On nonlinear adaptivity with heterogeneity

Jed Brown jed@jedbrown.org (CU Boulder)
Collaborators: Mark Adams (LBL), Matt Knepley (UChicago),
Dave May (ETH), Laetitia Le Pourhiet (UPMC),
Ravi Samtaney (KAUST)

Copper Mountain Multigrid Conference, 2017-03-30

This talk: https://jedbrown.org/files/20170330-AdaptHeterogeneous.pdf

DD6 (1992) in Como, Italy

Multigrid Solvers on Decomposed Domains

ACHI BRANDT and BORIS DISKIN

ABSTRACT. For general nonlinear elliptic problems with many gridpoints per processor, a domain-decomposed multigrid algorithm is described. It solves a problem in essentially the same work as needed for solving just once, by the fastest solver, a separate problem in each subdomain. During the entire solution process, only few episodes of data transfer between processors are needed, and the total amount of transferred data is small compared with the size of the decomposition interfaces. A mode analysis and numerical tests are reported.

Plan: ruthlessly eliminate communication

- Eliminate, not “aggregate and amortize”

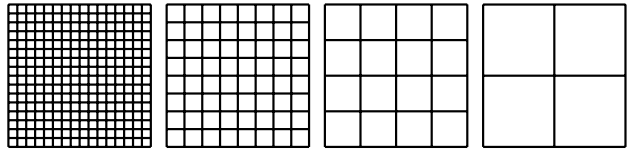
Why?

- Enables pruning unnecessary work
- More scope for dynamic load balance
- Tolerance for high-frequency load imbalance
  - From irregular computation or hardware error correction
- Local recovery despite global coupling

Requirements

- Must retain optimal convergence with good constants
- Flexible, robust, and debuggable

Multigrid Preliminaries



Multigrid is an $O(n)$ method for solving algebraic problems by defining a hierarchy of scale. A multigrid method is constructed from:

1. a sequence of discretizations
   - coarser approximations of problem, same or different equations
   - constructed algebraically or geometrically
2. intergrid transfer operators
   - residual restriction I_h^H (fine to coarse)
   - state restriction $\hat{I}_h^H$ (fine to coarse)
   - partial state interpolation I_H^h (coarse to fine, 'prolongation')
   - state reconstruction $\mathbb{I}_H^h$ (coarse to fine)
3. Smoothers (S)
   - correct the high frequency error components
   - Richardson, Jacobi, Gauss-Seidel, etc.
   - Gauss-Seidel-Newton or optimization methods
   - Compatible Monte Carlo, ...

τ formulation of Full Approximation Scheme (FAS)

- classical formulation: "coarse grid *accelerates* fine grid ↘↗
- τ formulation: "fine grid feeds back into coarse grid" ↗↘
- To solve $Nu = f$, recursively apply

$$\text{pre-smooth} \quad \tilde{u}^h \leftarrow S^h_{\text{pre}}(u^h_0, f^h)$$

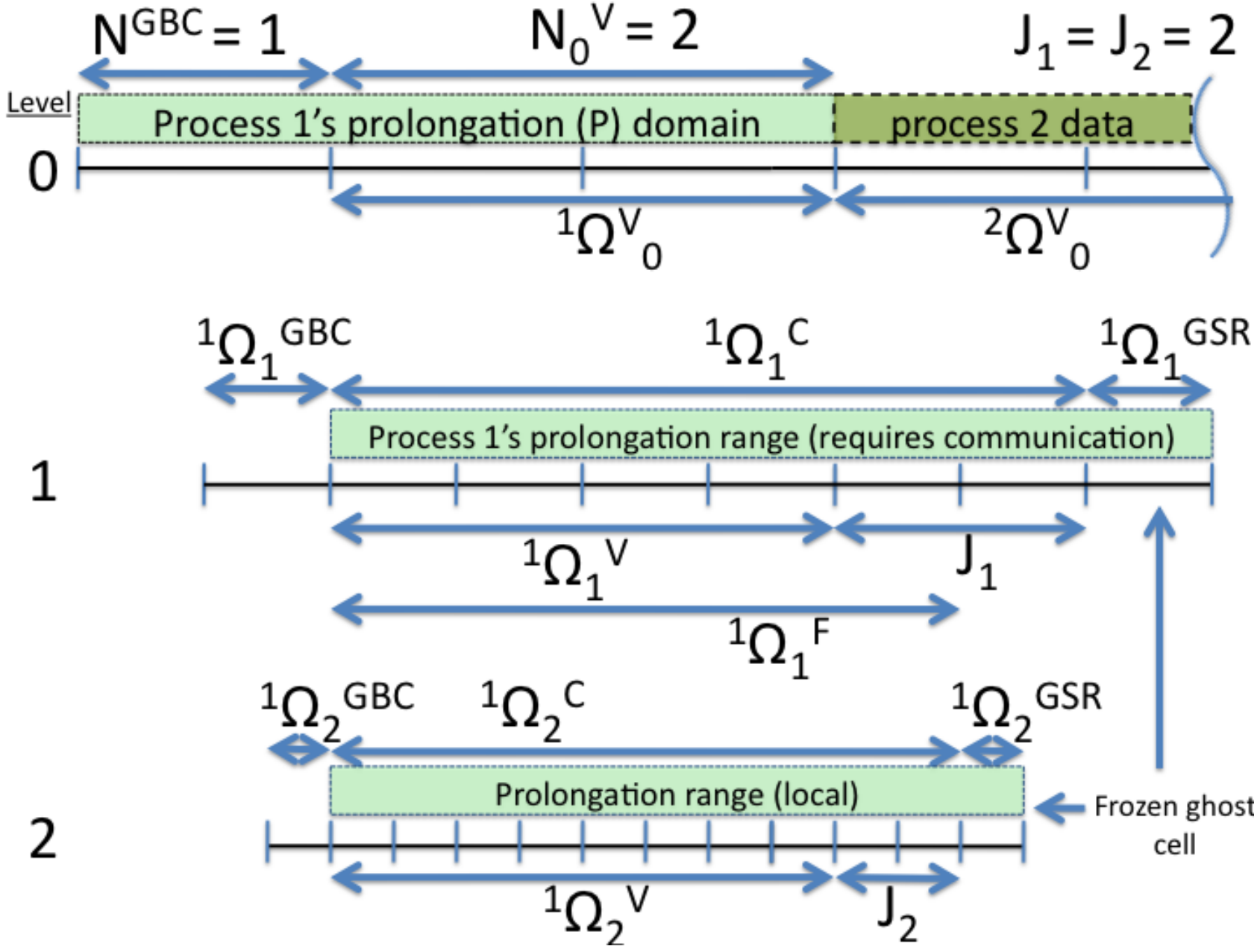
$$\text{solve coarse problem for } u^H \quad N^H u^H = \underbrace{I^H_h f^h}_{f^H} + \underbrace{N^H \hat{I}^H_h \tilde{u}^h - I^H_h N^h \tilde{u}^h}_{\tau^H_h}$$

$$\text{correction and post-smooth} \quad u^h \leftarrow S^h_{\text{post}}\left(\tilde{u}^h + I^h_H(u^H - \hat{I}^H_h \tilde{u}^h), f^h\right)$$

