Parallel High-Order Geometric Multigrid Methods on Adaptive Meshes for Highly Heterogeneous Nonlinear Stokes Flow Simulations of Earth's Mantle

Johann Rudi[1] Hari Sundar[2] Tobin Isaac[1] Georg Stadler[3]
Michael Gurnis[4] Omar Ghattas[1,5]

[1]Institute for Computational Engineering and Sciences (ICES), The University of Texas at Austin, USA

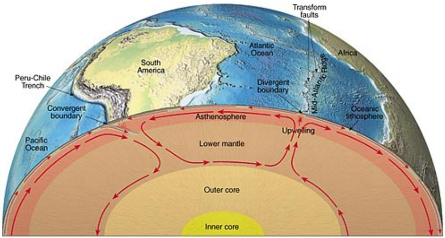
[2]School of Computing, The University of Utah, USA

[3]Courant Institute of Mathematical Sciences, New York University, USA

[4]Seismological Laboratory, California Institute of Technology, USA

[5]Jackson School of Geosciences and Department of Mechanical Engineering, The University of Texas at Austin, USA

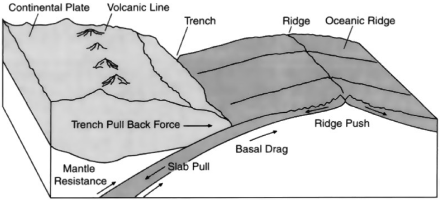
Introduction to mantle convection & plate tectonics



- Mantle convection is the thermal convection in the Earth's upper $\sim$3000 km
- It controls the thermal and geological evolution of the Earth
- Solid rock in the mantle moves like viscous incompressible fluid on time scales of millions of years

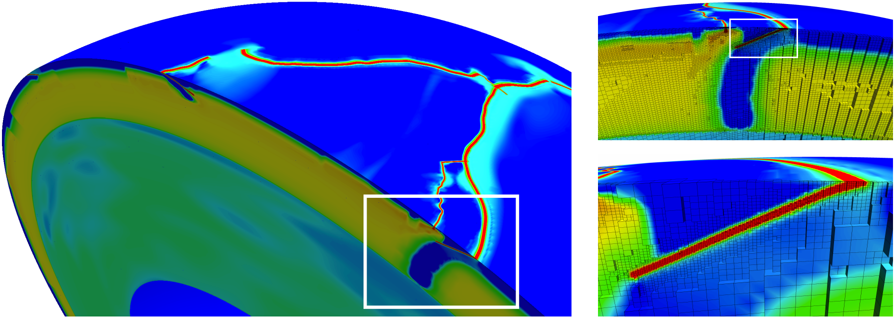
Main open questions:

- Energy dissipation in hinge zones
- Main drivers of plate motion: negative buoyancy forces or convective shear traction
- Role of slab geometries
- Accuracy of rheology extrapolations from experiments



Our research target:

Global simulation of the
Earth's mantle convection & associated plate tectonics with
realistic parameters & resolutions down to faulted plate boundaries.



Effective viscosity field and adaptive mesh resolving narrow plate boundaries (shown in red).

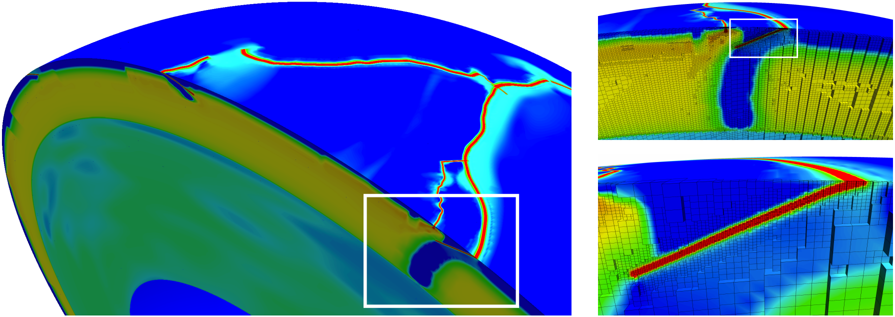
(Visualization by L. Alisic)

