Practical Multigrid Methods for Momentum Balance in Ice Sheets

This talk:
http://59A2.org/files/20150202-LIWGMultigrid.pdf

Jed Brown jed@jedbrown.org (ANL and CU Boulder)

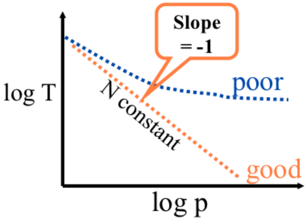
Land Ice Working Group, NCAR, 2015-02-02

Why do we need scalable solvers?

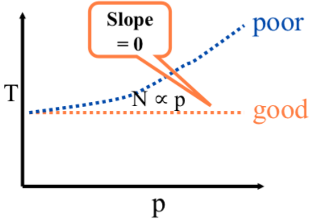
- Increasing resolution
  - larger problem sizes
  - more 3D effects visible
  - more time steps $\Longrightarrow$ smaller budget per time step
- Sequence of simulations – data assimilation, UQ
- All other costs typically linear in problem size

Review: two definitions of scalability

- **"Strong scaling"**
  - **execution time (*T*) decreases in inverse proportion to the number of processors (*p*)**
  - ***fixed size problem* (*N*) *overall***
  - **often instead graphed as reciprocal, "speedup"**



- **"Weak scaling" (memory bound)**
  - **execution time remains constant, as problem size and processor number are increased in proportion**
  - ***fixed size problem per processor***
  - **also known as "Gustafson scaling"**

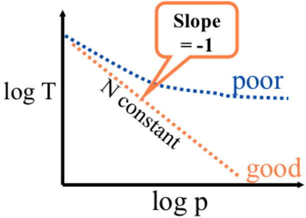


(c/o David Keyes)

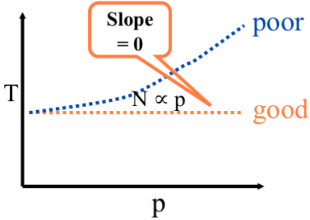
The easiest way to make software scalable is to make it sequentially inefficient. – Gropp 1999

Review: two definitions of scalability

- **"Strong scaling"**
  - **execution time (*T*) decreases in inverse proportion to the number of processors (*p*)**
  - ***fixed size problem* (*N*) *overall***
  - **often instead graphed as reciprocal, "speedup"**



- **"Weak scaling" (memory bound)**
  - **execution time remains constant, as problem size and processor number are increased in proportion**
  - ***fixed size problem per processor***
  - **also known as "Gustafson scaling"**



(c/o David Keyes)

The easiest way to make software scalable is to make it sequentially inefficient. – Gropp 1999

Is multigrid needed?

- Long-range coupling is slow to converge using local methods
  - Iteration count proportional to diameter of support of Green's functions
- How local are the Green's functions?
  - Columns of the inverse matrix

$$u(x) = \int_{y\in\Omega} G(x,y)f(y)$$

  - Sticky, flat bad – Green's functions local (and SIA is accurate)
  - Slippery bed (ice shelf), steep topography at high resolution
  - Pressure: surface is Dirichlet boundary condition, causes rapid decay

Bathymetry and stickyness distribution

- Bathymetry:
  - Aspect ratio $\varepsilon = [H]/[x] \ll 1$
  - Need surface *and* bed slopes to be small
- Stickyness distribution:
  - Limiting cases of plug flow versus vertical shear
  - Stress ratio: $\lambda = [\tau_{xz}]/[\tau_{\text{membrane}}]$
  - Discontinuous: frozen to slippery transition at ice stream margins
- Standard approach in glaciology:
  Taylor expand in ε and sometimes λ, drop higher order terms.

$\lambda \gg 1$ Shallow Ice Approximation (SIA), no horizontal coupling

$\lambda \ll 1$ Shallow Shelf Approximation (SSA), 2D elliptic solve in map-plane

  - Hydrostatic and various hybrids, 2D or 3D elliptic solves
- Bed slope is discontinuous and of order 1.
  - Taylor expansions no longer valid
  - Numerics require high resolution, subgrid parametrization, short time steps
  - Still get low quality results in the regions of most interest.
- Basal sliding parameters are discontinuous.

