

# Scalable Parallel Solvers for Highly Heterogeneous Nonlinear Stokes Flow Discretized with Adaptive High-Order Finite Elements

Johann Rudi<sup>1</sup>, Tobin Isaac<sup>1</sup>, Georg Stadler<sup>2</sup>, Michael Gurnis<sup>3</sup>, and Omar Ghattas<sup>1,4</sup>

<sup>1</sup>Institute for Computational Engineering and Sciences (ICES), The University of Texas at Austin, USA

<sup>2</sup>Courant Institute of Mathematical Sciences, New York University, USA

<sup>3</sup>Seismological Laboratory, California Institute of Technology, USA

<sup>4</sup>Jackson School of Geosciences and Department of Mechanical Engineering, The University of Texas at Austin, USA

## Summary of main results

- Geometric multigrid (GMG) for preconditioning Stokes systems
- Novel GMG based BFBT/LSC pressure Schur complement preconditioner
- Repartitioning on coarse GMG levels for load-balancing and MPI communicator reduction
- Algebraic multigrid (AMG) as coarse solver for GMG avoids full AMG setup cost and large matrix assembly
- High-order finite elements
- Adaptive meshes resolving heterogeneous viscosity with variations of up to 6 orders of magnitude
- Octree algorithms for handling adaptive meshes in parallel
- Parallel scalability results on up to 16,384 CPU cores (MPI)
- Inexact Newton-Krylov method for highly nonlinear rheology
- Global-scale simulation of Earth's mantle flow

## 1. Earth mantle flow

### Model equations for mantle convection with plate tectonics

Rock in the mantle moves like a viscous, incompressible fluid on time scales of millions of years. From conservation of mass and momentum, we obtain that the instantaneous flow velocity can be modeled as a **nonlinear Stokes** system.

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

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

$\mathbf{u}$  ... velocity  
 $p$  ... pressure  
 $T$  ... temperature  
 $\mu$  ... viscosity

The right-hand side forcing  $\mathbf{f}$  is derived from the Boussinesq approximation and depends on the temperature. The viscosity  $\mu$  depends exponentially on the temperature (via an Arrhenius relationship), on a power of the second invariant of the strain rate tensor, incorporates plastic yielding, and lower/upper bounds.

$$\mu(T, \mathbf{u}) = \max \left( \mu_{\min}, \min \left( \frac{\tau_{\text{yield}}}{2\dot{\varepsilon}(\mathbf{u})}, w \min \left( \mu_{\max}, a(T) \dot{\varepsilon}(\mathbf{u})^{\frac{1-n}{n}} \right) \right) \right)$$

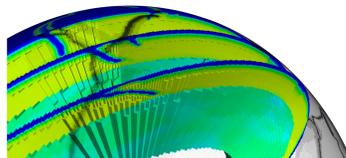
with exponentially on temperature dependent factor  $a(T)$ , plate decoupling  $w(\mathbf{x})$ , viscosity bounds  $0 < \mu_{\min} < \mu_{\max}$ , yielding stress  $0 < \tau_{\text{yield}}$ , exponent  $n \approx 3$ , and square root of the 2<sup>nd</sup> invariant of the strain rate tensor  $\dot{\varepsilon}(\mathbf{u})$ .

### Central open questions

- Strength of plate coupling & amount of energy dissipation in hinge zones
- Role of subducting slab geometries
- Accuracy of rheology extrapolations derived from laboratory experiments
- Main drivers of plate motion; negative buoyancy forces or convective shear traction?

### Research target

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



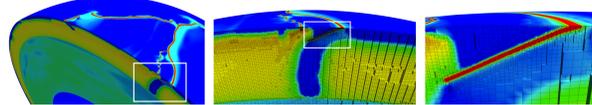
## 2. Solver challenges of global-scale mantle flow

### Inherent challenges of realistic Earth mantle flow simulations:

- Severe **nonlinearity**, **heterogeneity**, and **anisotropy** of the Earth's rheology with a wide range of spatial scales
- Highly localized features** with respect to Earth's radius (~6371 km), like plate thickness ~50 km and shearing zones at plate boundaries ~5 km
- 6 orders of magnitude** viscosity contrast within ~5 km thin plate boundaries

### Highly accurate numerical simulations require:

- Resolution down to ~1 km** at plate boundaries (uniform mesh of Earth's mantle would result in computationally prohibitive  $O(10^{12})$  degrees of freedom). Enabled by: **adaptive mesh refinement**
- Velocity approximation with high accuracy and local mass conservation. Enabled by: **high-order discretizations**



Effective viscosity field and adaptive mesh resolving narrow plate boundaries (in red). Visualization by L. Allics.

