Large-Scale Parallel Nonlinear Optimization for High-Resolution 3D-Seismic Imaging

Olaf Schenk
Computer Science Department
University of Basel

Tarje Nissen-Meyer
Institute of Geophysics
ETH Zurich

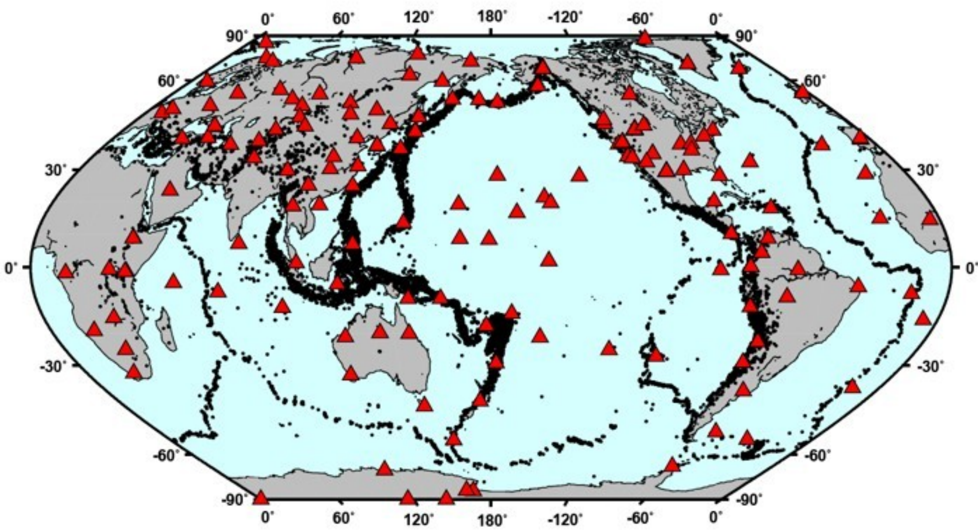
HP2C Kick-off Meeting, USI Lugano
March 16, 2010

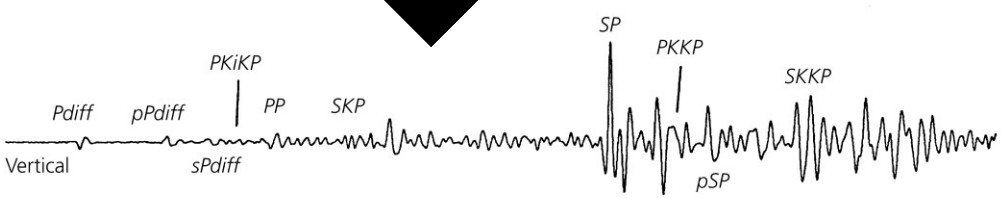
HP2C Project Partners

- **Large-Scale Parallel Nonlinear Optimization for 3D-Seismic Imaging**
  - ETHZ: Institute of Geophysics & Swiss Seismological Service
    - D. Giardini, L. Boschi, T. Nissen-Meyer (**Computational Seismology**)
  - University of Basel: Math&Computer Science Department
    - M. Grote (Math., **Computational Wave Propagation**)
    - O. Schenk, H. Burkhart (CS, **Computational Optimization** and **HPC)**
  - Academic and Industrial Cooperation Partners:
    - A. Wächter (IBM Watson, 6-month Sabbatical, **Computational Optimization)**
    - J. Tromp (Princeton 3-month Sabbatical, **Computational Seismology)**
    - NVIDIA Research, summer intern + consulting
- **HP2C Funding:**
  - 2 PostDocs & 1 PhD (U Basel), 1 PostDoc (ETHZ)

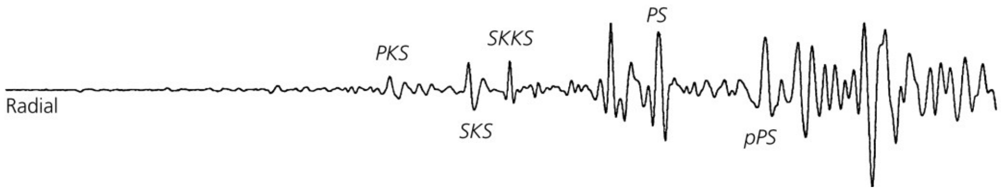
Seismic tomography: Mapping the Earth's structure and dynamics

