Petascale Earthquake Simulations for Improved Seismic Hazard Analysis

Yifeng Cui
San Diego Supercomputer Center

In collaboration with Kim Olsen and Steve Day of San Diego State University, Philip Maechling and Thomas Jordan of the Southern California Earthquake Center and Amit Chourasia of San Diego Supercomputer Center
The southernmost San Andreas fault has a high probability of rupturing in a large (>M 7.5) earthquake during the next two decades.

Historic earthquakes:
- 1906, enormous damage to SF
- 1857, 360km long rupture

Major events have not been seen on San Bernardino Mountains segment since 1812 and on Coachella Valley segment since ~1680. Average recurrence intervals on this segment is ~150 and ~220 years respectively.

1994 Northridge
- When: 17 Jan 1994
- Where: San Fernando Valley
- Damage: $20 billion
- Deaths: 57
- Injured: >9000
SCEC PetaSHA Computational Pathways

- Standard seismic hazard analysis
- Ground motion simulation
- Dynamic rupture modeling
- Ground-motion inverse problem

Unified Structural Representation

1. Intensity Measures
2. Attenuation Relationship
3. Earthquake Rupture Forecast
4. Ground Motions

- Other Data
  - Geology
  - Geodesy
- F3DT

Source: SCEC

KFR = Kinematic Fault Rupture
DFR = Dynamic Fault Rupture
AWP = Anelastic Wave Propagation
NSR = Nonlinear Site Response
F3DT = Full 3D Tomography

Source: SCEC
<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>TeraShake 1.x First large wave propagation simulations of Mw7 earthquakes on the southern San Andreas with maximum frequency of 0.5 Hz run using 53 billion mesh points based on the Denali earthquake. 744 SDSC Jaguar cores used, 53 TB outputs, largest simulation outputs recorded.</td>
</tr>
<tr>
<td>2005</td>
<td>TeraShake 2.x Simulations of Mw7.8 earthquakes in 2005-2006 using source descriptions generated by dynamic rupture simulations. The dynamic rupture simulations were based on boundary initial stress conditions, used 1024 NCSA TG cores.</td>
</tr>
<tr>
<td>2006</td>
<td>ShakeOut 2.x Simulations of Mw7.8 earthquakes with max 1Hz using source descriptions generated by ScDAC dynamic rupture simulations. The ShakeOut 2.x dynamic rupture simulations were introduced at the ShakeOut 2.x kinematic sources. 33K TACC Ranger cores used.</td>
</tr>
<tr>
<td>2007</td>
<td>Chino Hills 1.x Comparison of simulated and recorded ground motions for 2007 Mw 4.3 Chino Hills, five simulations were conducted using meshes extracted from D6M of five databases for CVM-H and CVM-T, 96K NICS Kraken cores used.</td>
</tr>
<tr>
<td>2008</td>
<td>BGW 96% Parallel efficiency on 40K TJ Watsonson BG L cores.</td>
</tr>
<tr>
<td>2008</td>
<td>M8 1.x Simulations of Mw8.0 scenario on SAF from the Salton Sea to Parkfield ('Wall-to-Wall'), up to 1 Hz. The source description was generated by combining several existing dynamic Mw7.8 dynamic source descriptions. 96K NICS Kraken cores used.</td>
</tr>
<tr>
<td>2008</td>
<td>ShakeOut verification with 3 models</td>
</tr>
<tr>
<td>2009</td>
<td>Chino Hills 1.x Comparison of simulated and recorded ground motions for 2009 Mw 5.4 Chino Hills, five simulations were conducted using meshes extracted from D6M of five databases for CVM-H and CVM-T, 96K NICS Kraken cores used.</td>
</tr>
<tr>
<td>2009</td>
<td>BGW 96% Parallel efficiency on 40K TJ Watsonson BG L cores.</td>
</tr>
<tr>
<td>2009</td>
<td>M8 2.x 40m spacing and 435 billion mesh points, 400 x to run on 230K NCCS Jaguar cores, the world's most powerful machine.</td>
</tr>
<tr>
<td>2010</td>
<td>M8 3.2 Improved source descriptions based Wave propagation simulation: dx=5m, Mel 8.0, 2 Hz; 2.94 bilion mesh points, 25% bigger than current runs.</td>
</tr>
<tr>
<td>2011</td>
<td>M8 3.1 New model under development to deal with complex geometry, topography and non-planar fault surfaces.</td>
</tr>
<tr>
<td>2012</td>
<td>Big 10 Simulation of 9.0 Megaquake in Pacific Northwest 15 Ms SU awarded, largest NSF TG allocation</td>
</tr>
<tr>
<td>2012</td>
<td>M8 3.2 Improved source descriptions based Wave propagation simulation: dx=5m, Mel 8.0, 2 Hz; 2.94 bilion mesh points, 25% bigger than current runs.</td>
</tr>
</tbody>
</table>

