Probabilistic Seismic Hazard Map Calculations using 3D Ground Motion Simulations

Abstract:
The purpose of this project is develop improved probabilistic seismic hazard maps for a region in southern California through the use of 3D ground motion simulations. Improved probabilistic seismic hazard analysis maps will provide scientists, engineering groups, and emergency responses organization with more accurate understanding of the seismic hazards in the mapped areas. In this proposal, SCEC Community Modeling Environment (SCEC/CME) Project collaborators request Service Units (SUs) on TeraGrid IA-64 machines. With these Service Units, we will run approximately 500 wave propagation simulations using a well-validated, high performance, finite difference, anelastic waveform modeling code (AWM-Olsen). By running these simulations, we will create a library of synthetic seismograms that will be post-processed into peak ground motion values for each simulation. The peak ground motion values will then be combined to create probabilistic seismic hazard maps for a region in southern California. Job submission and data management will be performed using the SCEC grid-based workflow tools that are based on National Middleware Initiative (NMI) software tools. We have identified the primary field of science for this project as earthquake hazard mitigation because our research is developing new techniques for seismic hazard analysis. The work proposed on this project is directed at improved understanding of the seismic hazards in Los Angeles, but the techniques we are using are applicable to other areas as well.
Project Summary:

Probabilistic Seismic Hazard Analysis (PSHA) is a method of analyzing the probability of strong ground motions at a given site. A probabilistic seismic hazard analysis provides a specification of the maximum intensity of shaking expected at a site during a fixed time interval at a given probability level. The results of a PSHA are often presented as seismic hazard maps (see Figure 1). Engineers use these maps to design buildings, emergency preparedness officials use them for planning purposes, and insurance companies use them to estimate potential losses.

![Figure 1: Probabilistic Seismic Hazard Map for Southern California showing Peak Ground Acceleration (PGA), Interval of 50 years, and Probability of Exceedance 2%](image)

This project will be conducted as part of the SCEC Community Modeling Environment (SCEC/CME) collaboration. The SCEC/CME Project includes collaborators from the Southern California Earthquake Center (SCEC), University of Southern California (USC), Carnegie Mellon University (CMU), Scripps Institution of Oceanography (SIO), San Diego Supercomputer Center (SDSC), and San Diego State University (SDSU). The SCEC ITR is funded by a National Science Foundation (NSF) Information Technology Research (ITR) award. The SCEC/CME Project is actively performing research on the development of probabilistic seismic hazard analysis tools and techniques, on grid-based workflow techniques, on development of information technology for the organization and management of simulation output and observational data, and on the creation of scientific digital libraries. The SCEC/CME Project explicitly assumed use of the distributed Terascale Facility for computations, visualization, and data storage.

The current methodology of probabilistic seismic hazard analysis combines an earthquake forecast model with attenuation relationships to provide probabilistic estimates of intensity measures. Intensity measures of interest might include the peak ground acceleration (PGA), peak ground velocity (PGV), or the response spectral densities at specified frequencies. The earthquake forecast, also known as an Earthquake Rupture Forecast (ERF), comprises a set of earthquake scenarios, each described by a
magnitude, a location, and the probability that the scenario will occur by some future date (e.g., a Poisson distribution). The attenuation relationship is a relatively simple analytical expression that relates each earthquake scenario to the intensity of shaking (e.g., PGA) at each site of interest; it usually accounts for the local geologic conditions at each site (e.g., sediment sites tend to shake more than rock sites). The analysis determines the intensity that will be exceeded at some specified probability over a fixed period of time (e.g., PGA with a 10% probability of exceedance during the 50-year life span of a building).

Researchers on the SCEC/CME have developed a general model for PSHA as shown in Figure 2. The primary elements in a PSHA calculation include an Earthquake Rupture Forecast (a list of probable events in the region of interest) and an Intensity Measure Relationship (IMR) that models how the ground motions produced by the ruptures decay with distance. Software tools developed on the SCEC/CME project provided implementation for each of these PSHA calculation elements.

![Fundamental elements for an OpenSHA Calculation](image)

**Figure 2: General Model for Probabilistic Seismic Hazard Analysis (PSHA) Calculations.**

Determination of the effects of a given earthquake at a specific site is modeled through an Intensity Measure Relationship (IMR). An attenuation relationship is one type of IMR. An important advancement in PSHA will be to develop IMRs that uses numeric simulations of wave propagation. Numerical simulations of wave propagation can now be done in three dimensions for models with sufficient realism (e.g., three-dimensional geology, propagating sources, frequencies approaching 1 Hz) to be of engineering interest.
There is strong interest in application of wave propagation simulations as intensity measure relationships because of the recent recognition that rupture directivity plays a strong role in ground motion. Figure 4 shows a “scenario” ground motion map showing predicted peak spectral accelerations for a hypothetical Mag 7.4 earthquake on the southern San Andreas. This map was calculated using the OpenSHA tools developed on the SCEC/CME Project. A “scenario” ground motion map is somewhat different from the probabilistic ground motion maps that we are proposing to develop on this project. A “scenario” ground motion map shows the predicted ground motion for a single earthquake. A probabilistic ground motion map shows the predicted peak ground motions over a specific period of time (such as 50 years) and it is created by combining many scenario ground motion data sets, hence our need to run a large number of earthquake simulations.