Seismicity and seismometers

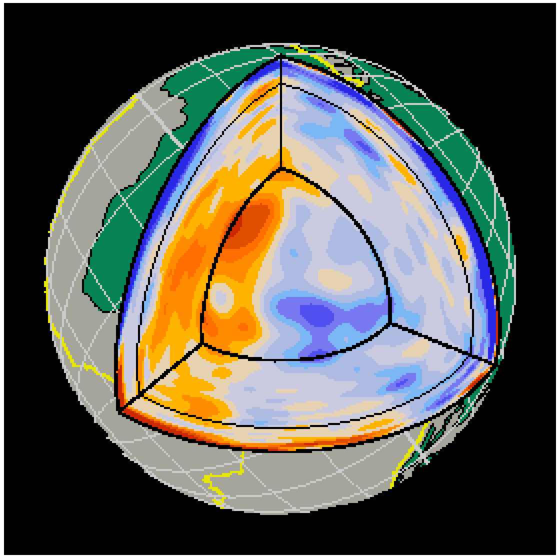




Seismograms: traveltimes & amplitudes



Earth models for seismic (compressional/shear) velocities



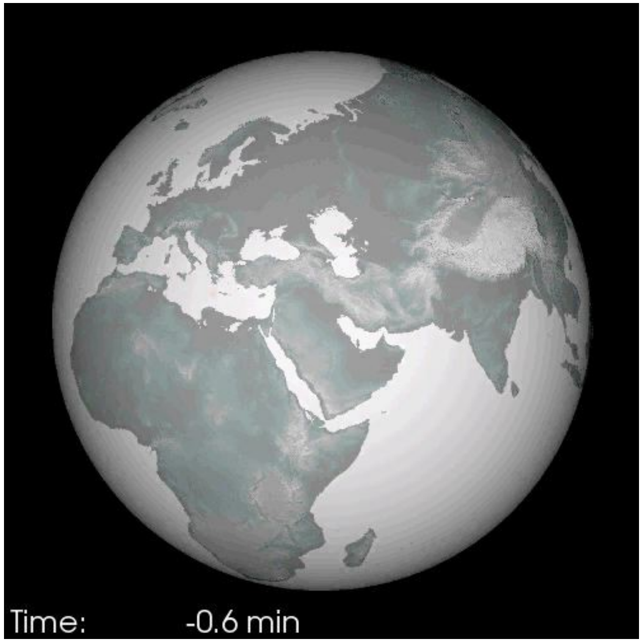
Left to right:
Solving the forward *and* inverse problem

The forward problem

Find ground displacement d_0 upon assumed earthquake source characteristics s and background earth model m_0:

$$F: (m_0, s_0) \rightarrow d_0$$

Solution to the (linear) elastodynamic wave equation F



Simplified model:
- Normal-mode summation
- Reflectivity methods

Complex model:
- Finite-difference methods
- Discontinuous Galerkin methods
- Spectral-element methods

Peter, Boschi, Woodhouse (2009)

The inverse problem

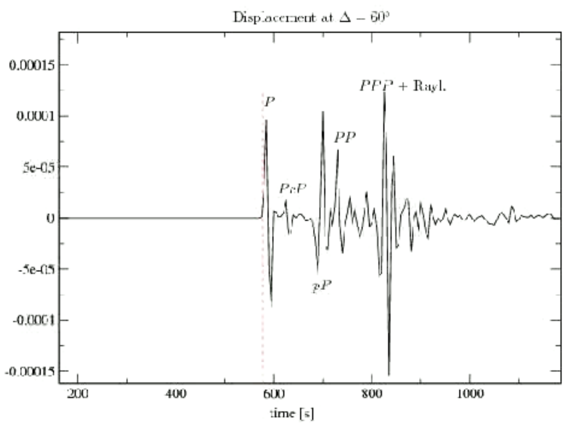
Find earth model *m* upon recorded seismic data *d*, assumed initial model m_0 and source characteristics s_0:

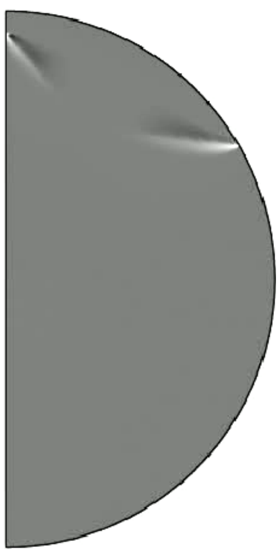
$$G^{-1}: (d, m_0, s_0, d_0) \rightarrow m$$

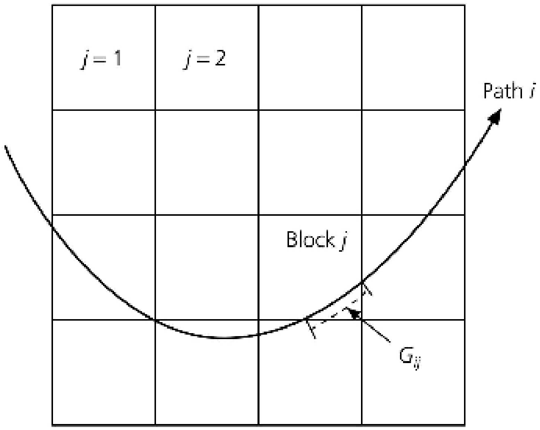
Nonlinear, overdetermined, non-unique, ill-posed, unverifiable

Calculation of *G* ($\partial d / \partial m$):

