An Extreme-Scale Implicit Solver for Highly Nonlinear and Heterogeneous Flow in Earth's Mantle

Johann Rudi[1] Georg Stadler[2] Omar Ghattas[1,3]

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

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

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

INSTITUTE FOR COMPUTATIONAL ENGINEERING & SCIENCES

Outline

Earth's mantle convection: The driving application & solver challenges

Weighted BFBT preconditioner for the Schur complement

Hybrid spectral–geometric–algebraic multigrid

Numerical results: Parallel scalability of the linear solver

Inexact Newton–Krylov method & primal-dual linearization

What we know: Observational data

- Current plate motion from GPS and magnetic anomalies
- Plate deformation obtained from dense GPS networks
- Average viscosity in regions affected by post-glacial rebound
- Topography indicating normal traction at earth's surface

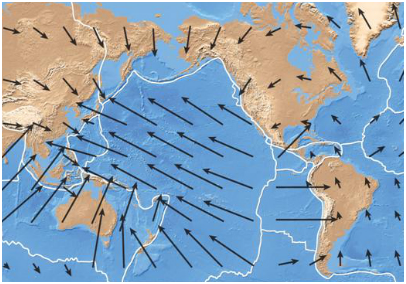


Plate motion (Credit: Pearson Prentice Hall, Inc.)

Additional knowledge contributing to mantle rheology:

- Location and geometry of plates, plate boundaries, and subducting slabs (from seismicity)
- Images of present-day earth structure (by correlating seismic wave speed with temperature)

Earth's mantle modeling via nonlinear rheology

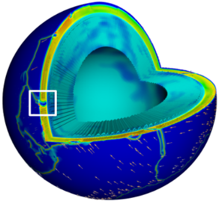
Nonlinear constitutive relationship / rheology due to:

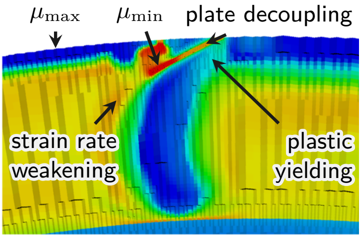
- Strain rate weakening exponent $n \geq 1$ ($\dot{\varepsilon}_{\mathrm{II}}(\boldsymbol{u})$ is 2nd invariant of strain rate)
- Yield strength $\tau_{\mathrm{yield}} > 0$ causing plastic yielding in lithosphere

Additionally, strong heterogeneity in the linear system is introduced by:

- Exponential temperature dependence $a(T)$ (Arrhenius relationship)
- Plate decoupling factor $0 < w(\boldsymbol{x}) \leq 1$ with orders-of-magnitude contrasts

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





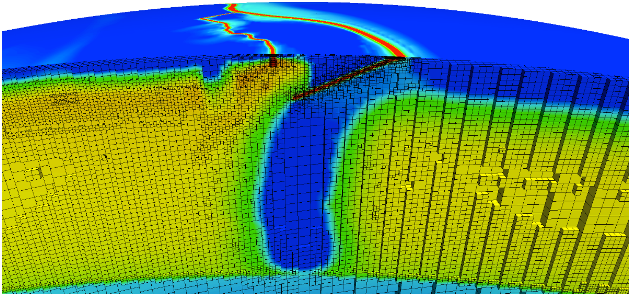
What we would like to learn: Rheological parameters

Globally constant parameters affecting viscosity and nonlinearity:

- Global scaling factor of the upper mantle viscosity (0–660 km depth)
- Stress exponent controlling severity of strain rate weakening
- Yield strength governing plastic yielding phenomena

Local, spatially varying parameters:

- Coupling strength / energy dissipation between plates



(Credit: Alisic)

Mantle flow governed by incompressible Stokes equations

Nonlinear incompressible Stokes PDE (w/ free-slip & no-normal flow BC):

$$-\nabla \cdot \left[\mu(\boldsymbol{u})\left(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^{\mathsf{T}}\right)\right] + \nabla p = \boldsymbol{f} \qquad \text{viscosity } \mu, \text{ RHS forcing } \boldsymbol{f}$$

$$-\nabla \cdot \boldsymbol{u} = 0 \qquad \text{seek: velocity } \boldsymbol{u}, \text{ pressure } p$$

Linearization (with Newton), then discretization (with inf-sup stable F.E.):

