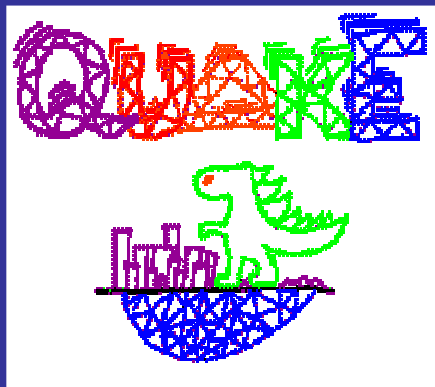
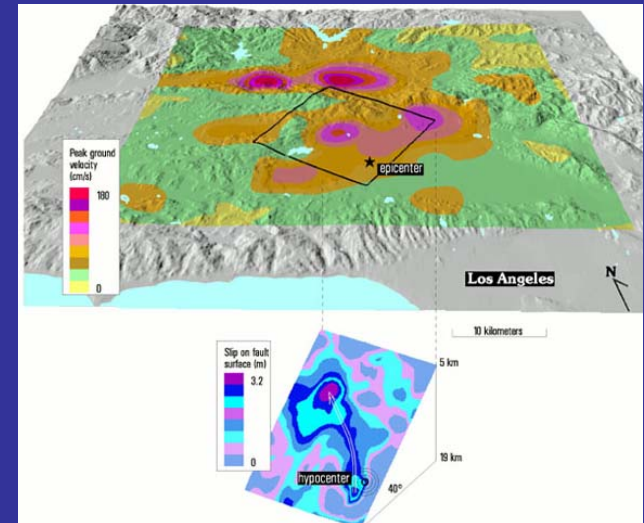


Large Scale Earthquake Inversion

Omar Ghattas
Carnegie Mellon University

Joint work with:

Volkan Akcelik, Jacobo Bielak, George Biros (UPenn),
Ioannis Epanomeritakis, Loukas Kallivokas (Texas),
Eui Joong Kim, David O'Hallaron, Tiankai Tu



Carnegie Mellon



“PDE-constrained” optimization

- Optimization of systems governed by PDEs and variational inequalities
- Brings together approaches from finite-dimensional large-scale optimization and infinite-dimensional analysis
- Work in this area dates back several decades
- Recent years have seen acceleration of interest and activity
 - VPI workshop 1994
 - ICASE workshop 1995
 - Santa Fe workshops 2001, 2004
 - Oberwolfach workshops 2003, 2004, 2005, 2006
 - IMA workshop 2003
 - Various workshops at Graz, Heidelberg, Trier, etc.
- See article by E. Sachs in 2003 SIAG/OPT News & Views

The “simulation problem” (forward, state, direct)

PDE model:

$$c(u, d) = 0$$

where

$u :=$ state variables

$d :=$ decision (design, control, inversion) variables

State problem: given d (material property, domain or boundary sources, initial condition, geometry, etc.) find u (velocity, temperature, flux, displacement, stress, concentration, magnetic field, electric field, etc.)

Often well-posed

The “optimization problem” (inverse, design, control)

Optimization (design, control, inverse) problem: Given desired goal and inequality constraints involving u and/or d , find optimal d :

$$\begin{aligned} &\text{minimize } \mathcal{J}(u, d) \\ &\text{subject to } c(u, d) = 0 \\ &\quad \quad \quad h(u, d) \geq 0 \end{aligned}$$

where

$u :=$ states

$d :=$ decision (control, design, inversion) variables

$\mathcal{J} :=$ objective functional

$c :=$ state equations

$h :=$ inequality constraints

Often ill-posed

PDE-constrained optimization vs. general NLP

- **Problem size:**
 - o N_u =up to $O(10^9)$ state variables (per time step)
 - o $N_d=O(1)$ — $O(10^9)$ decision variables
 - o Generally cannot afford more than small number of PDE "solves"
- **Structure of PDE constraints must be exploited**
 - o Iterative solvers necessary in 3D
 - o Parallelism often necessary
 - o Preconditioning essential
- **Solver requirements vs. optimizer requirements**
 - o Many PDE codes are often Jacobian-free
 - o PDE Jacobian often approximated in many codes
- **Infinite dimensional setting**
 - o Existence and irregularity of Lagrange multipliers
 - o Discretize-then-optimize vs. optimize-then-discretize
 - o Convergence theory
- **No such thing as a general-purpose PDE solver**
→ no general-purpose PDE optimizer!

Karush-Kuhn-Tucker first order optimality conditions

Return to PDE-constrained optimization formulation:

$$\begin{aligned} &\text{minimize } \mathcal{J}(u, d) \\ &\text{subject to } c(u, d) = 0 \end{aligned}$$

Define Lagrangian function(al):

$$\mathcal{L}(u, d, p) := \mathcal{J}(u, d) + \langle p, c(u, d) \rangle$$

where p is the Lagrange multiplier for c , or the adjoint or costate variable. Optimality conditions are stated by requiring stationarity of the Lagrangian w.r.t. p, u, d , resulting in:

$c(u, d) = 0$	state equation
$c_u^*(u, d)p = -\mathcal{J}_u(u, d)$	adjoint equation
$c_d^*(u, d)p = -\mathcal{J}_d(u, d)$	decision equation

Discretization issues

- Often $\text{DTO} \neq \text{OTD}$ (non-Galerkin discretization of optimality system, nonsymmetric treatment of time discretization, stabilization methods, subgrid scale models, shape optimization, nonsmoothness, etc.)
- OTD: aids in understanding nature of optimality equations, opens way to different (adaptive) discretizations for state and adjoint equations, avoids differentiating artifacts of the state discretization, provides guidance on stabilizations for adjoint, etc.
- Drawback of OTD is that resulting discretized gradient is not guaranteed to be the derivative of the discretized objective (but often within discretization error)
- Which is preferable is problem-dependent
- Best advice is to use infinite dimensional optimality system as a guide, but strive to discretize it in a way that is compatible with DTO

Solver issues

- Even in simplified setting of no inequalities, optimization problem may be very difficult
- State PDE constraints can be
 - nonlinear PDEs
 - time-dependent problems
 - vector unknowns, coupled systems
- **Full space methods** solve for states, controls, and adjoints simultaneously
- **Reduced space methods** solve in space of controls; for nonlinear problems, we have choice:
 - eliminate states and adjoints first, then linearize (unconstrained method)
 - linearize then eliminate states and adjoints (e.g. reduced SQP)

Illustrative Examples

- Artificial heart design
- Image-driven cardiac diagnosis
- Inverse contaminant transport
- Accelerator design
- Earthquake inversion

Multiscale blood flow modeling for artificial heart device design

Carnegie Mellon

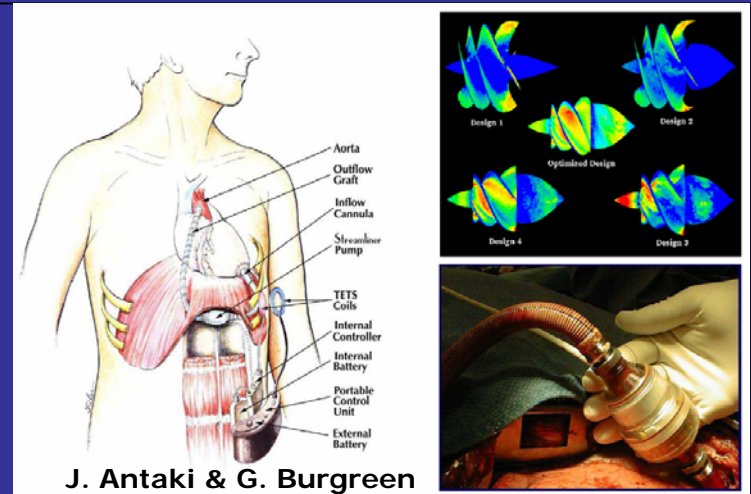


James Antaki, Guy Blesloch, Omar Ghattas, Judy Hill, Marina Kameneva (Pitt), Robert Kormos (Pitt), Ivan Malcevic (GE), Gary Miller, K. Rajagopal (Texas A&M), George Turkiyyah (Washington), Noel Walkington



At macroscopic (device) scales:

- Development of artificial heart assist device at Univ Pitt Med Center (Antaki)
- Numerous advantages (size, power, reliability, non-invasiveness)
- Design challenge: overcome tendency to damage red blood cells
- Need macroscopic blood flow theory that accounts for blood (cell) microstructure



J. Antaki & G. Burgreen

At microscopic (cell) scales:

- Macroscopic model fails in small-length-scale regions (blade tip, rotor bearing)
- Need modeling at cell scales to account for blood damage
- Our mesoscopic simulations resolve interaction of RBCs elastic membrane with plasma fluid dynamics
- Prospects for 3D simulation of blade-tip region: 1 week at sustained 1 petaflops/s

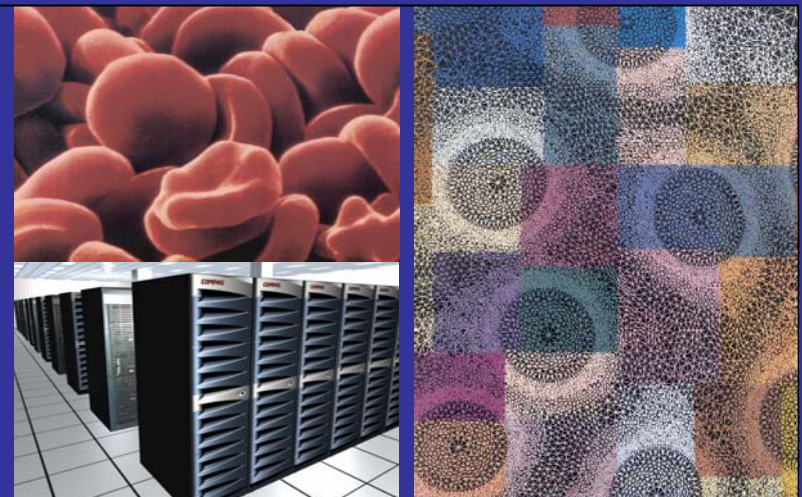


Image-based patient-specific inversion-based cardiac modeling



Volkan Akcelik (CMU) , George Biros (Penn), Alfio Borzi (Graz), Alex Cunha (CMU), Christos Davatzikos (Penn), Omar Ghattas (CMU), William Gropp (Argonne), Michael Hintermueller (Graz), Eldad Haber (Emory), David Keyes (Columbia), Jan Modersitzki (Lubek), Jennifer Schopf (Argonne)



medical
imaging



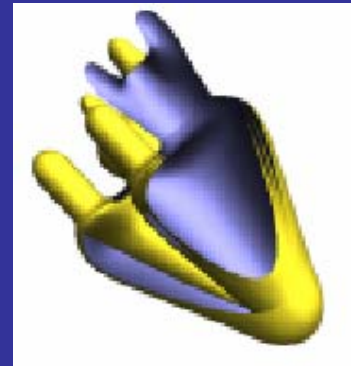
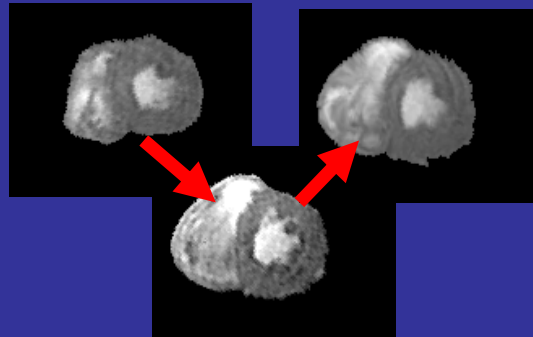
4D image
registration



5D model
inversion



diagnosis &
planning



imaging
lab server



institutional
cluster



regional
supercomputing
center



physician
desktop
center



Real time optimization for dynamic inversion & control

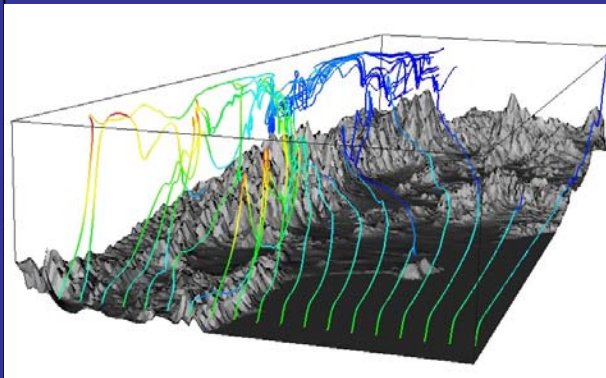
Carnegie Mellon



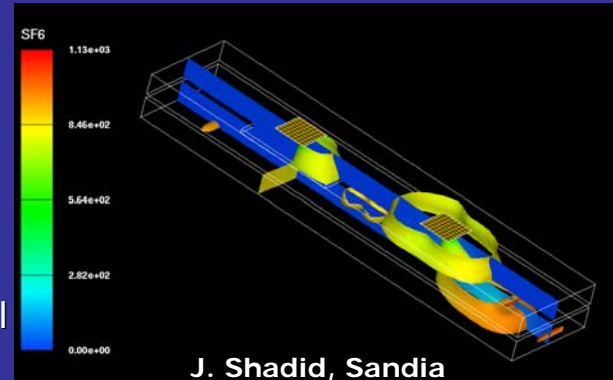
Volkan Akcelik (CMU), Roscoe Bartlett (Sandia), Lorenz Biegler (CMU), George Biros (UPenn), Andrei Dragenscu (Sandia), Frank Fendell (TRW), Omar Ghattas (CMU), Matthias Heinkenshloss (Rice), Judy Hill (CMU), David Keyes (Columbia), Carl Laird (CMU), John Shadid (Sandia), Bart van Bloemen Waanders (Sandia), Andreas Wachter (IBM), David Young (Boeing)



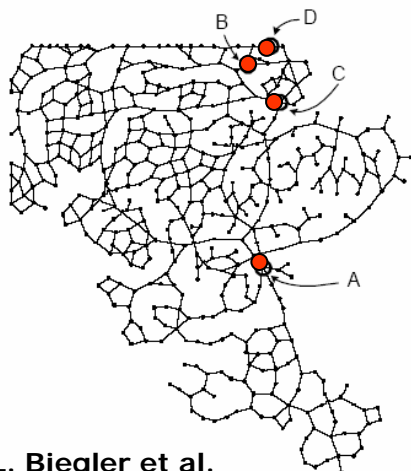
Inversion and control for airborne contaminant transport



- sensor data provides concentrations of hazardous agents
- inverse problem solved to reconstruct initial conditions
- control problem solved to find optimal remediation strategy



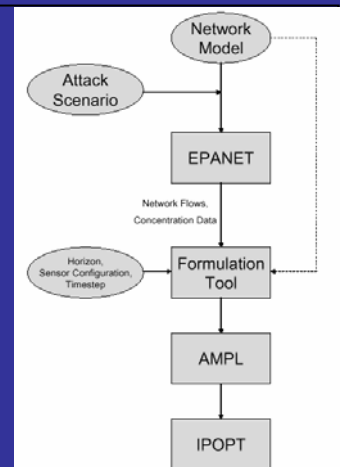
J. Shadid, Sandia



L. Biegler et al.

Water network contaminant inversion/control

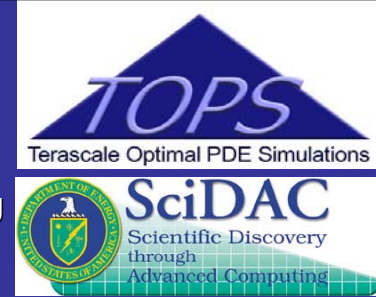
- Nonlinear optimization problem with >300K variables and >100k controls
- Solution time < 2 CPU minutes
→ real time source detection
- Algorithm successful on thousands of numerical tests on several municipal water networks
- Formulation tool links to existing modeling software (EPANET) and powerful NLP solver (IPOPT)



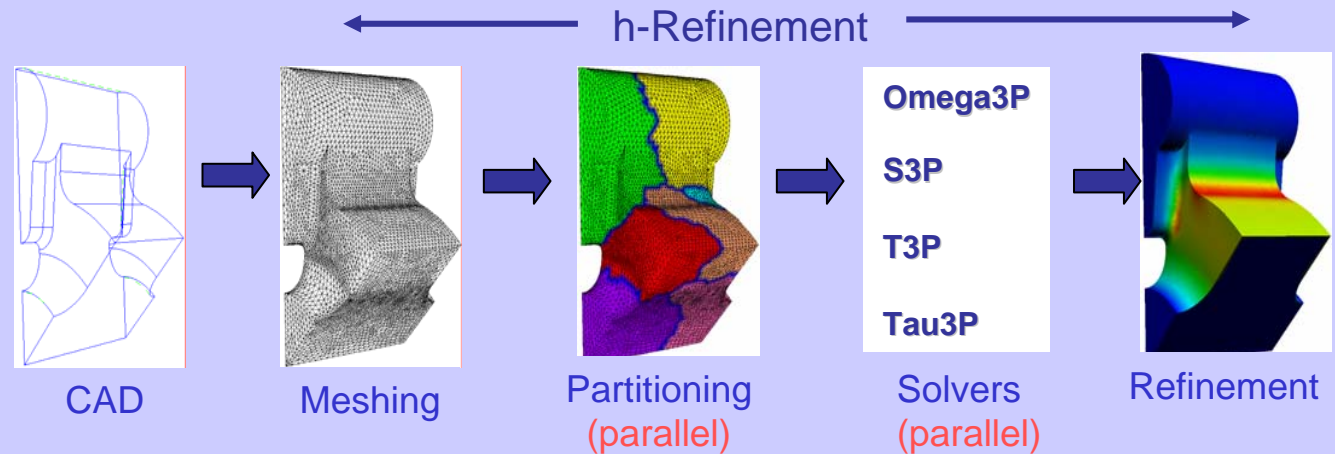
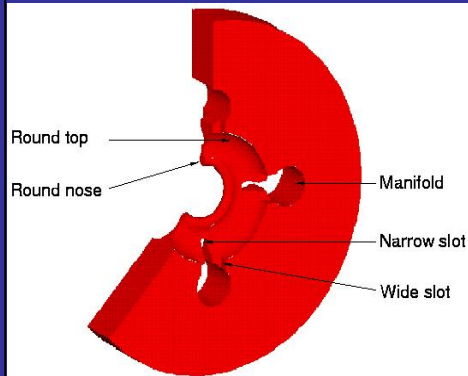
Shape optimization of accelerator structures



Volkan Akcelik (CMU), Lori Freitag (LLNL), Omar Ghattas (CMU), David Keyes (Columbia), Patrick Knupp (SNL), Kwok Ko (SLAC), Lie-Quan (Rich) Lee (SLAC), Esmond Ng (LBNL), Mark Shepherd (RPI), Tim Tautges (SNL)



Basic Analysis Loop for given Geometry



- Computer modeling has replaced trial and error prototyping
- Next generation accelerators have complex cavities that require shape optimization for improved performance and reduced cost

- Shape optimization problem governed by electromagnetic eigenvalue problem
- Cost functions involve target frequency, surface integrals of magnetic field, line integrals of electric field

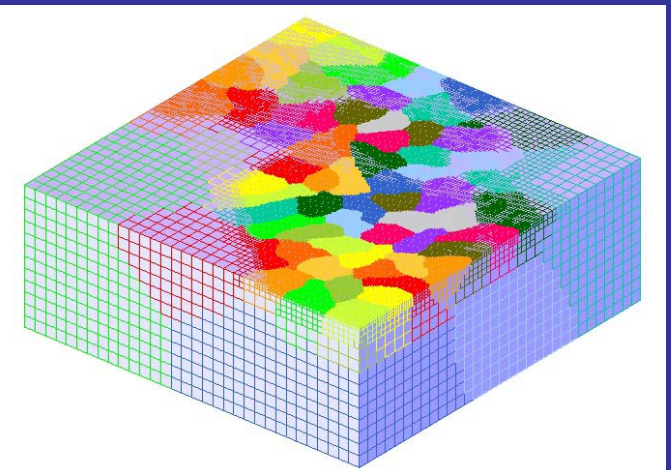
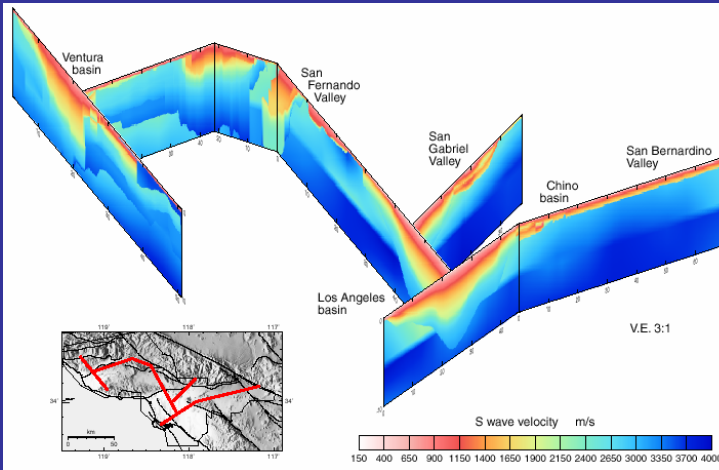
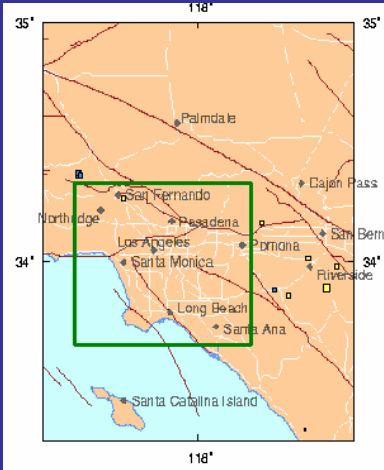


