

# TeraGrid- Successes and Functionalities from a User's Perspective

---

Ralph Roskies

Scientific Director, Pittsburgh Supercomputing Center

Professor of Physics, U. of Pittsburgh

May 10, 2005



# What is TeraGrid?

---

Arden Bement (Director of NSF) to President's Information Technology Advisory Committee (PITAC) (July/04)

“...the Teragrid—a next step in an evolutionary process that makes a rich mix of **cyberinfrastructure** resources broadly available to the science and engineering community.... pioneering entirely new ways of conducting investigations, from **computationally intensive** to large **data-driven** applications. The power and flexibility of these tools, combined with **progress in data collection** and observational tools, from sensors to satellites, continues to draw **new communities** of researchers into the computational fold.”



# TeraGrid Vision

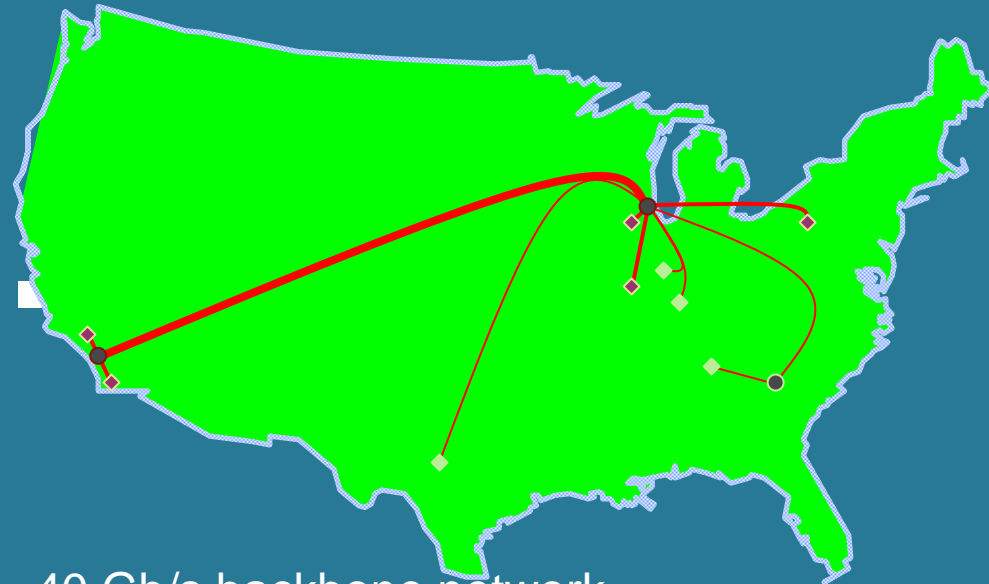
---

- TeraGrid *DEEP: Enabling the Nation's Terascale Science*
  - Make Science More Productive through a unified set of very-high capability resources.
- TeraGrid *WIDE: Empowering communities to leverage TeraGrid capabilities*
  - Bring TG capabilities to the broad science community (no longer just “big” science).
- Base TeraGrid Cyberinfrastructure: *Integrating the Nation's Most Powerful Resources*
  - Provide a unified, general purpose, reliable set of services and resources.



# TeraGrid Partners

- PSC, NCSA, SDSC (major resources)
- Argonne, (software, visualization)
- TACC (computation and visualization)
- Caltech, IU, Purdue (data collections, science gateways)
- ORNL (spallation neutron source)



40 Gb/s backbone network  
30 Gb/s to major computational resources  
10 Gb/s to ORNL, IU, Purdue, TACC



# Aggregate capability

---

- ~ 50 TF computational power
- ~ 1.5 PB on-line storage
- ~15 PB archival storage
- Visualization engines



# Unique resources- heterogeneous

---

- Tightly coupled machines like 6 TF 3000 processor alpha-based Terascale Computing System with 2 rails of Quadrics, and XT3
- Shared memory machines like the Altix- (2 512 processor Itanium processors, total 6.5 TF ; one has 1 TB memory, one has 2TB) and Marvel (EV7 based with outstanding memory bandwidth)
- 10 TF IA64 cluster, Myrinet linked (NCSA)
- 6.3 TF IA32 Dell cluster Myrinet linked (TACC)
- DataStar (1760 Power 4's) 10.4 TF, 3.8 TB of memory (SDSC)
- Online storage 600 TB (NCSA), 540 TB (SDSC)
- Instruments like SNS (Oak Ridge)
- Data collections



