SAND2016-3238C

Performance of smoothers for algebraic multigrid preconditioners for finite element variational multiscale incompressible magnetohydrodynamics

SIAM PP 2016

Paris, France

Apr 13, 2016

Paul Lin, John N. Shadid, Jonathan J. Hu
Center for Computing Research
Sandia National Labs

Paul Tsuji

Lawrence Livermore National Laboratory

Livermore CA

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.

Need for large-scale simulations

- Goal: high-fidelity solutions of transport phenomena for large-scale problems with complex physics and geometry
  - CFD, magnetohydrodynamic (MHD) simulations
  - DOE interested in: pulsed fusion reactors (e.g. z-pinch), magnetically confined fusion (e.g. ITER tokamak)
- FEM stabilized/VMS; unstructured meshes
- Fully-implicit Newton-Krylov solution approach
  - robust (promising for complex physics and chemistry)
  - preconditioner critical: robustness, efficiency, scalability
  - large-scale problems: multilevel/multigrid
- Talk focus: our fully-coupled Newton-Krylov multigrid preconditioned approach for large-scale FEM simulations
  - Drekar CFD/MHD application code (resistive MHD)
  - Trilinos linear solvers

Resistive MHD model

(J. Shadid, R. Pawlowski, E. Cyr, L. Chacon)

Navier-Stokes + Magnetic Induction

$$\rho\frac{\partial \mathbf{u}}{\partial t} + \rho(\mathbf{u}\cdot\nabla\mathbf{u}) - \nabla\cdot(\mathbf{T} + \boxed{\mathbf{T_M}}) - \rho\mathbf{g} = 0$$

$$\mathbf{T} = -(P + \frac{2}{3}\mu(\nabla\cdot\mathbf{u}))\mathbf{I} + \mu[\nabla\mathbf{u} + \nabla\mathbf{u}^T]$$

$$\boxed{\mathbf{T_M} = \frac{1}{\mu_0}\mathbf{B}\otimes\mathbf{B} - \frac{1}{2\mu_0}\|\mathbf{B}\|^2\mathbf{I}}$$

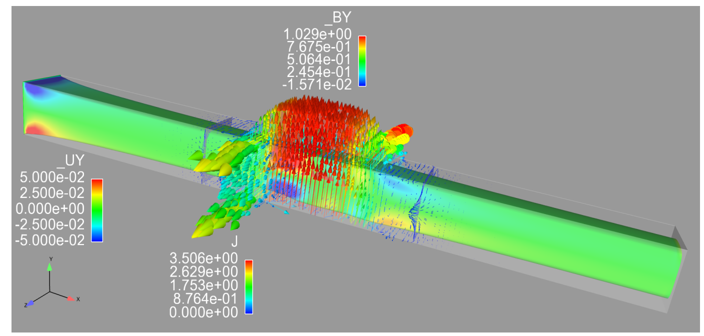
$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{u}) = 0$$

$$\rho C_p\left[\frac{\partial T}{\partial t} + \mathbf{u}\cdot\nabla T\right] + \nabla\cdot\mathbf{q} - \boxed{\eta\|\mathbf{J}\|^2} = 0$$

$$\boxed{\frac{\partial\mathbf{B}}{\partial t} - \nabla\times(\mathbf{u}\times\mathbf{B}) + \nabla\times(\frac{\eta}{\mu_0}\nabla\times\mathbf{B}) = \mathbf{0}}$$

Steady-state MHD generator

- Flow with external cross-stream B field
- 8 DOFs/mesh node



Drekar implicit/IMEX FE application

(J. Shadid, R. Pawlowski, E. Cyr, T. Smith, T. Wildey, E. Phillips, etc.)

- Navier-Stokes, MHD, LES, RANS
- stabilized FEM, unstructured hexahedral meshes
- fully-coupled multigrid preconditioned Newton-Krylov solve (Trilinos)

U.S. DEPARTMENT OF ENERGY NNSA National Nuclear Security Administration

Sandia National Laboratories

Numerical Solution Approach

- Resistive (incompressible) MHD model
- Stabilized FEM: variational multiscale (VMS) (Hughes 1995)
  - many authors employ VMS for CFD
- Most MHD turbulence simulation employ spectral methods
- MHD turbulence with VMS (Shadid et al, Sondak+Oberai; possibly Badia+Codina)
  - VMS MHD turbulence: advantage of being automatically adaptive
- Unstructured meshes
- Fully-coupled Newton-Krylov (multigrid preconditioned)
  - Robust (promising for complex physics and chemistry)
  - but depends on efficiency of sparse linear solver
  - preconditioner critical: robustness, efficiency, scalability
  - large-scale problems: multilevel/multigrid