- Geometrical ray theory
- Wave-based Fréchet derivatives



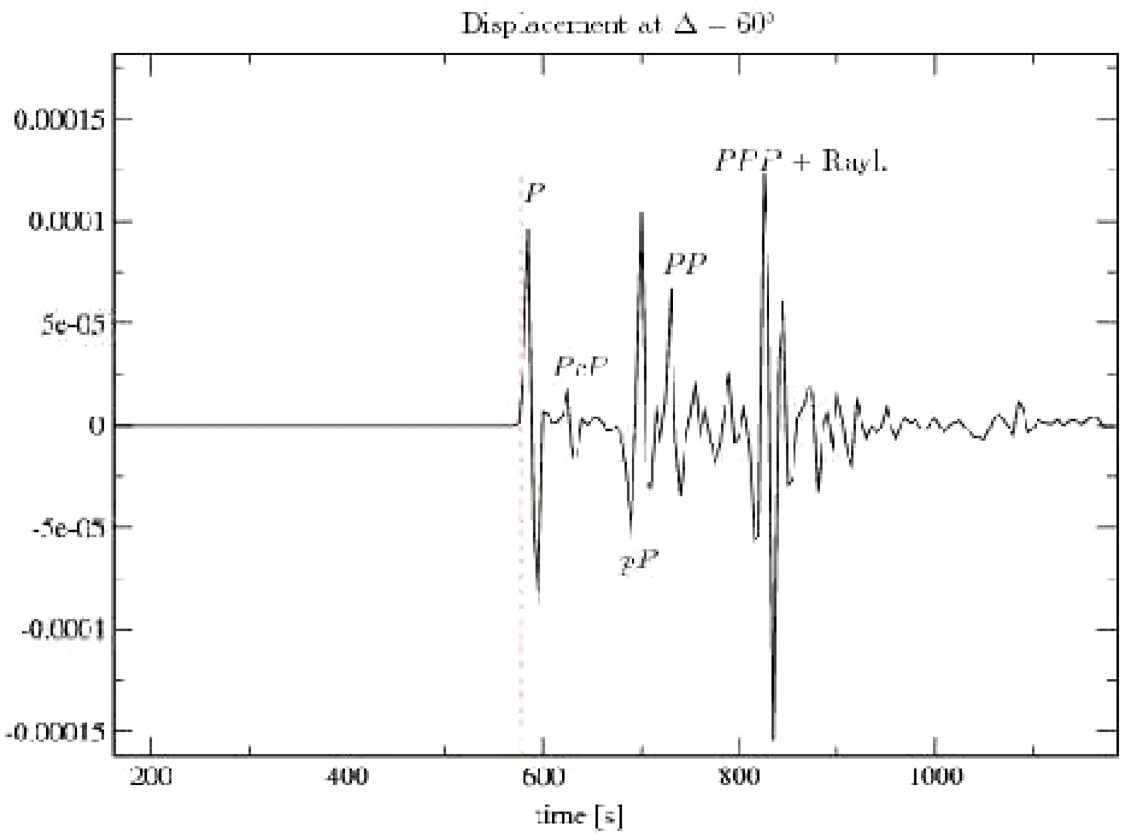


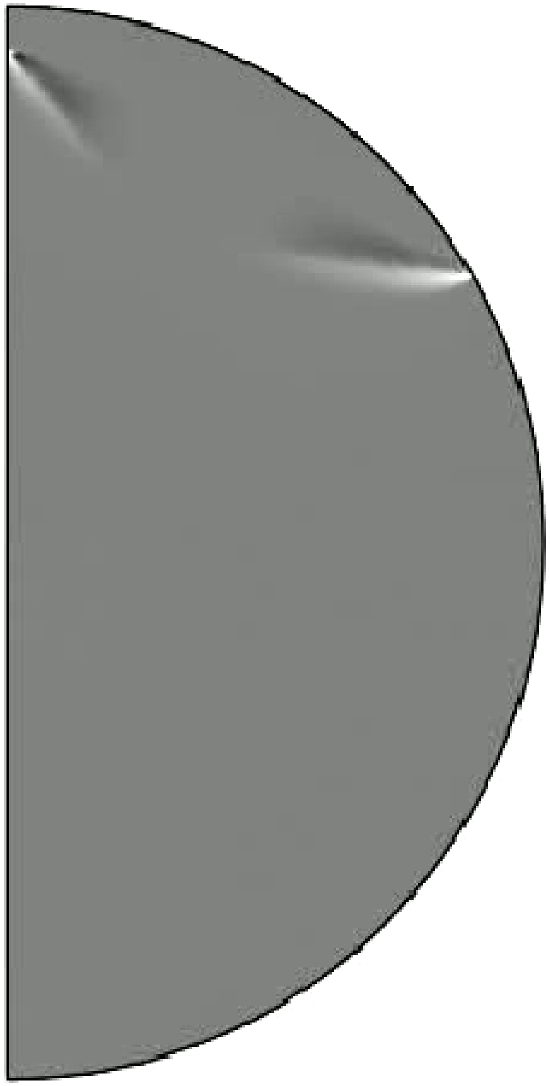


Inversion approaches:

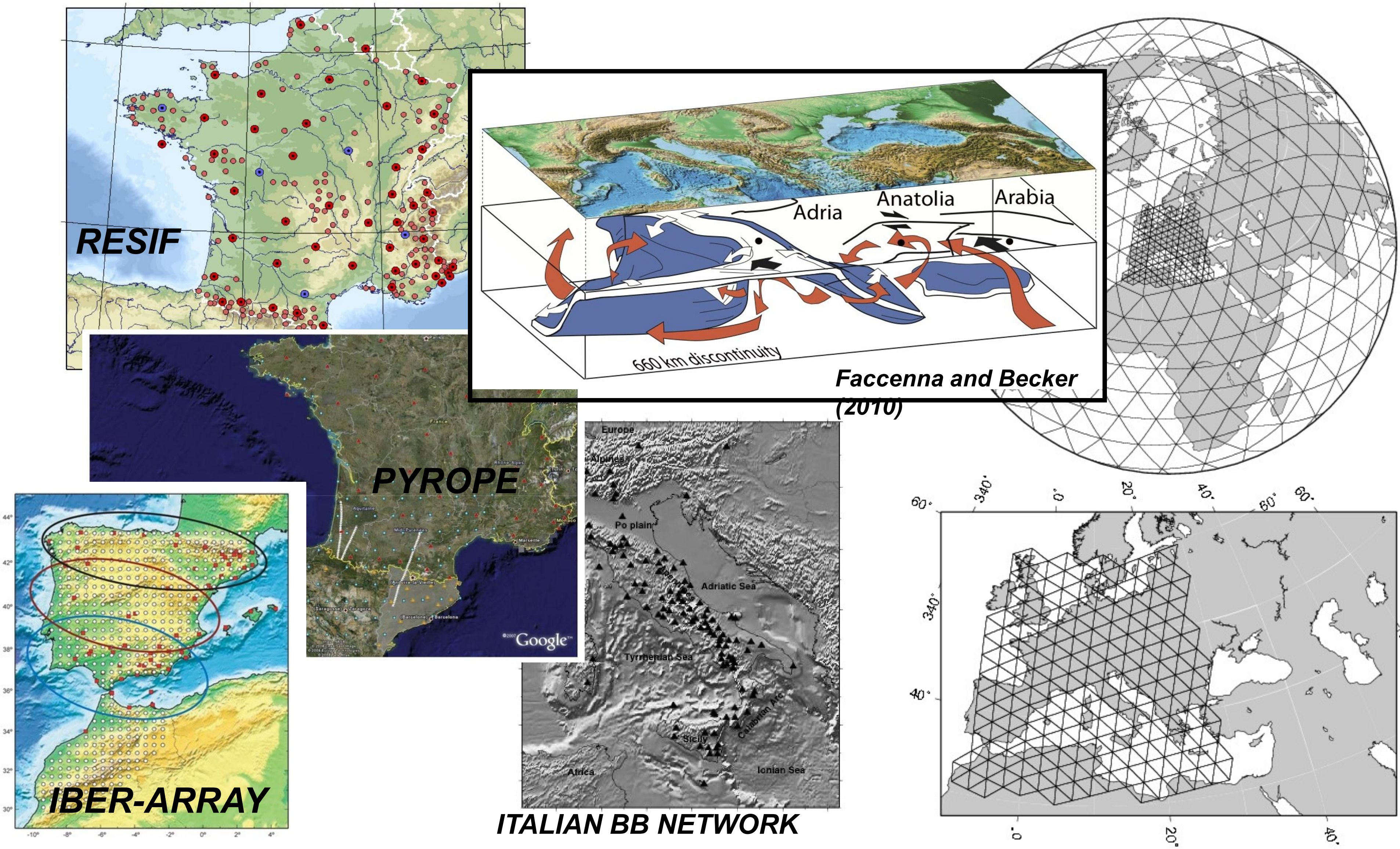
- linearization: $d_i = G_{ij}\, m_j$
- gradient techniques
- probabilistic methods