Earth's mantle flow, modeled as a nonlinear Stokes system

$$-\nabla \cdot \left[\mu(T, \boldsymbol{u}) \left(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^\top\right)\right] + \nabla p = \boldsymbol{f}(T)$$

$$\nabla \cdot \boldsymbol{u} = 0$$

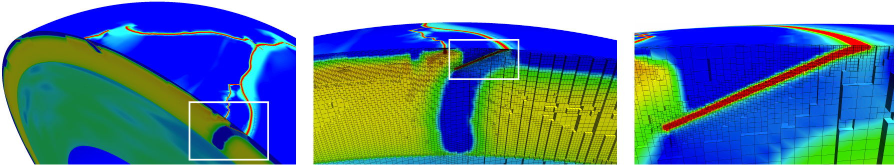
$\boldsymbol{u}$ . . . velocity
p . . . pressure
T . . . temperature
μ . . . viscosity



Effective viscosity field and adaptive mesh resolving narrow plate boundaries (shown in red).

(Visualization by L. Alisic)

Solver challenges



What causes the demand for scalable solvers for high-order discretizations on adaptive grids? — The severe nonlinearity, heterogeneity & anisotropy of the Earth's rheology:

- Up to 6 orders of magnitude viscosity contrast; sharp viscosity gradients due to decoupling at plate boundaries
- Wide range of spatial scales and highly localized features w.r.t. Earth radius (∼6371 km): plate thickness ∼50 km & shearing zones at plate boundaries ∼5 km
- Desired resolution of ∼1 km results in $O(10^{12})$ degrees of freedom on a uniform mesh of Earth's mantle, so adaptive mesh refinement is essential
- Demand for high accuracy leads to high-order discretizations

Summary of main results

I. Efficient methods/algorithms

- High-order finite elements
- Adaptive meshes, resolving viscosity variations
- Geometric multigrid (GMG) preconditioners for elliptic operators
- Novel GMG based BFBT/LSC pressure Schur complement preconditioner
- Inexact Newton-Krylov method
- H^{-1}-norm for velocity residual in Newton line search

II. Scalable parallel implementation

- Matrix-free stiffness/mass application and GMG smoothing
- Tensor product structure of finite element shape functions
- Octree algorithms for handling adaptive meshes in parallel
- Algebraic multigrid (AMG) only as coarse solver for GMG avoids full AMG setup cost and large matrix assembly
- Parallel scalability results up to 16,384 CPU cores (MPI)

Results covered in this talk

I. Efficient methods/algorithms

- High-order finite elements
- Adaptive meshes, resolving viscosity variations
- Geometric multigrid (GMG) preconditioners for elliptic operators
- Novel GMG based BFBT/LSC pressure Schur complement preconditioner
- Inexact Newton-Krylov method
- H^{-1}-norm for velocity residual in Newton line search

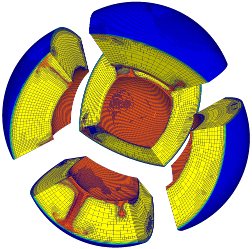
II. Scalable parallel implementation

- Matrix-free stiffness/mass application and GMG smoothing
- Tensor product structure of finite element shape functions
- Octree algorithms for handling adaptive meshes in parallel
- Algebraic multigrid (AMG) only as coarse solver for GMG avoids full AMG setup cost and large matrix assembly
- Parallel scalability results up to 16,384 CPU cores (MPI)

Scalable parallel Stokes solver

Parallel octree-based adaptive mesh refinement (p4est)

- Identify octree leaves with hexahedral elements
- Octree structure enables fast parallel adaptive octree/mesh refinement and coarsening
- Octrees and space filling curves enable fast neighbor search, repartitioning, and $2:1$ balancing in parallel
- Algebraic constraints on non-conforming element faces with hanging nodes enforce global continuity of the velocity basis functions
- Demonstrated scalability to $O(500\text{K})$ cores (MPI)



High-order finite element discretization of the Stokes system

$$\begin{cases} -\nabla \cdot \left[\mu \left(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^\top \right) \right] + \nabla p = \boldsymbol{f} \\ \nabla \cdot \boldsymbol{u} = 0 \end{cases} \xrightarrow{\text{discretize}} \begin{bmatrix} \mathbf{A} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}$$

- High-order finite element shape functions
- Inf-sup stable velocity-pressure pairings: $\mathbb{Q}_k \times \mathbb{P}_{k-1}^{\text{disc}}$ with $2 \leq k$
- Locally mass conservative due to discontinuous pressure space
- Fast, matrix-free application of stiffness and mass matrices
- Hexahedral elements allow exploiting the tensor product structure of basis functions for a high floating point to memory operations ratio