Smoothers multgrid preconditioned approach

- Effective smoothers critical for multigrid
  - Need to efficiently damp high frequency errors
  - Standard relaxation smoothers not sufficiently robust for our MHD problems
  - Our standard smoother ILU(0) overlap=1 is expensive
  - Interested in other smoothers
- Krylov smoothers
  - Lots of previous work for SPD problems
  - Far less previous work for nonsymmetric systems
  - Cannot use GS or ILU for problems with zeroes on diagonal
- Evaluate smoothers
  - Steady-state
  - Transient

Brief Trilinos overview

- Classic Trilinos (Epetra-based) (Heroux et al.):
  - Limited by 32-bit integer global objects
  - Most packages employ flat MPI-only; future architectures?
- Modern Trilinos solver stack (Tpetra-based):
  - No 32-bit limitation on global objects (employs C++ templated data types)
  - Path forward for future architectures: Trilinos Kokkos (Edwards, Trott, Sunderland; not part of this talk)

Functionality	Classic stack	Modern solver stack
Distributed linear alg	Epetra	Tpetra (Hoemmen,Trott, etc.)
Iterative linear solve	Aztec	Belos (Thornquist,Hoemmen,etc.)
Incomplete factor	Aztec, Ifpack	Ifpack2 (Hoemmen,Hu,Siefert, etc.)
Algebraic multigrid	ML	MueLu (Hu,Prokopenko,Tsuji,Siefert,Tuminaro,etc.)
Partition & load balance	Zoltan	Zoltan2 (Devine,Boman,Rajamanickam,Wolf,etc.)
Direct solve interface	Amesos	Amesos2 (Rajamanickam,etc.)

- PETSc is another well-known solvers library (ANL; Smith, Gropp, Knepley, Brown, McInnes, Balay, Zhang, et al.)

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

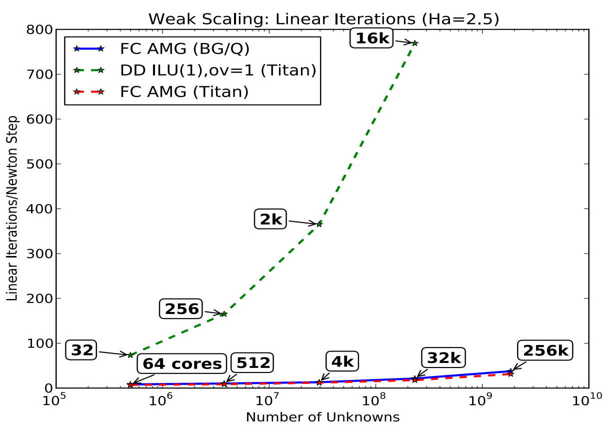
Sandia National Laboratories

Trilinos MueLu Library: algebraic multigrid preconditioners

(J. Hu, A. Prokopenko, J. Gaidamour, T. Wiesner, P. Tsuji, C. Siefert, R. Tuminaro)

- Smoothed aggregation; aggregates to produce a coarser operator
  - Create graph where vertices are block nonzeros in matrix A_k
  - Edge between vertices i and j added if block $B_k(i,j)$ contains nonzeros
  - Uncoupled aggregation
- Restriction/prolongation operator; $A_{k-1} = R_k A_k P_k$
- Repartition coarser level matrices (MueLu+Zoltan2) to reduce communication
- Coarsest level: serial direct solve (KLU; T. Davis) on 1 MPI process

Other approaches: LLNL Hypre (R. Falgout, U. Yang, T. Kolev, A. Baker, E. Chow, C. Tong, et al.), MLBDDC (S. Badia, A. Martín, J. Principe, et al.), etc.



- Weak scaling: MHD generator
- Re = 500, Re_m = 1, Ha = 2.5
- Cray XK7, IBM Blue Gene/Q

Additive Schwarz domain decomposition does not scale
Multigrid critical for performance and scaling

U.S. DEPARTMENT OF ENERGY
NNSA National Nuclear Security Administration
Sandia National Laboratories

Strong scaling: Poisson equation

(with J. Hu, J. Shadid, A. Prokopenko, E. Cyr, R. Pawlowski)

- 3D Poisson (1 DOF/mesh node)


(Image courtesy of LLNL)

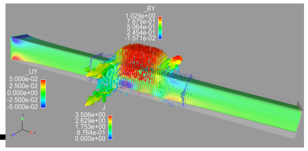
- Simple cube geometry, near uniform mesh
- Fixed problem size (2.4b DOFs); 1 MPI task/core BG/Q
- Optimal iteration count to 1.6 million cores (full-scale Sequoia BG/Q)

