Scalable Spectral-Geometric-Algebraic Multigrid and Schur Complement Preconditioning for Nonlinear, Multiscale, Heterogeneous Flow in Earth's Mantle

Johann Rudi

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

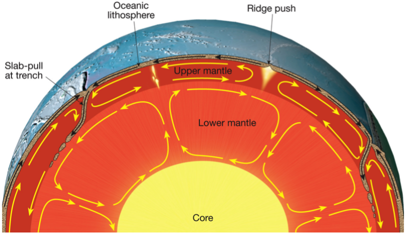
Co-Advisors
Omar Ghattas (UT Austin) and Georg Stadler (New York University)

Collaborators
Tobin Isaac (U Chicago), Michael Gurnis (Caltech), and from
IBM Research – Zurich: Cristiano I. Malossi, Peter W.J. Staar,
Yves Ineichen, Costas Bekas, Alessandro Curioni

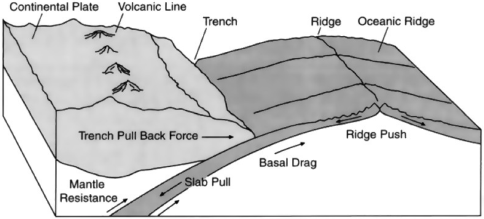
Outline

Mantle convection & plate tectonics

- Mantle convection is the thermal convection in earth's upper ~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
- Driver for plate tectonics, earthquakes, tsunamis, volcanos



Convection layering (Credit: Pearson Prentice Hall, Inc.)



Subducting slab (Credit: Schubert, Turcotte, Olsen)

- Main drivers of plate motion: negative buoyancy forces or convective shear traction?
- Key process governing the occurrence of great earthquakes: material properties between the plates or tectonic stress?

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

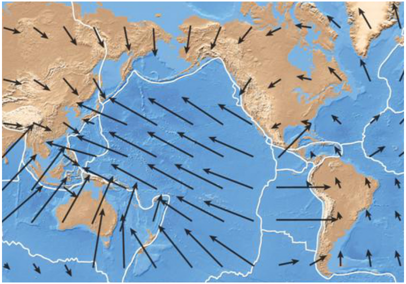


Plate motion (Credit: Pearson Prentice Hall, Inc.)

Additional knowledge contributing to mantle rheology:

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

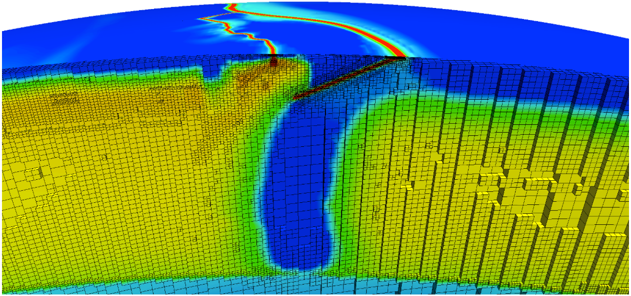
What we would like to learn: Rheological parameters

Globally constant parameters affecting viscosity and nonlinearity:

- Scaling factor of the upper mantle viscosity (down to ∼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.):

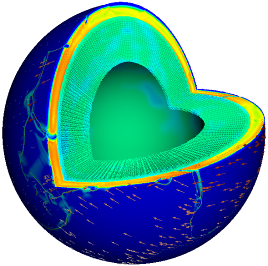
$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathsf{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{u}} \\ \tilde{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} -\mathbf{r}_1 \\ -\mathbf{r}_2 \end{bmatrix}$$

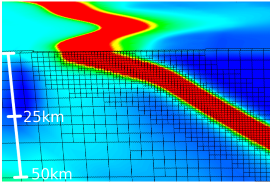
- High-order finite element shape functions
- Inf-sup stable velocity–pressure pairings: $\mathbb{Q}_k \times \mathbb{P}_{k-1}^{\text{disc}}$ with order $k \geq 2$
- Locally mass conservative due to discontinuous, modal pressure
- Adaptive mesh refinement resolving fine-scale features of mantle
- Non-conforming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

Severe challenges for parallel scalable solvers

. . . arising in Earth's mantle convection:

- Severe nonlinearity, heterogeneity, and anisotropy due to Earth's rheology (strain rate weakening, plastic yielding)
- 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 $\mathcal{O}(1\,\text{km})$ influence plate motion at continental scales of $\mathcal{O}(1000\,\text{km})$
- Adaptive mesh refinement is essential
- High-order finite elements $\mathbb{Q}_k \times \mathbb{P}_{k-1}^{\text{disc}}$, 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

Earth's mantle convection: Driving application & solver challenges

Preconditioner for the inverse Schur complement: Weighted BFBT

Preconditioning with Hybrid Spectral–Geometric–Algebraic Multigrid

Numerical results: Algorithmic & parallel scalability

Preconditioning of saddle-point systems from PDEs

Iterative scheme with upper triangular block preconditioning:

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

Commonly occurring preconditioning challenge in CS&E:

- Creeping non-Newtonian fluid modeled by incompressible Stokes equations with power-law rheology yields spatially-varying and highly heterogeneous viscosity μ after linearization.

Preconditioning of saddle-point systems from PDEs

Iterative scheme with upper triangular block preconditioning:

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

Commonly occurring preconditioning challenge in CS&E:

- Creeping non-Newtonian fluid modeled by incompressible Stokes equations with power-law rheology yields spatially-varying and highly heterogeneous viscosity μ after linearization.

Possible preconditioners for the inverse Schur complement $\mathbf{S}^{-1}$:

- Viscosity-weighted pressure mass matrix, $\mathbf{M}_p(1/\mu)$: [Burstedde, Ghattas, Stadler, et al., 2009], [Grinevich and Olshanskii, 2009]
- BFBT for Navier–Stokes: [Elman, 1999], [Silvester, Elman, Kay, Wathen, 2001], [Kay, Loghin, Wathen, 2002], [Elman, Tuminaro, 2009]
- BFBT for variable-viscosity Stokes: [May, Moresi, 2008]

Weighted BFBT: Inverse Schur complement approximation

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

Weighted BFBT: Inverse Schur complement approximation

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

From a "commutator relationship" leading to a least-squares minimization problem, we derive the BFBT approximation:

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

Weighted BFBT: Inverse Schur complement approximation

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

$\tilde{\mathbf{A}}^{-1} \approx \mathbf{A}^{-1} \rightarrow$ MG V-cycle

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

From a "commutator relationship" leading to a least-squares minimization problem, we derive the BFBT approximation:

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

Choice of diagonal weighting matrices $\mathbf{C}_w$, $\mathbf{D}_w$ is critical for efficacy & robustness with respect to viscosity variations.

- [May, Moresi, 2008] introduces $\mathbf{C}_w$, $\mathbf{D}_w$ based on entries of $\mathbf{A}$
- [Rudi, Malossi, Isaac, et al., 2015] uses $\mathbf{C}_w = \mathbf{D}_w := \mathrm{diag}(\mathbf{A})$
- [Rudi, Stadler, Ghattas, 2017] proposes $\mathbf{C}_w = \mathbf{D}_w := \tilde{\mathbf{M}}_{\boldsymbol{u}}(\sqrt{\mu})$

Benchmark problem: Multiple sinkers at random locations

Two parameters increase problem difficulty:

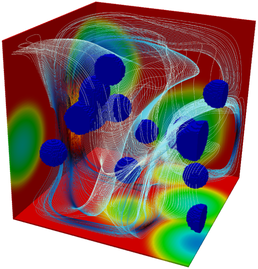
- Number of sinkers n at random points $\boldsymbol{c}_i$
- Dynamic ratio $\mathrm{DR}(\mu) \coloneqq \mu_{\max}/\mu_{\min}$

Smooth but highly varying viscosity μ is defined as:

$$\mu(\boldsymbol{x}) \coloneqq (\mu_{\max} - \mu_{\min})(1 - \chi_n(\boldsymbol{x})) + \mu_{\min}$$

