Acoustic and elastic inversion and imaging on future exascale machines

and the SPECFEM3D package

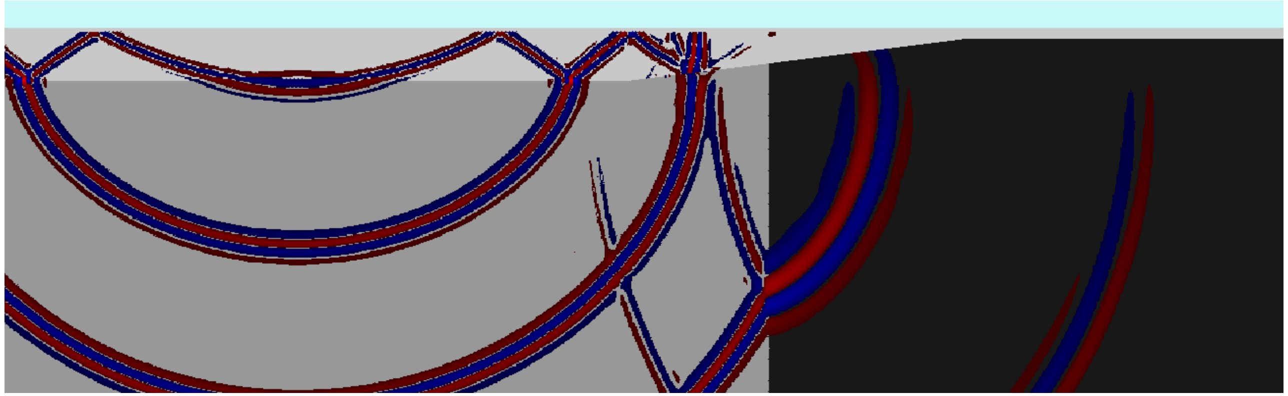
Dimitri Komatitsch (and many others)
Laboratory of Mechanics and Acoustics
CNRS, Marseille, France

ICTP, Trieste, Italy
November 13, 2014

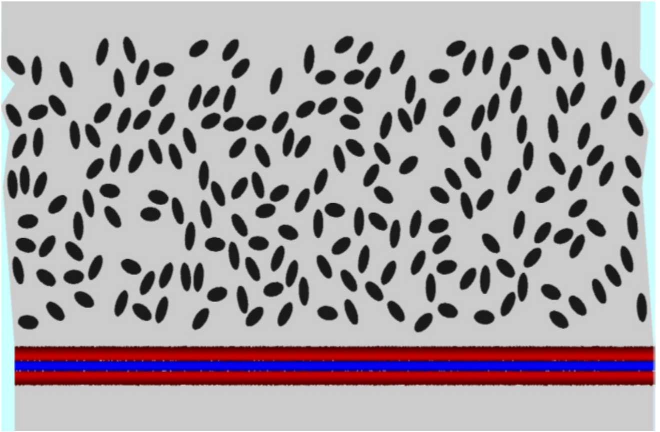
with some slides from Emanuele Casarotti (INGV Roma)
and from Matthieu Lefebvre (Princeton Univ, USA)

Application domains

Earthquakes



Ocean acoustics



Non destructive testing

Earthquake hazard assessment

Use parallel computing to simulate earthquakes

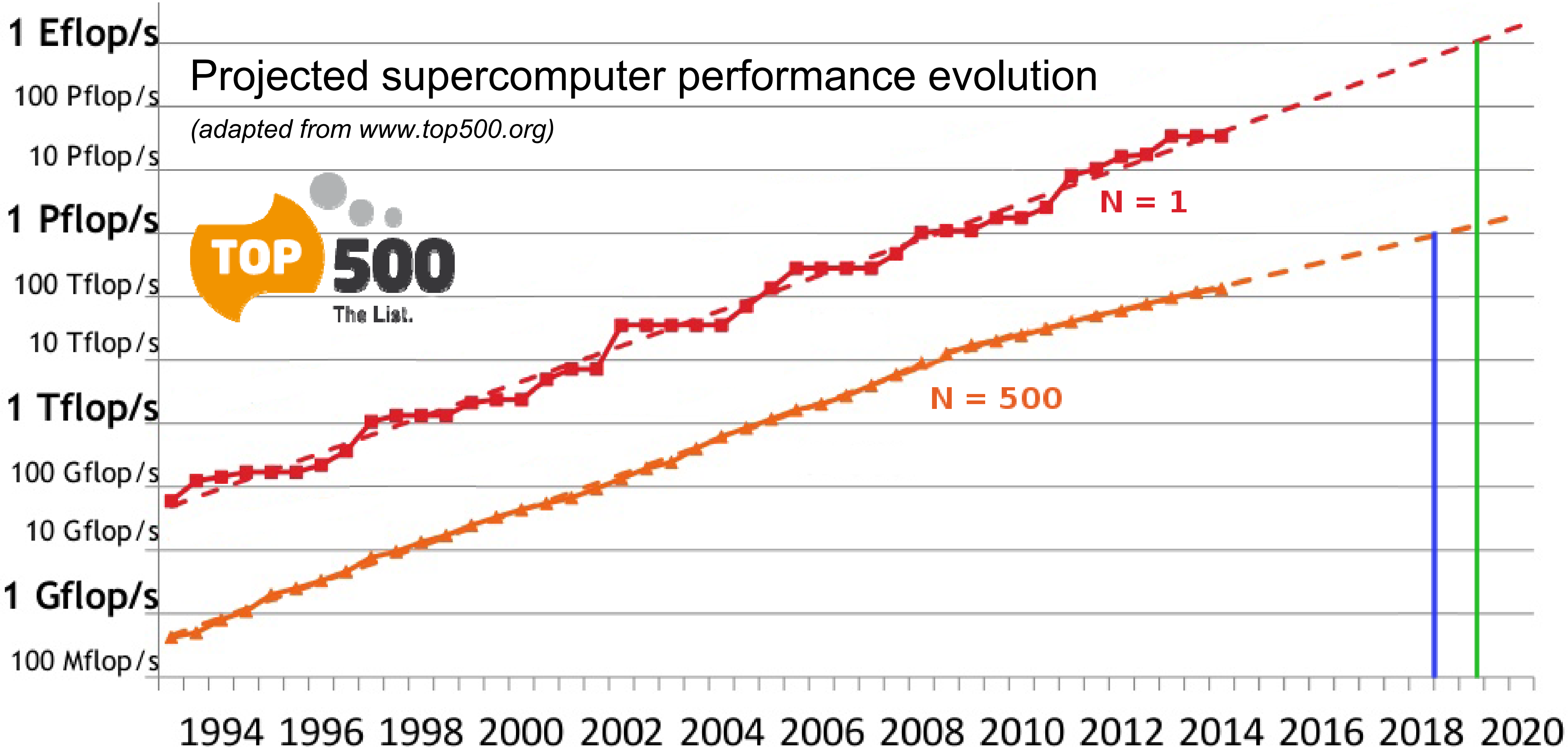
Learn about structure of the Earth based upon seismic waves (tomography)

Produce seismic hazard maps (local/regional scale) e.g. Los Angeles, Tokyo, Mexico City, Seattle

2001 Gujarati (M 7.7) Earthquake, India

20,000 people killed
167,000 injured
≈ 339,000 buildings destroyed
783,000 buildings damaged

About the path to exaflops



Projected supercomputer performance evolution

(adapted from www.top500.org)

- End of 2018 for exaflop/s, end of 2017 for petaflop/s easily everywhere, around 2027 for exaflop/s easily everywhere (?)
- For SPECFEM3D it is increasingly needed to perform a very large number (thousands!) of medium-size runs (500 to 2000 cores), rather than a single, very large grand-challenge run; this comes from solving imaging problems iteratively rather than a single forward problem once.

Equations of motion (solid)

Differential or *strong* form (e.g., finite differences):

$$\rho\,\partial_t^2\mathbf{u} = \nabla\cdot\sigma + \mathbf{f}$$

We solve the integral or *weak* form in the time domain:

$$\int \rho\,\mathbf{w}\cdot\partial_t^2\mathbf{u}\,\mathrm{d}^3\mathbf{r} = -\int \nabla\mathbf{w} : \sigma\,\mathrm{d}^3\mathbf{r}$$

$$+\,\mathbf{M} : \nabla\mathbf{w}(\mathbf{r}_\mathrm{s})\,S(t) - \int_{\mathrm{F-S}} \mathbf{w}\cdot\sigma\cdot\hat{\mathbf{n}}\,\mathrm{d}^2\mathbf{r}$$

+ attenuation (memory variables) and ocean load

Equations of motion (fluid)

Differential or *strong* form in the time domain:

$$\rho \partial_t \mathbf{v} = -\nabla p \qquad \partial_t p = -\kappa \nabla \cdot \mathbf{v}$$

with κ the adiabatic bulk modulus.

We use a scalar potential of ρ * displacement:

$$\rho \mathbf{u} = \nabla \chi \qquad p = -\partial_t^2 \chi$$

The integral or *weak* form is:

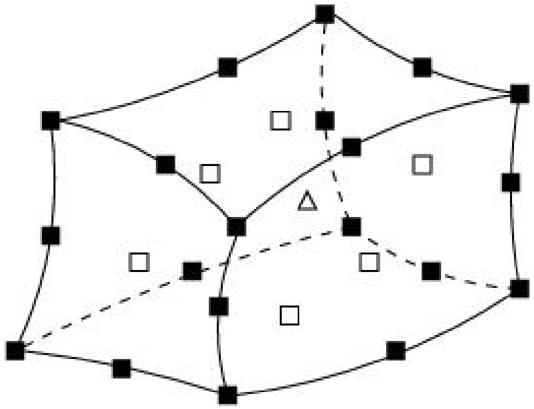
$$\int \kappa^{-1} w \partial_t^2 \chi \, \mathrm{d}^3\mathbf{r} = -\int \rho^{-1} \nabla w \cdot \nabla \chi \, \mathrm{d}^3\mathbf{r}$$

$$+\int_{\mathrm{F-S}} w \hat{\mathbf{n}} \cdot \mathbf{v} \, \mathrm{d}^2\mathbf{r}$$

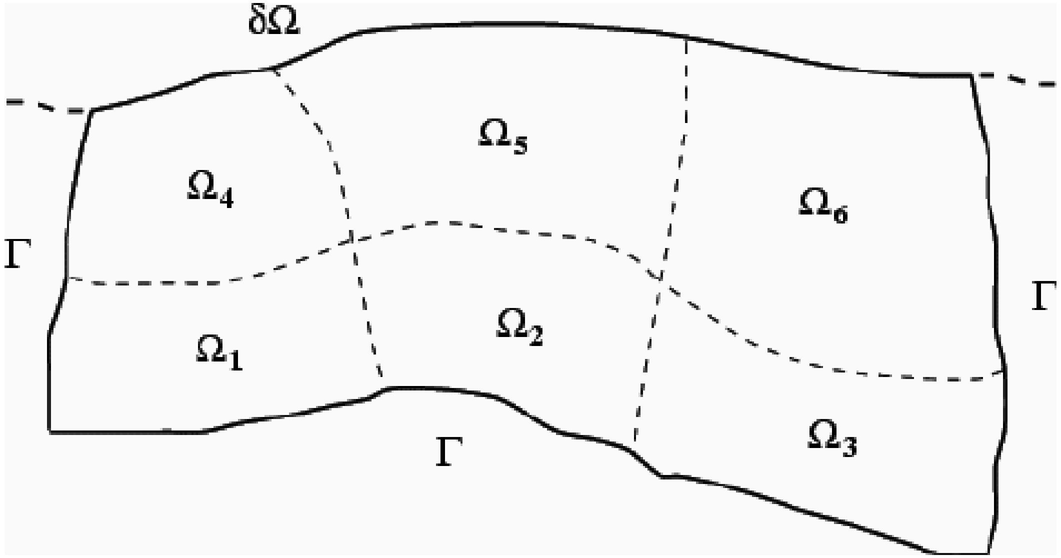
⇒ cheap (scalar potential)

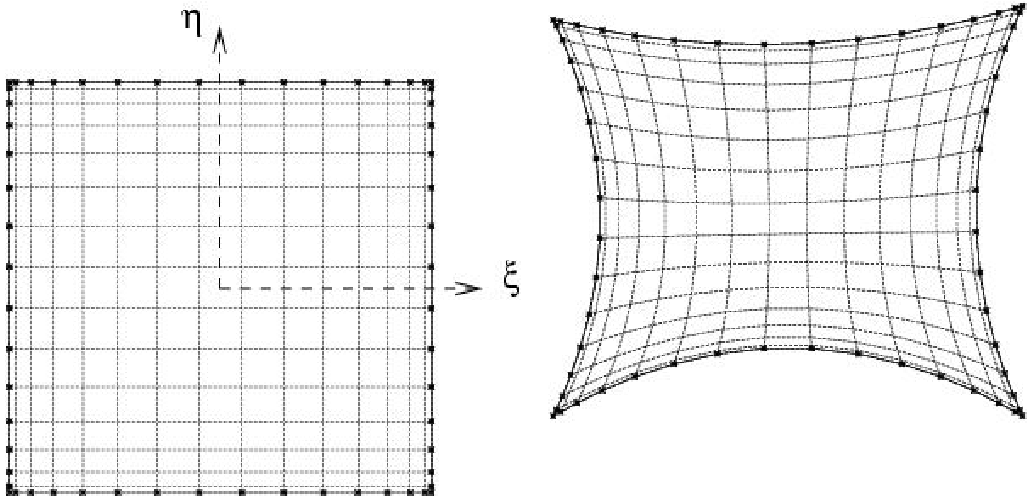
⇒ natural coupling with solid

Spectral-Element Method



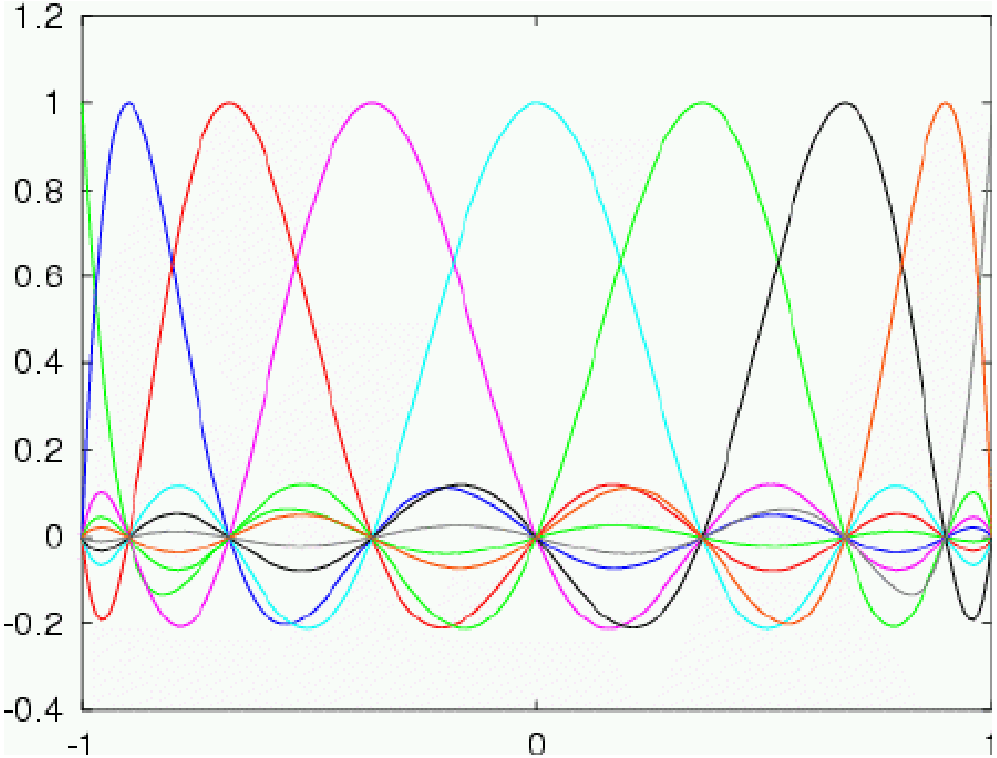
- Developed in Computational Fluid Dynamics (Patera 1984)
- Accuracy of a pseudospectral method, flexibility of a finite-element method
- Extended by Komatitsch and Tromp, Chaljub et al.
- Large curved “spectral” finite-elements with high-degree polynomial interpolation
- Mesh honors the main discontinuities (velocity, density) and topography
- Very efficient on parallel computers, no linear system to invert (diagonal mass matrix)