To create Figure 4, an attenuation relationship-based Intensity Measure Relationship was used to model the decay of ground motion with distance. Using this technique, the strong ground motions are predicted to be limited to a narrow region adjacent to the fault.

During 2004, members of the SCEC/CME Project ran an earthquake wave propagation simulation, called TeraShake, for a similar earthquake on the southern San Andreas. This earthquake wave propagation simulation used the AWM-Olsen software that we propose to use in this project. A peak ground motion map for the TeraShake simulation is shown in Figure 5. In this map, significant ground motions are predicted.
well away from the fault including in heavily populated areas like downtown Los Angeles.

Figure 4: Scenario ShakeMap for a M 7.4 southern San Andreas rupture using an attenuation relationship to model ground motion decay with distance.

Figure 5: Peak Ground Velocity map from the TeraShake simulation of a M 7.7 southern San Andreas earthquake using AWM-Olsen code. The results showed significant directivity effects and strong ground motions well away from the fault.
In our proposed project, we will begin the process of developing Probabilistic Seismic Hazard Maps using more accurate, but more computationally expensive, numeric wave propagation simulations as intensity measure relationships.

Use of numeric simulation of wave propagation have not been widely applied to PSHA for a number of reason including the following: (a) lack of access to validated, high performance, numeric simulation codes, (b) lack of validated geophysical models for the region of interest, (c) lack of access to high performance computing facilities, and (d) lack of computational and data management capabilities.

The SCEC/CME Project is in a unique position to address these problems and make the first steps in advancing the application of numeric earthquake wave propagation simulations to probabilistic seismic hazard analysis. The SCEC/CME Project has the scientific expertise in Probabilistic Seismic Hazard Analysis, and in numeric wave propagation simulations, as well as access to well validated and highly efficient codes, geophysical models, and computational and data management tools needed to perform this work provided that we can obtain significant computational allocation for this Project.

Summary of Proposed Research

We propose to perform a Probabilistic Seismic Hazard Analysis for 50km x 50km region in southern California using wave propagation simulations as an intensity measure relationship. The selected region is currently described in attenuation relationship-based USGS seismic hazard maps. The Los Angeles area has been selected as our initial region of interest due to its geological properties (including deep sedimentary basins and rock mountains) as well as for the significant societal infrastructure at risk in the area (including tall buildings that respond to long period seismic waves). Using SHA tools developed on the SCEC/CME system, we will analyze current USGS standard Earthquake Rupture Forecast (Frankel 2002) and identify the most menacing earthquakes in the forecast for our region of interest. For each of these ruptures, we will run a numeric wave propagation simulation to model ground motions produced in the selected region.

The resulting ground motion collection will be processed, using SCEC/CME software tools, into probabilistic seismic hazard maps for the region under study. These maps will be evaluated by comparison to existing, attenuation relationship-based, PSHA maps for the same region.

This project will use numeric wave propagation software developed by Kim Olsen (Olsen et al., 2003) and optimized for use on TeraGrid computers under the Strategic Application Collaboration (SAC) Project at San Diego Supercomputer Center. The job submission and data management will use the SCEC/CME grid-based workflow system that currently uses a National Middleware Initiative (NMI)-based software stack to submit jobs to the TeraGrid machines and to perform data management.

A unique characteristic of the project is the inter-disciplinary characteristics of the effort which strongly support the goals of the NSF ITR Program. This project involves geoscientists with expertise in Probabilistic Seismic Hazard Analysis and numeric wave propagation simulations as well as information technology researchers with expertise in high performance computing, grid-based workflow systems, data management, and visualization. Our proposed project leverages both the geosciences and IT research already performed under the SCEC/CME ITR award.
Computational Methodology

We have selected a 50km x 50km region of interest in southern California centered in the Los Angeles area for which we will calculate a probabilistic seismic hazard map. We used the current standard USGS Frankel 2002 Earthquake Rupture Forecast to identify potential earthquake that menace this region. The Frankel 2002 ERF was used to produce the current national seismic hazard maps. This ERF identifies approximately 1,000,000 possible future earthquakes for the California area. We have selected a subset of the total ERF by setting minimum Magnitudes, Probability of Occurrence, and distance from Los Angeles.

