SAND2020-13734PE

December 2020 ECP ST Project Review

ECP Project WBS 2.3.3.13 (Clover)

Business Sensitive Information

EXASCALE COMPUTING PROJECT

PI: Jack Dongarra (UTK)

Co-PIs for PEEKS: Erik Boman (Sandia), Hartwig Anzt (KIT/UTK)

Date: Dec. 9, 2020

National Nuclear Security Administration

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.

U.S. DEPARTMENT OF ENERGY | Office of Science

Clover/PEEKS-Trilinos: Exascale readiness

- Linear solves is a large fraction of the time in many applications
- Our focus: sparse solvers at large scale
- In particular, Krylov solvers (Trilinos/Belos)
- Reduce communication (CA, pipeline methods)
- Application-specific preconditioners are beyond scope of PEEKS
- Broader Trilinos functionality on exascale architectures will be addressed in Sake (new proj.)

- Worked on pre-exascale platforms:
  - Cori: Multicore CPU
  - Summit: Multi-GPU (Nvidia)
- Just getting started:
  - Tulip: AMD accelerators

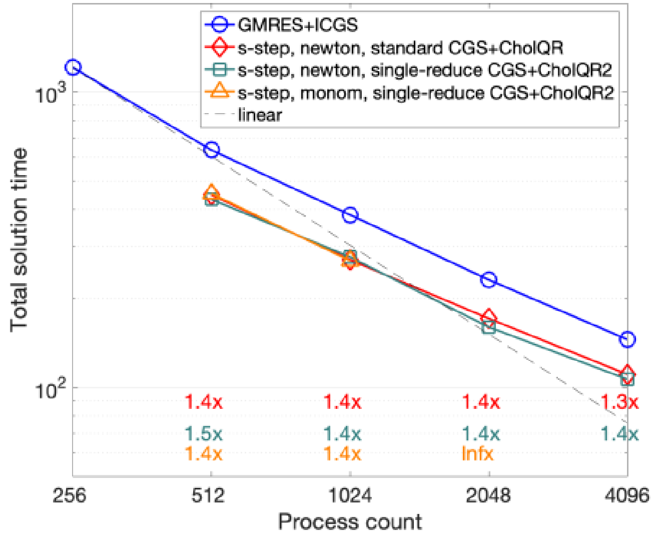
- Major challenges:
  - Overlapping communication and computation
    - Problematic when there are both global collectives and point-to-point
  - GPU-aware MPI, performance is lacking
- Critical dependencies:
  - Kokkos, KokkosKernels
  - Tpetra

- Implemented and deployed several new communication-reducing solvers in Trilinos/Belos
  - S-step and pipelined CG/GMRES
  - Single-reduce CGS/MGS orthogonalization
- Low-synch Householder (in progress)
- Polynomial preconditioning (in progress)

Clover/PEEKS-Trilinos: Capability integration

- ExaWind: Integrated new Belos solvers to reduce communication in linear solves
- ExaGraph: Integrating polynomial preconditioning in spectral partitioner (LOBPCG eigensolver)
- ATDM/SNL: Discussions with Empire and Sparc application teams, already using Trilinos solvers
- xSDK: Any improvements in Trilinos benefit xSDK

- New s-step GMRES on ExaWind NaLu-Wind problem show 1.4X speedup on up to 4096 cores (Cori)



- Challenges:
  - Get software stack running on new platforms
  - AD teams currently not using Trilinos are reluctant to try (high startup cost?)

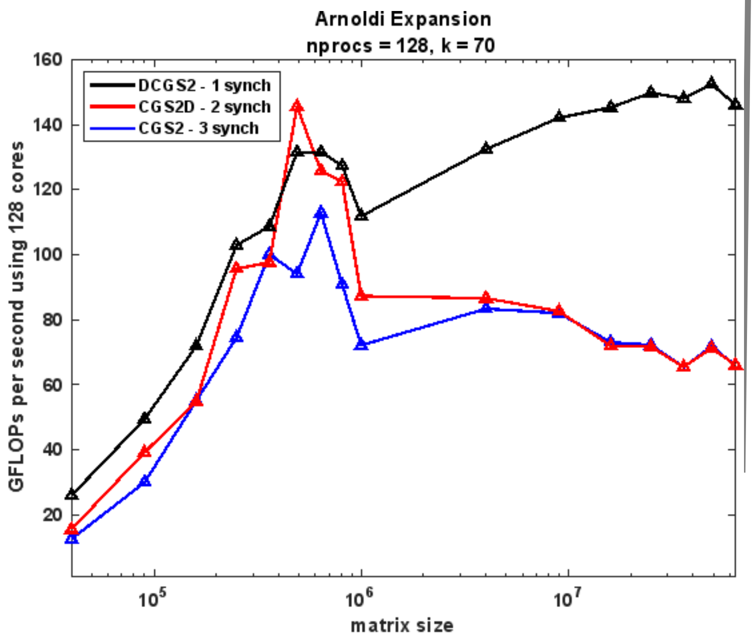
- KPP-3 values
  - Passing value: 4
  - Present value: 1 (ExaWind)
  - Coming soon: ExaGraph/Sphynx