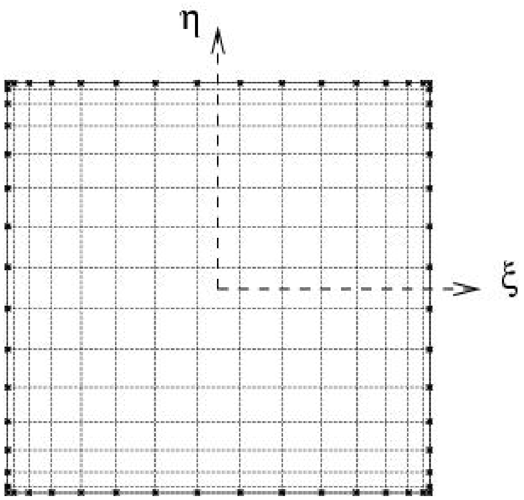


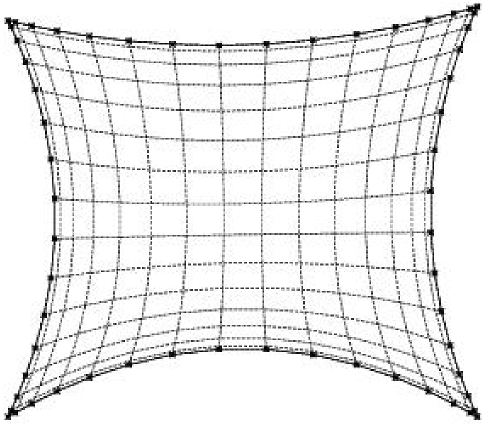


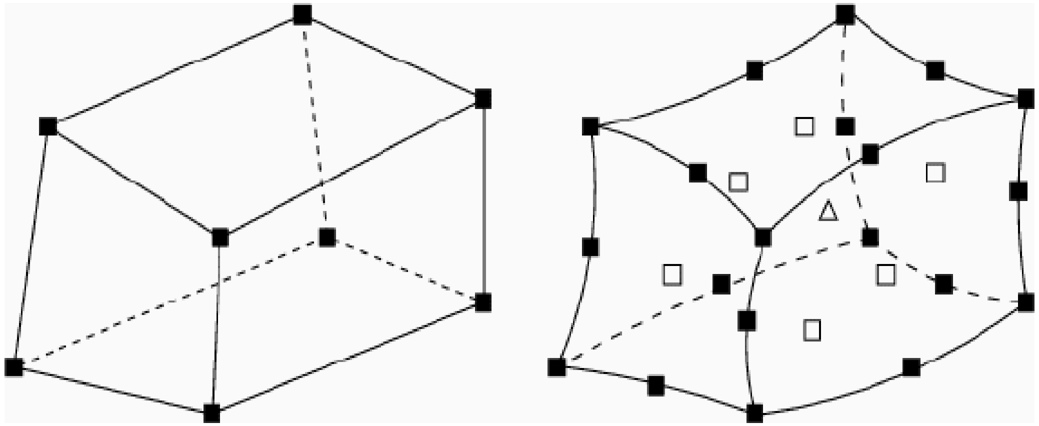
Finite elements

- High-degree pseudospectral finite elements
- N = 5 to 8 usually
- *Strictly* diagonal mass matrix
- No linear system to invert
- Fully explicit time scheme



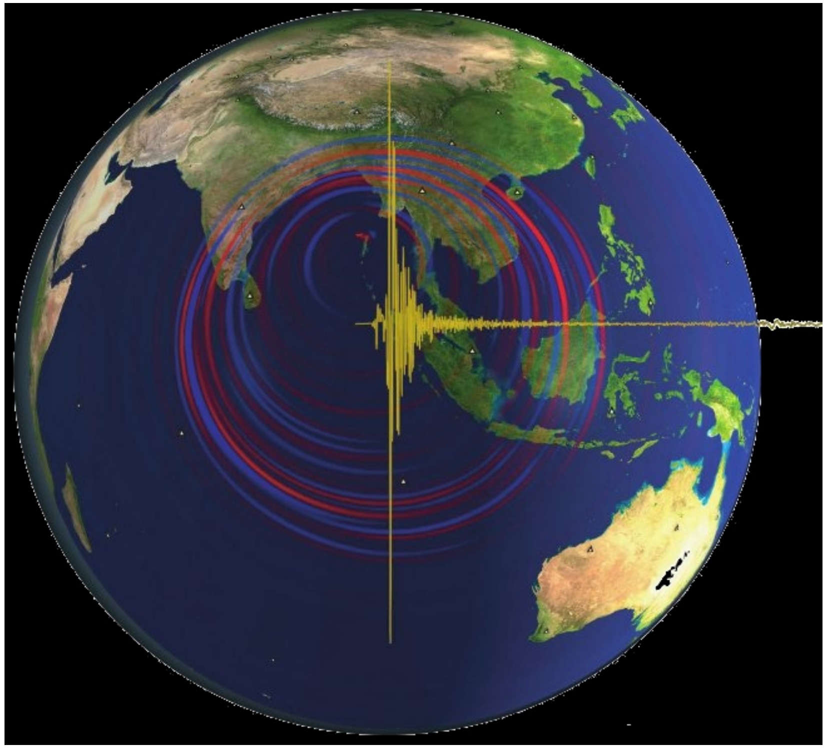






Our SPECFEM3D software package

User download map



Goal: model acoustic / elastic / viscoelastic / poroelastic / seismic wave propagation in in non destructive testing, in ocean acoustics, in the Earth (earthquakes, oil industry)…

The SPECFEM3D source code is open (GNU GPL v2)

Initially Komatitsch and Vilotte at IPG Paris (France), mostly developed by Dimitri Komatitsch and Jeroen Tromp at Harvard University, then Caltech, Princeton (USA) and CNRS (France) since 1996.

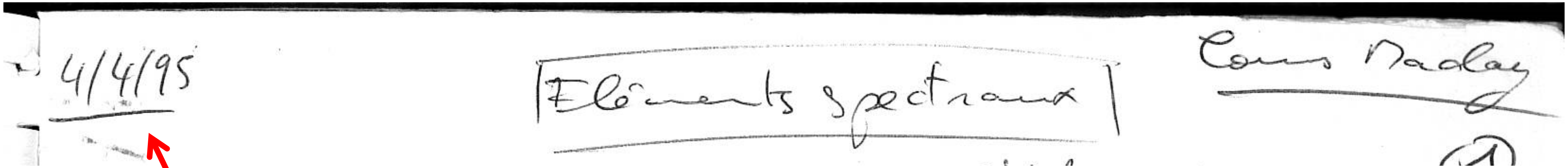
Improved with INRIA and University of Pau (France), ETH Zürich and University of Basel (Switzerland), the Barcelona Supercomputing Center (Spain), NVIDIA…

OGS February 1995

G. Seriani, E. Priolo,
J. Carcione

Non diagonal
mass matrix

Yvon Maday, Paris, April 4, 1995



4/4/95

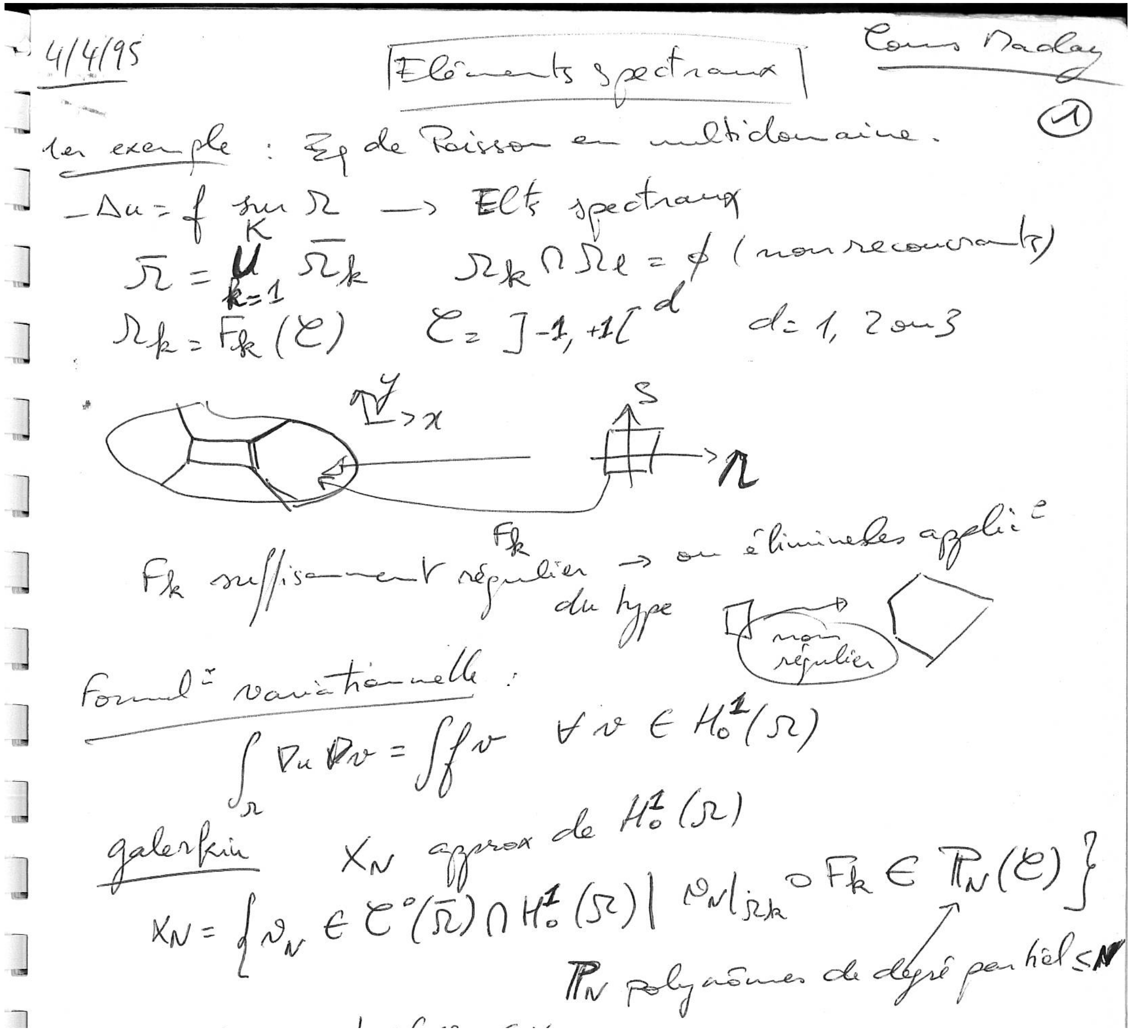
Eléments spectraux

Cours Maday

①

April 4, 1995

Diagonal
mass matrix



4/4/95

Eléments spectraux

Cours Maday

①

1er exemple : Eq de Poisson en multidomaine.

$-\Delta u = f$ sur Ω → Elts spectraux

$\bar{\Omega} = \bigcup_{k=1}^{K} \bar{\Omega}_k \qquad \Omega_k \cap \Omega_\ell = \phi$ (non recouvrants)

$\Omega_k = F_k(\mathcal{C}) \qquad \mathcal{C} =]-1,+1[^d \qquad d = 1, 2$ ou 3

F_k suffisamment régulier → on élimine les applic$^{\text{e}}$ F_k du type

Formul$^{\text{n}}$ variationnelle :

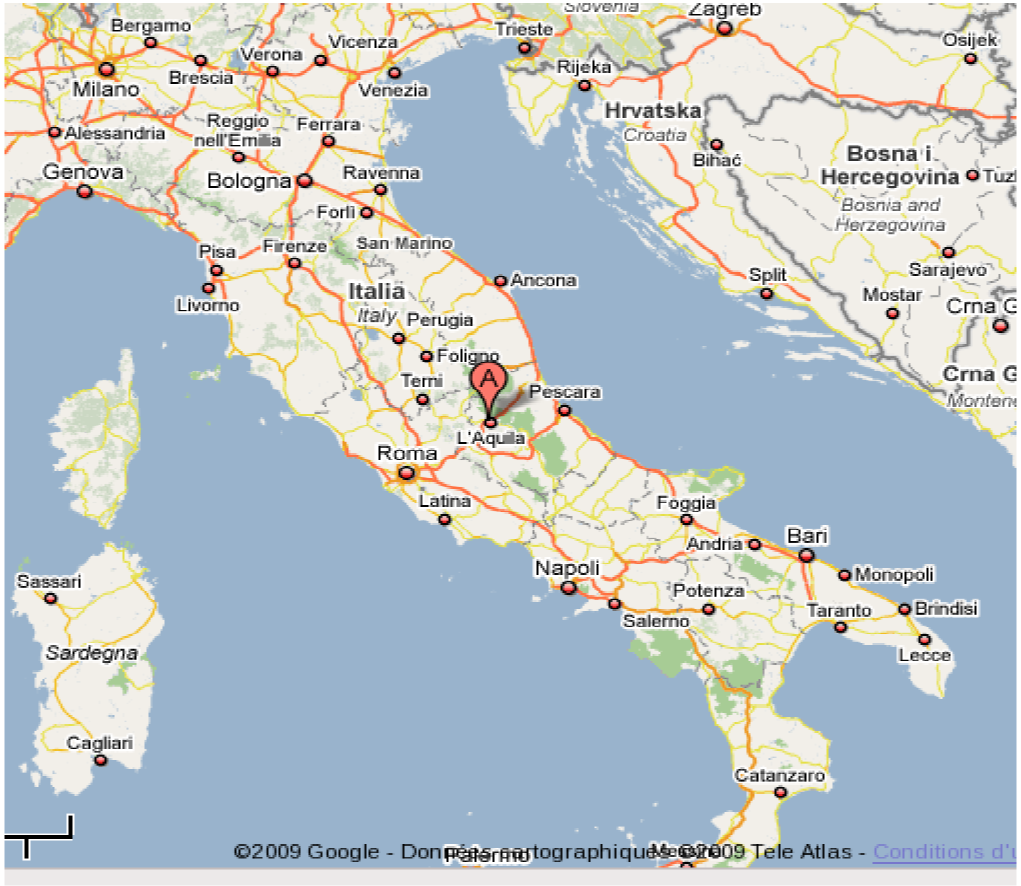
$$\int_\Omega \nabla u \nabla v = \int f v \quad \forall v \in H_0^1(\Omega)$$

galerkin $\quad X_N$ approx de $H_0^1(\Omega)$

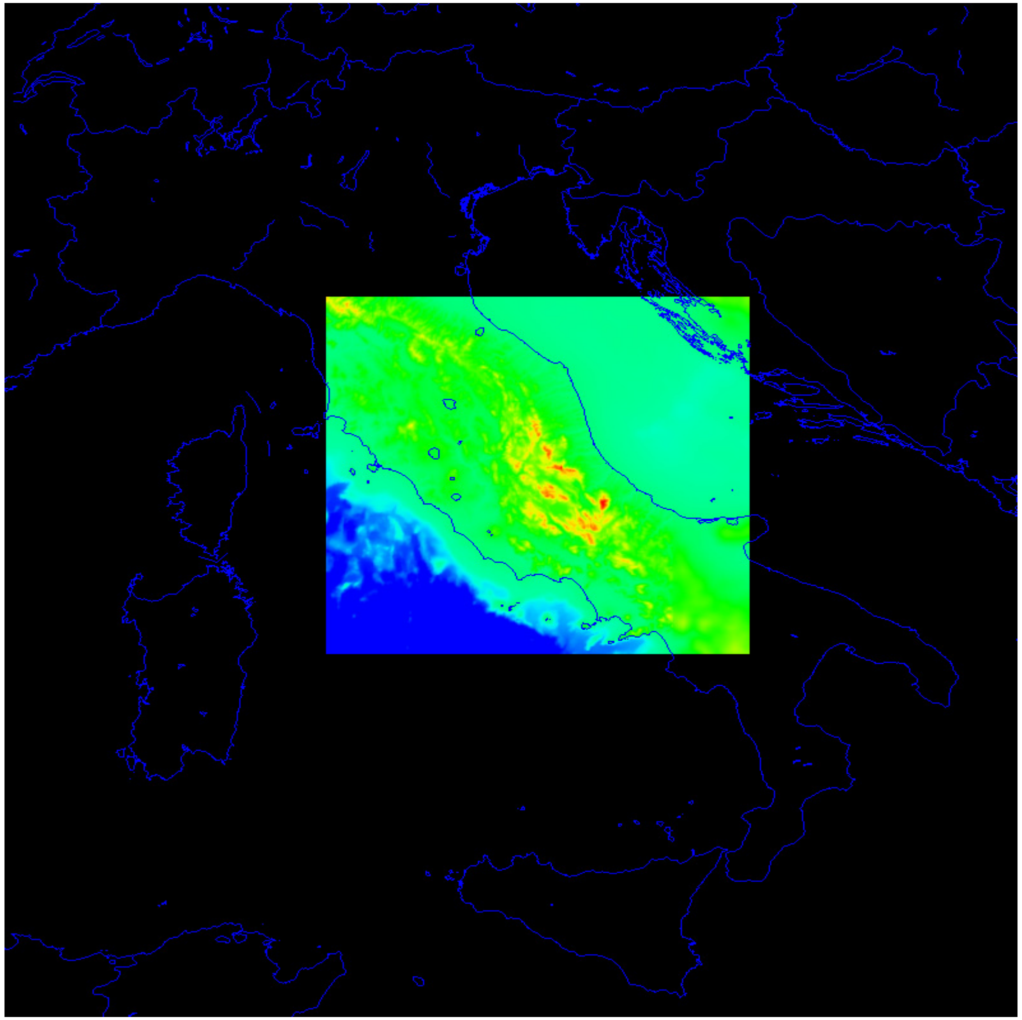
$$X_N = \{ v_N \in \mathcal{C}^0(\bar{\Omega}) \cap H_0^1(\Omega) \mid v_N|_{\Omega_k} \circ F_k \in \mathbb{P}_N(\mathcal{C}) \}$$

$\mathbb{P}_N$ polynômes de degré par var$^{\text{el}}$ $\leq N$

L’Aquila, Italy, April 6, 2009 (Mw = 6.2)



Location of the epicenter
(© Google Maps)



Mesh defined on the
JADE supercomputer
on April 7, 2009

Earthquakes

6 April 2009
M_w 6.2 L'Aquila (Italy)

310 casualties
~ 1000 injured
~ 26000 homeless

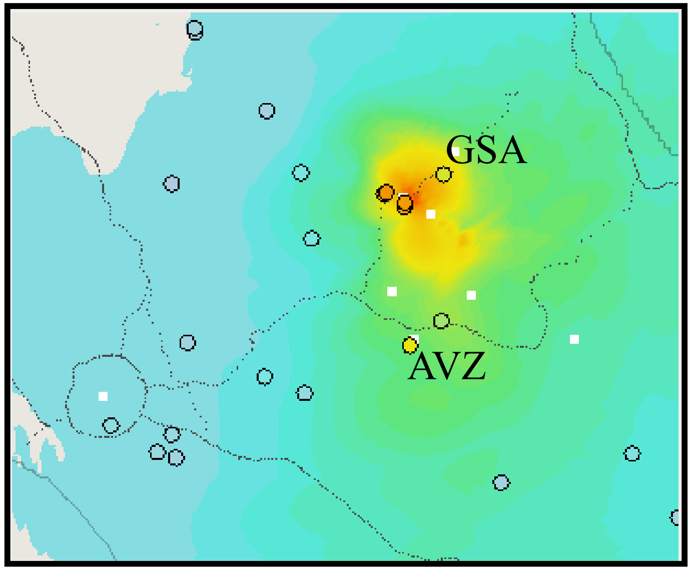
Istituto Nazionale di Geofisica e Vulcanologia