<table>
<thead>
<tr>
<th>Minimum Magnitude</th>
<th>Minimum Probability of occurrence in 50 years (in %)</th>
<th>Maximum Distance to Los Angeles (in Km)</th>
<th>Num of Ruptures in Frankel-2002 Earthquake Rupture Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some value &gt;0</td>
<td>Some value &gt;0</td>
<td>200</td>
<td>~1,000,000</td>
</tr>
<tr>
<td>5.0</td>
<td>0.5</td>
<td>200</td>
<td>2106</td>
</tr>
<tr>
<td>5.0</td>
<td>1.0</td>
<td>200</td>
<td>481</td>
</tr>
<tr>
<td>5.0</td>
<td>2.0</td>
<td>200</td>
<td>57</td>
</tr>
<tr>
<td>6.0</td>
<td>0.5</td>
<td>200</td>
<td>301</td>
</tr>
<tr>
<td>6.0</td>
<td>1.0</td>
<td>200</td>
<td>159</td>
</tr>
<tr>
<td>6.0</td>
<td>2.0</td>
<td>200</td>
<td>57</td>
</tr>
</tbody>
</table>

We believe that the selection criteria of Minimum Magnitude = 5.0, POE=1%, and Max Distance=200km are the minimum number of simulations required to provide a meaningful PSHA map for our Los Angeles area region of interest.

We propose to run a numerical simulation for each of these 481 ruptures and save the resulting surface seismograms on a grid with 200 meter grid spacing. This spacing is sufficiently dense to produce a hazard map comparable to the national seismic hazard maps. Our simulations will be run at sufficient temporal resolution, that is, it will produce synthetics at high enough frequency, to produce meaningful values for intensity measure of spectral acceleration at 3.0 seconds. This intensity measure is of interest to the engineering community and can be produced by existing numeric simulation codes. Our output data collections will be 3 components synthetic seismograms at the surface for our region of interest. Post processing (including conversion from velocity seismograms to peak spectral acceleration at 3.0 seconds) of these surface seismograms will be performed using methodologies and software tools developed and validated by SCEC/CME scientists.

While the region of interest for which we will produce a probabilistic seismic hazard map is 50km x 50km, a significant number of earthquakes that menace this region are actually outside the region of interest. The volume for which we will run our wave propagation simulation will be a 300km x 100km x 40km region that surrounds the 50km x 50km region of interest. This size simulation volume ensures that all the ruptures originate well within the simulation volume and that the region of interest has a small buffer zone and the seismograms in the region of interest will be not affected by boundary effects. Figure 6 shows a 50km x 50km region of interest embedded within a larger 300km x 100km simulation volume.
We will use a 200m grid spacing in order to resolve frequencies of interest. With 200 meter spacing and 100 time steps per second, we expect the frequency resolution of our output data to be .5Hz. Our simulation volumes will contain \((300\text{km} \times 5/\text{km}) \times (100\text{km} \times 5/\text{km}) \times (40\text{km} \times 5/\text{km})\) = 150,000,000 nodes. The S wave propagation time across this region is approximately 65 seconds. We will run each simulation for 70 seconds at 100 time steps per second in order to capture S waves for all events. This will result in 7000 time steps per simulation.

In some cases, we may be able to run a wave propagation simulation on a smaller grid. For example, if one of the simulation ruptures is within the 50km x 50km region of interest, we can run on 50km x 50km grid. However, our initial analysis of the location of the 481 ruptures indicates that a large majority of them are outside of our Los Angeles region of interest, so we will use the 300 km x 100 km as our planning region with the understanding we will using a smaller grid if possible. If we run on a smaller grid, we will also reduce the number of simulation time steps.

We will use the SCEC Community Velocity Model to provide the geological structure for the wave propagation simulation. This velocity model has been carefully validated by SCEC scientists for the Los Angeles basin.

The Frankel 2002 Earthquake Rupture Forecasts contain source descriptions including hypocenter, magnitude, rake, and dip as well as fault dimension parameters. We will convert these ERF source parameters into double-couple point source.
descriptions that can be used as earthquake source descriptions by the AMW-Olsen wave propagation code.

To run the wave propagation simulations we will use the AWM-Olsen (Olsen et al., 2003) anelastic wave propagation software developed by Kim Olsen at San Diego State University. This code has been developed and validated by SCEC researchers. It has been run at several supercomputer centers including USC HPCC, SDSC on Blue Horizon, DataStar, and at SDSC and NCSA on TeraGrid IA-64 computers.

During 2004, the AMW-Olsen software was the focus of a Strategic Application Collaboration by the SCEC/CME and SDSC. The code was validated and optimized for execution on both DataStar and the TeraGrid systems with the intention that it be used by a wider geophysical community.