Clover/PEEKS-Trilinos: Portability, performance

- We rely on Kokkos/KokkosKernels for performance portability: CPU, Nvidia GPU, AMD GPU, Intel accelerators
- “Write once, run anywhere” – minimal code changes for new platforms
- Saves valuable human resources (software development time)

- Focus performance on these areas
  - Reduce inter-node communication
  - Overlap communication and computation
  - Use optimized kernels on node (KokkosKernels, BLAS)
- For Krylov methods, most flops are in orthogonalization (Gram-Schmidt)
  - This can benefit from accelerators

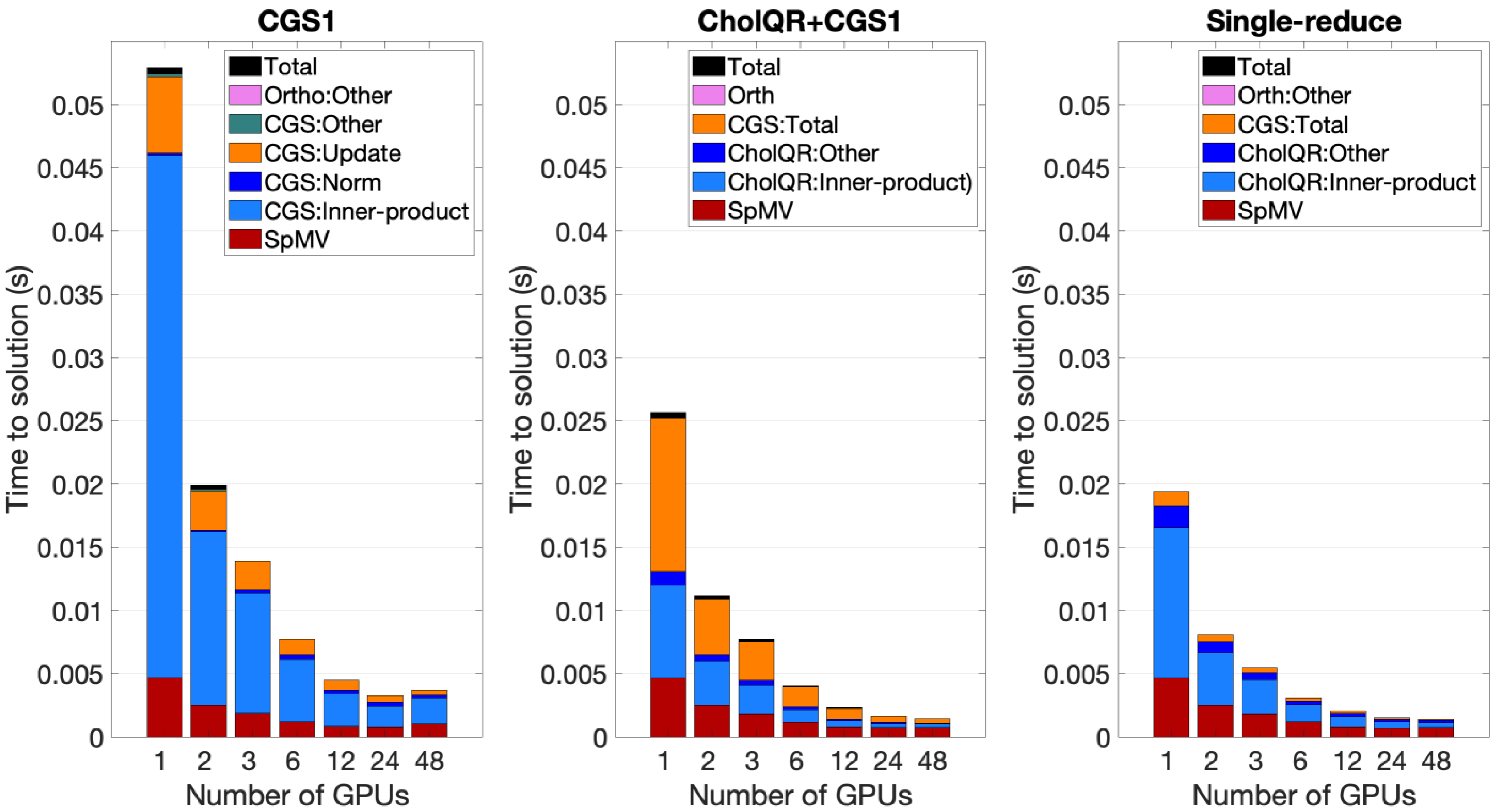
- Collaboration with Daniel Bielich & Julien Langou (CU Denver) on low-synch QR/ortho.
- Currently CPU, GPU in progress