Earthquake modeling for seismic hazard assessment

Carnegie Mellon

M A C Mechanics,
Algorithms,
& Computing

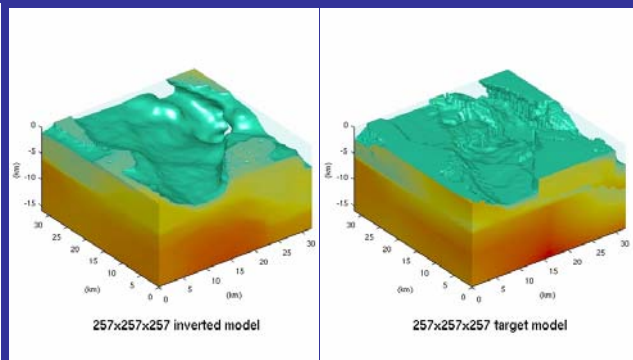
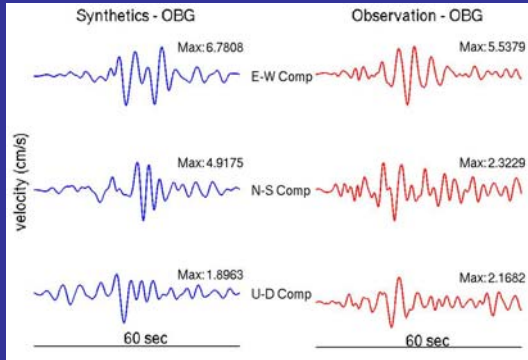
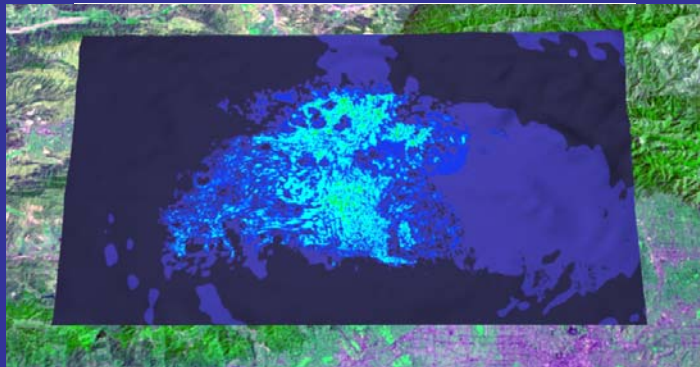
Aysegul Askan, Volkan Akcelik, Jacobo Bielak, George Biros (UPenn), Steven Day (SDSU), Omar Ghattas, Loukas Kallivokas (Texas), Harold Magistrale (SDSU), David O'Hallaron, Leonardo Ramirez, Tiankai Tu



Region of interest for
1994 Northridge
earthquake simulation

SCEC geological model provides 3D soil
properties in Greater LA Basin

Adaptive grid resolves up to 1Hz freq.
w/100 million grid pts; uniform grid
would require 2000x more points

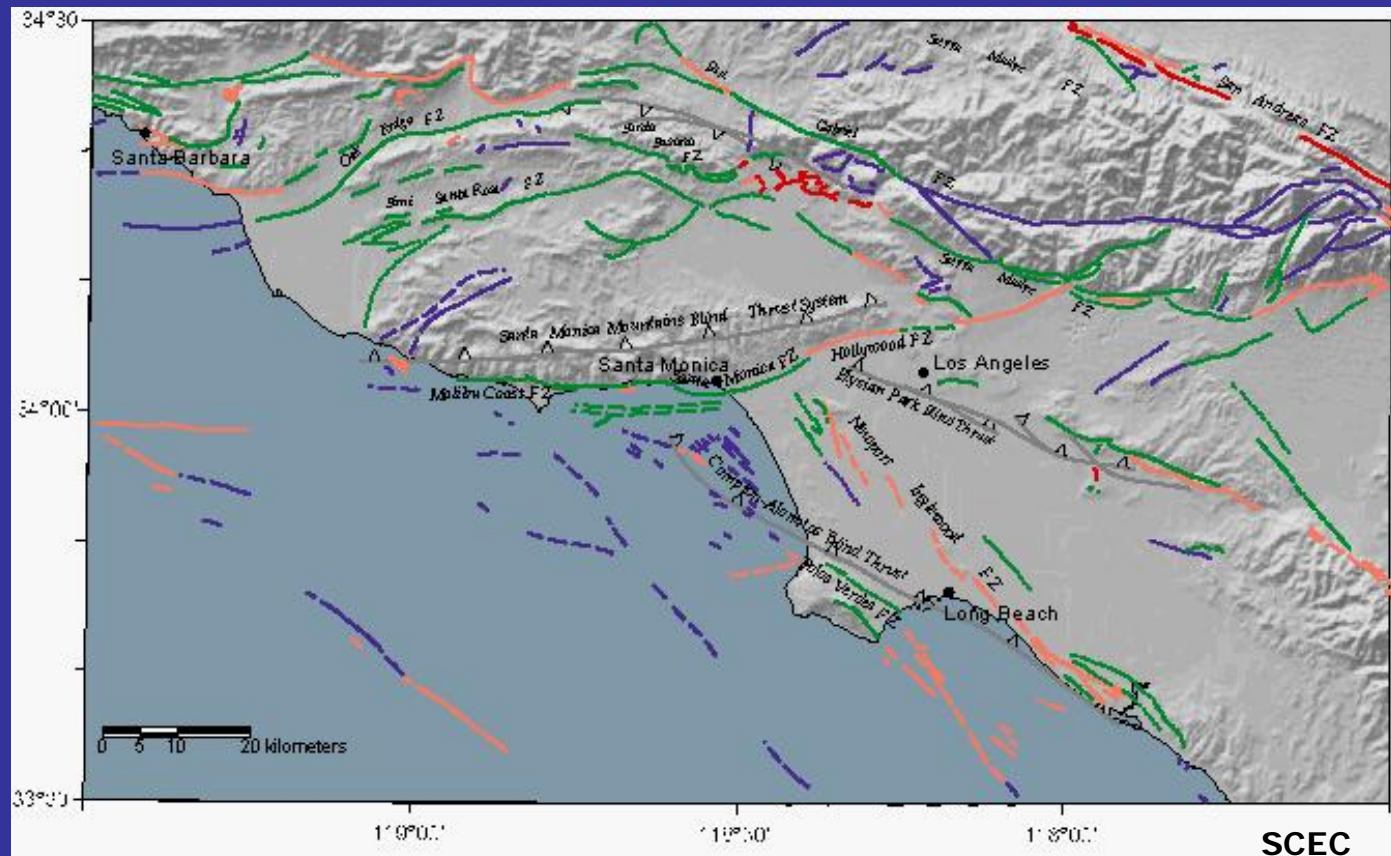


Snapshot of simulated ground motion
(simulation requires 3hr on 6Tflops PSC
machine, running at >80% parallel eff)

Comparison of observation with
simulation (improved prediction
requires petaflops capability)

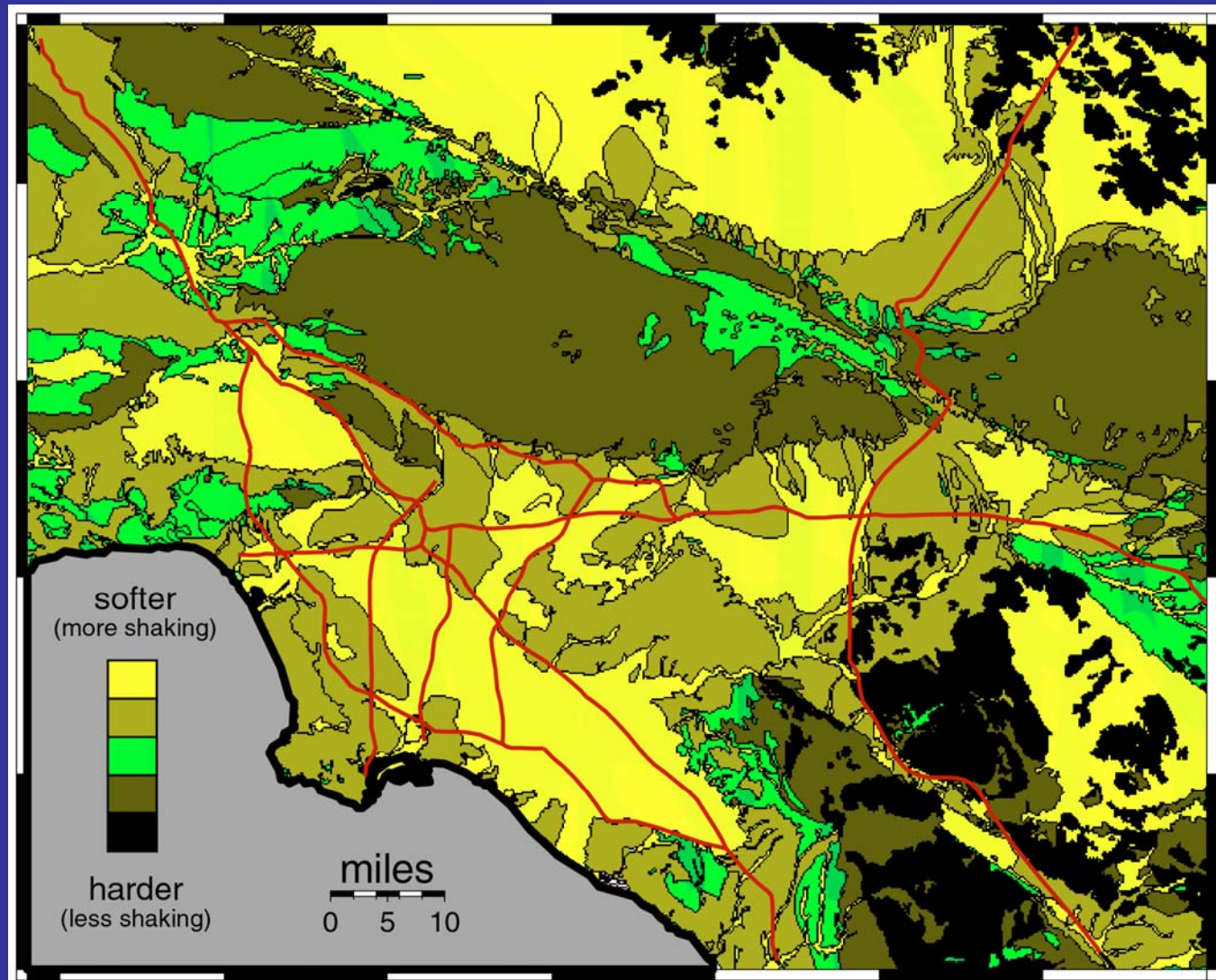
Inversion of surface observations
for 17 million elastic parameters
(right: target; left: inversion result)

Overall goal:
Assess seismic hazard by computer simulation of
earthquake scenarios



The digital fault and fold map for southern California highlights blind thrust systems and other principal faults. Faults exposed at the surface are color-coded according to slip rates and earthquake recurrence intervals, with the red and orange features having higher rates of activity than the green and blue. Specific data on all faults are included in the geologic database.

Surface geology



<http://www.scec.org/phase3/images.html>

Complexity of earthquake ground motion simulation

- multiple spatial scales
 - wavelengths vary from $O(10\text{m})$ to $O(1000\text{m})$
 - Basin/source dimensions are $O(100\text{km})$
- multiple temporal scales
 - $O(0.01\text{s})$ to resolve highest frequencies of source
 - $O(10\text{s})$ to resolve of shaking within the basin
- highly irregular basin geometry
- highly heterogeneous soils material properties
- geology and source parameters observable only indirectly

Earthquake wave propagation model

$$\begin{aligned}\nabla \cdot \left[\mu (\nabla \mathbf{u} + \nabla \mathbf{u}^T) + \lambda (\nabla \cdot \mathbf{u}) \mathbf{I} \right] &= \rho \ddot{\mathbf{u}} - \mathbf{b} \text{ in } \Omega \times (0, T) \\ \left[\mu (\nabla \mathbf{u} + \nabla \mathbf{u}^T) + \lambda (\nabla \cdot \mathbf{u}) \mathbf{I} \right] \mathbf{n} &= \mathbf{L}^{AB} \mathbf{u} \text{ on } \partial\Omega \times (0, T) \\ \mathbf{u} &= \mathbf{0} \text{ on } \Omega \times \{t = 0\} \\ \dot{\mathbf{u}} &= \mathbf{0} \text{ on } \Omega \times \{t = 0\}\end{aligned}$$

+Rayleigh attenuation model

$\mathbf{u}(\mathbf{x}, t)$:= displacement
 ρ := material density
 μ, λ := elastic parameters
 $\mathbf{b}(\mathbf{x}, t)$:= rupture force, e.g. for point source
 $\mathbf{b}(\mathbf{x}, t) := -\mu v A f(t) \mathbf{M} \nabla \delta(\mathbf{x} - \boldsymbol{\xi})$

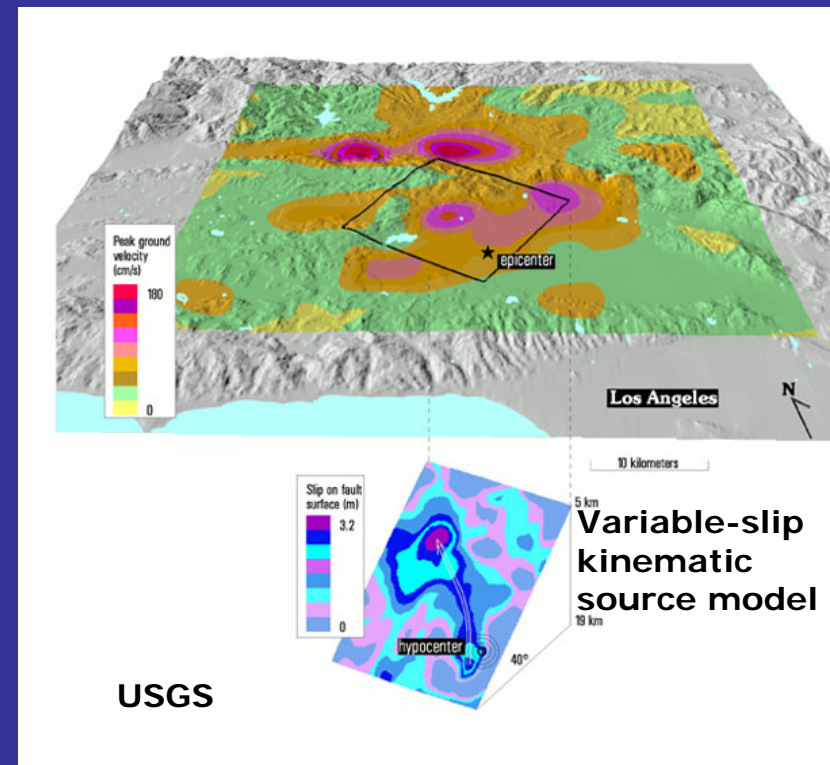
\mathbf{L}^{AB} is 0 on free surfaces, and is given by Stacy's absorbing boundary condition on truncated surfaces:

$$\mathbf{L}^{AB} \mathbf{u} \equiv \begin{bmatrix} -d_1 \frac{\partial}{\partial t} & c_1 \frac{\partial}{\partial \tau_1} & c_1 \frac{\partial}{\partial \tau_2} \\ -c_1 \frac{\partial}{\partial \tau_1} & -d_2 \frac{\partial}{\partial t} & 0 \\ -c_1 \frac{\partial}{\partial \tau_2} & 0 & -d_2 \frac{\partial}{\partial t} \end{bmatrix} \begin{Bmatrix} u_n \\ u_{\tau_1} \\ u_{\tau_2} \end{Bmatrix}$$

$$c_1 = -2\mu + \sqrt{\mu(\lambda + 2\mu)},$$

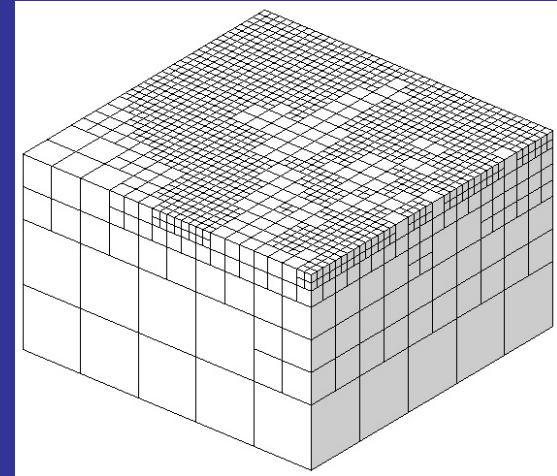
$$d_1 = \sqrt{\rho(\lambda + 2\mu)},$$

$$d_2 = \sqrt{\rho\mu}.$$



Wavelength-adaptive octree-based wave propagation solver

- Galerkin trilinear finite elements in space
- explicit central differences in time
- octree wavelength-adaptive meshes
 - o typical 10^3 X reduction in # grid pts vs. structured grid
 - o wavelength-adaptivity insures that CFL-limited time step of order of accuracy-driven time step
 - o low memory of stencil-based methods
 - o adaptivity of unstructured mesh methods
- algebraic constraints at hanging grid pts to maintain continuity of finite element approximation
- element-based matvecs results in good cache performance (25% scalar efficiency on EV68 Alpha)
- MPI implementation (87% parallel efficiency on 2K PEs)
- extensively verified with Green's functions & FD codes



Performance of forward octree-based earthquake modeling code on PSC HP AlphaServer cluster

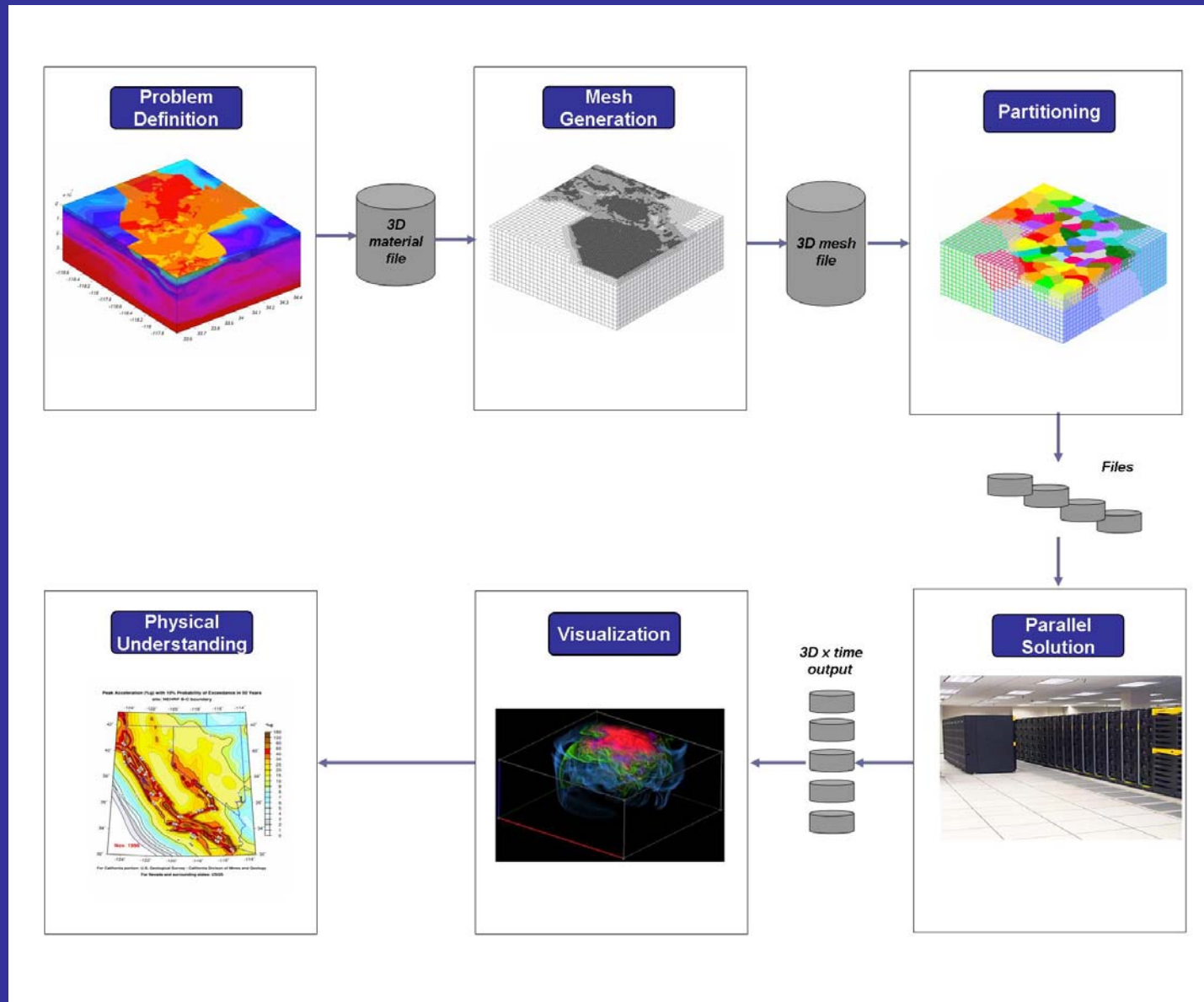
PEs	model	grids pts	pts/PE	Gflops	Mflops/PE	efficiency
1	LA10S	134,500	134,500	0.505	505	1.00
16	LA5S	618,672	38,667	7.85	491	0.972
128	LA2S	14,792,064	115,563	60.0	469	0.929
512	LA1HA	47,556,096	92,883	231	451	0.893
1024	LA1HB	101,940,152	99,551	460	450	0.891
2048	LA1HB	101,940,152	49,775	907	443	0.874
3000	LA1HB	101,940,152	33,980	1,210	403	0.800

- Largest (partial) simulation
 - o 28 Oct 2001 Compton aftershock in Greater LA Basin
 - o maximum resolved frequency: 1.85Hz
 - o 100m/s min shear wave velocity
 - o physical size: 100x100x37.5 km³
 - o # of elements: 899,591,066
 - o # of grid points: 1,023,371,641
 - o # of slaves: 125,726,862
 - o 25 sec wallclock/time step on 1024 PEs
 - o 65 Gb input