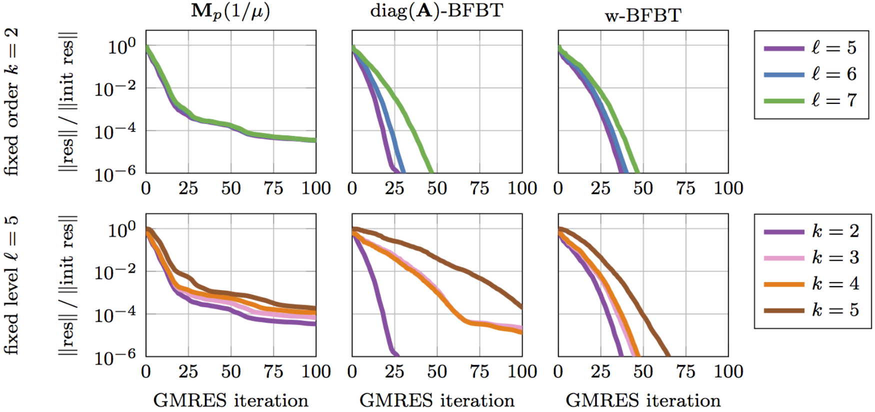
$$\chi_n(\boldsymbol{x}) \coloneqq \prod_{i=1}^{n} 1 - \exp\left[-d \max\left(0, |\boldsymbol{c}_i - \boldsymbol{x}| - \frac{w}{2}\right)^2\right]$$

(where $\mu_{\min}, \mu_{\max}, d, w$ are constant)



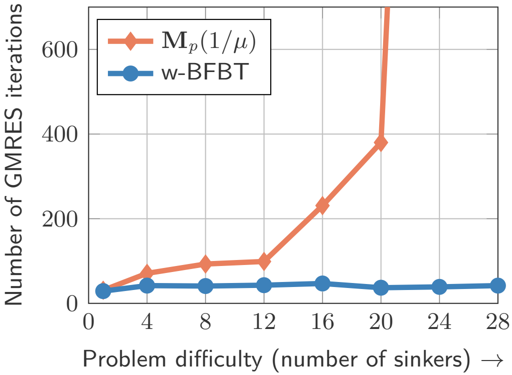
Smooth viscosity *(colors)* with highest value *(blue)* assumed inside spheres; streamlines show velocity field.

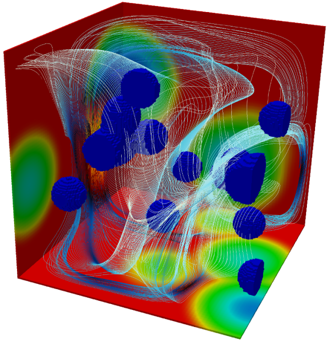
Comparison of Schur complement preconditioners



- Convergence of GMRES for benchmark problem with challenging viscosity μ
- k is velocity discretization order and ℓ is refinement level of uniform mesh
- w-BFBT, where $\mathbf{C}_w = \mathbf{D}_w := \tilde{\mathbf{M}}_{\boldsymbol{u}}(\sqrt{\mu})$, combines robust convergence of $\mathrm{diag}(\mathbf{A})$-BFBT with improved algorithmic scalability when order k increases

Robustness of w-BFBT w.r.t. viscosity variations





- Graph shows excerpt from more extensive numerical study
- Preconditioner $\mathbf{M}_p(1/\mu)$ becomes ineffective as sinker count increases
- w-BFBT is largely unaffected by viscosity variations, which makes it advantageous for highly heterogeneous problems

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.

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

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

Proof idea (Spectral equivalence for w-BFBT)

1. Establish a "sup-form" for approx. and exact Schur complements:

$$\left(\tilde{S}_{\text{w-BFBT}}\, q\,,\, q\right) = \sup_{p} \frac{\left(B^* p\,,\, wB^* q\right)^2}{\left(wAwB^* p\,,\, B^* p\right)}$$

$$\left(Sq\,,\, q\right) = \sup_{\boldsymbol{v}} \frac{\left(\boldsymbol{v}\,,\, wB^* q\right)^2}{\left(wAw\boldsymbol{v}\,,\, \boldsymbol{v}\right)}$$