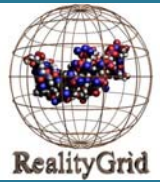
# Success stories

---

- Biology and materials science
- Oil Exploration
- Astronomy
- Earthquake Modeling and Civil Engineering
- Tornado Modeling



# UK – Teragrid HPC-Grid Experiment

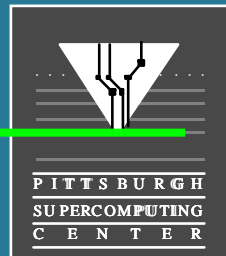


TeraGyroid: Lattice-Boltzmann simulations of defect dynamics in amphiphilic liquid crystals

- Peter Coveney (University College London),
- Richard Blake (Daresbury Lab)
- Stephen Pickles (Manchester).
- Bruce Boghosian (Tufts)



ANL



# Project Partners

---

## Reality Grid partners:

- University College London (Application, Visualisation, Networking)
- University of Manchester (Application, Visualisation, Networking)
- Edinburgh Parallel Computing Centre (Application)
- Tufts University (Application)

## UK High-End Computing Services

- HPCx run by the University of Edinburgh and CCLRC Daresbury Laboratory (Compute, Networking, Coordination)
- CSAR run by the University of Manchester and CSC (Compute and Visualisation)

## Teragrid sites at:

- Argonne National Laboratory (Visualization, Networking)
- National Center for Supercomputing Applications (Compute)
- Pittsburgh Supercomputing Center (Compute, Visualisation)
- San Diego Supercomputer Center (Compute)



# Defect pathways and dynamics in gyroid self-assembly

- Amphiphiles are chemicals with hydrophobic (water-avoiding) tails and hydrophilic (water attracting) heads. When dispersed in solvents or oil/water mixtures, self assemble into complex shapes; some (gyroids) are of particular interest in biology.
- Shapes depend sensitively on parameters like
  - abundance and initial distribution of each component
  - the strength of the surfactant-surfactant coupling,
- Desired structures can sometimes only be seen in very large systems. E.g. smaller region form gyroids in different directions and how they then interact is of major significance.
- Implications for drug delivery, prevention of crystallization in oil pipelines, design better shampoos