Collaboration with
Emanuele Casarotti and Federica Magnoni (INGV Roma, Italy)

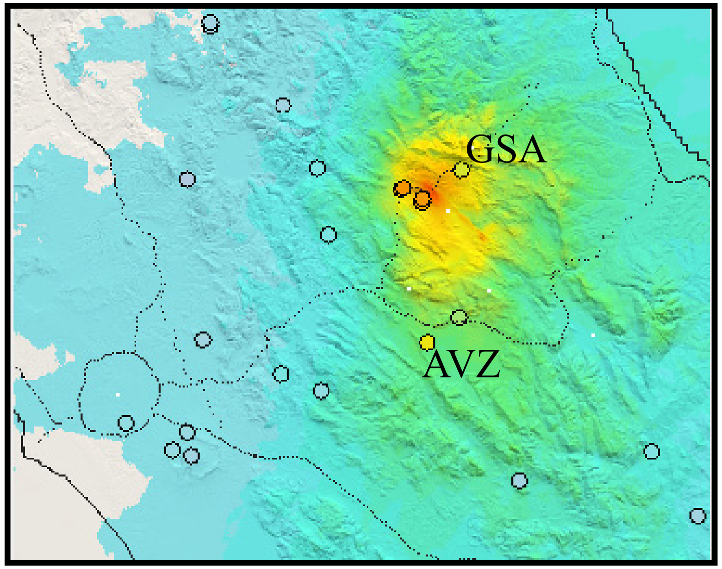
M_W 6.2 L'Aquila

Istituto Nazionale di
Geofisica e Vulcanologia

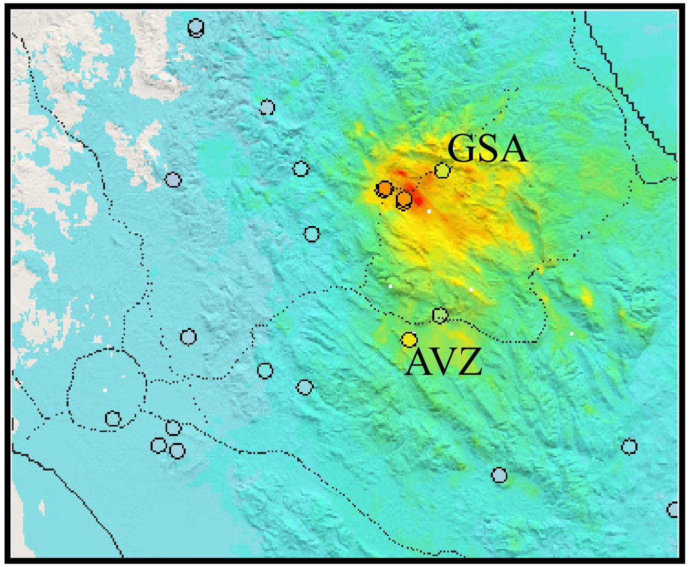
Scenario



1D flat - max PGV 45 cm/s



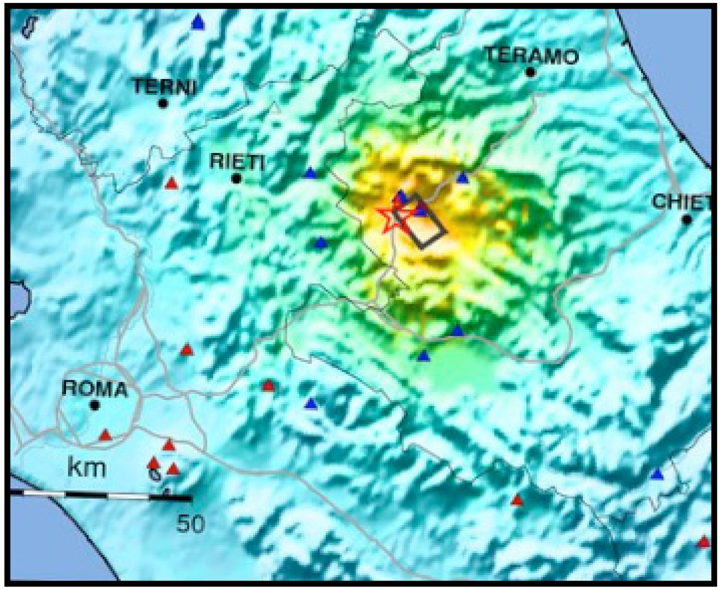
1D w topo - max PGV 48 cm/s



3D - max PGV 74 cm/s

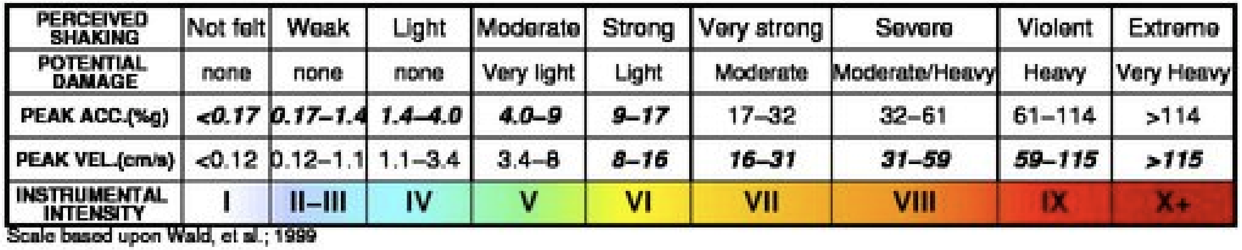
Max PGV [illegible] a ~ 65 cm/s

Istituto Nazionale di
Geofisica e Vulcanologia



(Faenza et al., 2011)

INGV ShakeMap : CENTRAL ITALY – AQUILANO



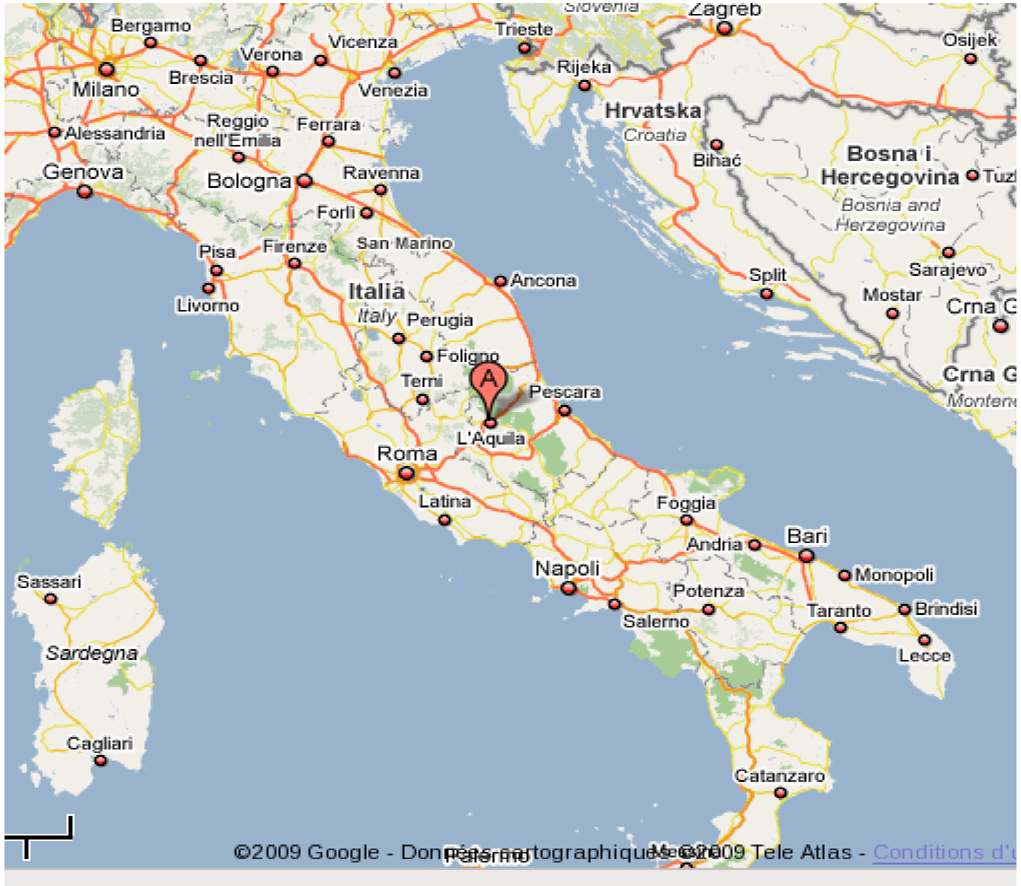
PERCEIVED SHAKING	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
POTENTIAL DAMAGE	none	none	none	Very light	Light	Moderate	Moderate/Heavy	Heavy	Very Heavy
PEAK ACC.(%g)	<0.17	0.17–1.4	1.4–4.0	4.0–9	9–17	17–32	32–61	61–114	>114
PEAK VEL.(cm/s)	<0.12	0.12–1.1	1.1–3.4	3.4–8	8–16	16–31	31–59	59–115	>115
INSTRUMENTAL INTENSITY	I	II–III	IV	V	VI	VII	VIII	IX	X+

Scale based upon Wald, et al.; 1999

Our joint current PRACE project

PRACE project with INGV Roma (E. Casarotti, F. Magnoni, D. Melini, A. Michelini) + Princeton University, USA (J. Tromp) + University of Fairbanks, Alaska (C. Tape) to image the Italian lithosphere: 40 million core hours on CURIE (PRACE / TGCC, France)

“IMAGINE_IT: 3D full-wave tomographic IMAGINg of the Entire ITalian lithosphere”

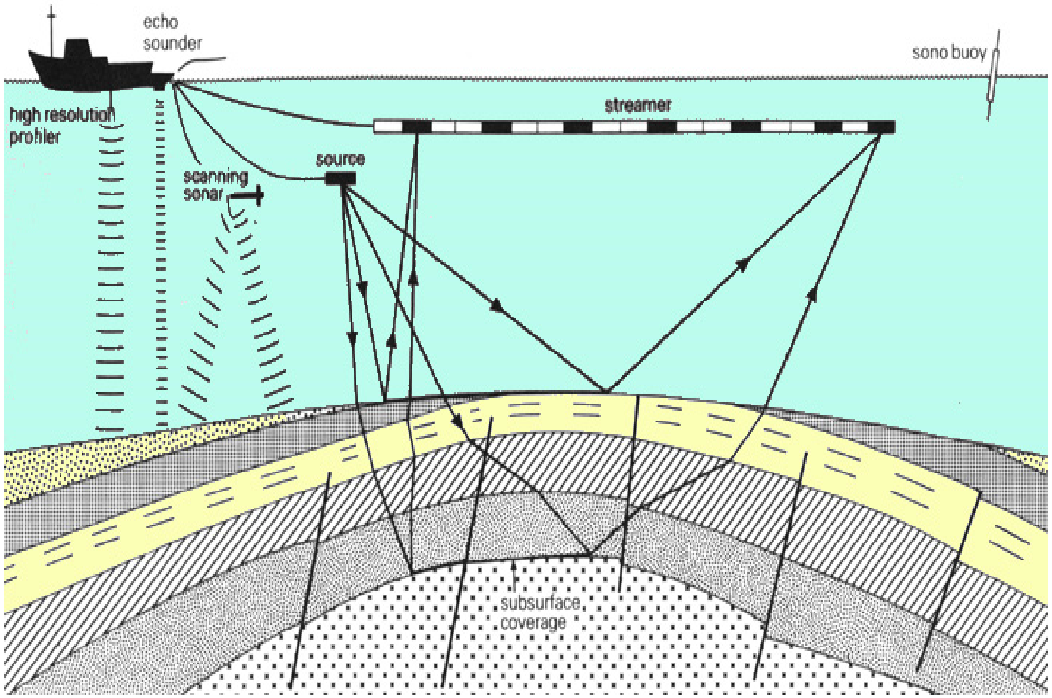


Istituto Nazionale di
Geofisica e Vulcanologia

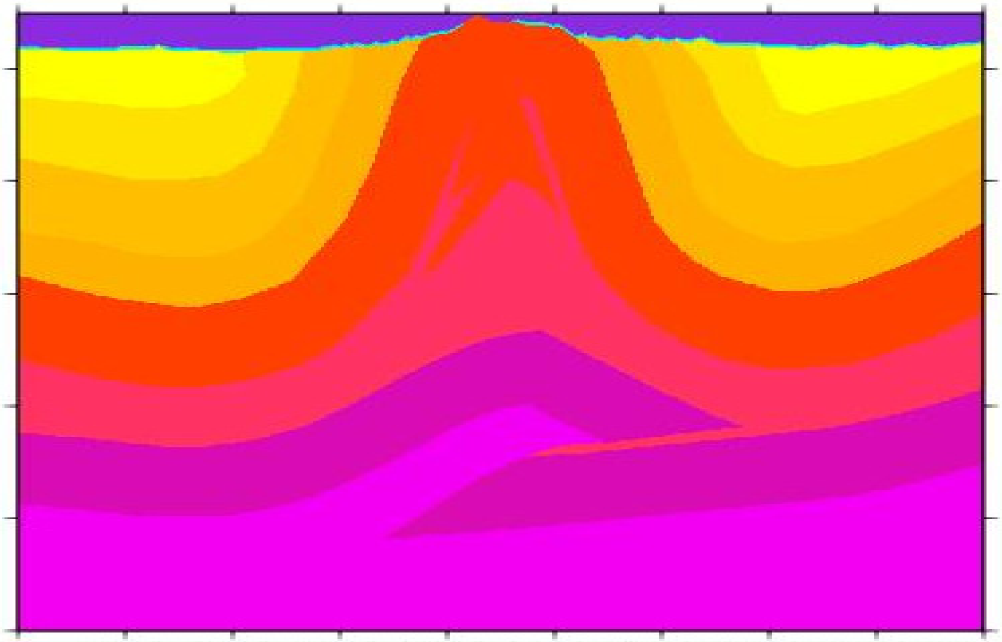
cnrs

Oil industry applications

TOTAL S.A.



Offshore

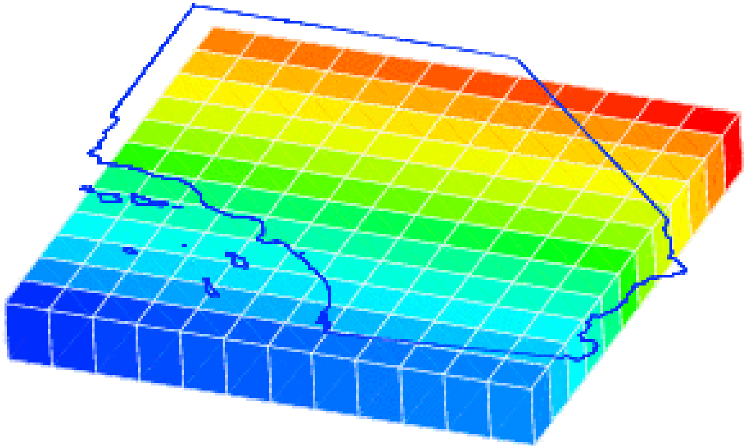


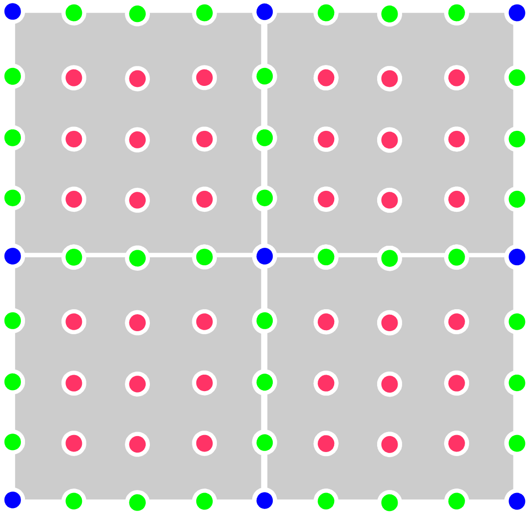
In foothill regions

- Elastic wave propagation in complex 3D structures,
- Often fluid / solid problems: many oil fields are located offshore (deep offshore, or shallower).
- Anisotropic rocks, geological faults, cracks, bathymetry / topography…
- Thin weathered zone / layer at the surface ⇒ model dispersive surface waves.

Building a cluster

Year 2000, Caltech (USA).





Parallel calculations
with message passing (MPI).
320 processors, 160 Gb of memory, Linux.

Earth Simulator – Japan (2002 - 2003)

Earth Simulator Center
Japan Marine Science and Technology Center
Yokohama Institute For Earth Science

Pictures and data taken from www.es.jamstec.go.jp

640 processor nodes, each consisting of eight vector processors are connected as a high speed interconnection network. The Earth Simulator was in 2004 the fastest supercomputer in the world.