MPI process	CG iterations	Solve t (s)	MG setup (s)	DOFs/proc
131,072	6.3	1.17	7.67	~18,800
262,144	6.0	1.08	12.35	~9400
524,288	6.3	1	25.43	~4700
1,048,576	7.3	0.91	53.04	~2400
1,572,864	7.0	0.94	128.9	~1500

U.S. DEPARTMENT OF ENERGY NNSA National Nuclear Security Administration

Sandia National Laboratories

Weak scaling: fully-coupled multigrid MHD



(with J. Hu, J. Shadid, A. Prokopenko, E. Cyr, R. Pawlowski)

MPI	DOFs	GMRES iterations /Newton	Time/Newton step (s)		
			Multigrid setup		Solve
			Hier+smoo	Smoother	
128	845,000	14.0	12.4	11.0	4.7
1024	6,473,096	20.0	14.7	13.0	6.6
8192	50,658,056	30.8	16.9	14.2	10.1
65,536	400,799,240	53.4	20.3	16.1	17.9
524,288	3,188,616,200	98.7	45.3	19.1	40.1

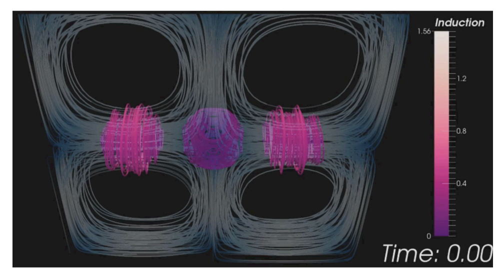
- BG/Q: 1 MPI/core
- Multigrid prec setup time/Newton step
- Smoother: ILU(0) overlap=1

- Drekar 3D MHD generator on BG/Q (simple geometry)
- Algorithmic scaling challenging for nonsymmetric matrices
  - 4096x increase in size: 6.0x iterations, 7.3x time
  - Petrov-Galerkin or energy minimization approaches promising
  - Need better aggregation, better smoothers, etc.
- Sparse matrix-matrix multiply (A_c=R*A*P)
- Employ reuse of construction of hierarchy and smoothers (Prokopenko)
  - Application dependent (e.g. cannot reuse for adaptive mesh)
  - Critical for transient simulations (10^4 or 10^5 time steps)

Results (Transient)

Transient Taylor-green MHD vortex decay (resistive MHD VMS turbulence) (8 DOF/mesh node)

- Performance comparison of different smoothers (CFL ~0.5)
  - 3 problem sizes: 16.8M, 134M, 1.07b DOFs
  - Aztec GMRESR "outer" Krylov solve
  - Jacobi, block Jacobi, GS, block GS failed for even small problems and were not included in following tables
  - ILU(0) overlap=0,1
  - GMRES (no prec, Gauss-Seidel, block Jacobi, block GS)
  - Schwarz/domain decomposition GMRES smoother
    - no prec, Gauss-Seidel, block Jacobi, block GS
    - Overlap=0,1
- Robustness for higher CFL study
  - 2 problem sizes: 16.8M,1.07b DOFs
  - Focus on more promising smoothers



Transient Taylor-Green MHD Smoother Comparison (16.8M DOFs)

- Transient Taylor-green MHD vortex decay (VMS resistive MHD)
- Cube domain, 128^3 elem, 20 time steps, dt=0.025 (CFL ~0.5)
- 256 MPI; linux cluster dual-socket SNB, IB fat-tree(TLCC2)

smoother		iters/dt	Prec setup(s)	Solve(s)	Prec+solve(s)	Total(s)	Mem(MB)
SGS		87.7	18	730	748	1108	1050
ILU	overlap=0	20.9	127	110	237	606	1157
	overlap=1	14.2	259	97	356	725	1436
GMRES	noprec	15.4	23	260	283	696	917
	ptGS	13.6	24	413	437	828	917
	bkJac	13.1	35	238	273	665	927
	bkGS	13.6	24	413	437	828	917
DD-GMRES	noprec ov0	21.4	35	530	565	929	917
	noprec ov1	15.9	125	486	611	995	1107
	ptGS ov0	20.2	51	986	1037	1396	917
	ptGS ov1	12.1	142	759	901	1263	1102
	bkJac ov0	20.2	45	535	581	942	925
	bkJac ov1	15.5	151	512	663	1069	1129
	bkGS ov0	20.2	51	986	1037	1396	917
	bkGS ov1	11.9	162	902	1063	1422	1129

- DD-GMRES smoother not competitive (particularly GS and block GS prec)

Transient Taylor-Green MHD Smoother Comparison (134M DOFs)