# Networking



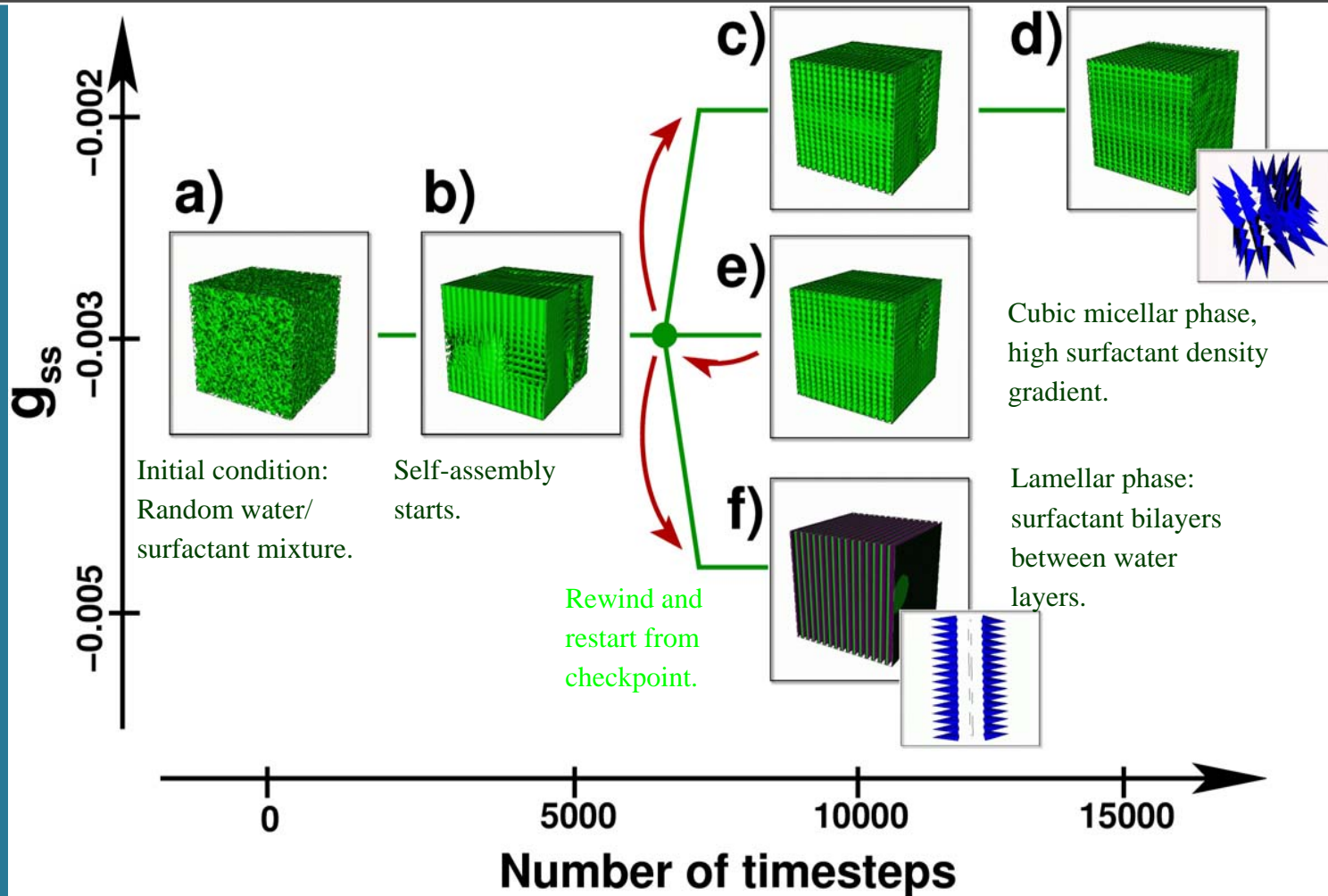
# Distribution of function

---

- Computations ran at HPCx, CSAR, SDSC, PSC and NCSA. (7 TB memory - 5K processors in integrated resource) One Gigabit of LB3D data generated per simulation time-step.
- $128^3$  at many places,  $512^3$  HPCx,  $1024^3$  TCS
- Visualisation ran at Manchester/ UCL/ Argonne
- Scientists steered calculations from UCL and Boston over Access Grid. Steering requires reliable near-real time data transport across the Grid to visualization engines.
- Visualisation output and collaborations multicast to SC03 Phoenix and visualised on the show floor in the University of Manchester booth