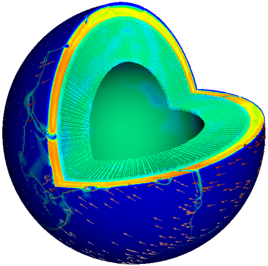
$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathsf{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{u}} \\ \tilde{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} -\mathbf{r}_1 \\ -\mathbf{r}_2 \end{bmatrix}$$

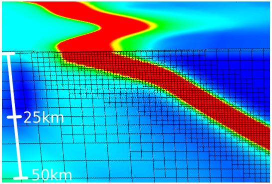
- High-order finite element shape functions
- Inf-sup stable velocity–pressure pairings: $\mathbb{Q}_k \times \mathbb{P}_{k-1}^{\mathrm{disc}}$ with order $k \geq 2$
- Locally mass conservative due to discontinuous, modal pressure
- Non-conforming hexahedral meshes with "hanging nodes"
- Adaptive mesh refinement resolving fine-scale features of mantle

Severe challenges for parallel scalable solvers

. . . arising in Earth's mantle convection:

- Severe nonlinearity, heterogeneity, and anisotropy imposed by Earth's rheology
- Sharp viscosity gradients in narrow regions (6 orders of magnitude drop in $\sim$5 km)
- Wide range of spatial scales and highly localized features, e.g., plate boundaries of size $\mathcal{O}(1\,\mathrm{km})$ influence plate motion at continental scales of $\mathcal{O}(1000\,\mathrm{km})$
- Adaptive mesh refinement is essential
- High-order finite elements $\mathbb{Q}_k \times \mathbb{P}_{k-1}^{\mathrm{disc}}$, order $k \geq 2$, with local mass conservation; yields a difficult to deal with discontinuous, modal pressure approximation





Viscosity *(colors)* and locally refined mesh.

Outline

w-BFBT: Robust inverse Schur complement approximation

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^\mathsf{T} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{B}^\mathsf{T} \\ \mathbf{0} & \tilde{\mathbf{S}} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\tilde{\mathbf{u}}} \\ \tilde{\tilde{\mathbf{p}}} \end{bmatrix} = \begin{bmatrix} -\mathbf{r}_1 \\ -\mathbf{r}_2 \end{bmatrix}$$

$$\tilde{\mathbf{A}}^{-1} \approx \mathbf{A}^{-1}$$

$$\tilde{\mathbf{S}}^{-1} \approx (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^\mathsf{T})^{-1}$$

w-BFBT: Robust inverse Schur complement approximation

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^\mathsf{T} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{B}^\mathsf{T} \\ \mathbf{0} & \tilde{\mathbf{S}} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\tilde{\mathbf{u}}} \\ \tilde{\tilde{\mathbf{p}}} \end{bmatrix} = \begin{bmatrix} -\mathbf{r}_1 \\ -\mathbf{r}_2 \end{bmatrix} \qquad \begin{aligned} \tilde{\mathbf{A}}^{-1} &\approx \mathbf{A}^{-1} \\ \tilde{\mathbf{S}}^{-1} &\approx (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^\mathsf{T})^{-1} \end{aligned}$$

$$\tilde{\mathbf{S}}^{-1}_{\text{w-BFBT}} := \underbrace{\left(\mathbf{B}\mathbf{C}_w^{-1}\mathbf{B}^\mathsf{T}\right)^{-1}}_{\text{Poisson solve}} \left(\mathbf{B}\mathbf{C}_w^{-1}\mathbf{A}\mathbf{D}_w^{-1}\mathbf{B}^\mathsf{T}\right) \underbrace{\left(\mathbf{B}\mathbf{D}_w^{-1}\mathbf{B}^\mathsf{T}\right)^{-1}}_{\text{Poisson solve}}$$

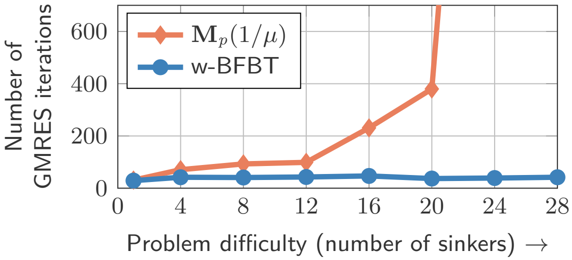
Choice of diagonal weighting matrices $\mathbf{C}_w = \mathbf{D}_w := \tilde{\mathbf{M}}_{\boldsymbol{u}}(w)$ is critical for efficacy & robustness. [Rudi, Stadler, Ghattas, 2017] proposes $w = \sqrt{\mu}$.