The inverse problem



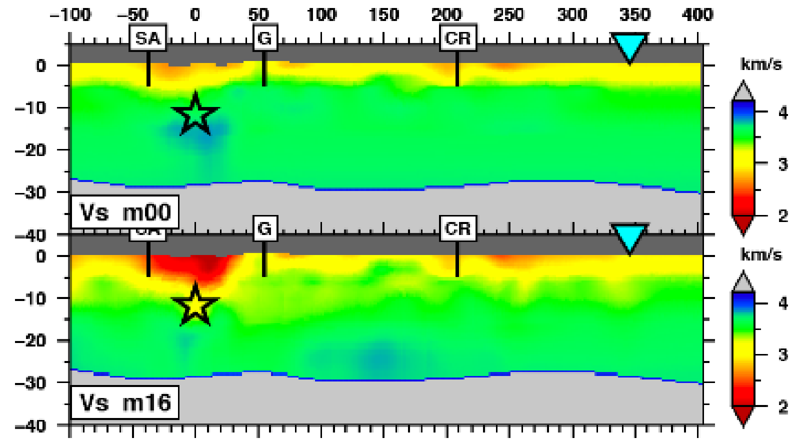


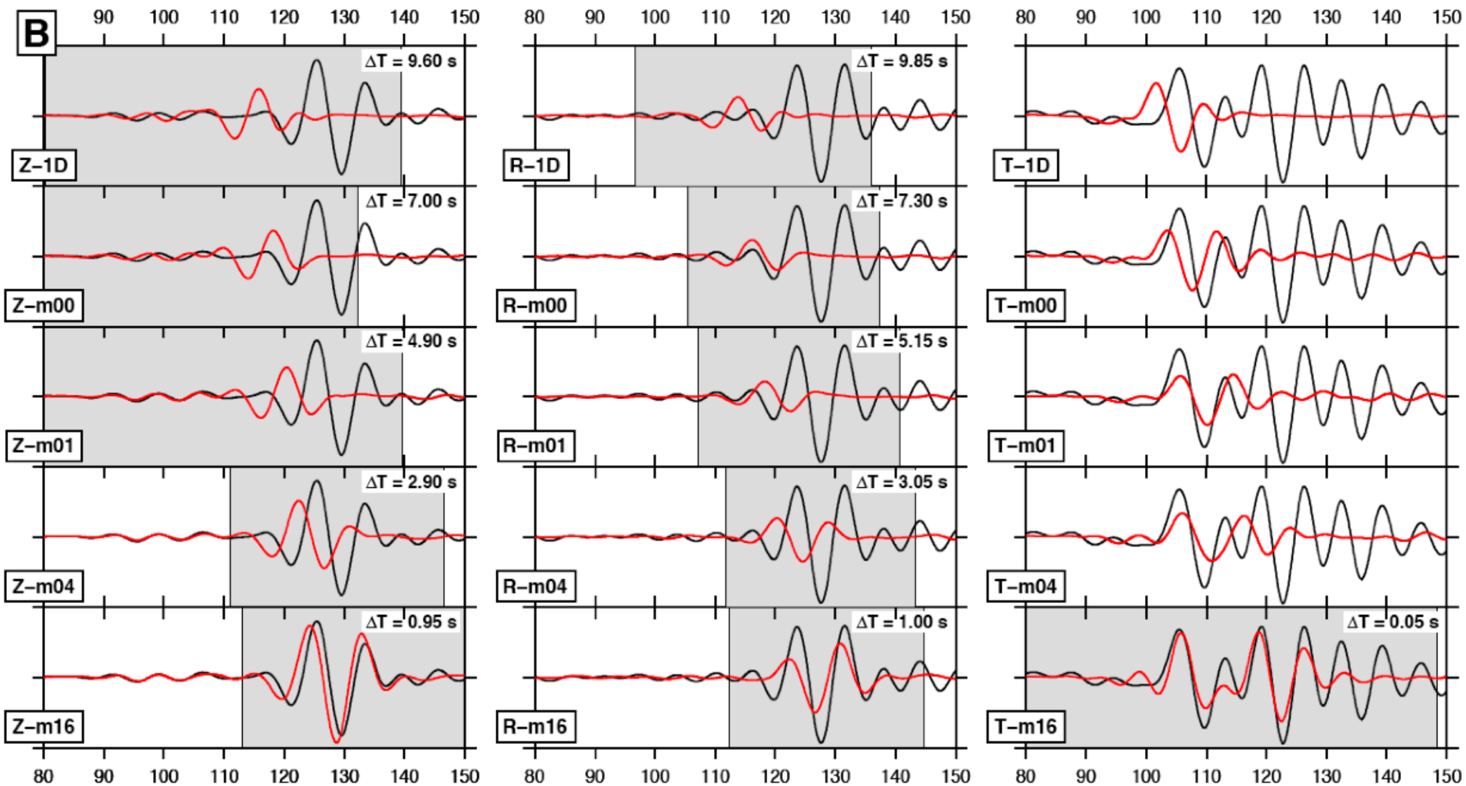
The Mediterranean basin as a natural laboratory