- Transient Taylor-green MHD vortex decay (VMS resistive MHD)
- Cube domain, 256^3 elem, 20 time steps, dt=0.0125 (CFL ~0.5)
- 2048 MPI; linux cluster dual-socket SNB, IB fat-tree (TLCC2)

smoother		iters/dt	Prec setup(s)	Solve(s)	Prec+solve(s)	Total(s)	Mem(MB)
SGS		Failed					
ILU	overlap=0	18.6	680	154	834	680	1164
	overlap=1	11.9	862	134	995	862	1440
GMRES	noprec	11.7	741	328	1069	741	920
	ptGS	10.9	921	516	1437	921	920
	bkJac	9.5	749	306	1055	749	933
	bkGS	11.0	1376	954	2330	1376	936
DD-GMRES	noprec ov0	20.4	1272	842	2114	1272	926
	noprec ov1	14.5	1657	1094	2751	1657	1111
	ptGS ov0	18.2	2083	1655	3738	2083	939
	ptGS ov1	10.9	1948	1383	3331	1948	1111
	bkJac ov0	18.1	1218	789	2007	1218	931
	bkJac ov1	10.8	1455	872	2327	1455	1134
	bkGS ov0	18.3	2319	1881	4200	2319	939
	bkGS ov1	10.9	2166	1582	3748	2166	1134

- DD-GMRES smoother not competitive (particularly GS and block GS prec)

Transient Taylor-Green MHD Smoother Comparison (1.07b DOFs)

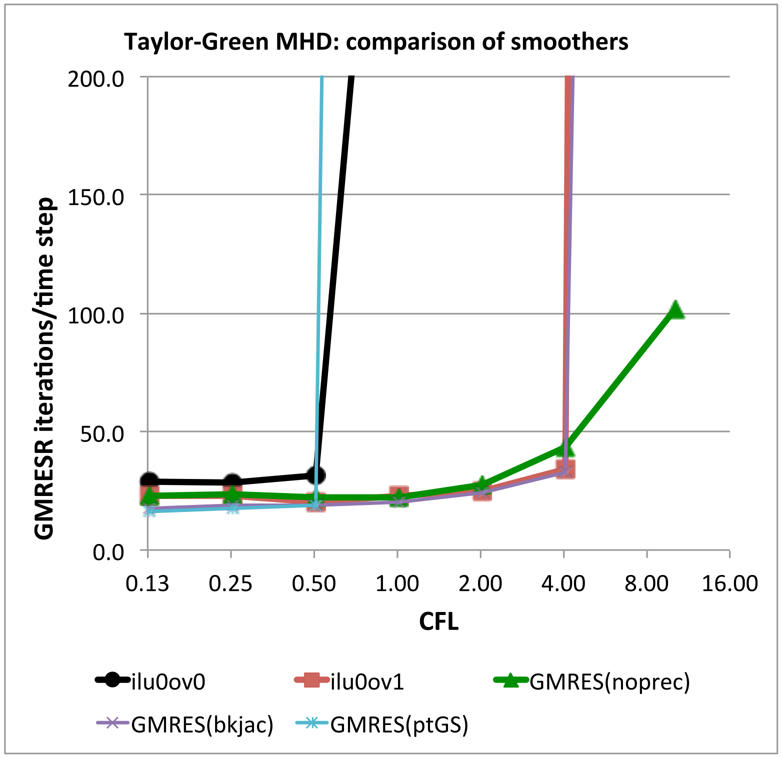
- Transient Taylor-green MHD vortex decay (VMS resistive MHD)
- Cube domain, 512^3 elem, 20 time steps, dt=0.0625 (CFL ~0.5)
- 16384 MPI; linux cluster dual-socket SNB, IB fat-tree (TLCC2)