AWM-Olsen

AWM-Olsen is a 4th order finite difference wave propagation and rupture dynamics code. Its 3D numerical modeling uses a structured (equi-spaced) staggered-grid of velocity-stress finite differences (FD) (Olsen, 1994; Olsen et al., 1995; Marcinkovich and Olsen, 2003). The AWM-Olsen code was written and tested by Kim Olsen, a member of the SCEC/CME Project. The FD approach has in recent years established itself as a robust and accurate algorithm for large-scale wave propagation in heterogeneous models. Figure 7 illustrates how a 3-D volume of the model is decomposed over 8 parallel processors using domain decomposition. Each processor is responsible for performing stress and velocity calculations for its portion of the grid, as well as dealing with boundary conditions at the external edges of each volume. At the internal edges, where neighboring portions of the earth volume are contained on separate nodes, the processors must exchange stress and velocity information to propagate the waves correctly.

![Figure 7: Illustration of the decomposition of the 3D basin model, decomposed over a number (here, eight) of parallel processors](image_url)
The AWM-Olsen code has been carefully validated through previous simulations for historic scenario earthquakes in the Los Angeles area. To validate the results of a simulation for a historic earthquake, the synthetic seismograms output by the simulation are compared against the observed seismograms for the real event. Figure 8 shows an example of this validation approach for a Sept, 2002 earthquake in Yorba Linda, CA.

Figure 8: Comparisons for 09/03/02 Yorba Linda Earthquake. Observed data is shown in black. AWM-Olsen synthetics in blue, SEM-Tromp synthetics in red.

The AWM-Olsen code was the focus of an evaluation and optimization effort by the NPACI Strategic Applications Collaborations (SAC) project at SDSC. As a part of the SCEC TeraShake effort performed in 2004, the scientific program group at SDSC performed a number of important evaluations and improvements in the existing AWM-Olsen code. The SAC work included the following:

- Porting AWM-Olsen code onto IBM DataStar machine
- Porting of AWM-Olsen code onto TeraGrid IA64 machines
- Modifying the MPI to improve scalability
- Solving memory usage problem related to the large mesh scale.
- Implemented efficient use of MPI I/O for large simulations
- Performed single-processor tuning as well as scaling improvement
- Introduced restart/checkpoint capability
- Added MD5 check summing capability
- Improved output format preferences
- Validating the result
The resulting AWM-Olsen code was used for a several large scale wave propagation simulation runs on DataStar in a set of simulation at SDSC in 2004 termed the TeraShake simulations. We propose to use the same code, the AWM-Olsen code, used for TeraShake, on the TeraGrid IA-64 machines for our set of earthquake simulations.

This AWM-Olsen code is a community code for 3D numeric earthquake simulations. Our proposed project applies this community code to an innovative improvement in traditional seismic hazard analysis and is consistent with the goals our NSF SCEC/CME ITR Project.

The project we propose here is an application of the AWM-Olsen community software on TeraGrid resources. This proposed project does not have a significant software development element. Both the wave propagation codes, as well as the post-processing codes that we will use to create the seismic hazard maps, have already been developed and validated by SCEC/CME researchers.

This project will require significant job submission, monitoring, and data management capabilities. To help us with job-submission, data communication between sites, and job monitoring, the SCEC/CME project has developed a workflow system that utilizes NMI software including Globus, Condor-g, and Storage Resource Broker clients. We will use the SCEC/CME workflow system on this Project to facilitate the job submission, data communications, and data management elements of the work.

Preliminary Progress

This proposed project leverages a significant amount of earlier work done by SCEC/CME collaborators, as well as work done by the Scientific Computing Group at SDSC. The key elements in this effort include (a) Probabilistic Seismic Hazard Analysis software tools to identify the ruptures to simulate and to combine the resulting seismograms into meaningful PSHA maps, (b) a highly scalable wave propagation code, (c) NMI-based, grid-based workflow system to manage the job submission and file management, and (d) digital library tools to manage the large number of resulting files.

The SCEC/CME project has developed the PSHA tools that enable us to identify the appropriate ruptures for our simulations and to combine the simulation results into probabilistic hazard maps. Our PSHA software library includes data processing codes to extract Intensity Measures of Interest, including Spectral Acceleration at 3.0 seconds, from the synthetic seismograms produced by the simulations. One of the SCEC/CME working group leaders (Edward Field) leads the OpenSHA seismic hazard analysis development in the Pasadena Office of the US Geological Survey. He will assist in the production of the PSHA maps when the simulations are completed. Figure 9 shows a probabilistic seismic hazard map created for southern California using the OpenSHA tools. The OpenSHA tools will be used to create the PSHA maps once the simulations have been run and the library of synthetic seismograms are complete.
We have well-validated, highly scalable, earthquake wave propagation software available in the AWM-Olsen code. The AWM-Olsen software has been ported to DataStar and TeraGrid IA64 computers. This software was written by a SCEC/CME collaborator, Kim Olsen, who will assist in the simulation planning and data analysis.