Huge progress in 10 years

SC2003 IGNITING INNOVATION

SC2003 CONFERENCE PHOENIX, AZ NOVEMBER 15-21, 2003

SC2003 Gordon Bell Award

Dimitri Komatitsch

California Institute of Technology

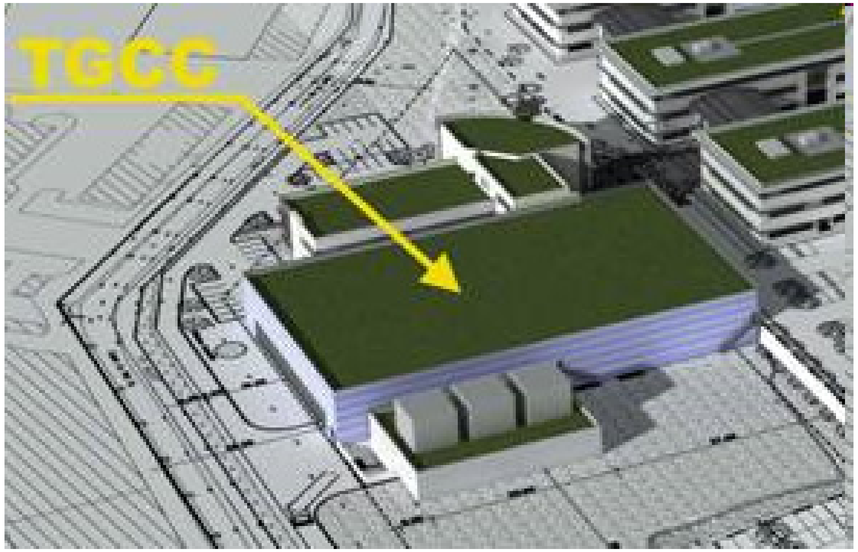
A 14.6 Billion Degrees of Freedom, 5 Teraflops, 2.5 Terabyte Earthquake Simulation on the Earth Simulator

SC2003 GENERAL CHAIR

SC2003 TECHNICAL PROGRAM CO-CHAIR

Earth Simulator: Peak 40 Teraflops; we won the Gordon Bell supercomputing award with SPECFEM3D for a run at 5 teraflops sustained (!!) (OK, with 15 billion degrees of freedom…)

PRACE + TGCC



Petaflop for the European infrastructure

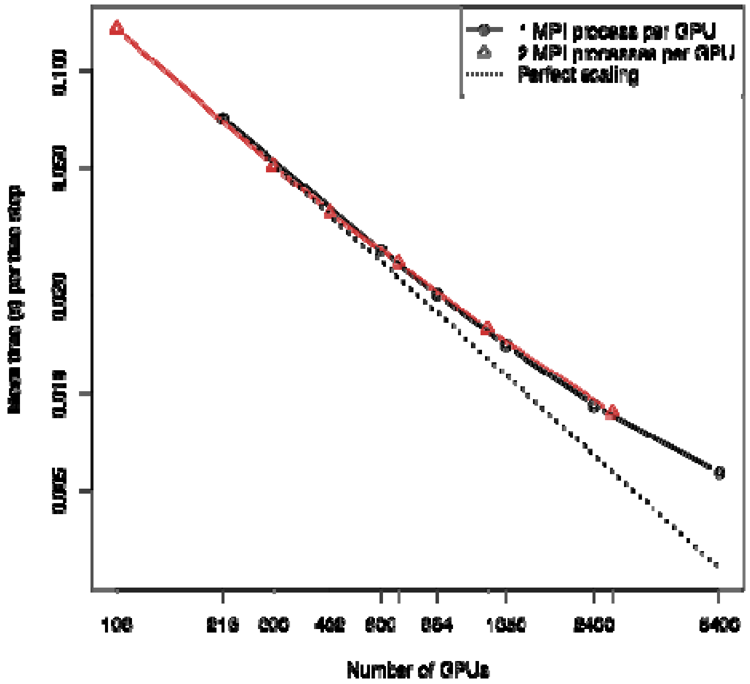
The TGCC (Très Grand Centre de Calcul / “Very Big Computing Center”) hosts the PRACE “CURIE” European machine

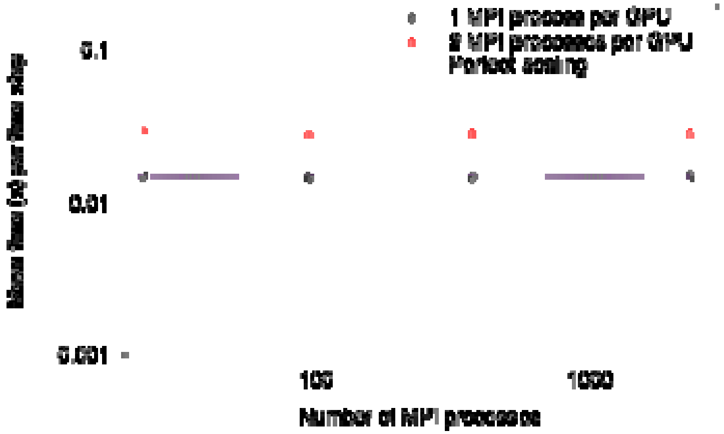
GENCI in France, CINECA / CASPUR in Italy.

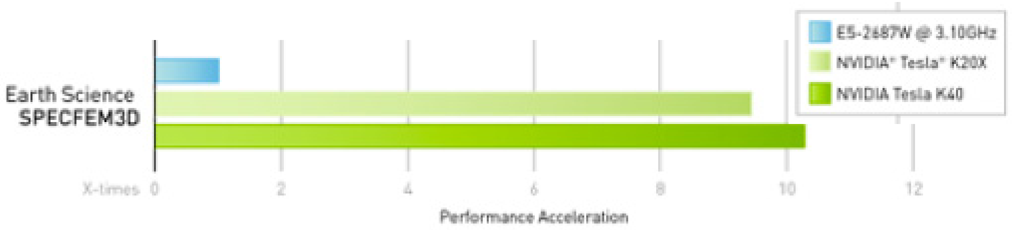

COMPUTATIONAL INFRASTRUCTURE FOR GEODYNAMICS (CIG)
PRINCETON UNIVERSITY (USA)
CNRS and UNIVERSITY OF MARSEILLE (FRANCE)
ETH ZÜRICH (SWITZERLAND)

SPECFEM 3D
Globe
User Manual
Version 5.1

Solver Performance







http://www.nvidia.com/object/gpu-test-drive.html

February 05, 2013

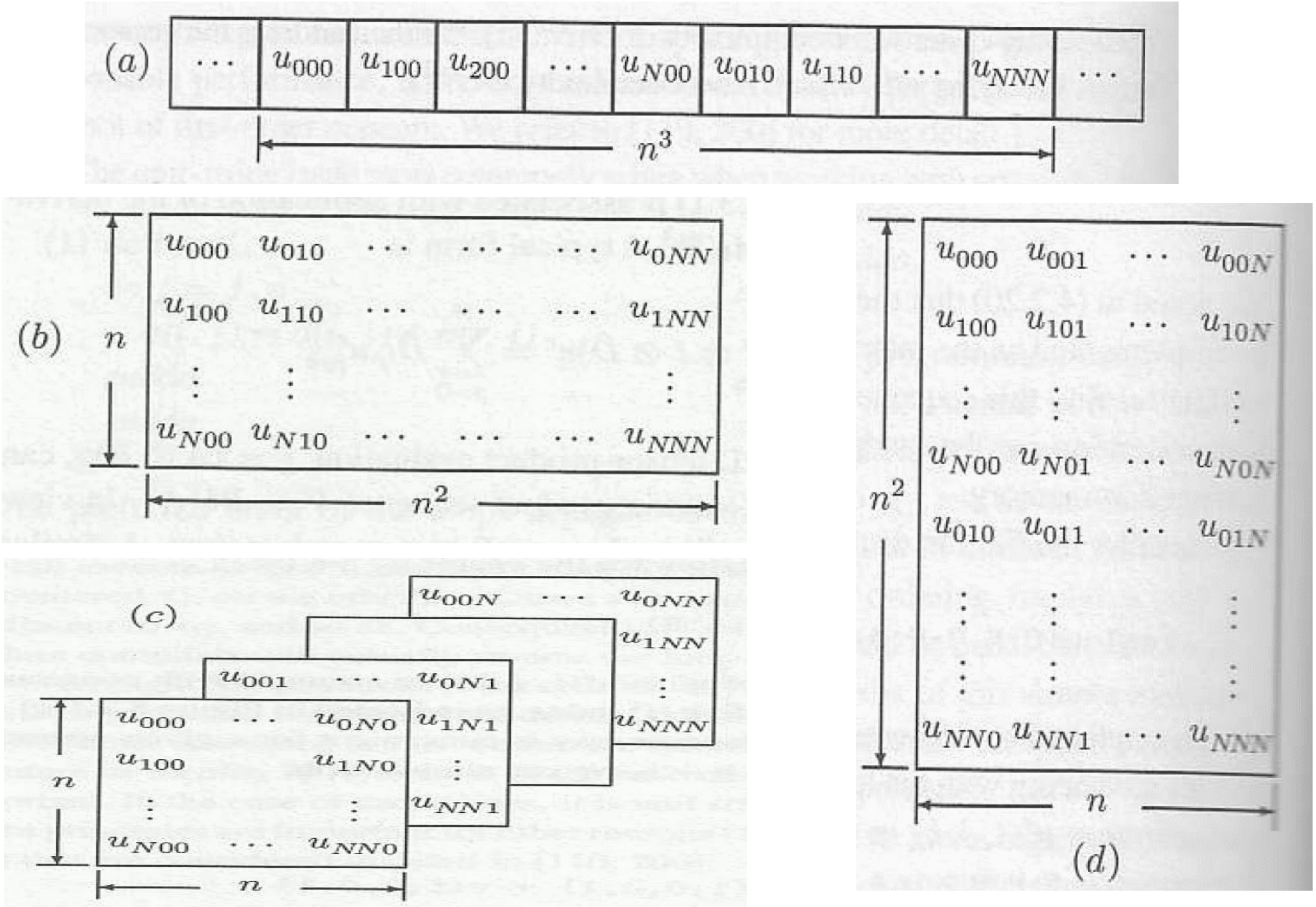
Four Applications Sustain One Petaflop on Blue Waters

HPC wire

July 18, 2012

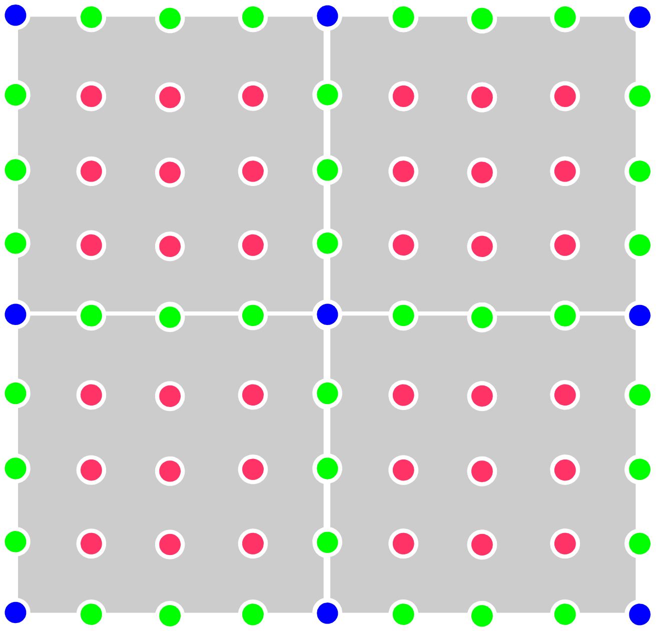
Researchers Squeeze GPU Performance from 11 Big Science Apps

Deville (2002) unrolling algorithm

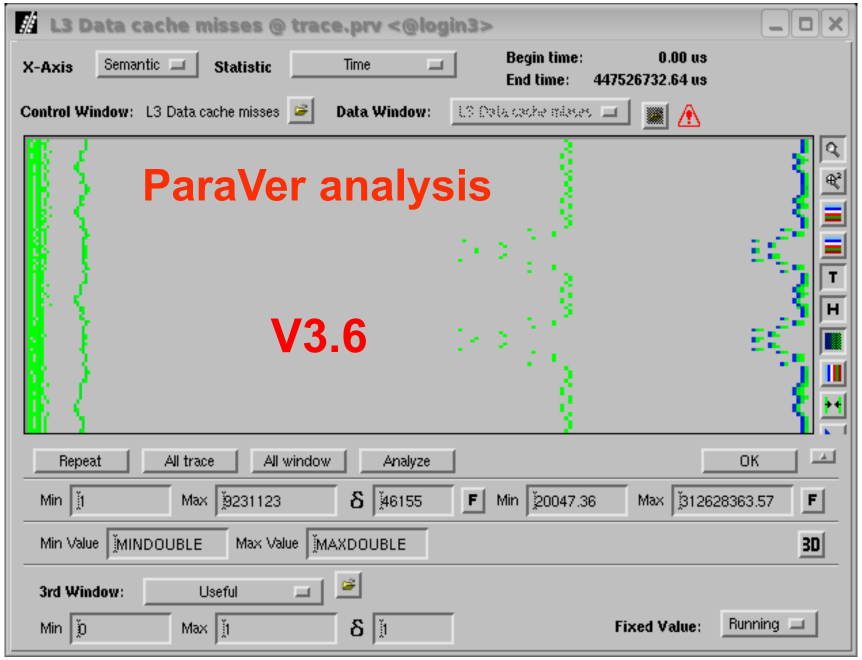


Each element has 5 x 5 x 5 points. Minimize the total number of memory accesses to make the code less memory bound.

Results for load balancing: cache misses (J. Labarta, BSC)



⇒ it is crucial to reuse common points by keeping them in the cache

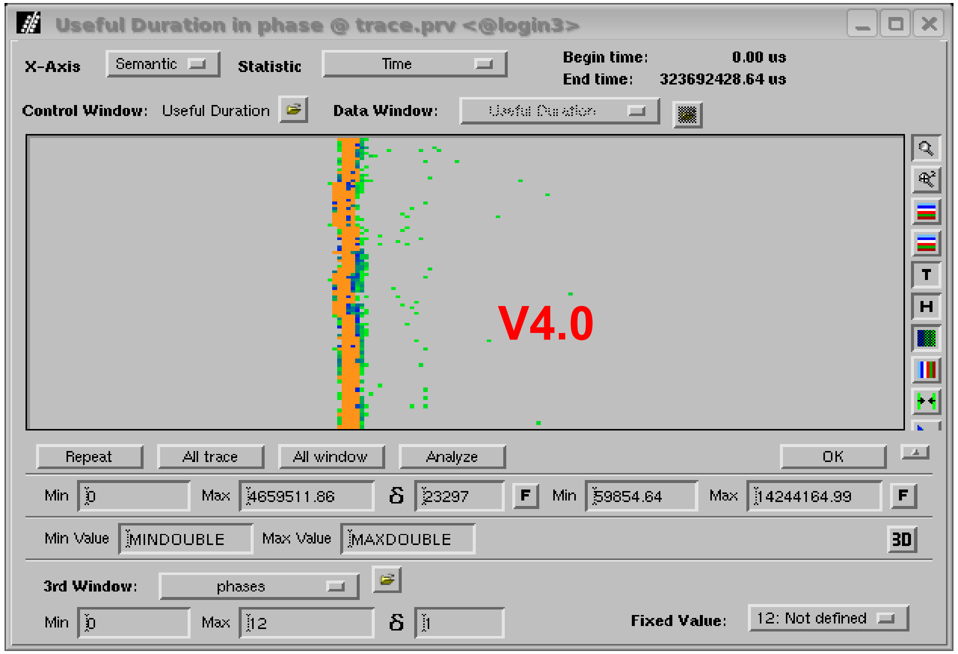


ParaVer analysis

V3.6

L3 Data cache misses @ trace.prv <@login3>

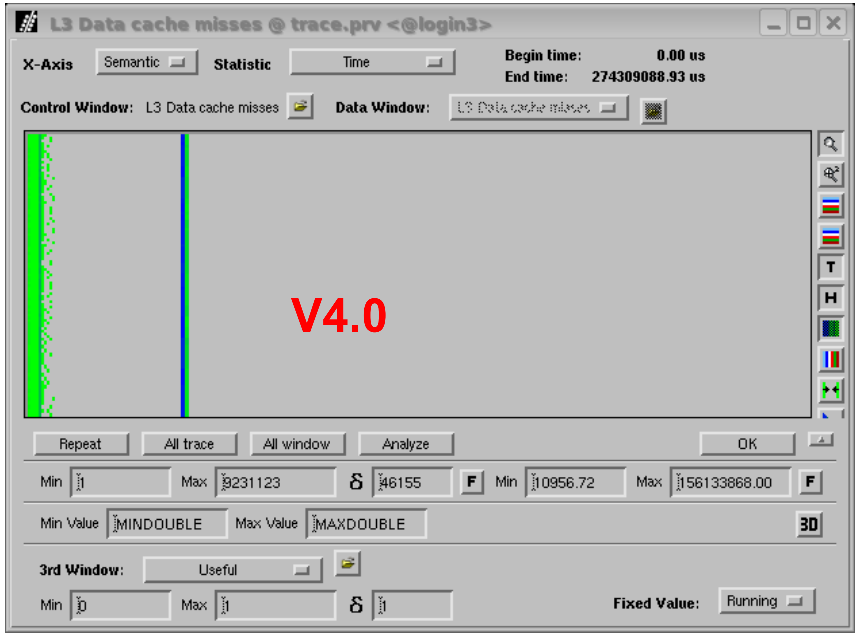
X-Axis: Semantic
Statistic: Time
Begin time: 0.00 us
End time: 447526732.64 us
Control Window: L3 Data cache misses
Data Window: L3 Data cache misses
Min 1 Max 9231123 δ 46155 F Min 20047.36 Max 312628363.57 F
Min Value MINDOUBLE Max Value MAXDOUBLE 3D
3rd Window: Useful
Min 0 Max 1 δ 1 Fixed Value: Running



