FASTMath: Kokkos Kernels and Linear Solvers

Performance portable kernels and solvers for large-scale, scientific simulations

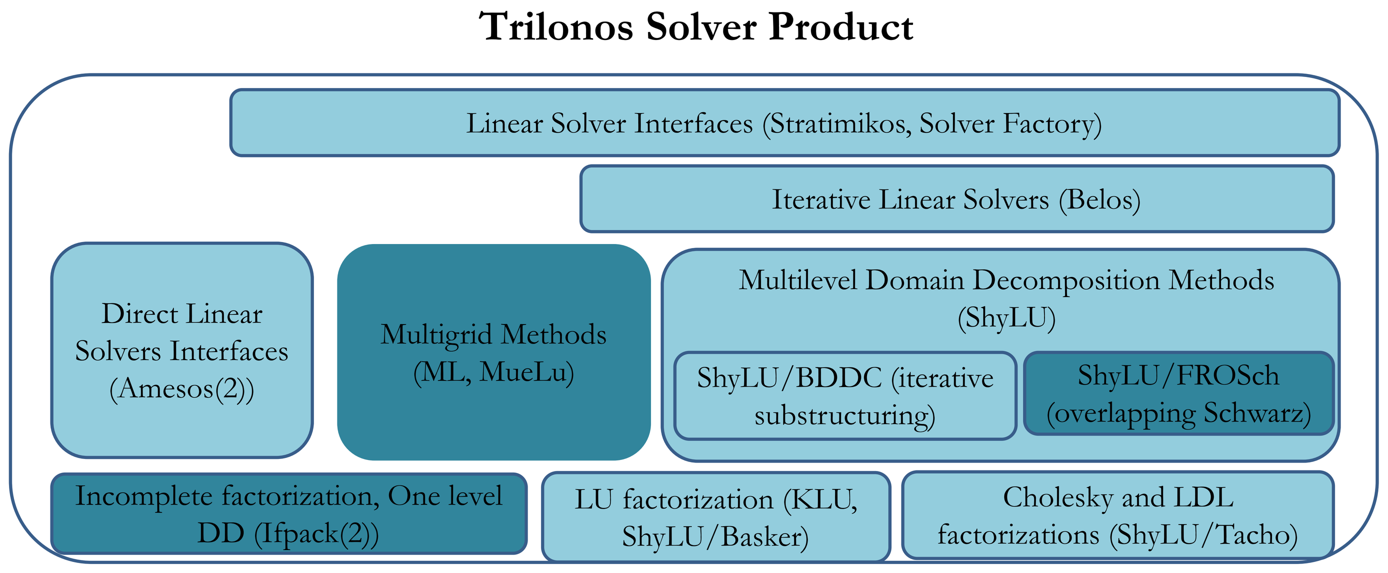
Ian Bogle, Karen Devine, Jonathan Hu, Kyungjoo Kim, Mauro Perego, Siva Rajamanickam, George Slota
Center for Computing Research, Sandia National Laboratories, NM, USA
Rensselaer Polytechnic Institute, Troy, New York

Sandia National Laboratories

Trilinos Scalable Solvers in FASTMath support SciDAC Applications

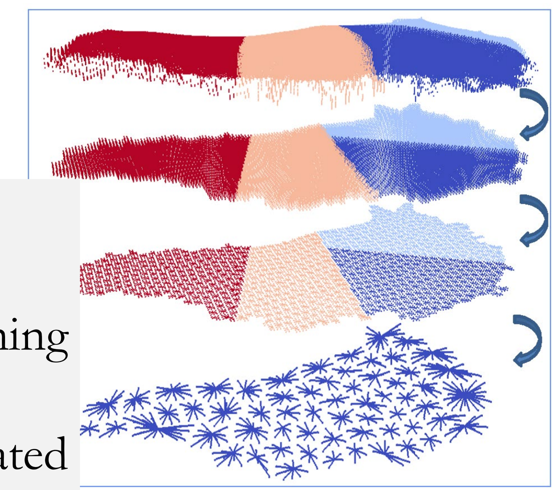
- Trilinos Solver Product delivers high-performance linear solvers to ASCR, SciDAC, NNSA and ECP customers,
- New solver stack is performance portable to several architectures CPUs, GPUs, and KNLs
- SciDAC Applications use several components of the Trilinos solvers
  - Algebraic multigrid (AMG), associated smoothers, coarse solvers
  - Domain Decomposition solvers, associated direct solvers and kernels