2. Lower estimate (with constant one) follows immediately.
3. For the upper estimate, derive that

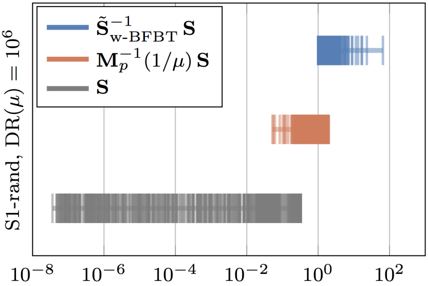
$$\frac{1}{2\,C_{\mu,w}} \left\| wB^* q \right\|^2_{(H^{-1}(\Omega))^d} \le \left(\tilde{S}_{\text{w-BFBT}}\, q\,,\, q\right),$$

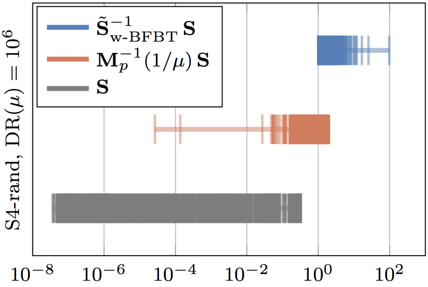
$$\left(Sq\,,\, q\right) \le \sup_{\boldsymbol{v}} \frac{\left\| w^{-1}\boldsymbol{v} \right\|^2_{(H^1(\Omega))^d} \left\| wB^* q \right\|^2_{(H^{-1}(\Omega))^d}}{2\left\| \sqrt{\mu}\, \frac{1}{2}(\nabla \boldsymbol{v} + \nabla \boldsymbol{v}^{\mathsf{T}}) \right\|^2_{(L^2(\Omega))^{d\times d}}}.$$

Result follows with weighted Poincaré–Friedrichs' and Korn's ineq.

Spectrum comparisons of preconditioned Schur matrices

2D Stokes problem discretized with $\mathbb{P}_2^{\text{bubble}} \times \mathbb{P}_1^{\text{disc}}$ elements (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 with respect to viscosity variations.

Outline

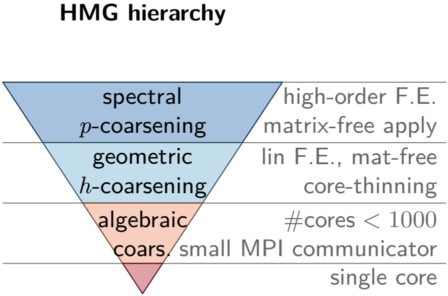
Earth's mantle convection: Driving application & solver challenges

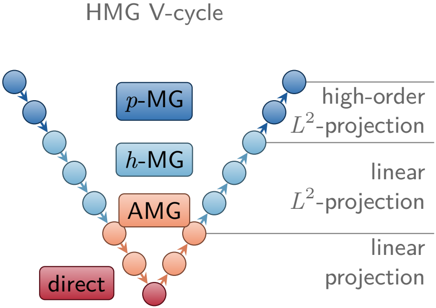
Preconditioner for the inverse Schur complement: Weighted BFBT

Preconditioning with Hybrid Spectral–Geometric–Algebraic Multigrid

Numerical results: Algorithmic & parallel scalability

HMG: Hybrid spectral–geometric–algebraic multigrid

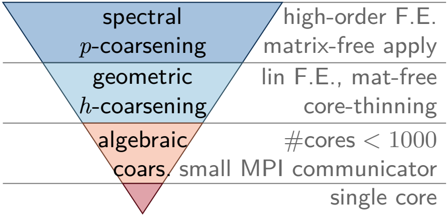




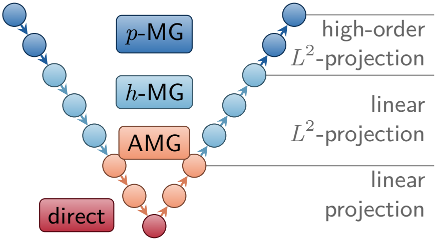
- Multigrid hierarchy of nested meshes is generated from an adaptively refined octree-based mesh via spectral–geometric coarsening
- Re-discretization of PDEs at coarser levels
- Parallel repartitioning of coarser meshes for load-balancing (crucial for AMR); sufficiently coarse meshes occupy only subsets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

