# Plate Boundary-Resolving Nonlinear Global Mantle Flow Simulations Using Parallel High-Order Geometric Multigrid Methods on Adaptive Meshes

Johann Rudi[1] Hari Sundar[2] Tobin Isaac[1] Georg Stadler[1]
Michael Gurnis[3] Omar Ghattas[14]

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

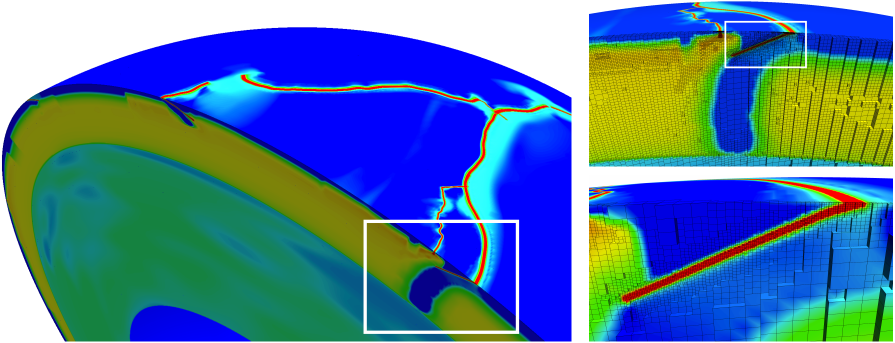
[2]School of Computing, The University of Utah, USA

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

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

# The global mantle flow problem

## A multiscale problem, globally coupled, locally high resolution



(Visualization by L. Alisic)

# Main results summary

Two essential components for scalable mantle flow solvers:

## I. Efficient methods/algorithms

- high-order finite elements
- adaptive meshes, resolving viscosity variations
- inexact Newton-Krylov method
- $H^{-1}$-norm for velocity comp. for Newton line search
- multigrid preconditioners for elliptic operators
- BFBT/LSC type pressure Schur complement preconditioner

## II. Scalable parallel implementation

- matrix-free stiffness/mass application
- tensor product structure of FE shape functions
- octree algorithms for handling adaptive mesh in parallel
- high-order GMG with linear AMG as coarse solve
- AMG on sparsified matrix using trilinear FE at high-order nodes
- scalability up to 16384 cores

# Main results covered in this talk

Two essential components for scalable mantle flow solvers:

## I. Efficient methods/algorithms

- high-order finite elements
- adaptive meshes, resolving viscosity variations
- inexact Newton-Krylov method
- $H^{-1}$-norm for velocity comp. for Newton line search
- multigrid preconditioners for elliptic operators
- BFBT/LSC type pressure Schur complement preconditioner

## II. Scalable parallel implementation

- matrix-free stiffness/mass application
- tensor product structure of FE shape functions
- octree algorithms for handling adaptive mesh in parallel
- high-order GMG with linear AMG as coarse solve
- AMG on sparsified matrix using trilinear FE at high-order nodes
- scalability up to 16384 cores

# Design of a plate tectonics & mantle convection benchmark problem

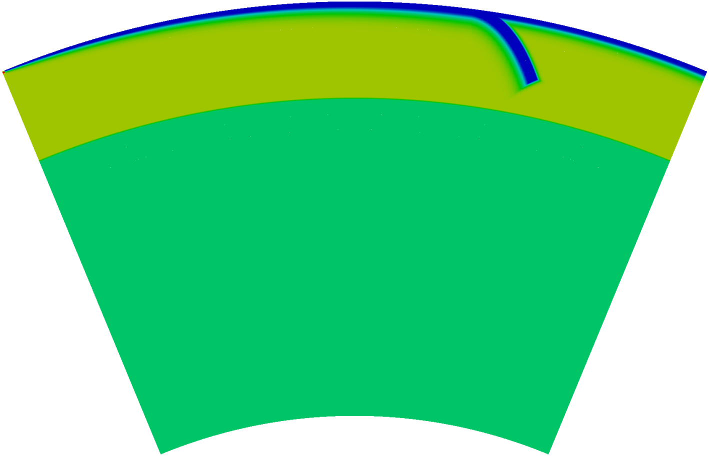
# Plate tectonics & mantle convection benchmark problem

## Setup and major challenges

1. Curved slice domain
2. Temperature of lithosphere derived from half-space cooling model
3. Viscosity depends exponentially on temperature
   $\rightsquigarrow$ contrast of $> 10^{15}$ $\rightsquigarrow$ viscosity bounds s.t. $\mu_{\mathrm{max}}/\mu_{\mathrm{min}} \leq 10^6$
4. Viscosity contrast with bounds:
   lithosphere $\xrightarrow{\times 10^{-4}}$ asthenosphere $\xrightarrow{\times 10^{2}}$ lower mantle

# Plate tectonics & mantle convection benchmark problem

lithosphere $\xrightarrow{\times 10^{-4}}$ asthenosphere $\xrightarrow{\times 10^{2}}$ lower mantle



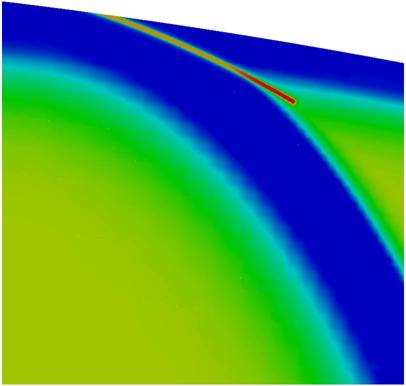
# Plate tectonics & mantle convection benchmark problem

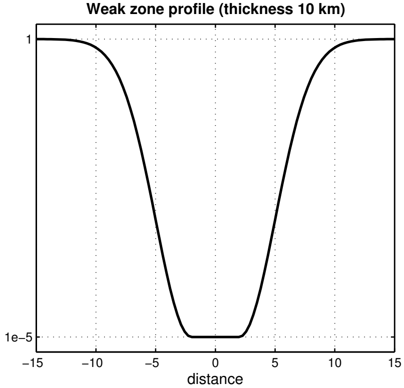
## Setup and major challenges

1. Curved slice domain
2. Temperature of lithosphere derived from half-space cooling model
3. Viscosity depends exponentially on temperature
   $\rightsquigarrow$ contrast of $> 10^{15}$ $\rightsquigarrow$ viscosity bounds s.t. $\mu_{\mathrm{max}}/\mu_{\mathrm{min}} \leq 10^6$
4. Viscosity contrast with bounds:
   lithosphere $\xrightarrow{\times 10^{-4}}$ asthenosphere $\xrightarrow{\times 10^{2}}$ lower mantle
5. Decouple plates by weak zone factor at plate boundaries;
   gives $10^6$ viscosity contrast and sharp gradients