Trilonos Solver Product

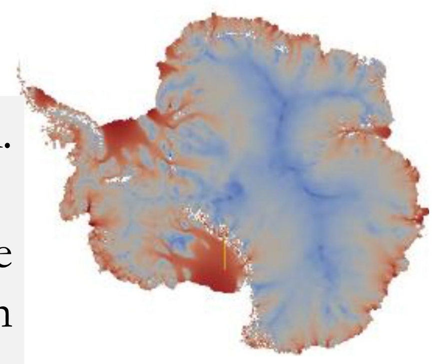


Trilinos provides comprehensive suite of solvers that cover entire spectrum of DOE application solver needs

Improving Multigrid Solver Performance for Multiphysics Simulations



- Starting new collaboration with SciDAC ProSPect project
  - Goal: develop performant line-solver kernels
  - Current line solvers are bottleneck in an existing semicoarsening algebraic multigrid (AMG) solver*
  - Impact: Enable ProSPect to transition to Trilinos 64-bit templated stack and leverage its performance portability.



- New thrust with J. Shadid (Tokamak Disruption Simulation project) and R. Tuminaro to help develop fast solvers for structured meshes
  - Leverages research from Tuminaro's multilevel ASCR project. Enable high-resolution 3D studies of high Z number neutral gas injection into Deuterium plasmas.
  - Impact: Ability to track greater number of ionic species at higher charge levels over longer time scales.

*Tuminaro, Perego, Tezaur, Salinger, Price, *A matrix dependent/algebraic multigrid approach for extruded meshes with applications to ice sheet modeling*, SISC 38(1), 2016

Batched Linear Algebra Kernels Support BES Applications

- **Batched Linear Algebra critical for performance of dense linear algebra operations on modern architectures**
- Community effort to develop a standard for batched BLAS and LAPACK operations
- Primary focus is on BLAS and LAPACK operations on lot of small matrices
- **Kokkos Kernels team part of the community effort for batched linear algebra standards**

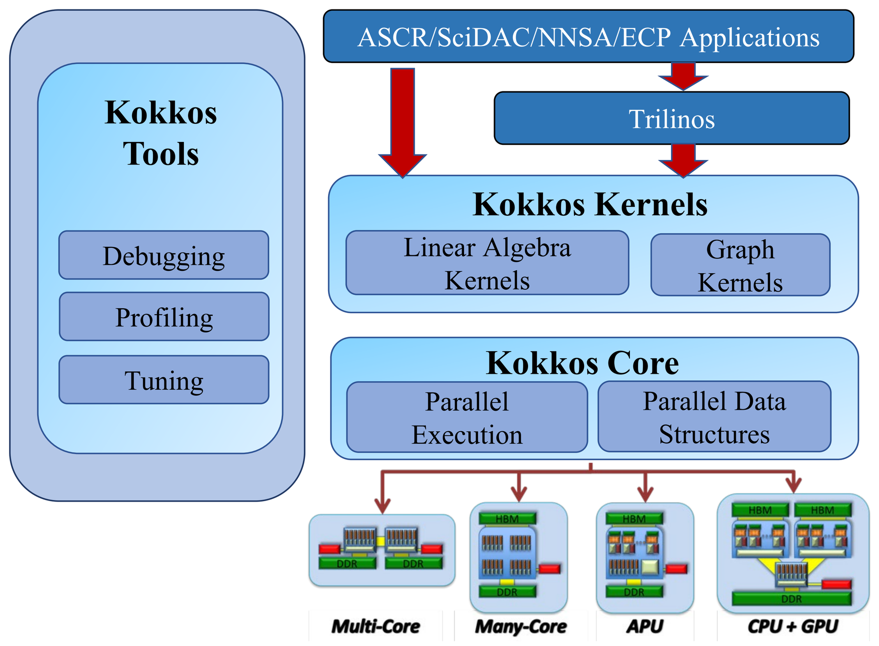
- Exascale Catalytic Chemistry project (BES) required batched eigen solvers
  - The Computational Singular Perturbation method was developed (late 80s) for analysis and reduction of stiff chemical kinetic models for gas phase combustion
    - Enables identification of cause-and-effect relations in chemical models
    - Enables stiffness reduction and stable explicit large-time step time integration
  - The dominant computational cost in CSP is the eigensolution for many typically-small (50x50) to mid-size (1000x1000) Jacobian matrices

- Joint Work with BES project : Kokkos Kernels and Exascale Catalytic Chemistry (Kyungjoo Kim, Rob Fazle, Habib Najm)

- New Kokkos Kernels functionality for Householder transformations, Givens rotations, QR, Hessenberg, Schur decomposition, and Eigen Value decomposition

- Integration and Evaluation of the Kokkos Kernels batched functionality in the CSP project is in progress

Kokkos Ecosystem provides performance portability for SciDAC applications



Kokkos Ecosystem provides performance portability with **Kokkos Core** programming model, **Kokkos Kernels** for performance portable kernels and **Kokkos Tools** for profiling and debugging.