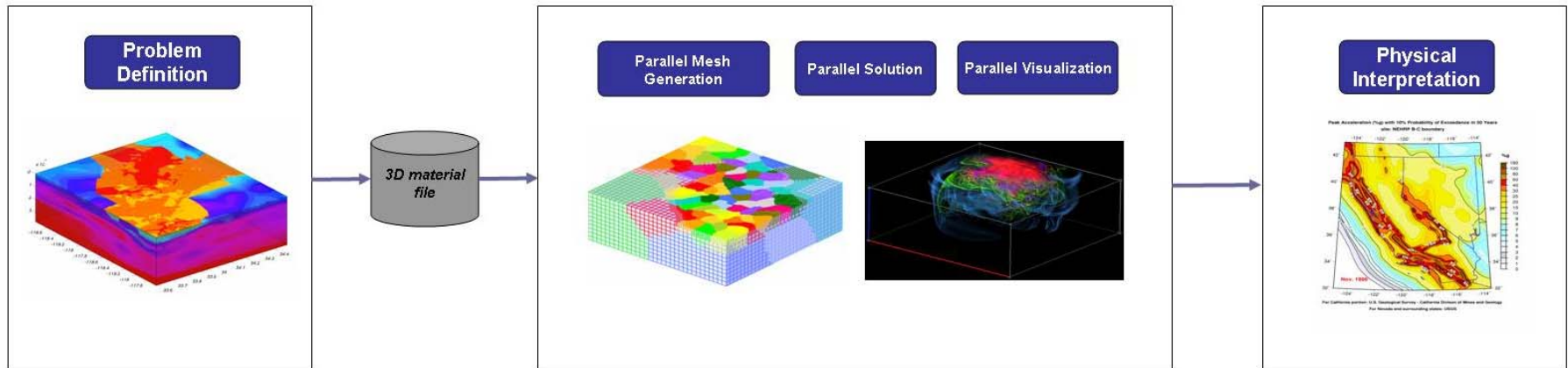


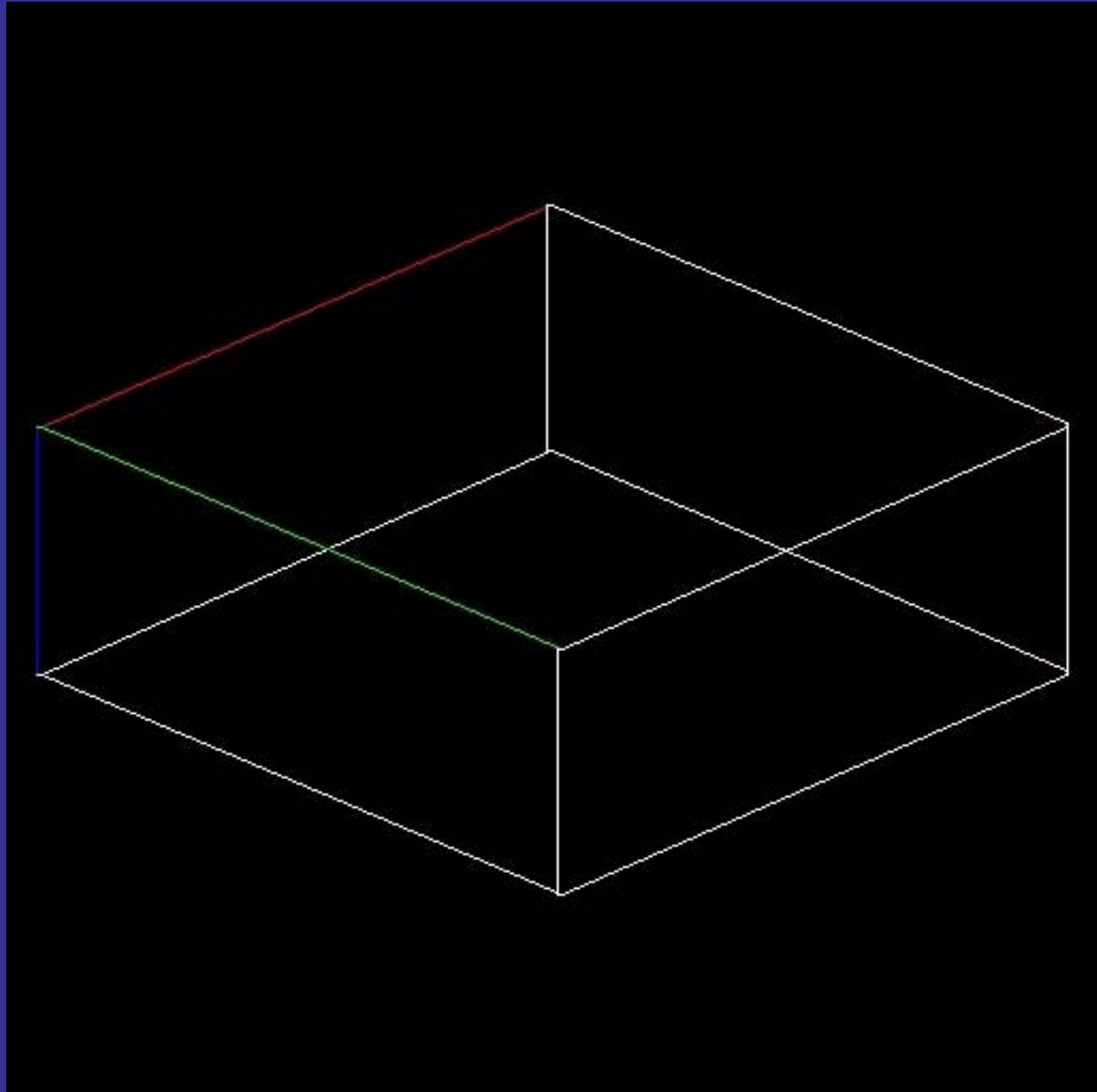
Iemieux at PSC

Old forward simulation: offline, file-based



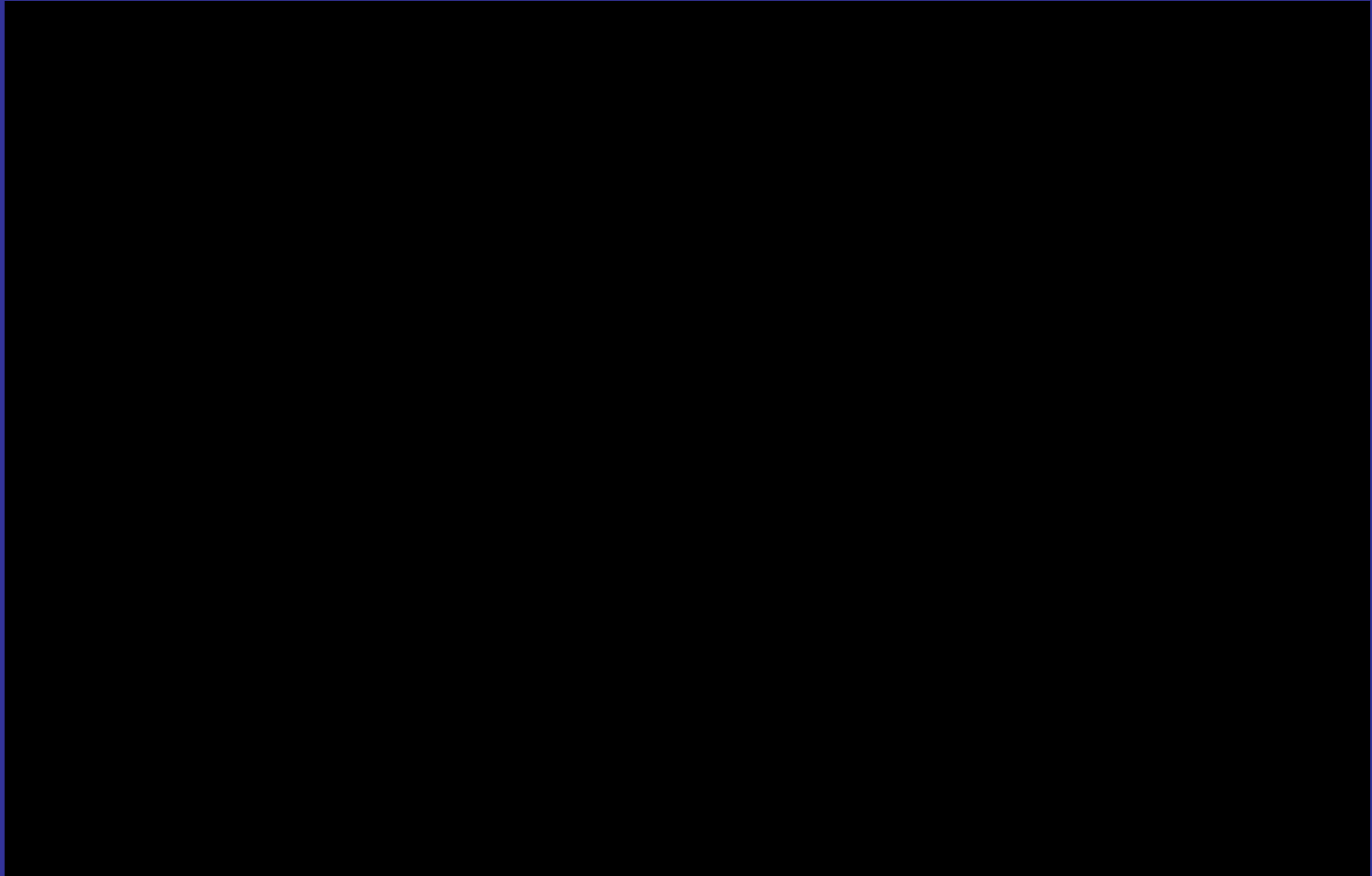
New forward simulation: online, parallel, lightweight





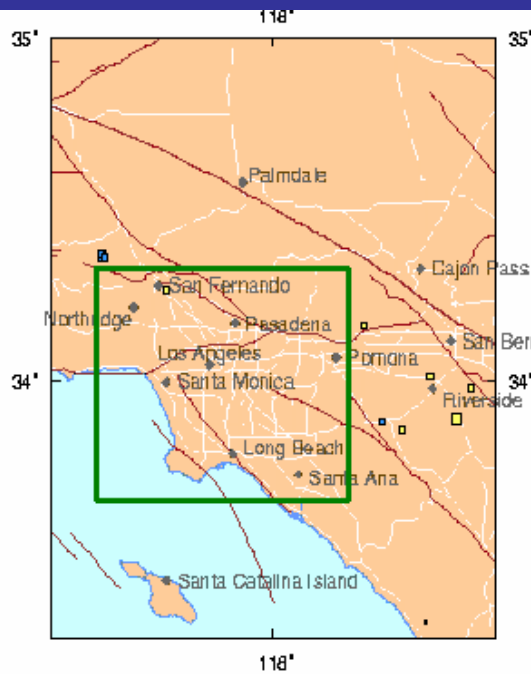
Joint work with K-L. Ma
and H. Fu, UC Davis

Surface visualization, aftershock

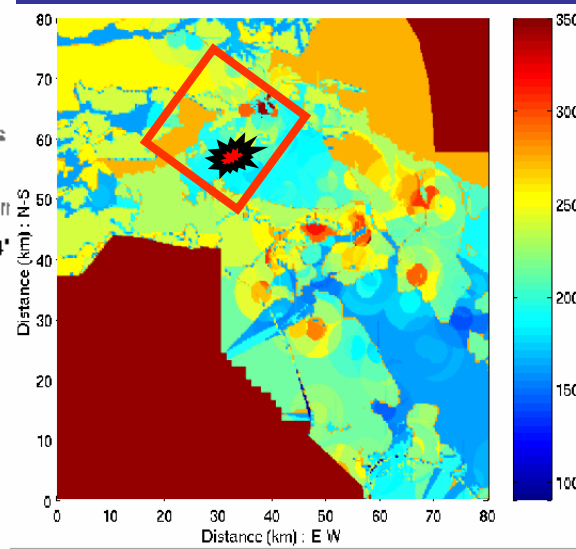


Animation by Greg Foss, PSC

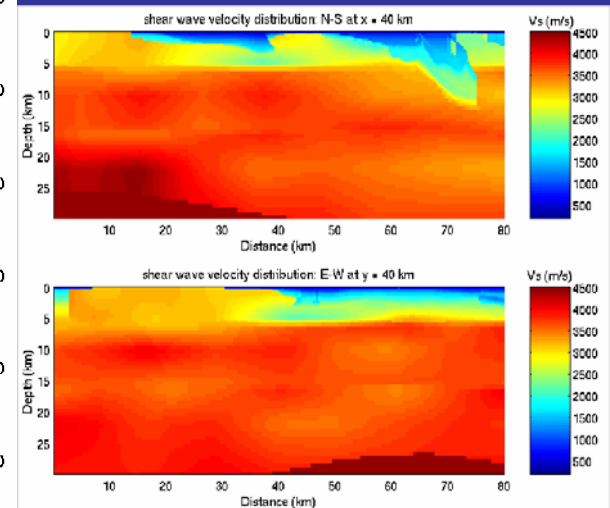
1994 Northridge earthquake



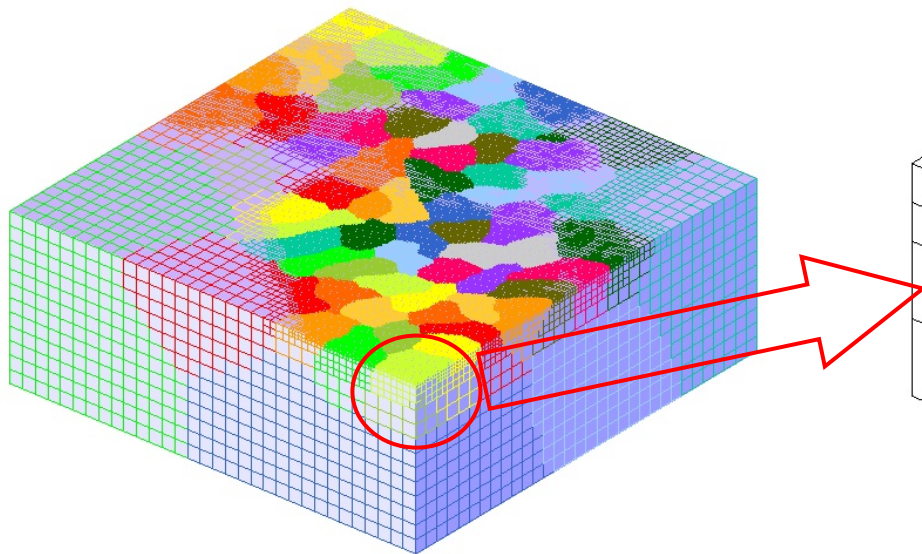
Computational domain



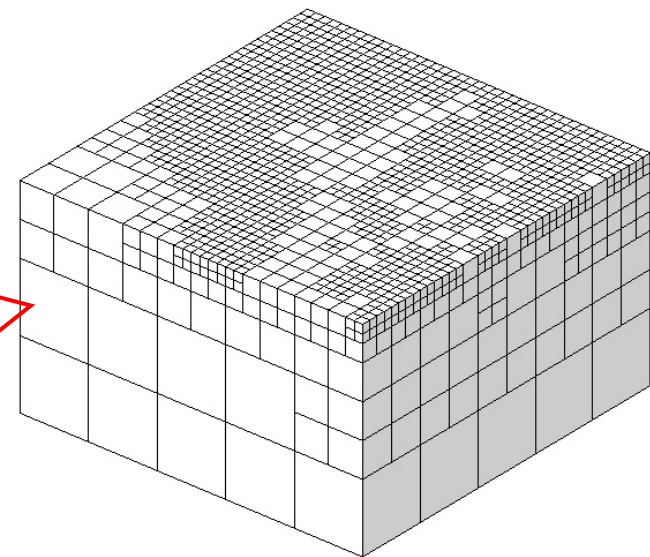
Surface shear wave velocity



Shear wave velocity at depth



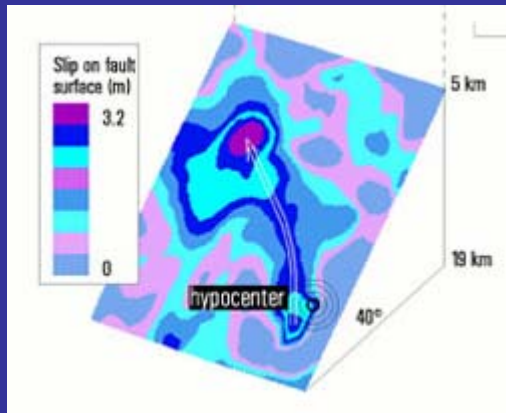
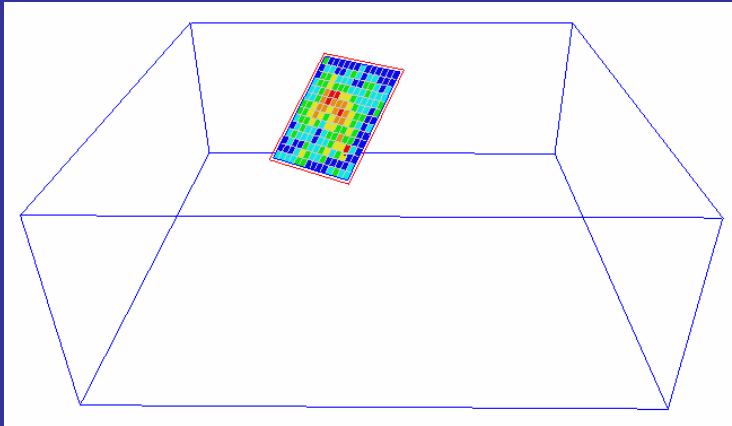
Octree-based hex mesh, partitioned by ParMetis



Detail of hexahedral mesh

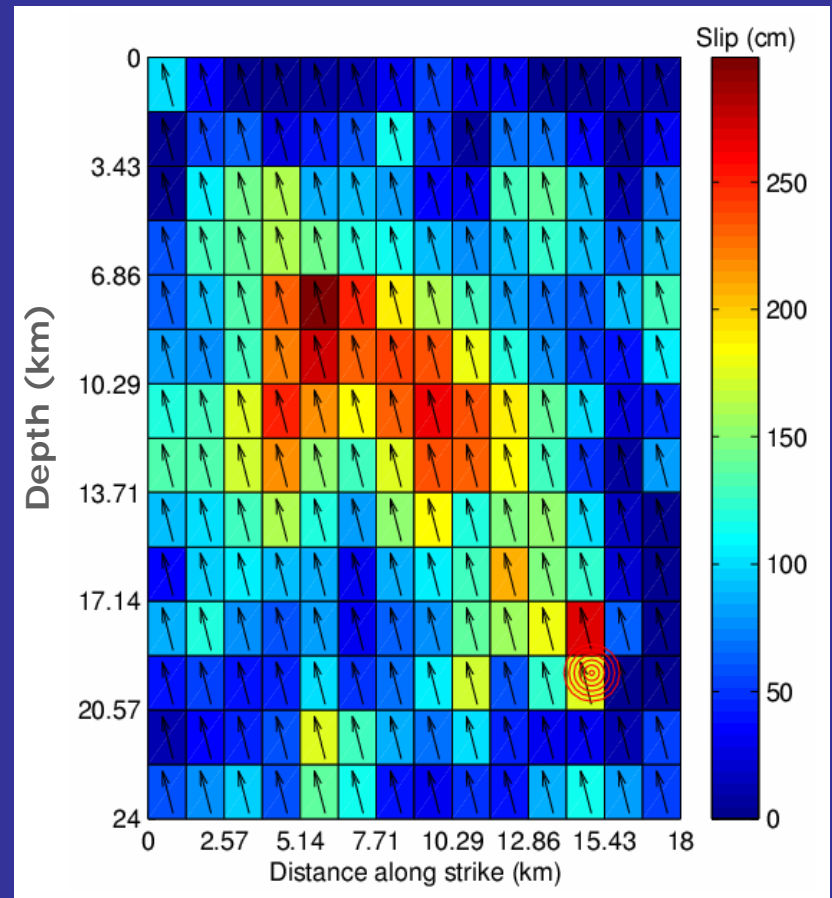
Rupture Model

Wald et al. (1996)