Linear solver: Preconditioned Krylov subspace method

Fully coupled iterative solver: GMRES with upper triangular block preconditioning

$$\underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{0} \end{bmatrix}}_{\text{Stokes operator}} \underbrace{\begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{B}^\top \\ \mathbf{0} & -\tilde{\mathbf{S}} \end{bmatrix}^{-1}}_{\text{preconditioner}} \begin{bmatrix} \mathbf{u}' \\ \mathbf{p}' \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}$$

Approximating the inverse, $\tilde{\mathbf{A}}^{-1} \approx \mathbf{A}^{-1}$, is well suited for multigrid.

Inverse Schur complement approximation, $\tilde{\mathbf{S}}^{-1} \approx \mathbf{S}^{-1} := (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^\top)^{-1}$, with improved BFBT / Least Squares Commutator (LSC) method:

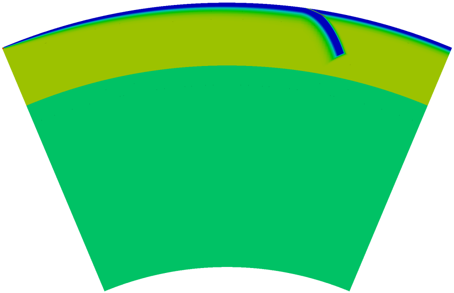
$$\tilde{\mathbf{S}}^{-1} = (\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top)^{-1}(\mathbf{B}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top)(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top)^{-1}$$

with diagonal scaling, $\mathbf{D} := \mathrm{diag}(\mathbf{A})$. Here, approximating the inverse of the discrete pressure Laplacian, $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top)$, is well suited for multigrid.

Stokes solver robustness with scaled BFBT Schur complement approximation

Stokes solver robustness with scaled BFBT Schur complement approximation

The subducting plate model problem on a cross section of the spherical Earth domain serves as a benchmark for solver robustness.



Subduction model viscosity field.

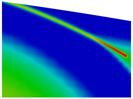
Multigrid parameters:

GMG for $\tilde{\mathbf{A}}$: 1 V-cycle, 3+3 smooth

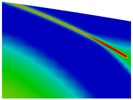
AMG (PETSc's GAMG) for $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^{\top})$: 3 V-cycles, 3+3 smooth

Robustness with respect to plate coupling strength

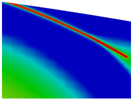
10^{-4}
(mild decoupling)

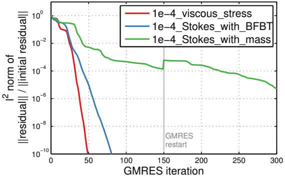


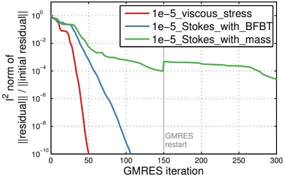
10^{-5}

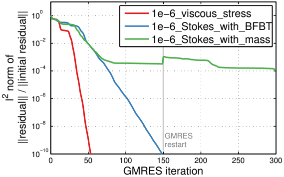


10^{-6}
(strong decoupling)





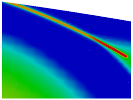




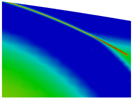
Convergence for solving $\mathbf{A}\mathbf{u} = \mathbf{f}$ (red), Stokes system with BFBT (blue), Stokes system with viscosity weighted mass matrix as Schur complement approximation (green) for comparison to conventional preconditioning.

Robustness with respect to plate boundary thickness

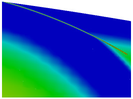
10 km

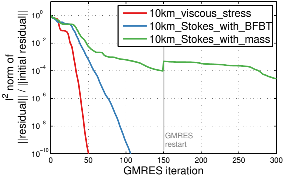


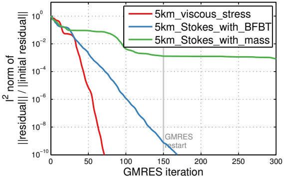
5 km

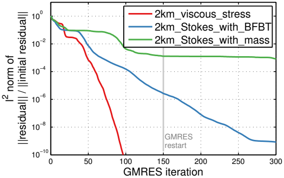


2 km









Convergence for solving $\mathbf{Au} = \mathbf{f}$ (red), Stokes system with BFBT (blue), Stokes system with viscosity weighted mass matrix as Schur complement approximation (green) for comparison to conventional preconditioning.