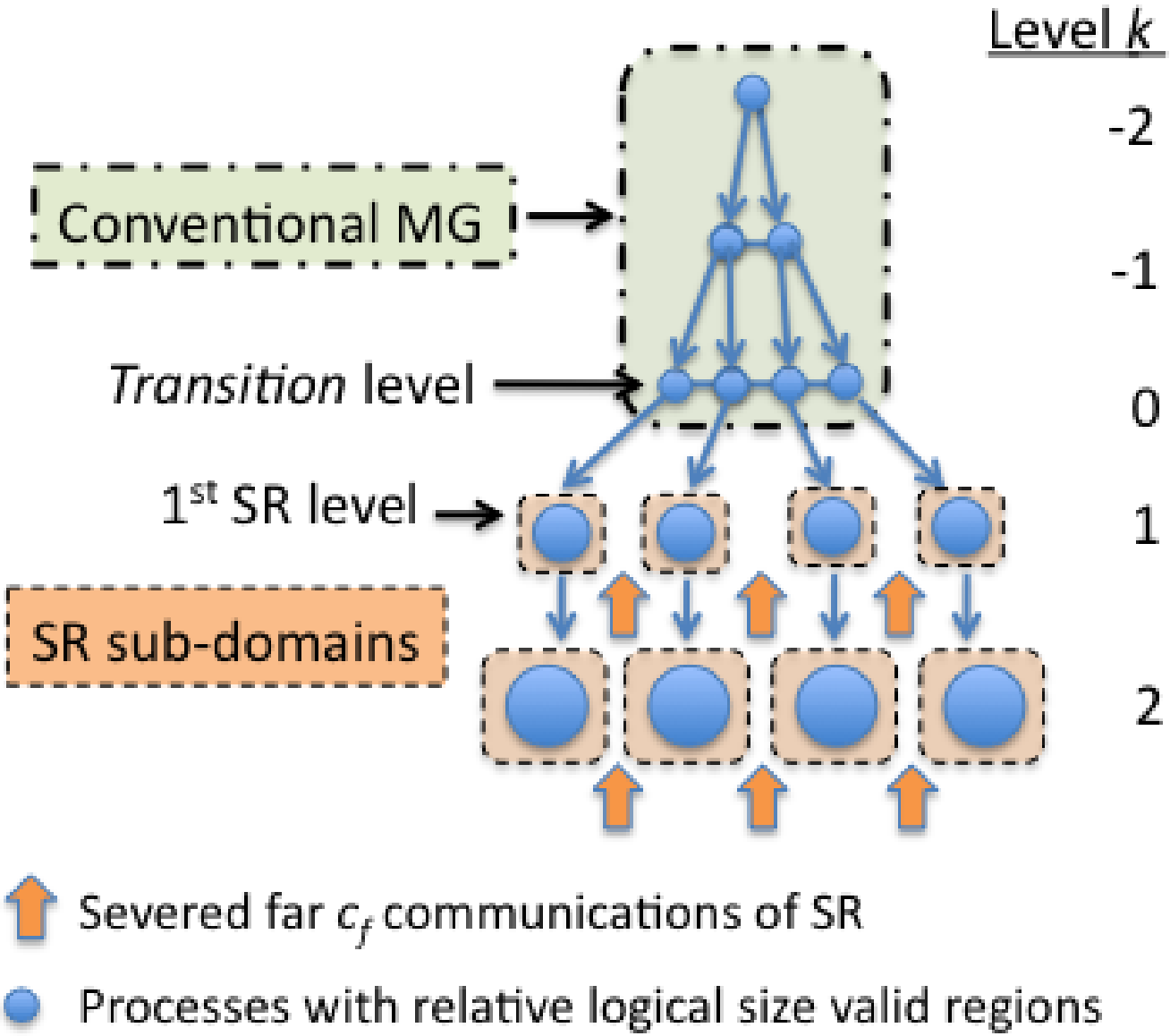
I^H_h	residual restriction	$\hat{I}^H_h$	solution restriction
I^h_H	solution interpolation	$f^H = I^H_h f^h$	restricted forcing
$\{S^h_{\text{pre}}, S^h_{\text{post}}\}$	smoothing operations on the fine grid		

- At convergence, $u^{H*} = \hat{I}^H_h u^{h*}$ solves the τ-corrected coarse grid equation $N^H u^H = f^H + \tau^H_h$, thus τ^H_h is the "fine grid feedback" that makes the coarse grid equation accurate.
- τ^H_h is *local* and need only be recomputed where it becomes stale.
- Interpretation by Achi Brandt in 1977, many tricks followed

Segmental refinement: dependencies

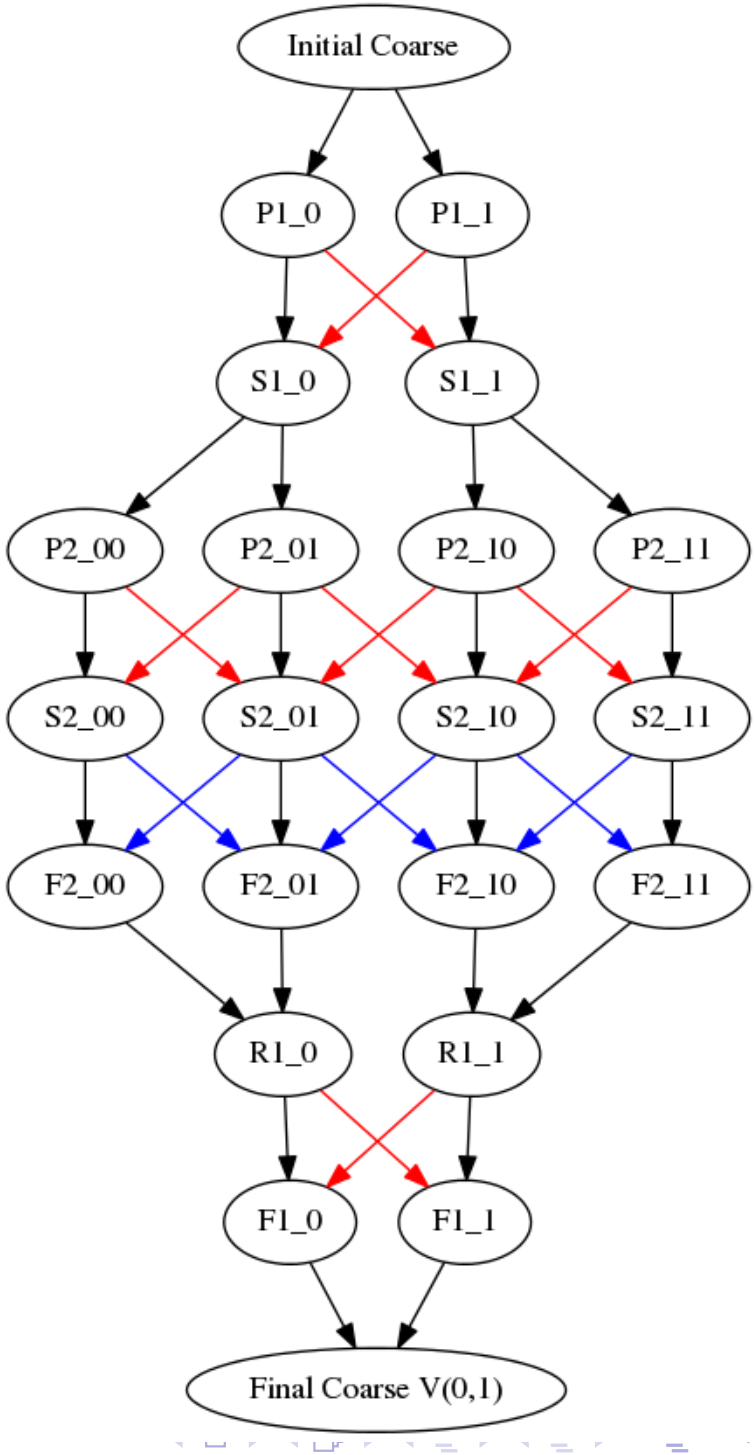


Segmental refinement: parallel



Low communication MG

- red arrows can be removed by τ-FAS with overlap
- blue arrows can also be removed, but then algebraic convergence stalls when discretization error is reached
- no simple way to check that discretization error is obtained
- if fine grid state is not stored, use compatible relaxation to complete prolongation P
- “Segmental refinement” by Achi Brandt (1977)
- 2-process case by Brandt and Diskin (1994)