**SciDAC OASCR Award**

**TeraGrid Viz Award**

**96% Parallel efficiency on 40K TJ Watsonson BG L cores.**

**M8 2.x**

40m spacing and 435 billion mesh points, 400 x to run on 230K NCCS Jaguar cores, the world’s most powerful machine.

**M8 3.2**

Improved source descriptions based Wave propagation simulation: dx=5m, Mel 8.0, 2 Hz; 2.94 bilion mesh points, 25% bigger than current runs.

**M8 3.1**

New model under development to deal with complex geometry, topography and non-planar fault surfaces.

**96% Parallel efficiency on 40K TJ Watsonson BG L cores.**

**M8 2.x**

Simulations of Mw7.8 earthquakes with max 1Hz using source descriptions generated by ScDAC dynamic rupture simulations. The ShakeOut 2.x dynamic rupture simulations were introduced at the ShakeOut 2.x kinematic sources. 33K TACC Ranger cores used.

**ShakeOut verification with 3 models**

**Chino Hills 1.x**

Comparison of simulated and recorded ground motions for 2009 Mw 4.3 Chino Hills, five simulations were conducted using meshes extracted from D6M of five databases for CVM-H and CVM-T, 96K NICS Kraken cores used.

**BGW 96% Parallel efficiency on 40K TJ Watsonson BG L cores.**

**M8 1.x**

Simulations of Mw8.0 scenario on SAF from the Salton Sea to Parkfield (Wall-to-Wall), up to 1 Hz. The source description was generated by combining several existing dynamic Mw7.8 dynamic source descriptions. (‘ShakeOut-D’). 96K NICS Kraken cores used.

**ShakeOut 1.x**

Simulations of Mw7.8 with max frequency of 0.1 Hz run using kinematic source descriptions based on geological observations. 1024 TACC Lonestar cores.

**ShakeOut 2.x**

Simulations of Mw7.8 earthquakes with max 1Hz using source descriptions generated by ScDAC dynamic rupture simulations. The ShakeOut 2.x dynamic rupture simulations were introduced for the first time to simulate the ShakeOut 2.x kinematic sources. 33K TACC Ranger cores used.

**M8 3.1**

New model under development to deal with complex geometry, topography and non-planar fault surfaces.

**Big 10**

Simulation of 9.0 Megaquake in Pacific Northwest 15 Ms SU awarded, largest NSF TG allocation.
TeraShake Platform

- 600km x 300km x 80km
- Spatial resolution = 200m
- Mesh Dimensions
  - 3000 x 1500 x 400 = 1.8 billion mesh points
- Simulated time = 3 minutes
- Number of time steps = 22,728 (0.011 sec time step)
- 60 sec source duration from Denali
- 3D Crustal structure: subset of SCEC CVM3.0
- Near-surface S-wave velocity truncated at 500m/s, up to 0.5 Hz
M$_{w}$ 7.8 ‘ShakeOut’ Scenario