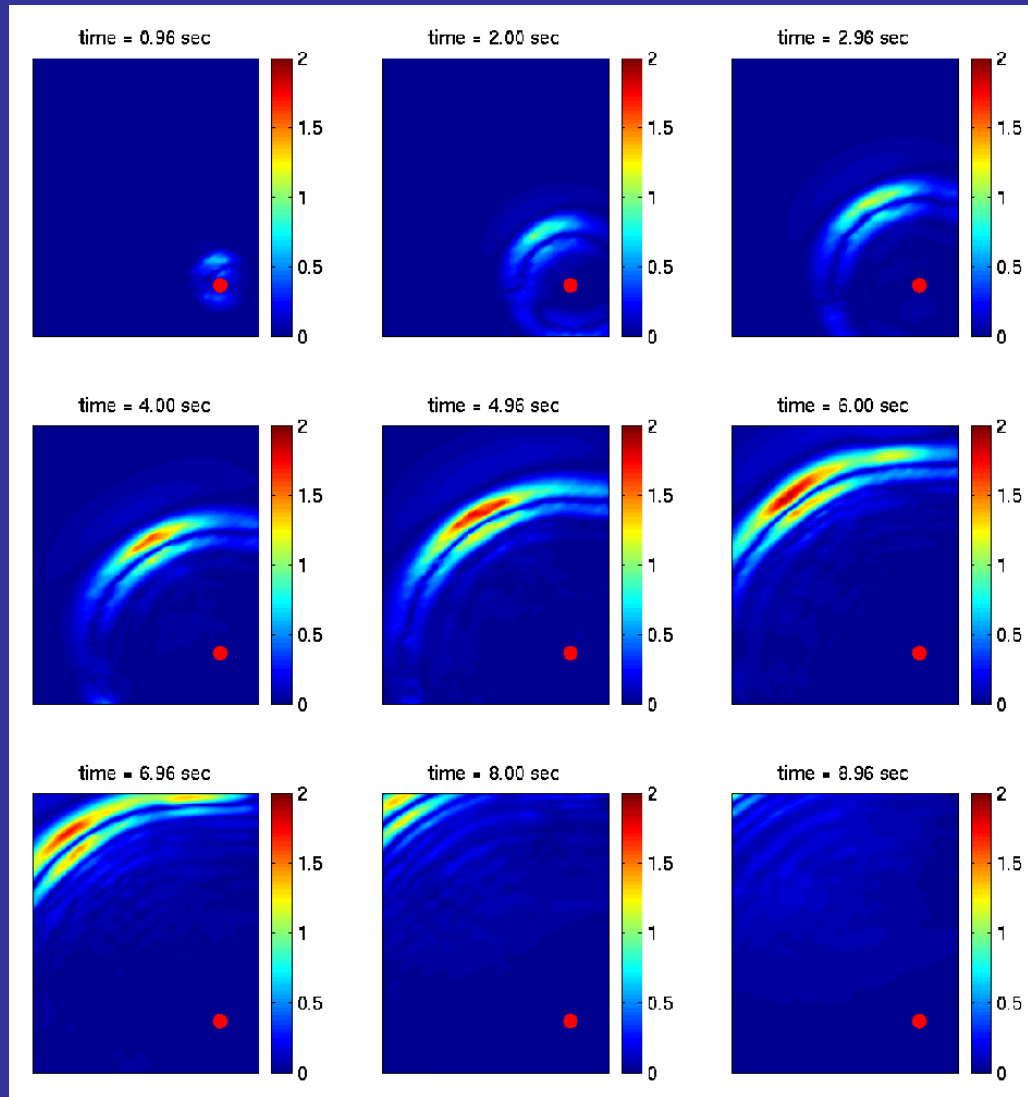


USGS

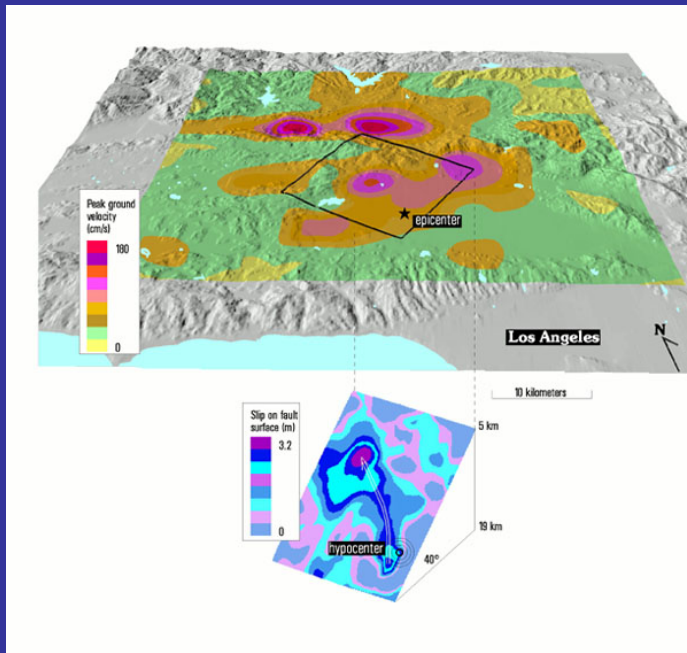
Strike=122 (S58E), Dip=40 (S32W), Rake=101



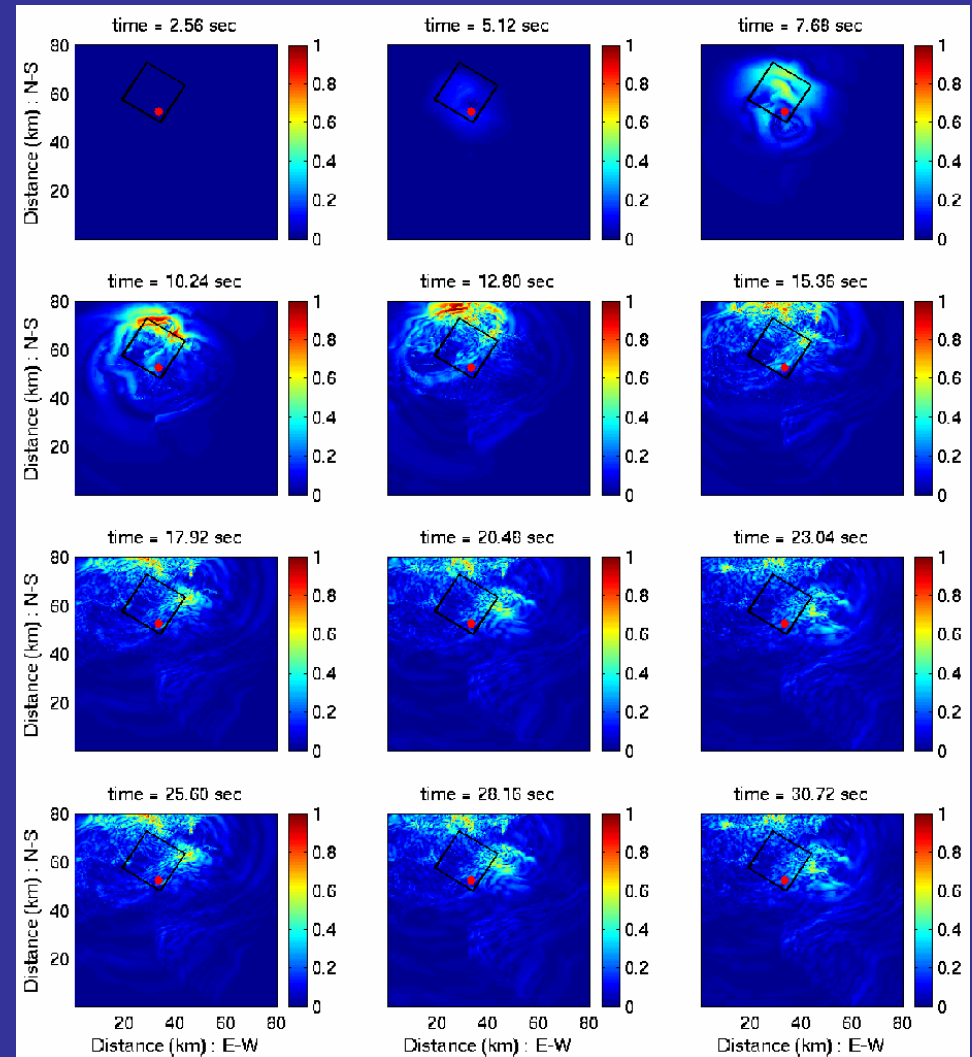
Rupture propagation (velocity)



Snapshots of surface velocity



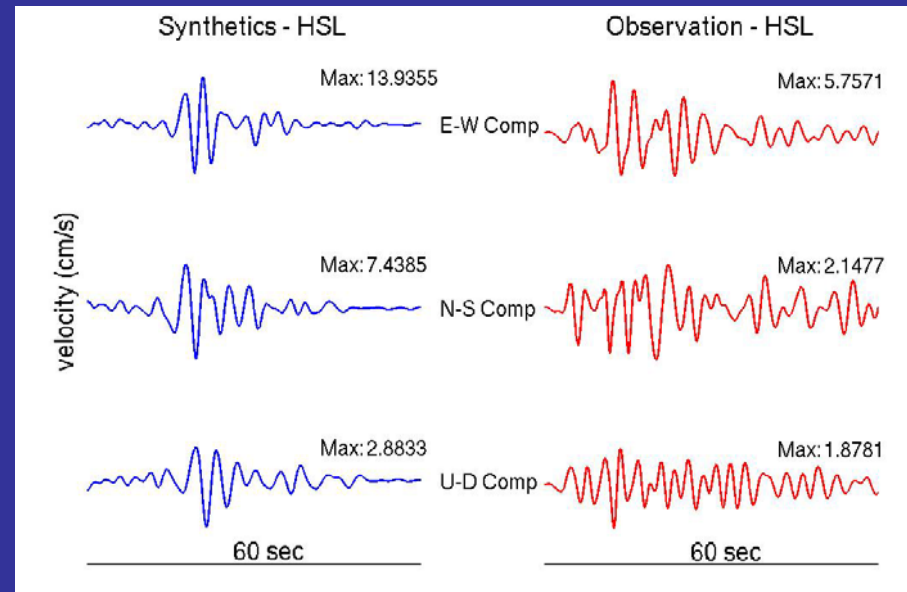
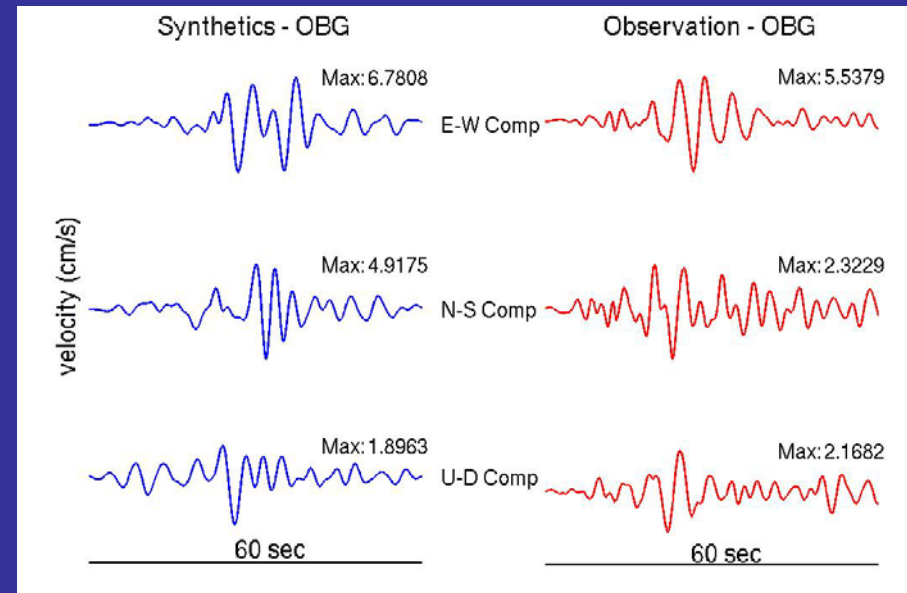
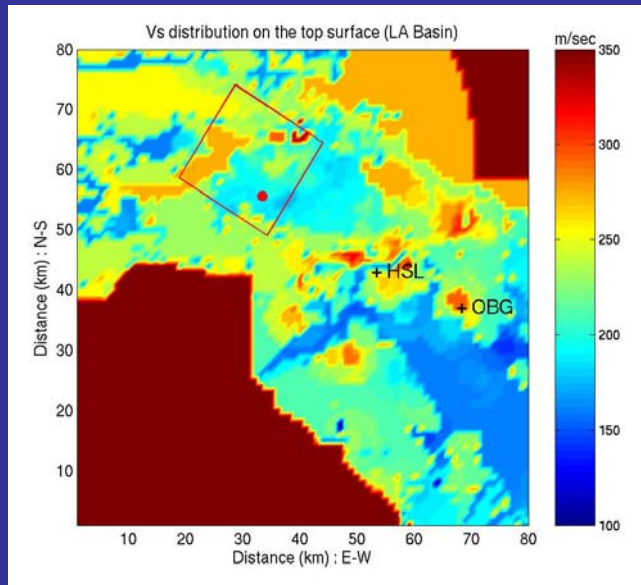
Peak observed
ground velocity
(USGS)

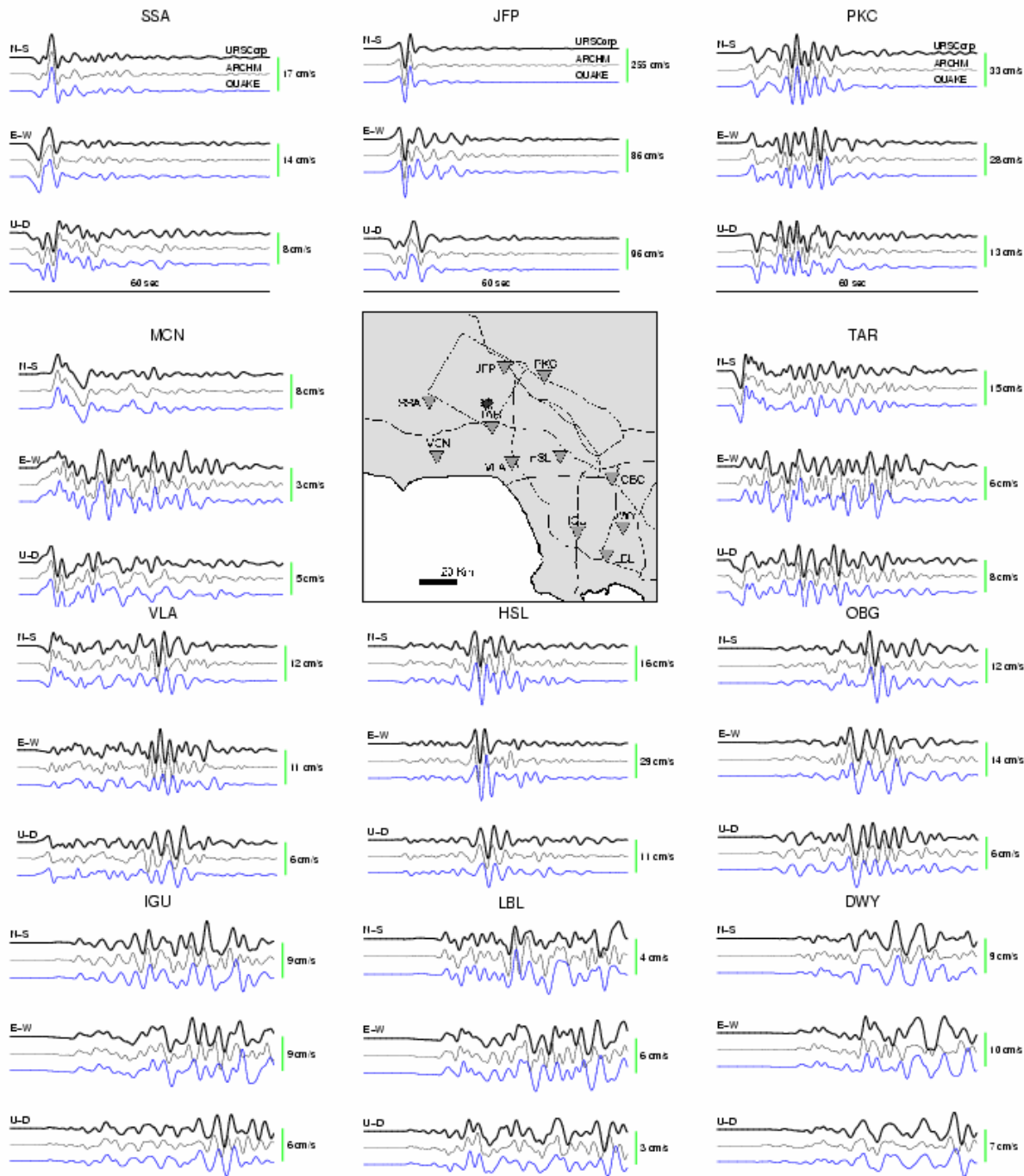


Comparison with observations

OBG: Obregon Park

HSL: Hollywood Storage

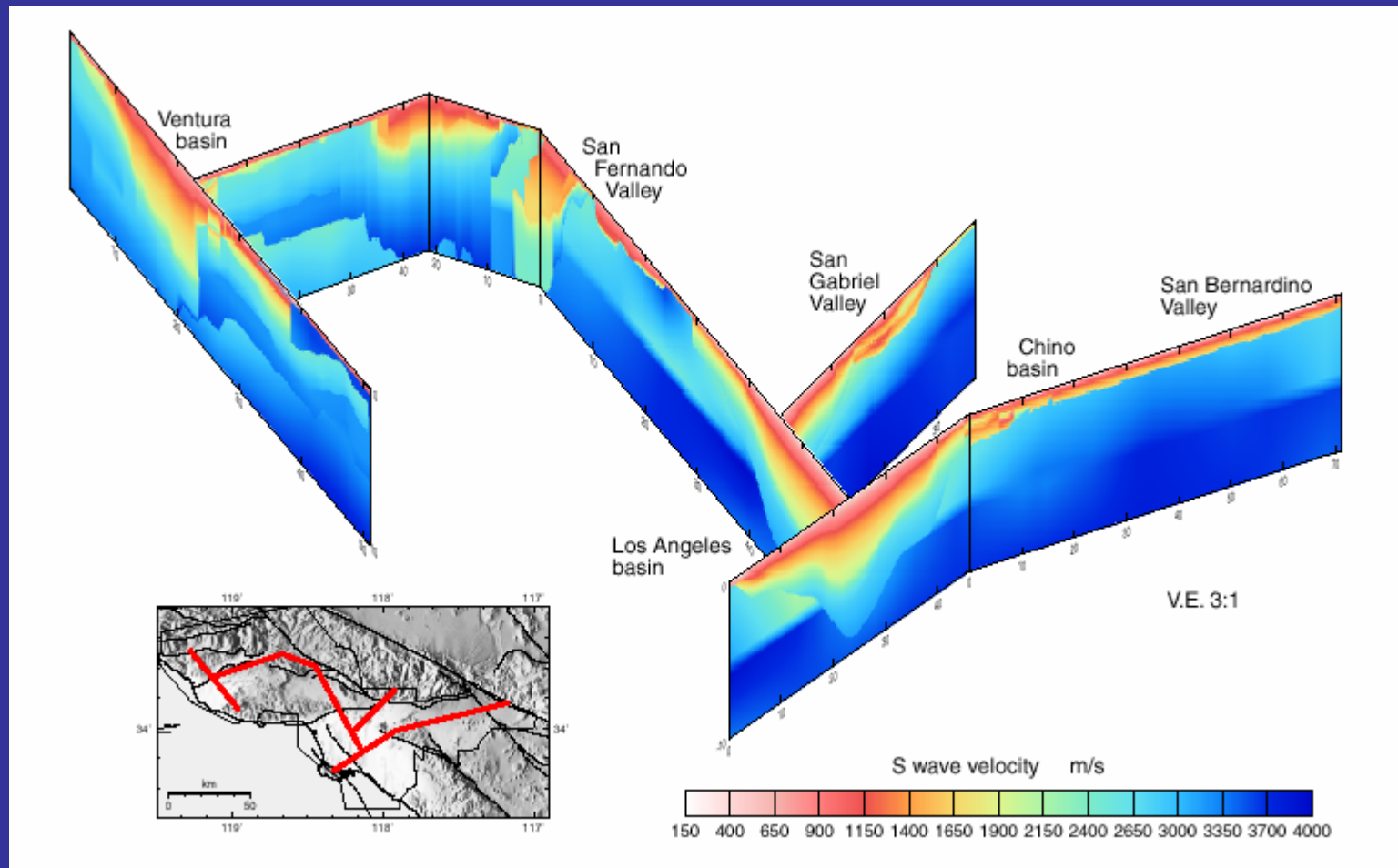




Verification
against
other codes

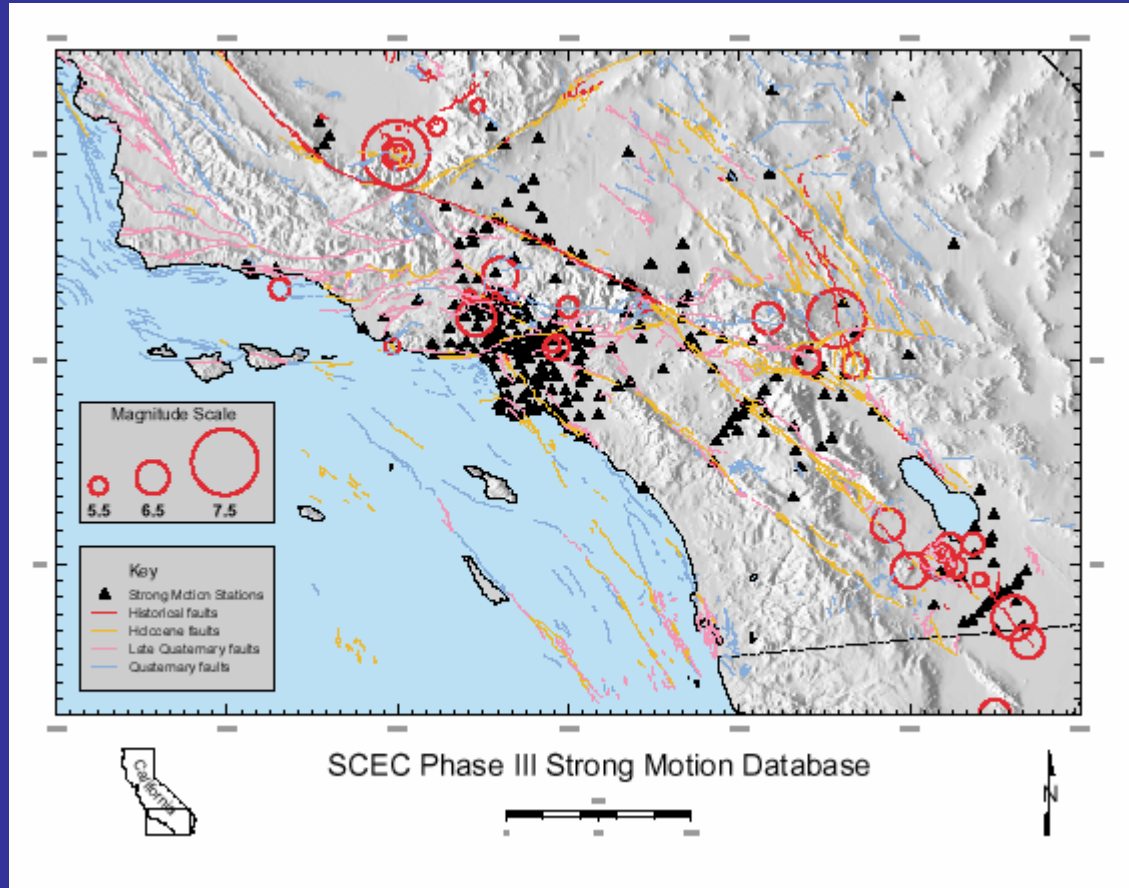
- R. Graves
- Archimedes
- Quake

SCEC Community Velocity Model for SoCal, v.3 (H. Magistrale, S. Day, R. Clayton, R. Graves)



Harold Magistrale, SDSU

Inverse problem: Use records of past seismic events to improve velocity model



SCEC Phase III strong motion database:

Observations from 28 earthquakes and 281 stations

Least squares parameter estimation formulation of inverse wave propagation

sources receivers data misfit

inversion fields

$$\min_{\mathbf{u}_i, \lambda, \mu} \mathcal{F} = \frac{1}{2} \sum_{i=1}^{N_s} \sum_{j=1}^{N_r^i} \int_0^T \int_{\Omega} (\mathbf{u}_i^* - \mathbf{u}_i)^2 \delta(\mathbf{x} - \mathbf{x}_{ij}) d\mathbf{x} dt$$

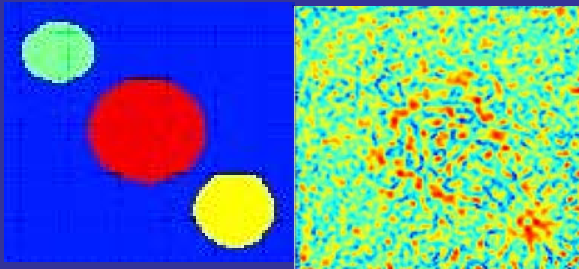
displacements

$$+ \int_{\Omega} \beta_{\mu} (\nabla \mu \cdot \nabla \mu + \epsilon)^{\frac{1}{2}} + \beta_{\lambda} (\nabla \lambda \cdot \nabla \lambda + \epsilon)^{\frac{1}{2}} d\mathbf{x}$$

subject to $\forall i$:

$$\left. \begin{aligned} \rho \ddot{\mathbf{u}}_i - \nabla \cdot [\mu (\nabla \mathbf{u}_i + \nabla \mathbf{u}_i^T) + \lambda (\nabla \cdot \mathbf{u}_i) \mathbf{I}] &= \mathbf{b}_i \quad \text{in } \Omega \times (0, T) \\ [\mu (\nabla \mathbf{u}_i + \nabla \mathbf{u}_i^T) + \lambda (\nabla \cdot \mathbf{u}_i) \mathbf{I}] \mathbf{n} &= \mathbf{L}_{AB} \mathbf{u}_i \quad \text{on } \partial\Omega \times (0, T) \\ \mathbf{u}_i &= 0 \quad \text{on } \Omega \times \{t = 0\} \\ \dot{\mathbf{u}}_i &= 0 \quad \text{on } \Omega \times \{t = 0\} \end{aligned} \right\}$$

forward
wave
propagation
model

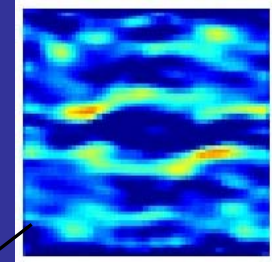


target

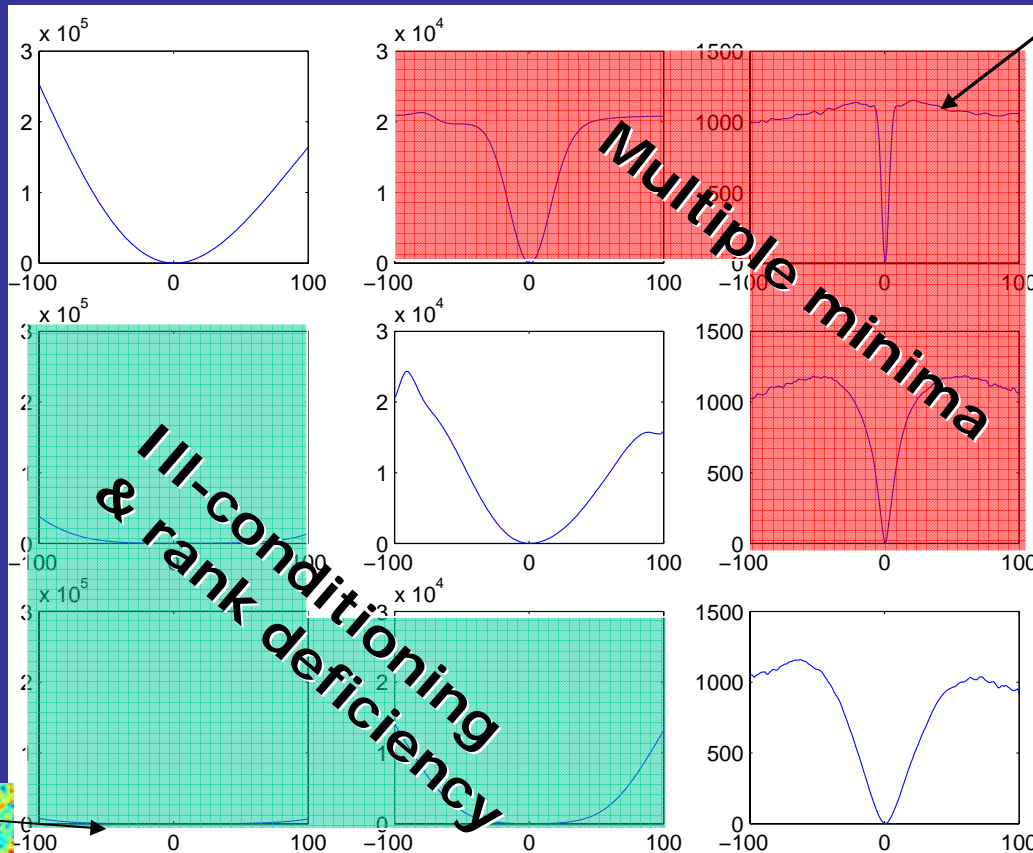
unregularized
solution

Behavior of misfit function F in direction of material perturbation

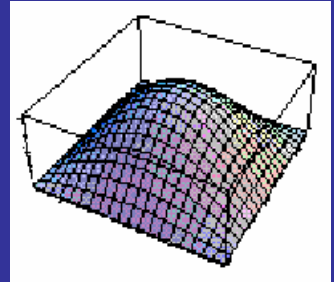
Source frequency



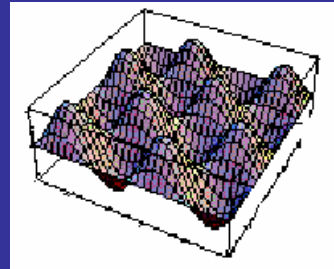
Frequency of perturbation of material field



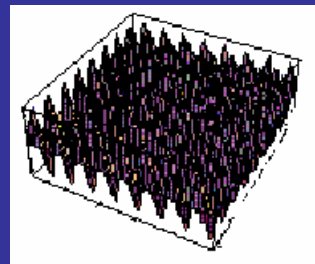
low



med



high

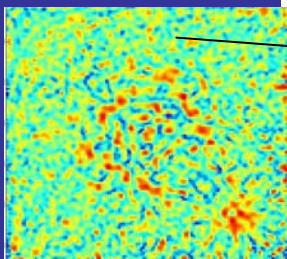


low

medium

high

$$\mathcal{F} = \frac{1}{2} \sum_{i=1}^{N_s} \sum_{j=1}^{N_r^i} \int_0^T \int_{\Omega} (u_i^* - u_i)^2 \delta(x - x_{ij}) dx dt$$



Least squares parameter estimation formulation of inverse wave propagation

sources receivers data misfit

inversion fields

$$\min_{u_i, \lambda, \mu} \mathcal{F} = \frac{1}{2} \sum_{i=1}^{N_s} \sum_{j=1}^{N_r^i} \int_0^T \int_{\Omega} (u_i^* - u_i)^2 \delta(x - x_{ij}) dx dt$$

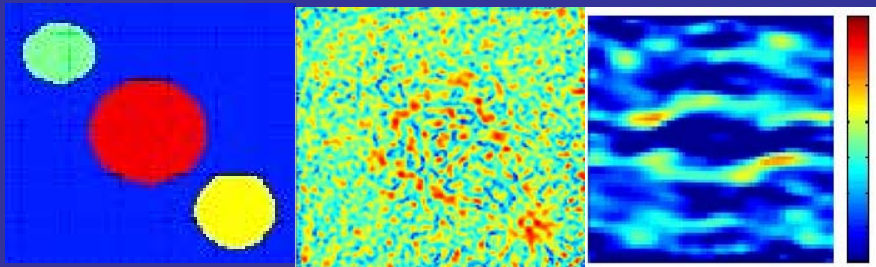
displacements

$$+ \int_{\Omega} \beta_{\mu} (\nabla \mu \cdot \nabla \mu + \epsilon)^{\frac{1}{2}} + \beta_{\lambda} (\nabla \lambda \cdot \nabla \lambda + \epsilon)^{\frac{1}{2}} dx$$

subject to $\forall i$:

$$\left. \begin{aligned} \rho \ddot{u}_i - \nabla \cdot [\mu (\nabla u_i + \nabla u_i^T) + \lambda (\nabla \cdot u_i) I] &= b_i \quad \text{in } \Omega \times (0, T) \\ [\mu (\nabla u_i + \nabla u_i^T) + \lambda (\nabla \cdot u_i) I] n &= L_{AB} u_i \quad \text{on } \partial \Omega \times (0, T) \\ u_i &= 0 \quad \text{on } \Omega \times \{t = 0\} \\ \dot{u}_i &= 0 \quad \text{on } \Omega \times \{t = 0\} \end{aligned} \right\}$$

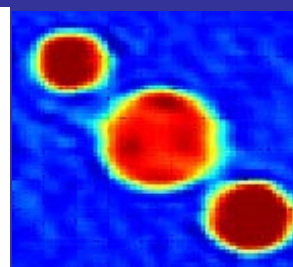
forward
wave
propagation
model



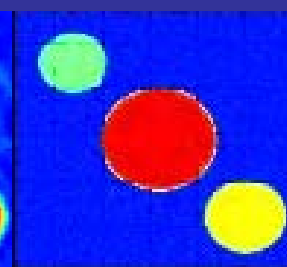
target

unregularized
solution

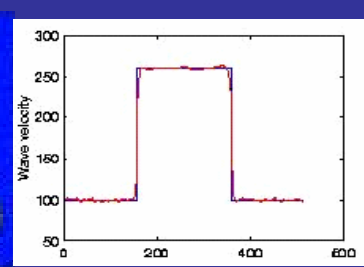
local minimum



Tikhonov reg.
+ multiscale



TV regular.
+ multiscale



TV reg., section

Least squares parameter estimation formulation of inverse wave propagation

sources
receivers

inversion fields
data misfit

$\min_{u_i, \lambda, \mu}$

displacements

subject to $\forall i$:

second variation

$$\mathcal{F} = \frac{1}{2} \sum_{i=1}^{N_s} \sum_{j=1}^{N_r^i} \int_0^T \int_{\Omega} (u_i^* - u_i)^2 \delta(x - x_{ij}) dx dt$$

$$+ \int_{\Omega} \left(\beta_{\mu} (\nabla \mu \cdot \nabla \mu + \varepsilon)^{\frac{1}{2}} + \beta_{\lambda} (\nabla \lambda \cdot \nabla \lambda + \varepsilon)^{\frac{1}{2}} \right) dx$$

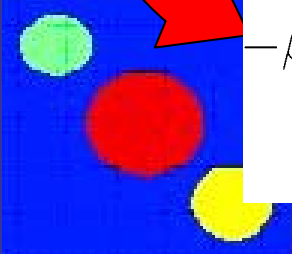
$$\rho \ddot{u}_i - \nabla \cdot \left[\mu (\nabla u_i + \nabla u_i^T) + \lambda (\nabla \cdot u_i) I \right] = b_i \quad \text{in } \Omega \times (0, T)$$

$$\left[\mu (\nabla u_i + \nabla u_i^T) + \lambda (\nabla \cdot u_i) I \right] n = L_{AB} u_i \quad \text{on } \partial \Omega \times (0, T)$$

$$u_i = 0 \quad \text{on } \Omega \times \{t = 0\}$$

$$\dot{u}_i = 0 \quad \text{on } \Omega \times \{t = 0\}$$

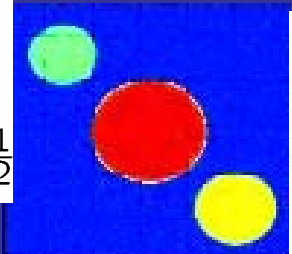
forward
wave
propagation
model



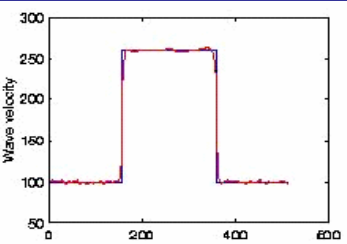
target

$$-\beta_{\mu} \nabla \cdot \left[\frac{1}{|\nabla \mu|_{\varepsilon}} \left(I - \frac{\nabla \mu \otimes \nabla \mu}{|\nabla \mu|_{\varepsilon}^2} \right) \nabla \tilde{\mu} \right],$$

where $|\nabla \mu|_{\varepsilon} := (\nabla \mu \cdot \nabla \mu + \varepsilon)^{\frac{1}{2}}$



TV regular.
+ multiscale



TV reg., section

Simpler problem: inverse scalar wave propagation (antiplane shear)

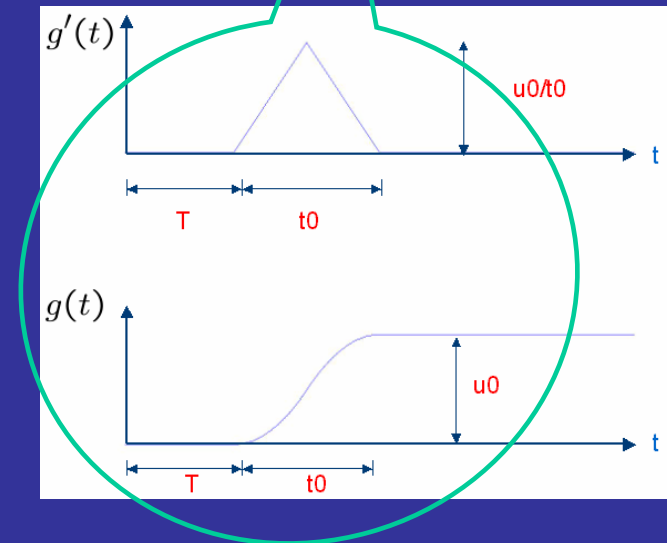
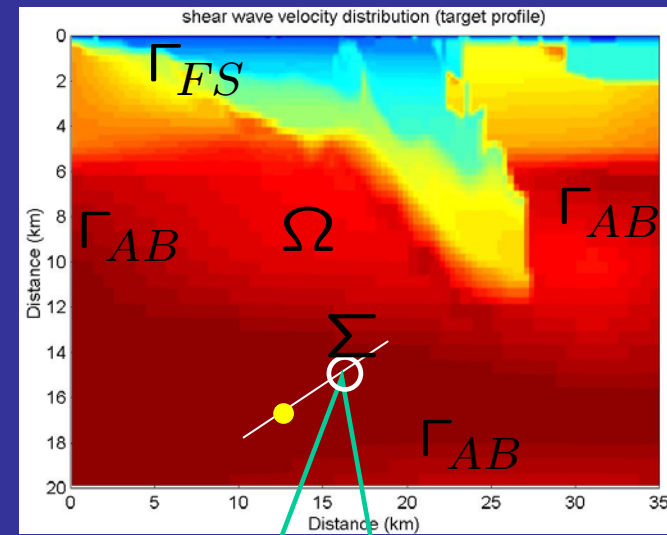
Minimize w.r.t. μ

$$\frac{1}{2} \sum_{j=1}^{N_r} \int_{\Omega} \int_0^T (u - u^*)^2 \delta(x - x_j) dx dt + \beta \int_{\Omega} |\nabla \mu|_{\varepsilon} dx$$

subject to:

$$\begin{aligned} \rho u_{tt} - \nabla \cdot \mu \nabla u &= -\nabla \cdot (\mu u_0 g(t_0, T) \delta(\Sigma) \mathbf{n}_{\Sigma}) && \text{in } \Omega \times (0, T) \\ \mu \nabla u \cdot \mathbf{n} &= 0 && \text{on } \Gamma_{FS} \times (0, T) \\ \mu \nabla u \cdot \mathbf{n} &= \sqrt{\rho \mu} u_t && \text{on } \Gamma_{AB} \times (0, T) \\ u = u_t &= 0 && \text{in } \Omega \times \{t = 0\} \end{aligned}$$

$u(x, t) :=$ transverse displacement
 $\rho(x) :=$ density
 $\mu(x) :=$ shear modulus
 $T(\Sigma) :=$ delay time
 $t_0(\Sigma) :=$ rise time
 $u_0(\Sigma) :=$ dislocation magnitude
 $\mathbf{n}_{\Sigma} :=$ fault normal



Lagrangian and weak form of optimality system

$$\begin{aligned}\mathcal{L}(u, p, \mu) := & \frac{1}{2} \sum_{j=1}^{N_r} \int_0^T \int_{\Omega} (u - u^*)^2 \delta(\mathbf{x} - \mathbf{x}_j) d\mathbf{x} dt + \beta \int_{\Omega} |\nabla \mu|_{\varepsilon} d\mathbf{x} \\ & + \int_0^T \int_{\Omega} (\mu \nabla u \cdot \nabla p - \rho u_t p_t) d\mathbf{x} dt - \int_0^T \int_{\Sigma} \mu u_0 g \nabla p \cdot \mathbf{n}_{\Sigma} ds dt - \int_0^T \int_{\Gamma_{AB}} u_t p \sqrt{\rho \mu} ds dt\end{aligned}$$



Strong form of first order necessary conditions

State equation

$$\begin{aligned}
 \rho u_{tt} - \nabla \cdot \mu \nabla u &= -\nabla \cdot (\mu u_0 g(t_0, T) \delta(\Sigma) \mathbf{n}_\Sigma) && \text{in } \Omega \times (0, T) \\
 \mu \nabla u \cdot \mathbf{n} &= 0 && \text{on } \Gamma_{FS} \times (0, T) \\
 \mu \nabla u \cdot \mathbf{n} &= \sqrt{\rho \mu} u_t && \text{on } \Gamma_{AB} \times (0, T) \\
 u = u_t &= 0 && \text{in } \Omega \times \{t = 0\}
 \end{aligned}$$

Adjoint equation

$$\begin{aligned}
 \rho p_{tt} - \nabla \cdot \mu \nabla p &= - \sum_{j=1}^{N_r} (u - u^*) \delta(\mathbf{x} - \mathbf{x}_j) && \text{in } \Omega \times (0, T) \\
 \mu \nabla p \cdot \mathbf{n} &= 0 && \text{on } \Gamma_{FS} \times (0, T) \\
 \mu \nabla p \cdot \mathbf{n} &= -\sqrt{\rho \mu} p_t && \text{on } \Gamma_{AB} \times (0, T) \\
 p = p_t &= 0 && \text{in } \Omega \times \{t = T\}
 \end{aligned}$$

Shear modulus equation

$$\begin{aligned}
 -\nabla \cdot \left(\frac{\beta}{|\nabla \mu|_\varepsilon} \nabla \mu \right) &= - \int_0^T (\nabla u \cdot \nabla p + u_0 g \nabla p \cdot \mathbf{n}_\Sigma \delta(\Sigma)) dt && \text{in } \Omega \\
 \frac{\beta}{|\nabla \mu|_\varepsilon} \nabla \mu \cdot \mathbf{n} &= 0 && \text{on } \Gamma_{FS} \\
 \frac{\beta}{|\nabla \mu|_\varepsilon} \nabla \mu \cdot \mathbf{n} &= \frac{1}{2} \int_0^T u_t p \sqrt{\frac{\rho}{\mu}} dt && \text{on } \Gamma_{AB}
 \end{aligned}$$

The Newton step



The Newton step, symbolically

$$\begin{bmatrix} \mathcal{B} & \mathcal{C}^*(p) & \mathcal{A}^*(\mu) \\ \mathcal{C}(p) & \mathcal{R}(\mu) & \mathcal{D}^*(u) \\ \mathcal{A}(\mu) & \mathcal{D}(u) & 0 \end{bmatrix} \begin{bmatrix} \tilde{u} \\ \tilde{\mu} \\ \tilde{p} \end{bmatrix} = - \begin{bmatrix} \mathcal{L}_u(u, \mu, p) \\ \mathcal{L}_\mu(u, \mu, p) \\ \mathcal{L}_p(u, \mu) \end{bmatrix}$$

where

$$(\hat{p}, \mathcal{A}(\mu) \tilde{u}) = (\tilde{u}, \mathcal{A}^*(\mu) \hat{p}) = \int_0^T \int_{\Omega} (\mu \nabla \tilde{u} \cdot \nabla \hat{p} - \rho \tilde{u}_t \hat{p}_t) d\mathbf{x} dt - \int_0^T \int_{\Gamma_{AB}} \tilde{u}_t \hat{p} \sqrt{\rho \mu} ds dt$$

$$(\hat{u}, \mathcal{B} \tilde{u}) = \sum_{j=1}^{N_r} \int_0^T \int_{\Omega} \tilde{u} \hat{u} \delta(\mathbf{x} - \mathbf{x}_j) d\mathbf{x} dt$$

$$(\hat{\mu}, \mathcal{C}(p) \tilde{u}) = (\tilde{u}, \mathcal{C}^*(p) \hat{\mu}) = \int_0^T \int_{\Omega} (\hat{\mu} \nabla \tilde{u} \cdot \nabla p d\mathbf{x} dt - \int_0^T \int_{\Gamma_{AB}} \hat{\mu} \tilde{u}_t p \sqrt{\frac{\rho}{4\mu}} ds dt$$

$$\begin{aligned} (\hat{\mu}, \mathcal{R}(\mu) \tilde{\mu}) &= \int_{\Omega} \frac{\beta}{|\nabla \mu|_{\varepsilon}} \nabla \hat{\mu} \cdot \left(\mathbf{I} - \frac{\nabla \mu \otimes \nabla \mu}{|\nabla \mu|_{\varepsilon}^2} \right) \cdot \nabla \tilde{\mu} d\mathbf{x} \\ &+ \int_0^T \int_{\Gamma_{AB}} \hat{\mu} u_t p \tilde{\mu} \sqrt{\frac{\rho}{16\mu^3}} ds dt \end{aligned}$$

$$\begin{aligned} (\hat{p}, \mathcal{D}(u) \tilde{\mu}) &= (\tilde{\mu}, \mathcal{D}^*(u) \hat{u}) = \int_0^T \int_{\Omega} \tilde{\mu} \nabla u \cdot \nabla \hat{p} d\mathbf{x} dt - \int_0^T \int_{\Sigma} \tilde{\mu} u_0 g \nabla \hat{p} \cdot \mathbf{n}_{\Sigma} ds dt \\ &- \int_0^T \int_{\Gamma_{AB}} u_t \hat{p} \tilde{\mu} \sqrt{\frac{\rho}{4\mu}} ds dt \end{aligned}$$