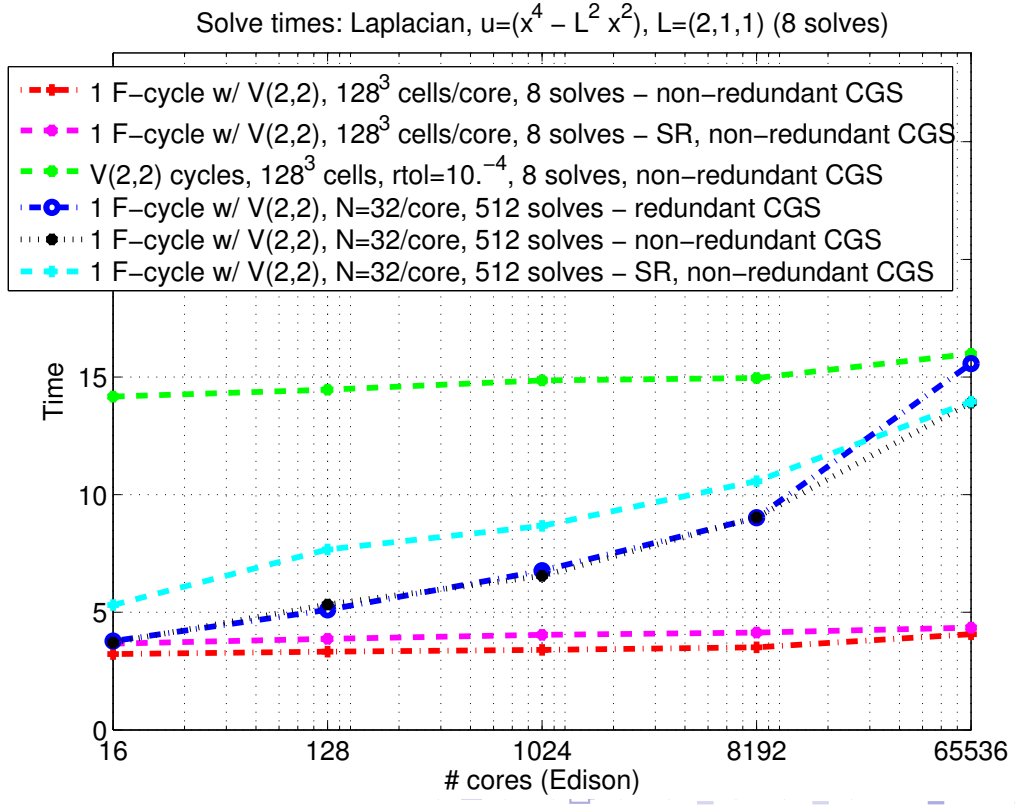


Segmental refinement: no horizontal communication

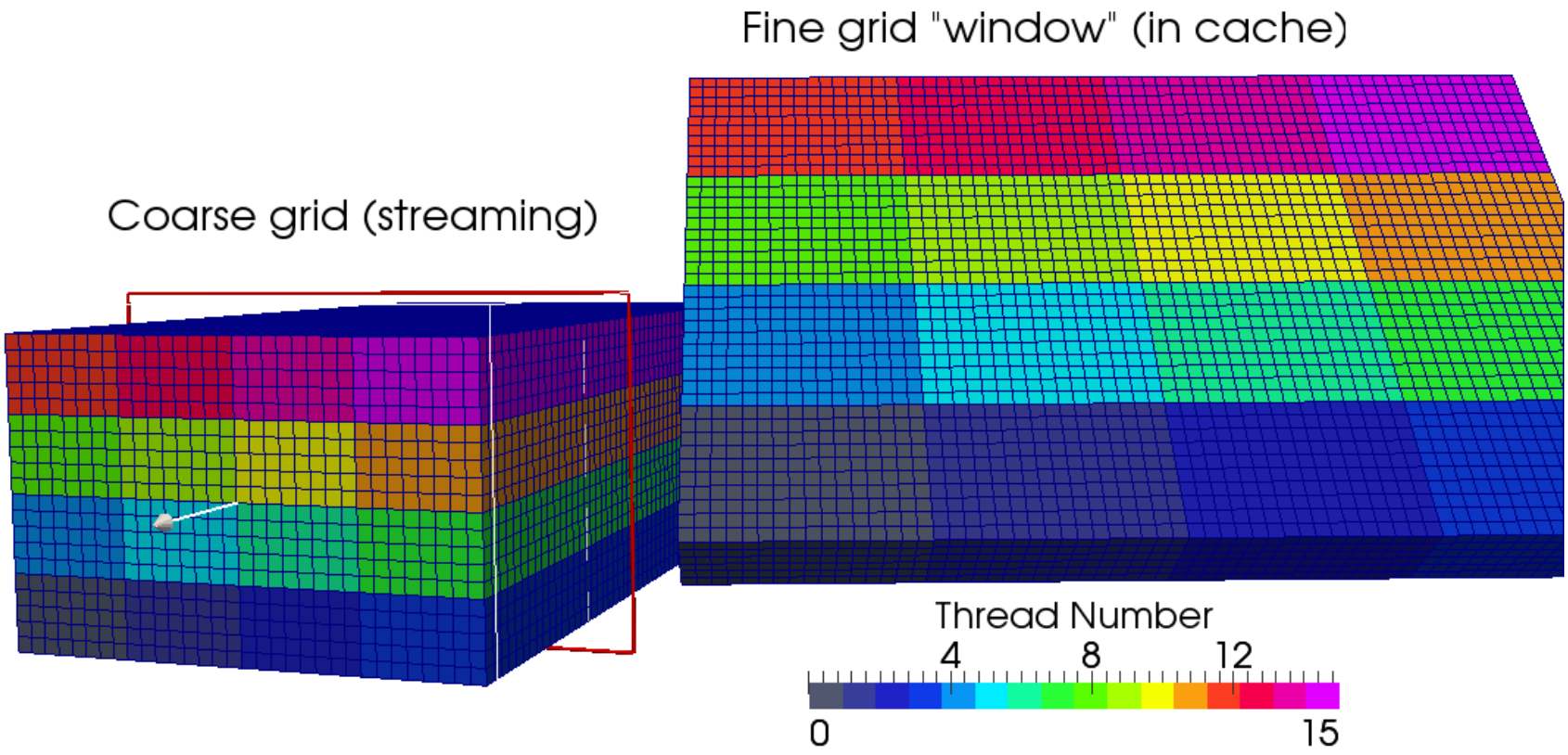
- Adams, Brown, Knepley, Samtaney (SISC 2016)
- 27-point second-order stencil, manufactured analytic solution
- 5 SR levels: 16^3 cells/process local coarse grid
- Overlap $=$ Base $+ (L - \ell)$Increment
  - Implementation requires even number of cells—round down.
- FMG with $V(2,2)$ cycles

Table: $\|e_{SR}\|_\infty / \|e_{FMG}\|_\infty$

	Base		
Increment	1	2	3
1	1.59	2.34	1.00
2	1.00	1.00	1.00
3	1.00	1.00	1.00

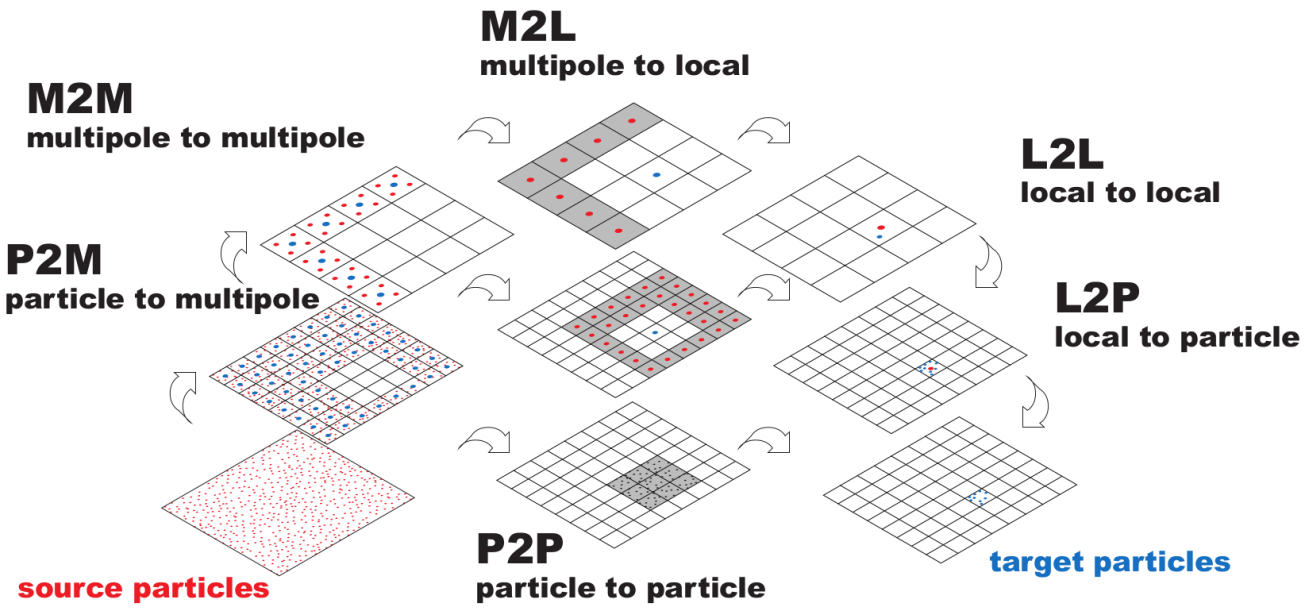


Reducing memory bandwidth



- Sweep through “coarse” grid with moving window
- Zoom in on new slab, construct fine grid “window” in-cache
- Interpolate to new fine grid, apply pipelined smoother (*s*-step)
- Compute residual, accumulate restriction of state and residual into coarse grid, expire slab from window

