Efficient algorithms and software for quasidynamic earthquake simulators

Andrew M. Bradley

ambrad@cs.stanford.edu

{www|cs}.stanford.edu/~ambrad

CDFM Group, Dept. of Geophysics
Stanford University

Outline

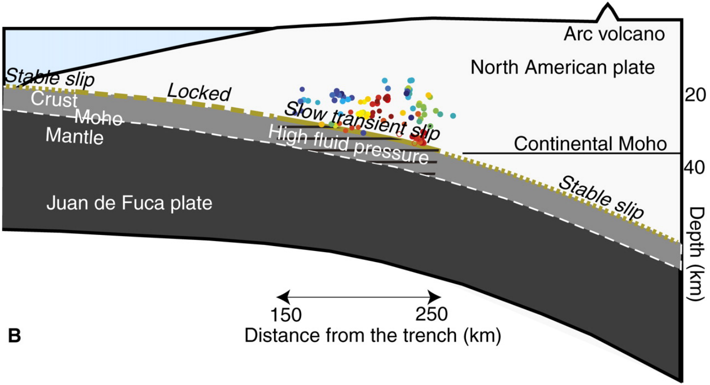
1. Quasidynamic fault simulation
2. H-matrix compression (`hmmvp`)
3. A 3D Displacement Discontinuity Method for a nonuniform mesh (`dc3dm`)

Outline

1. Quasidynamic fault simulation
2. H-matrix compression (`hmmvp`)
3. A 3D Displacement Discontinuity Method for a nonuniform mesh (`dc3dm`)

Cascadia subduction zone

Illustration of northern Cascadia and 2007 episodic tremor and slip (ETS) event.

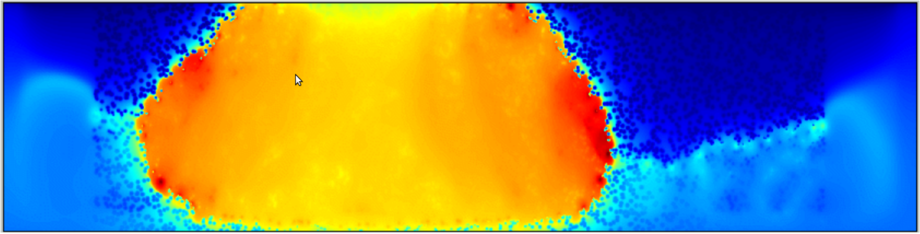


Gomberg J , and the Cascadia 2007 and Beyond Working Group Geological Society of America Bulletin 2010;122:963-978

©2010 by Geological Society of America

THE GEOLOGICAL SOCIETY OF AMERICA

Time-dependent quasidynamic simulation of slow slip (movie)



Snapshot of $\log_{10}$ slip speed from a quasidynamic simulation of a slow slip event.

Quasidynamic simulation of slip

Traction due to slip	$\tau = G\,s : \tau_{\text{shear}} = G_{\text{shear}} s,\ \tau_{\text{normal}} = G_{\text{normal}} s$
Rate-state friction	$f(v, \theta)$
State evolution	$\dot{\theta} = E(v, \theta)$
Pressure evolution	$\dot{p} = P(\cdots)$
Effective normal stress	$\sigma_{\text{eff}} \equiv \sigma_{\text{normal}} + G_{\text{normal}} s - p$
Fault strength	$\sigma_{\text{eff}}\, f(v, \theta)$
Radiation damping	$\eta v \quad \text{with} \quad \eta \equiv \frac{\mu}{2V_s}$
Momentum balance	$\tau_{\text{shear}} = \sigma_{\text{eff}}\, f(v, \theta) - \eta v$

$\tau = Gs$:

- Assume optimal linear algebra. Assume a BEM can be used.
  A BEM is practically and asymptotically faster than a volume-discretization method.
- A fault is a crack with a high aspect ratio.
- Displacement Discontinuity Method (DDM), a type of Boundary Element Method (BEM):
  - Model crack has infinite aspect ratio.
  - Mesh the fault.
  - An element represents both sides of a crack.
  - Many Green's functions.
- G is numerically dense.

Outline

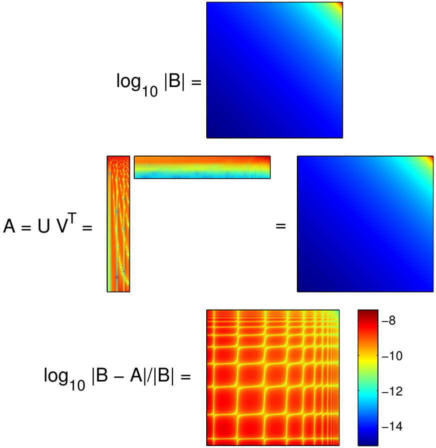
1. Quasidynamic fault simulation
2. H-matrix compression (`hmmvp`)
3. A 3D Displacement Discontinuity Method for a nonuniform mesh (`dc3dm`)

Off-diagonal block approximate low-rank structure

- Low-rank approximation:

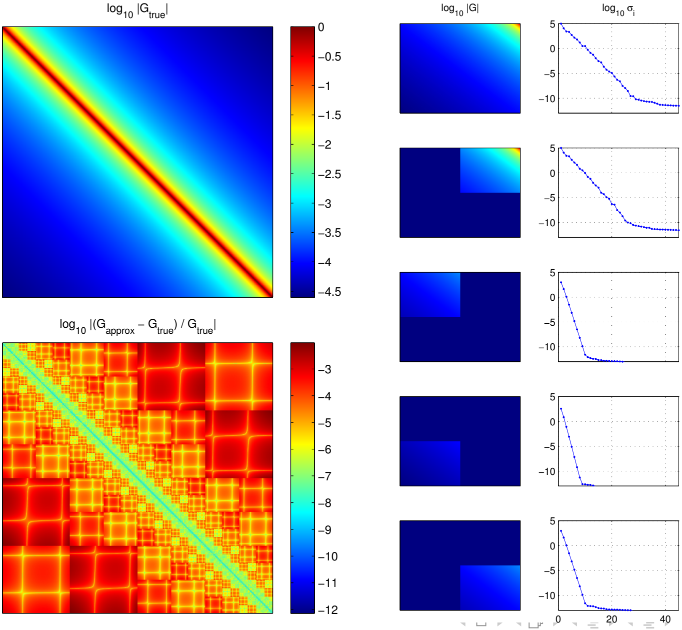
  $B \in \mathbb{R}^{m \times n}$, $U \in \mathbb{R}^{m \times r}$, $V \in \mathbb{R}^{n \times r}$
  $B \approx UV^T$ (outer product)
  $O(mn)$ *vs.* $O(r(m+n))$ work, memory
  Ideally, $r \ll \min\{m, n\}$

- Algorithm building blocks:
  - Permutation matrices to reveal block approximate low-rank structure.
    - ★ Hierarchical spatial decomposition.
  - Approximation error control.
  - Low-rank approximation.
  - Numerical linear algebra operations.

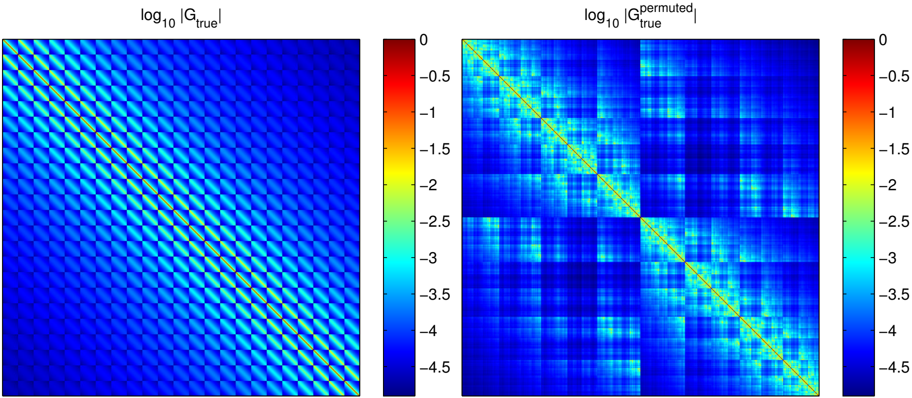


- Many methods: Barnes-Hut [1986]; Fast Multipole Method [Greengard & Rokhlin 1987]; kernel-independent or blackbox FMM [Ying, Biros, & Zorin 2004, Fong & Darve 2009]; Hierarchical Matrix (H-Matrix) [Hackbusch 1999, Börm, Grasedyck, & Hackbusch 2003, Bebendorf 2008], H^2-Matrix [Hackbusch, Khoromskij, & Sauter 2000], Hierarchically Semiseparable (HSS) [Chandrasekaran, Dewilde, Gu, Pals, Sun, van der Veen, & White 2006]; p-HSS, HOLDR [Ambikasaran & Darve 2013].

Visualization: A line



Visualization: A square



Approximation error control

Let $\bar{B}$ be an approximation to $B \in \mathbb{R}^{M \times N}$.
B is partitioned into blocks $\{B_i\}_i$.
Assume B is already permuted to reveal block approximate low-rank structure.
Let $\delta B \equiv B - \bar{B}$.
Basic error tolerance: $\|\delta B\|_\mathrm{F} \le \varepsilon \|B\|_\mathrm{F}$.
Then for $y = Bx$, $\|\delta y\|_2 \le \varepsilon \|B\|_\mathrm{F} \|x\|_2$.

Method B (standard method):

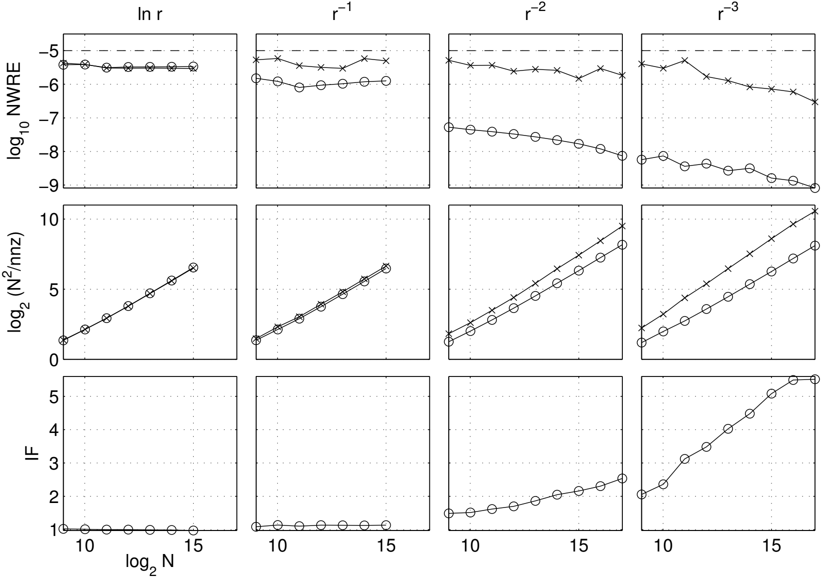
- $\|\delta B_i\|_\mathrm{F} \le \varepsilon \|B_i\|_\mathrm{F}$.
- Proof: $\|\delta B\|_\mathrm{F}^2 = \sum_i \|\delta B_i\|_\mathrm{F}^2 \le \varepsilon^2 \sum_i \|B_i\|_\mathrm{F}^2 = \varepsilon^2 \|B\|_\mathrm{F}^2$.