V4.0

Useful Duration in phase @ trace.prv <@login3>

X-Axis: Semantic
Statistic: Time
Begin time: 0.00 us
End time: 323692428.64 us
Control Window: Useful Duration
Data Window: Useful Duration
Min 0 Max 4659511.86 δ 23297 F Min 59854.64 Max 14244164.99 F
Min Value MINDOUBLE Max Value MAXDOUBLE 3D
3rd Window: phases
Min 0 Max 12 δ 1 Fixed Value: 12: Not defined



V4.0

L3 Data cache misses @ trace.prv <@login3>

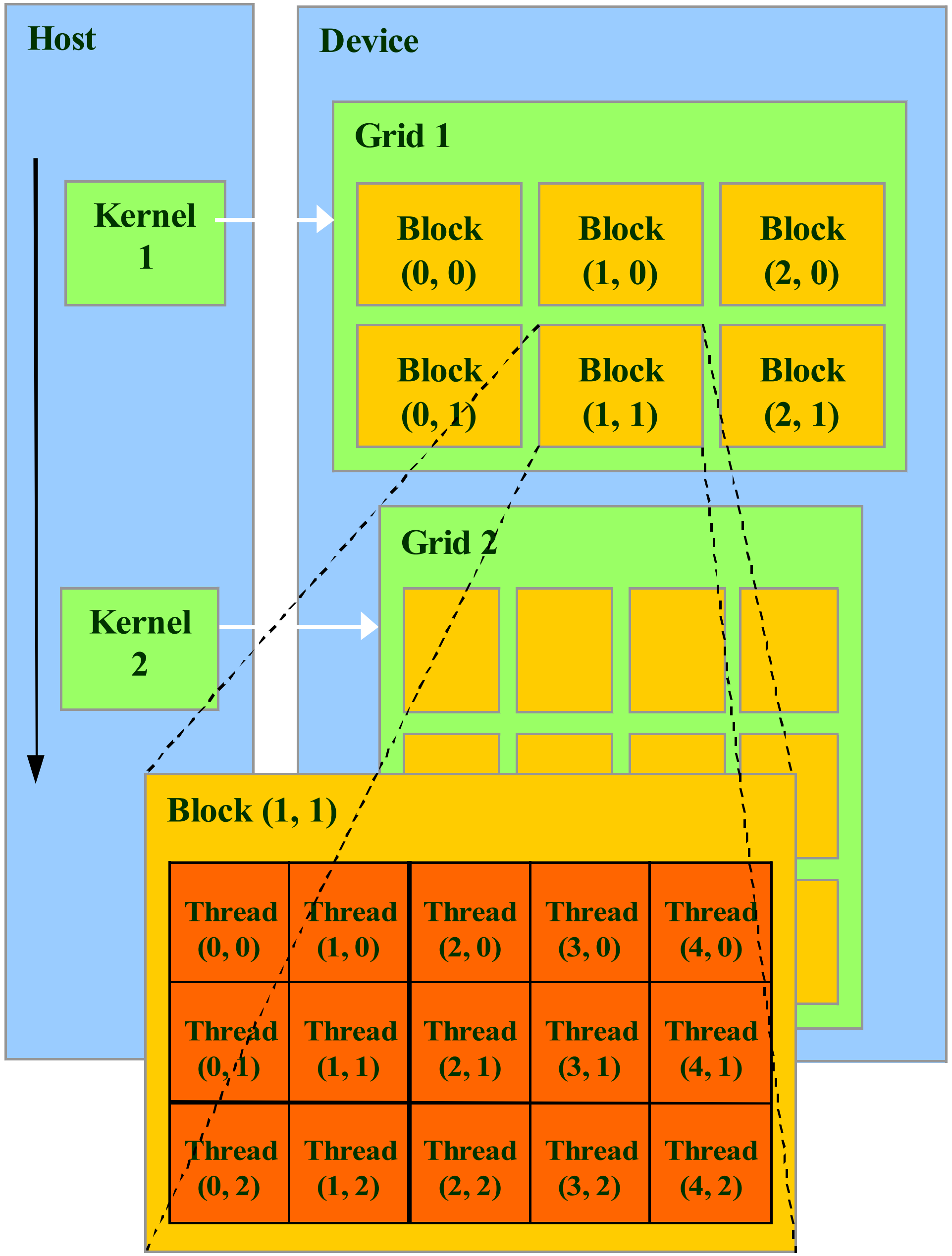
X-Axis: Semantic
Statistic: Time
Begin time: 0.00 us
End time: 274309088.93 us
Control Window: L3 Data cache misses
Data Window: L3 Data cache misses
Min 1 Max 9231123 δ 46155 F Min 10956.72 Max 156133868.00 F
Min Value MINDOUBLE Max Value MAXDOUBLE 3D
3rd Window: Useful
Min 0 Max 1 δ 1 Fixed Value: Running

GPU graphics cards

NVIDIA

Why are they so powerful for scientific computing?

Massive lightweight multithreading.



Host

Kernel 1

Kernel 2

Device

Grid 1

Block (0, 0)	Block (1, 0)	Block (2, 0)
Block (0, 1)	Block (1, 1)	Block (2, 1)

Grid 2

Block (1, 1)

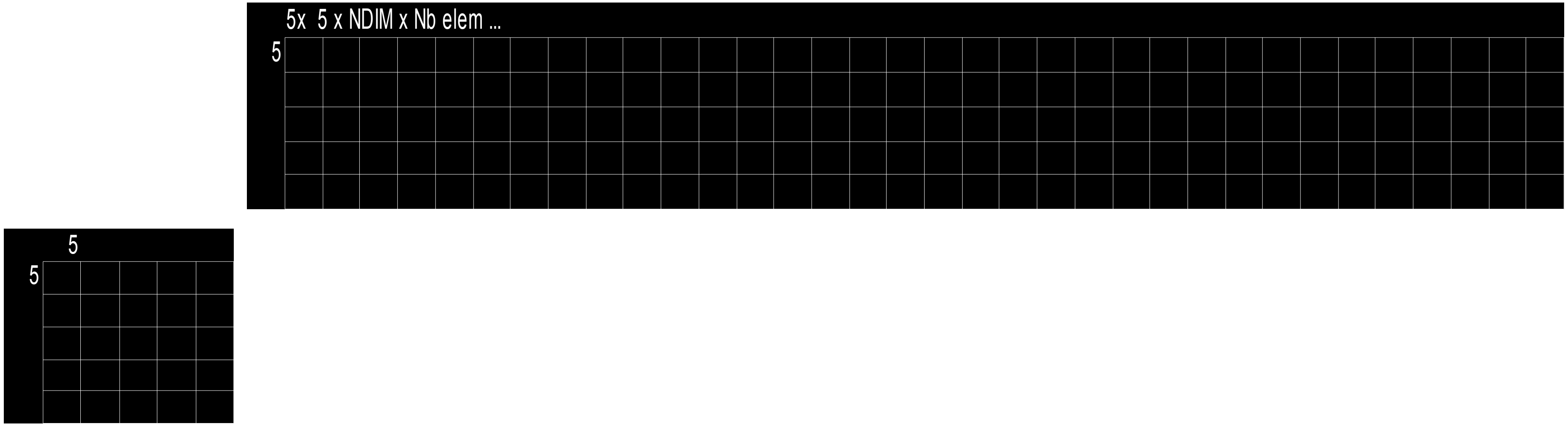
Thread (0, 0)	Thread (1, 0)	Thread (2, 0)	Thread (3, 0)	Thread (4, 0)
Thread (0, 1)	Thread (1, 1)	Thread (2, 1)	Thread (3, 1)	Thread (4, 1)
Thread (0, 2)	Thread (1, 2)	Thread (2, 2)	Thread (3, 2)	Thread (4, 2)

Collaboration with D. Peter (ETH), P. Messmer (NVIDIA), D. Göddeke (Dortmund)

Porting SPECFEM3D on GPUs

- At each iteration of the serial time loop, three main types of operations are performed:
  - **update (with no dependency) of some global arrays composed of the unique points of the mesh**
  - **purely local calculations of the product of predefined derivative matrices with a local copy of the displacement vector along cut planes in the three directions (i, j and k) of a 3D spectral element**
  - **update (with no dependency) of other global arrays composed of the unique points of the mesh**

BLAS 3 (Basic Linear Algebra Subroutines)



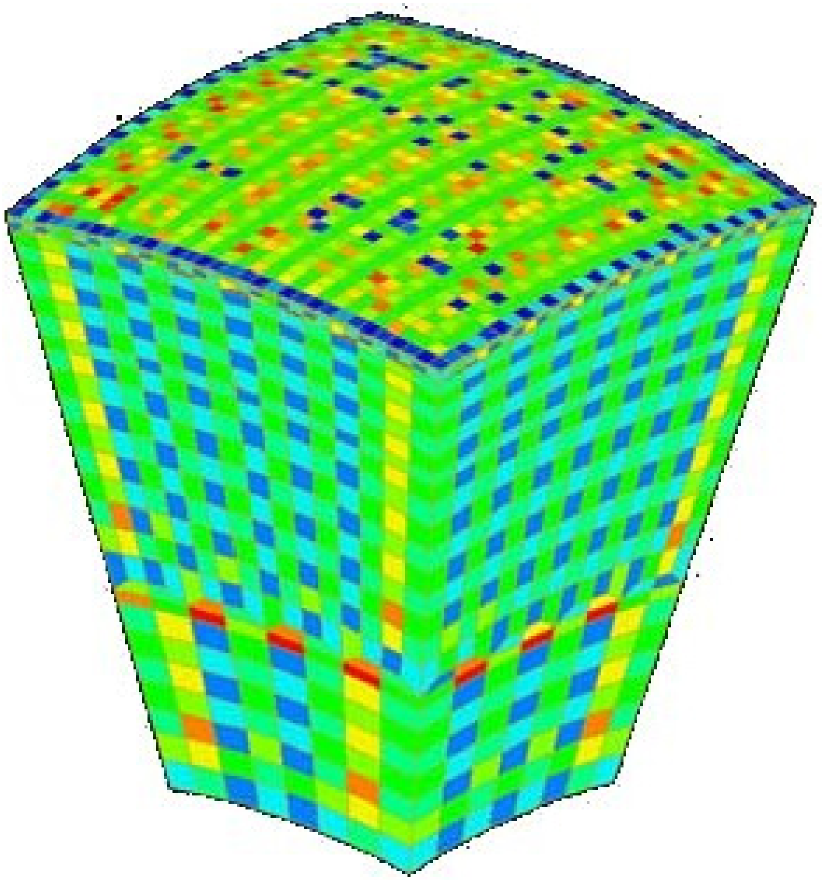
Can we use highly optimized BLAS matrix/matrix products (90% of computations)?

- For one element: matrices (5x25, 25x5, 5 x matrices of (5x5)), BLAS is not efficient: overhead is too expensive for matrices smaller than 20 to 30 square.
- If we build big matrices by appending several elements, we have to build 3 matrices, each having a main direction (x,y,z), which causes a lot of cache misses due to the global access because the elements are taken in different orders, thus destroying spatial locality.
- Since all arrays are static, the compiler already produces a very well optimized code.

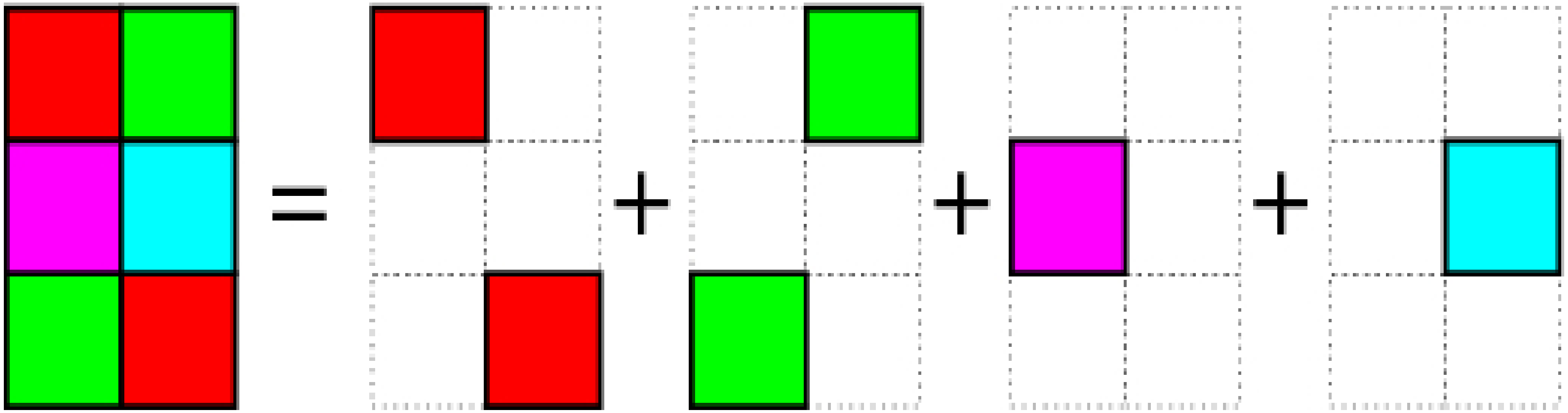
⇒ No need to, and cannot easily use BLAS

⇒ Compiler already does an excellent job for small static loops

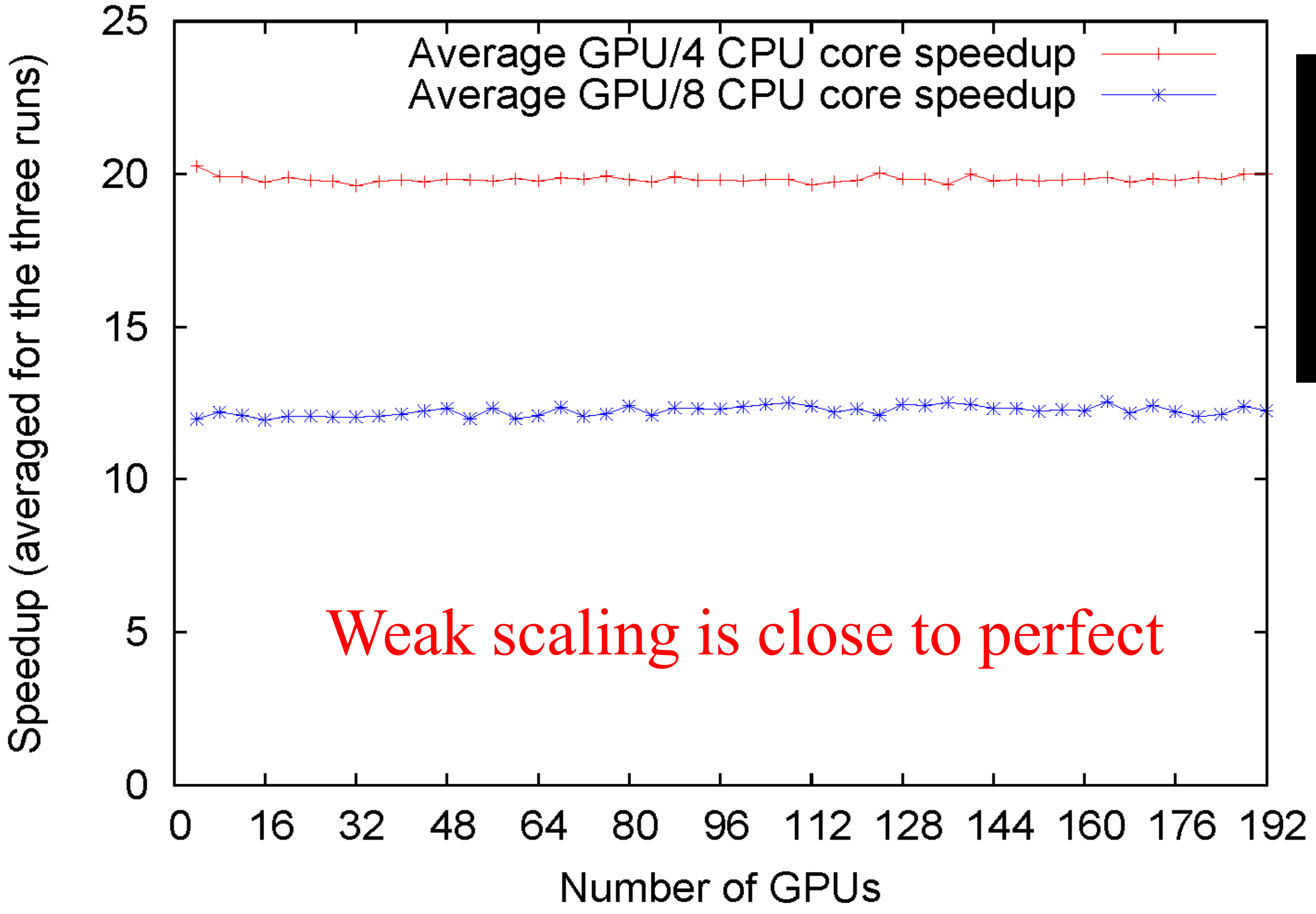
Porting to GPUs: mesh coloring



- **Key challenge: ensure that contributions from two local nodes never update the same global value from different warps**
- **Use of mesh coloring: suppress dependencies between mesh points inside a given kernel**
- **Use of "atomic" leads to slower code**



Multi-GPU weak scaling (up to 192 GPUs)



Weak scaling is close to perfect

Prix Bull - Joseph Fourier

en association avec GENCI

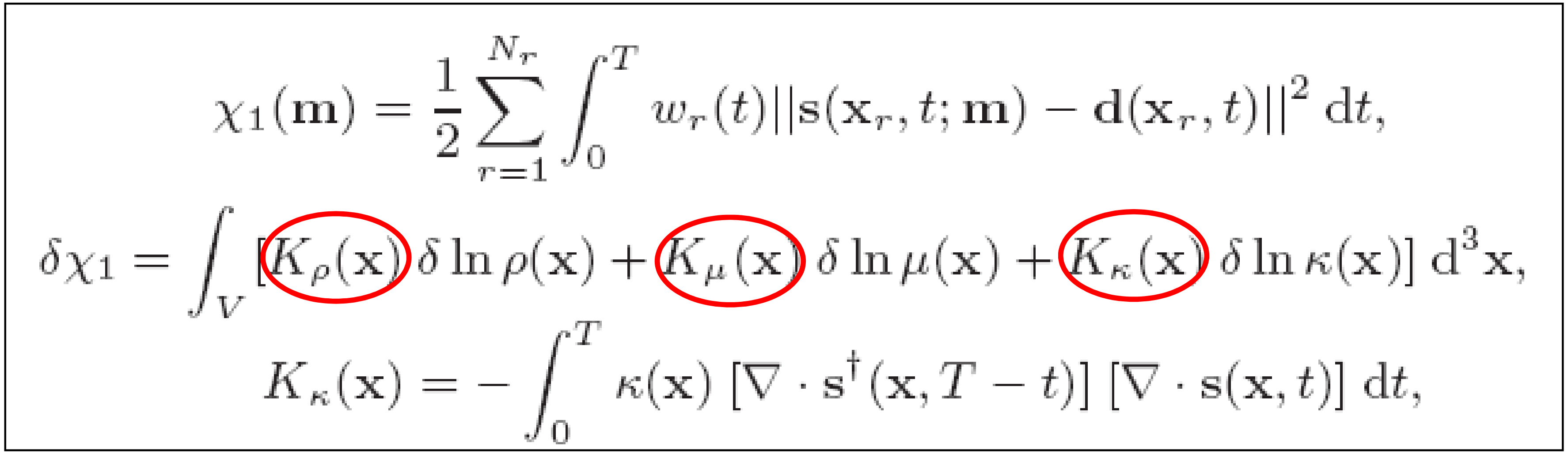
Pour le développement de
la simulation numérique

High-frequency ocean acoustics, inverse problems in seismology, acoustic tomography, reverse-time migration in seismics: high resolution needed, and/or large iterative problems to solve ⇒ Large calculations to perform.