The SCEC/CME project has developed an NMI-based workflow system and has successfully demonstrated the ability to submit jobs to TeraGrid system and to transfer data between SCEC, TeraGrid, and SRB-based Project storage at SDSC. This workflow system demonstrated TeraGrid jobs submissions from the SCEC system, grid-based data transfers, and automatic registration of AWM-Olsen simulation results into the Storage Resource Broker. SCEC/CME Workflow system is shown schematically in Figure 10.

**Figure 9:** Probabilistic Seismic Hazard Map for southern California created using OpenSHA tools.

**Figure 10:** The SCEC/CME Workflow systems uses grid-based workflow tools to manage a large number of jobs and data files.
The SCEC/CME digital library uses the Storage Resource Broker (SRB) for long term storage of simulation results including data and metadata description of each data set. SCEC/CME collaborators at SDSC include Reagan Moore and Marcio Faerman will assist in the data and metadata management elements of this project. An overview of the SCEC/CME SRB-based digital library system is shown in Figure 11.

Figure 11: Simulation results will be stored in the SCEC Digital Library which uses the Storage Resource Broker (SRB).

AWM-Olsen Scaling Information

SCEC/CME project members and members of the SDSC scientific applications group have tested the scaling efficiency of AWM-Olsen on two different platforms: the DataStar system located at the San Diego Supercomputer Center (SDSC), and the TeraGrid IA64 cluster at SDSC.

Our scaling measurements were performed as part of the SCEC-SDSC TeraShake effort which was a Strategic Application Collaboration (SAC) during 2004 with Kim Olsen, Yifeng Cui, Marcio Faerman and others. The AWM-Olsen code was tested with TeraShake scale mesh dimensions (3000 x 1500 x 400 = 1.8 billion nodes), as well as smaller scale mesh dimensions (1500 x 750 x 200 = 225 million nodes) on both DataStar and TeraGrid IA64.

Generally, the algorithm shows very good scaling as a function of the number of processors involved. For simulation utilizing 30 nodes, or less, the TeraGrid systems performed better. For simulations utilizing more than 30 nodes, the DataStar performance was better.

Although the AWM-Olsen performance is better on the DataStar system, we are requesting TeraGrid IA-64 allocations because we need to run a large number of jobs,
and we believe we can obtain higher simulation throughput by using a number of TeraGrid sites and because the standard TeraGrid software tools include the NMI software that our workflow system uses.

**Measured Performance of AWM-Olsen (1500x750x200) on IBM SP4 and TeraGrid IA-64 Cluster**

For this Project, the SDSC Scientific Application Group’s measurements of AWM-Olsen performance on a smaller box (1500 x 750 x 200) are the most applicable measurements to our proposed project. The Total Elapsed time for 11 Time Steps for the AWM-Olsen code running this simulation for both the DataStar and TeraGrid are shown below.

These measurements are for a somewhat larger simulation than the simulations we propose to run in our Project. However, these simulations match ours in other respects. For example, these are for a double-couple point source, which we plan to use, and these are for a simulation that saves only surface data, not volume data. We also plan to only save surface data for our simulations.

These measurements are the Elapsed Time in Seconds for 11 Time Steps of a simulation that uses a double couple point source, that is configured to do check pointing, that outputs surface data only, and that does not calculate an MD5 checksum for each output data files.

### Performance of AWM-Olsen (1500x750x200) on IBM SP4 and Teragrid IA-64 Cluster for 11 Time Steps for a Double Couple Point Source Simulation.

<table>
<thead>
<tr>
<th>Processors</th>
<th>DataStar DS - Elapsed Time (Seconds)</th>
<th>DataStar Speedup</th>
<th>TG IA-64 TG - Elapsed Time (Seconds)</th>
<th>TG - Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1210</td>
<td></td>
<td>93.18</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>101.5</td>
<td>16</td>
<td>61.27</td>
<td>24</td>
</tr>
<tr>
<td>30</td>
<td>53.88</td>
<td>30</td>
<td>33.28</td>
<td>36</td>
</tr>
<tr>
<td>60</td>
<td>26.07</td>
<td>62</td>
<td>30.39</td>
<td>41</td>
</tr>
<tr>
<td>120</td>
<td>18.65</td>
<td>87</td>
<td>27.46</td>
<td>43</td>
</tr>
<tr>
<td>240</td>
<td>12.91</td>
<td>126</td>
<td></td>
<td></td>
</tr>
<tr>
<td>480</td>
<td>20.01</td>
<td>81</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Performance of AWM-Olsen (1500x500x200) on TeraGrid IA-64 Cluster**