A Gauss-Newton-Schur-CG method

The Gauss-Newton step:

$$\begin{bmatrix} \mathcal{B} & \mathcal{C}^*(p) & \mathcal{A}^*(\mu) \\ \mathcal{C}(p) & \mathcal{R}(\mu) & \mathcal{D}^*(u) \\ \mathcal{A}(\mu) & \mathcal{D}(u) & 0 \end{bmatrix} \begin{Bmatrix} \tilde{u} \\ \tilde{\mu} \\ \tilde{p} \end{Bmatrix} = - \begin{Bmatrix} \mathcal{L}_u(u, \mu, p) \\ \mathcal{L}_\mu(u, \mu, p) \\ \mathcal{L}_p(u, \mu) \end{Bmatrix}$$

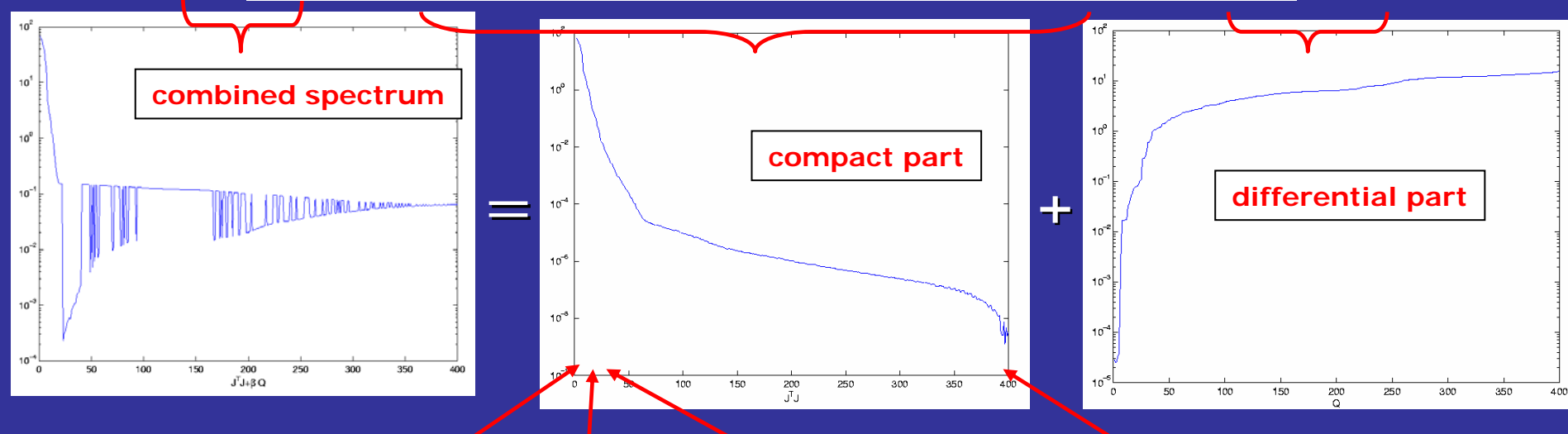
A reduced space approach:

$$\left(\mathcal{D}^* \mathcal{A}^{*-1} \mathcal{B} \mathcal{A}^{-1} \mathcal{D} - \mathcal{D}^* \mathcal{A}^{*-1} \mathcal{C} - \mathcal{C} \mathcal{A}^{-1} \mathcal{D} + \mathcal{R} \right) \tilde{\mu} = \mathcal{D}^* \mathcal{A}^{*-1} \mathcal{L}_u - \mathcal{L}_\mu$$

- Instead, use Gauss-Newton approximation and solve by matrix-free conjugate gradients:
 - o Intractable, e.g. for largest problem we solve:
 - o form Hessian-vector products on the fly, terminate early
 - o 1/7 million wave propagations to set up linear system per Eisenstat-Walker
 - o 2 petabytes to store it
 - o Hessian guaranteed to be positive definite
 - o quadratic convergence for good fit problems, linear otherwise
 - o 3 hours on an exascale machine for one Newton iteration
 - o each CG iteration requires 1 forward, 1 adjoint wave propagation -> parallelizes as well as forward problem (N_s needed, arrives in 2050 per Denni Moore's law)
 - o need good preconditioner (but difficult, since Hessian not available)

Spectrum of reduced Hessian & preconditioning issues

$$\mathcal{S} := \mathcal{D}^* \mathcal{A}^{*-1} \mathcal{B} \mathcal{A}^{-1} \mathcal{D} + \mathcal{R}$$



Preconditioner:

- CG rapidly eliminates error components associated with large eigenvalues (smooth eigenvectors)
- capture these via limited memory BFGS preconditioner (Nocedal-Morales)
- initialize LBFGS inverse approximation with several iterations of Frankel's two-step stationary iterative solver on $\alpha \mathcal{I} + \beta \mathcal{R}$



Solution algorithm:

Multiscale-Gauss-Newton-CG-LMBFGS

- Multiscale continuation over grid and source frequency (Chavent '95)
 - Inexact Gauss-Newton nonlinear iteration with Armijo backtracking line search
- Matrix-free conjugate gradient solution of reduced Hessian system (each matvec requires N_s forward & adjoint wave propagation solutions)
 - Preconditioner:
 - » limited memory BFGS (Morales-Nocedal '00)
 - » initialized with several iterations of Frankel's method (two-step stationary method) to "invert" $\alpha \mathcal{I} + \beta \mathcal{F}$

Algorithmic scalability for 3D acoustic inversion example

material grid	Picard-Gauss-Newton-Krylov iter, no PC			Picard-Gauss-Newton-Krylov iter, LBFGS/2SR PC		
	nonlinear iter	total linear iter	avg linear iter	nonlinear iter	total linear iter	avg linear iter
2^3	6	31	5.2	6	13	2.2
3^3	11	121	11.0	11	39	3.5
5^3	18	321	17.8	17	144	8.5
9^3	13	614	47.2	12	249	21.0
17^3	11	1413	128.5	12	396	33.0
33^3	17	1445	85.0	25	439	17.6
65^3	19	1923	101.2	19	370	19.5
129^3	21	2003	95.4	22	436	19.8

Mesh independence of
nonlinear iterations

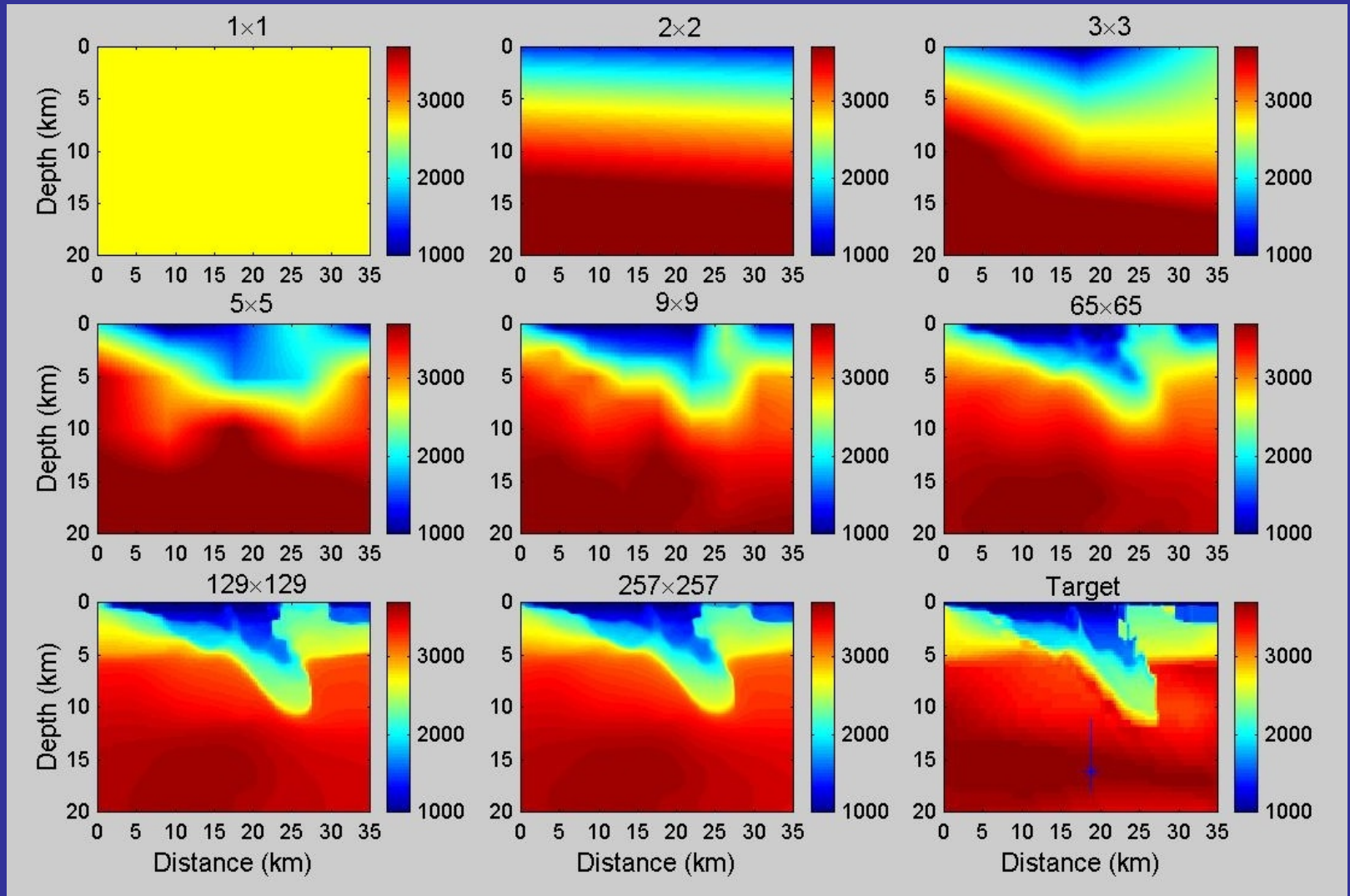
Mesh independence
of linear iterations

But even with mesh
independence, # of wave
propagations still large!

Inversion examples

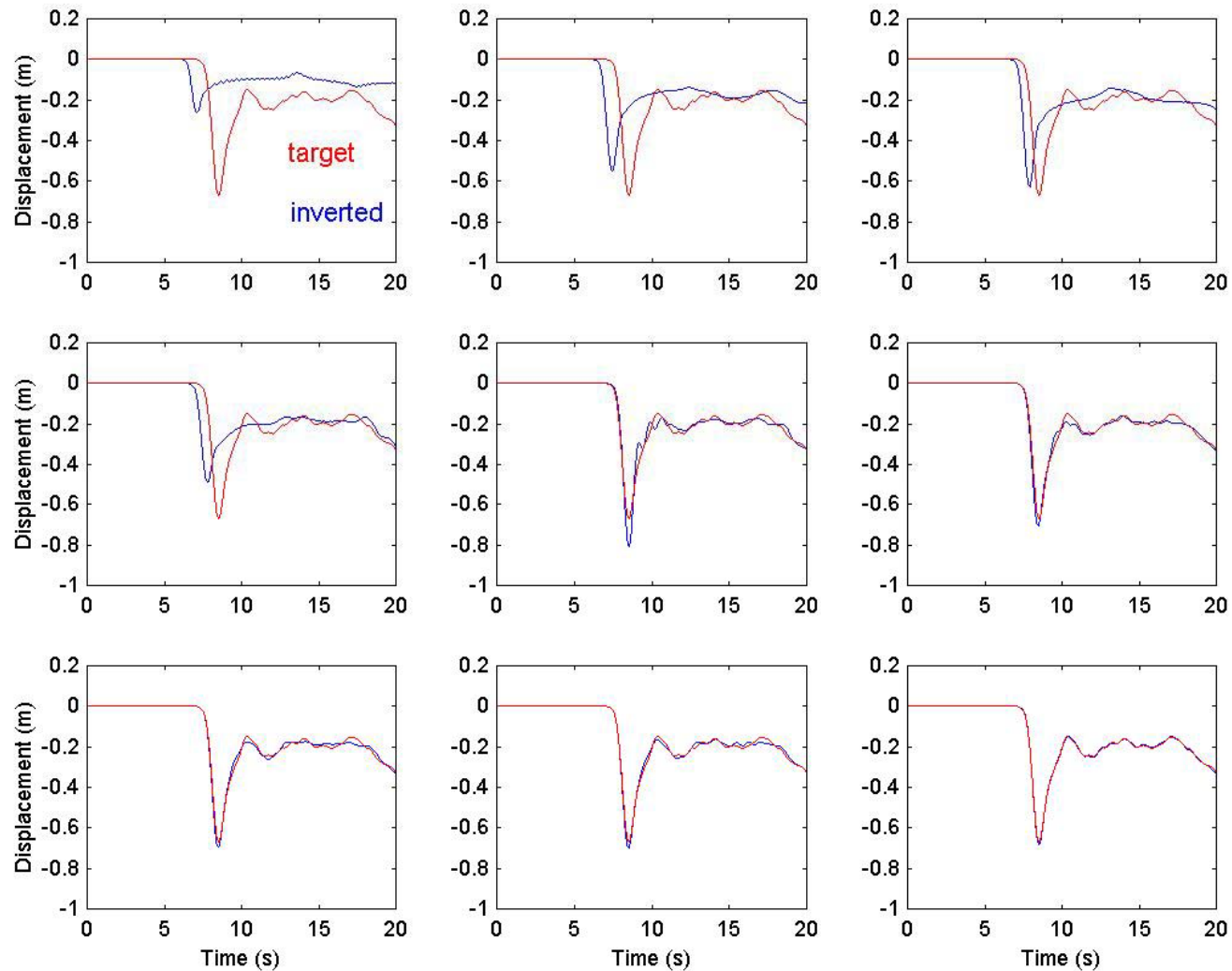
- 2D shear, 3D acoustic, and 3D elastic models
- Synthetic inversion (some with 5% added noise) using SCEC community velocity model
- Piecewise bi/trilinear finite element approximation of state, adjoint, and material property in space
- Explicit central difference time approximation in time
- PETSc (www.petsc.anl.gov) parallel implementation
- Up to $257 \times 257 \times 257$ grid (17 million inversion parameters) on 2048 processors (~ 12 h)
- Up to 225 surface receivers

Material inversion: multiscale continuation (64 receivers)

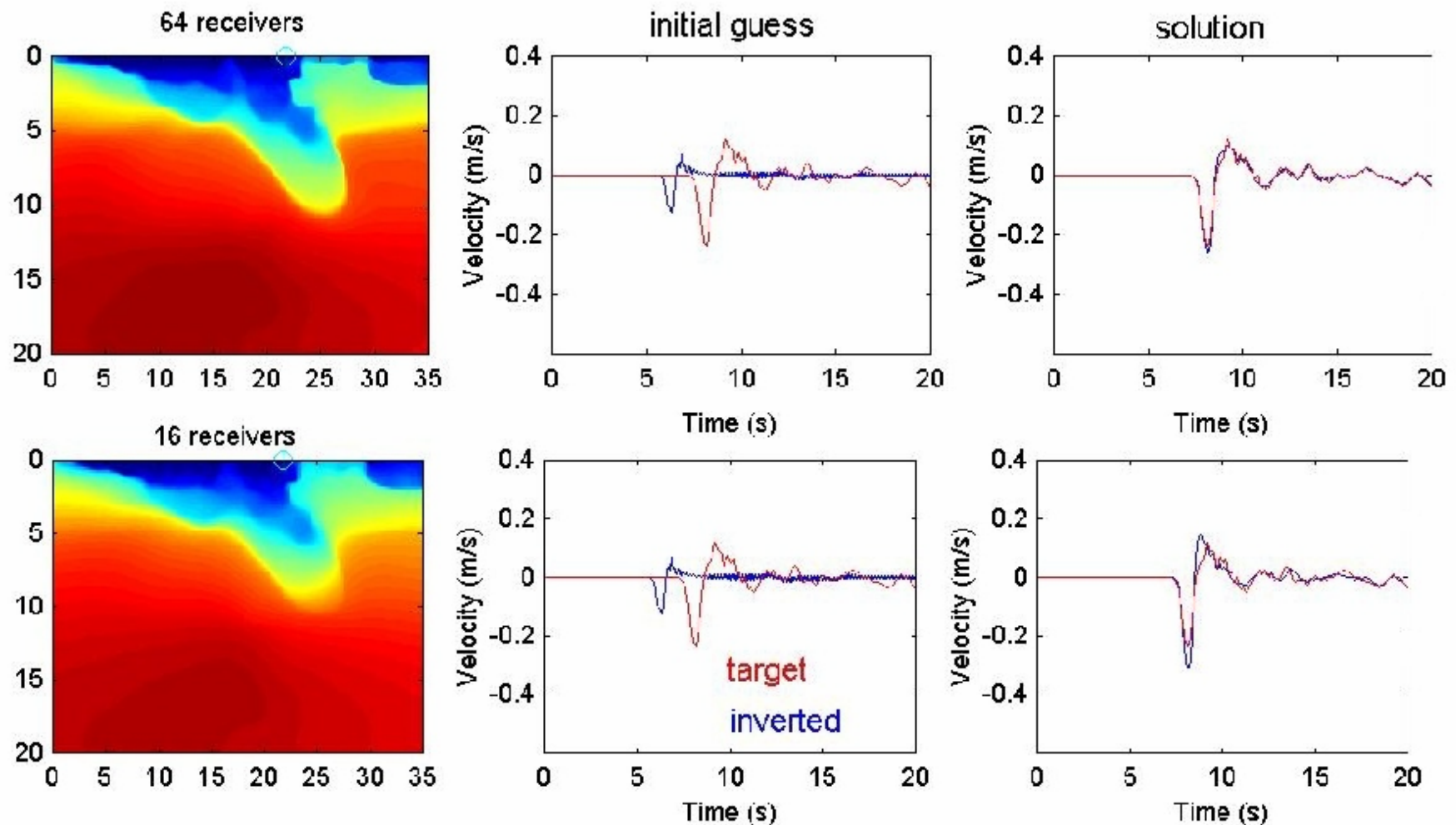


Material inversion: target vs. inverted displacement history at a receiver

Inverted and target dynamic response at $x = 220$ m.

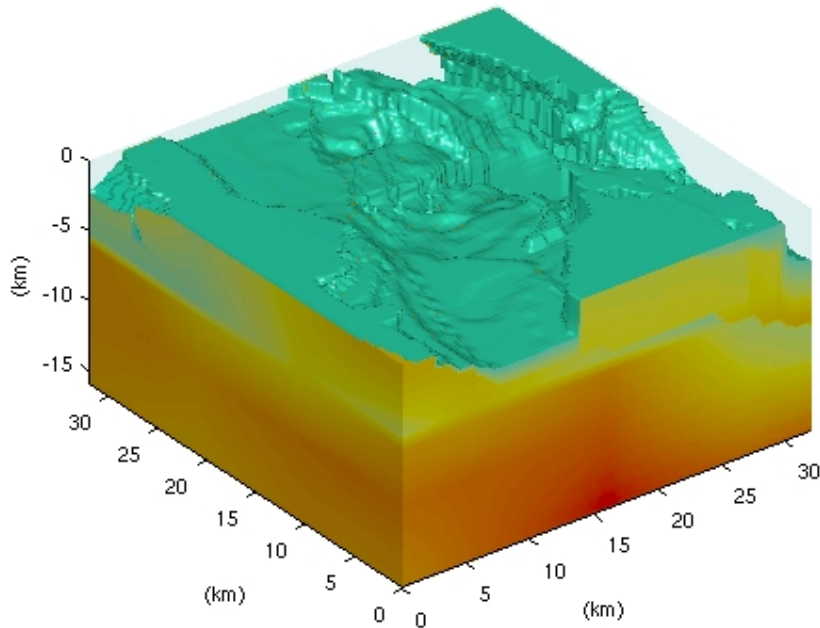


Material inversion: target vs. inverted displacement history at non-receiver location