# Plate tectonics & mantle convection benchmark problem

## decouple plates by weak zone factor at plate boundaries





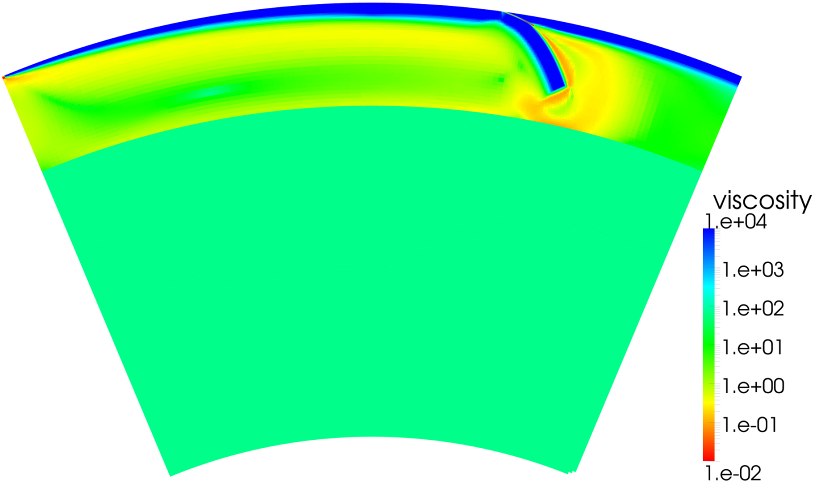
# Plate tectonics & mantle convection benchmark problem

## Setup and major challenges

1. Curved slice domain
2. Temperature of lithosphere derived from half-space cooling model
3. Viscosity depends exponentially on temperature
   $\rightsquigarrow$ contrast of $> 10^{15}$ $\rightsquigarrow$ viscosity bounds s.t. $\mu_{\max}/\mu_{\min} \leq 10^6$
4. Viscosity contrast with bounds:
   lithosphere $\xrightarrow{\times 10^{-4}}$ asthenosphere $\xrightarrow{\times 10^{2}}$ lower mantle
5. Decouple plates by weak zone factor at plate boundaries;
   gives $10^6$ viscosity contrast and sharp gradients
6. No outflow boundary conditions and right-hand side
7. Highly nonlinear rheology: strain rate dependent viscosity (power law), yielding at high stresses

# Plate tectonics & mantle convection benchmark problem

## strain rate dependent viscosity, yielding at high stresses



# The nonlinear Stokes system

# Nonlinear Stokes model PDE for mantle flow with plates

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 the nonlinear Stokes system for velocity and pressure:

$$\begin{cases} -\nabla \cdot \left[ \mu(T, \boldsymbol{u})(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^\top) \right] + \nabla p = \mathrm{Ra}\,(T - T_0)\boldsymbol{e}_r \\ \nabla \cdot \boldsymbol{u} = 0 \end{cases}$$

**Variables:**

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

**Parameters:**

- $\mu(T, \boldsymbol{u})$ ... viscosity
- $T_0$ ... background temperature
- $\mathrm{Ra} \sim 10^6 - 10^9$ ... Rayleigh number
- $\boldsymbol{e}_r$ ... radial direction

# Rheology

**Nonlinear viscosity:**

viscosity = upper bound → weak zone → yielding → lower bound

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

**Given:**

- $a(T)$ ... temperature dependent viscosity factor
- $w(\boldsymbol{x})$ ... weak zone factor
- $0 < \mu_{\min} < \mu_{\max} < \infty$ ... viscosity bounds
- $0 < \tau_{\text{yield}}$ ... yielding stress
- $n \approx 3$ ... stress exponent

**Definitions:**

- $\nabla_s \boldsymbol{u} := \frac{1}{2}(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^\top)$,
- $II_{\dot{\varepsilon}}(\boldsymbol{u}) := \frac{1}{2}(\nabla_s \boldsymbol{u} : \nabla_s \boldsymbol{u})$, $\quad \dot{\varepsilon}(\boldsymbol{u}) := \sqrt{II_{\dot{\varepsilon}}(\boldsymbol{u})}$

# Linearization: The Newton step

Nonlinear Stokes PDE:

$$\begin{cases} -\nabla \cdot [2\,\mu(\dot{\varepsilon})\nabla_s \boldsymbol{u}] + \nabla p = \boldsymbol{f} \\ \nabla \cdot \boldsymbol{u} = 0 \end{cases}$$

1st order variation w.r.t. $(\boldsymbol{u}, p)$ to get Newton step $(\tilde{\boldsymbol{u}}, \tilde{p})$:

$$\begin{cases} -\nabla \cdot \left[ 2 \left( \mu(\dot{\varepsilon})\mathbf{I} + \dot{\varepsilon}\frac{\partial \mu(\dot{\varepsilon})}{\partial \dot{\varepsilon}} \frac{\nabla_s \boldsymbol{u} \otimes \nabla_s \boldsymbol{u}}{\|\nabla_s \boldsymbol{u}\|_F^2} \right) \nabla_s \tilde{\boldsymbol{u}} \right] + \nabla \tilde{p} = -\boldsymbol{r}_{\mathrm{mom}} \\ \nabla \cdot \tilde{\boldsymbol{u}} = -r_{\mathrm{mass}} \end{cases}$$

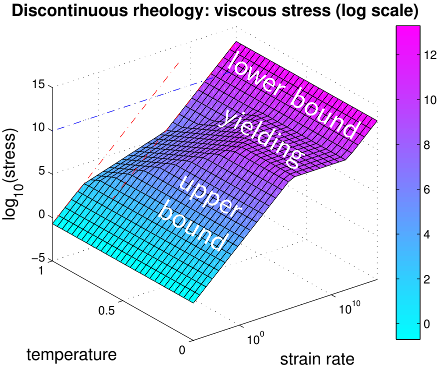
**Q:** Is 4th order tensor term bounded?. . . yes
**Q:** Is the viscosity continuously differentiable w.r.t. $\dot{\varepsilon}$ ?. . . no
Newton update:

$$(\boldsymbol{u}_{\mathsf{new}}, p_{\mathsf{new}}) = (\boldsymbol{u}, p) + \alpha(\tilde{\boldsymbol{u}}, \tilde{p})$$

**Q:** How does nonlinearity affect line search for step length $\alpha$?

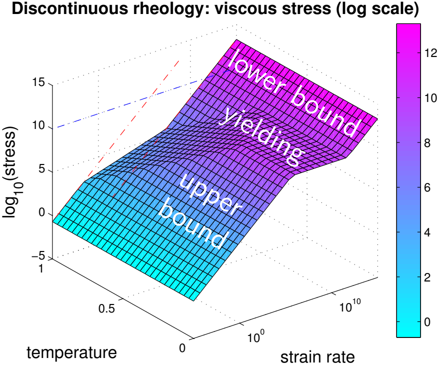
# Regularizing the rheology for Newton's method