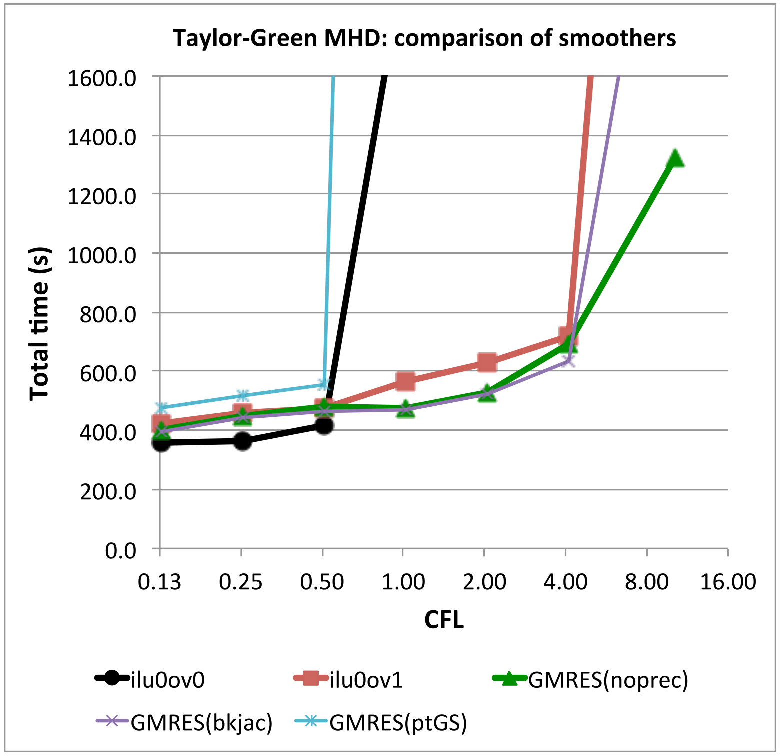
smoother		iters/dt	Prec setup(s)	Solve(s)	Prec+solve(s)	Total(s)	Mem(MB)
ILU	overlap=0	68.1	251	411	662	1380	1287
	overlap=1	37.3	407	280	687	1280	1519
GMRES	noprec	31.0	92	610	702	1326	1002
	ptGS	21.5	94	728	822	1374	1002
	bkJac	21.9	60	458	518	1028	1017
DD-GMRES	noprec ov0	69.0	61	2040	2101	2689	1031
	bkJac ov0	69.7	135	2227	2362	3088	1040

- DD-GMRES smoother not competitive
- GMRES smoother with either no preconditioner or block Jacobi are most competitive to ILU
  - GMRES can significantly lower iterations, but cost/iteration expensive
  - Trade-off between expensive ILU factorization for setup vs. more mat-vecs and higher communication during solve of GMRES smoother
  - ILU smoother requires more memory

Transient Taylor-Green MHD
Smoother Comparison: Robustness with CFL (16.8M DOFs)

- Transient Taylor-green MHD vortex decay (VMS resistive MHD)
- Cube domain, 128^3 elem, 10 time steps
- 256 MPI (TLCC2 dual-socket SNB; IB fat-tree)





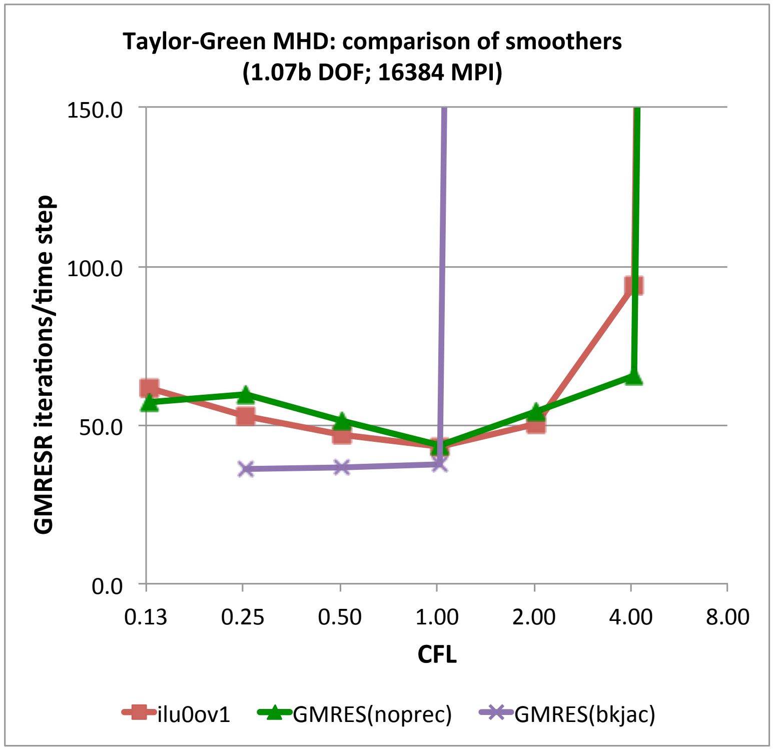
- ILU(0) overlap=0, GMRES(no prec),GMRES(bkGS prec) smoothers not competitive
- Need to compare ILU(0) overlap=1 with GMRES smoothers