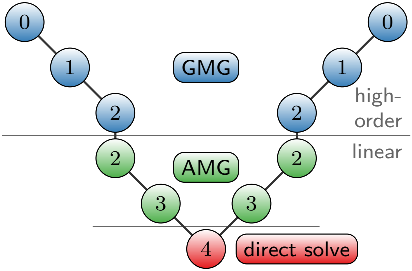
Parallel adaptive high-order geometric multigrid

Parallel adaptive high-order geometric multigrid

The multigrid hierarchy of nested meshes is generated from an adaptively refined octree-based mesh via geometric coarsening:

- Parallel repartitioning of coarser meshes for load-balancing; repartitioning of sufficiently coarse meshes on subsets of cores
- High-order L^2-projection of coefficients onto coarser levels; re-discretization of differential eqn's at coarser geometric multigrid levels

Multigrid hierarchy of viscous stress $\tilde{\mathbf{A}}$



Multigrid for pressure Laplacian:

Geometric multigrid for the pressure Laplacian is problematic due to the discontinuous modal pressure discretization $\mathbb{P}_{k-1}^{\text{disc}}$.

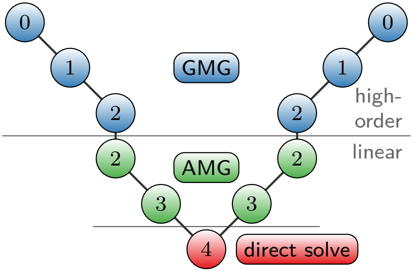
Here, a novel approach is taken by re-discretizing with continuous nodal $\mathbb{Q}_k$ basis functions.

Parallel adaptive high-order geometric multigrid

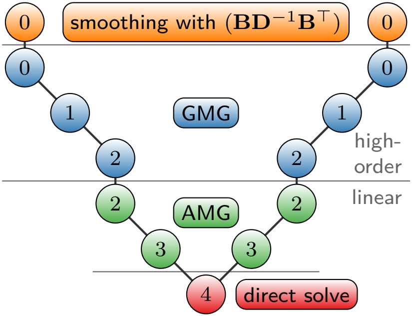
The multigrid hierarchy of nested meshes is generated from an adaptively refined octree-based mesh via geometric coarsening:

- Parallel repartitioning of coarser meshes for load-balancing; repartitioning of sufficiently coarse meshes on subsets of cores
- High-order L^2-projection of coefficients onto coarser levels; re-discretization of differential eqn's at coarser geometric multigrid levels

Multigrid hierarchy of viscous stress $\tilde{\mathbf{A}}$



Multigrid hierarchy of pressure Laplacian



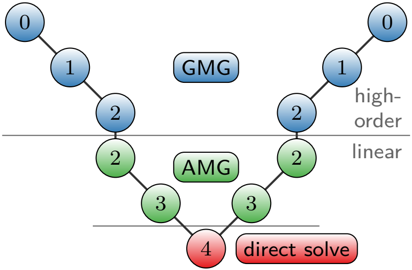
Parallel adaptive high-order geometric multigrid

GMG smoother: Chebyshev accelerated Jacobi (PETSc) with matrix-free high-order stiffness apply, assembly of high-order diagonal only.

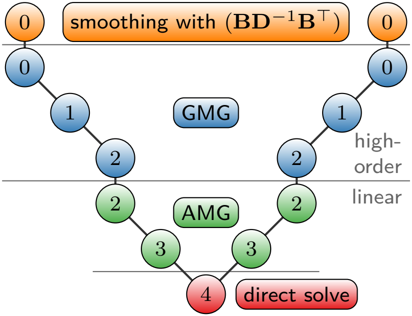
GMG restriction & interpolation: High-order L^2-projection; restriction and interpolation operators are adjoints of each other in L^2-sense.

No collective communication in GMG cycles needed; as the coarse solver for GMG, AMG (PETSc's GAMG) is invoked on only small core counts.