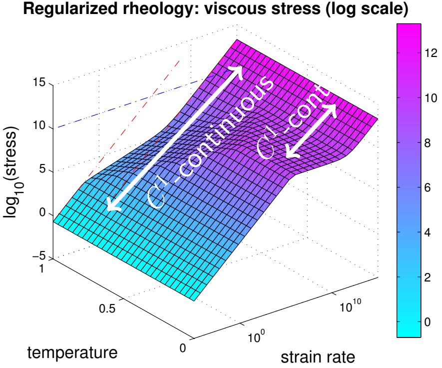
**Idea:** Construct continuously differentiable $\dot{\varepsilon} \mapsto \tau$ relationship



# Regularizing the rheology for Newton's method

**Idea:** Construct continuously differentiable $\dot{\varepsilon} \mapsto \tau$ relationship




**Create smooth transitions:**

upper bound plane $\leftrightarrow$ power law plane $\leftrightarrow$ lower bound plane

yielding plane $\leftrightarrow$ lower bound plane

# Regularizing the rheology for Newton's method

**Original viscosity:**

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

- Modify upper bound:

  find shift $d$ s.t. $\tau_d(\dot{\varepsilon}) := \begin{cases} 2\mu_{\max}\dot{\varepsilon}, & \mu_{\max} < \mu \\ 2a(T)(\dot{\varepsilon}-d)^{\frac{1}{n}}, & \text{otherwise} \end{cases}$ is $C^1$

- Modify lower bound:

  instead of cut-off: $\mu \leftarrow \max(\mu_{\min}, \mu)$, use sum: $\mu \leftarrow \mu + \mu_{\min}$

  This lower bound regularization is consistent for power law and yielding viscosity!

**Regularized viscosity:**

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

# The discrete Stokes system

# Finite element discretization of the Stokes system

$$\begin{cases} -\nabla \cdot [2\,\mu\,\nabla_s \boldsymbol{u}] + \nabla p = \boldsymbol{f} \\ \nabla \cdot \boldsymbol{u} = 0 \end{cases} \xrightarrow{\text{discretize}} \begin{bmatrix} \mathbf{A} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}$$

- Hexahedral meshes with non-conforming elements; algebraic constraints on element faces with hanging nodes enforce continuity of the global velocity basis functions
- High-order finite element shape functions
- Inf-sup stable velocity-pressure pairings: $\mathbb{Q}_k \times \mathbb{P}_{k-1}^{\mathrm{disc}}$ with $2 \leq k$
- Locally mass conservative due to discontinuous pressure space
- Fast, matrix-free application of stiffness and mass matrices
- Hexahedral elements allow for the basis functions derivatives to be calculated efficiently using tensor products

# Linear solver: Preconditioned Krylov method

Fully coupled iterative solver: GMRES with right preconditioning with an upper triangular block matrix

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

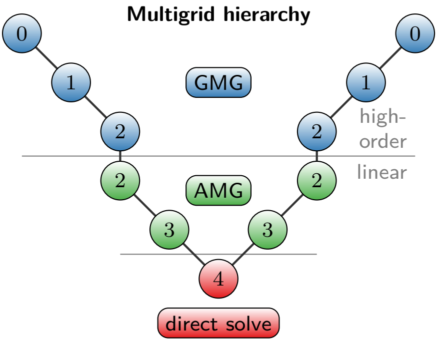
Next, we seek:

- Approximation for viscous stress block: $\tilde{\mathbf{A}}^{-1} \approx \mathbf{A}^{-1}$
- Approximation for Schur complement: $\tilde{\mathbf{S}}^{-1} \approx \mathbf{S}^{-1} := (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^\top)^{-1}$

# Viscous stress block & Schur complement block preconditioner for Krylov methods

# Hybrid geometric-algebraic multigrid for $\tilde{\mathbf{A}}^{-1}$

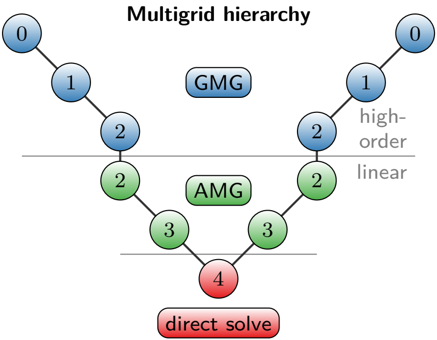
- Finest level: adaptively refined mesh
- Repartition coarser meshes for load-balancing
- Repartition onto fewer cores for small meshes
- AMG (PETSc GAMG) is only used for small problems on small core counts




- High-order $L^2$-projection of viscosity onto coarser levels (equal to discretization order)
- AMG uses sparsified matrix with trilinear FE at high-order nodes
- High-order GMG smoothing & linear AMG smoothing at transition

# Hybrid geometric-algebraic multigrid for $\tilde{\mathbf{A}}^{-1}$

- **GMG smoother:** Chebyshev accelerated Jacobi (PETSc) with matrix-free high-order stiffness apply and assembled (point-block) diagonal of high-order stiffness matrix
- **AMG smoother:** Chebyshev accelerated Jacobi (PETSc) with assembled linear matrix at high-order nodes




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

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

**Goal:** Effective and robust preconditioning of the Schur complement in Stokes systems with high viscosity variations.

Use an improved version of BFBT / Least Squares Commutator methods

$$\tilde{\mathbf{S}}^{-1} = (\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top)^{-1}(\mathbf{B}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top)(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top)^{-1}, \quad \mathbf{D} := \mathrm{diag}(\mathbf{A})$$

based on [May, Moresi, 2008].

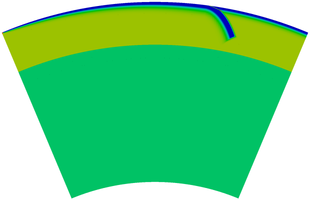
Derived from the solution of the least squares problem:

$$\min_{\mathbf{X}} \left\| \mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top \mathbf{e}_j - \mathbf{B}^\top \mathbf{X} \mathbf{e}_j \right\|_{\mathbf{D}^{-1}}^2 \text{ for all } j$$

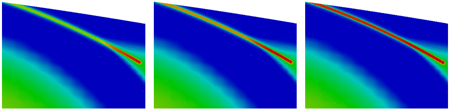
In practice, approximate $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top)^{-1}$ by AMG V-cycles with Chebyshev accelerated Jacobi smoother (PETSc).

