Tensor-Optimized Hardware Accelerates Fused Discontinuous Galerkin Simulations
Alexander Breuer, Alexander Heinecke (Intel), Yifeng Cui

The International Conference for High Performance Computing, Networking, Storage, and Analysis
Dallas, Texas
November 11-16, 2018

Extreme-scale Discontinuous Galerkin Environment (EDGE)

- Focus: Problem settings with high geometric complexity, e.g., mountain topography
- Unique support for fused simulations exploiting inter-solver parallelism
- Rapid prototyping through support for: Line elements, quad triangles, hexes, etc.
- Parallelization: Assembly kernels for AVX, AVX2, AVX512 and AVX512_4FMA extensions, utilizing all 86 CPUs of the last five years optimally
- OpenMP-MPI (custom and overlapping)
- World record seismic wave propagation: 10.4 PPFLOPs on Cori [1]
- Continuity: Continuous Integration (sanity checks), Continuous Delivery (automated convergence + benchmarks runs), code coverage, license checks, container bootstrap
- License: BSD 3-Clause (software), CCO for supporting files, e.g., user guide

Fused Simulations Exploit Inter-Solver Parallelism

- "Similar simulations?"
- EDGE imposes restrictions on fused simulations: 1) Identical mesh for all fused simulations 2) Identical simulation parameters: start and end time, convergence rate, frequency of wave field output, location and orientation of seismic receivers 3) Identical material parameters (velocity model)
- 4) Sources mostly arbitrary: Arbitrary initial DOS, seismic sources: arbitrary location and moment rates, spontaneous rupture: identical friction law, other initial parameters arbitrary

Contribution #1: Verification of 32bit Precision - Beyond Artificial Convergence Benchmarks

- Benchmarking is key to assessing the accuracy of seismic wave propagation solvers
- EDGE has a multitude of modeling parameters:
  1) Fused vs. non-fused simulations
  2) Single vs. double precision
  3) Convergence rate in space and time
  4) Feature- and velocity-aware mesh refinement
  5) Source discretization
  6) Topography: Flat vs. DEM-derived
  7) Velocity models: Layered vs. data-input
- Choosing the right modeling parameters is crucial for best time-to-solution
- Our verification study considers the entire modeling and simulation space and covers essential modeling decisions (1-3) above for best practices

General Applicability and Availability

- Verification efforts transfer directly to DG-FEM solvers, relying on unstructured tetrahedral meshes, e.g., SeisSol or DSGCack
- LBXSMM optimizations transfer easily to community codes, using the library: e.g., CFP, NReXs, NReXso0 or SeisSol
- Parallelization through fused simulations / multiple right hand sides is a novel technique for improved efficiency. Approach is applicable to other software, facing challenging efficient SIMD utilization due to small matrix kernels, e.g., SpecFEM3D or N-ReXs5000
- Commercial libraries, e.g., Intel’s MKL, recently added compact vectorized routines, running patched BLAS routines. The underlying idea is similar, but would annihilate the presented speedup due to required element matrix duplication and zero padding. Suggests adoption on the library level

Artifact appendix of this poster contains details on the availability of all presented software (BDD-3) and data (CCO)

References and Support

Sumatra megathrust earthquake - C. Upho
[2] This research used resources of the National Energy Research Scientific Computing
[3] DB/SP
[4] This work was supported by the Department of Energy under award number DE-AC02-05CH11231, and the Stock CARMO project (45716). SDSC is a DOE Office of Science User Facility supported by the Office of Advanced Scientific Computing Research, and the National Science Foundation, under the auspices of the National Science Board. SDSC is a DOE Office of Science User Facility supported by the Office of Advanced Scientific Computing Research, and the National Science Foundation, under the auspices of the National Science Board.