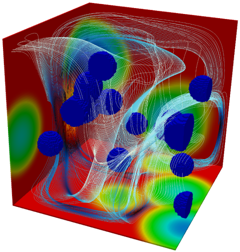
w-BFBT: Robust inverse Schur complement approximation

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^\mathsf{T} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{B}^\mathsf{T} \\ \mathbf{0} & \tilde{\mathbf{S}} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\tilde{\mathbf{u}}} \\ \tilde{\tilde{\mathbf{p}}} \end{bmatrix} = \begin{bmatrix} -\mathbf{r}_1 \\ -\mathbf{r}_2 \end{bmatrix} \qquad \begin{aligned} \tilde{\mathbf{A}}^{-1} &\approx \mathbf{A}^{-1} \;\rightarrow\; \text{MG V-cycle} \\ \tilde{\mathbf{S}}^{-1} &\approx (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^\mathsf{T})^{-1} \end{aligned}$$

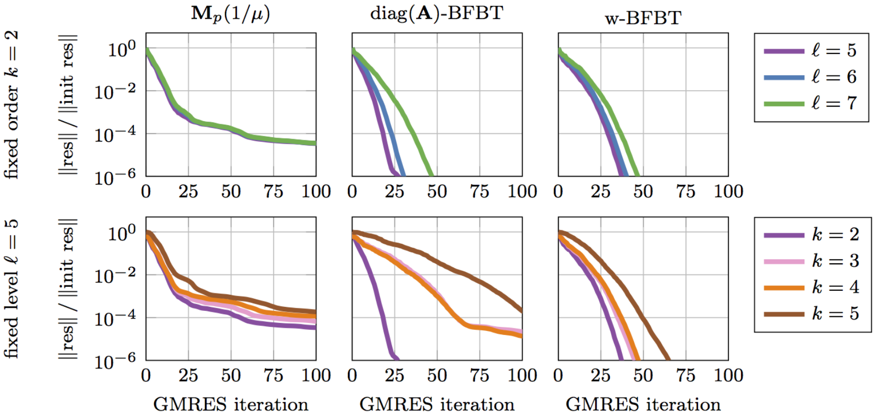
$$\tilde{\mathbf{S}}^{-1}_{\text{w-BFBT}} := \underbrace{\left(\mathbf{B}\mathbf{C}_w^{-1}\mathbf{B}^\mathsf{T}\right)^{-1}}_{\rightarrow\ \text{MG V-cycle}} \left(\mathbf{B}\mathbf{C}_w^{-1}\mathbf{A}\mathbf{D}_w^{-1}\mathbf{B}^\mathsf{T}\right) \underbrace{\left(\mathbf{B}\mathbf{D}_w^{-1}\mathbf{B}^\mathsf{T}\right)^{-1}}_{\rightarrow\ \text{MG V-cycle}}$$

Choice of diagonal weighting matrices $\mathbf{C}_w = \mathbf{D}_w := \tilde{\mathbf{M}}_{\boldsymbol{u}}(w)$ is critical for efficacy & robustness. [Rudi, Stadler, Ghattas, 2017] proposes $w = \sqrt{\mu}$.





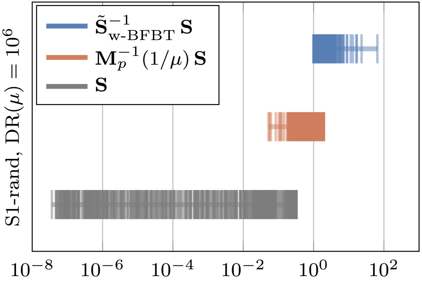
Comparison of Schur complement preconditioners

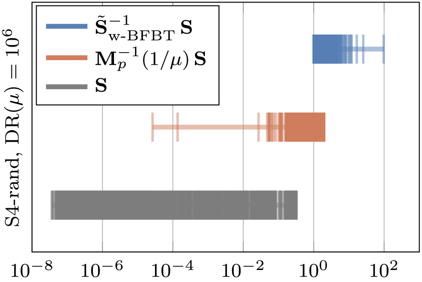


w-BFBT combines robust convergence of $\mathrm{diag}(\mathbf{A})$-BFBT [May, Moresi, 2008] with improved algorithmic scalability when order k increases.