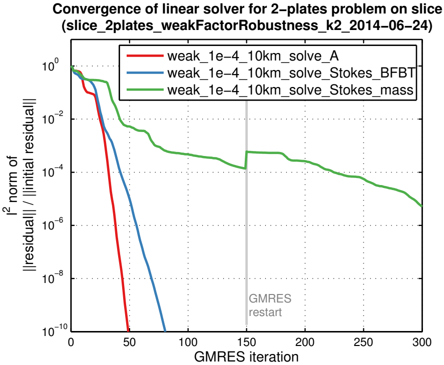
# Solver robustness

# Solver robustness w.r.t. weak zone factor



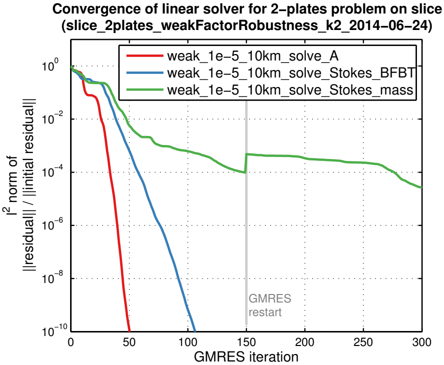
Vary only the weak zone factor:
$w_{\min} = 10^{-4}, 10^{-5}, 10^{-6}$



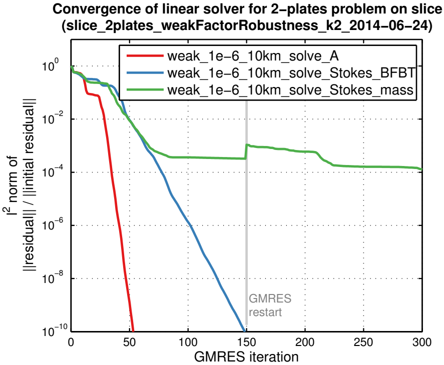



$w_{\min} = 10^{-4}$

# Solver robustness w.r.t. weak zone factor

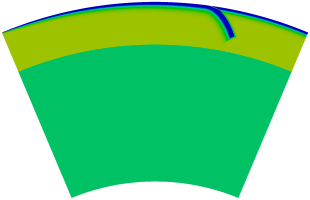



$$w_{\min} = 10^{-5}$$

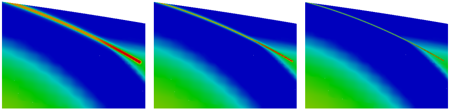


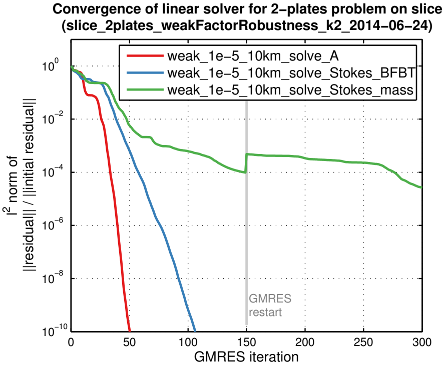

$$w_{\min} = 10^{-6}$$

# Solver robustness w.r.t. weak zone thickness



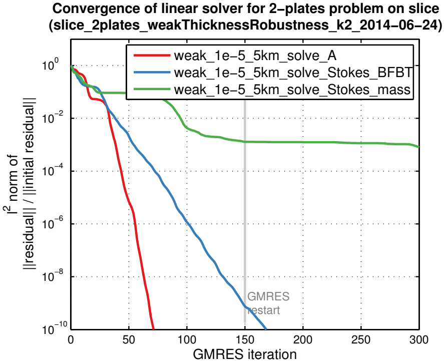
Vary only the weak zone thickness:
$10, 5, 2$ km



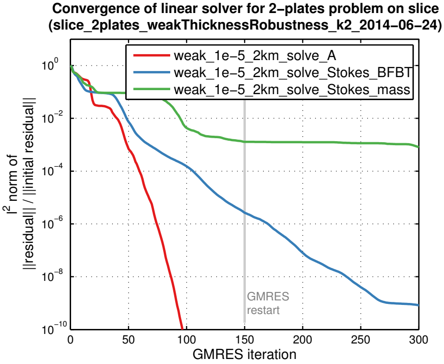



$10$ km thickness

# Solver robustness w.r.t. weak zone thickness




5 km thickness



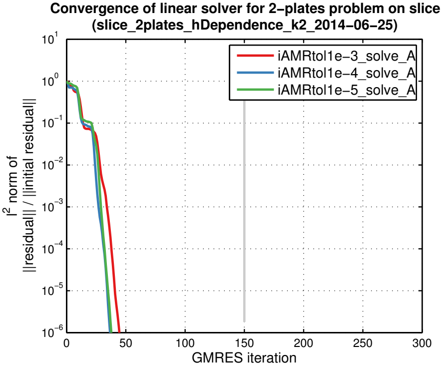

2 km thickness

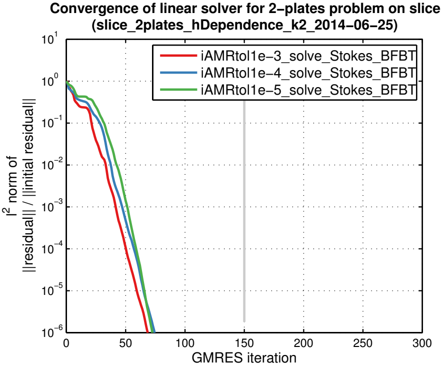
# Solver $h$-dependence and $p$-dependence

# Solver $h$-dependence

## Dependence of the number of Krylov iterations on the mesh resolution

mesh sizes: $224788$, $661980$, $1565232$; quadratic FE for velocity



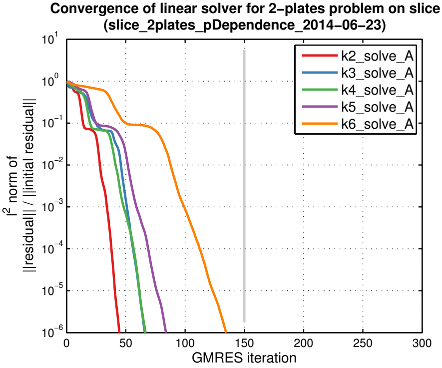

Elliptic solve: $\mathbf{A}\mathbf{u} = \mathbf{f}$

Stokes solve

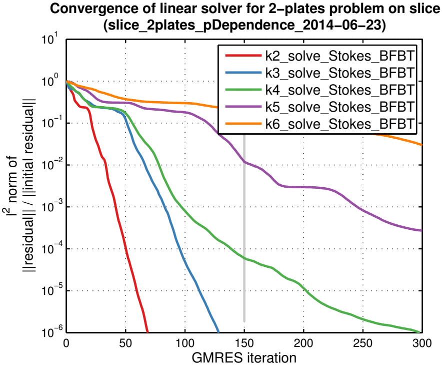
# Solver $p$-dependence

## Dependence of the number of Krylov iterations on the discretization order

FE orders for velocity discretization: $2, 3, 4, 5, 6$; mesh fixed




Elliptic solve: $\mathbf{Au} = \mathbf{f}$




Stokes solve

# Parallel scalability of GMG

# Computing environment

## Resources provided by the Texas Advanced Computing Center (TACC)

### Stampede supercomputer

- 16 CPU cores per node,
  $2 \times 8$ core Intel Xeon E5-2680
- 32GB main memory per node,
  $8 \times 4$GB DDR3-1600MHz
- 6400 nodes, 102,400 cores total
- InfiniBand FDR network