The AWM-Olsen computational requirements for our simulations are based on measurements of the code on TeraGrid IA-64 machines. The simulations for which we have measurements are larger than the simulations we propose to run in this Project. We will assume that Total Elapsed Time per Time Step performance scales linearly down to our somewhat smaller size simulations. We will scale from 225 million grid point measurements to the proposed 150 million grid points by scaling the Elapsed Time.
Measurements by a factor of 1.5 (225 million/150 million = 1.5). Our TeraGrid SU allocation request justifications are based on the following scaled estimates:

<table>
<thead>
<tr>
<th>Processors</th>
<th>TeraGrid IA-64</th>
<th>TeraGrid IA-64</th>
<th>TeraGrid IA-64</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TG - Elapsed Time (seconds) for 225 million nodes measured for 11 Time Steps</td>
<td>TG – Elapsed Time (seconds) for 150 million nodes – estimated for 11 Time Steps</td>
<td>TG – Elapsed Time (seconds) for 150 million nodes – estimated for 1 Time Step</td>
</tr>
<tr>
<td>16</td>
<td>93.18</td>
<td>62</td>
<td>5.6</td>
</tr>
<tr>
<td>30</td>
<td>61.27</td>
<td>41</td>
<td>3.7</td>
</tr>
<tr>
<td>60</td>
<td>33.28</td>
<td>22</td>
<td>2.0</td>
</tr>
<tr>
<td>120</td>
<td>30.39</td>
<td>20</td>
<td>1.8</td>
</tr>
<tr>
<td>240</td>
<td>27.46</td>
<td>18</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Memory requirements with AMM-Olsen code

We measured the memory usage for the AWM-Olsen code on a mesh size of 1500x500x200. In the case of using 60 processors, 388 MB memory is allocated per processor which is within the capabilities of the TeraGrid systems.

Table 1: AWM-Olsen Memory Usage for 1500x500x200 mesh

<table>
<thead>
<tr>
<th>Processors</th>
<th>MemA/p (MB)</th>
<th>MemA(MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>656</td>
<td>20992</td>
</tr>
<tr>
<td>60</td>
<td>388</td>
<td>21960</td>
</tr>
<tr>
<td>120</td>
<td>187</td>
<td>22440</td>
</tr>
<tr>
<td>240</td>
<td>98</td>
<td>23520</td>
</tr>
</tbody>
</table>

where:
MemA/p: memory allocation per processor
MemA : memory allocation in total
Note: when using sub-domain source, memory allocated is similar as point source.

Publications Describing the AWM-Olsen Software:


Justification of the Number of Service Units Requested

The proposed use of the TeraGrid facilities is to conduct a large number of simulations that are not feasible on any other platform. We request time on the TeraGrid IA-64 platform because our code has been tested and optimized to run on that architecture. Also, the TeraGrid software stack will allow us to use our NMI software tools to submit jobs and manage the resulting data.

We request 125,000 SUs on the TeraGrid IA-64 platforms to perform earthquake simulations, to create time varying map view visualizations, and to manage transfer of the data to disk and to the HPSS archive. The following estimates are based upon timing measurements made under a prior award to Kim Olsen for the SCEC Rupture Dynamic Project (TG-EAR030004N). The time estimates are for simulations consisting of a 1500 x 500 x 200 (150 million points) mesh with 7000 time steps per simulation. We assume that each data element generates 3 fields of output – 3 velocity components. Each field value is represented by a 4 byte float.

Service Units Required for Ground Motion Simulations

We calculate the number of SUs required to run the proposed wave propagation simulations by using the total execution per time step that were derived from testing on the TeraGrid IA-64 system. We selected 60 CPU’s as optimal number of CPU’s per simulation. We believe that the TeraGrid systems can meet the memory and computational requirements of our proposed simulations using the AWM-Olsen code with this number of CPU’s per simulation. The total SUs for all simulation will be = (Number of Processors x Time Step Execution Time x Number of Time Steps).

We believe a small number of testing simulations will be required to prototype the workflow capabilities and validate the code on various sites. We will not create long term archives for these initial test simulations.