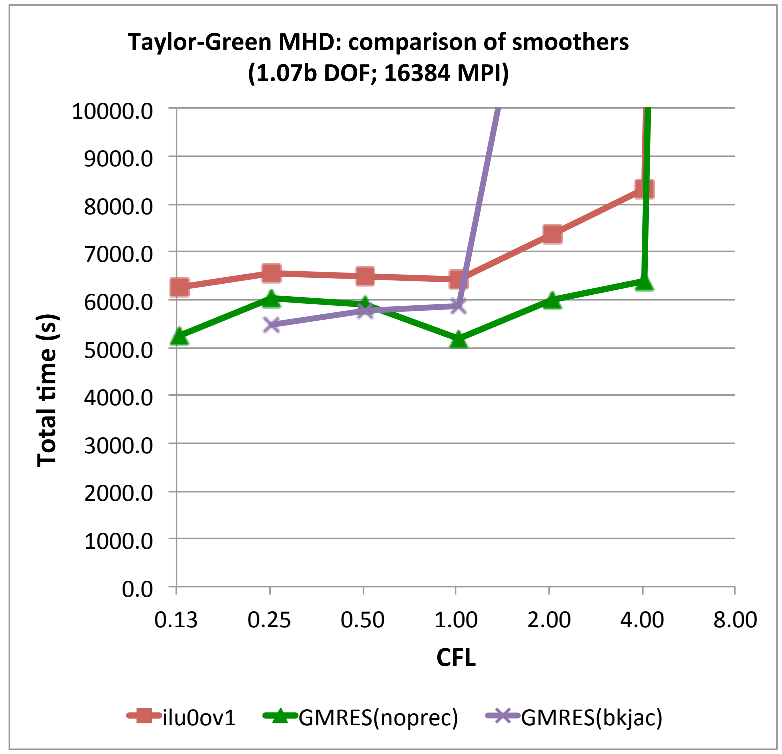
U.S. DEPARTMENT OF ENERGY NNSA National Nuclear Security Administration

Sandia National Laboratories

Transient Taylor-Green MHD
Smoother Comparison: Robustness with CFL (1.07b DOF)

- Transient Taylor-green MHD vortex decay (VMS resistive MHD)
- Cube domain, 512^3 elem, 10 time steps; 16384 MPI (BG/Q)





- Did not even run ILU(0) overlap=0 because fails at low CFL (unfortunate because ILU overlap=1 requires ~40% more memory than GMRES smoother)
- GMRES(bkGS prec) fails CFL ~2
- ILU(0) overlap=1, GMRES(no prec) fail CFL ~8

Concluding remarks and future work

- Performed an initial evaluation of GMRES as an alternative smoother to our current standard (ILU)
  - Initial empirical study (one test case so far)
  - Shows promise: can solve an initial class of relevant problems (appears competitive; expensive, but so is ILU)
  - Memory usage benefits (ILU requires ~40% more)
  - Krylov smoother should work on indefinite systems, nonsymmetric matrices with zero diagonal entries (but still need to demonstrate)
- Issues
  - Need to go back and try to analyze method more carefully
    - Think about which Krylov methods better at killing off high frequency modes
  - AMG notorious for too many adjustable parameters
    - Krylov smoothers exacerbate this issue (e.g. precondition? sweeps? tolerance? Etc.)

Additional future work

- Many challenges for multigrid-preconditioned linear solve
  - algorithmic scaling
  - multigrid preconditioner setup (sparse mat-mat)
- Trilinos Kokkos for manycore and accelerators (“X” for MPI+X)
  - Tpetra with Kokkos implementation is ongoing work
  - other Trilinos packages need to being implemented with Kokkos: MueLu setup, additional Ifpack2 smoothers, etc.
- Drekar progress depends on above Kokkos work
  - Trininos components for assembly ported (I. Demeshko)
- Additional physics and discretization issues include, e.g.
  - strong convection effects, hyperbolic systems
  - non-uniform FE aspect ratios

Thanks For Your Attention!

Paul Lin (ptlin@sandia.gov)
John Shadid, Jonathan Hu, Paul Tsuji

Acknowledgments:

- Andrey Prokopenko, Ray Tuminaro (Trilinos ML library--P.I.), Chris Siefert
- Mark Hoemmen, Eric Phipps
- LLNL LC BG/Q team (especially John Gyllenhaal and Scott Futral)

Funding Acknowledgment:
The authors gratefully acknowledge funding from the DOE NNSA Advanced Simulation & Computing (ASC) program and ASCR Applied Math Program