### Compiler and libraries

- Compiler: Intel 14.0.1.106
- MPI: MVAPICH2 2.0b
- Linear algebra library: PETSc 3.4.3

# Scalability problem: Elliptic solve on adaptive Earth mesh

- solve for velocity: $\mathbf{Au} = \mathbf{f}$
- adaptive mesh (p4est library) with up to $\sim 0.5$ km resolution
- quadratic FE velocity discretization
- weak zone factor $10^{-5}$, thickness $20$ km
- viscosity variation is as high as in physically realistic simulation

# Weak scaling for $\mathbf{Au} = \mathbf{f}$ on adaptive Earth mesh

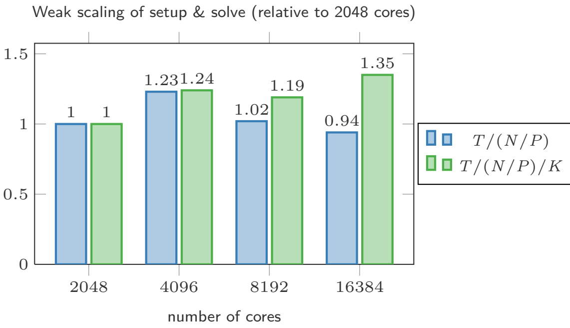
| #cores | #elems | vel DOF | #levels GMG, AMG | setup GMG, AMG, total | solve | total | #iter |
|---|---|---|---|---|---|---|---|
| 2048 | 31.4e6 | 640.1e6 | 7, 4 | 13.2, 14.4, 27.7 | 2348.6 | 2421.6 | 392 |
| 4096 | 56.8e6 | 1157.7e6 | 7, 4 | 13.3, 37.2, 50.5 | 2585.4 | 2702.0 | 389 |
| 8192 | 120.5e6 | 2440.1e6 | 8, 4 | 18.9, 43.0, 61.9 | 2241.2 | 2377.6 | 334 |
| 16384 | 260.7e6 | 5374.9e6 | 8, 4 | 34.7, 98.0, 132.7 | 2220.3 | 2488.6 | 275 |

Algorithmic & implementation scalability: $T/(N/P)$

Implementation scalability: $T/(N/P)/K$

where

- $T$ ... setup + solve time
- $N$ ... velocity DOF
- $P$ ... #cores
- $K$ ... #iterations

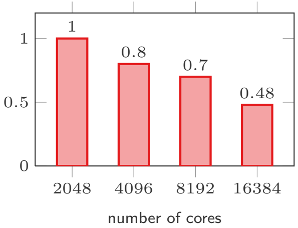


# Strong scaling for $\mathbf{Au} = \mathbf{f}$ on adaptive Earth mesh

Problem size: $N = 640.1\text{e}6$ velocity DOF

| **#cores** | **setup** GMG, AMG, total | **solve** | **total** | **#iter** |
|---|---|---|---|---|
| 2048 | 13.2, 14.4, 27.7 | 2348.6 | 2421.6 | 392 |
| 4096 | 9.9, 27.0, 36.9 | 1492.1 | 1564.7 | 400 |
| 8192 | 11.3, 41.6, 52.8 | 864.8 | 947.1 | 404 |
| 16384 | 14.9, 80.7, 95.6 | 621.0 | 746.1 | 402 |

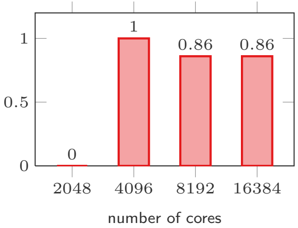
Strong scaling of solve (efficiency rel. to 2048)




Problem size: $N = 1157.7\text{e}6$ velocity DOF

| **#cores** | **setup** GMG, AMG, total | **solve** | **total** | **#iter** |
|---|---|---|---|---|
| 4096 | 13.3, 37.2, 50.5 | 2585.4 | 2702.0 | 389 |
| 8192 | 12.4, 65.6, 77.9 | 1498.5 | 1621.9 | 386 |
| 16384 | 13.7, 99,4, 113.2 | 989.6 | 1137.9 | 389 |

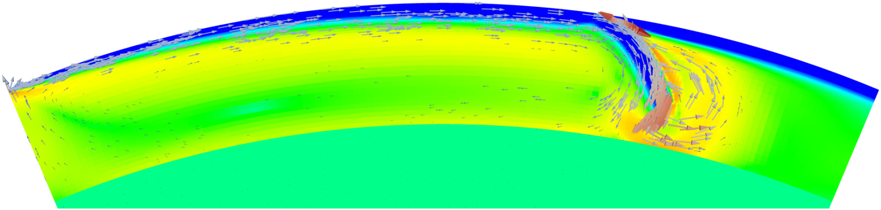
Strong scaling of solve (efficiency rel. to 4096)



# Nonlinear solver convergence

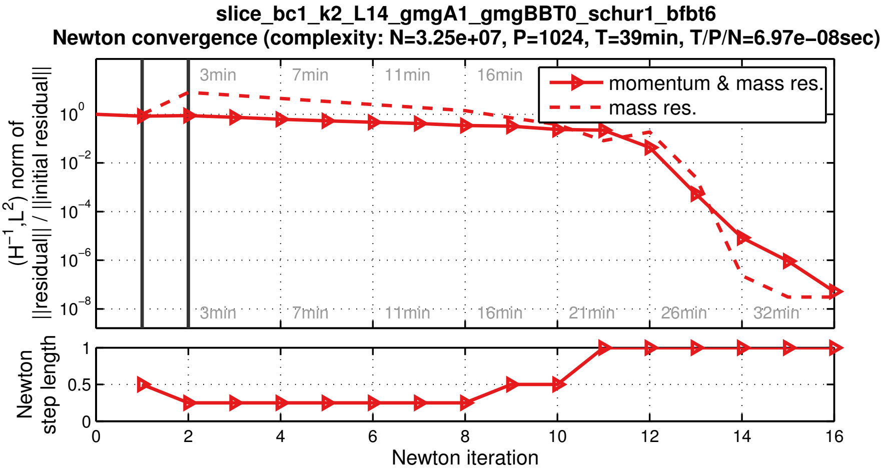
# Nonlinear convergence for slice problem

- Weak zone factor $10^{-5}$, $10$ km thickness
- Yielding at high stresses
- Adaptive mesh refinement after the first two Newton steps
- Residual measured in $H^{-1}$-norm for backtracking line search; important to not have overly conservative backtracking steps $\ll 1$ (requires 3 scalar constant coefficient Laplace solves)
- Complexity: 33M velocity & pressure DOF, 1024 processor cores, 39 min total runtime