# Exploring parameter space through computational steering



# Results

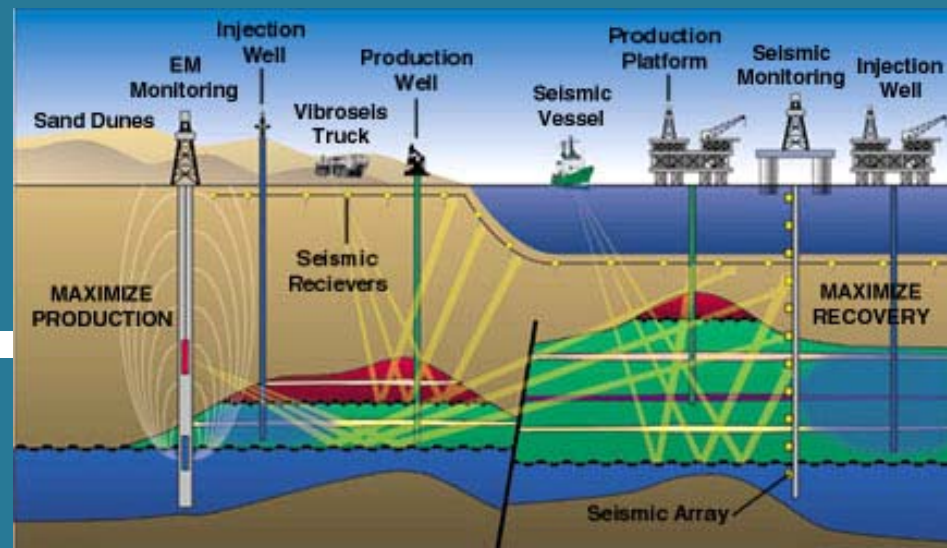
---

- Linking these resources allowed computation of the largest set of lattice-Boltzmann (LB) simulations ever performed, involving lattices of over one billion sites.



# Oil exploration

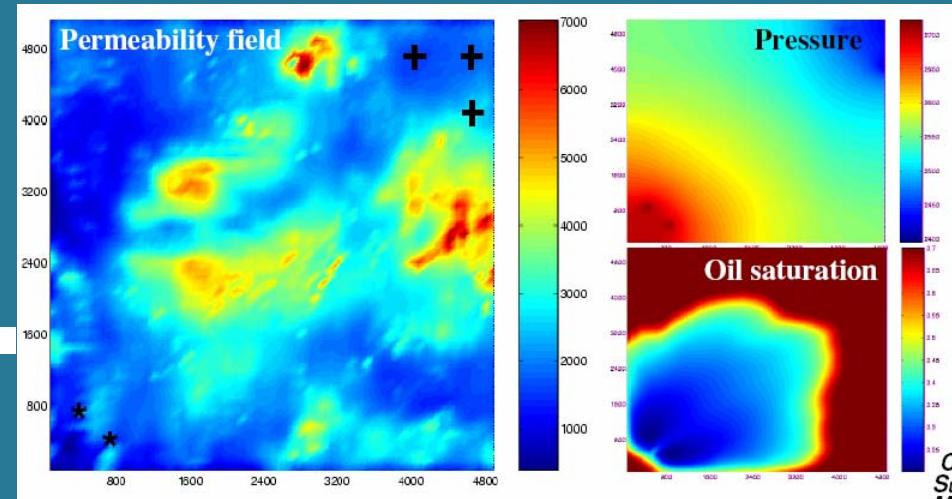
- Need intelligent ways to choose ideal places for drilling and to surmise the geological features below the surface.
- Team – Mary Wheeler (U. of Texas), Joel Saltz (Ohio State U.), Manish Parashar (Rutgers)
  - Pick drilling site;
  - Solve approximate model of fluid dynamics (IPARS, Wheeler) to estimate oil extraction (each run takes about 2 hours, one processor);
  - Move equipment around and repeat, and then optimize



Middleware tools (STORM, DataCutter) from Saltz's team manage the large amounts of distributed data. Dynamic steering and collaborative tools (Parashar) allow on the fly adaptive refinements.

# Oil exploration

- Bruce Loftis at NCSA built private toolkit that allow him, from his workstation to examine queues on remote machines, submit jobs under Condor to them, and monitor which have completed, which have failed, which are running
- Ran 25,000 reservoir optimization single processor runs in less than a week (averaging over 300 jobs at all times)
- Used TeraGrid systems at NCSA, SDSC, Caltech, and TACC which all had network accessible nodes
- Managing 8 TB of data at NCSA, >30TB at OSC



# GridShell

---

- Ed Walker (TACC) developed GridShell which among other things allows machines like TCS, without individual network-accessible nodes, to act as a single Condor pool.



