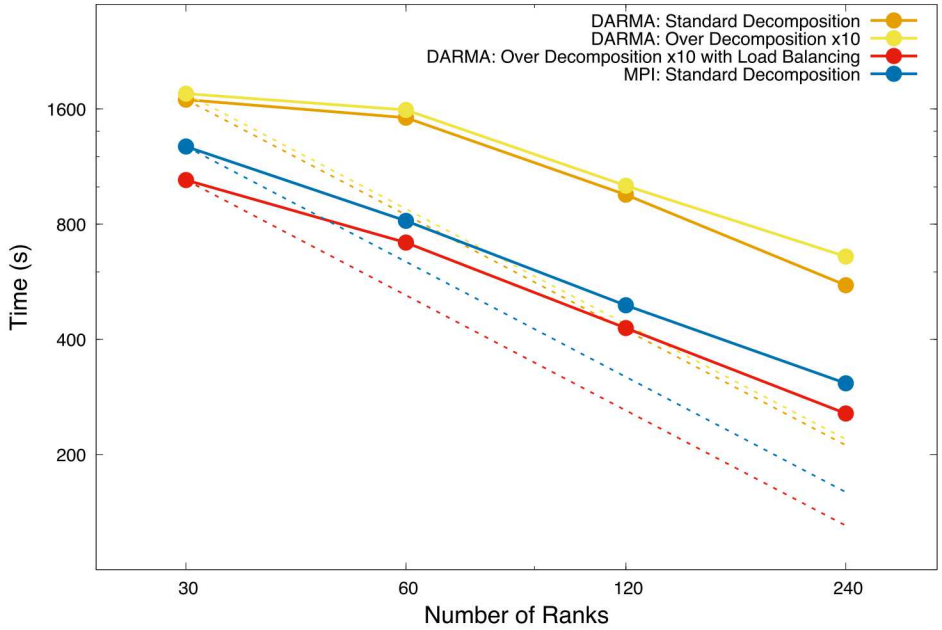
Tube with hole under tension

Poor load balance



High-fidelity model applied only in high-stress regions



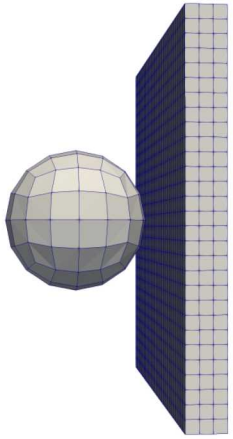


Adapting Contact Algorithms for Next-Gen Hardware

Sandia National Laboratories

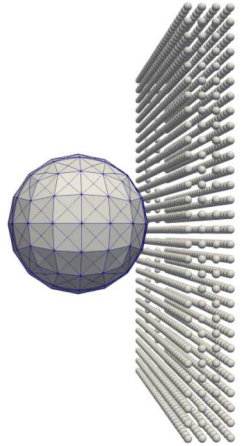
- Vetting of Kokkos-based search algorithms in tech demonstrator
  - Kokkos-based proximity search algorithms developed in Sierra/SM
- Morton Bounding Box unit test
  - GPU = 0.01728 sec
  - Single CPU thread = 0.16111 sec
  - 16 CPU core = 0.02039 sec
- Closest Point Projection unit test
  - GPU = 0.036942 sec
  - Singe CPU thread = 1.05114 sec

Finite element mesh

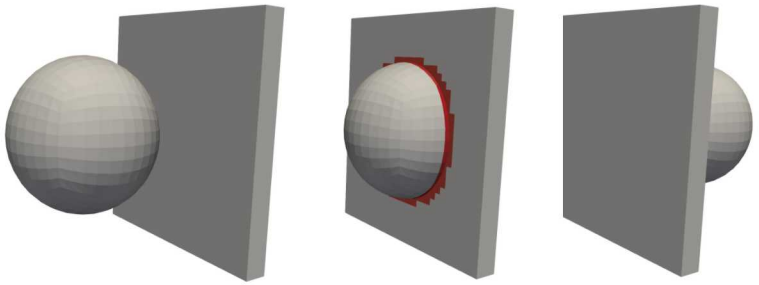


3D hexahedral elements

Contact submodel



Surface patches and nodes



Algorithm for detecting interpenetration

Sandia National Laboratories

Questions?

David Littlewood
djlittl@sandia.gov