Hydrostatic equations for ice sheet flow

- Valid when $w_x \ll u_z$, independent of basal friction (Schoof&Hindmarsh 2010)
- Eliminate p and w from Stokes by incompressibility:
  3D elliptic system for $u = (u, v)$

$$-\nabla \cdot \left[\eta \begin{pmatrix} 4u_x + 2v_y & u_y + v_x & u_z \\ u_y + v_x & 2u_x + 4v_y & v_z \end{pmatrix} \right] + \rho g \overline{\nabla} h = 0$$

$$\eta(\theta, \gamma) = \frac{B(\theta)}{2}(\gamma_0 + \gamma)^{\frac{1-\mathfrak{n}}{2\mathfrak{n}}}, \qquad \mathfrak{n} \approx 3$$

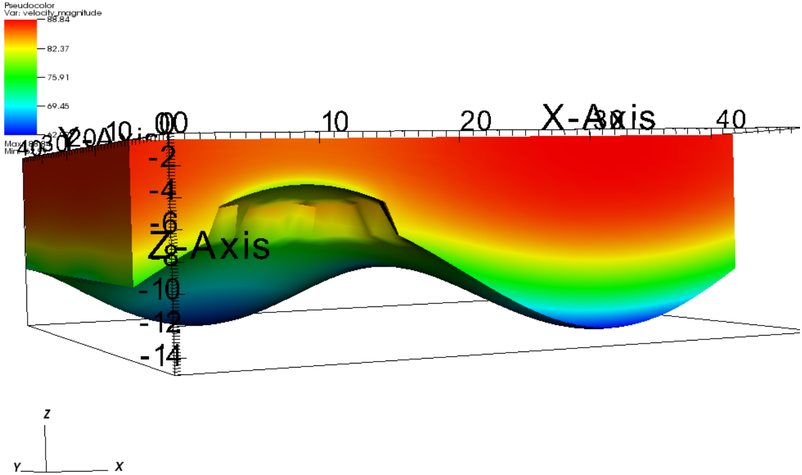
$$\gamma = u_x^2 + v_y^2 + u_x v_y + \frac{1}{4}(u_y + v_x)^2 + \frac{1}{4}u_z^2 + \frac{1}{4}v_z^2$$

and slip boundary $\sigma \cdot n = \beta^2 u$ where

$$\beta^2(\gamma_b) = \beta_0^2(\varepsilon_b^2 + \gamma_b)^{\frac{\mathfrak{m}-1}{2}}, \qquad 0 < \mathfrak{m} \leq 1$$

$$\gamma_b = \frac{1}{2}(u^2 + v^2)$$

- Q_1 FEM with Newton-Krylov-Multigrid solver in PETSc: `src/snes/examples/tutorials/ex48.c`



- Bathymetry is essentially discontinuous on any grid
- Shallow ice approximation produces oscillatory solutions
- Nonlinear and linear solvers have major problems or fail
- Grid sequenced Newton-Krylov multigrid works as well as in the smooth case

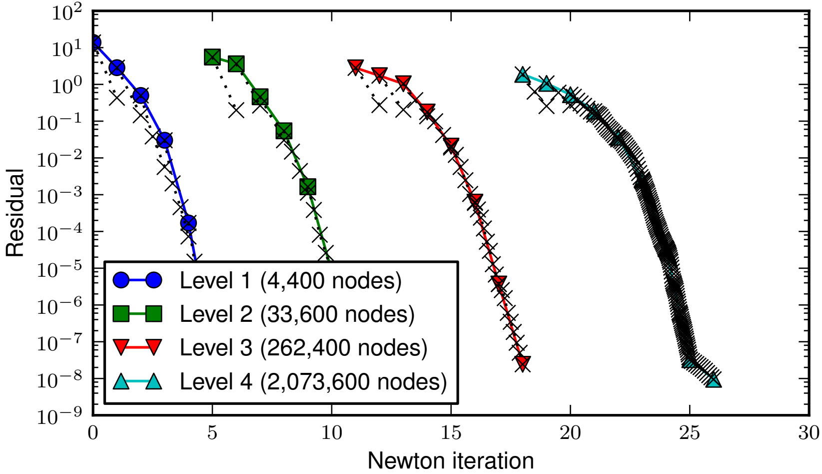


Figure: Grid sequenced Newton-Krylov convergence for test Y. The "cliff" has 58° angle in the red line (12×125 meter elements), 73° for the cyan line (6×62 meter elements).

Strong scaling on Blue Gene/P (Shaheen)