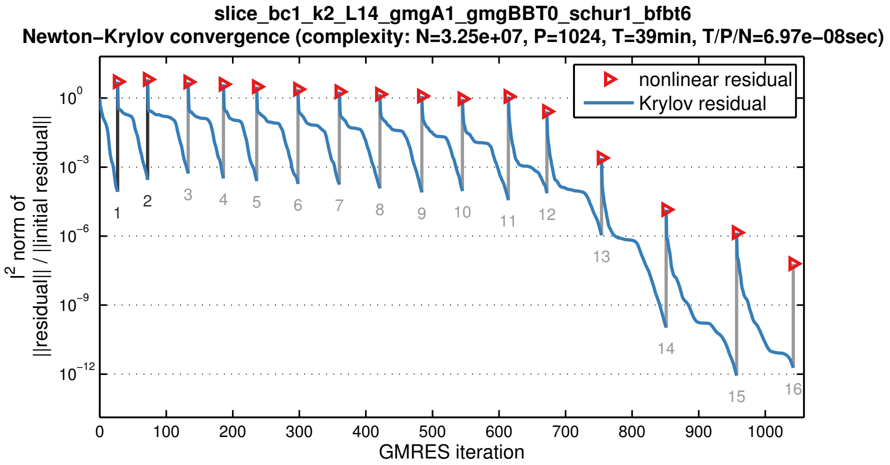
# Inexact Newton-Krylov convergence for slice problem

## Reduction of $H^{-1}$-norm of velocity residual and Newton step length



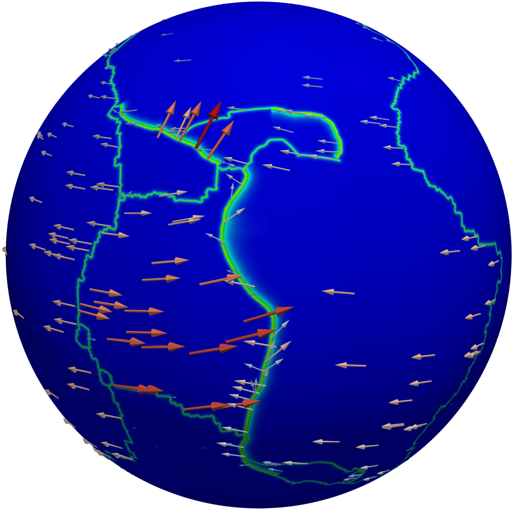
# Inexact Newton-Krylov convergence for slice problem

## Reduction of $l^2$-norm of residual at Newton and Krylov iterations



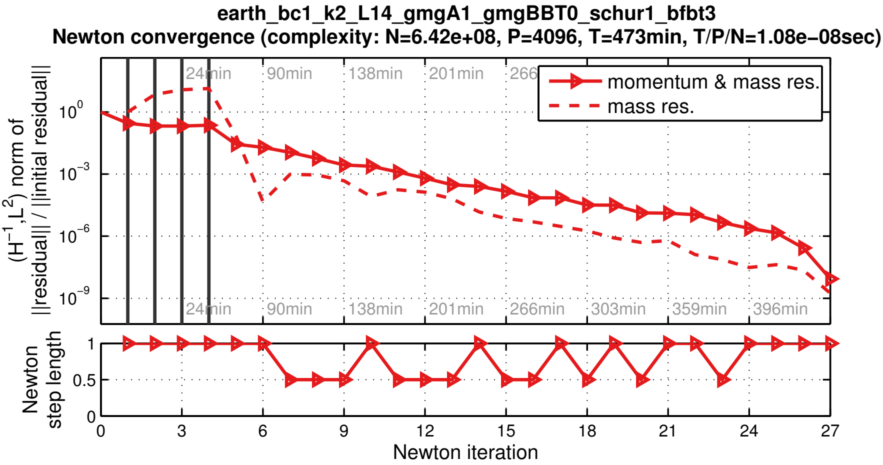
# Nonlinear convergence for Earth problem

- Weak zone factor $10^{-4}$, $100$ km thickness
- Yielding at high stresses
- AMR after the first four Newton steps
- Residual measured in $H^{-1}$-norm for backtracking line search
- Complexity: 642M velocity & pressure DOF, 4096 processor cores, 473 min total runtime



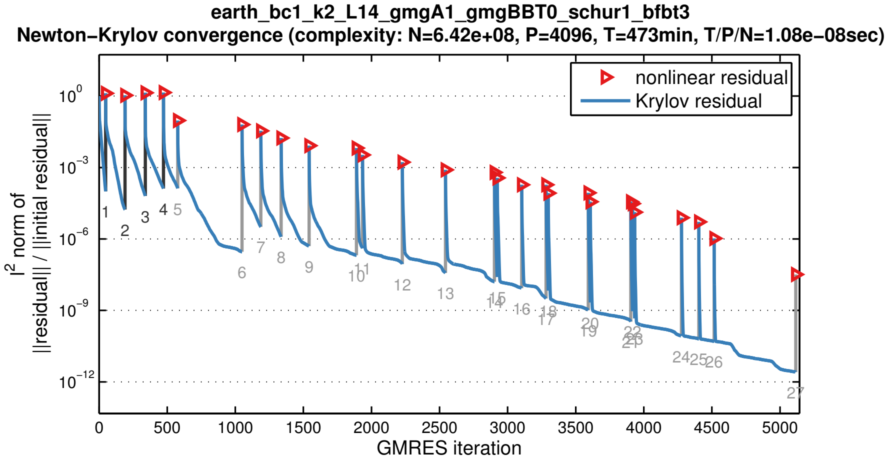
# Inexact Newton-Krylov convergence for Earth problem

## Reduction of $H^{-1}$-norm of velocity residual and Newton step length



# Inexact Newton-Krylov convergence for Earth problem

## Reduction of $l^2$-norm of residual at Newton and Krylov iterations



Thank you

# Main results summary

Two essential components for scalable mantle flow solvers:

## I. Efficient methods/algorithms

- high-order finite elements
- adaptive meshes, resolving viscosity variations
- inexact Newton-Krylov method
- $H^{-1}$-norm for velocity comp. for Newton line search
- multigrid preconditioners for elliptic operators
- BFBT/LSC type pressure Schur complement preconditioner

## II. Scalable parallel implementation

- matrix-free stiffness/mass application
- tensor product structure of FE shape functions
- octree algorithms for handling adaptive mesh in parallel
- high-order GMG with linear AMG as coarse solve
- AMG on sparsified matrix using trilinear FE at high-order nodes
- scalability up to 16384 cores