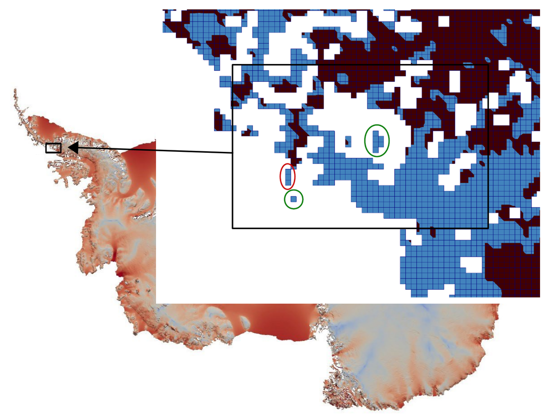
Kokkos Kernels functionality includes
- **Sparse linear algebra kernels** – matrix-matrix addition, matrix-matrix multiplication, matrix transpose
- **Graph Algorithms** – Distance-1 graph coloring, Distance-2 graph coloring, deterministic coloring, triangle counting
- **Batched BLAS and LAPACK** – LU factorization, matrix-matrix multiplication, triangular solves, and eigen solvers
- **BLAS interface –** BLAS kernels for non-standard data types and interface to vendor BLAS

Kokkos Ecosystem addresses complexity of supporting numerous many/multi-core architectures that are central to DOE HPC enterprise

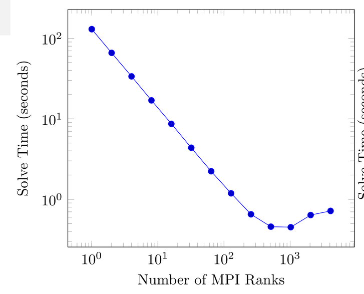
Parallel Graph Kernels for detecting degenerate features on Ice sheet Meshes

Simulation of ice sheets using Albany results in "islands" and "hinges" that result in solver convergence issues
Past Solution: Climate scientists manually remove these degenerate features using sequential code

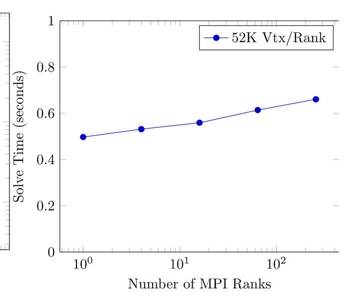
FASTMath contribution:
- Reformulate the problem as graph connected and biconnected components
- Develop distributed-memory parallel graph algorithms and implement it in Zoltan2 package of Trilinos
- Integrate with Albany for removal of degenerate features as soon as they form
- When compared to the past sequential code **speed up varies from 59x on a 16km mesh (6 MPI ranks) and 46,880x on 1 KM mesh (1536 MPI ranks)**



Strong scaling: 1 km mesh



Weak scaling: 16km, 8km, 4km, 2km, 1km meshes



- **Best Paper Award**: "A Parallel Graph Algorithm for Detecting Mesh Singularities in Distributed Memory Ice Sheet Simulations", Ian Bogle, Karen Devine, Mauro Perego, Siva Rajamanickam, George Slota. International Conference on Parallel Processing, 2019.

- **Developed parallel graph algorithms to detect degenerate features on ice sheet meshes**
- **Improved workflow of climate scientists by in-situ removal of the degenerate features during the simulations**

Ongoing Work and Future Directions

- **Multigrid Solvers**
  - Outreach – SNL multigrid developers will present hand-on training at Argonne Training Program on Extreme-Scale Computing, August 2019.
  - ProSPect: Investigate multigrid preconditioners for fully-coupled flow+temperature model

- **Domain Decomposition Solvers**
  - Integrate ShyLU/FROSch with Albany for land-ice simulations
  - Develop and evaluate methods using accelerator-aware subdomain solvers for FROSch

- **Kokkos Kernels**
  - Develop batched linear algebra kernels to support SciDAC applications
  - Develop Sparse linear algebra kernels for supporting the needs of domain decomposition solvers
  - Develop graph kernels to support solver and application needs
  - New requests: Discussions on ML needs of ASCR applications and plans to support them.

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-NA-0003525.
SAND2019-0331