- San Andreas “Big One”
- Initiation near Bombay Beach (unilateral rupture to the NW)
- Slip of 4.5 meters at Cajon Pass
- Slip distribution based on geology and time-predictable model and dynamic rupture simulations
- 6000 x 3000 x 800 grids, upper to 1-Hz, 300km fault rupture zone, velocity properties SCEC CVM4.0
- Runs on TACC Ranger and NICS Kraken
- Largest emergency response in US history with 5.3 million participants, demonstrated that existing disaster plans are inadequate for an event of this scale, USGS collaborated with SCEC and the City of LA Emergency Preparedness Department
Dynamic Rupture Simulations

- Dynamic Rupture simulation results from simulation performed on the TeraGrid resource showing distribution of slip and other rupture parameters on a model of a 300km section of the southern San Andreas Fault. Multiple traces in a single diagram help scientists determine if region of peak horizontal slip occurs correlates with regions of peak vertical slip and if peak slip rate occurs at the position of maximum final displacement (Dalguer & Day, 2007)
Comparison of ShakeOut Simulations

Comparison of ShakeOut Simulations performed at different sites using different codes. The good agreement between the results helped to build confidence in the ground motions projected by the simulations for the ShakeOut scenario event. TeraGrid resources at TACC and PSC were used to run these 1Hz simulations, (Bielak et al., SCEC'08)
Chino Hills Validation Simulations

- Comparison of the goodness-of-fit at 0.1-0.5 Hz for synthetics relative to data from the 2008 Mw5.4 Chino Hills earthquake

- The synthetics were compared to strong-motion data at 33 stations.

- Goodness-of-fit algorithm helps identify critical areas in need of improvement in the Community Velocity Model (CVM), and rate the accuracy of different models.

- Seismogram comparisons of recorded data (black traces), CVM-S synthetics (red traces) and CVM-H synthetics (blue traces)

- 180x125x60km, 50m resolution, 10.8 billion grids, 80k time steps, upper frequency limit to 2-hz, using both SCEC CVM4 and CVM-H velocity models, run on NICS Kraken

(Source: K. Olsen, 2009)
SCEC M8 1.0

- Mw8.0, 550 km long rupture, worst scenario, biggest Earthquake Simulation on San Andreas Fault
- Simulated in a 32 billion grid point subset of the SCEC Community Velocity Model (CVM) V4 with a minimum shear-wave velocity of 500 m/s up to a maximum frequency of 1 Hz.
- Simulated 350s of wave propagation for 3 wall-to-wall source realizations, namely two uni-lateral (southeast-to-northwest and northwest-to-southeast) ruptures, and a bi-lateral rupture starting in the center of the fault.
- The source descriptions were generated by combining several dynamic Mw7.8 ShakeOut-D dynamic source descriptions.
- 96,000 processor cores used for wall-to-wall runs on Kraken, 2.6 hrs WCT, 53 sustained TeraFlop/s

(Movie by A. Chourasia of SDSC)
SCEC Big 10 Project

- Generate a hierarchy of simulations for (roughly) ten of the most probable large (M > 7) ruptures in Southern California, with the objective of understanding how source directivity, rupture complexity, and basin effects control ground motions.

Source: SCEC
AWP-ODC Finite Difference Code

- Structured 3D with 4\textsuperscript{th} order staggered-grid finite differences for velocity and stress originally developed by Olsen et al. of SDSU, Staggered-grid split-node DFM integrated by Dalguer and Day of SDSU, and overall integration and enhancements by Cui et al. of SDSC, extensively validated for a wide range of problems.

- Perfected Matched Layers absorbing boundary conditions on the side and bottom of the grid, zero-stress free surface condition at the top.

- Fortran 90, message passing done with MPI using domain decomposition, I/O using MPI-IO, point-to-point and collective communication.

- A rich end-to-end package including mesh/source generator and partitioners, solver and post-processing tools.
Fault Model Representation:
Staggered-Grid Split-Node (SGSN) Method

(for Velocity-stress staggered-grid finite difference scheme)

Fault Model Representation:
Staggered-Grid Split-Node (SGSN) Method

Fault Plane ($\Sigma$)

Dalguer and Day (2007), JGR
Adapting AWP-ODC to Different Architectures

Adaptive algorithms used to map application to different architectures, and look ahead of new architectures.