Adjoint tomography: Southern California

143 earthquakes
6,864 simulations
168 cores per simulation
45 minutes of wall-clock time per simulation
864,864 core hours





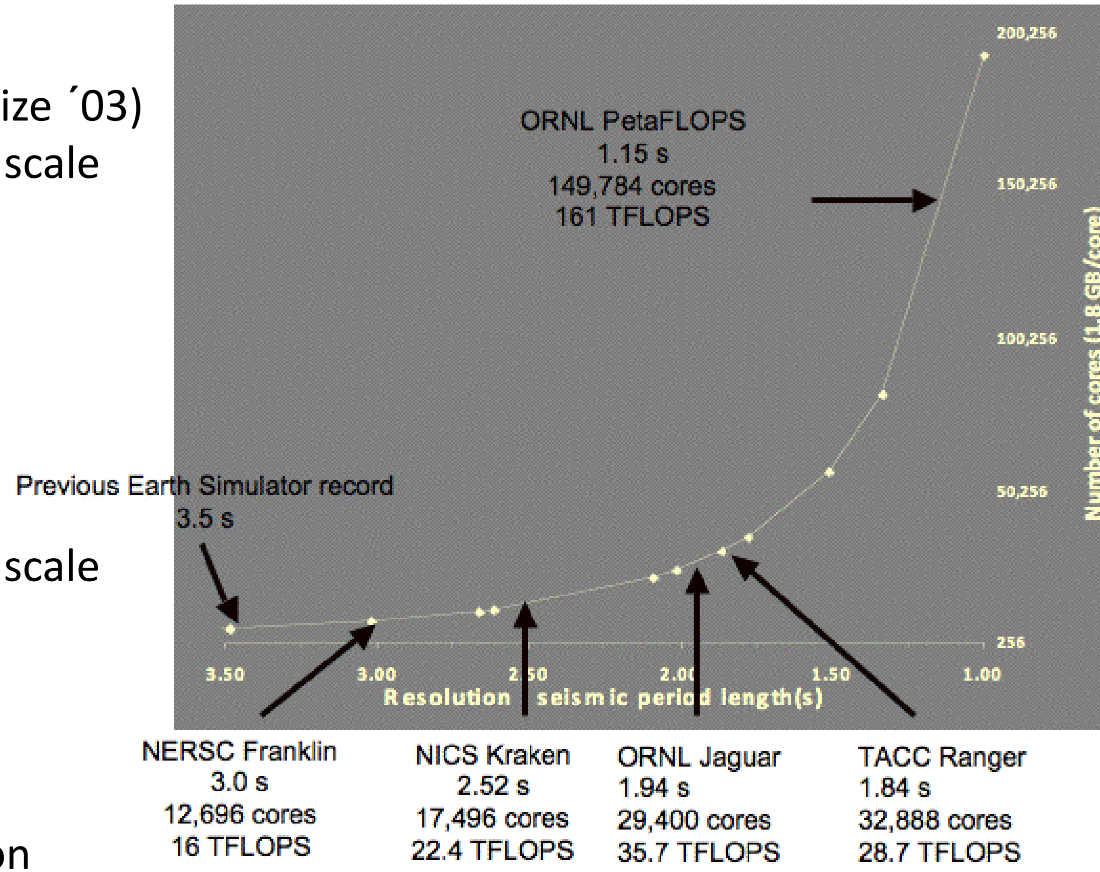
Computational tasks

Global spectral-element method

- highly efficient, explicit (Gordon Bell prize ´03)
- solves all relevant physics at the global scale
- open-source (geodynamics.org)
- typical global simulation: 10 hours on 400 cores

Local discontinuous Galerkin method

- highly flexible (tetrahedral elements)
- solves all relevant physics at the global scale
- computationally expensive, not open-access yet



Adjoint tomography

- 3 simulations per earthquake & iteration
- cost independent of # seismometers
- 16 iterations needed in Southern California
- No Hessian preconditioning available