HMG: Hybrid spectral–geometric–algebraic multigrid

HMG hierarchy

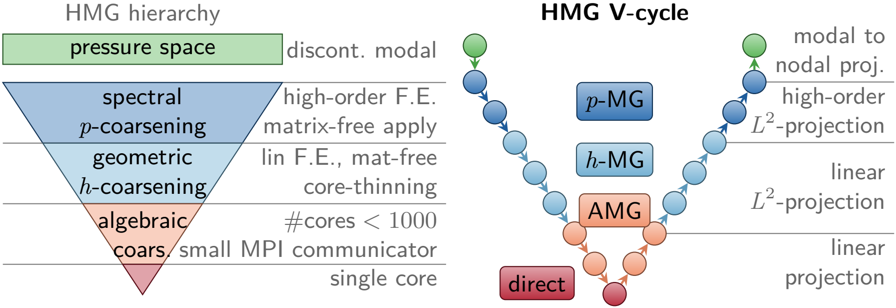


HMG V-cycle



- 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

HMG: Hybrid spectral–geometric–algebraic multigrid

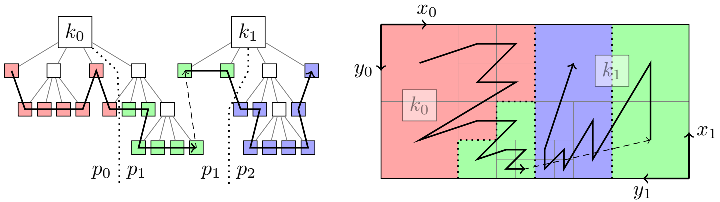


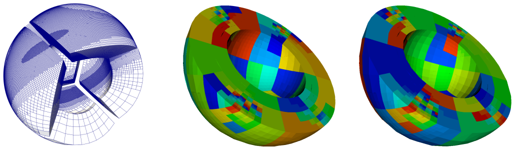
- High-order L^2-projection onto coarser levels; restriction & interpolation are adjoints of each other in L^2-sense
- Chebyshev accelerated Jacobi smoother (Cheb. from PETSc) with tensorized matrix-free high-order stiffness apply; assembly of high-order diagonal only
- Efficacy, i.e., error reduction,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

p4est: Parallel forest-of-octrees AMR library [p4est.org]

Scalable geometric multigrid coarsening due to:

- Forest-of-octree based meshes enable fast refinement/coarsening
- Octrees and space filling curves used for fast neighbor search, mesh repartitioning, and 2:1 mesh balancing in parallel

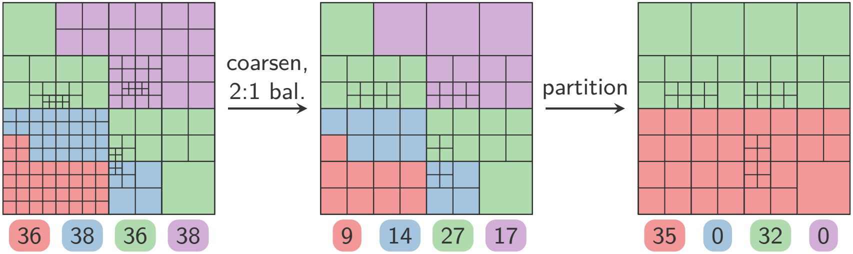




Colors depict different processor cores. (Credit: Burstedde, et al.)

Geometric coarsening: Repartitioning & core-thinning

- Parallel repartitioning of locally refined 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



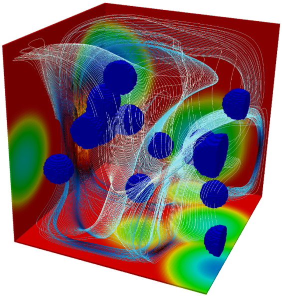
Colors depict different processor cores, *numbers* indicate element count on each core.