Multigrid hierarchy of viscous stress $\tilde{\mathbf{A}}$



Multigrid hierarchy of pressure Laplacian



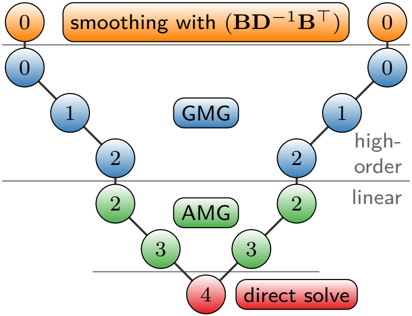
Parallel adaptive high-order geometric multigrid

GMG smoother for $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^{\top})$, discontinuous modal: Chebyshev accelerated Jacobi (PETSc) with matrix-free apply and assembled diagonal.

GMG restriction & interpolation for $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^{\top})$: L^2-projection between discontinuous modal and continuous nodal spaces.

No collective communication in GMG cycles needed; as the coarse solver for GMG, AMG (PETSc's GAMG) is invoked on only small core counts.

Multigrid hierarchy of pressure Laplacian

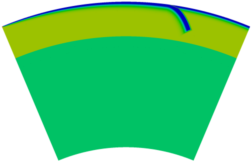


Convergence dependence on mesh size and discretization order

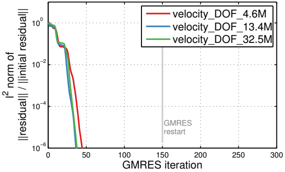
h-dependence using geometric multigrid for $\tilde{\mathbf{A}}$ and $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^{\top})$

The mesh is increasingly refined while the discretization stays fixed to $\mathbb{Q}_2 \times \mathbb{P}_1^{\text{disc}}$.

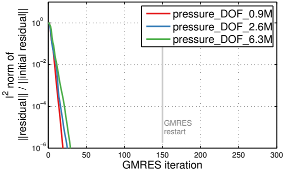
(Multigrid parameters:
GMG for $\tilde{\mathbf{A}}$: 1 V-cycle, 3+3 smoothing;
GMG for $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^{\top})$: 1 V-cycle, 3+3 smoothing)



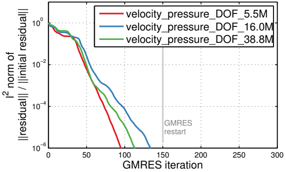
Solve $\mathbf{A}\mathbf{u} = \mathbf{f}$



Solve $\left(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^{\top}\right)\mathbf{p} = \mathbf{g}$

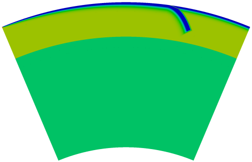


Solve Stokes system

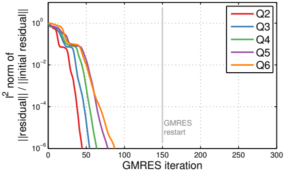


p-dependence using geometric multigrid for $\tilde{\mathbf{A}}$ and $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top)$

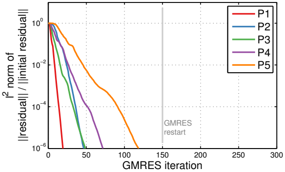
The discretization order of the finite element space increases while the mesh stays fixed.
(Multigrid parameters:
GMG for $\tilde{\mathbf{A}}$: 1 V-cycle, 3+3 smoothing;
GMG for $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top)$: 1 V-cycle, 3+3 smoothing)



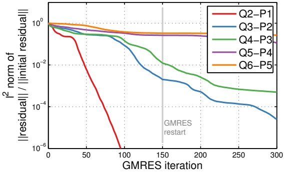
Solve $\mathbf{A}\mathbf{u} = \mathbf{f}$



Solve $\left(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top\right)\mathbf{p} = \mathbf{g}$



Solve Stokes system

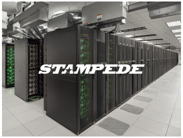


Remark: The deteriorating Stokes convergence with increasing order is due to a deteriorating approximation of the Schur complement by the BFBT method and not the multigrid components.