Forward Problem in 3D Seismic Imaging

Second-order Wave Equation:

$$\frac{\partial^2 u}{\partial t^2} - \nabla \cdot (c^2 \nabla u) = f \quad \text{in} \quad \Omega \times [0, T],$$

$$c^2 \nabla u \cdot n = 0 \quad \text{on} \quad \Gamma \times [0, T],$$

$$u = 0 \quad \text{at} \quad t = 0,$$

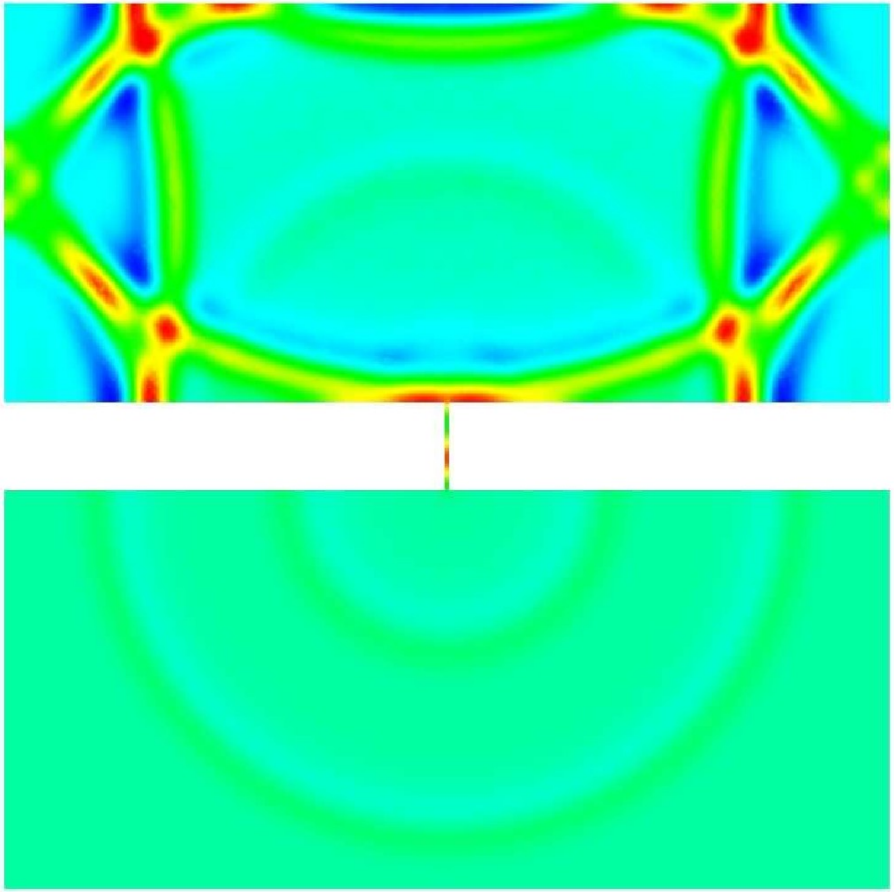
$$\dot{u} = 0 \quad \text{at} \quad t = 0,$$

Discontinuous Galerkin Methods (M. Grote, A. Schneebeli, D.Schötzau, 2007)

- Flexibility in mesh-design (tetrahedral elements)
- easily handles varying polynomial degree (hp-adaptivity)
- (Block-)Diagonal mass matrix & local time stepping → **truly explicit&parallel**
- GPU Code **HEDGE** (A. Klöckner, JCP 2009)

Discontinuous Galerkin (DG) method

Numerical Experiments in 2D (4th order in time, M. Grote, D. Schötzau, 2009)



h^{coarse}=0.0125, h^{fine}=7.62.10^{-5} = h^{coarse} /170

Seismic Inversion

minimize $\frac{1}{2} \sum_{j=1}^{N_r} \int_0^T \int_\Omega [u^* - u]^2 \, \delta(x - x_j) \, d\Omega \, dt + \beta R(q)$

s. t. $\frac{\partial^2 u}{\partial t^2} - \nabla \cdot c^2 \nabla u = f \quad \text{in} \quad \Omega \times [0, T],$

$$c^2 \nabla u \cdot n = 0 \quad \text{on} \quad \Gamma \times [0, T],$$

$$u = 0 \quad \text{at} \quad t = 0,$$

$$\dot{u} = 0 \quad \text{at} \quad t = 0,$$

Large-Scale Optimization

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
$$s.t. \quad c(x) = 0$$
$$d(x) \leq 0$$

Nonlinear Optimization	Small-Scale (10^5 variables)	Large-Scale (10^6 to 10^8 variables)
Multicores (<16 cores)	- Randomized metaheuristics (Evolutionary Algorithms, Simulated Annealing, Ant Colony) - Derivative-free optimization - Interior-point optimization	- **Interior-point optimization** **+ Fast convergence** - Need derivate information **(Jacobian, Hessian matrices)** - Matrices are indefinite and highly ill-conditioned
Manycores (~1‘000 cores)	———	- **Interior-point optimization**

Large-Scale Optimization

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
$$s.t. \quad c(x) = 0$$
$$d(x) \leq 0$$