# BFBT/LSC methods for Schur complement preconditioning

# Review: BFBT/LSC methods for Schur complement $\tilde{\mathbf{S}}^{-1}$

**BFBT method** [Elman, 1999]: pseudoinverse

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

**Least Squares Commutators (LSC)** [Elman, et al., 2006]:
Find commutator matrix $\mathbf{X}$ s.t. $(\mathbf{A}\mathbf{B}^\top - \mathbf{B}^\top\mathbf{X}) \approx \mathbf{0}$, by solving the least squares problem:

$$\text{Find columns } \mathbf{x}_j \text{ of } \mathbf{X} \text{ s.t. } \quad \min_{\mathbf{x}_j} \left\| [\mathbf{A}\mathbf{B}^\top]_j - \mathbf{B}^\top\mathbf{x}_j \right\|_2^2$$

$$\Rightarrow \ \mathbf{X} = (\mathbf{B}\mathbf{B}^\top)^{-1}(\mathbf{B}\mathbf{A}\mathbf{B}^\top)$$

$$(\mathbf{A}\mathbf{B}^\top - \mathbf{B}^\top\mathbf{X}) \approx \mathbf{0} \ \Rightarrow \ (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^\top)^{-1} \approx (\mathbf{B}\mathbf{B}^\top)^{-1}(\mathbf{B}\mathbf{A}\mathbf{B}^\top)(\mathbf{B}\mathbf{B}^\top)^{-1}$$

LSC gives same result for $\tilde{\mathbf{S}}^{-1}$ as pseudoinverse.

**Q:** Does this work for FE discretizations?. . .

# Review: BFBT/LSC methods for Schur complement $\tilde{\mathbf{S}}^{-1}$

**BFBT method** [Elman, 1999]: pseudoinverse

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

**Least Squares Commutators (LSC)** [Elman, et al., 2006]:
Find commutator matrix $\mathbf{X}$ s.t. $(\mathbf{A}\mathbf{B}^\top - \mathbf{B}^\top\mathbf{X}) \approx \mathbf{0}$, by solving the least squares problem:

$$\text{Find columns } \mathbf{x}_j \text{ of } \mathbf{X} \text{ s.t. } \quad \min_{\mathbf{x}_j} \left\| [\mathbf{A}\mathbf{B}^\top]_j - \mathbf{B}^\top\mathbf{x}_j \right\|_2^2$$

$$\Rightarrow \; \mathbf{X} = (\mathbf{B}\mathbf{B}^\top)^{-1}(\mathbf{B}\mathbf{A}\mathbf{B}^\top)$$

$$(\mathbf{A}\mathbf{B}^\top - \mathbf{B}^\top\mathbf{X}) \approx \mathbf{0} \;\Rightarrow\; (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^\top)^{-1} \approx (\mathbf{B}\mathbf{B}^\top)^{-1}(\mathbf{B}\mathbf{A}\mathbf{B}^\top)(\mathbf{B}\mathbf{B}^\top)^{-1}$$

LSC gives same result for $\tilde{\mathbf{S}}^{-1}$ as pseudoinverse.

**Q:** Does this work for FE discretizations?... no

# Review: BFBT/LSC methods for Schur complement $\tilde{\mathbf{S}}^{-1}$

**Diagonally scaled BFBT method** [Elman, et al., 2006]:

Find columns $\mathbf{x}_j$ of $\mathbf{X}$ s.t. $\min_{\mathbf{x}_j} \left\| \mathbf{M}_1^{-1/2} [\mathbf{A}\mathbf{M}_2^{-1}\mathbf{B}^\top]_j - \mathbf{M}_1^{-1/2}\mathbf{B}^\top \mathbf{x}_j \right\|_2^2$

$$\Rightarrow \ \mathbf{X} = (\mathbf{B}\mathbf{M}_1^{-1}\mathbf{B}^\top)^{-1}(\mathbf{B}\mathbf{M}_1^{-1}\mathbf{A}\mathbf{M}_2^{-1}\mathbf{B}^\top)$$

$$\Rightarrow \ \tilde{\mathbf{S}}^{-1} = (\mathbf{B}\mathbf{M}_1^{-1}\mathbf{B}^\top)^{-1}(\mathbf{B}\mathbf{M}_1^{-1}\mathbf{A}\mathbf{M}_2^{-1}\mathbf{B}^\top)(\mathbf{B}\mathbf{M}_2^{-1}\mathbf{B}^\top)^{-1}$$

Proposed scaling: For FE, use “diagonalized” velocity mass matrix,

diagonal: $\mathbf{M}_1 = \mathbf{M}_2 = \mathrm{diag}(\mathbf{M}_{\boldsymbol{u}})$ or lumped: $\mathbf{M}_1 = \mathbf{M}_2 = \tilde{\mathbf{M}}_{\boldsymbol{u}}$

Since $\mathbf{B}\mathbf{M}_1^{-1}\mathbf{B}^\top$ can be understood as a Laplace operator for the pressure, approximate $(\mathbf{B}\mathbf{M}_1^{-1}\mathbf{B}^\top)^{-1}$ by a multigrid V-cycle.

**Q:** Is mass scaled BFBT effective for high viscosity variations?. . .

# Review: BFBT/LSC methods for Schur complement $\tilde{\mathbf{S}}^{-1}$

**Diagonally scaled BFBT method** [Elman, et al., 2006]:

$$\text{Find columns } \mathbf{x}_j \text{ of } \mathbf{X} \text{ s.t. } \min_{\mathbf{x}_j} \left\| \mathbf{M}_1^{-1/2} [\mathbf{A}\mathbf{M}_2^{-1}\mathbf{B}^\top]_j - \mathbf{M}_1^{-1/2}\mathbf{B}^\top \mathbf{x}_j \right\|_2^2$$

$$\Rightarrow \ \mathbf{X} = (\mathbf{B}\mathbf{M}_1^{-1}\mathbf{B}^\top)^{-1}(\mathbf{B}\mathbf{M}_1^{-1}\mathbf{A}\mathbf{M}_2^{-1}\mathbf{B}^\top)$$

$$\Rightarrow \ \tilde{\mathbf{S}}^{-1} = (\mathbf{B}\mathbf{M}_1^{-1}\mathbf{B}^\top)^{-1}(\mathbf{B}\mathbf{M}_1^{-1}\mathbf{A}\mathbf{M}_2^{-1}\mathbf{B}^\top)(\mathbf{B}\mathbf{M}_2^{-1}\mathbf{B}^\top)^{-1}$$

Proposed scaling: For FE, use "diagonalized" velocity mass matrix,

$$\text{diagonal: } \mathbf{M}_1 = \mathbf{M}_2 = \mathrm{diag}(\mathbf{M}_{\boldsymbol{u}}) \quad \text{or lumped: } \mathbf{M}_1 = \mathbf{M}_2 = \tilde{\mathbf{M}}_{\boldsymbol{u}}$$

Since $\mathbf{B}\mathbf{M}_1^{-1}\mathbf{B}^\top$ can be understood as a Laplace operator for the pressure, approximate $(\mathbf{B}\mathbf{M}_1^{-1}\mathbf{B}^\top)^{-1}$ by a multigrid V-cycle.