Determining how fundamental system attributes affect application performance.

Cui et al, Acta Geotechnica, 2008
### Adaptive Mesh Partitioning

M8 mesh as example: 32 billion mesh points, 596 GB in size, divided 96K files

<table>
<thead>
<tr>
<th>Mesh inputs</th>
<th>Serial (part-serial)</th>
<th>Serial (part-parallel)</th>
<th>MPI-IO scattered read</th>
<th>MPI-IO Contiguous read</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Performance</th>
<th>Serial IO</th>
<th>Serial Pre-partitioned IO</th>
<th>MPIIO (scattered)</th>
<th>MPIIO (contiguous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System dependence</th>
<th>Serial IO</th>
<th>Serial Pre-partitioned IO</th>
<th>MPIIO (scattered)</th>
<th>MPIIO (contiguous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scalability</th>
<th>Serial IO</th>
<th>Serial Pre-partitioned IO</th>
<th>MPIIO (scattered)</th>
<th>MPIIO (contiguous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>Poor</td>
<td>dependents</td>
<td>Good</td>
<td>Good</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of files</th>
<th>Serial IO</th>
<th>Serial Pre-partitioned IO</th>
<th>MPIIO (scattered)</th>
<th>MPIIO (contiguous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>npx<em>npy</em>npz</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Memory requirement (elements)</th>
<th>Serial IO</th>
<th>Serial Pre-partitioned IO</th>
<th>MPIIO (scattered)</th>
<th>MPIIO (contiguous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>nxt<em>nyt</em>nzt/ core</td>
<td>nxt<em>nyt</em>nzt/ core</td>
<td>nxt<em>nyt</em>nzt/ core</td>
<td>nx*ny/core - sender (nz cores)</td>
<td>nx<em>nyt</em>nzt/ core - receiver (all cores)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Communication overhead</th>
<th>Serial IO</th>
<th>Serial Pre-partitioned IO</th>
<th>MPIIO (scattered)</th>
<th>MPIIO (contiguous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>None</td>
<td>None</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Collective IO</th>
<th>Serial IO</th>
<th>Serial Pre-partitioned IO</th>
<th>MPIIO (scattered)</th>
<th>MPIIO (contiguous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stripe number (recommended)</th>
<th>Serial IO</th>
<th>Serial Pre-partitioned IO</th>
<th>MPIIO (scattered)</th>
<th>MPIIO (contiguous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>Small</td>
<td>Large</td>
<td>Large</td>
<td>Large</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stripe size (recommended)</th>
<th>Serial IO</th>
<th>Serial Pre-partitioned IO</th>
<th>MPIIO (scattered)</th>
<th>MPIIO (contiguous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>Small</td>
<td>Big</td>
<td>Bigger (nx*ny)</td>
<td></td>
</tr>
</tbody>
</table>
Adaptive Mesh Partitioning

Serial Read

Serial Read, pre-partitioned

MPI-IO Scattered Read

MPI-IO Contiguous Read

(K. Lee et al., Best Poster Finalist, SC’09)
Source Partitioning

M8 source as example: 879382 nodes, each 24K steps, 393 GB in size

(Cui, Y. et al, e-book, 2009)
Single CPU Optimization

- Reduce the use of expensive division as much as possible, store reciprocals of media parameters $\mu$ and $\lambda$
- Avoid the expensive formulas because of the large difference in execution time between addition and division that degrade on-chip parallelism
- Avoid function calls inside loop, avoid branching in inner loops – this leads to cache misses and also prevents the compiler from vectorizing loops and make use of the multiple floating point units.
- Optimal cache blocking, loop unrolling to improve cache utilization
- Overall improvement of single CPU optimization is 30%