<table>
<thead>
<tr>
<th>SU Needed for Initial Testing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Processors/Simulation</td>
<td>60</td>
</tr>
<tr>
<td>Time Step Total Execution Time (11 Seconds)</td>
<td>22</td>
</tr>
<tr>
<td>Time Step Total Execution Time (secs)</td>
<td>2</td>
</tr>
<tr>
<td>Time Step Total Execution Time (hours)</td>
<td>0.000555556</td>
</tr>
<tr>
<td>Number of Timesteps/Simulations</td>
<td>7000</td>
</tr>
<tr>
<td><strong>SUs/Simulation</strong></td>
<td><strong>233.3333333</strong></td>
</tr>
<tr>
<td>Number of Simulations</td>
<td>15</td>
</tr>
<tr>
<td><strong>Total SUs required for Initial Testing</strong></td>
<td><strong>3500</strong></td>
</tr>
</tbody>
</table>
The simulations derived from the Frankel-2002 Earthquake Rupture Forecast will be used to generate a Probabilistic Seismic Hazard Map for our region of interest. Surface seismograms will be saved for each of these simulations.

**SU Needed to Run Simulations**

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Processors/Simulation</td>
<td>60</td>
</tr>
<tr>
<td>Time Step Total Execution Time (11 seconds)</td>
<td>22</td>
</tr>
<tr>
<td>Time Step Total Execution Time (secs)</td>
<td>2</td>
</tr>
<tr>
<td>Time Step Total Execution Time (hours)</td>
<td>0.000555556</td>
</tr>
<tr>
<td>Number of Timesteps/Simulations</td>
<td>7000</td>
</tr>
<tr>
<td><strong>SUs/Simulation</strong></td>
<td><strong>233.3333333</strong></td>
</tr>
<tr>
<td>Number of Simulations</td>
<td>481</td>
</tr>
<tr>
<td><strong>Total SUs required for Simulations</strong></td>
<td><strong>112233.3333</strong></td>
</tr>
</tbody>
</table>

**Time Varying Map Visualization**

We have surface rendering CPU estimates from the TeraShake simulation visualization work done by the SDSC Visualization Group coordinated by Steve Cutchin. The following rendering estimates are based on the TeraShake rendering measurements. However, we have scaled the TeraShake rendering SU measurements down by a factor of 12 because our map area is a factor of 12 smaller than the TeraShake map area.

**SUs Needed for Visualizations**

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computational Time per Time Step (scaled down from TeraShake)</td>
<td>1</td>
</tr>
<tr>
<td>CPU's used</td>
<td>8</td>
</tr>
<tr>
<td>Number of Time Steps per Simulation</td>
<td>7000</td>
</tr>
<tr>
<td>Computational Time per Map</td>
<td>56000</td>
</tr>
<tr>
<td>SUs per Map</td>
<td>15.555555556</td>
</tr>
<tr>
<td>Number of Simulations</td>
<td>481</td>
</tr>
<tr>
<td><strong>Total SUs for Visualization</strong></td>
<td><strong>7482.22222</strong></td>
</tr>
</tbody>
</table>

**Data Archiving:**

The data of each simulation will be transferred to remote archival storage and will be managed by the SDSC Storage Resource Broker. Our estimate storage requirements are approximately 2.5 TB.

**Data Output Per Simulation**

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grid Points per X</td>
<td>250</td>
</tr>
<tr>
<td>Grid Points per Y</td>
<td>250</td>
</tr>
<tr>
<td>Total Grid Points in Region of Interest</td>
<td>62500</td>
</tr>
<tr>
<td>Number of Fields</td>
<td>3</td>
</tr>
<tr>
<td>Bytes Per Field</td>
<td>4</td>
</tr>
<tr>
<td>Bytes Per Time Step</td>
<td>750000</td>
</tr>
<tr>
<td>Number of Time Steps</td>
<td>7000</td>
</tr>
<tr>
<td>Total Data Per Simulation (bytes)</td>
<td>5,250,000,000</td>
</tr>
<tr>
<td><strong>Total Data Per Simulation (Mbytes)</strong></td>
<td><strong>5250</strong></td>
</tr>
<tr>
<td>Total Data for Project (Mbytes)</td>
<td>2525250</td>
</tr>
<tr>
<td><strong>Total Data For Project (Tbytes)</strong></td>
<td><strong>2.52</strong></td>
</tr>
</tbody>
</table>
We estimate that the following number of service units will be necessary for data archiving. We assume that on average we will be able to obtain 100Mbytes/s of bandwidth between local disk storage at the TeraGrid and the remote archival system. We also assume that one CPU can handle this data transfer rate. Under this assumption we estimate that the number of CPU hours required for archiving are:

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output per Simulation (MBytes)</td>
<td>5250</td>
</tr>
<tr>
<td>Estimated Bandwidth to Data Archive (MB/Sec)</td>
<td>100</td>
</tr>
<tr>
<td>SU's for to write Data to Archive per Simulation</td>
<td>52.5</td>
</tr>
<tr>
<td>SU's for to write Data to Archive per Simulation</td>
<td>0.014583333</td>
</tr>
<tr>
<td>Number of Simulations</td>
<td>481</td>
</tr>
<tr>
<td>Total SUs for Writing Data to Archive</td>
<td>7.014583333</td>
</tr>
<tr>
<td>Total SUs for Project (test + psha + visualization + archiving)</td>
<td>123222.5701</td>
</tr>
</tbody>
</table>

Based on our TeraGrid IA-64 performance measurements of AMW-Olsen code, and these calculations, we request 125,000 SU’s on TeraGrid IA-64 systems in order to perform the research outlined in this proposal.