**Q:** Is mass scaled BFBT effective for high viscosity variations?. . . no

# Review: BFBT/LSC methods for Schur complement $\tilde{\mathbf{S}}^{-1}$

**BFBT for scaled Stokes systems** that arise in geodynamics [May, Moresi, 2008]:

$$\begin{bmatrix} \mathbf{D}_{\boldsymbol{u}}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_p^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{D}_{\boldsymbol{u}}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_p^{-1/2} \end{bmatrix}$$

Then the standard BFBT method yields its scaled version,

$$\Rightarrow \; \tilde{\mathbf{S}}^{-1} = (\mathbf{B}\mathbf{D}_{\boldsymbol{u}}^{-1}\mathbf{B}^\top)^{-1}(\mathbf{B}\mathbf{D}_{\boldsymbol{u}}^{-1}\mathbf{A}\mathbf{D}_{\boldsymbol{u}}^{-1}\mathbf{B}^\top)(\mathbf{B}\mathbf{D}_{\boldsymbol{u}}^{-1}\mathbf{B}^\top)^{-1}$$

Proposed scaling: heuristic, motivated by scaling of dimensional systems

$$[\mathbf{D}_{\boldsymbol{u}}]_{i,i} = \max_j |[\mathbf{A}]_{i,j}|$$

**Q:** Is BFBT with this scaling effective for high viscosity variations?...

# Review: BFBT/LSC methods for Schur complement $\tilde{\mathbf{S}}^{-1}$

**BFBT for scaled Stokes systems** that arise in geodynamics [May, Moresi, 2008]:

$$\begin{bmatrix} \mathbf{D}_{\boldsymbol{u}}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_p^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{D}_{\boldsymbol{u}}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_p^{-1/2} \end{bmatrix}$$

Then the standard BFBT method yields its scaled version,

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

Proposed scaling: heuristic, motivated by scaling of dimensional systems

$$[\mathbf{D}_{\boldsymbol{u}}]_{i,i} = \max_j |[\mathbf{A}]_{i,j}|$$

**Q:** Is BFBT with this scaling effective for high viscosity variations?... yes

# New view on BFBT/LSC methods

Let $\mathbf{C}$ be symm. pos. def. and let $\mathbf{D}$ be arbitrary,

$$\text{Find } \mathbf{X} \text{ s.t.} \quad \min_{\mathbf{X}} \left\| \mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top \mathbf{e}_j - \mathbf{B}^\top \mathbf{X} \mathbf{e}_j \right\|^2_{\mathbf{C}^{-1}} \quad \text{for all } j$$

$$\Rightarrow \; \mathbf{X} = (\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\top)^{-1}(\mathbf{B}\mathbf{C}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top)$$

And we have a $\mathbf{C}^{-1}$-orthogonal projection, i.e., the residual satisfies

$$\left\langle \mathbf{B}^\top \mathbf{e}_i, (\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top - \mathbf{B}^\top \mathbf{X})\mathbf{e}_j \right\rangle_{\mathbf{C}^{-1}} = \mathbf{0} \quad \text{for all } i, j,$$

therefore

$$\left( \mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top - \mathbf{B}^\top \mathbf{X} \right) \mathbf{e}_j \perp_{\mathbf{C}^{-1}} \mathrm{Ran}(\mathbf{B}^\top) \quad \text{for all } j$$

# New view on BFBT/LSC methods

**Goal:** Effective and robust preconditioning of the Schur complement in Stokes systems with high viscosity variations

**Recall:** Condition for optimal preconditioning $(\mathbf{B}\tilde{\mathbf{A}}^{-1}\mathbf{B}^\top)\tilde{\mathbf{S}}^{-1} = \mathbf{I}$.
By choosing $\mathbf{C} = \tilde{\mathbf{A}}$, we obtain equivalence between orthogonality and the condition for optimal preconditioning:

$$\left\langle \mathbf{B}^\top \mathbf{e}_i, (\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top - \mathbf{B}^\top \mathbf{X})\mathbf{e}_j \right\rangle_{\tilde{\mathbf{A}}^{-1}} = \mathbf{0} \quad \forall i, j \quad \Leftrightarrow \quad \tilde{\mathbf{S}} = \mathbf{B}\tilde{\mathbf{A}}^{-1}\mathbf{B}^\top$$

# New view on BFBT/LSC methods

**Goal:** Effective and robust preconditioning of the Schur complement in Stokes systems with high viscosity variations

**Recall:** Condition for optimal preconditioning $(\mathbf{B}\tilde{\mathbf{A}}^{-1}\mathbf{B}^\top)\tilde{\mathbf{S}}^{-1} = \mathbf{I}$.
By choosing $\mathbf{C} = \tilde{\mathbf{A}}$, we obtain equivalence between orthogonality and the condition for optimal preconditioning:

$$\left\langle \mathbf{B}^\top \mathbf{e}_i, (\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top - \mathbf{B}^\top\mathbf{X})\mathbf{e}_j \right\rangle_{\tilde{\mathbf{A}}^{-1}} = \mathbf{0} \quad \forall i,j \quad \Leftrightarrow \quad \tilde{\mathbf{S}} = \mathbf{B}\tilde{\mathbf{A}}^{-1}\mathbf{B}^\top$$

Choices of $\mathbf{C}, \mathbf{D}$ that are computationally feasible are limited.
Our choice: $\mathbf{C} = \mathbf{D} := \mathrm{diag}(\mathbf{A})$, thus

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

Approximate $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top)^{-1}$ with AMG V-cycles (PETSc GAMG).

**Q:** Is this $\tilde{\mathbf{S}}^{-1}$ robust at high viscosity variations?...

# New view on BFBT/LSC methods

**Goal:** Effective and robust preconditioning of the Schur complement in Stokes systems with high viscosity variations

**Recall:** Condition for optimal preconditioning $(\mathbf{B}\tilde{\mathbf{A}}^{-1}\mathbf{B}^\top)\tilde{\mathbf{S}}^{-1} = \mathbf{I}$.
By choosing $\mathbf{C} = \tilde{\mathbf{A}}$, we obtain equivalence between orthogonality and the condition for optimal preconditioning:

$$\left\langle \mathbf{B}^\top \mathbf{e}_i, (\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top - \mathbf{B}^\top\mathbf{X})\mathbf{e}_j \right\rangle_{\tilde{\mathbf{A}}^{-1}} = \mathbf{0} \quad \forall i,j \quad \Leftrightarrow \quad \tilde{\mathbf{S}} = \mathbf{B}\tilde{\mathbf{A}}^{-1}\mathbf{B}^\top$$

Choices of $\mathbf{C}, \mathbf{D}$ that are computationally feasible are limited.
Our choice: $\mathbf{C} = \mathbf{D} := \mathrm{diag}(\mathbf{A})$, thus

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

Approximate $(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top)^{-1}$ with AMG V-cycles (PETSc GAMG).

**Q:** Is this $\tilde{\mathbf{S}}^{-1}$ robust at high viscosity variations?... see results