Connections to Fast Multipole Method / $\mathscr{H}$-matrices



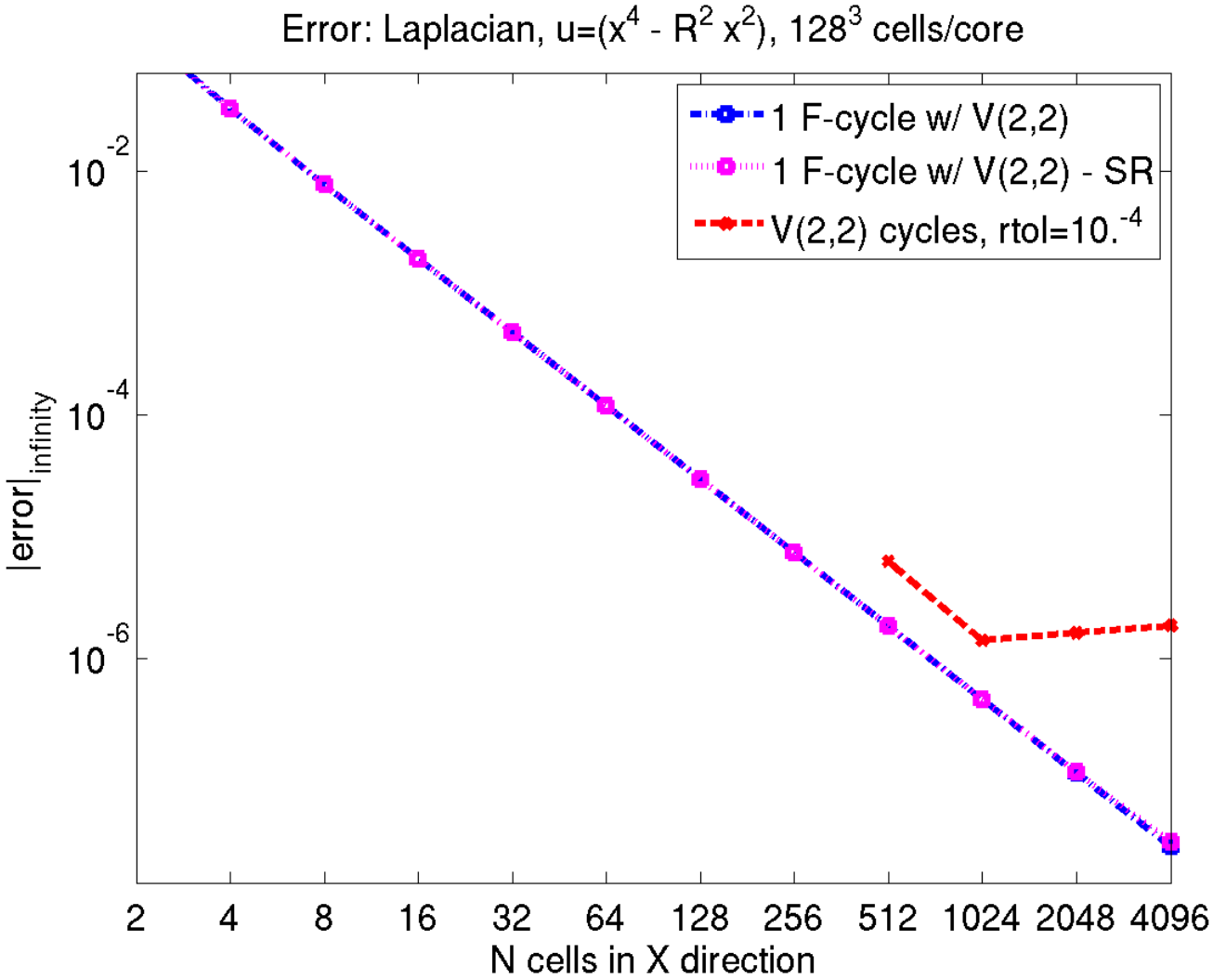
[Yokota, Barba (2011)]

- Can evaluate solution of nearby problems without solving everywhere
  - need τ correction from everywhere
- Segmental Refinement buffer regions $\sim$ separation criteria
- No need for Green's functions

Other uses of segmental refinement

- Compression of solutions, local decompression, resilience
- Transient adjoints
  - Adjoint model runs backward-in-time, needs state from solution of forward model
  - Status quo: hierarchical checkpointing
  - Memory-constrained and requires computing forward model multiple times
  - If forward model is stiff, each step has global dependence
  - Compression via τ-FAS accelerates recomputation, can be local
- Visualization and analysis
  - Targeted visualization in small part of domain
  - Interesting features emergent so can't predict where to look