⇒ GPU computing: code needs to be rewritten, but large speedup can be obtained (around 20x-30x for our finite-element codes, but it is difficult to define speedup).

Adjoint methods for tomography and imaging

Problem is self-adjoint, thus no need
for automatic differentiation (AD, autodiff)



$$\chi_1(\mathbf{m}) = \frac{1}{2}\sum_{r=1}^{N_r}\int_0^T w_r(t)||\mathbf{s}(\mathbf{x}_r,t;\mathbf{m}) - \mathbf{d}(\mathbf{x}_r,t)||^2\,\mathrm{d}t,$$

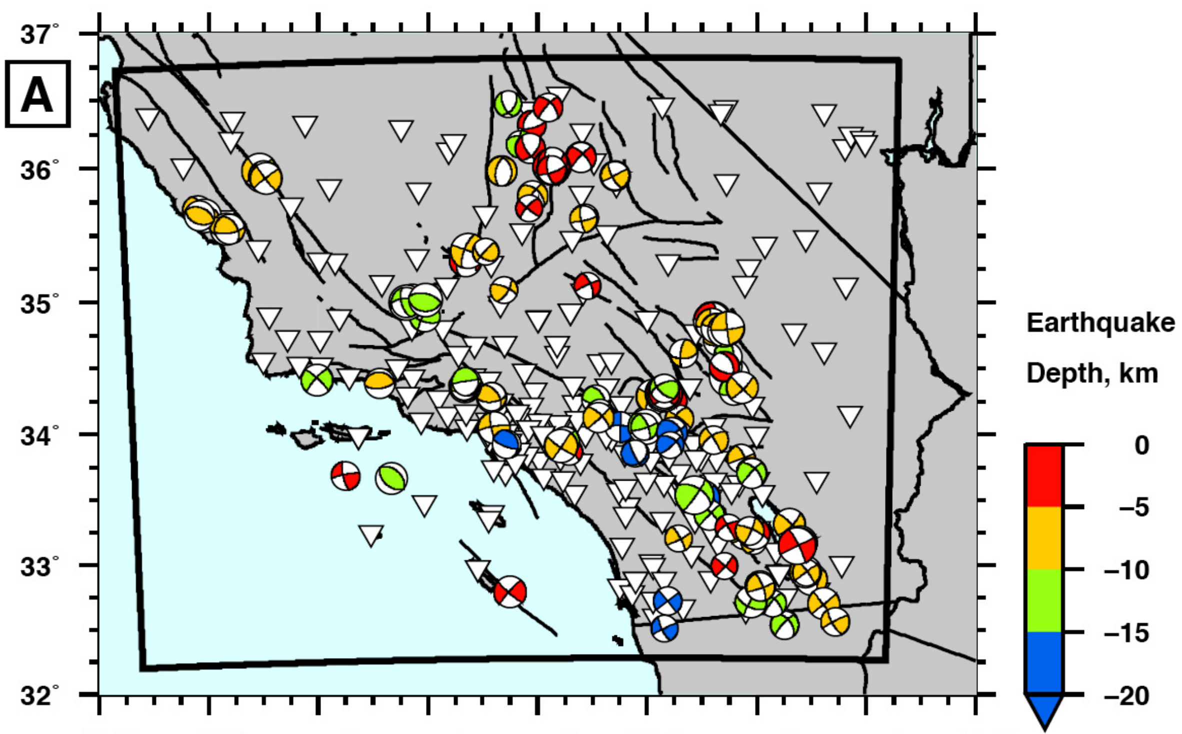
$$\delta\chi_1 = \int_V [K_\rho(\mathbf{x})\,\delta\ln\rho(\mathbf{x}) + K_\mu(\mathbf{x})\,\delta\ln\mu(\mathbf{x}) + K_\kappa(\mathbf{x})\,\delta\ln\kappa(\mathbf{x})]\,\mathrm{d}^3\mathbf{x},$$

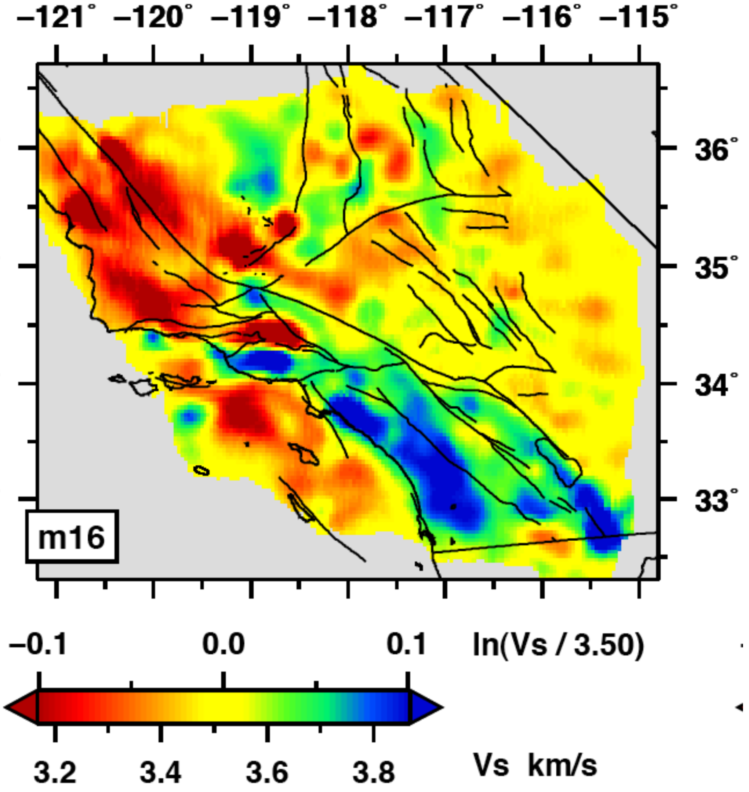
$$K_\kappa(\mathbf{x}) = -\int_0^T \kappa(\mathbf{x})\,[\nabla\cdot\mathbf{s}^\dagger(\mathbf{x},T-t)]\,[\nabla\cdot\mathbf{s}(\mathbf{x},t)]\,\mathrm{d}t,$$

Theory: A. Tarantola, Talagrand and Courtier.
'Banana-Donut' kernels (Tony Dahlen et al., Princeton)
Close to time reversal (Mathias Fink et al.) but not identical,
thus interesting developments to do.

Idea: apply this to tomography of the full Earth
(current ANR / NSF contract with Princeton University, USA), and in acoustic
tomography: ocean acoustics, non destructive testing.

Tape et al. (2009): 143 earthquakes used in inversion

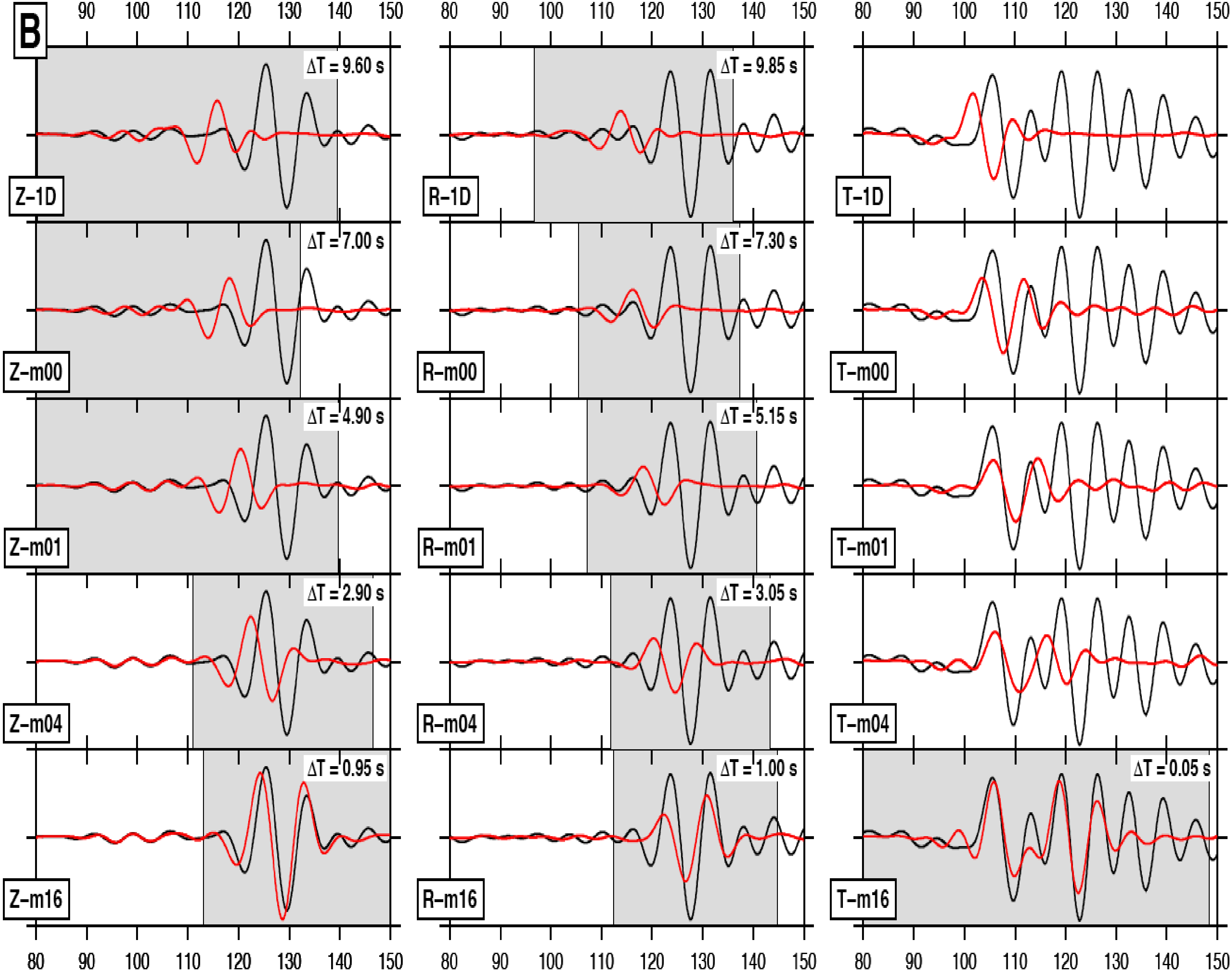




Depth 10 km

- 3 simulations per earthquake per iteration
- 16 iterations
- 6,864 simulations
- 168 processor cores per simulation
- 45 minutes of wall-clock time per simulation
- 864,864 processor core hours

Princeton, USA



Princeton, USA

L-BFGS method

- Iterative Gauss-Newton algorithm

$$m_{k+1} = m_k + \underbrace{(G_k^t G_k + C_m^{-1})^{-1} \nabla J(m_k)}_{\delta m_k}$$

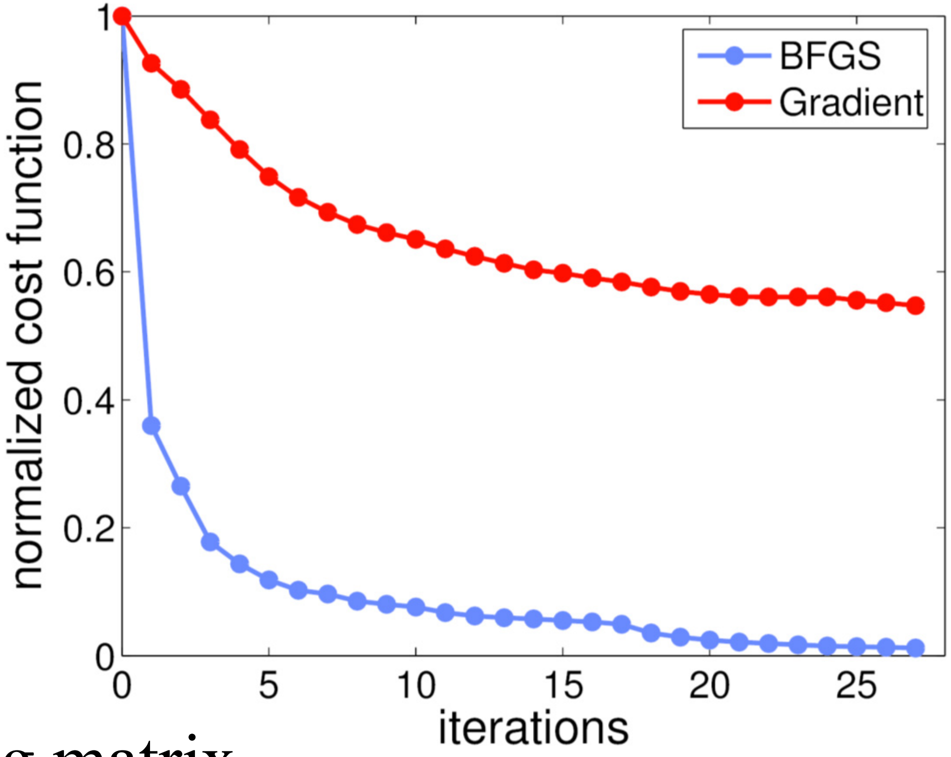
- L-BFGS (Low-memory Broyden–Fletcher–Goldfarb–Shanno): Approximate

$$\delta m_k$$

from:

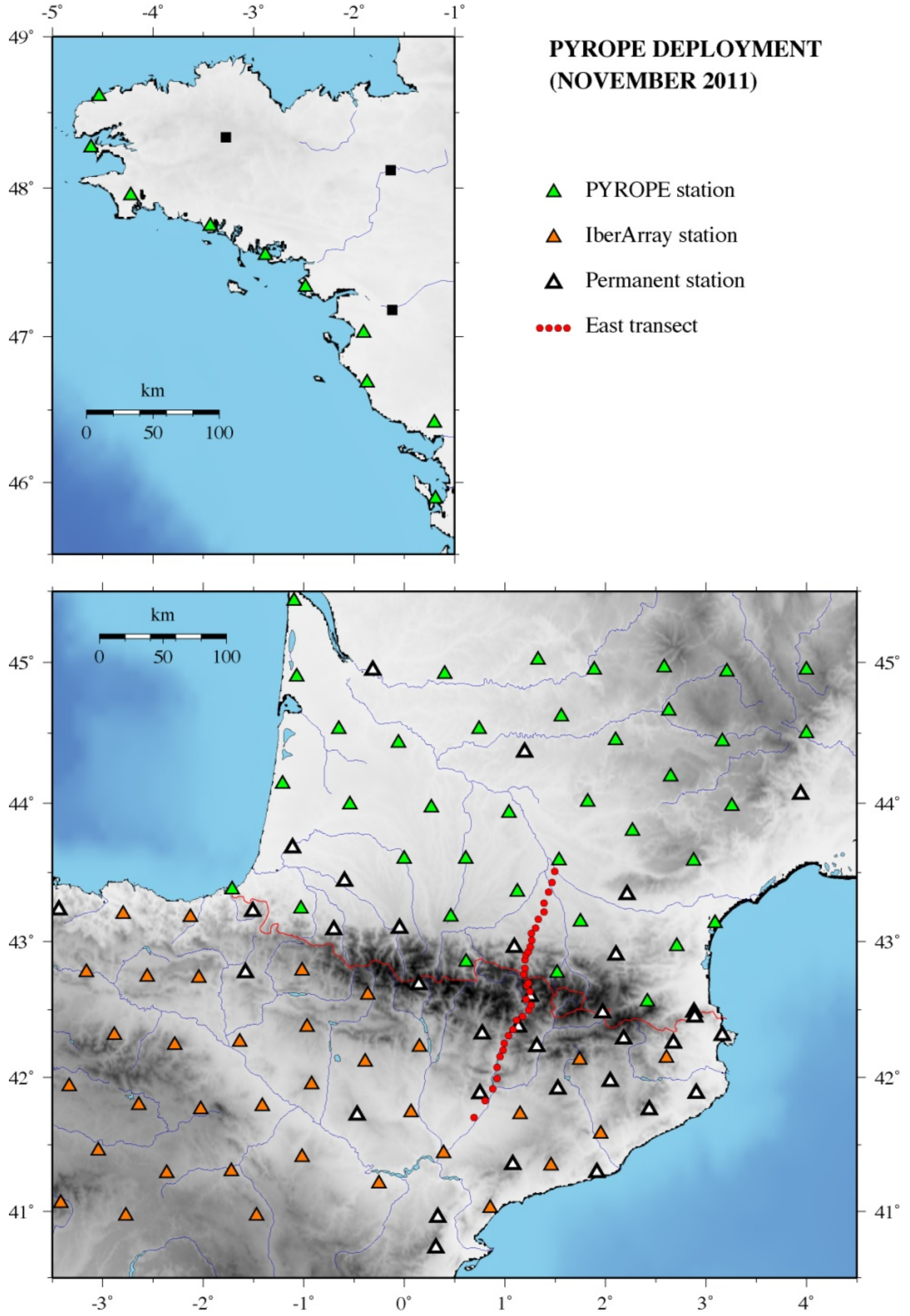
$$m_{k-1}, m_{k-2}, m_{k-3}, \ldots, m_0$$

$$\nabla J(m_{k-1}), \nabla J(m_{k-2}), \ldots, \nabla J(m_0)$$



→ no need to invert or even build a big matrix

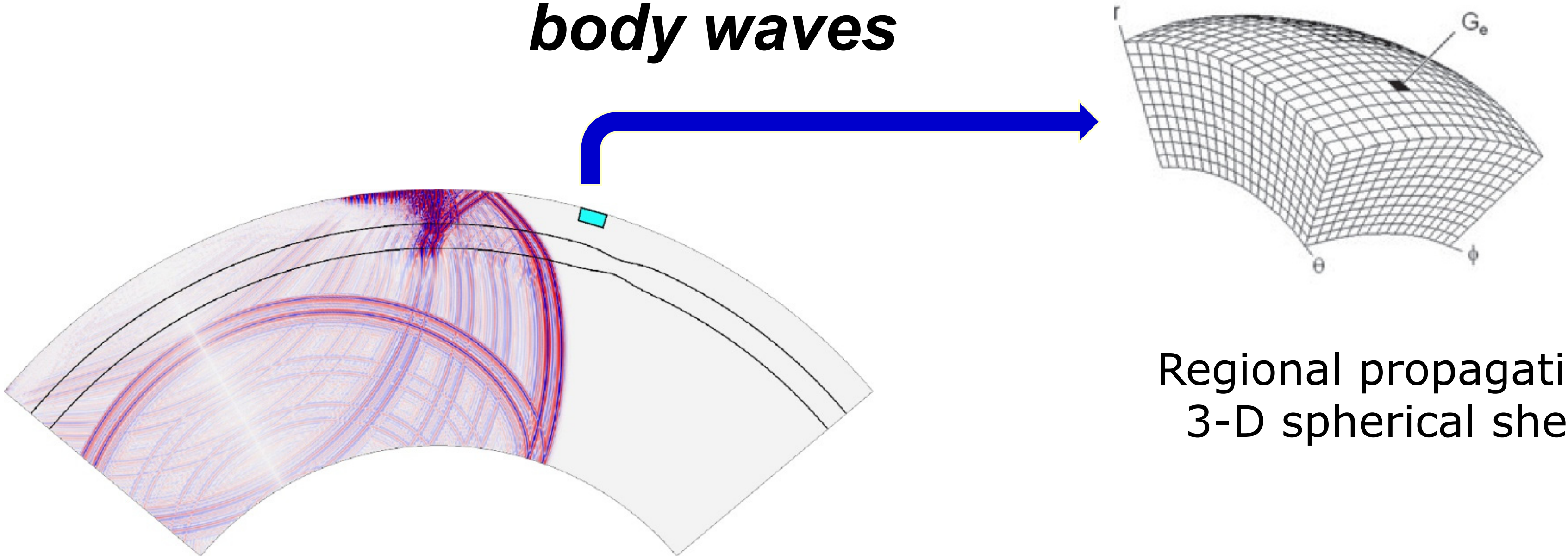
The PYROPE experiment



- ➢ French/Spanish initiative, supported by the French ANR
- ➢ ~150 temporary + 50 permanent BB stations
- ➢ Interstation spacing ~ 60 km
- ➢ Dense transects across the Pyrénées

A hybrid approach: Coupling global and regional propagations

A hybrid technique for 3-D waveform modeling and inversion of high frequency teleseismic body waves

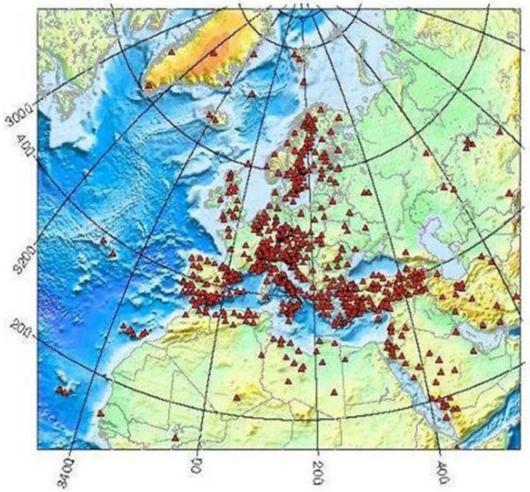


Regional propagation
3-D spherical shell

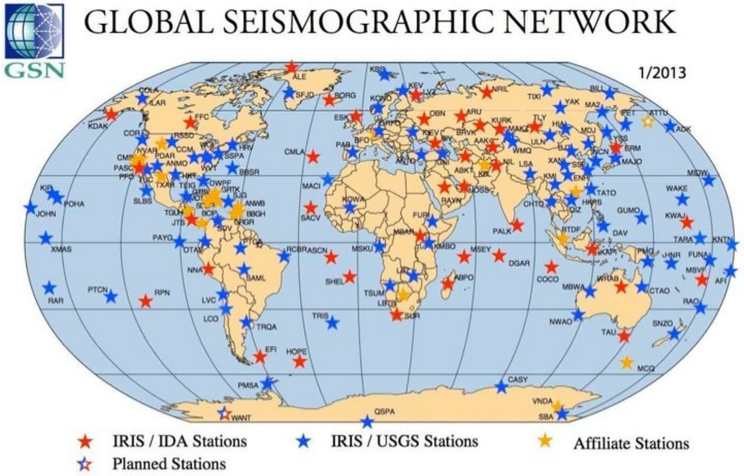
Global propagation
Spherically symmetric Earth model

S. Chevrot, V. Monteiller, D. Komatitsch & N. Fuji
Geophysical Journal International, 2014

Data in Regional & Global Seismology

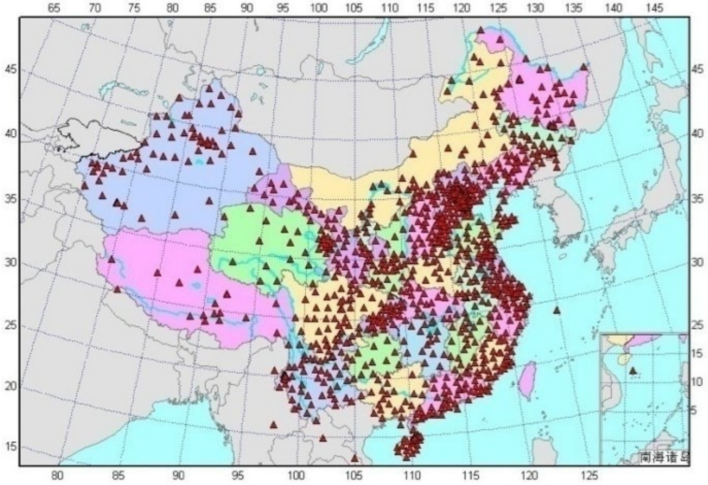


[www.geo.uib.no]

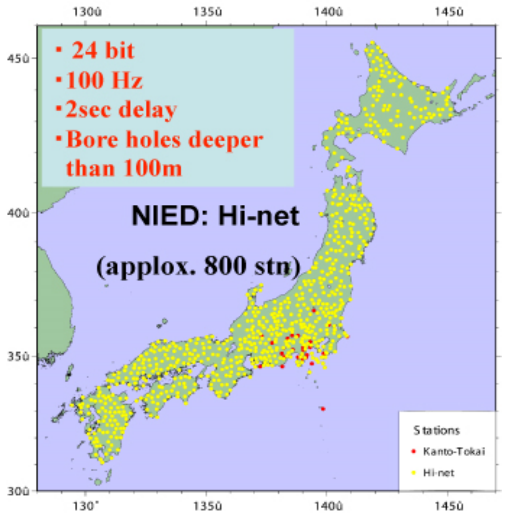


[www.iris.edu]

[web.mst.edu]

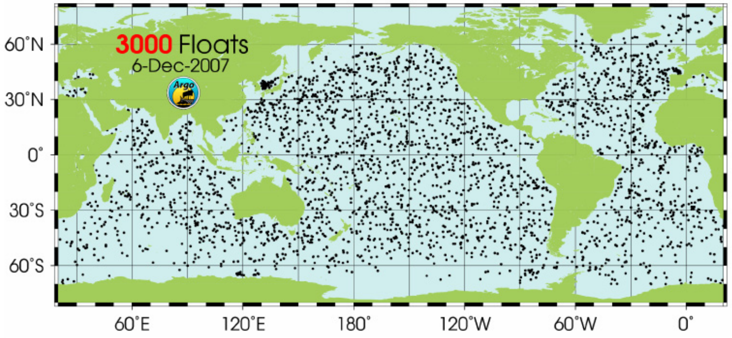


[data.earthquake.cn]



[drh.edm.bosai.go.jp]

MERMAID/MariScope



[Simons et al, 2006]

Data in Exploration Seismology

3D marine surveys can involve 5,000 shots and 50,000 recorders

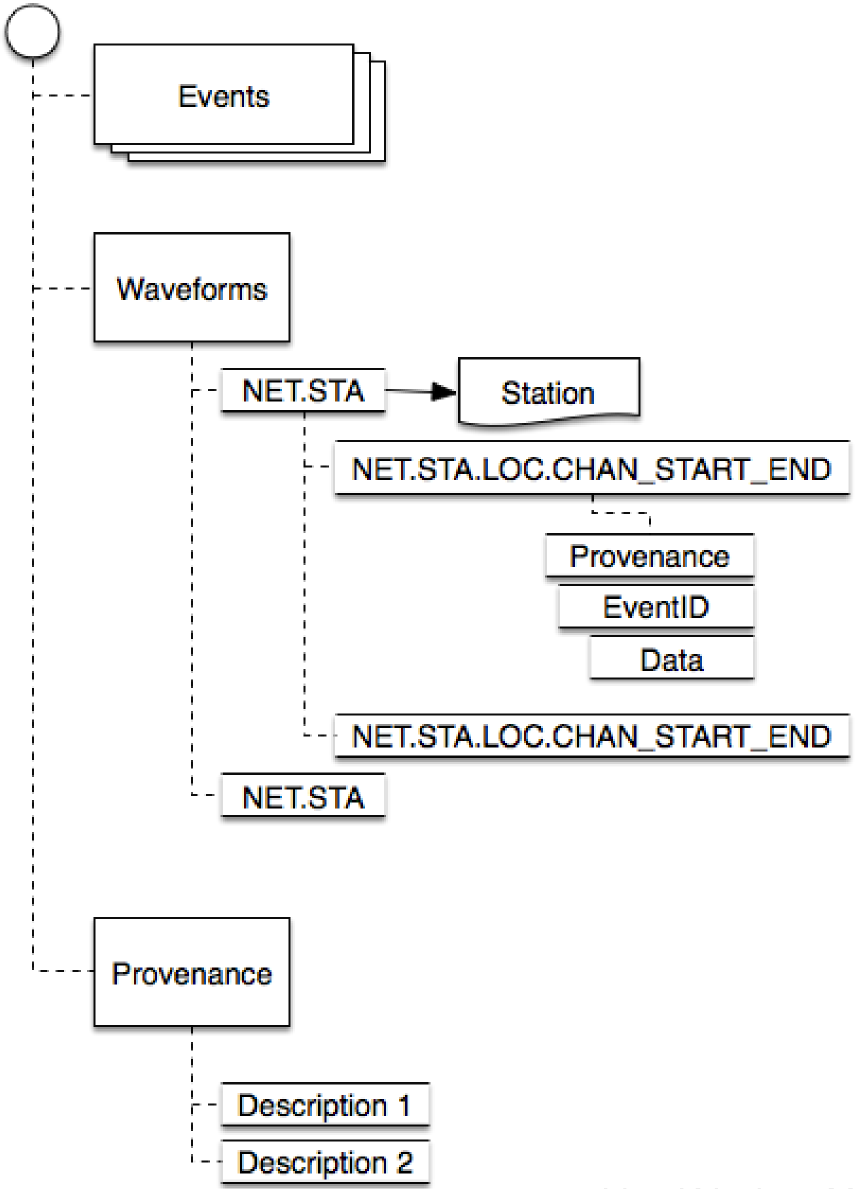
- Petabytes of data
- SEG-Y is the current standard
- Variable SEG-Y file structure
- SEG-Y programs do not always follow specifications

http://www.earth.northwestern.edu/people/seth/107/Seismology/mcsresult.htm

ASDF: an Adaptable Seismic Data Format

- Collaboration involving Princeton University, Munich University (ObsPy) and Oak Ridge National Laboratory
- Increase I/O performance by combining all the time series for a single shot or earthquake into one file
- Take advantage of parallel processing
- Use modern file format as container (e.g., HDF5 or ADIOS)
- Store provenance inside the file for reproducibility
- Use existing standards when possible (e.g., XML)