(Nguyen et al., SCEC'09)
Optimization of Communication

Synchronous Communication

Asynchronous Communication

(K. Lee, et al, Best Poster Finalist, SC’09)
Communication/Computing Overlap

- Collaboration with OSU D.K. Panda team on this effort
- Reduced synchronization using two-sided non-blocking
  - Initiate exchange of all components of velocity and then wait on the completion of all transfers at once using a MPI Waitall.
  - Overlapping computation-communication
    - the exchange of an independent velocity and stress or component with the calculation of the next.
    - Split the velocity computation and swap functions into component level granularity.
    - Group the transfers to all neighbors to retain the advantage of concurrent progress.
- Developed multiple one-sided implementations using the active synchronization mechanisms supported in MPI-2.
  - Window Design
  - Fence
  - Post-Wait/Start-Complete
- Compared our Fence-based and Post/Wait/Start/Complete-base version with equivalently restructured ISend/IRecv/Waitall-based versions, show that these two mechanisms provide similar benefit to our application, with a 7% in runtime on 4096 TACC Ranger cores.

(S. Potluri et al, submitted to ISC 2010)
Coordinated Executions on TeraGrid: 2008 ShakeOut-D as an Example

Dynamic Rupture Runs
- up to 2,000 cores
- Dynamic Source Partitioning and split
- Mesh Partitioning

Wave Propagation Runs
- On up to 60k cores
- Velocity mag. & cum peak
- Displace. mag & cum peak
- Seismograms

Analysis

Visualization
Registered to Digital Library

Registered to Digital Library

Network

SDSC Datastar p655
SDSC Datastar p690
SDSC IA-64

GPFS-WAN

Lustre

TACC Ranger

SAM-QFS

SDSC IA-64

Datastar

GPFS

gridFTP
150-450 MB/s

iRODs

SAM-QFS

HPSS

Input
-177 MB/s
Automatic End-to-end Workflow

- Execute different jobs on different TeraGrid machines and control all execution sequences remotely.
- Validate correctness of input and output data and detect errors occurred during the simulation process and recover them automatically.
- High performance data through parallel approach as well as using Globus toolkit and GridFTP features.
Data Transfer, Archive and Management

- Input/output data transfer between SDSC disk/HPSS to Ranger disk at the transfer rate up to 450 MB/s using Globus GridFTP
- 90k – 120k files per simulation, 150 TBs generated on Ranger, organized as a separate sub-collection in iRODs
- Direct data transfer using iRODs from Ranger to SDSC SAM-QFS up to 177 MB/s using our data ingestion tool PIPUT
- Sub-collections published through SCEC digital library (168 TB in size)
- integrated through SCEC portal into seismic-oriented interaction environments

![Data Transfer Diagram](image)

(Zhou et al., CSO’10)
Visualization Using SDSC Vista

- Used SDSC’s volume rendering tool Vista, based on Scalable Visualization Toolkit. Vista employs ray casting for performing volumetric rendering.
- Surface Viz utilized Adobe’s After Effects used for compositing and encoding.
- Topography used Autodesk Maya.
- Web Portal uses LAMP (Linux, Apache, MySQL, PHP) and Java technology for web middleware.
- TeraShake Viz alone consumed 40K CPU-hours in computation, Millions images in total.

(Chourasia et al., 2007)
NCCS Jaguar, World Most Powerful Supercomputer

- The Jaguar XT5 partition contains 224,256 compute cores in addition to dedicated login/service nodes.
- Each compute node contains two hex-core AMD Opteron processors, 16GB memory, and a SeaStar 2+ router.
- The SeaStar2+ router (XT5 partition) has a peak bandwidth of 57.6GB/s.
- The routers are connected in a 3D torus topology, which provides an interconnect with very high bandwidth, low latency, and extreme scalability.
Sustained Performance of SCEC Capability Simulations