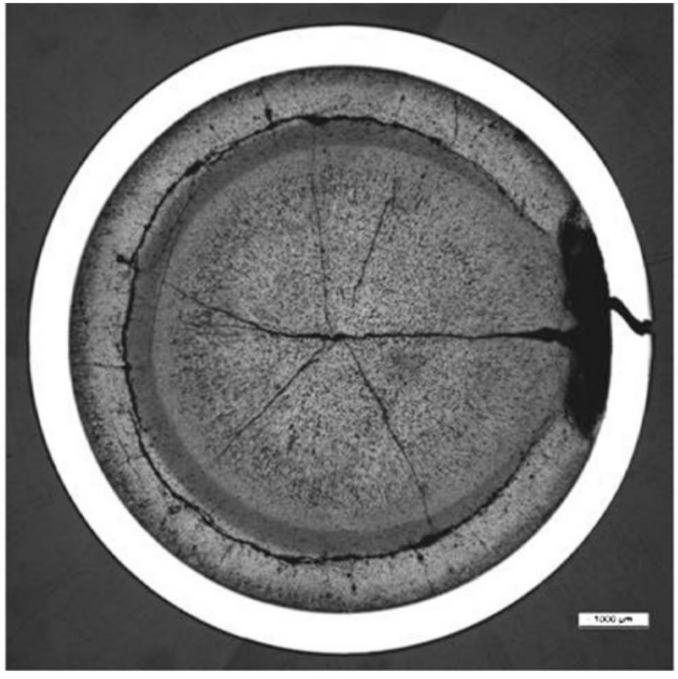
Tolerances and FMG



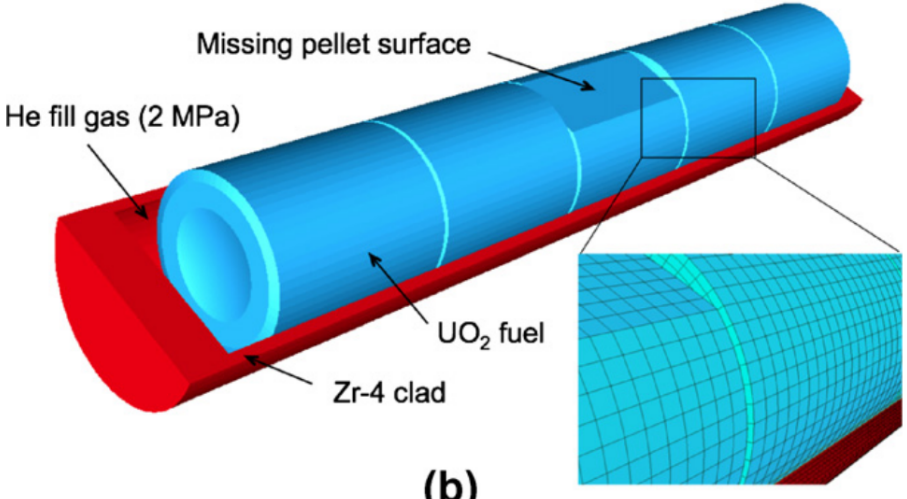
Continental rifting

Rifting Video

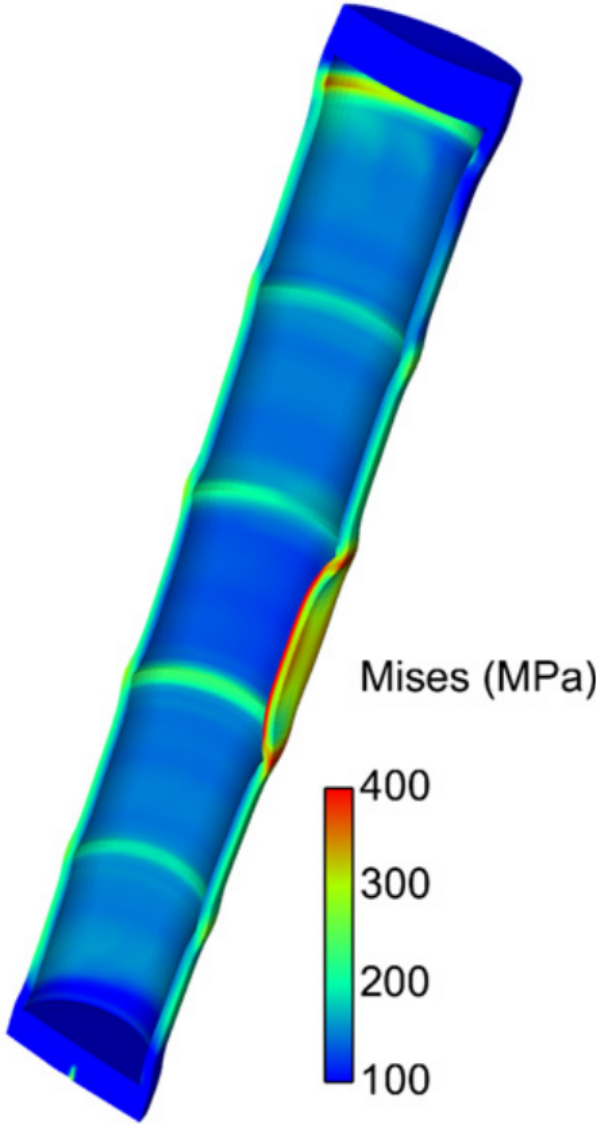
Nuclear fuel pellet-cladding mechanics



(a)



(b)



[Williamson et al (2012)]

Model problem: $\mathfrak{p}$-Laplacian with slip boundary conditions

- 2-dimensional model problem for power-law fluid cross-section

$$-\nabla\cdot\left(|\nabla u|^{\mathfrak{p}-2}\nabla u\right) - f = 0, \qquad 1 \leq \mathfrak{p} \leq \infty$$

  Singular or degenerate when $\nabla u = 0$

- Regularized variant

$$-\nabla\cdot(\eta\nabla u) - f = 0$$

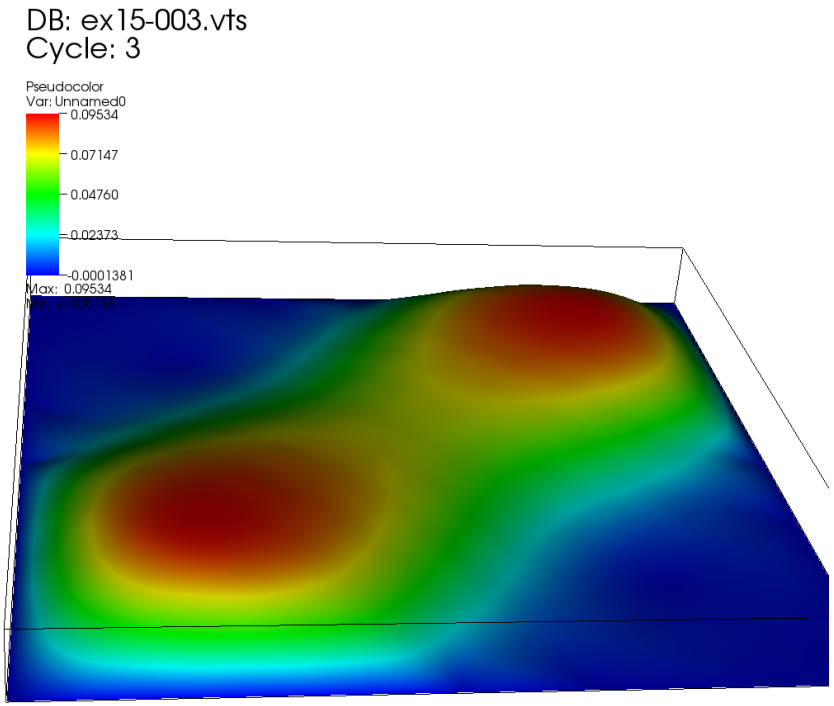
$$\eta(\gamma) = (\varepsilon^2 + \gamma)^{\frac{\mathfrak{p}-2}{2}} \qquad \gamma(u) = \frac{1}{2}|\nabla u|^2$$

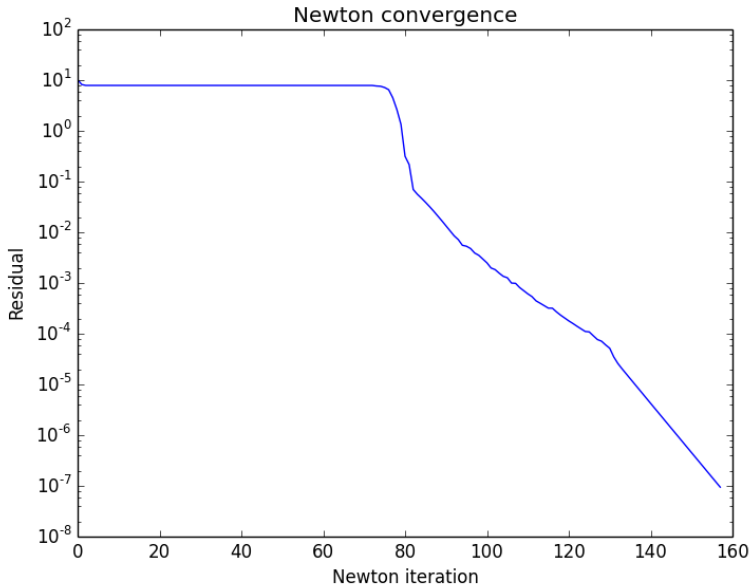
- Friction boundary condition on one side of domain

$$\nabla u \cdot n + A(x)\,|u|^{q-1}\,u = 0$$

Model problem: $\mathfrak{p}$-Laplacian with slip boundary conditions

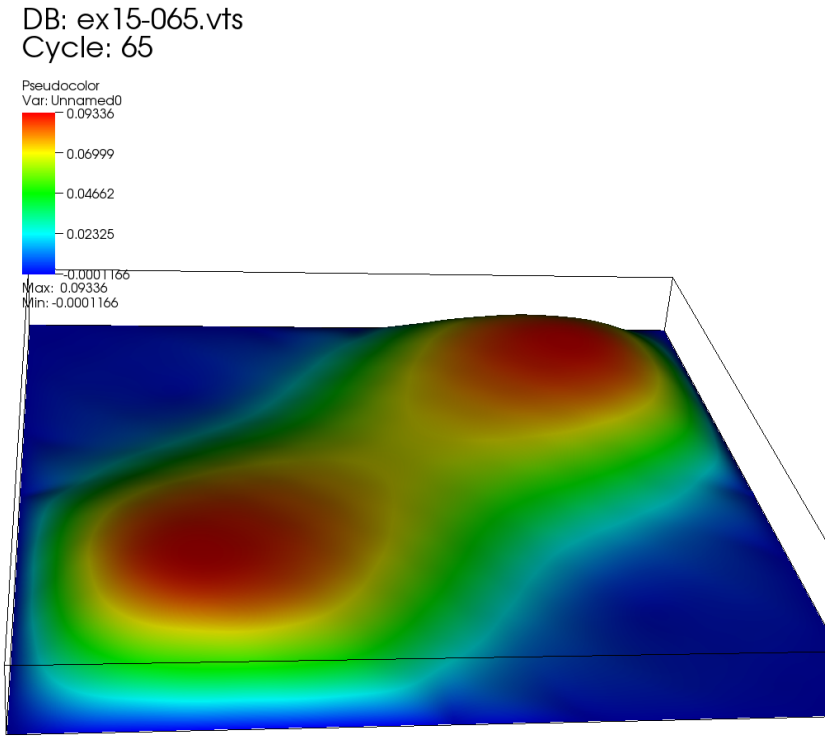
- $\mathfrak{p} = 1.3$ and $q = 0.2$, checkerboard coefficients $\{10^{-2}, 1\}$
- Friction coefficient $A = 0$ in center, 1 at corners