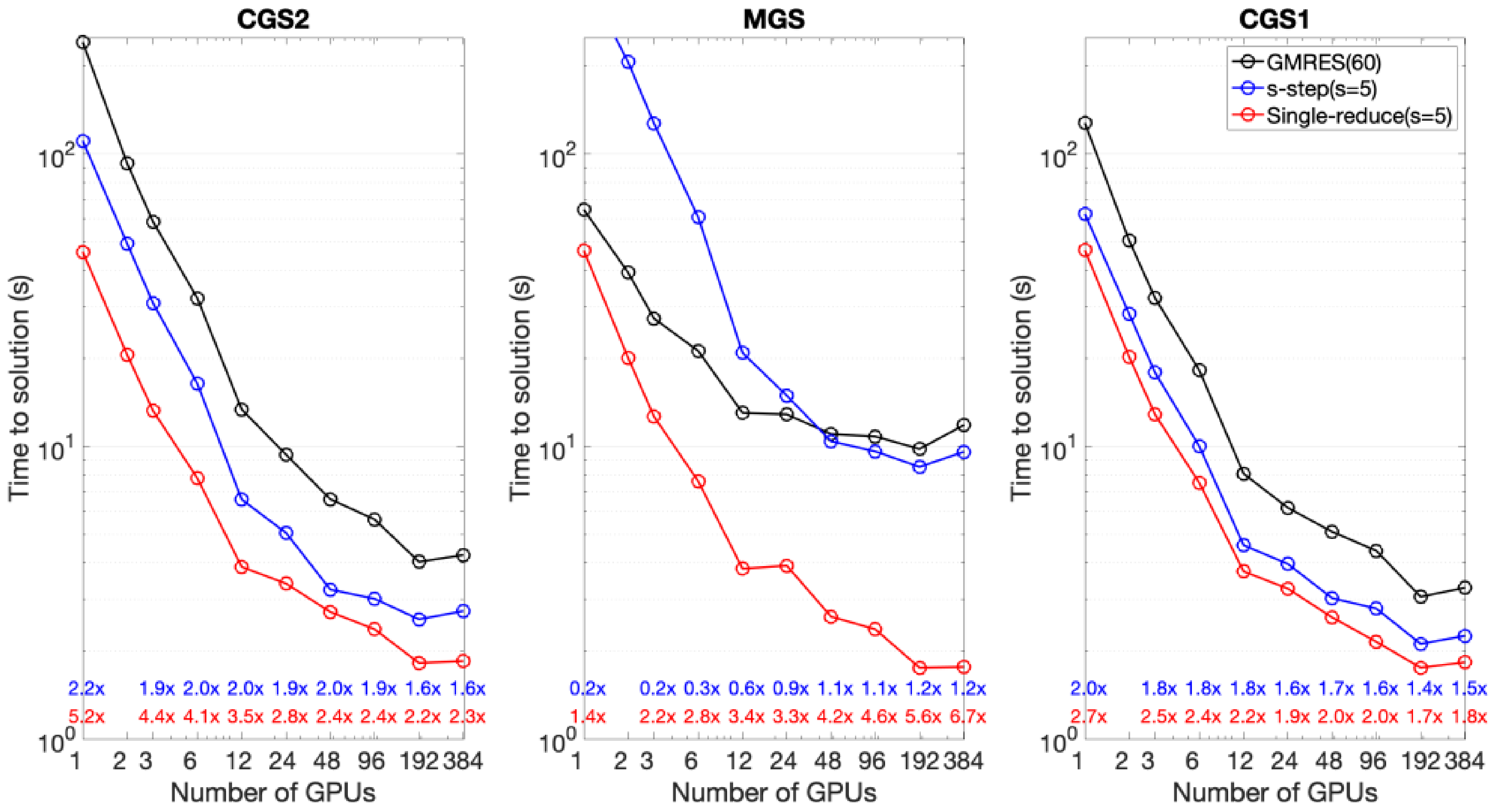
- Future challenges
  - Porting code is not enough
  - Algorithm research is crucial to get optimal performance

Backup slides

GMRES timing breakdown (GPU)