- Wall-to-Wall-2 (1Hz, 40m, Vs200m/s)
- Chino Hills 15m (3.3Hz, Min Vs250m/s)
- Wall-to-Wall-1 (1Hz, 100m, Min Vs500m/s)
- Chino Hills 50m (2Hz, Min Vs500m/s)
- ShakeOut-2 (1Hz, 100m, Min Vs500m/s)
- ShakeOut-1 (1Hz, 100m, Min Vs500m/s)
- TeraShake-2 (0.5-Hz, 200m, Min Vs500m/s)
- TeraShake-1 (0.5-Hz, 200m, Min Vs500m/s)
AWP-ODC Strong Scaling on TeraGrid and INCITE Machines

TS: TeraShake (3000x1500x400), SO: ShakeOut (6000x3000x800), W2W: Wall-to-Wall (150m: 5312x2656x520, 40m: 20250x10125x2125)

- SO-100m, Kraken-XT5 (Async Comm)
- SO-100m, Intrepid (Async Comm)
- SO-100m, Ranger (Async Comm)
- TS1-200m, DataStar (Sync Comm)
- SO-100m, Kraken-XT4 (Sync Comm)
- TS2-200m, DataStar (Sync Comm before I/O Tuning)
- W2W-150m, BG/L (Sync Comm)
- W2W-40m, Jaguar (Async Comm)
- W2W-40m, Jaguar-ideal

Wall-to-Wall on Jaguar, Single CPU Optimization
Superlinear Speedup
Sustained 200 Tflop/s
## SCEC Milestone Capability Runs

<table>
<thead>
<tr>
<th>Milestone Runs</th>
<th>TS1</th>
<th>TS2</th>
<th>DS2</th>
<th>SO1</th>
<th>SO2</th>
<th>CH50m</th>
<th>W2W-1</th>
<th>CH15m*</th>
<th>W2W-2*</th>
<th>W2W-3**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine</td>
<td>SDSC DataStar</td>
<td>SDSC DataStar</td>
<td>NCSA IA-64</td>
<td>TACC LoneStar</td>
<td>TACC Ranger</td>
<td>NICS Kraken</td>
<td>NICS Kraken</td>
<td>NICS Kraken</td>
<td>NCCS Jaguar</td>
<td>NCSA Blue Water</td>
</tr>
<tr>
<td>Outer scale (km)</td>
<td>600</td>
<td>600</td>
<td>299</td>
<td>600</td>
<td>600</td>
<td>180</td>
<td>800</td>
<td>183</td>
<td>810</td>
<td>800</td>
</tr>
<tr>
<td>Inner (m)</td>
<td>200</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>50</td>
<td>100</td>
<td>15</td>
<td>40</td>
<td>25</td>
</tr>
<tr>
<td>Max Frequency</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.0</td>
<td>1</td>
<td>3.3</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Min Surface Vel (m/s)</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>250</td>
<td>200</td>
<td>250</td>
<td></td>
</tr>
<tr>
<td>Mesh Points</td>
<td>1.8E+09</td>
<td>1.8E+09</td>
<td>9.6E+08</td>
<td>1.4E+10</td>
<td>1.4E+10</td>
<td>1.1E+10</td>
<td>3.1E+10</td>
<td>3.0E+11</td>
<td>4.4E+11</td>
<td>2.0E+12</td>
</tr>
<tr>
<td>Time Steps</td>
<td>22,768</td>
<td>22,768</td>
<td>13,637</td>
<td>45,456</td>
<td>50,000</td>
<td>80,000</td>
<td>60,346</td>
<td>100,000</td>
<td>120,000</td>
<td>320,000</td>
</tr>
<tr>
<td>Vel. Model Input (TB)</td>
<td>0.05</td>
<td>0.05</td>
<td>0.03</td>
<td>0.42</td>
<td>0.42</td>
<td>0.31</td>
<td>0.89</td>
<td>6.87</td>
<td>12.68</td>
<td>59.60</td>
</tr>
<tr>
<td>Storage w/o ckpt (TB)</td>
<td>53.0</td>
<td>10.0</td>
<td>9.5</td>
<td>0.5</td>
<td>0.5</td>
<td>1.9</td>
<td>0.3</td>
<td>66.4</td>
<td>39.9</td>
<td>400.0</td>
</tr>
<tr>
<td>Cores used</td>
<td>240</td>
<td>1,920</td>
<td>1,024</td>
<td>1,920</td>
<td>32,000</td>
<td>64,000</td>
<td>96,000</td>
<td>96,000</td>
<td>223,080</td>
<td>320K**</td>
</tr>
<tr>
<td>Wall-Clock-Time (hrs)</td>
<td>66.8</td>
<td>6.7</td>
<td>35.2</td>
<td>32.0</td>
<td>6.9</td>
<td>2.3</td>
<td>2.5</td>
<td>24</td>
<td>20</td>
<td>45**</td>
</tr>
<tr>
<td>Sustained TeraFlop/s</td>
<td>0.04</td>
<td>0.43</td>
<td>0.68</td>
<td>1.44</td>
<td>7.29</td>
<td>26.86</td>
<td>50.00</td>
<td>87.00</td>
<td>200.00</td>
<td>1,000**</td>
</tr>
</tbody>
</table>