Method M:

- $\|\delta B_i\|_\mathrm{F} \le \varepsilon \frac{\sqrt{m_i n_i}}{\sqrt{MN}} \|B\|_\mathrm{F}$.
- Proof: $MN = \sum_i m_i n_i$. $\|\delta B\|_\mathrm{F}^2 = \sum_i \|\delta B_i\|_\mathrm{F}^2 \le \varepsilon^2 (MN)^{-1} \|B\|_\mathrm{F}^2 \sum_i m_i n_i = \varepsilon^2 \|B\|_\mathrm{F}^2$.

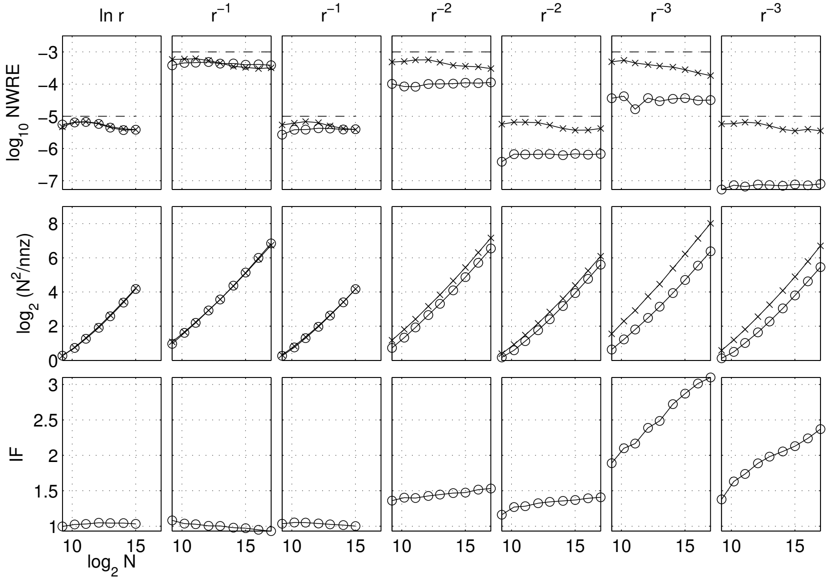
$\varepsilon^2 \|B_i\|_\mathrm{F}^2$	$\varepsilon^2 \frac{m_i n_i}{MN} \|B\|_\mathrm{F}^2$
$(m_i n_i)^{-1} \|B_i\|_\mathrm{F}^2$	$(MN)^{-1} \|B\|_\mathrm{F}^2$
With increasing order of singularity, increasingly large near diagonal, increasingly small away from diagonal.	Constant.

Numerical experiment: Approximation error control



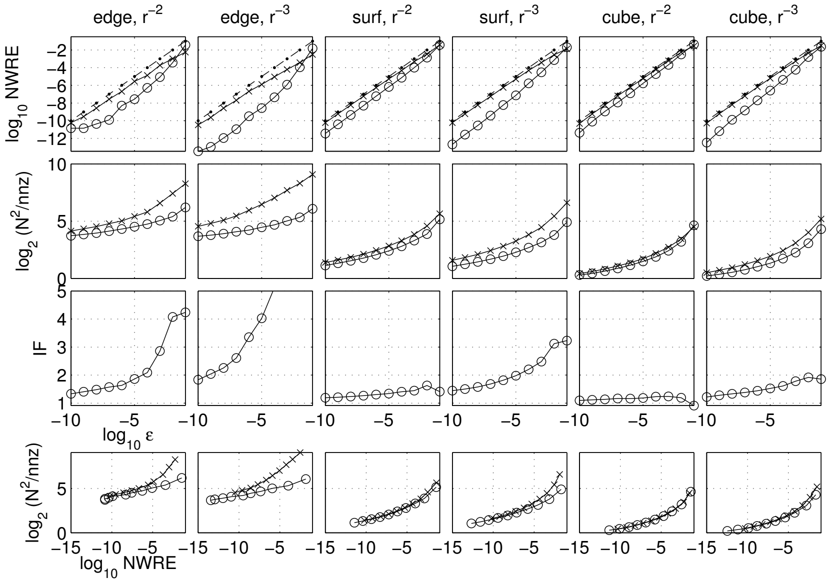
Points distributed on edges of a cube.

Numerical experiment: Approximation error control



Points distributed on faces of a cube.

Numerical experiment: Approximation error control



$\sim 2^{13}$ particles.

Block compression

Singular value decomposition (SVD) is optimal but requires too much work:

- SVD itself is expensive.
- Ideally only a small fraction of the elements in a block are computed.