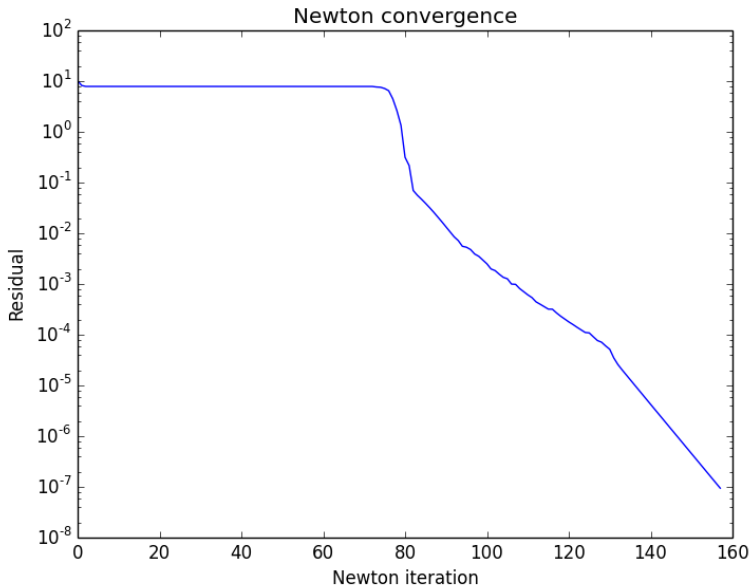




Model problem: $\mathfrak{p}$-Laplacian with slip boundary conditions

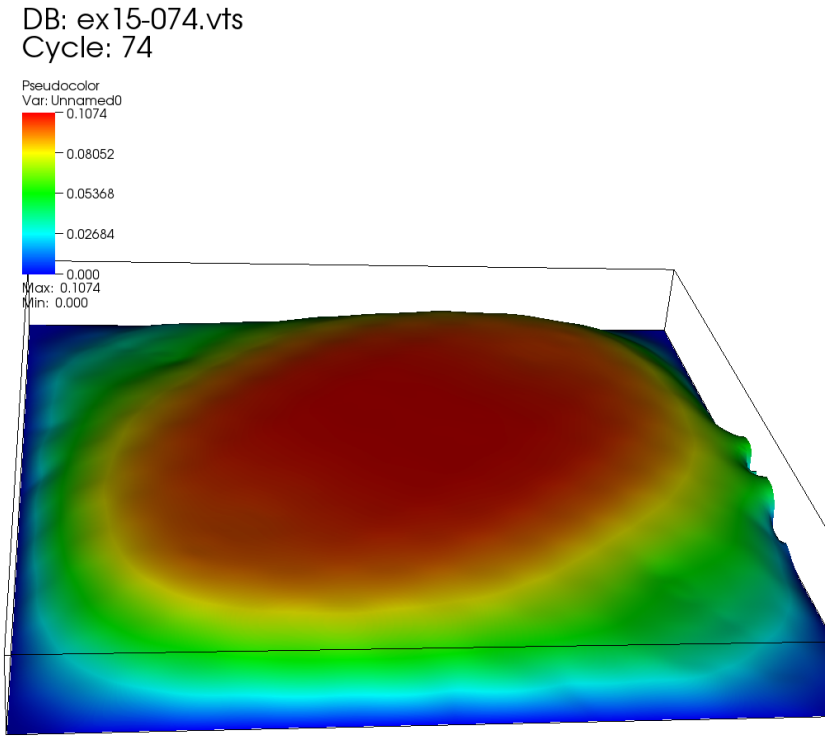
- $\mathfrak{p} = 1.3$ and $q = 0.2$, checkerboard coefficients $\{10^{-2}, 1\}$
- Friction coefficient $A = 0$ in center, 1 at corners

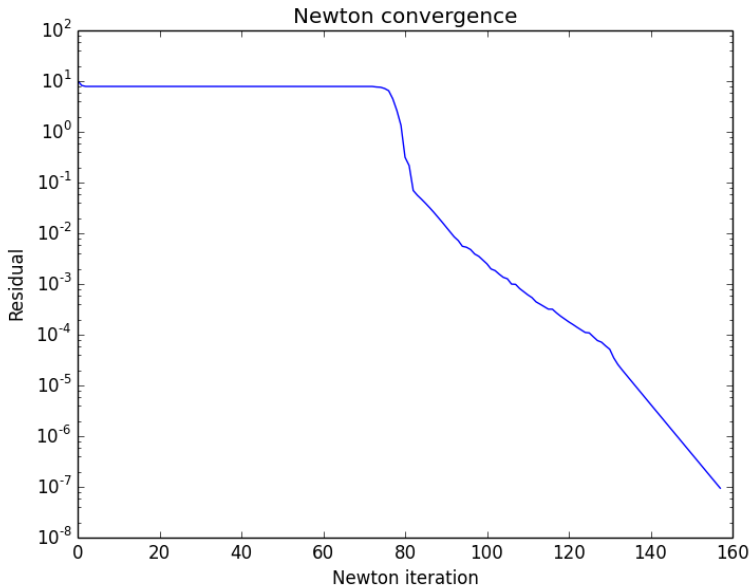




Model problem: $\mathfrak{p}$-Laplacian with slip boundary conditions

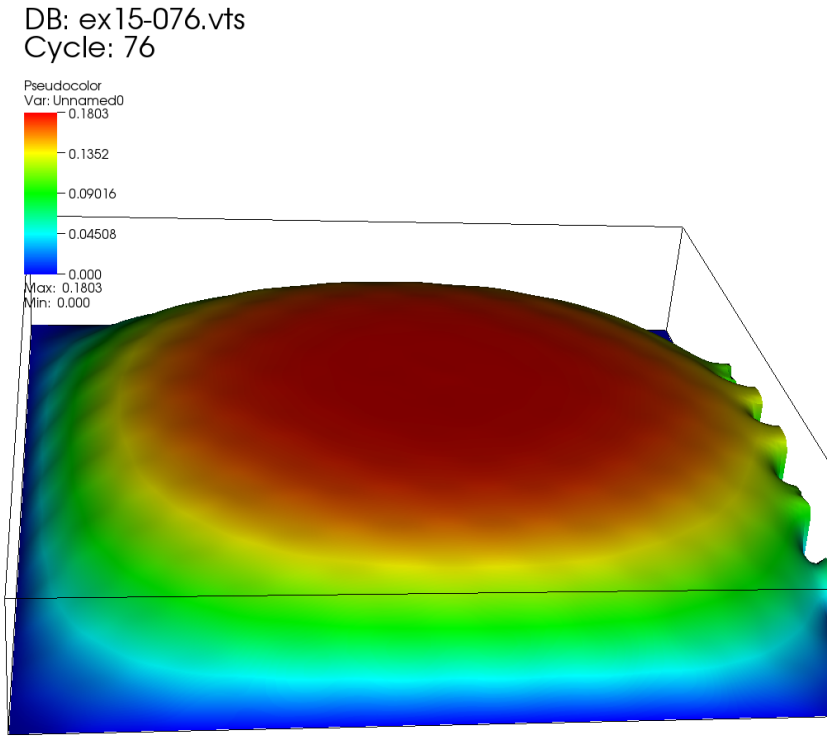
- $\mathfrak{p} = 1.3$ and $q = 0.2$, checkerboard coefficients $\{10^{-2}, 1\}$
- Friction coefficient $A = 0$ in center, 1 at corners