[Sundar, Biros, Burstedde, Rudi, Ghattas, Stadler, 2012]

Outline

Algorithmic scalability for HMG+w-BFBT

Number of iterations for solving elliptic sub-systems $\mathbf{Au} = \mathbf{f}$, $(\mathbf{BD}_w^{-1}\mathbf{B}^\mathsf{T})\mathbf{p} = \mathbf{Kp} = \mathbf{g}$, and full Stokes system for benchmark sinker problem.



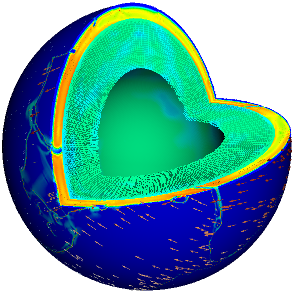
Vary mesh refinement ℓ for fixed order $k = 2$

ℓ	$\boldsymbol{u}$-DOF $[\times 10^6]$	It. **A**	p-DOF $[\times 10^6]$	It. **K**	DOF $[\times 10^6]$	It. **Stokes**
4	0.11	18	0.02	8	0.12	40
5	0.82	18	0.13	7	0.95	33
6	6.44	18	1.05	6	7.49	33
8	405.02	18	67.11	6	472.12	34
10	25807.57	18	4294.97	6	30102.53	34

Vary order k for fixed mesh refinement $\ell = 5$

k	$\boldsymbol{u}$-DOF $[\times 10^6]$	It. **A**	p-DOF $[\times 10^6]$	It. **K**	DOF $[\times 10^6]$	It. **Stokes**
2	0.82	18	0.13	7	0.95	33
3	2.74	20	0.32	8	3.07	37
4	6.44	20	0.66	7	7.10	36
6	21.56	23	1.84	12	23.40	50
8	50.92	22	3.93	10	54.86	67

Parallel scalability: Global mantle convection problem setup



Discretization parameters to test parallel scalability:

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

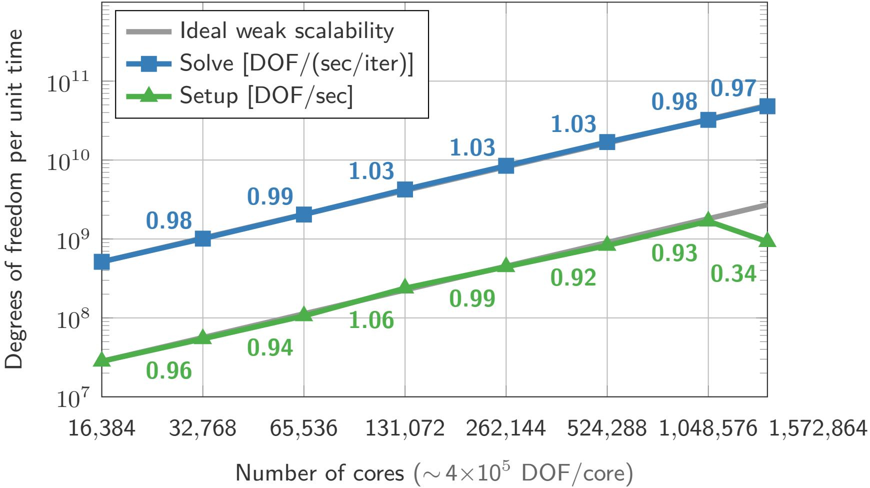
Multigrid parameters for elliptic blocks $\mathbf{A}$ and $\mathbf{K}$:

- 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]

Summary & References

Summary of results:

- Weighted BFBT preconditioner for the for the Schur complement; scalable HMG-based BFBT algorithms, heterogeneity-robust weighting of BFBT and theoretical foundation.
- Hybrid spectral–geometric–algebraic multigrid; based on p4est library.
- Optimal or nearly optimal algorithmic scalability.
- Parallel scalability of solvers to 1.6 million cores.

References:

- Rudi, Stadler, Ghattas, SIAM J. Sci. Comput.(2017), to appear.
- Rudi, Malossi, Isaac, Stadler, Gurnis, Ineichen, Bekas, Curioni, and Ghattas, Proceedings of SC15 (2015), Gordon Bell Prize.
- Sundar, Biros, Burstedde, Rudi, Ghattas, and Stadler, Proceedings of SC12 (2012).