But: Reduced SVD [textbook method]:

- Given $B = UV^T$.
- Compute $Q_U R_U = \text{qr}(U, 0)$. Similarly for V.
- Compute $W \Sigma Z^T = \text{svd}(R_U R_V^T)$.
- Then $(Q_U W) \Sigma (Q_V Z)^T$ is the SVD of B.

And: Adaptive Cross Approximation (ACA) [Bebendorf & Rjasanow 2001].

- $O(r^2(m+n))$.
- Decently robust termination.
- Decently efficient compression.

Together:

- Set $\eta \equiv \varepsilon \frac{\sqrt{m_i n_i}}{\sqrt{MN}} \|B\|_\text{F}$.
- ACA: $\bar{B}_i^{(0)} \equiv U^{(0)} (V^{(0)})^T$ so that

$$\|\bar{B}_i^{(0)} - B_i\|_\text{F} \le \frac{\eta}{2} \; (\times 0.1 \text{ empirical factor}).$$

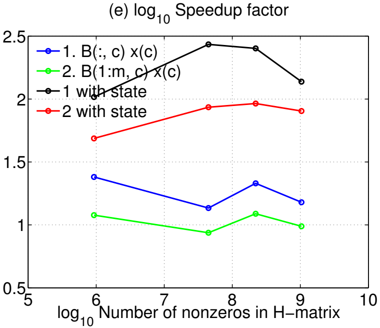
- Reduced SVD: $\bar{B}_i^{(0)} = Q_U \Sigma Q_V^T$.
- Choose smallest k so that $\sum_{i=k+1}^r \sigma_i^2 \le \frac{\eta^2}{4}$.
- Set

$$\begin{aligned} U &\equiv Q_U(:, 1{:}k)\; \Sigma(1{:}k, 1{:}k) \\ V &\equiv Q_V(:, 1{:}k) \\ \bar{B}_i &\equiv UV^T. \end{aligned}$$

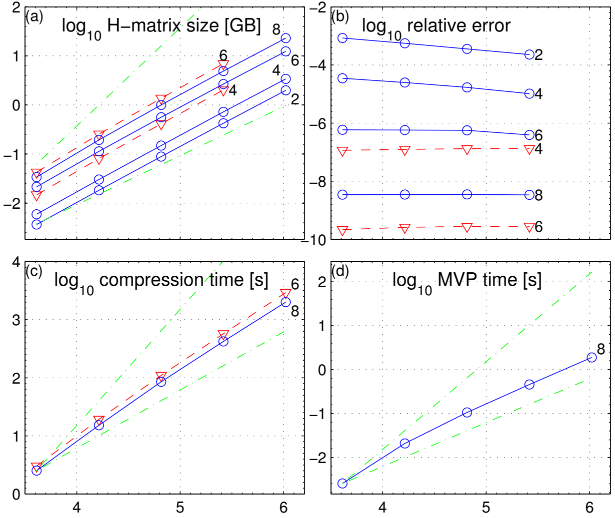
- Then $\|\bar{B}_i - B_i\|_\text{F} \le \eta$.

hmmvp

- `C++` library. Matlab, Fortran (limited) interfaces.
- MPI, OpenMP.
- Also:
  $y = B(:, c)x(c)$
  $y(r) = B(r, :)x$
  $y(r) = B(r, c)x(c)$.
- `hmmvpex`: In-depth example code: `dc3d` [Okada 1992], `tgf` [Gimbutas, Greengard, Barall, Tullis 2012].



Numerical experiment: `hmmvp` with [Okada 92] Green's function



$N = 1024^2$, $\varepsilon = 10^{-6}$, method M: 12.4 GB (330$\times$ compression).

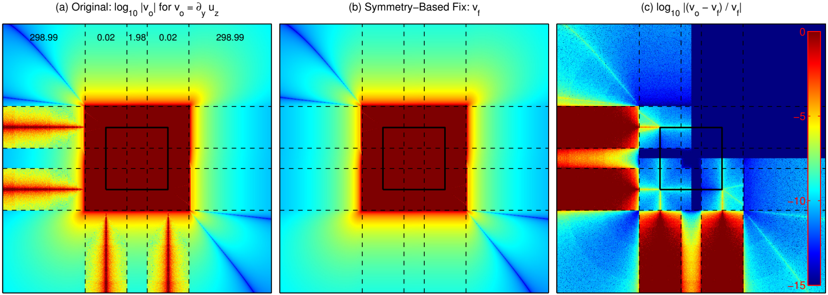
$N = 512^2$, $\varepsilon = 10^{-6}$: method M 2.5$\times$ more efficient than method B.

$N = 1024^2$, method M: compression took 34 minutes (compared with tens of hours w/o comp.).

$N = 1024^2$, $\varepsilon = 10^{-8}$, method (M): MVP takes 1.9 seconds.

An aside: Accuracy in DC3D [Okada 1992]

Analyzing compression for a Green's function (GF) can reveal numerical errors in the GF.



$\log_{10} |\partial_y u_z|$.

Numerical cancellation in expressions of the form $R + y$ for $y = \eta < 0$ or $y = \xi < 0$, where $R = (\xi^2 + \eta^2 + q^2)^{1/2}$ and ξ, η, q are element coordinate directions.

Use symmetry.