Spectrum comparisons of preconditioned Schur matrices

2D Stokes problem discretized with $\mathbb{P}_2^{\mathrm{bubble}} \times \mathbb{P}_1^{\mathrm{disc}}$ finite elements on a uniform triangular mesh consisting of 512 triangles (FEniCS library).





- As the problem difficulty (i.e., sinker counts) increases, the spreading of small eigenvalues for $\mathbf{M}_p(1/\mu)$ becomes more severe, which is disadvantageous for Krylov solver convergence.
- w-BFBT remains largely unaffected by increased difficulty, which results in convergence that is robust w.r.t. viscosity variations.

Spectral equivalence for w-BFBT

Theorem: [Rudi, Stadler, Ghattas, 2017] Assume an infinite-dimensional w-BFBT approximation of the Schur complement:

$$\tilde{S}_{\text{w-BFBT}} := K_w^* (Bw\,A\,wB^*)^{-1} K_w, \quad K_w^* := BwB^*, \quad w \equiv \mu^{-\frac{1}{2}}$$

Then $\tilde{S}_{\text{w-BFBT}}$ is equivalent to $S = BA^{-1}B^*$,

$$\left(\tilde{S}_{\text{w-BFBT}}\, q \,,\, q\right) \leq (Sq \,,\, q) \leq C_{\text{w-BFBT}} \left(\tilde{S}_{\text{w-BFBT}}\, q \,,\, q\right) \quad \text{for all } q,$$

with a constant based on weighted Poincaré–Friedrichs' and Korn's ineq.

$$C_{\text{w-BFBT}} := \left(1 + \frac{1}{4} \left\|\nabla\mu\right\|^2_{L^\infty(\Omega)^d}\right) \left(C^2_{P,\mu} + 1\right) C^2_{K,\mu}$$

Remark: For a constant viscosity $\mu \equiv 1$ the equivalence relationship holds with classical Poincaré–Friedrichs' and Korn's inequalities.

Outline

Earth's mantle convection: The driving application & solver challenges

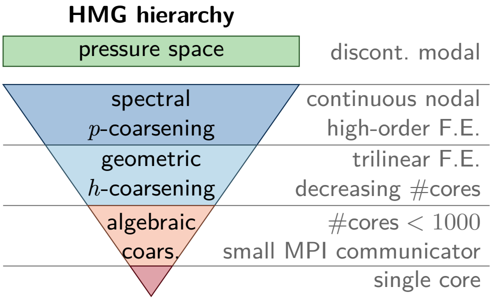
Weighted BFBT preconditioner for the Schur complement

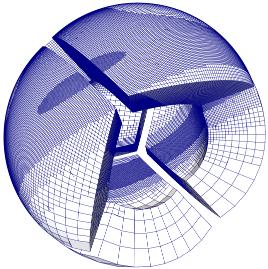
Hybrid spectral–geometric–algebraic multigrid

Numerical results: Parallel scalability of the linear solver

Inexact Newton–Krylov method & primal-dual linearization

HMG: Hybrid spectral–geometric–algebraic multigrid

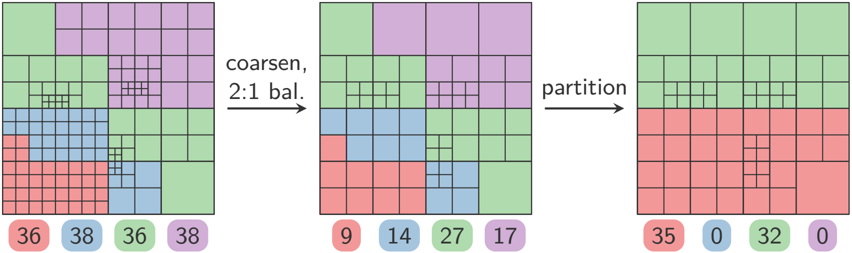




- Multigrid hierarchy of nested meshes is generated from an adaptively refined octree-based mesh via spectral–geometric coarsening
- Re-discretization of PDEs at coarser levels
- Parallel repartitioning of coarser meshes for load-balancing (crucial for AMR); sufficiently coarse meshes occupy only subsets of cores
- Coarse grid solver: AMG (from PETSc) invoked on small core counts