ASDF Internal Structure



Events

Waveforms

NET.STA

Station

NET.STA.LOC.CHAN_START_END

Provenance

EventID

Data

NET.STA.LOC.CHAN_START_END

NET.STA

Provenance

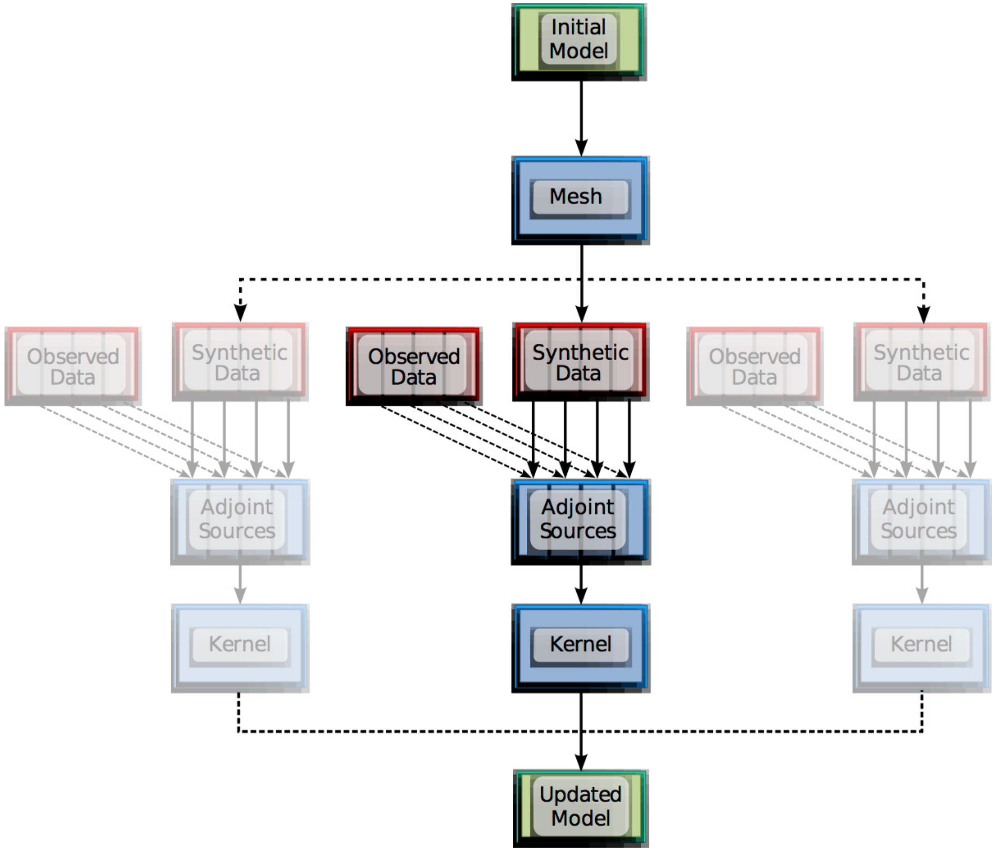
Description 1

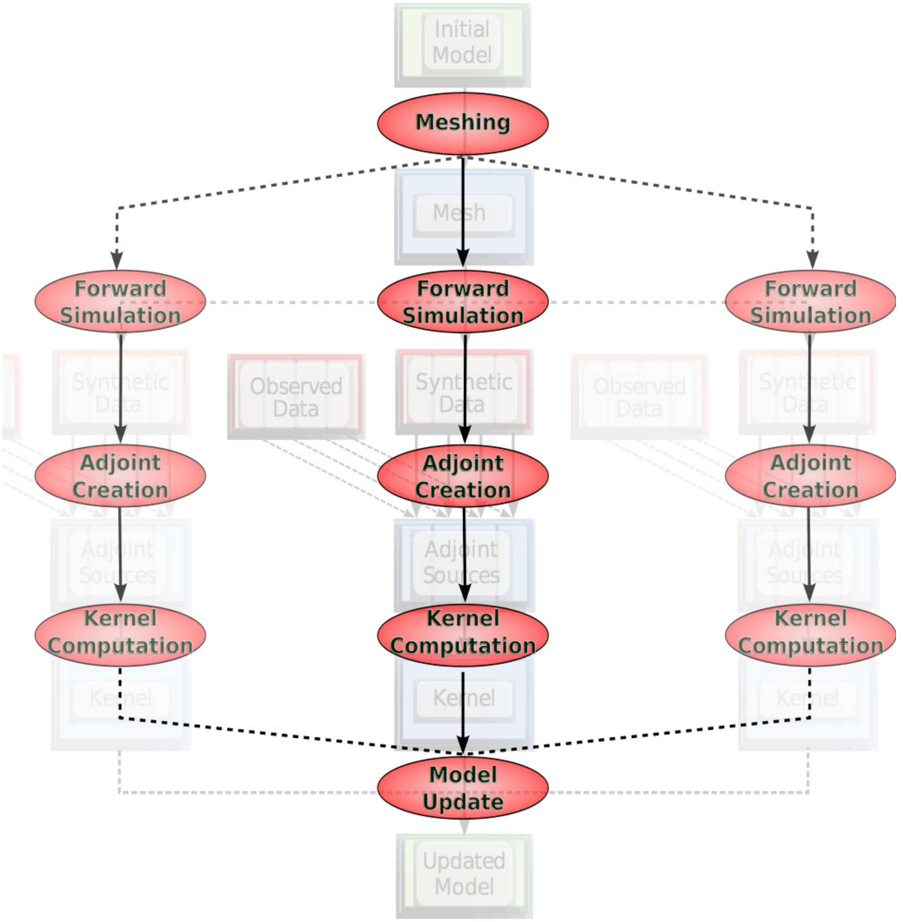
Description 2

1000 Stations	Number of SAC Files	Number of ADIOS Files
255 Earthquakes	1,275,00	255
6,000 Earthquakes	30,000,000	6,000

Lion Krischer, Munich University, Germany

Data and processing flows





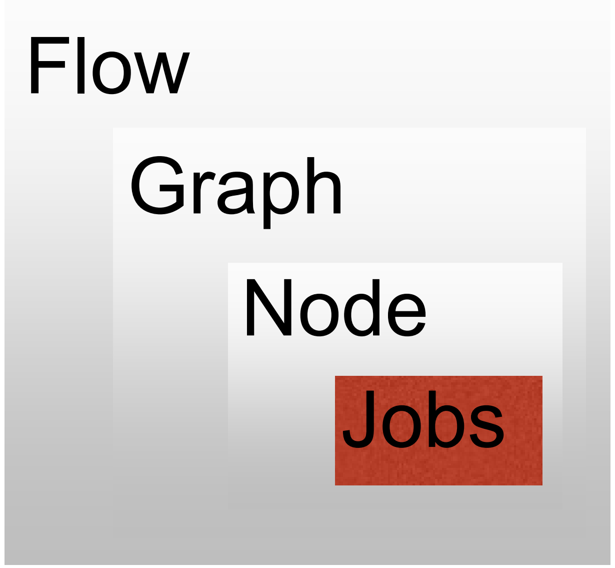
Each processing step contains more than one atomic operation

Workflow Requirements

- Least action
  - The generic workflow should run with a minimum number of steps
  - Minimum user interaction to modify the workflow
- High abstraction level
  - Job description and dependencies should stay simple
  - Computational details should be hidden

Linking Jobs: FlowJobs

- Flow
  - Organize jobs in a DAG
  - Keep track of jobs' status
    - Done, Failed, Scheduled, Pending, …
  - State can be saved to be relaunched
- Jobs encapsulate
  - SAGA-Python job definition
  - Pure python functions
  - (BigJob job definition)
- Targets stand-alone clusters, will be extended to allow distributed environments


Flow

Graph

Node

Jobs

FlowJobs

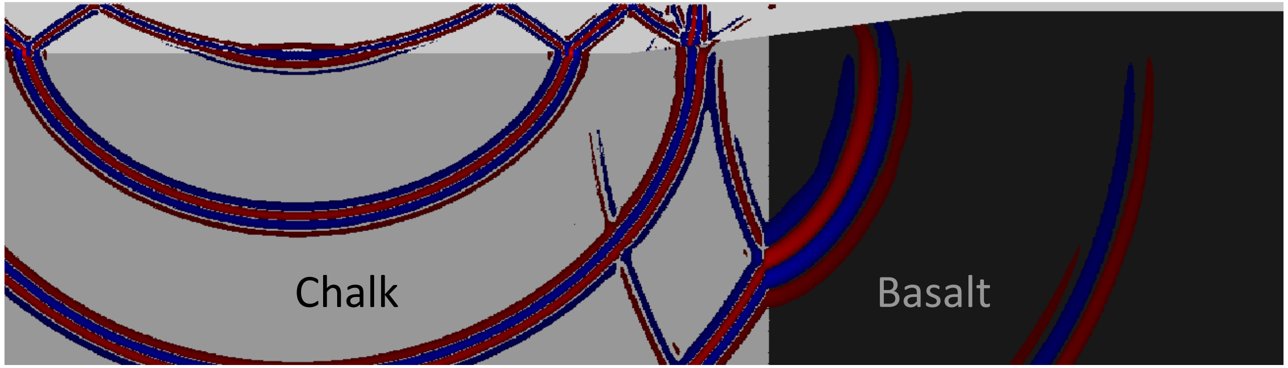
Ocean acoustics

Numerical simulation

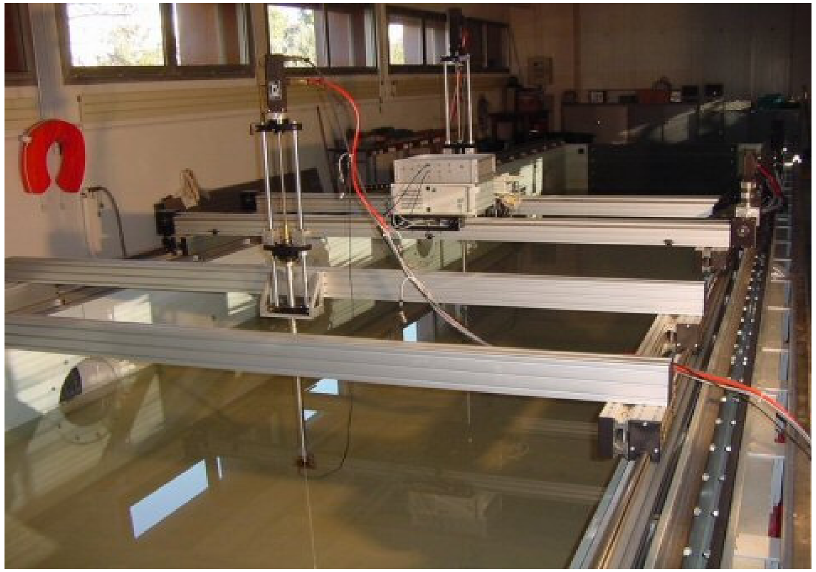
Collaboration with Paul Cristini.

Wave propagation across an impedance discontinuity.

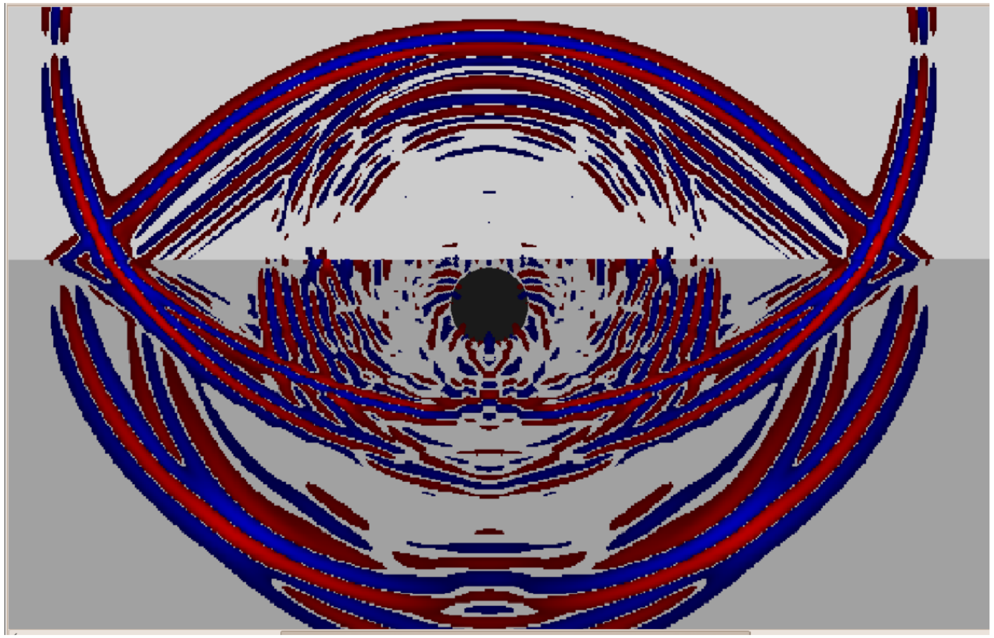
Influence on interface waves.



Experiments performed in tanks



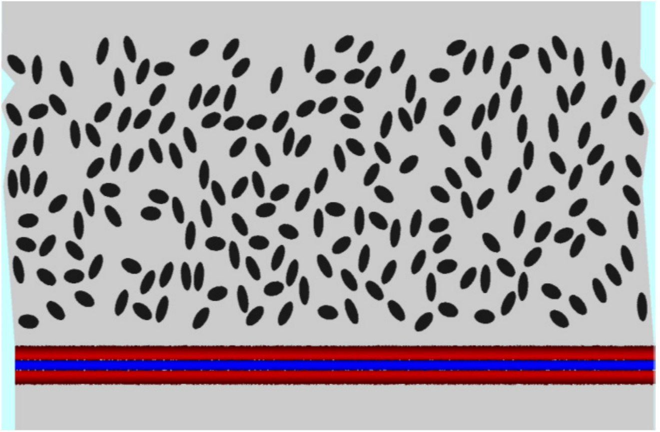
Experimental tanks in Marseille

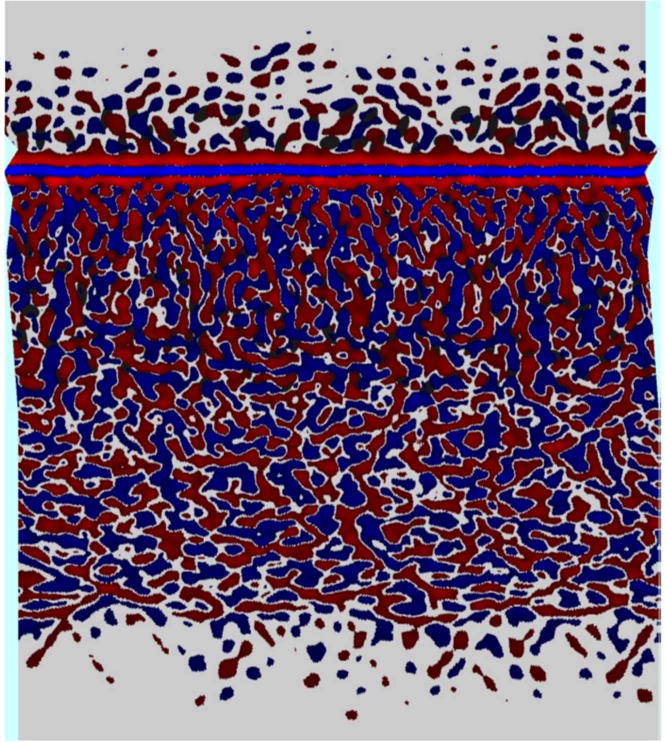


Experiments in known environment / setup

Perform experimental benchmarks

Non destructive testing of materials





Collaboration with Non Destructive Testing Lab in Marseille.

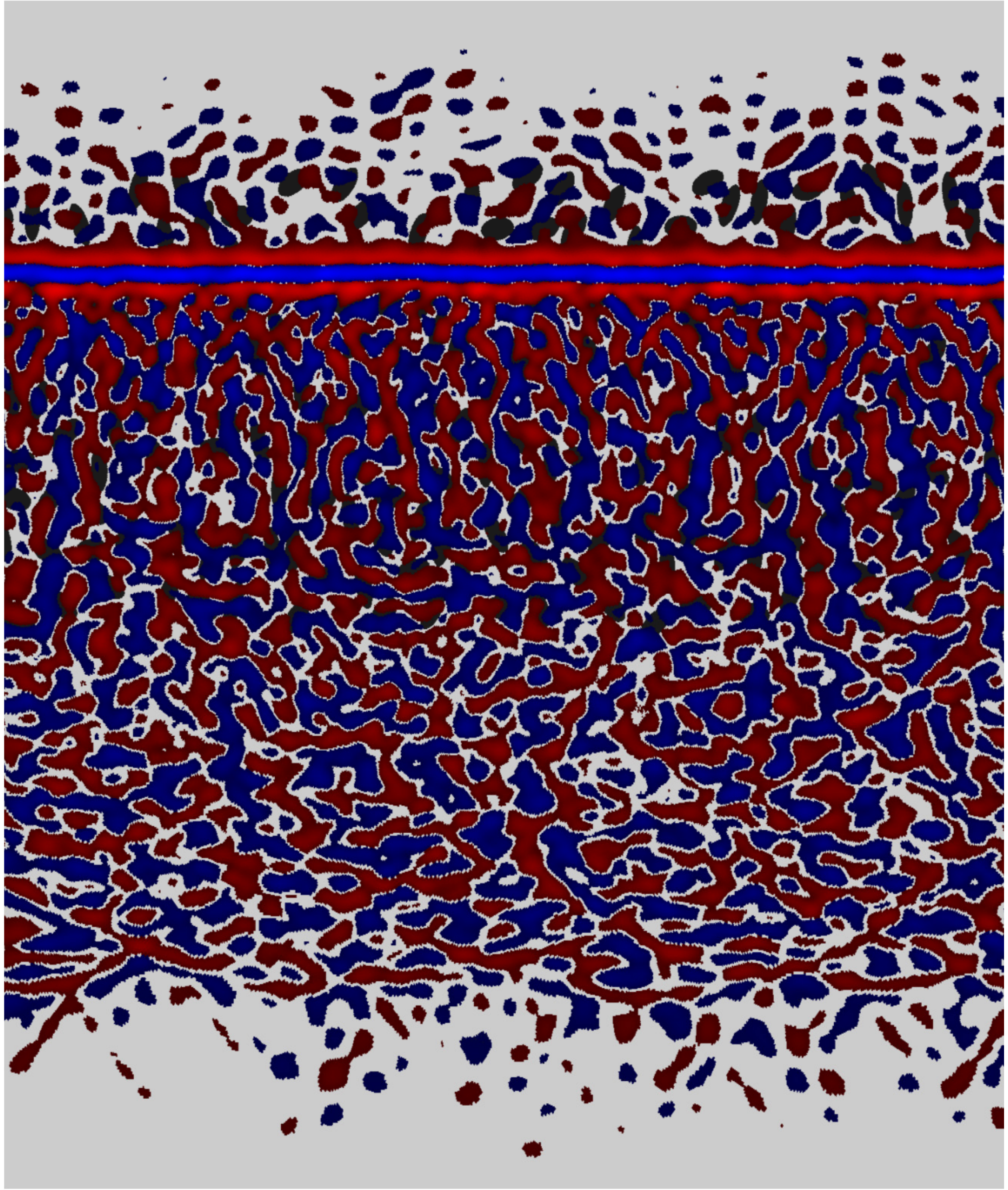
Currently: Physical modeling based on diffusion functions for objects of complex shape, cracks or multiple cavities in concrete, metals, or composite materials. Experiments on samples.

Very accurate calculations without homogenization can validate (or not) these diffusion functions and extend them beyond their domain of validity.

Reliable modeling of the “coda” part of the signal, which contains useful information on the medium.

Movie of the numerical simulation of multiple diffusion in a concrete block

Plane shear wave incident on the small rocks that are present in the concrete block



Conclusions and future work

- On modern computers, large 3D full-waveform forward modeling problems can be solved at high resolution in the time domain for acoustic / elastic / viscoelastic / poroelastic / seismic waves
- Inverse (adjoint) tomography / imaging problems can also be studied, although the cost is still high
- Useful in different industries in addition to academia: oil and gas, medical imaging, ocean acoustics / sonars, non destructive testing (concrete, composite media, fractures, cracks)
- Hybrid (GPU) computing is useful to solve inverse problems in seismic wave propagation and imaging
- PRACE project with INGV Roma to image the Italian lithosphere: 40 million core hours on a petaflop machine
- Some future trends: high-frequency ocean acoustics, tomography of buried objects, wavelet compression

About the path to exascale

- We are highly interested and involved in the effort, but we are not 100% experts (we are in acoustics or geophysics labs, not computer science)
- In most cases we will run hundreds of semi-independent runs on different parts of the machine rather than a single big run; sharing or processing data then becomes a big issue ; big data can also become an issue (see VERCE project)
- We are in the process of adding OpenMP support in addition to MPI; not too challenging in our application, only a few critical routines impacted
- We tried higher-level directive models (OpenAcc, StarSs and OmpSs from Barcelona BSC). So far the code we get is always significantly slower than our pure MPI code, but the programming models are flexible and interesting
- We successfully used GPUs
- INRIA (Franck Cappello) and we added fault-tolerant MPI (SC'11 paper)
- We also recently used ARM boards (MONTBLANC European project) to target lower energy-to-solution models.

The SPECFEM3D code is freely available open source at http://www.geodynamics.org