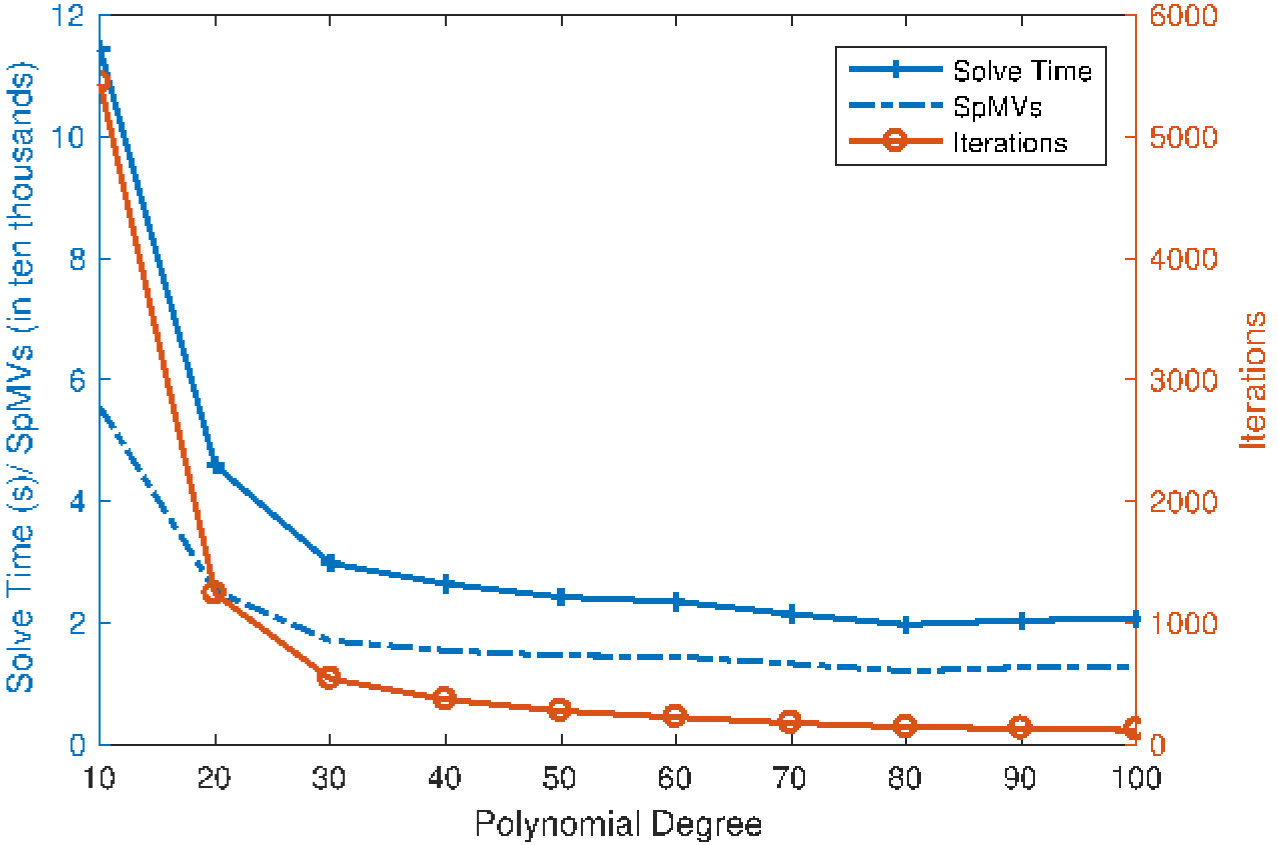


CA-GMRES on Summit



EXASCALE COMPUTING PROJECT

Polynomial Preconditioning



Matrix is cfd2, n=123440. Running GMRES(50) with random right-hand side. 32 MPI processes over 1 node.

$Ax = b \rightarrow p(A)x = p(A)b$

General polynomial preconditioner based upon the GMRES polynomial.**

Root-adding makes it stable for very high degrees.

Can reduce SpMVs, iterations, and solve time- less time in orthogonalization kernels.

Stand-alone or combine with other standard preconditioners, such as ILU.

Available now in the Trilinos linear solvers package Belos!

For this example, degree 80 polynomial gives 57x speedup over no preconditioning, with over 14x reduction in SpMVs.

**For details, see:
Jennifer Loe, Erik Boman, and Heidi Thornquist. *Polynomial Preconditioned GMRES in Trilinos: Practical Considerations for High-Performance Computing* In Proceedings of the 2020 SIAM Conference on Parallel Processing for Scientific Computing, pages 35–45.