## 3. Scalable Stokes solver

### High-order finite element discretization of the Stokes system

$$\begin{cases} -\nabla \cdot [\mu (\nabla \mathbf{u} + \nabla \mathbf{u}^T)] + \nabla p = \mathbf{f} \\ \nabla \cdot \mathbf{u} = 0 \end{cases} \quad \text{discretize with high-order FE} \quad \begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ p \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}$$

- 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 to greatly reduce the number of floating point operations

### Linear solver: Preconditioned Krylov subspace method

Coupled iterative solver: GMRES with upper triangular block preconditioning

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{B}^T \\ \mathbf{0} & -\tilde{\mathbf{S}} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{u} \\ p \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}$$

Stokes operator preconditioner

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

### BFBT/LSC Schur complement approximation $\tilde{\mathbf{S}}^{-1}$

Improved BFBT / Least Squares Commutator (LSC) method:

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

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

### Derivation: Consider the least squares problem of a commutation relationship

$$\text{Find minimizing matrix } \mathbf{X} \text{ for: } \min_{\mathbf{X}} \|\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^T \mathbf{e}_j - \mathbf{B}^T \mathbf{X} \mathbf{e}_j\|_{C^{-1}}^2 \text{ for all } j,$$

where matrix  $\mathbf{C}$  is s.p.d., matrix  $\mathbf{D}$  is invertible but arbitrary for now, and  $\mathbf{e}_j$  is the  $j$ -th unit vector. The solution  $\mathbf{X} = (\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T)^{-1}(\mathbf{B}\mathbf{C}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^T)$  gives a  $\mathbf{C}^{-1}$ -orthogonal projection, i.e.,

$$(\mathbf{B}^T \mathbf{e}_i, (\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^T - \mathbf{B}^T \mathbf{X}) \mathbf{e}_j)_{C^{-1}} = 0 \text{ for all } i, j.$$

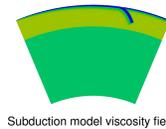
From the choice  $\mathbf{C}^{-1} = \tilde{\mathbf{A}}^{-1}$ , e.g., a multigrid V-cycle, we obtain

$$(\mathbf{B}^T \mathbf{e}_i, (\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^T - \mathbf{B}^T \mathbf{X}) \mathbf{e}_j)_{\tilde{\mathbf{A}}^{-1}} = 0 \text{ for all } i, j \Leftrightarrow \tilde{\mathbf{S}} = \mathbf{B}\tilde{\mathbf{A}}^{-1}\mathbf{B}^T,$$

which represents an optimal preconditioner for the right-preconditioned discrete Stokes system. A computationally feasible choice is  $\mathbf{C}^{-1} = \mathbf{D}^{-1} = \text{diag}(\mathbf{A})^{-1}$ .

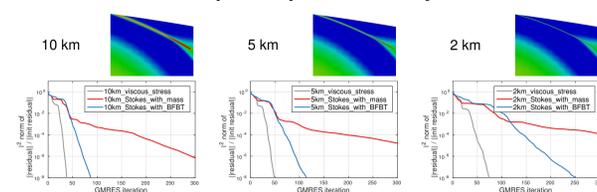
## 4. Stokes solver robustness with scaled BFBT Schur complement approximation

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



Multigrid parameters: GMG for  $\tilde{\mathbf{A}}$ : 1 V-cycle, 3+3 smoothing; GMG for  $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T)$ : 1 V-cycle, 3+3 smoothing, and additional 6+6 smoothing in discontinuous, modal pressure space.

### Robustness with respect to plate boundary thickness



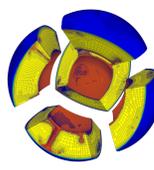
Convergence for solving  $\mathbf{A}\mathbf{u} = \mathbf{f}$  (gray), Stokes system with BFBT (blue), Stokes system with viscosity weighted mass matrix as Schur complement approximation (red) for comparison to conventional preconditioning.