U.S. DEPARTMENT OF ENERGY NNSA National Nuclear Security Administration

Sandia National Laboratories

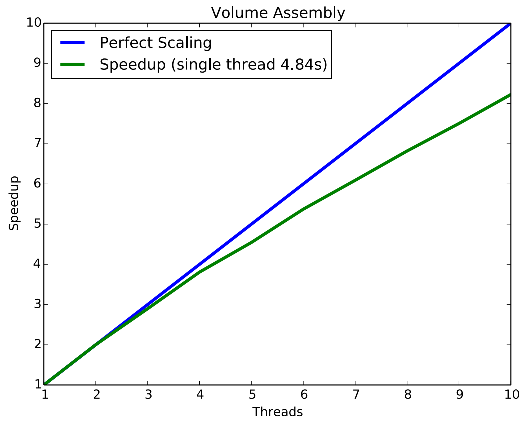
Extra Slides

U.S. DEPARTMENT OF ENERGY NNSA National Nuclear Security Administration

Sandia National Laboratories

Future looking slide: Kokkos for FEM matrix assembly

- Matrix assembly for Maxwell's eqns (2nd order form); edge-based
  - Results courtesy of Cyr, Bettencourt, Demeshko, Pawlowski
  - Initial results for Kokkos implementation of the Trilinos Phalanx library: directed acyclic graph (DAG) based assembly abstraction
  - Template-based embedded C++ data types used within DAG to assemble Jacobian



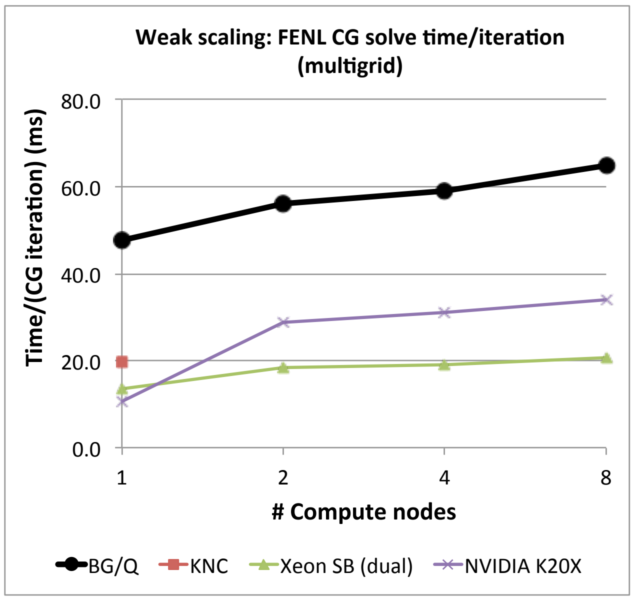
- Strong scaling 40^3 cube of hex elements
- OpenMP threading on dual Intel Xeon (Westmere)

- Drekar employs Phalanx for assembly; will leverage this work

U.S. DEPARTMENT OF ENERGY
NNSA National Nuclear Security Administration
Sandia National Laboratories

Future looking slide: Kokko for linear solve

- Finite Element Nonlinear (FENL) “miniapp” (C. Edwards): 3D nonlinear heat equation using Kokkos
- MueLu setup and Tpetra mat-mat multiply not yet implemented with Kokkos (setup on CPU)
- MueLu apply using Kokkos (Tpetra converted to Kokkos)
- Results courtesy of E. Phipps



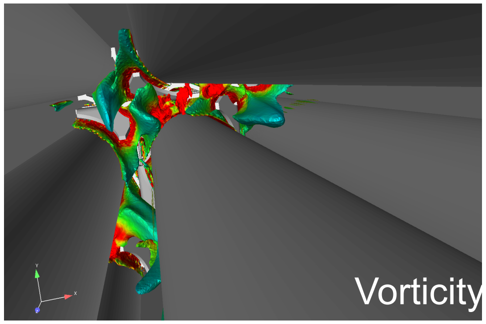
- Preliminary results
- Solve times (setup not included)
- BG/Q:1 MPI rank/node, 64 threads/rank
- KNC: Intel Xeon Phi 224 threads (multi-node results need further investigation)
- Xeon SB (dual): dual-socket Sandy Bridge 1MPI rank/node, 16 threads/rank
- NVIDIA K20X GPU 1 MPI rank/node

FENL with Kokkos runs on CPU, GPU, and Xeon Phi (FENL with MueLu apply)

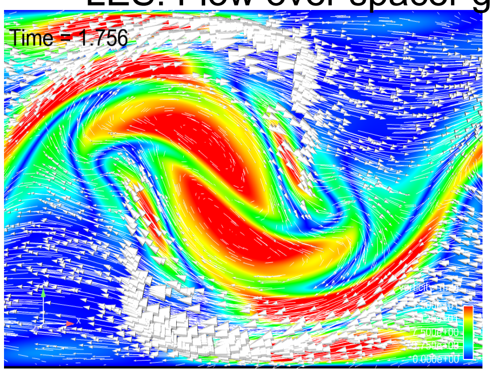
- Drekar will leverage this work

Drekar

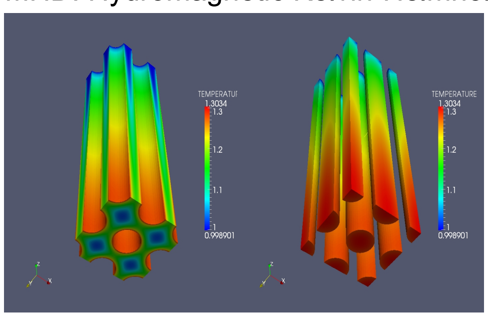
(J. Shadid, R. Pawlowski, E. Cyr, T. Smith, T. Wildey)