# National Virtual Observatory

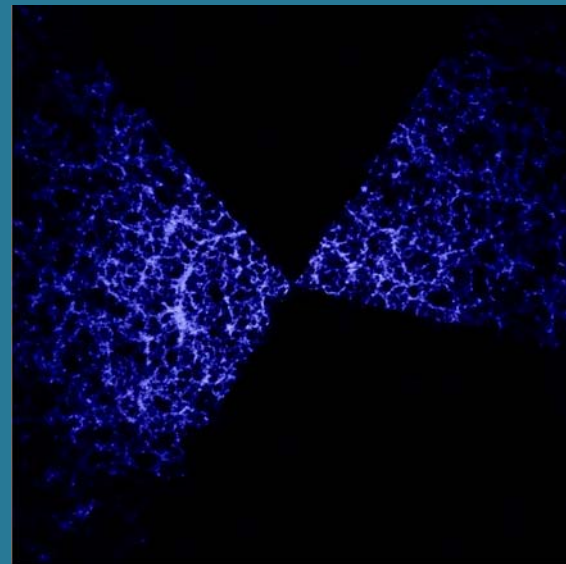
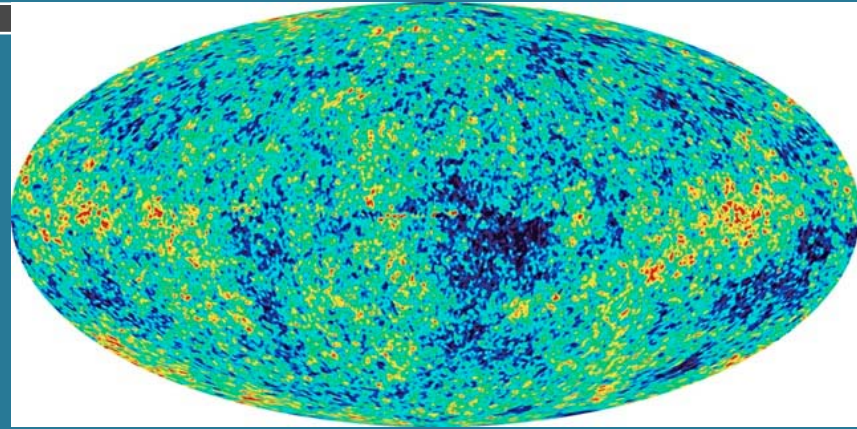


- Breakthroughs in telescope, detector, and computer technology allow astronomical surveys to produce terabytes of images and catalogues, in different wavebands, from gamma- and X-rays, optical, infrared, through radio.
- Soon it will be easier to "dial-up" a part of the sky than wait many months to access a telescope.
- Need multi-terabyte on-line databases interoperating seamlessly, interlinked catalogues, sophisticated query engines
- *research results from on-line data will be just as rich as that from "real" telescopes*



# Two critical datasets

- Wilkinson Microwave Anisotropy Probe – temperature of the Cosmic Microwave Background
- Sloan Digital Sky Survey-map one-quarter of the sky, determining the positions and absolute brightnesses of more than 100 million celestial objects