*benchmarked,  **estimated
Fundamental Earthquake Physics Problem to be Investigated on Blue Waters: Dynamic Ruptures

1. Our Blue Waters research will initially focus on improving the earthquake source descriptions used in ground motion prediction calculations.

2. We will improve our earthquake source descriptions by integrating more realistic friction laws into our dynamic rupture simulations and computing at large scales including inner-scale of friction processes and outer-scale of large faults.

3. New insights into the physics of earthquake ruptures will lead to improved source descriptions which will improve all of the existing ground motion prediction calculations with a very broad impact.
HPGeoC Lab Activities

- Developing parallel Finite Element code (MaFE) to simulate dynamic ruptures on complex fault geometries and topography, development funded through TeraGrid ASTA, will be used for SCEC capability computing.
- SORD I/O optimization for scalability on petascale architectures funded by SCEC.
- AWP-ODC development and optimization for Blue Waters Project (NSF PRAC).
- Petascale Earthquake Simulations funded through PetaApps-1 and PetaApps-2.
- MPI 1-sided Communication Project funded by NSF HECURA (PI: OSU DK Panda).

*Ma et al (2006) Effects of Large-Scale Surface Topography on Ground Motions*
Challenges for Petascale Computing

- Scalability, efficient cache utilization and data locality
- Efficient use of multiple-core
- Topology-aware MPI communication, mapping and scheduling required
- Coordinated parallel I/O support for large-scale simulations
- Dynamic and Adaptive algorithms needed for different architectures
- Data virtualization and trust virtualization needed to share and manage distributed data
- Automatic end-to-end simulations essential
- Fault tolerance techniques with low overhead necessary
SAN DIEGO SUPERCOMPUTER CENTER, UCSD

(Movie by A. Chourasia of SDSC)
Acknowledgements

- HPGeoC is mainly funded by SCEC
- Computing resources are from NCCS Jaguar, NICS Kraken, TACC Ranger, ALCF Intrepid, SDSC Triton
- Following contributed to the project:
  - SCEC: Philip Machleing, Thomas Jordan, David Okaya, Patrick Small, Geoff Ely, Kevin Milner
  - SDSU: Kim Olsen, Daniel Roten, Steve Day, Luis Angle Dalguer
  - SDSC: Dong Ju Choi, Kwangyoon Lee, Amit Majumdar, Mahidhar Tatineni, Jun Zhou, Joey Reed, Amit Chourasia
  - TACC: Tommy Minyard, Karl Schultz
  - NICS: Bruce Loftis, Kwai Wong
  - OSU: DK Panda, Sreeram Porluri, Karen Tomko, Ping Lai
  - CMU: Tim Kasier, Zizhong Chen
- ANL: James Osborn, Robert Latham
  - Cray: John Levesque
  - NCCS: Trey White