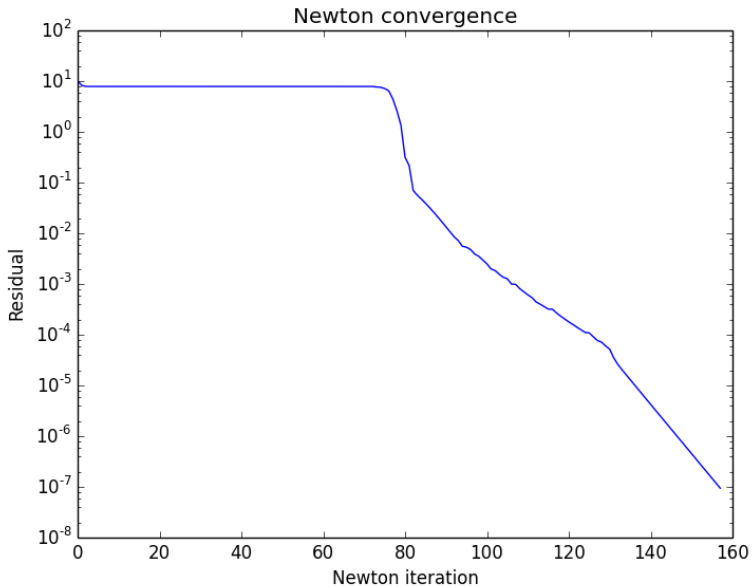




Model problem: $\mathfrak{p}$-Laplacian with slip boundary conditions

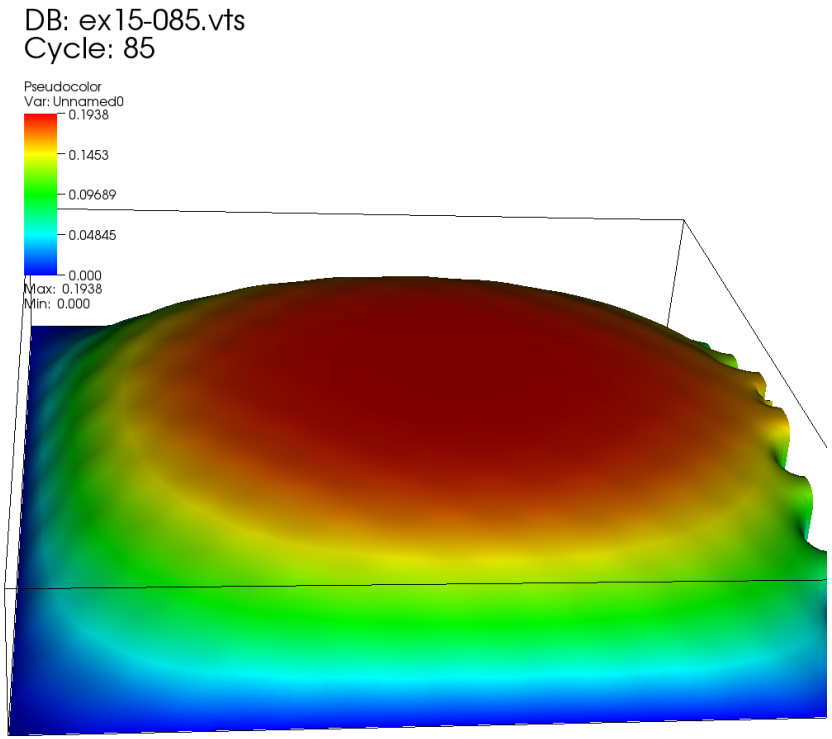
- $\mathfrak{p} = 1.3$ and $q = 0.2$, checkerboard coefficients $\{10^{-2}, 1\}$
- Friction coefficient $A = 0$ in center, 1 at corners

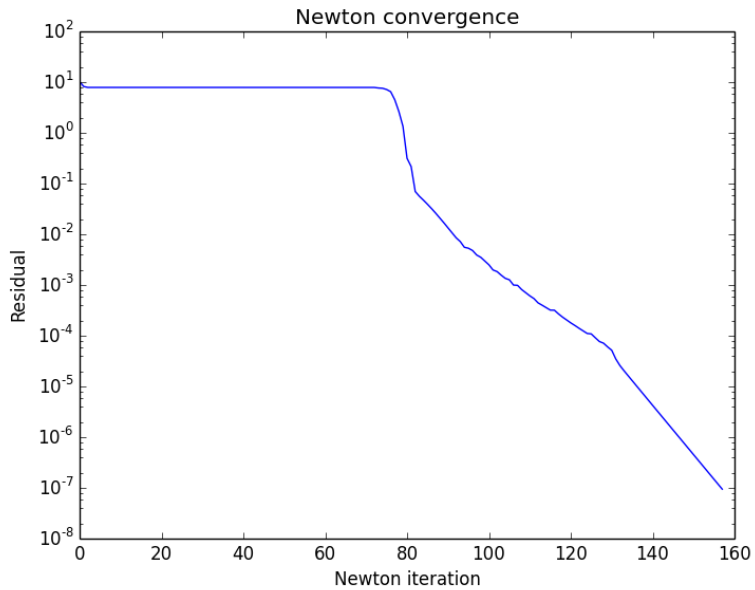




Model problem: $\mathfrak{p}$-Laplacian with slip boundary conditions

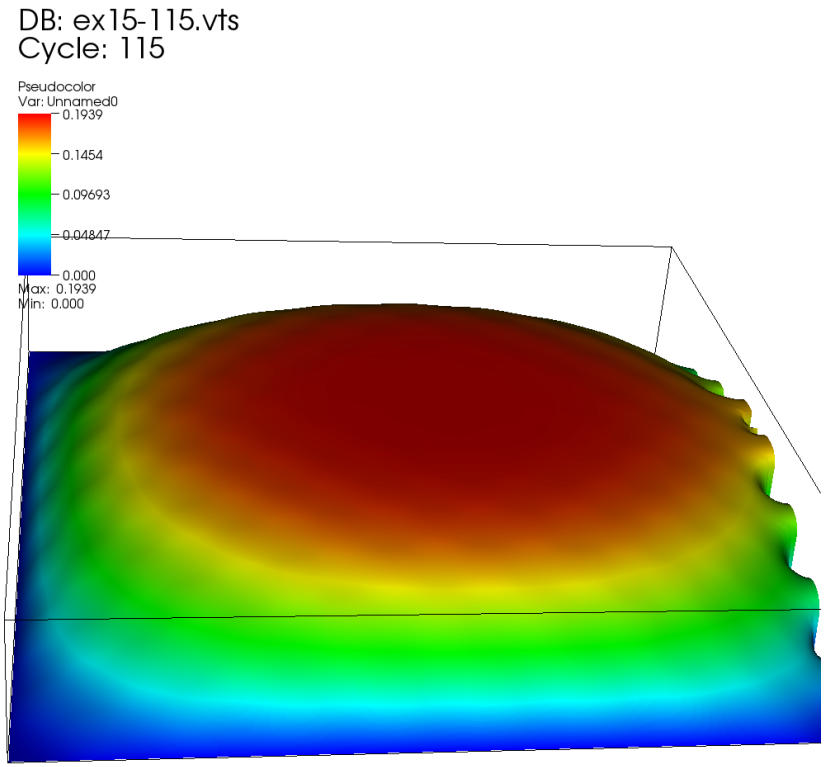
- $\mathfrak{p} = 1.3$ and $q = 0.2$, checkerboard coefficients $\{10^{-2}, 1\}$
- Friction coefficient $A = 0$ in center, 1 at corners