## 5. Parallel octree-based adaptive mesh refinement

Idea: Identify octree leaves with hexahedral elements.

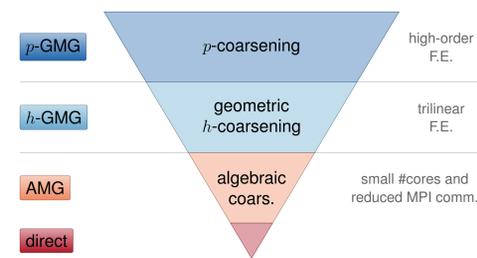
- Octree structure enables fast parallel adaptive octree/mesh refinement and coarsening
- Octrees and space filling curves enable fast neighbor search, repartitioning, and 2:1 balancing in parallel
- Algebraic constraints on non-conforming element faces with hanging nodes enforce global continuity of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

### p4est library



## 6. Parallel adaptive high-order geometric multigrid

The hybrid multigrid hierarchy: Coarsen **adaptive octree-based mesh**



### Geometric multigrid method: p-GMG and h-GMG

- Parallel repartitioning of coarser h-GMG meshes is important to maintain load-balancing of the adaptive meshes
- Sufficiently coarse meshes are repartitioned on **subsets of cores**, the MPI communicator is reduced to the nonempty cores
- High-order  $L^2$ -projection of coefficients onto coarser levels
- Re-discretization of differential equations at each coarser p- and h-GMG level
- Smoother: Chebyshev accelerated Jacobi (PETScs) with **matrix-free** differential operator-apply functions; avoiding full matrix assembly
- 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

### Coarse solver for geometric multigrid: AMG, PETSc's GAMG

- Coarse problems use only **small core counts**, usually  $O(100)$
- The **MPI communicator is reduced** to the nonempty cores

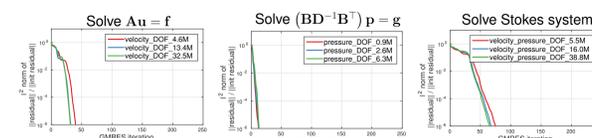
### GMG for $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T)$ on discontinuous, modal pressure space

- Novel approach:** Re-discretize the underlying variable coefficient Laplace operator with continuous, nodal high-order finite elements in  $\mathbb{Q}_k$ .
- Coefficient of Laplace operator is derived from diagonal scaling  $\mathbf{D}^{-1}$
- Apply GMG as described above to the continuous, nodal  $\mathbb{Q}_k$  re-discretization of the pressure Laplace operator
- On finest level, additionally apply smoother in the space  $\mathbb{P}_{k-1}^{\text{disc}}$

## 7. Convergence dependence on mesh size and discretization order

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

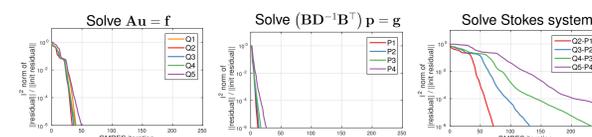
The mesh is increasingly refined while the discretization stays fixed to  $\mathbb{Q}_2 \times \mathbb{P}_1^{\text{disc}}$ . Performed with subducting plate model problem (see above).



Multigrid parameters: GMG for  $\tilde{\mathbf{A}}$ : 1 V-cycle, 3+3 smoothing; GMG for  $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T)$ : 1 V-cycle, 3+3 smoothing, and additional 6+6 smoothing in discontinuous, modal pressure space.

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

The discretization order of the finite element space increases while the mesh stays fixed. Performed with subducting plate model problem (see above).



Multigrid parameters: GMG for  $\tilde{\mathbf{A}}$ : 1 V-cycle, 3+3 smoothing; GMG for  $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T)$ : 1 V-cycle, 3+3 smoothing, and additional 6+6 smoothing in discontinuous, modal pressure space.