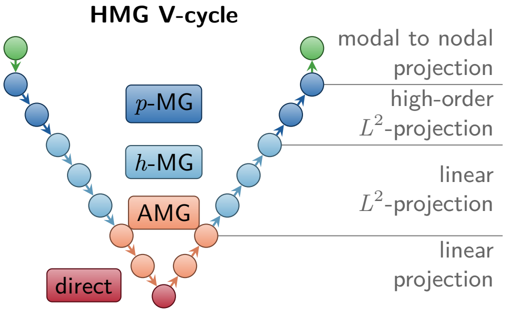
Geometric coarsening: Repartitioning & core-thinning

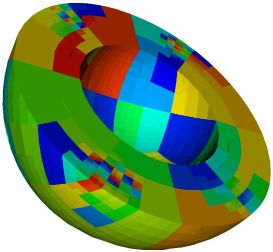
- Parallel repartitioning (p4est library) of adapted meshes for load balancing
- Core-thinning to avoid excessive communication in multigrid cycle
- Reduced MPI communicators containing only non-empty cores
- Ensure coarsening across core boundaries: Partition families of octants/elements on same core for next coarsening sweep



Colors depict different processor cores, *numbers* indicate element count on each core.
[Sundar, Biros, Burstedde, Rudi, Ghattas, Stadler, 2012]

HMG V-cycle: Smoothing & interpolation/restriction

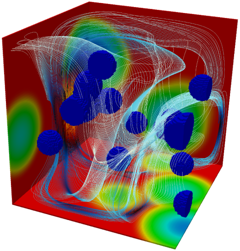




- High-order L^2-projection onto coarser levels; restriction & interpolation are adjoints of each other in L^2-sense
- Chebyshev accelerated Jacobi smoother (Cheb. from PETSc) with tensorized matrix-free high-order stiffness apply; assembly of high-order diagonal only
- Efficacy, i.e., error reduction, of HMG V-cycles is independent of core count
- No collective communication needed in spectral–geometric MG cycles

Outline

Problem setup for scalability tests on Lonestar 5



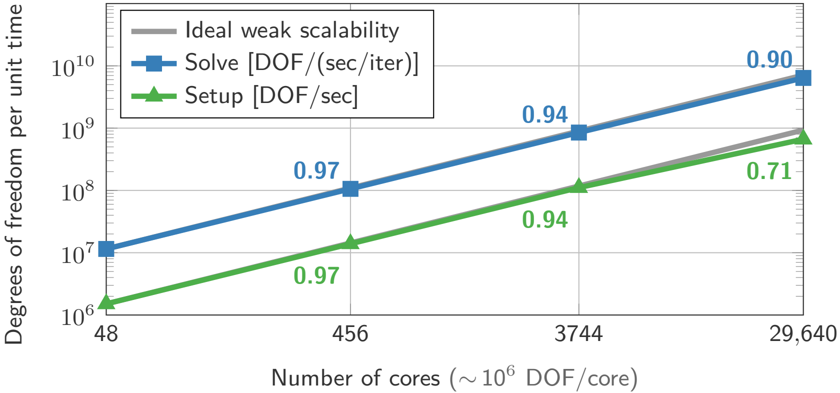
Discretization parameters to test parallel scalability:

- Finite element order $k = 2$ is fixed $(\mathbb{Q}_k \times \mathbb{P}_1^{\mathrm{disc}})$
- Vary mesh refinement level ℓ for weak scalability

Multigrid parameters for $\tilde{\mathbf{A}}^{-1}$ and $\left(\mathbf{B}\mathbf{C}_w^{-1}\mathbf{B}^{\mathsf{T}}\right)^{-1}$ are:

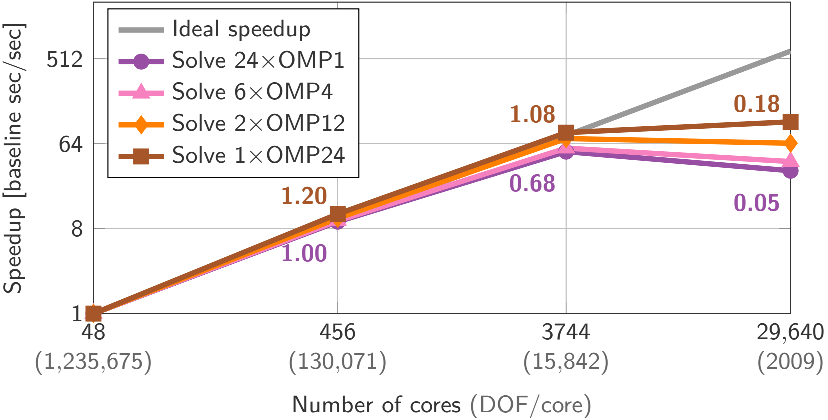
- 1 HMG V-cycle with 3+3 smoothing

Weak scalability for HMG+w-BFBT on Lonestar 5



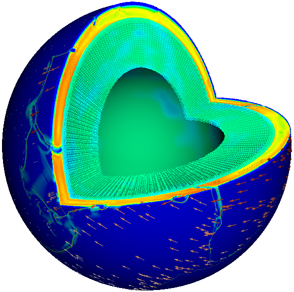
Performed on TACC's Lonestar 5: Cray XC40 with 1252 compute nodes, each contains 2 Intel Haswell 12-core processors and 64 GBytes of memory.
[Rudi, Stadler, Ghattas, 2017]

Strong scalability for HMG+w-BFBT on Lonestar 5



Performed on TACC's Lonestar 5: Cray XC40 with 1252 compute nodes, each contains 2 Intel Haswell 12-core processors and 64 GBytes of memory.
[Rudi, Stadler, Ghattas, 2017]

Parallel scalability: Global mantle convection problem setup



Discretization parameters to test parallel scalability:

- Finite element order $k = 2$ is fixed ($\mathbb{Q}_k \times \mathbb{P}_{k-1}^{\mathrm{disc}}$)
- Increase max mesh refinement $\ell_{\max}$
- Refinement down to $\sim$75 m local resolution
- Resulting mesh has 9 levels of refinement

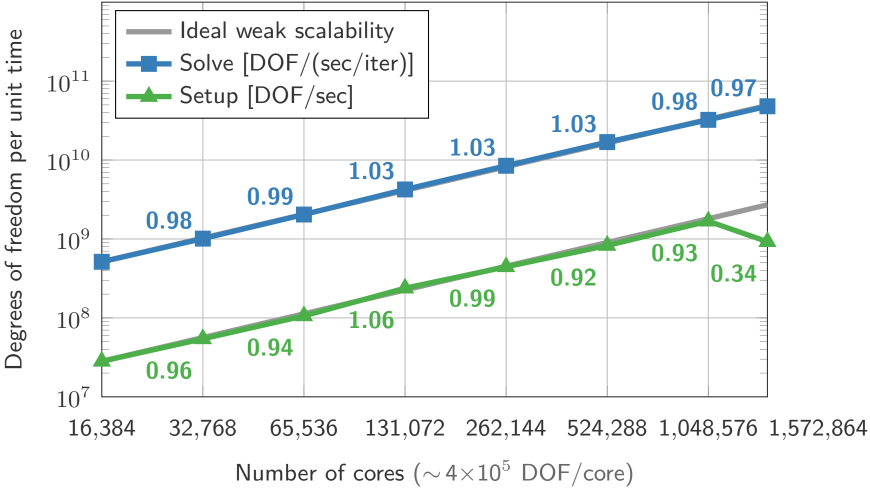
Multigrid parameters for elliptic blocks $\mathbf{A}$ and $\mathbf{K}$:

- 1 HMG V-cycle with 3+3 smoothing

Hardware and target system:

- IBM Blue Gene/Q architecture
- Lawrence Livermore National Lab's Sequoia
- 96 racks resulting in 98,304 nodes and 1,572,864 cores

Extreme weak scalability on Sequoia supercomputer



[Rudi, Malossi, Isaac, Stadler, Gurnis, Staar, Ineichen, Bekas, Curioni, Ghattas, 2015]

Outline

Inexact Newton–Krylov method

Compute Newton update $(\tilde{\boldsymbol{u}}, \tilde{p})$ via inexact solution: [w/ $\nabla_{\mathrm{s}} \boldsymbol{u} := \frac{1}{2}(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^{\mathsf{T}})$]

$$
\begin{aligned}
-\nabla_{\mathrm{s}} \cdot (2\mu' \, \nabla_{\mathrm{s}} \tilde{\boldsymbol{u}}) + \nabla \tilde{p} &= -\mathbf{r}_1 \\
\nabla \cdot \tilde{\boldsymbol{u}} &= -r_2
\end{aligned}
\quad \text{with} \quad \mu' = \mu \, \mathbf{I} + \dot{\varepsilon}_{\mathrm{II}} \frac{\partial \mu}{\partial \dot{\varepsilon}_{\mathrm{II}}} \frac{\nabla_{\mathrm{s}} \boldsymbol{u} \otimes \nabla_{\mathrm{s}} \boldsymbol{u}}{\|\nabla_{\mathrm{s}} \boldsymbol{u}\|_F^2}
$$