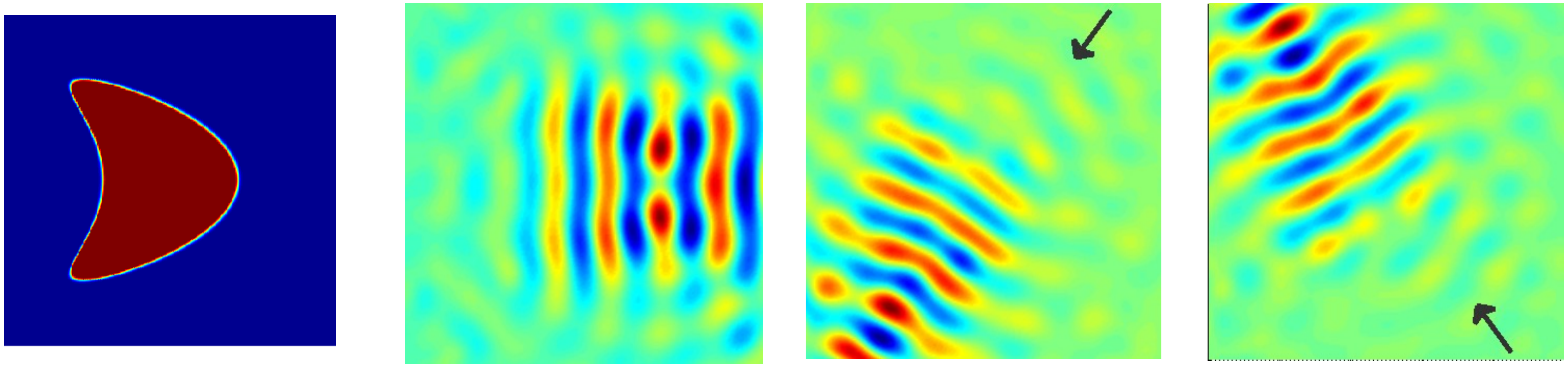
Second-Order Wave Equation

Additional information

- Nonlinear, overdetermined, non-unique, ill-posed, nonconvex and LARGE-SCALE
- Interior point methods are good candidates for very large problems.
- **Optimization Algorithm:**
  - Primal-dual interior point method (A. Wächter, ACM TOMS, 2006)
  - Line-search filter method using **inexact steps** to ensure **global convergence** (Curtis, Wächter, S., SISC 2010)
  - **Massively parallel implementation** on up to 1'000 cores (A. Wächter, S., 2010)
  - Parallel **inexact** sparse linear algebra kernel (A. Sameh, S., A. Wächter, 2010)

Why Optimization with Inequalities?

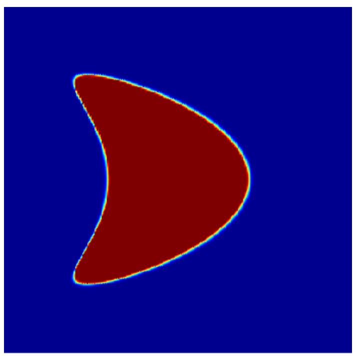
Inverse Problems: 2D Time-harmonic Wave Equation



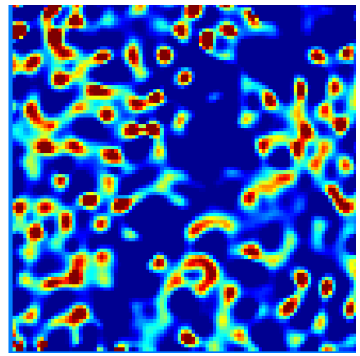
Source: SNF Project Multiscale Analysis and Simulation for Waves in Strongly Heterogeneous Media (Grote, Schenk, 2010)

Why Optimization with Inequalities?

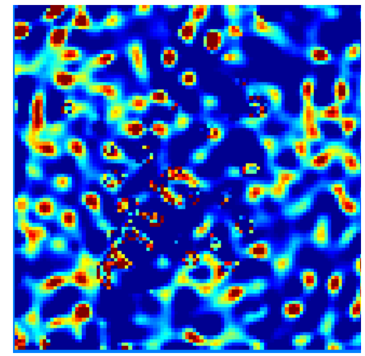
Inverse Problems: 2D Time-harmonic Wave Equation



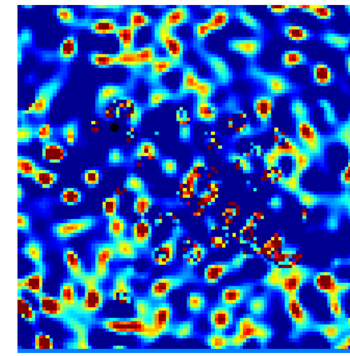
$$\min_{x \in \mathbb{R}^n} \quad f(x) \qquad s.t. \quad c(x) = 0$$



10 it.

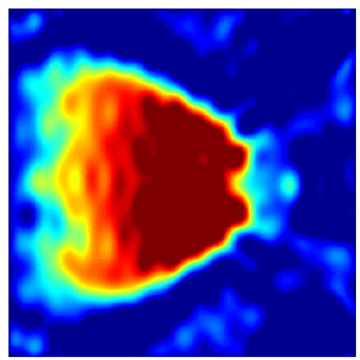


500 it.

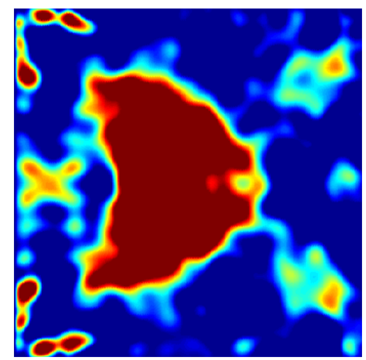


1000 it.

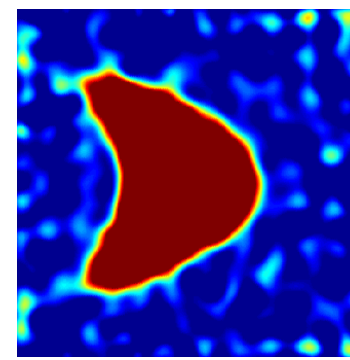
$$\min_{x \in \mathbb{R}^n} \quad f(x) \qquad s.t. \quad c(x) = 0, \quad d(x) \leq 0$$



3 it.



6 it.



11 it.