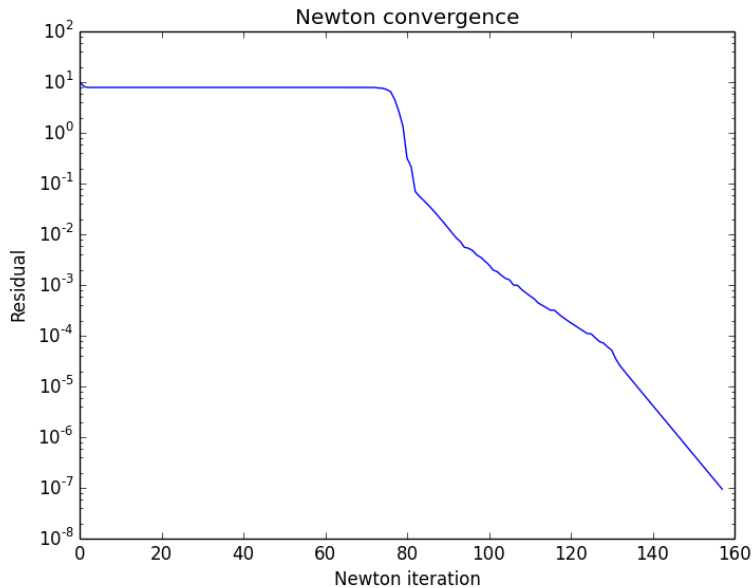




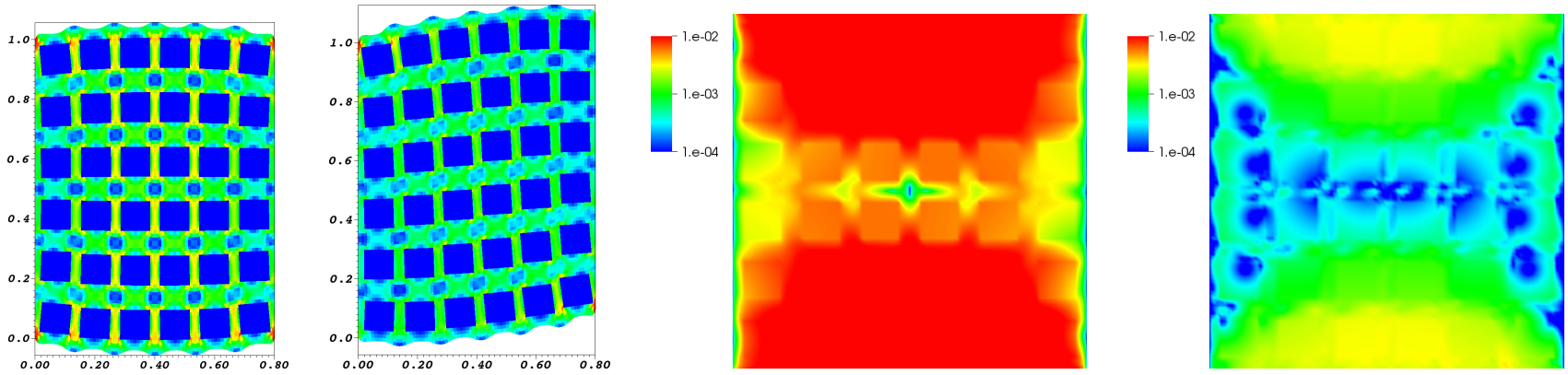
Model problem: $\mathfrak{p}$-Laplacian with slip boundary conditions

- $\mathfrak{p} = 1.3$ and $q = 0.2$, checkerboard coefficients $\{10^{-2}, 1\}$
- Friction coefficient $A = 0$ in center, 1 at corners





τ corrections

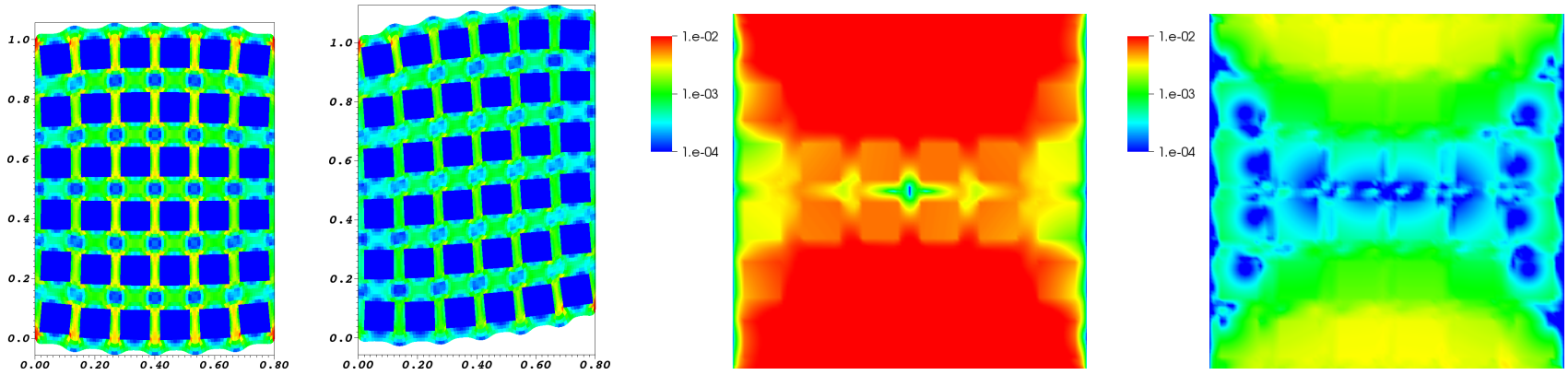


- Plane strain elasticity, $E = 1000, \nu = 0.4$ inclusions in $E = 1, \nu = 0.2$ material, coarsen by 3^2.
- Solve initial problem everywhere and compute $\tau_h^H = A^H \hat{I}_h^H u^h - I_h^H A^h u^h$
- Change boundary conditions and solve FAS coarse problem

$$N^H \acute{u}^H = \underbrace{I_h^H \acute{f}^h}_{\acute{f}^H} + \underbrace{N^H \hat{I}_h^H \tilde{u}^h - I_h^H N^h \tilde{u}^h}_{\tau_h^H}$$

- Prolong, post-smooth, compute error $e^h = \acute{u}^h - (N^h)^{-1} \acute{f}^h$
- Coarse grid *with* τ is nearly $10\times$ better accuracy

τ corrections



- Plane strain elasticity, $E = 1000, \nu = 0.4$ inclusions in $E = 1, \nu = 0.2$ material, coarsen by 3^2.
- Solve initial problem everywhere and compute $\tau_h^H = A^H \hat{I}_h^H u^h - I_h^H A^h u^h$
- Change boundary conditions and solve FAS coarse problem

$$N^H \acute{u}^H = \underbrace{I_h^H \acute{f}^h}_{\acute{f}^H} + \underbrace{N^H \hat{I}_h^H \tilde{u}^h - I_h^H N^h \tilde{u}^h}_{\tau_h^H}$$

- Prolong, post-smooth, compute error $e^h = \acute{u}^h - (N^h)^{-1} \acute{f}^h$
- Coarse grid *with* τ is nearly $10\times$ better accuracy

τ adaptivity: an idea for heterogeneous media

- Applications with localized nonlinearities
  - Subduction, rifting, rupture/fault dynamics
  - Carbon fiber, biological tissues, fracture
  - Frictional contact
- Adaptive methods fail for heterogeneous media
  - Rocks are rough, solutions are not “smooth”
  - Cannot build accurate coarse space without scale separation
- τ adaptivity
  - Fine-grid work needed everywhere at first
  - Then τ becomes accurate in nearly-linear regions
  - Only visit fine grids in “interesting” places: active nonlinearity, drastic change of solution

Comparison to nonlinear domain decomposition

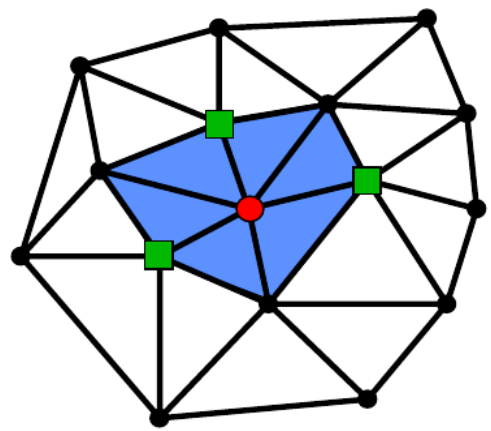
- ASPIN (Additive Schwarz preconditioned inexact Newton)
  - Cai and Keyes (2003)
  - More local iterations in strongly nonlinear regions
  - Each nonlinear iteration only propagates information locally
  - Many real nonlinearities are activated by long-range forces
    - locking in granular media (gravel, granola)
    - binding in steel fittings, crack propagation
  - Two-stage algorithm has different load balancing
    - Nonlinear subdomain solves
    - Global linear solve
- τ adaptivity
  - Minimum effort to communicate long-range information
  - Nonlinearity sees effects as accurate as with global fine-grid feedback
  - Fine-grid work always proportional to “interesting” changes

Nonlinear and matrix-free smoothing

- matrix-based smoothers require global linearization
- nonlinearity often more efficiently resolved locally
- nonlinear additive or multiplicative Schwarz
- nonlinear/matrix-free is good if

$$C = \frac{(\text{cost to evaluate residual at one “point”}) \cdot N}{(\text{cost of global residual})} \sim 1$$

  - finite difference: $C < 2$
  - finite volume: $C \sim 2$, depends on reconstruction
  - finite element: $C \sim$ number of vertices per cell
- larger block smoothers help reduce C
- additive correction (Jacobi/Chebyshev/multi-stage)
  - global evaluation, as good as $C = 1$
  - but, need to assemble corrector/scaling
  - need spectral estimates or wave speeds



Outlook

- τ adaptivity: benefits of AMR without fine-scale smoothness
- Coarse-centric restructuring is a major interface change
  - Algebraic coarsening?
- Nonlinear smoothers (and discretizations)
  - Smooth in neighborhood of “interesting” fine-scale features
  - Which discretizations can provide efficient matrix-free smoothers?
  - Does there exist an efficient smoother based on element Neumann problems?
- Weakening data dependencies enables dynamic load balancing
- Reliability of error estimates for refreshing τ
  - We want a coarse indicator for whether τ needs to change
  - Phase fields can provide such information

Thanks

DOE Office of Advanced Scientific Computing Research