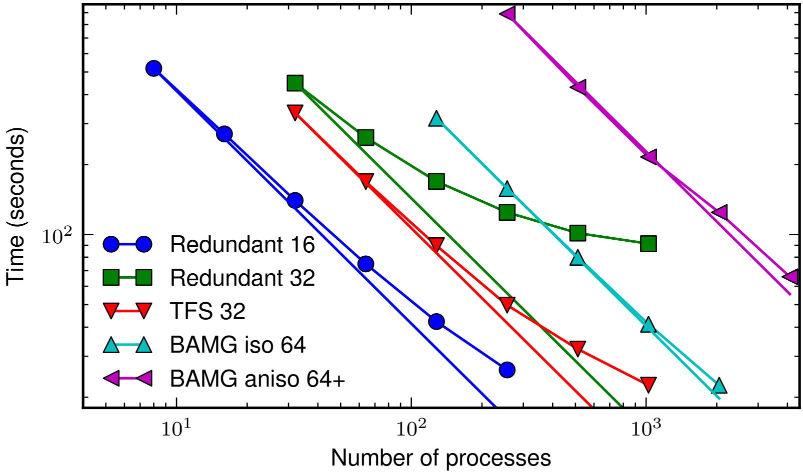


Figure: Strong scaling on Shaheen for different size coarse levels problems and different coarse level solvers. The straight lines on the strong scaling plot have slope -1 which is optimal.

Weak scaling on Blue Gene/P (Shaheen)

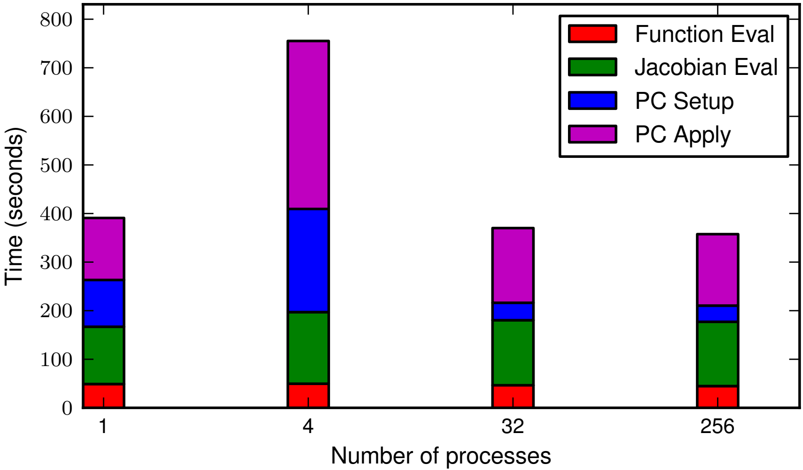


Figure: Weak scaling on Shaheen with a breakdown of time spent in different phases of the solution process. Times are for the full grid-sequenced problem, not just the finest level solve.

One high-accuracy solve costs 30 times as much as a residual evaluation

about 15 to reach truncation error

1000 times faster than some popular methods
e.g. Lemieux, Price, Evans, Knoll, Salinger, Holland, Payne 2011
(J. Computational Physics)
— Actual speedup subject to Amdahl's Law

(Brown, Smith, Ahmadia 2013, SIAM J. Scientific Computing)

One high-accuracy solve costs 30 times as much as a residual evaluation

about 15 to reach truncation error

1000 times faster than some popular methods

e.g. Lemieux, Price, Evans, Knoll, Salinger, Holland, Payne 2011 (J. Computational Physics)

— Actual speedup subject to Amdahl's Law

(Brown, Smith, Ahmadia 2013, SIAM J. Scientific Computing)

Algebraic multigrid for Hydrostatic

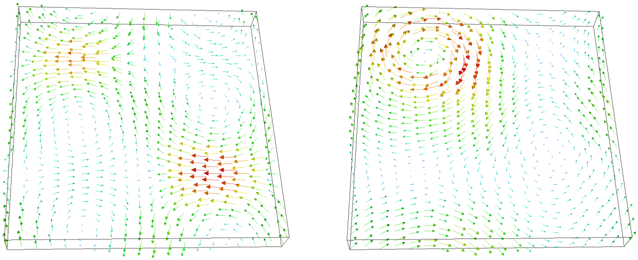
- Easy to use: assemble a matrix and throw it over the wall
- Higher setup costs, lower arithmetic intensity
- AMG uses heuristics to diagnose anisotropy; varies by discretization
- Need to represent rotational modes
  - Smoothed aggregation takes a “near null space” (translation plus rotation)

Eigen-analysis plugin for solver design

Hydrostatic ice flow (nonlinear rheology and slip conditions)