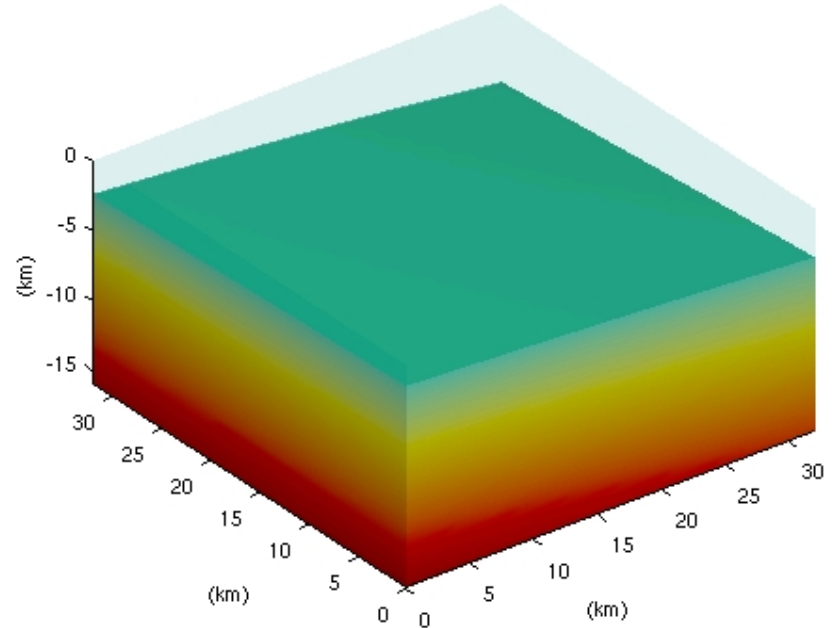


Multiscale inversion: Target vs. inverted isosurfaces, level 1

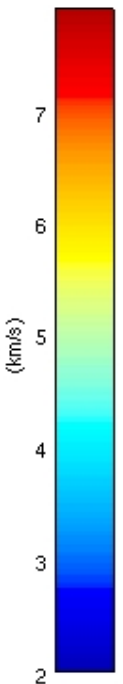
$V_p = 4.50$ km/s isosurface



257x257x257 target model

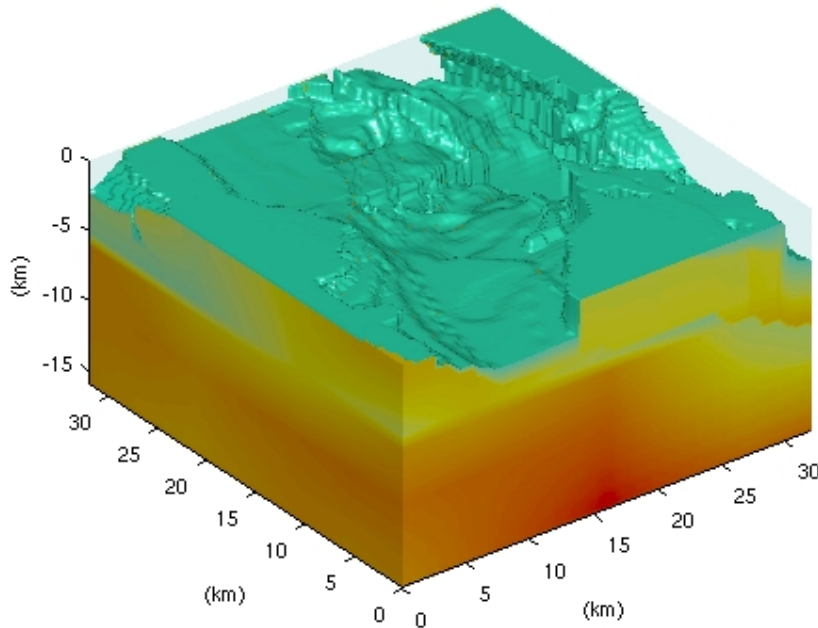


2x2x2 inverted model

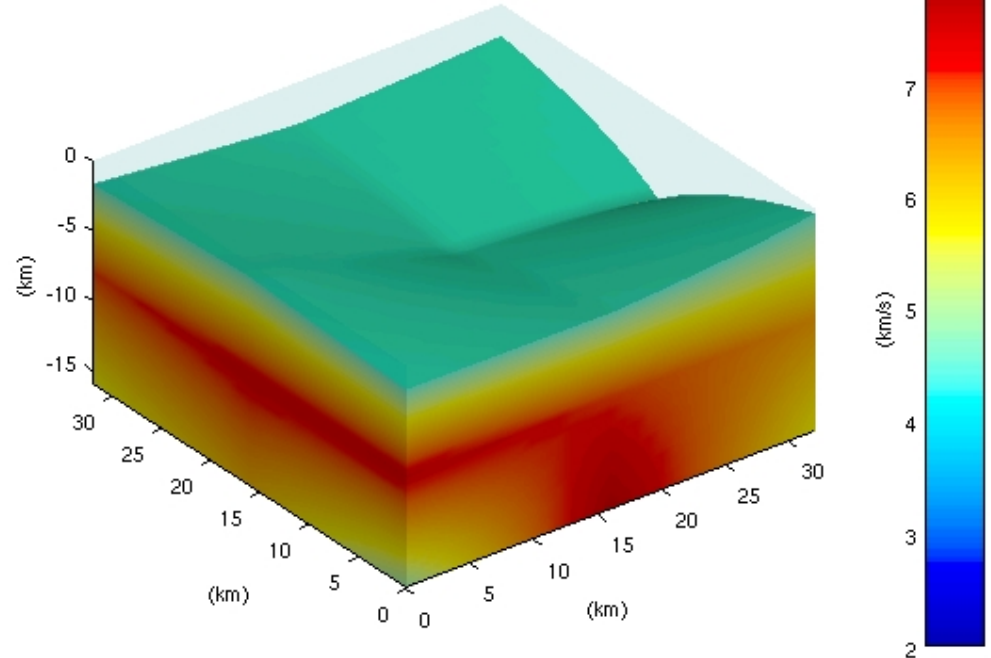


Multiscale inversion: Target vs. inverted isosurfaces, level 2

$V_p = 4.50$ km/s isosurface



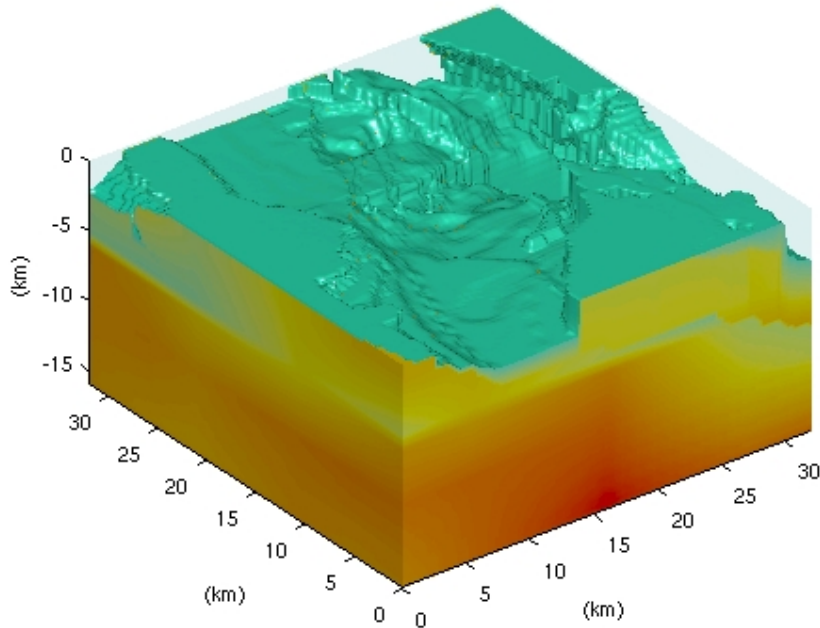
257x257x257 target model



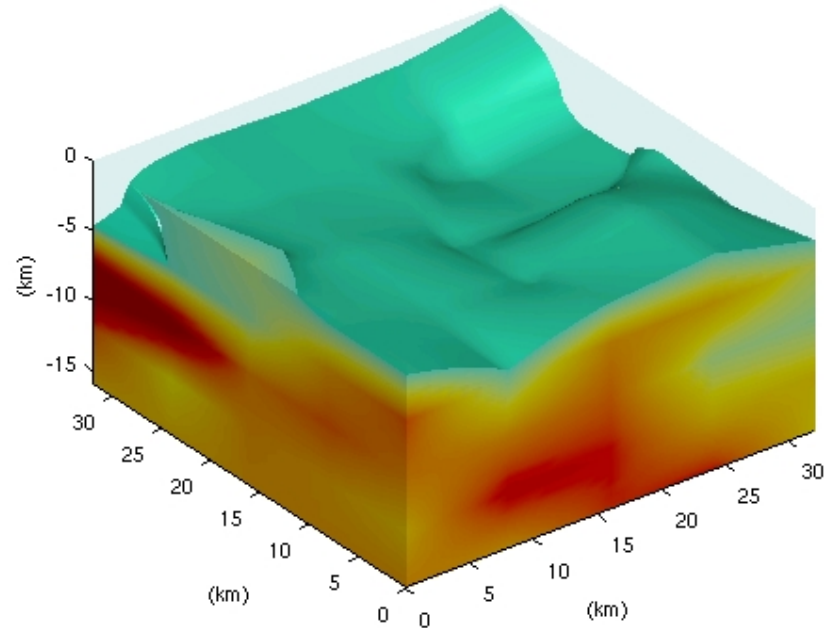
3x3x3 inverted model

Multiscale inversion: Target vs. inverted isosurfaces, level 3

$V_p=4.50$ km/s isosurface



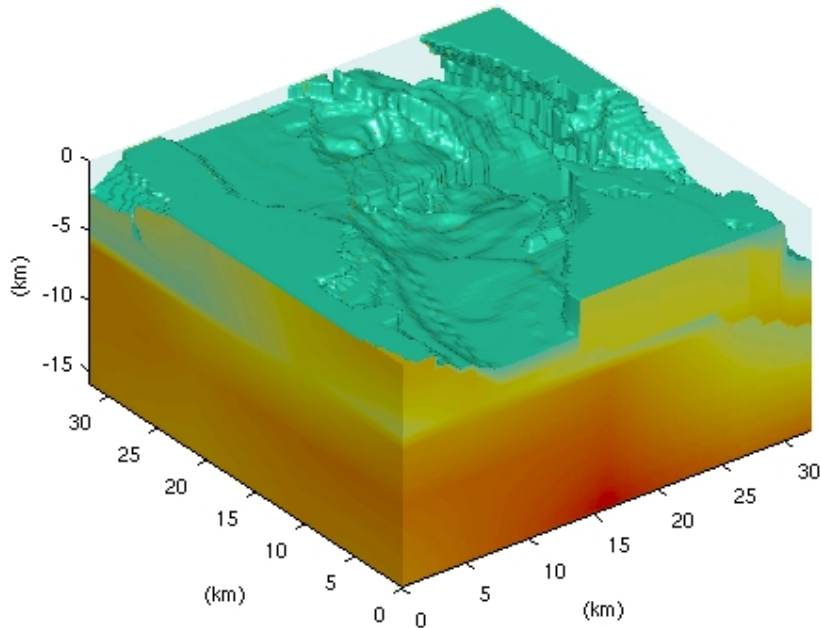
257x257x257 target model



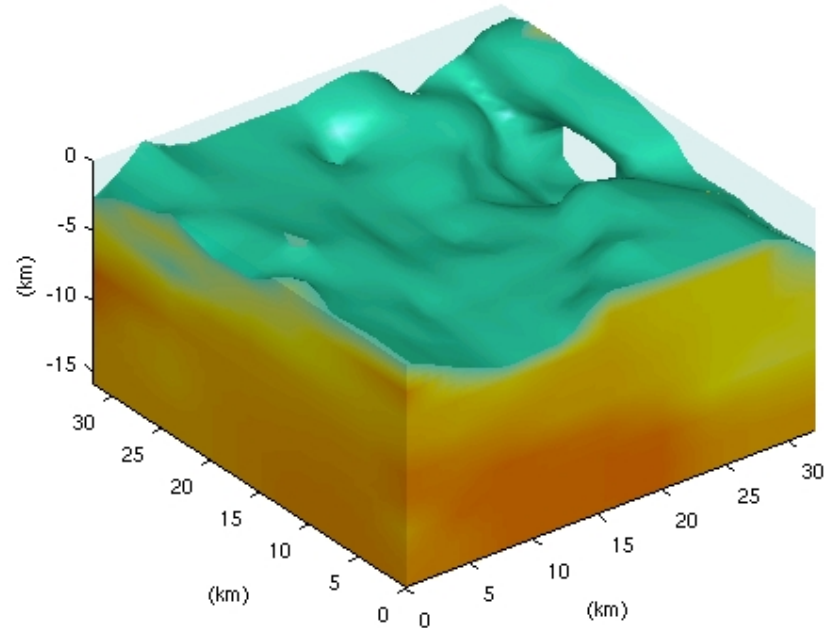
5x5x5 inverted model

Multiscale inversion: Target vs. inverted isosurfaces, level 4

$V_p=4.50$ km/s isosurface



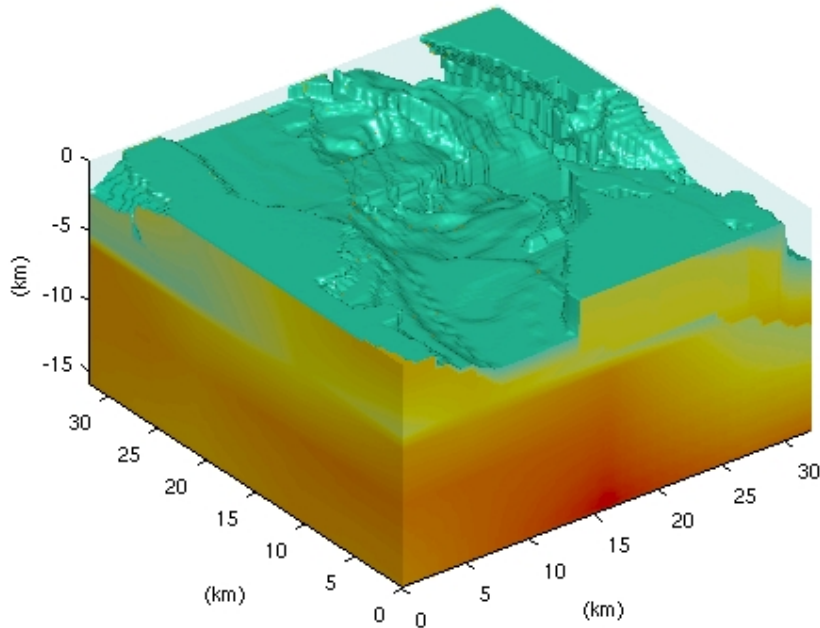
257x257x257 target model



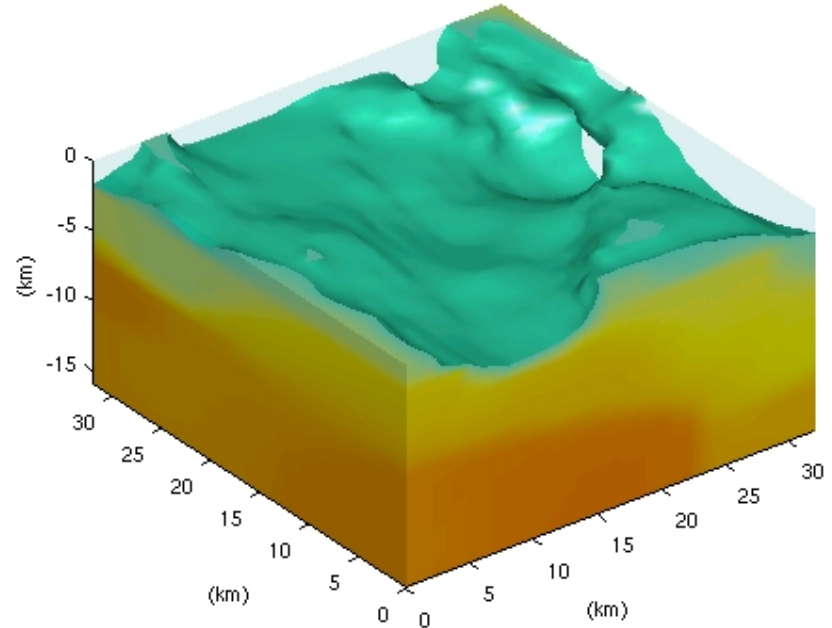
9x9x9 inverted model

Multiscale inversion: Target vs. inverted isosurfaces, level 5

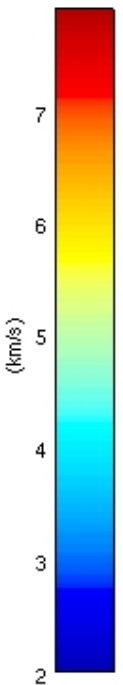
$V_p = 4.50$ km/s isosurface



257x257x257 target model

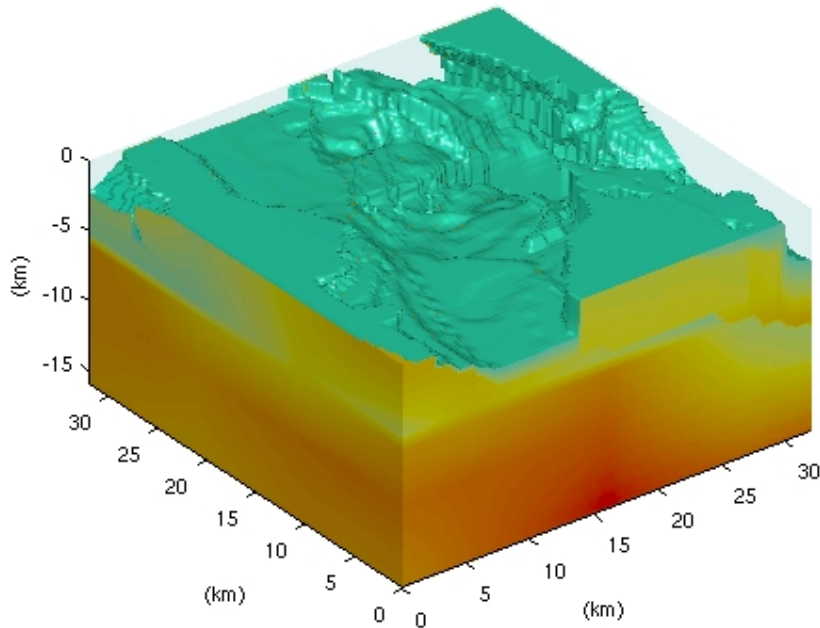


17x17x17 inverted model

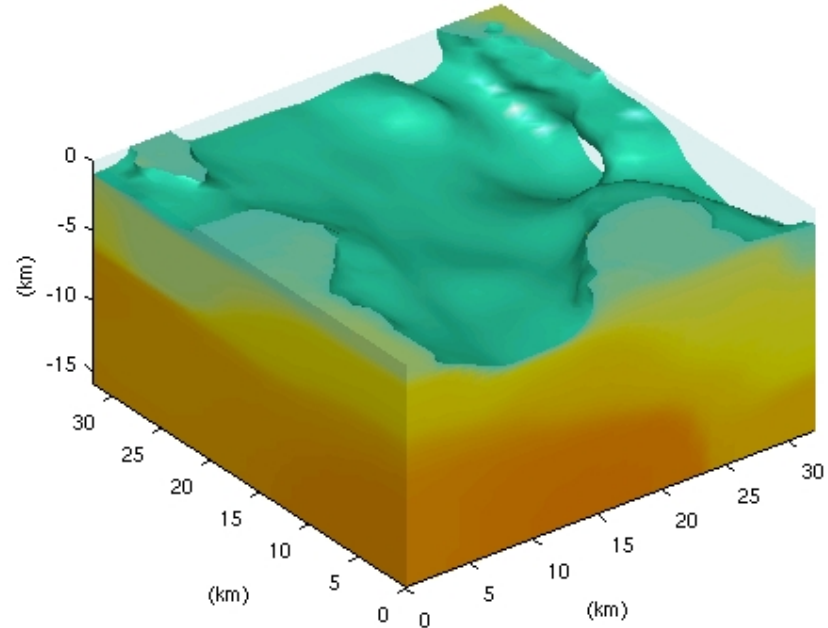


Multiscale inversion: Target vs. inverted isosurfaces, level 6

$V_p = 4.50$ km/s isosurface



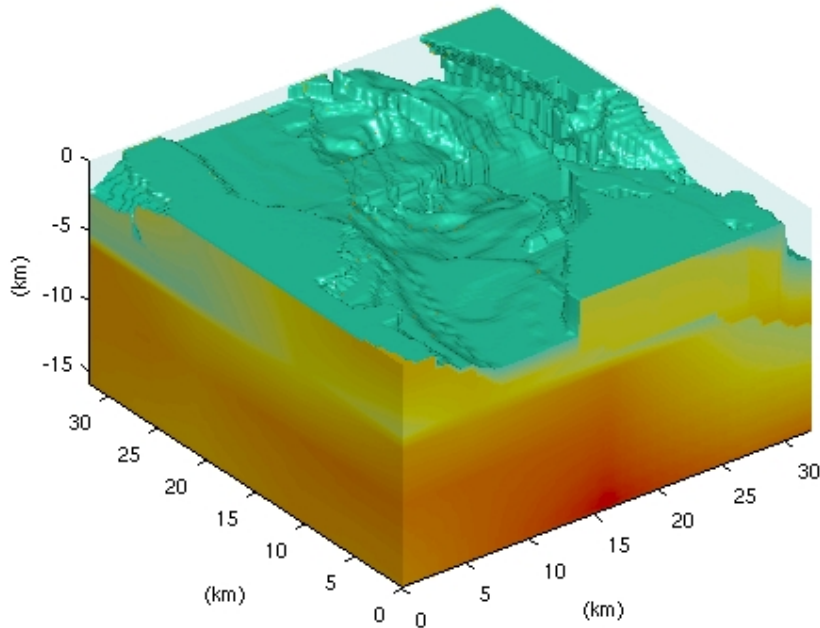
257x257x257 target model



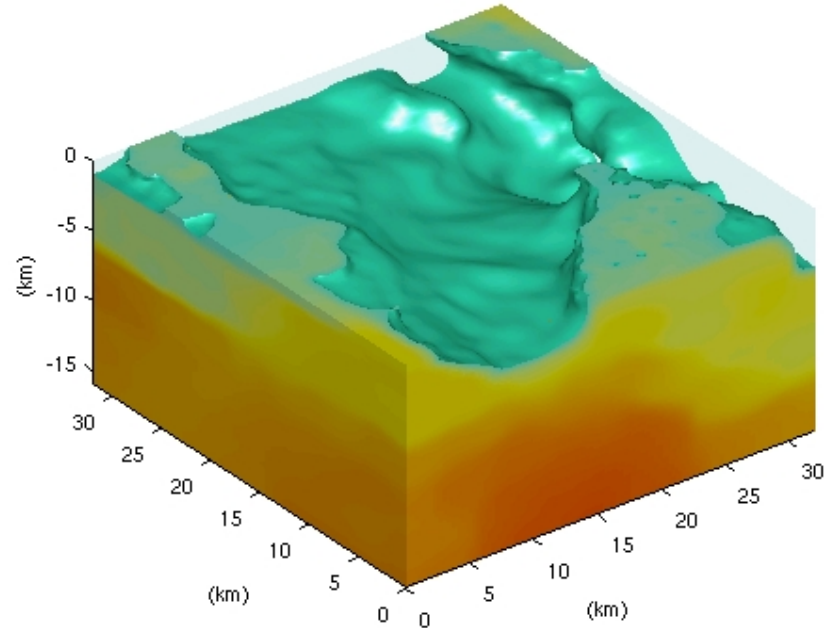
33x33x33 inverted model

Multiscale inversion: Target vs. inverted isosurfaces, level 7

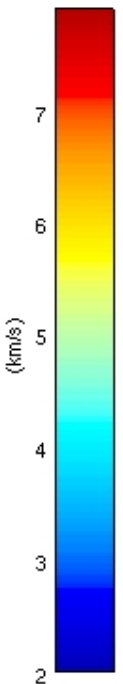
$V_p = 4.50$ km/s isosurface



257x257x257 target model

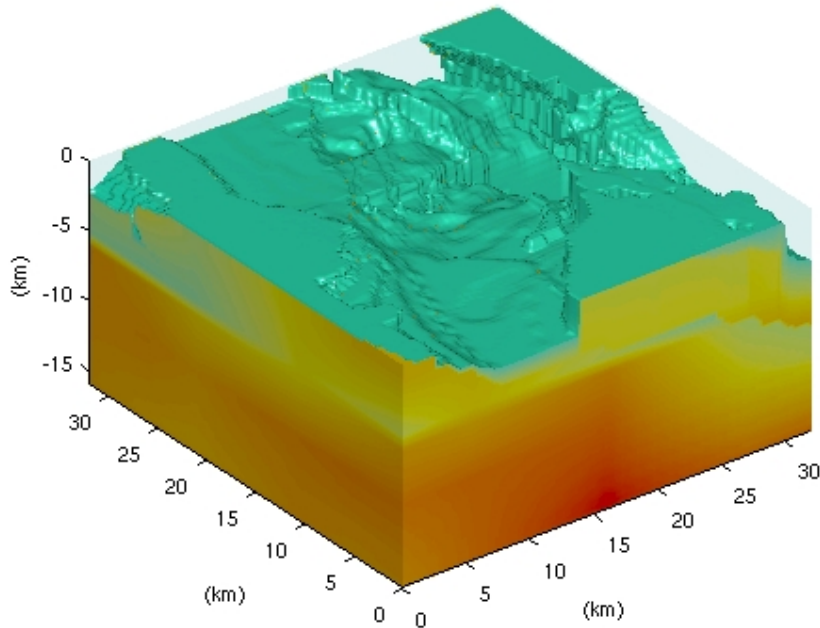


65x65x65 inverted model

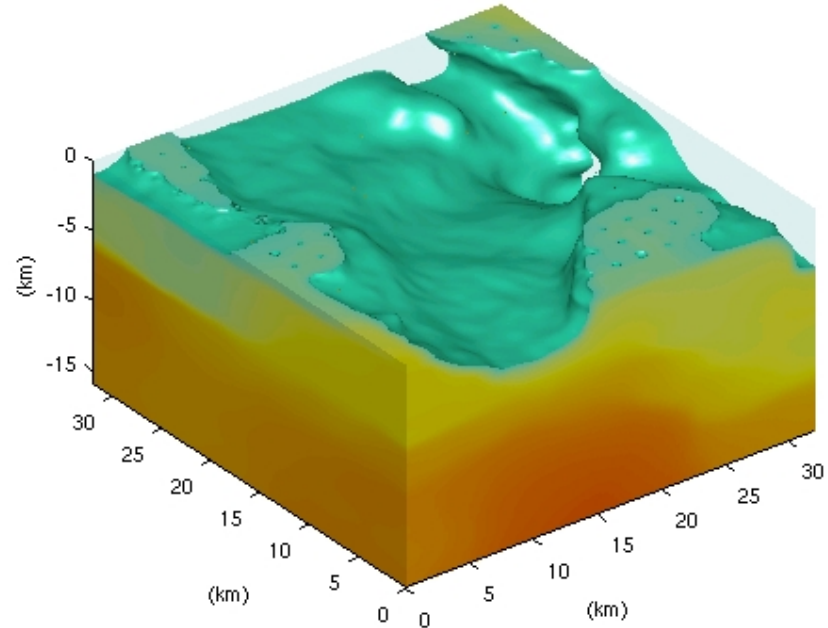


Multiscale inversion: Target vs. inverted isosurfaces, level 8

$V_p = 4.50$ km/s isosurface



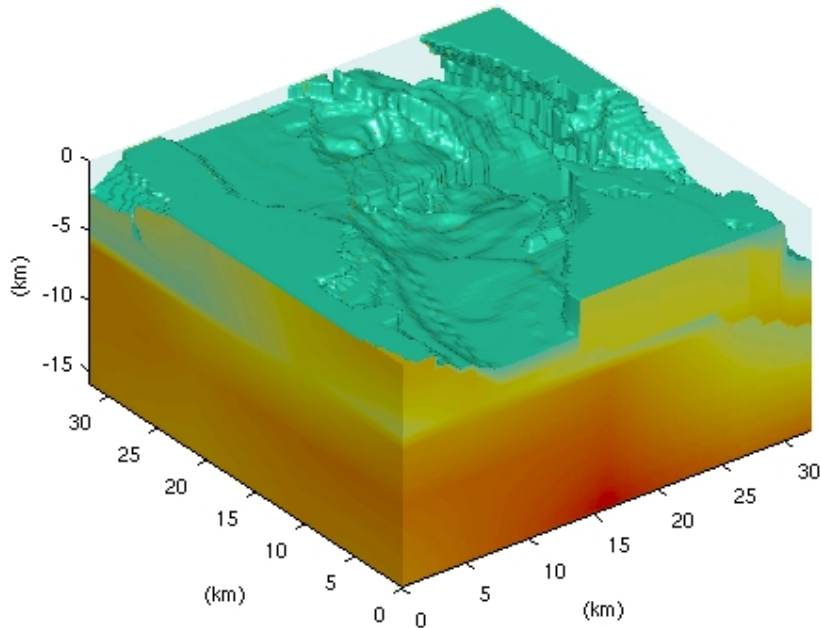
257x257x257 target model



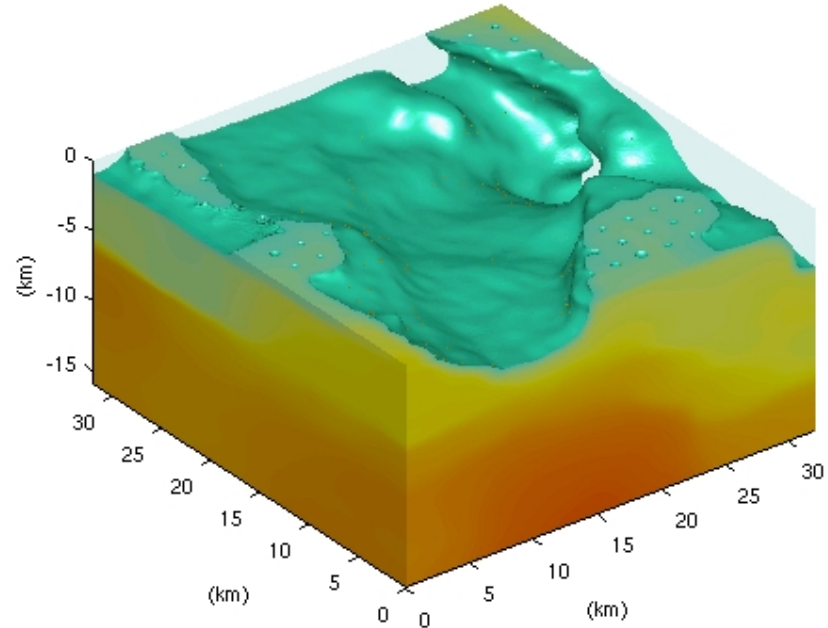
129x129x129 inverted model

Multiscale inversion: Target vs. inverted isosurfaces, level 9

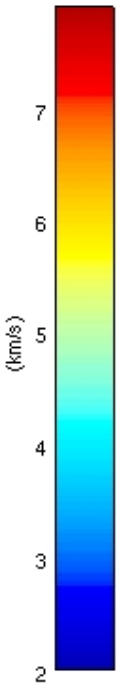
$V_p = 4.50$ km/s isosurface



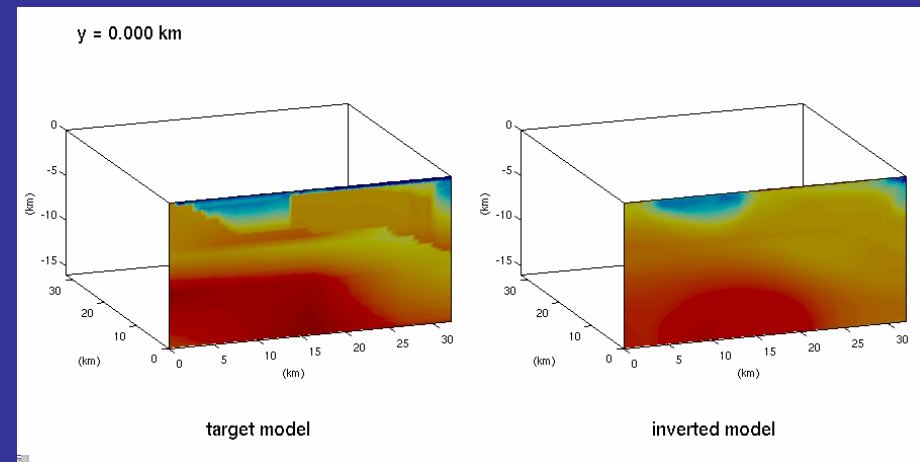
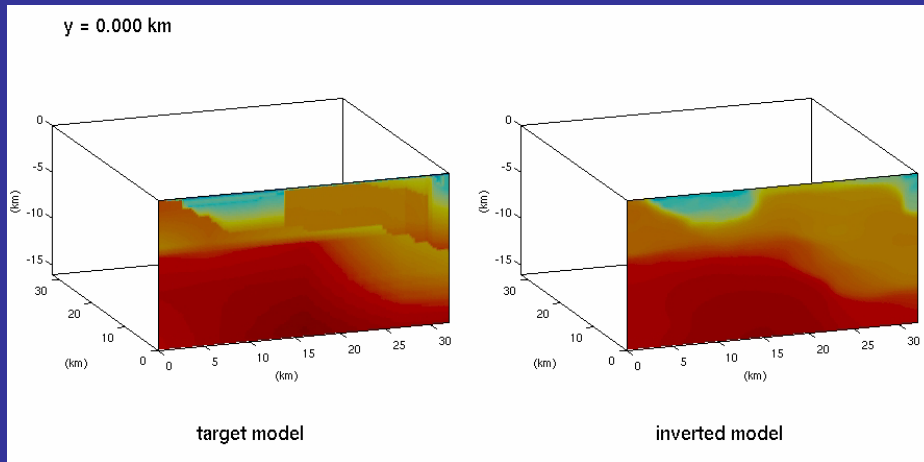
257x257x257 target model



257x257x257 inverted model

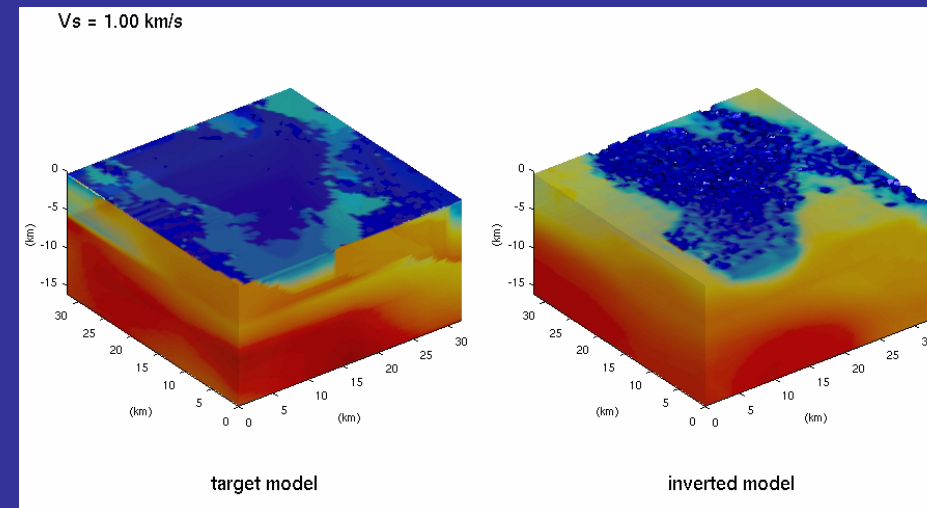
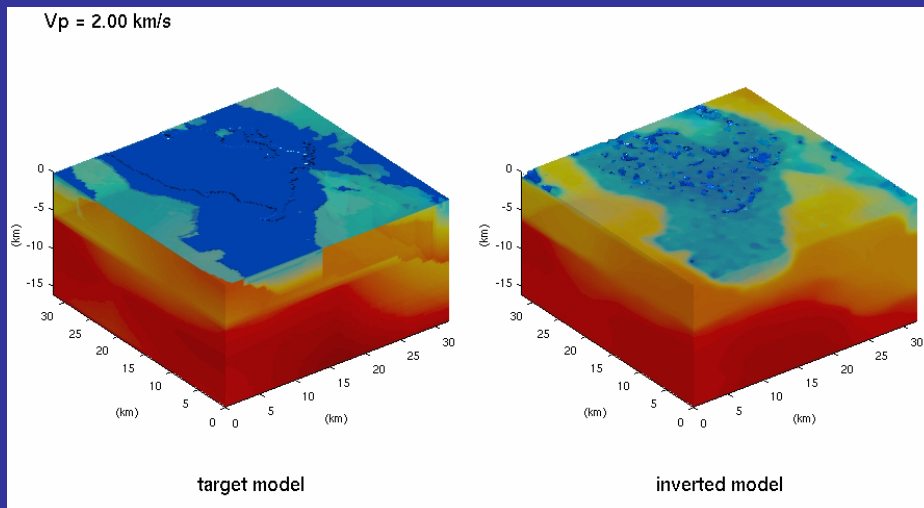


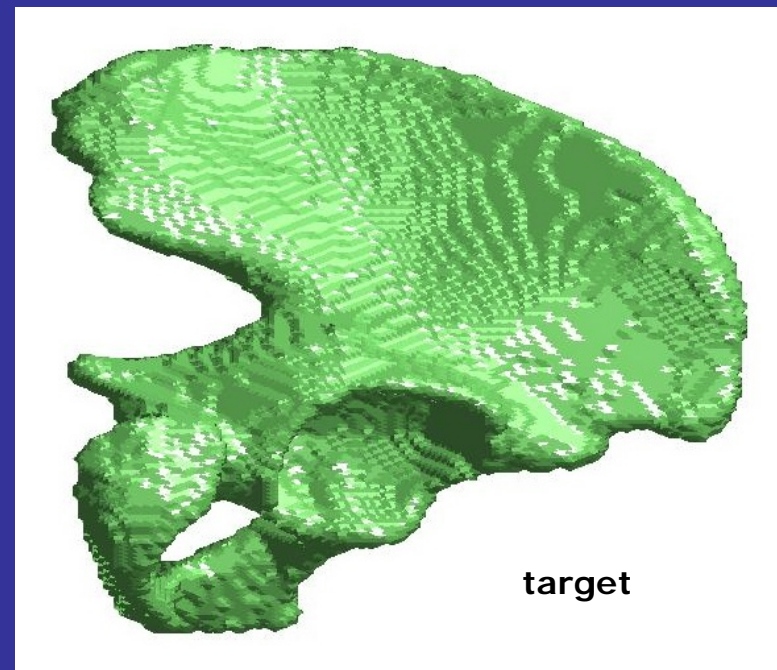
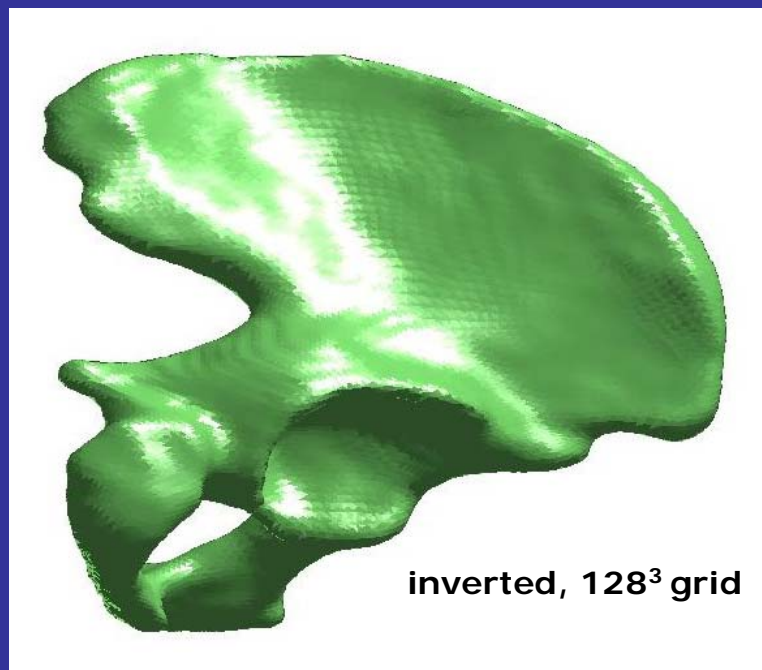
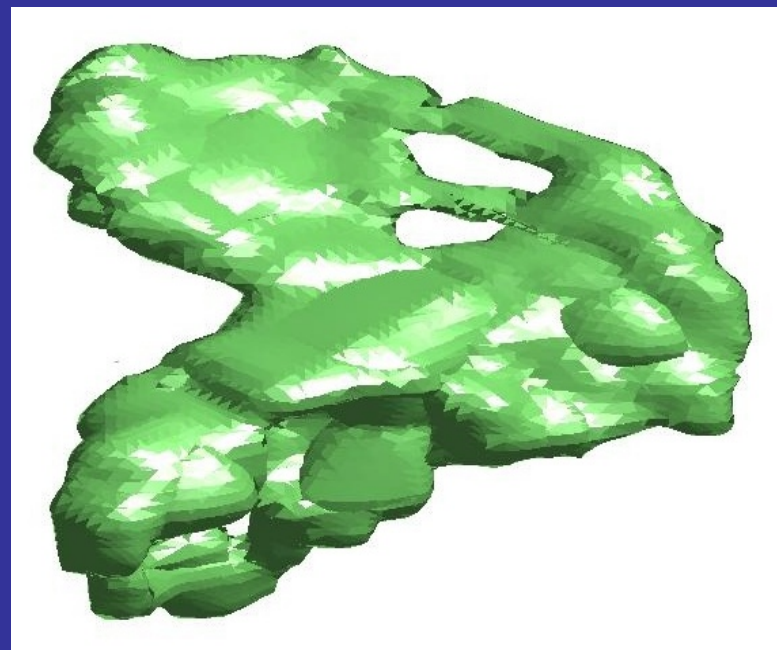
Comparison of target and inverted material models: 3D acoustic and elastic



Acoustic medium, p-wave velocity

Elastic medium, s-wave velocity





Prospects for 3D elastodynamic inversion with observations from multiple events?

- 5x2 earthquake simulations per CG iteration
- 20 CG iterations per Gauss-Newton iteration
- 20 Gauss-Newton iterations
- Inversion costs 4000x a single forward simulation
- Assume petaflops machine has 100,000 x 60 Gflops PEs (i.e. 30x number of 30x faster PEs)