Source: SNF Project Multiscale Analysis and Simulation for Waves in Strongly Heterogeneous Media (Grote, Schenk, 2010)

The KKT conditions for the Second-Order Wave Equation

- Lagrangian functional

$$\mathcal{L}(u,\lambda,q) = \sum_{j=1}^{N_r} \int_0^T \int_\Omega [u^* - u]^2\, \delta(x - x_j)\, d\Omega\, dt + \beta R(q) + \int_0^T \int_\Omega \lambda\left(\frac{\partial^2 u}{\partial t^2} - \nabla \cdot q \nabla u - f\right) d\Omega\, dt,$$

- First order optimality condition

$$A(q)u = f \quad \text{in} \quad \Omega \times [0,T],$$

$$A^*(q)\lambda = \sum_{j=1}^{N_r} [u^* - u]$$

$$\int_0^T \nabla u \cdot \nabla \lambda\, dt + \beta\, G($$

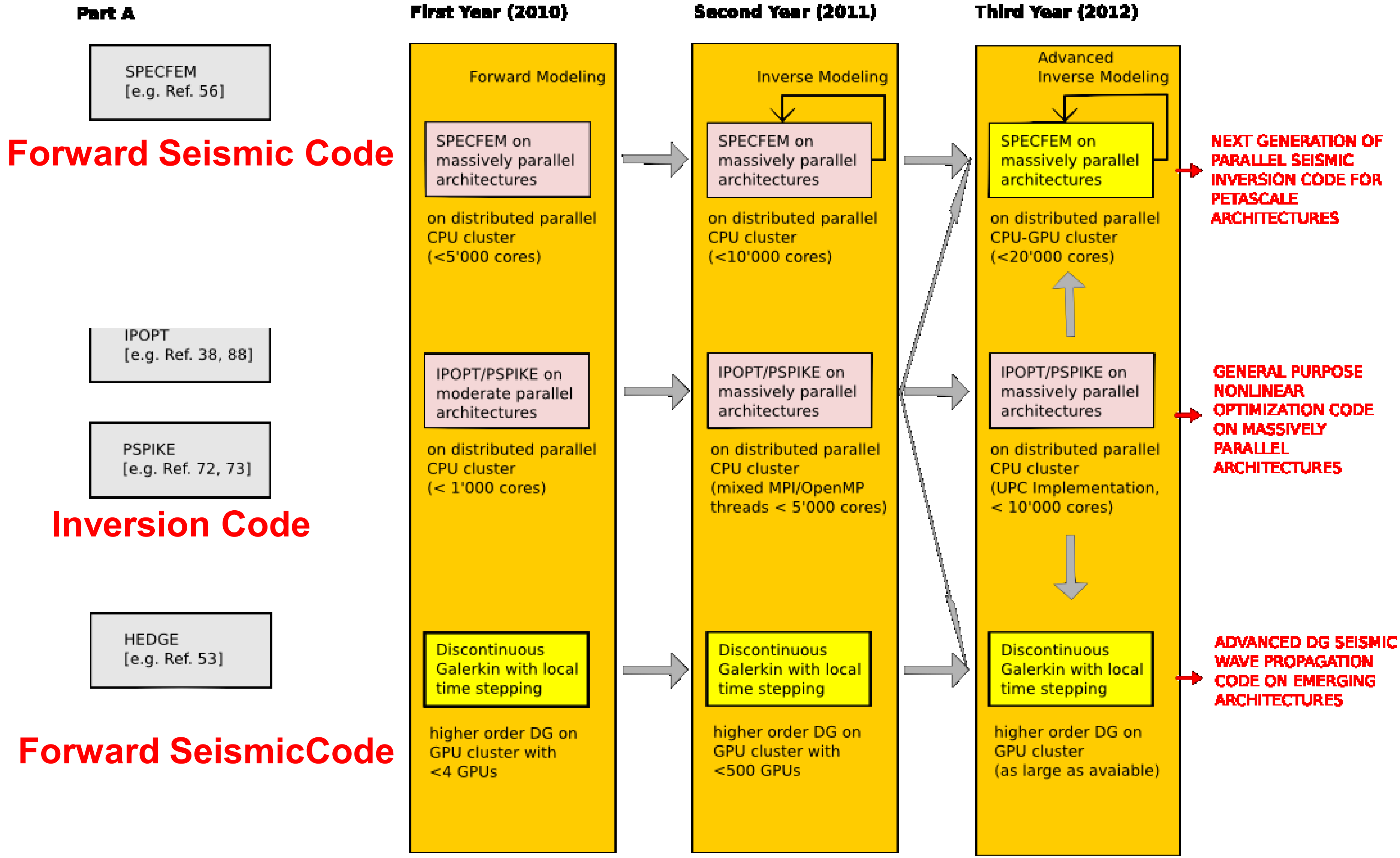
… too large to invert.

$$\begin{bmatrix} \nabla^2_{uu}\mathcal{L} & \nabla^2_{uq}\mathcal{L} & \nabla^2_{u\lambda}\mathcal{L} \\ \nabla^2_{qu}\mathcal{L} & \nabla^2_{qq}\mathcal{L} & \nabla^2_{q\lambda}\mathcal{L} \\ \nabla^2_{\lambda u}\mathcal{L} & \nabla^2_{\lambda q}\mathcal{L} & \nabla^2_{\lambda\lambda}\mathcal{L} \end{bmatrix} \begin{Bmatrix} \bar{u} \\ \bar{q} \\ \bar{\lambda} \end{Bmatrix} = - \begin{Bmatrix} \nabla_u \mathcal{L} \\ \nabla_q \mathcal{L} \\ \nabla_\lambda \mathcal{L} \end{Bmatrix}$$

KKT matrix is of size $2(N_n N_t)+N_n$. Here N_n is the number nodes, and N_t are the time steps.

- G. Biros and O. Ghattas, *Parallel Lagrange-Newton-Krylov-Schur Methods for PDE-Constrained Optimization. Part I: The Krylov-Schur Solver*, SIAM SISC, 2005
- Final Goal is to include inequality constraints into the inversion process (research with A. Wächter, IBM Watson)

3D Seismic Imaging Software Stack



Thank you for your attention!

Questions?