Dark matter density distribution  
inferred from SDSS catalog

# Understanding Dark Energy

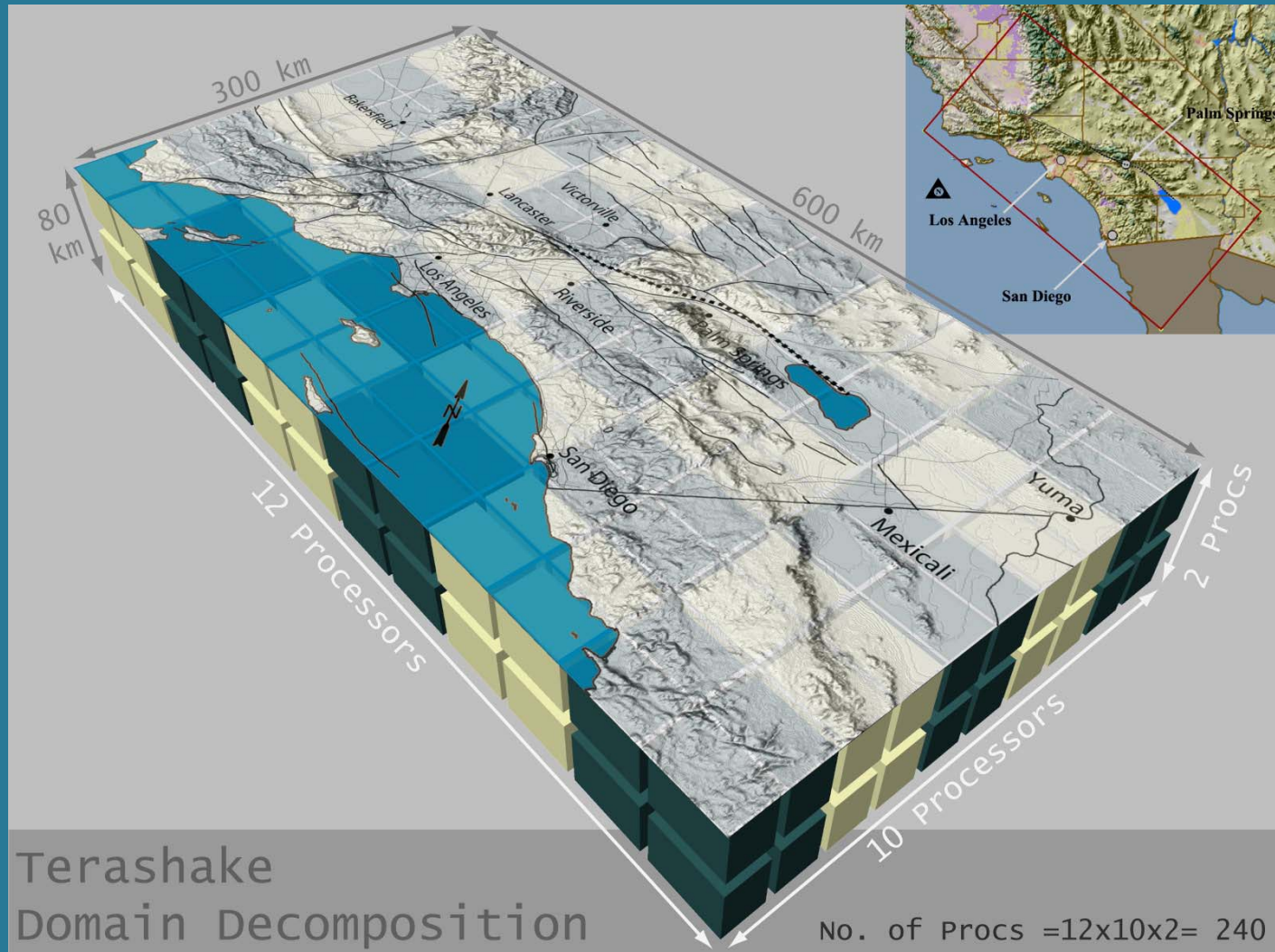
---

- Integrated Sachs Wolfe effect –dark energy makes light from the CMB blue-shifted when passing by large-scale structure (galaxies, clusters, filaments)
- Detecting this requires large galaxy survey comparing CMB with SDSS
- Assessing errors on the measurements due to coincidental overlaps requires generating tens of thousands of statistically similar realizations of the CMB and cross correlating them with the SDSS.
- Andy Connolly and Ryan Scranton (U. of Pittsburgh) and Jeff Gardner(PSC) used GridShell for this work. Ran on TCS at PSC, IA64 at NCSA and IA32 on TACC.



# Earthquake Modeling-SCEC (Southern California Earthquake Center)

- 48 organizations
- Develop seismic hazard assessment for southern CA for use by FEMA



# Simulated magnitude 7.7 earthquake in southern part of San Andreas fault

- Bernard Minster (Scripps Oceanographic Institute) et al simulated magnitude 7.7 earthquake (6 meters of ground motion)
- No earthquake here since 1680
- Resolution much better than previously possible, allow capturing whole basin and frequencies of interest to structural engineers (up to 0.5 Hz)
- Generated 43 TB; using TeraGrid SAN moved 10TB/day to Storage Resource Broker at SDSC
- Data is now being examined (mined) by other groups



240 processors of DataStar at SDSC (IBM Power 4)  
AWM (anelastic wave propagation model)  
-Kim Olsen (SDSU)  
Fourth order parallel finite difference code  
**1.8x10<sup>9</sup> grid points, regular mesh**



# New Insights

---

- First time able to model earthquake this big (7.7)
- Found directivity effects larger than anticipated, because seismic energy is guided by geological structures subparallel to San Andreas fault
- A rupture propagating from NW to SE would generate huge amplitudes and long duration in Mexicali, which now has 2 million people and a fragile building environment. This would be a major disaster.
- Next steps to model the earthquake sources better, and to add models of built structures to understand impact of such long shaking on these buildings.