Original Newton linearization vs. primal-dual: Introduce $\boldsymbol{T} = \nabla_{\mathrm{s}} \boldsymbol{u} / \|\nabla_{\mathrm{s}} \boldsymbol{u}\|_F$

$$
H_{1,1}(\boldsymbol{u}) \tilde{\boldsymbol{u}} := -\nabla_{\mathrm{s}} \cdot \left(2\mu \left(\mathbf{I} - \theta \frac{\nabla_{\mathrm{s}} \boldsymbol{u} \otimes \nabla_{\mathrm{s}} \boldsymbol{u}}{\|\nabla_{\mathrm{s}} \boldsymbol{u}\|_F^2} \right) \nabla_{\mathrm{s}} \tilde{\boldsymbol{u}} \right), \quad 0 \le \theta < 1,
$$

$$
H_{1,1}(\boldsymbol{u}, \boldsymbol{T}) \tilde{\boldsymbol{u}} := -\nabla_{\mathrm{s}} \cdot \left(2\mu \left(\mathbf{I} - \theta \frac{\nabla_{\mathrm{s}} \boldsymbol{u} \otimes \boldsymbol{T}}{\|\nabla_{\mathrm{s}} \boldsymbol{u}\|_F} \right) \nabla_{\mathrm{s}} \tilde{\boldsymbol{u}} \right), \quad \|\boldsymbol{T}\|_F \le 1.
$$

No additional DOF or solves, since dual variable is computed explicitly in each Newton step: $\boldsymbol{T} \leftarrow \boldsymbol{T} + \widetilde{\boldsymbol{T}}$

Primal-dual Newton for mantle's yielding rheology

Fast & stable nonlinear convergence with primal-dual Newton linearization:

Yielding volume	level ℓ	Original Newton #Newton	Original Newton #backtr.	Primal-dual Newton #Newton	Primal-dual Newton #backtr.
~45%	4	33	20	10	0
~45%	5	36	25	12	0
~45%	6	57	49	13	0
~65%	4	29	21	18	10
~65%	5	37	26	17	9
~65%	6	48	39	20	9
~90%	4	35	25	19	11
~90%	5	40	32	21	11
~90%	6	32	21	23	11

- Number of Newton steps (#Newton)
- Number of steps with backtracking line search (#backtr.)

Interpretation of primal-dual Newton

Observations:

- Primal-dual linearization only requires to modify the Hessian.
- No additional solve required, hence the computational cost per Newton step is negligible.
- Nonlinear convergence in the presence of yielding rheology is significantly improved.
- As mesh elements are refined, stable nonlinear convergence is maintained.

Interpretation:

- Primal-dual linearization acts as a regularization/preconditioner far from solution while enabling super-linear Newton convergence close to solution.
- Regularization has a large magnitude far from solution and goes to zero as solution is assumed.
- Regularization can be interpreted as "model error dependent," where "model" refers to the 2^{nd}-order approximation used by Newton.

Summary of completed research and outlook

Summary:

- Hybrid spectral–geometric–algebraic multigrid (based on p4est library; extended by a coarsening correction to enable coarsening across core boundaries).
- Weighted BFBT preconditioner for the for the Schur complement; scalable HMG-based BFBT algorithms, heterogeneity-robust weighting of BFBT and theoretical foundation.
- Inexact Newton–Krylov with grid-continuation for highly nonlinear mantle rheology.
- Parallel scalability of solvers to 1.6 million cores (collaboration with IBM Research – Zurich).

References:

- Rudi, Stadler, and Ghattas, SISC, 2017.
- Rudi, Stadler, and Ghattas, 14th Copper Mountain Conference on Iterative Methods (winner of student paper competition, unpublished competition paper), 2016.
- Rudi, Malossi, Isaac, Stadler, Gurnis, Staar, Ineichen, Bekas, Curioni, and Ghattas, Proceedings of SC15 (winner of Gordon Bell Prize), 2015.
- Sundar, Biros, Burstedde, Rudi, Ghattas, and Stadler, Proceedings of SC12, 2012.

Outlook:

- Mathematically rigorous derivation for primal-dual Newton.
- Bayesian inversion for rheological parameters of global mantle convection models.