LES: Flow over spacer grid



MHD: Hydromagnetic Kelvin-Helmholtz



Conjugate Heat Transfer

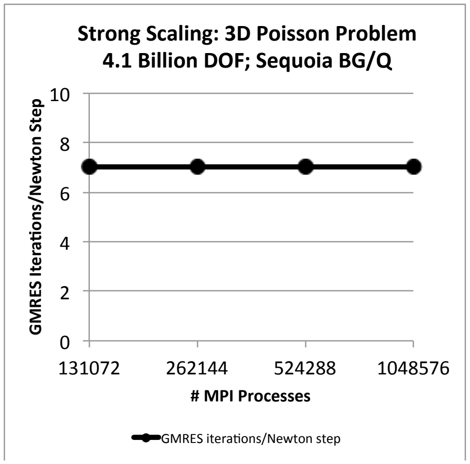
Scalable parallel implicit/IMEX FE code

- Includes: Navier-Stokes, MHD, LES, RANS
- Architecture admits new coupled physics
- Support for advanced discretizations
  - mixed, physics compatible and high-order basis functions
  - multi-physics capable (conjugate heat transfer)
- Advanced UQ tools/techniques
  - Adjoint based sensitivities and error-estimates
- Advanced solution methods
  - Parallel solvers from SNL's Trilinos framework
  - Physics-based preconditioning
  - Fully-coupled multigrid for monolithic systems

How many MPI tasks needed for multigrid?

- Our experience (app dependent): really bad multigrid setup scaling not until > couple 10,000s of MPI tasks (definitely ~100k)
- Do we need to worry about multigrid with one million MPI tasks?
  - Flat MPI-only clearly not the way to go
- Multigrid with $O(10^5)$ MPI tasks relevant for MPI+X approaches ?
  - MPI+X: number of MPI tasks the same as number of compute nodes
  - Number is clearly app dependent
  - May need minimum number per node that is > 1 to utilize NIC bandwidth
  - Future DOE machines (2016-2019) $O(10^3)$ to $O(10^5)$ nodes
  - Exascale machines will be $O(10^4)$ to $O(10^6)$ nodes
- Good multigrid performance on $O(10^5)$ tasks appears to be still relevant
  - Need to continue to work on algorithmic scaling
- Potential approach for "X" in MPI+X -> SNL Trilinos Kokkos

  - Drekar 3D Poisson problem
  - Simple cube geometry
  - Optimal iteration count to 1 million cores
  - Preconditioner setup does not scale



Preliminary weak scaling BG/Q: CFD jet (Re = 10^6, CFL ~0.25)

Drekar/Epetra/Aztec/ML/Ifpack SGS (classic Trilinos)

cores	DOF	Newt /dt	Iter/ Newt	Iter/ dt	Total time	Time/Newt (sec)		
						Prec	Solve	Jac
32	901056	3.40	8.41	28.6	1329	1.79	7.76	27.72
256	6931504	3.50	10.8	37.8	1435	1.94	9.56	27.85
2048	54,723,004	3.60	15.72	56.6	1657	2.95	13.6	27.92
16,384	434,886,004	3.60	22.42	80.7	1855	3.02	19.3	27.94
131,072	3,467,532,004	Cannot run problems with DOF > 2.1b						

Drekar/Tpetra/Belos/MueLu/Ifpack2 SGS (modern Trilinos)

cores	DOF	Newt /dt	Iter/ Newt	Iter/ dt	Total time	Time/Newt (sec)		
						Prec	Solve	Jac
32	901,056	3.40	8.44	28.7	1533	2.13	7.29	32.76
256	6,931,504	3.60	11.92	42.9	1729	2.28	10.1	32.86
2048	54,723,004	3.70	16.32	60.4	1932	2.53	14.0	32.89
16,384	434,886,004	3.90	24.67	96.2	2328	3.34	20.7	32.9
131,072	3,467,532,004	4.00	36.23	144.9	3421	14.9	30.4	32.9

- Iterations higher with Tpetra
- Keep improving setup: 131,072 cores Mar 2015 163s, Apr 2015 19s
- Tpetra/MueLu enables simulations > 2.1b (32-bit)
- Solved 40 billion DOF for 10 time steps