Parallel scalability of geometric multigrid

Global problem on adaptive mesh of the Earth

- Viscosity is generated from real Earth data
- Heterogeneous viscosity field exhibits 6 orders of magnitude variation
- Adaptively refined mesh (p4est library) down to ~0.5 km local resolution; $\mathbb{Q}_2 \times \mathbb{P}_1^{\mathrm{disc}}$ discretization
- Distributed memory parallelization (MPI)

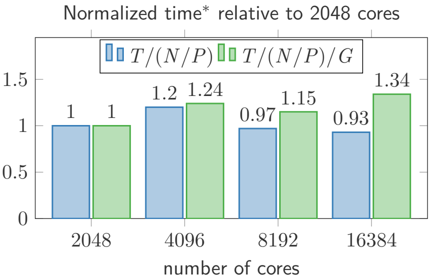


Stampede at the Texas Advanced Computing Center
16 CPU cores per node (2 × 8 core Intel Xeon E5-2680)
32GB main memory per node (8 × 4GB DDR3-1600MHz)
6,400 nodes, 102,400 cores total, InfiniBand FDR network

Weak scalability using adaptively refined Earth mesh

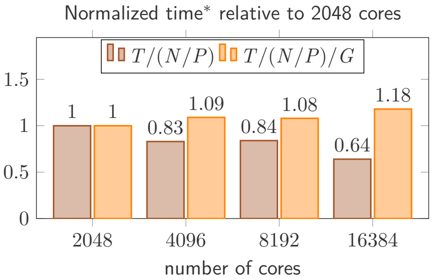
Normalized time* based on the setup and solve times for solving for velocity $\mathbf{u}$ in:

$$\mathbf{A}\mathbf{u} = \mathbf{f}$$



Normalized time* based on the setup and solve times for solving for pressure $\mathbf{p}$ in:

$$\left(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^{\top}\right)\mathbf{p} = \mathbf{g}$$



*Normalization explanation:

Scalability of algorithms & implementation: $T/(N/P)$

Scalability of implementation: $T/(N/P)/G$

T ... setup + solve time
N ... degrees of freedom (DOF)
P ... number of CPU cores
G ... number of GMRES iterations

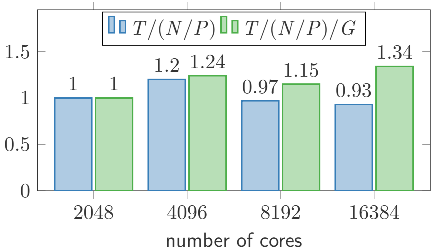
Weak scalability using adaptively refined Earth mesh

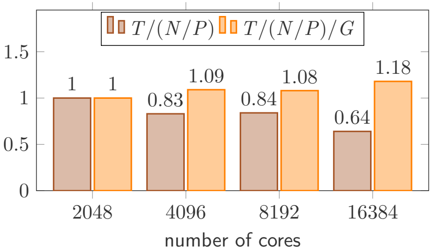
Normalized time* based on the setup and solve times for solving for velocity $\mathbf{u}$ in:

$$\mathbf{A}\mathbf{u} = \mathbf{f}$$

Normalized time* based on the setup and solve times for solving for pressure $\mathbf{p}$ in:

$$\left(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^{\top}\right)\mathbf{p} = \mathbf{g}$$





	velocity DOF	#levels geo,alg	setup time geo,alg,tot	solve time	#iter
2K	637M	7, 4	10, 14, 25	2298	402
4K	1155M	7, 4	13, 29, 41	2483	389
8K	2437M	8, 4	15, 16, 31	2130	339
16K	5371M	8, 4	29, 51, 80	2198	279

	pressure DOF	#levels geo,alg	setup time geo,alg,tot	solve time	#iter
2K	125M	7, 3	11, 1, 12	857	125
4K	227M	7, 4	12, 2, 15	638	95
8K	482M	8, 3	18, 2, 20	684	97
16K	1042M	8, 4	27, 9, 36	546	68

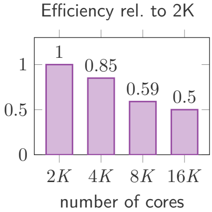
Strong scalability using fixed adaptive Earth mesh

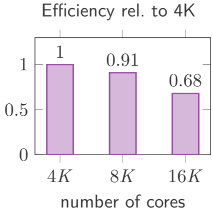
Efficiency based on the setup and solve times for solving for velocity $\mathbf{u}$ in:

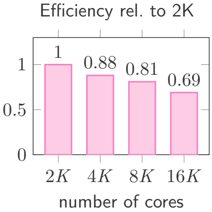
$$\mathbf{A}\mathbf{u} = \mathbf{f}$$

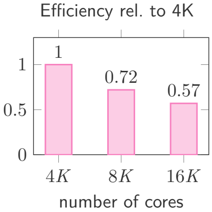
Efficiency based on the setup and solve times for solving for pressure $\mathbf{p}$ in:

$$\left(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^{\top}\right)\mathbf{p} = \mathbf{g}$$









Problem size: 637M
#iterations: 401 (± 1)

	setup time geo,alg,tot	**solve time**
2K	10, 14, 25	2298
4K	9, 16, 25	1328
8K	13, 27, 40	938
16K	11, 24, 36	545

Problem size: 1155M
#iterations: 388 (± 1)

	setup time geo,alg,tot	**solve time**
2K	—	—
4K	13, 29, 41	2483
8K	10, 39, 49	1326
16K	17, 48, 65	859

Problem size: 125M
#iterations: 125 (± 2)

	setup time geo,alg,tot	**solve time**
2K	11, 1, 12	857
4K	8, 1, 9	487
8K	8, 2, 9	256
16K	8, 2, 10	148

Problem size: 227M
#iterations: 96 (± 1)

	setup time geo,alg,tot	**solve time**
2K	—	—
4K	12, 2, 15	638
8K	17, 10, 27	431
16K	14, 4, 18	269

Scalable nonlinear Stokes solver: Inexact Newton-Krylov method

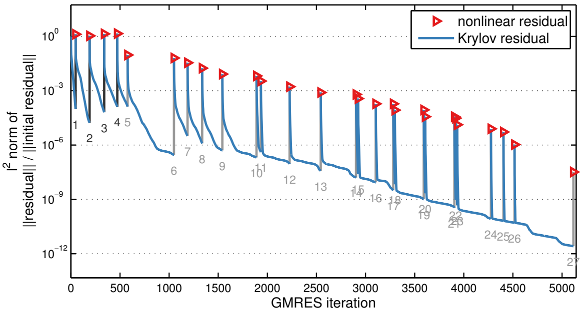
Inexact Newton-Krylov method

Newton update $(\tilde{\boldsymbol{u}}, \tilde{p})$:
$$-\nabla \cdot \left[\mu'(T, \boldsymbol{u})\left(\nabla\tilde{\boldsymbol{u}} + \nabla\tilde{\boldsymbol{u}}^\top\right)\right] + \nabla\tilde{p} = -\boldsymbol{r}_{\mathrm{mom}}$$
$$\nabla \cdot \tilde{\boldsymbol{u}} = -r_{\mathrm{mass}}$$

- Newton update is computed inexactly via Krylov subspace iterative method
- Krylov tolerance decreases with subsequent Newton steps to guarantee superlinear convergence
- Number of Newton steps is independent of the mesh size
- Velocity residual is measured in H^{-1}-norm for backtracking line search; this avoids overly conservative update steps $\ll 1$
- Grid continuation at initial Newton steps: Adaptive mesh refinement to resolve increasing viscosity variations arising from the nonlinear dependence on the velocity

Inexact Newton-Krylov method

Convergence of inexact Newton-Krylov (4096 cores)



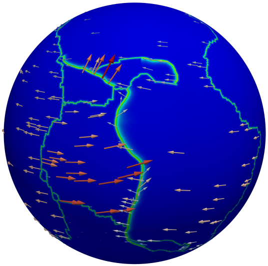


Plate velocities at nonlinear solution.

Adaptive mesh refinements after the first four Newton steps are indicated by black vertical lines. 642M velocity & pressure DOF at solution, 473 min total runtime on 4096 cores.

Thank you

Summary of main results

I. Efficient methods/algorithms

- High-order finite elements
- Adaptive meshes, resolving viscosity variations
- Geometric multigrid (GMG) preconditioners for elliptic operators
- Novel GMG based BFBT/LSC pressure Schur complement preconditioner
- Inexact Newton-Krylov method
- H^{-1}-norm for velocity residual in Newton line search

II. Scalable parallel implementation

- Matrix-free stiffness/mass application and GMG smoothing
- Tensor product structure of finite element shape functions
- Octree algorithms for handling adaptive meshes in parallel
- Algebraic multigrid (AMG) only as coarse solver for GMG avoids full AMG setup cost and large matrix assembly
- Parallel scalability results up to 16,384 CPU cores (MPI)