**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.

## 8. Parallel scalability of geometric multigrid

Global problem on adaptively refined mesh of the Earth's mantle

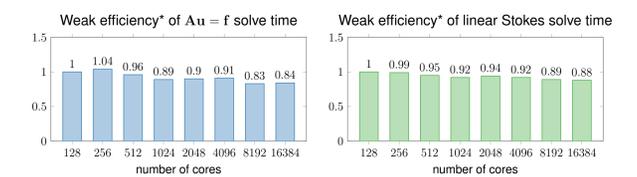
- Locally refined mesh with up to 6 refinement levels difference
- $\mathbb{Q}_2 \times \mathbb{P}_1^{\text{disc}}$  discretization
- Constant AMG setup time throughout all core counts, accounting for < 10 percent of total setup



### Stampede at the Texas Advanced Computing Center

16 CPU cores per node (2 x 8 core Intel Xeon E5-2680)  
32GB main memory per node (8 x 4GB DDR3-1600MHz)  
1,024 nodes or 16,384 cores used for scalability (MPI)

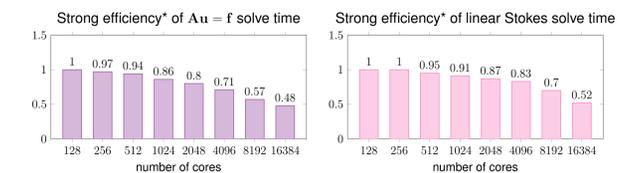
### Weak scalability with increasingly locally refined Earth mesh



Detailed timings for solving $\mathbf{A}\mathbf{u} = \mathbf{f}$				Detailed timings for solving linear Stokes system			
#cores	velocity DOF	setup time (s)	solve time (s)	#cores	total DOF velocity-pressure	setup time (s)	solve time (s)
128	21M	0.3, 3.0	64.7	128	25M	6.4	256.1
256	42M	0.5, 3.3	62.5	256	50M	7.5	258.7
512	82M	0.5, 3.8	65.1	512	97M	7.3	262.1
1024	162M	0.6, 4.6	69.4	1024	191M	8.1	269.1
2048	329M	0.3, 5.3	69.4	2048	386M	9.6	266.0
4096	664M	0.5, 8.0	69.8	4096	782M	11.2	274.1
8192	1333M	0.7, 12.9	76.6	8192	1567M	17.6	284.2
16384	2668M	0.3, 21.6	76.1	16384	3131M	26.1	287.2

\*Weak efficiency baseline is 128 cores

### Strong scalability with a fixed locally refined Earth mesh



\*Strong efficiency baseline is 128 cores

## 9. Scalable nonlinear Stokes solver: Inexact Newton-Krylov method

Newton update  $(\tilde{\mathbf{u}}, \tilde{p})$  is computed as the inexact solution of

$$-\nabla \cdot \left[ \left( \mu \mathbf{I} + \varepsilon \frac{\partial \mu}{\partial \dot{\varepsilon}} \frac{\partial \mu}{\partial \dot{\varepsilon}} \otimes (\nabla \mathbf{u} + \nabla \mathbf{u}^T) \right) (\nabla \tilde{\mathbf{u}} + \nabla \tilde{\mathbf{u}}^T) \right] + \nabla \tilde{p} = -\mathbf{r}_{\text{nonlin}}, \quad \nabla \cdot \tilde{\mathbf{u}} = -r_{\text{nonlin}}$$

- Krylov tolerance for the inexact update computation decreases with subsequent Newton steps to achieve 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  $\ll 1$  (evaluation of residual norm requires 3 scalar constant coefficient Laplace solves, which are performed by PCG with GMG preconditioning)
- Grid continuation at initial Newton steps: Adaptive mesh refinement to resolve increasing viscosity variations arising from the nonlinear dependence on the velocity

Convergence of inexact Newton-Krylov (16,384 cores)

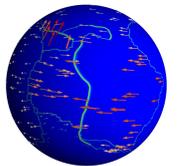
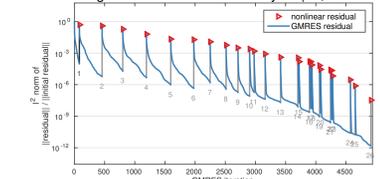


Plate velocities at nonlinear solution.

Adaptive mesh refinement after the first Newton step is indicated by **black vertical line**. 2.3B velocity & pressure DOF at solution, 459 min total runtime on 16,384 cores.