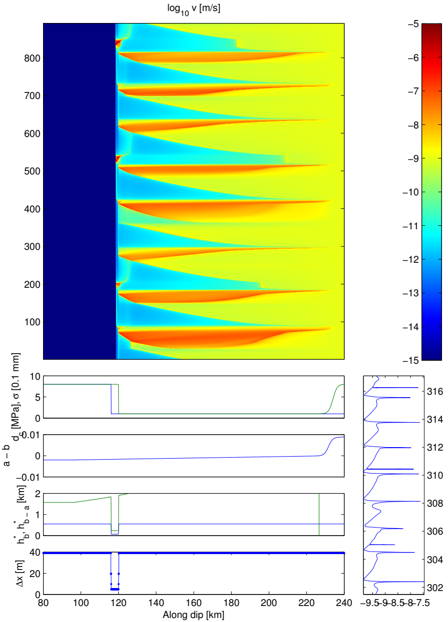
Two simulation regimes

- A few faults, high-resolution discretization.
  $\Rightarrow$ Directly compress G_{ij}^{mn}, the ith traction component due to the jth slip component for the interaction between fault m and n.
- Many faults, low-resolution discretization.
  $\Rightarrow$ Compress low-level matrices that are combined to produce G_{ij}.

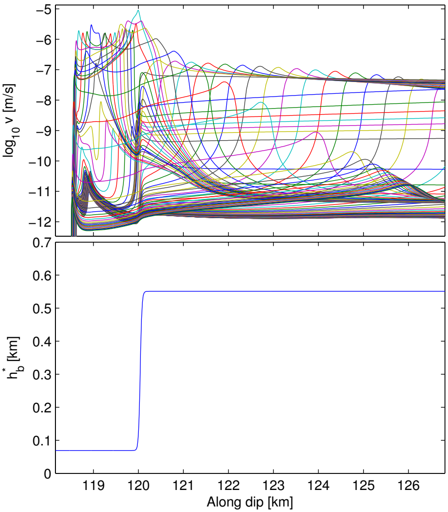
Outline

1. Quasidynamic fault simulation
2. H-matrix compression (`hmmvp`)
3. A 3D Displacement Discontinuity Method for a nonuniform mesh (`dc3dm`)

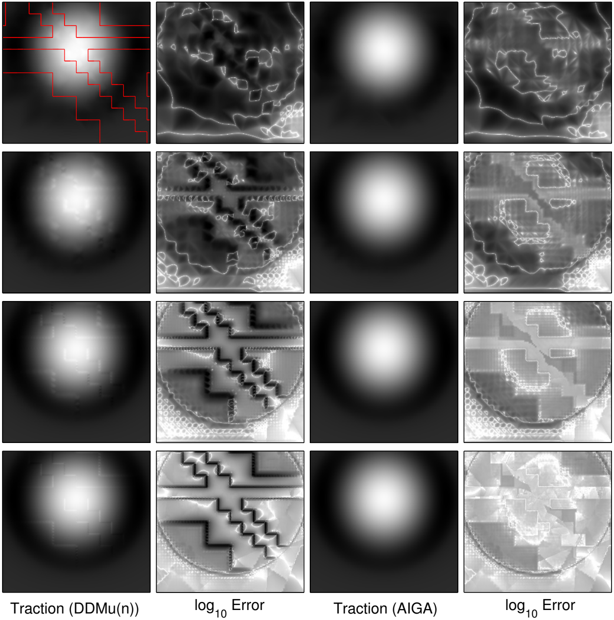
Motivation (plane strain)



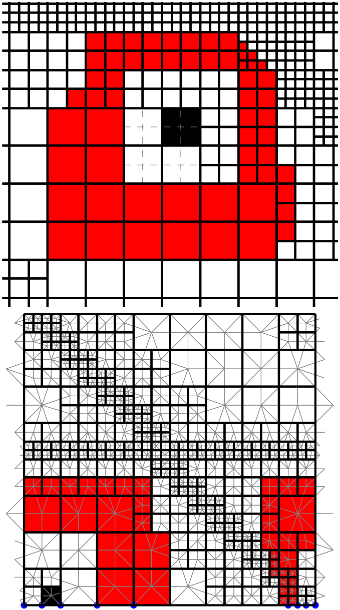
In this example, rupture tip length $\propto \frac{1}{\sigma_{\text{eff}}}$



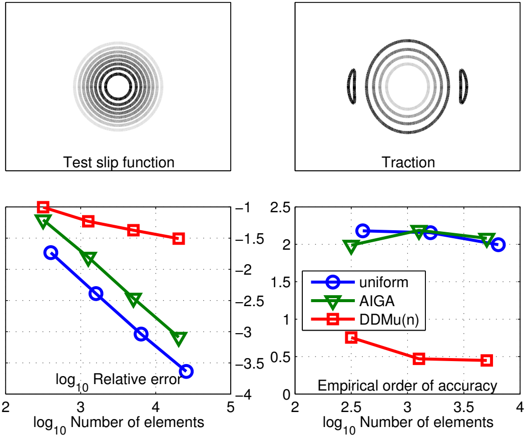
Standard DDM has artifacts at size boundaries that compromise the OOA