# Carnegie Mellon Adaptive Finite Element Code

---

- Multiple spatial scales
  - wavelengths vary from  $O(10\text{m})$  to  $O(1000\text{m})$
  - Basin/source dimensions are  $O(100\text{km})$
- Highly irregular basin geometry
- Highly heterogeneous soils material properties
- So need unstructured grids even though good parallel performance harder to achieve



# Performance of forward earthquake modeling code on PSC Terascale system

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

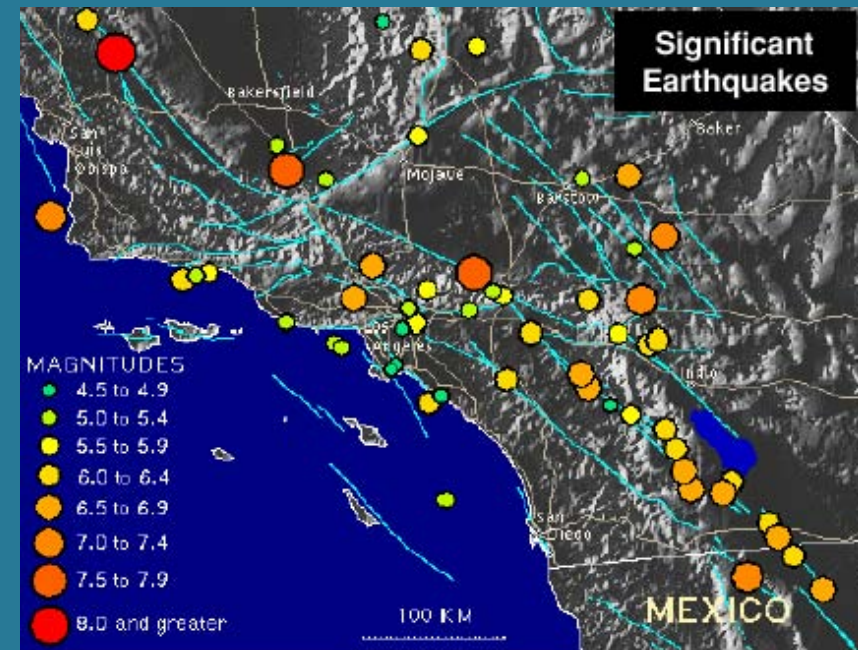
## Largest simulation

- 28 Oct 2001 Compton aftershock in Greater LA Basin
- **maximum resolved frequency: 1.85Hz**
- 100m/s min shear wave velocity
- physical size: 100x100x37.5 km<sup>3</sup>
- # of elements: 899,591,066
- **# of grid points: 1,023,371,641**
- # of slaves: 125,726,862
- 25 sec wallclock/time step on 1024 PEs
- 65 Gb input



# TeraGrid value added

- Architectural diversity-Use large shared memory machines for the grid generation, and large distributed memory machines for the forward computation
- Further goal is the inverse problem- use data from multiple earthquakes to refine geological models. Will require many independent runs on highly parallel machines.



S. CA significant earthquakes since 1812

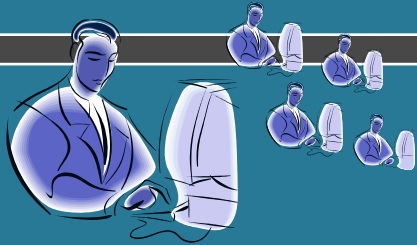
# WIDE-Science Gateways

---

- Web interfaces that invoke TeraGrid resources invisibly  
Involves many more users than the traditional HPC community



# nanoHUB: extensive cyberenvironment from NSF Network for Computational Nanotechnology



<http://nanoHUB.org>