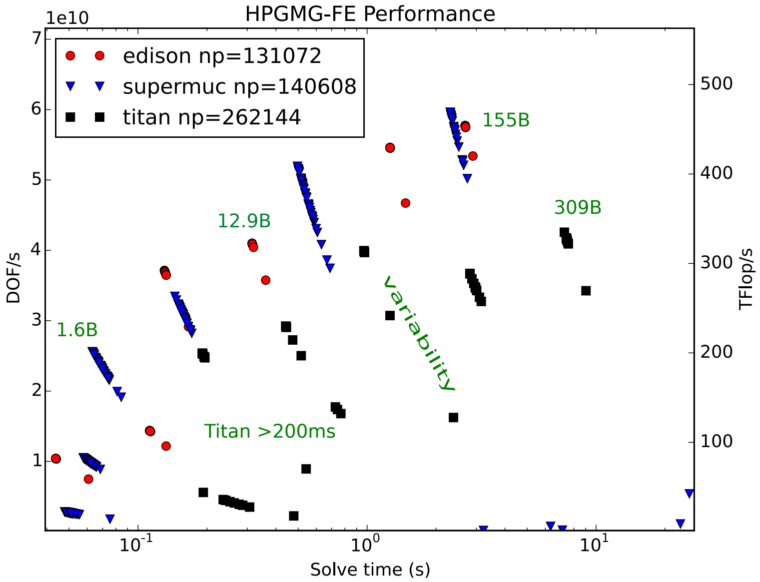
$$-\nabla\left[\eta\begin{pmatrix}4u_x+2v_y & u_y+v_x & u_z\\ u_y+v_x & 2u_x+4v_y & v_z\end{pmatrix}\right]+\rho g\nabla s=0, \qquad (1)$$

- Many solvers converge easily with no-slip/frozen bed, more difficult for slippery bed (ISMIP HOM test C)
- Geometric MG is good: $\lambda\in[0.805,1]$ (SISC 2013)



(a) $\lambda_0=0.0268$ (b) $\lambda_1=0.0409$

HPGMG-FE https://hpgmg.org



Conservative (non-Boussinesq) two-phase ice flow

Find momentum density ρu, pressure p, and total energy density E:

$$(\rho u)_t + \nabla\cdot(\rho u \otimes u - \eta D u_i + p1) - \rho g = 0$$
$$\rho_t + \nabla\cdot\rho u = 0$$
$$E_t + \nabla\cdot\big((E+p)u - k_T \nabla T - k_\omega \nabla\omega\big) - \eta D u_i : D u_i - \rho u\cdot g = 0$$

- Solve for density ρ, ice velocity u_i, temperature T, and melt fraction ω using constitutive relations.
- This and many other formulations lead to a Stokes problem

The Great Solver Schism: Monolithic or Split?

Monolithic

- Direct solvers
- Coupled Schwarz
- Coupled Neumann-Neumann (need unassembled matrices)
- Coupled multigrid

X Need to understand local spectral and compatibility properties of the coupled system

Split

- Physics-split Schwarz (based on relaxation)
- Physics-split Schur (based on factorization)
  - approximate commutators SIMPLE, PCD, LSC
  - segregated smoothers
  - Augmented Lagrangian
  - "parabolization" for stiff waves

X Need to understand global coupling strengths

- Preferred data structures depend on which method is used.
- Interplay with geometric multigrid.

Stokes

Weak form of the Newton step

Find (u, p) such that

$$\int_\Omega (Dv)^T \left[\eta 1 + \eta' Dw \otimes Dw\right] Du$$
$$- p\nabla \cdot v - q\nabla \cdot u = -v \cdot F(w) \qquad \forall (v, q)$$

Matrix

$$\begin{bmatrix} A(w) & B^T \\ B & \end{bmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = -\begin{pmatrix} F_u(w) \\ 0 \end{pmatrix}$$

Block factorization

$$\begin{bmatrix} A & B^T \\ B & \end{bmatrix} = \begin{bmatrix} 1 & \\ BA^{-1} & 1 \end{bmatrix} \begin{bmatrix} A & B^T \\ & S \end{bmatrix} = \begin{bmatrix} A & \\ B & S \end{bmatrix} \begin{bmatrix} 1 & A^{-1}B^T \\ & 1 \end{bmatrix}$$

where the Schur complement is

$$S = -BA^{-1}B^T.$$

Hardware Arithmetic Intensity

Operation	Arithmetic Intensity (flops/B)
Sparse matrix-vector product	1/6
Dense matrix-vector product	1/4
Unassembled matrix-vector product, residual	$\gtrsim 8$

Processor	STREAM Triad (GB/s)	Peak (GF/s)	Balance (F/B)
E5-2680 8-core	38	173	4.5
E5-2695v2 12-core	45	230	5.2
E5-2699v3 18-core	60	660	11
Blue Gene/Q node	29.3	205	7
Kepler K20Xm	160	1310	8.2
Xeon Phi SE10P	161	1060	6.6
KNL (DRAM)	100	3000	30
KNL (MCDRAM)	500	3000	6

Outlook

- Choose suitable technology
- Geometric multigrid is simple and has low setup cost
- Algebraic multigrid has higher setup, more finicky to discover anisotropy
- Stokes problems
  - block factorization is easiest (all run-time options in PETSc)
  - coupled MG is worth considering
- Newton linearization of sliding
- Mind the external factors