Nonuniform tiling squares; induced triangulation; Clough-Tocher interpolation



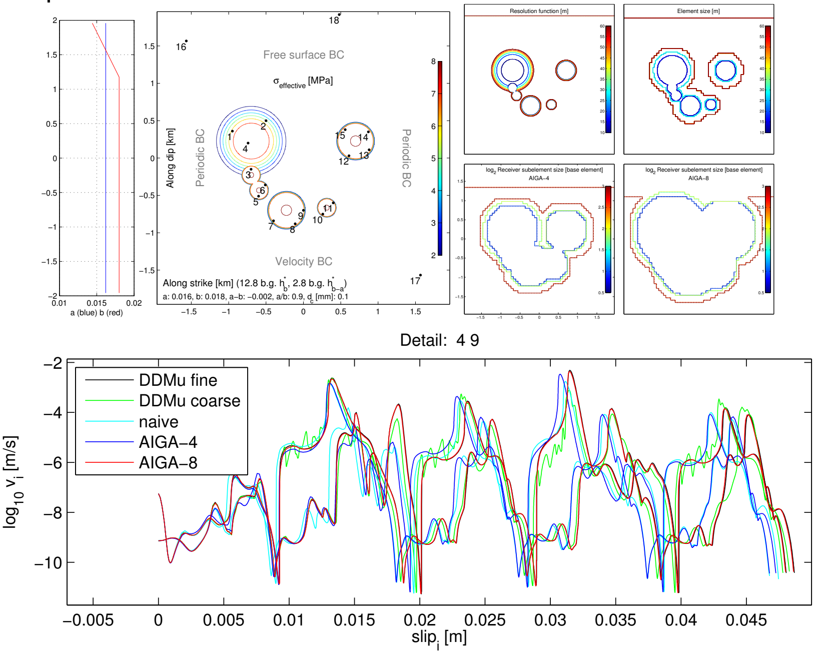
$$G_n \equiv A_{u \to n} G_u I_{n \to u}$$



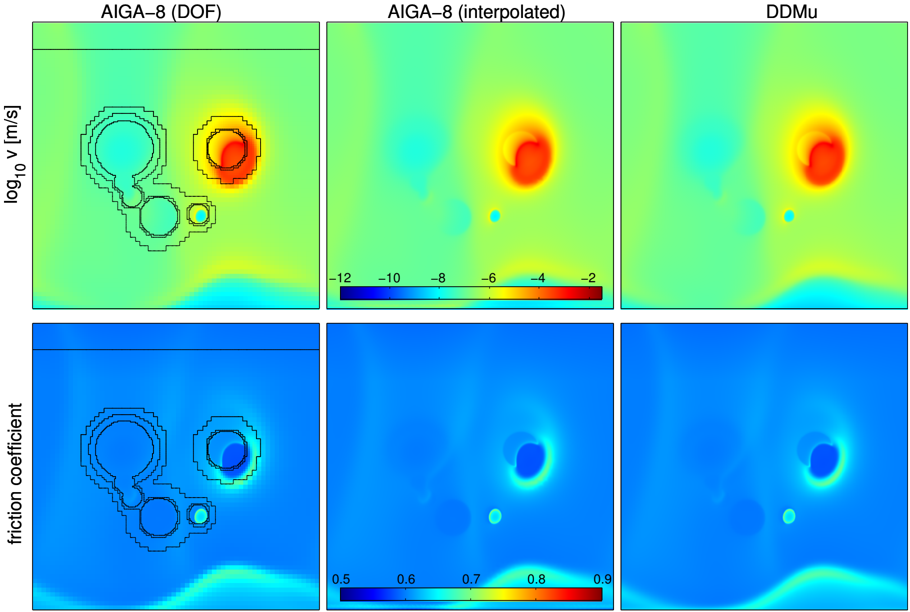
dc3dm

- C++ program on top of hmmvp. Matlab, Fortran (limited) interfaces.
- OpenMP compression. MPI, OpenMP MVP.

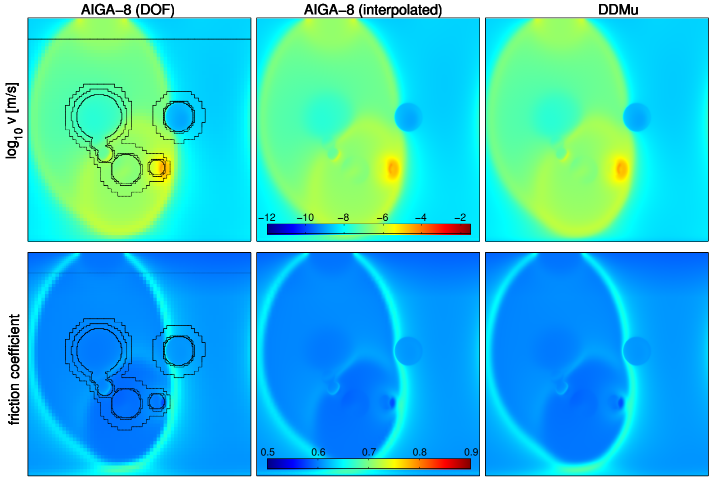
Time-dependent results



Time-dependent results (movie)