- Inverse problem can easily absorb 30x increase in PEs (assuming network keep up with faster processors; granularity will be 2.5k pts/PE)
- Therefore inverse problem can be solved in $4000 \times 3 / 1000$ hr, or ~ 12 h on 6 Pflops machine (!)
- Important role for Grid computing: loose coupling of tightly-coupled wave propagations

Conclusions:

Forward earthquake modeling

- Octree-based wavelength-adaptive method scaled to billion element simulations
- Excellent parallel scalability and good scalar performance on thousands of processors of commodity-based machine
- Permits us to perform earthquake simulations to frequencies of engineering interest on today's terascale machines
- Critical issue is to address uncertainties in material and source models

Conclusions:

Inverse earthquake modeling

- Multilevel continuation appears to force successive iterates to remain within basin of attraction of global minimum
- Total variation regularization very effective at localizing sharp material interfaces
- Outer and inner iterations are mesh-independent, once nonlinearities have been resolved
- Algorithmic, parallel, and overall scalability follow
- Despite algorithmic and parallel scalability, number of forward/adjoint solutions is large (equivalent to ~ 800 wave propagations for 129^3 grid)
- High-fidelity inverse earthquake modeling w/multiple earthquake sources is a petaflops-level challenge

Ongoing and future work

- Incorporation of parallel adaptive octree grids
- Hessian preconditioner and nonlinear solver improvements necessary
- Regularization parameter selection for real data
- Bound inequalities for material properties
- Treatment of correlated variables
- Estimation of uncertainty in parameters
- Incorporation of prior (SCEC community velocity model)
- Inversion for attenuation parameters
- Inversion for fault parameters
- Inversion for fault location (shape optimization problem)

Acknowledgments

- **Quake Project: Earthquake Modeling in Large Basins**
(www.cs.cmu.edu/~quake)
 - NSF/KDI CMS-9980063, NSF/ITR ATM-0326449
 - Other Quake group members: S. Day and H. Magistrale (SDSU), J. Shewchuk (Berkeley)
- **TOPS Center: Terascale Optimal PDE Simulations**
(www.tops-scidac.org)
 - Supported under DOE SciDAC/ISIC program
 - Collaboration with LLNL, ANL, LBNL + 8 universities
- **Caliente Project: Dynamic Inversion and Control**
(www.cs.cmu.edu/~caliente)
 - NSF/ITR ACI-0121667
 - Collaborators: other L. Biegler (CMU), D. Keyes (ODU), M. Heinkenschloss (Rice), R. Bartlett, K. Long, and B. van Bloemen Waanders (Sandia), D. Young (Boeing), F. Fendell (TRW)
- **Southern California Earthquake Center (SCEC)**
- **Computer Science Research Institute, Sandia**
- **Institute for Scientific Computing Research, Lawrence Livermore**
- Special thanks to staff at **Pittsburgh Supercomputing Center** (Computing time supported by grants ASC-010025P, ASC-010036P, MCA01S002P, BCS020001P, MCA04N026P)
- Special thanks to **PETSc group** at Argonne (S. Balay, K. Buschelman, W. Gropp, D. Kaushik, M. Knepley, L. McInnes, B. Smith, H. Zhang)