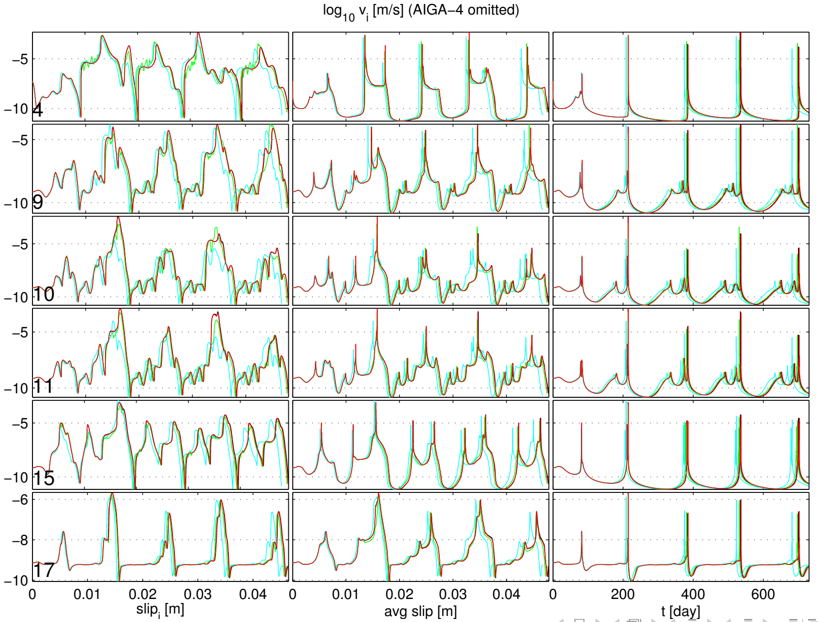


Time-dependent results (movie)

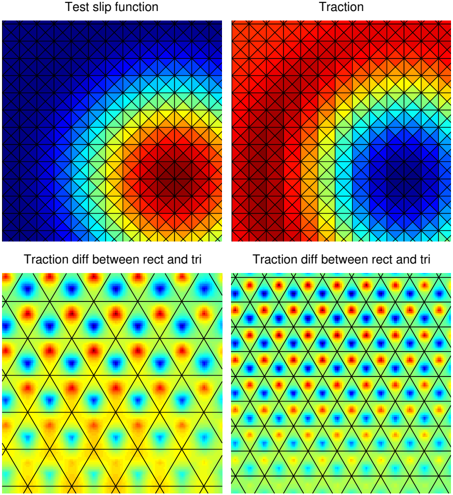


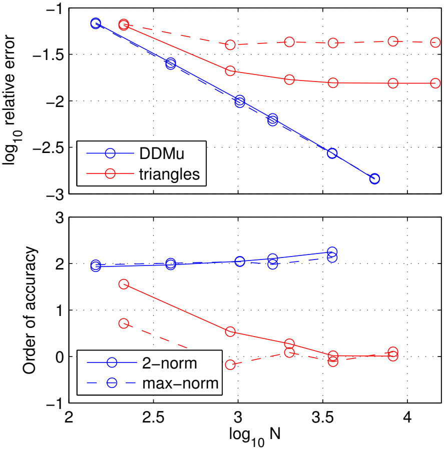
Uniform mesh: 91× matrix compression.
Nonuniform mesh: further 13.1× problem size reduction for over 1000× reduction.
Nonuniform mesh time-dependent simulation 14.1× faster.

Time-dependent results

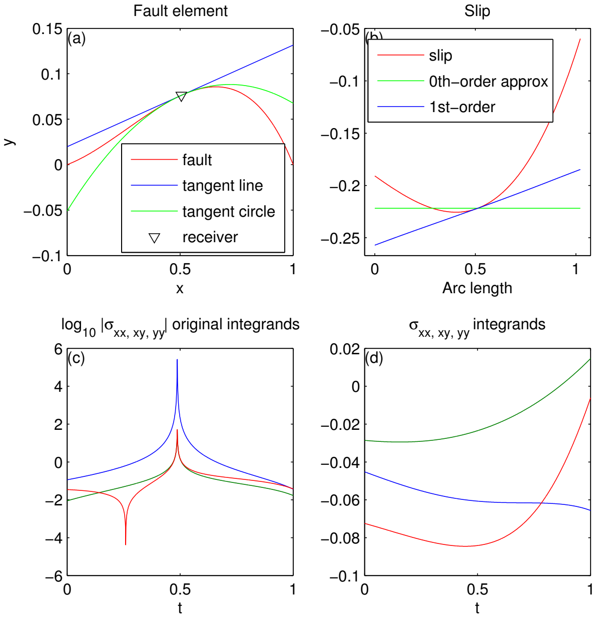


Triangles (a different kind of error)

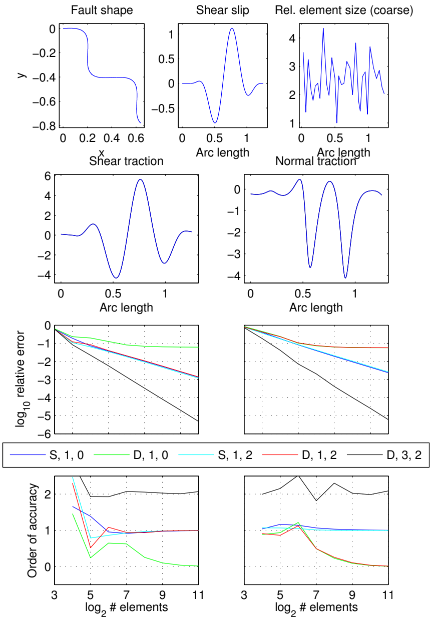




The future 1: Self-interaction integrals



The future 2: Sufficiently smooth basis functions



Summary and future work

- 3D quasidynamic simulations are computationally intensive.
- DDM for a nonuniform mesh (implemented in `dc3dm`) + optimal linear algebra (implemented in `hmmvp`) ⇒ practical speedup of 100s to 1000s.

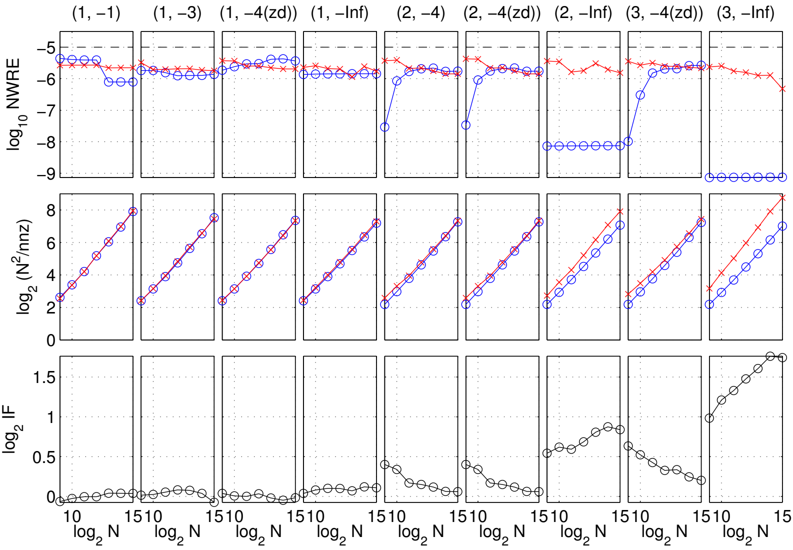
- Higher-order elements for nonuniformly discretized, nonplanar faults.
- `dc3dm`-like software using these elements.
- Many faults–low resolution case.

Thanks!

- Crustal Deformation and Fault Mechanics Group, Dept. of Geophysics, Stanford.
- NSF grant EAR-0838267, USGS grant G12AP20030, and SCEC grant 13096.

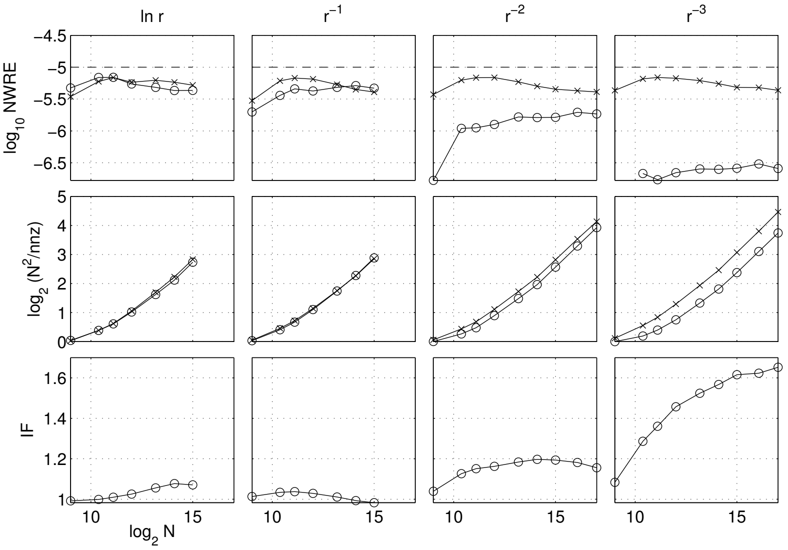
Backup slides.

Error control: Numerical experiment 1



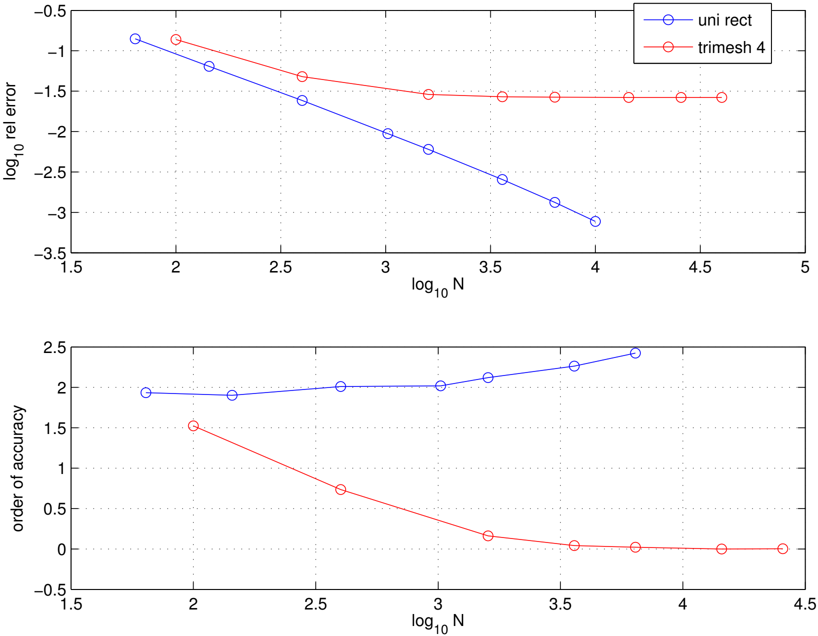
Line.

Error control: Numerical experiment 1



Filled cube.

Triangles



uni rect
trimesh 4
log10 rel error
log10 N
order of accuracy
log10 N