**Local Computing Environment**

The local computing environment available at SCEC consists of two shared-memory Sun workstations, two shared-memory Linux servers, and an 8 node Rocks-based Linux Cluster.

Our current need for computation to carry out the numerical simulations of our proposed research, in particular memory size, disk space and compute hours is far larger than the resources locally available.

**Other Supercomputer Support**

This proposal requests Service Units on the TeraGrid IA-64 systems. The TeraGrid systems will be used to run the simulations, to create map-based visualizations of the simulation results, and to archive the data into the SCEC SRB-based digital library.

Initial prototyping, workflow testing, and post-processing will be performed using the SCEC allocation on the USC High Performance Computing and Communication Linux Cluster under an existing allocation (June 2004 – June 2005) of 60,000 SUs termed “SCEC Community Modeling Environment”. Philip Maechling is the PI for this USC HPCC Allocation.

Also, the local computing environment at SCEC and USC High Performance Computing and Communication (HPCC) will be utilized during the creation of the Probabilistic Seismic Hazard Maps. We will access the output data collection from the SCEC SRB-based digital library and use computing time on SCEC and USC computers in order post-processes our simulation results into PSHA maps.
Qualifications

Principal Investigator on this project will be David Okaya, and co-PI is Philip Maechling.

David Okaya will have primary responsibility for the seismological aspects of the project insuring that simulations are configured properly so that they produce meaningful ground motion records, and that the seismic hazard analysis, including the post processing of the simulated ground motion data, is performed properly.

David Okaya is a Research Associate Professor in the Earth Sciences department at USC with significant high performance computing software development and earthquake wave propagation simulation experience. David is a working group lead on the SCEC/CME Project. He is a participant on the USC HPCC SCEC Community Modeling Environment allocation of 60,000 SUs.

Philip Maechling will have primary responsibility for the workflow environment, for the data and metadata management, and for ensuring that simulations are scheduled by the SCEC workflow system at the rate required to meet our computational throughput requirements.

Philip Maechling is an MIS Director in the Earth Sciences department at the University of Southern California. He is the Southern California Earthquake Center Information Technology Architect and the Project Manager on the SCEC/CME Project. Philip was PI on a prior PSCA grant (EAR030001P) of 10,000 SU’s. He participated on Kim Olsen’s TeraGrid grants (MCA03S012P and TG-EAR030004N). He participated on a NCSA grant (STA970001N) with PI Donald Frederick. He has run on Blue Horizon at SDSC and on TeraGrid machines at Argonne, Caltech, NCSA, and SDSC, and on Lemieux at PSC. In 2003, he was PI on a 20,000 SU allocation at USC HPCC. In 2004, Philip was PI on SCEC/CME group allocation of 60,000 SUs at the USC High Performance Computing and Communication Center (HPCC) for grid-based workflow software development.

Vita for David Okaya:
Biographical Sketch: David Akiharu Okaya:
3651 Trousdale Parkway
Science Hall, Room 169
Los Angeles, California 90089-0742
okaya@usc.edu, 213-740-5843, Fax 213-740-0011

Professional Preparation:
Princeton University (Geology & Geophysics), B.A. 1978
Stanford University M.S. (Exploration Geophysics), M.S. 1981
Stanford University (Geophysics), Ph.D. 1985

Appointments:
1997-present Research Associate Professor, University of Southern California.
1995-present Adjunct Professor, San Diego State University.
1988-1997 Research Assistant Professor, University of Southern California.
1984-1991 Guest Scientist, Lawrence Berkeley Laboratory.
1985, 1986  Acting Assistant Professor, Stanford University.
1978-1979  Associate Geophysicist, Conoco, Houston, Texas.

Honors:
1995  Fellow, Geological Society of America.
2002  Visiting Fellowship, University of Tokyo.

Publications:

Selected Relevant Publications:

Selected Other Publications:

**Synergistic Activities:**  
*Development and/or refinement of research tools*

- IRIS active source seismology software and data open exchange via IRIS web site (1999).

*Service on national boards and committees*

- IRIS Board of Directors member (2005-present).
- National Science Foundation, Continental Dynamics Program review panel member (2005-present).

**Collaborators & Other Affiliations:**

**Collaborators**
Mark Brandon (Yale Univ), Nik Christensen (Wisconsin), Fred Davey (Inst. Geol. Nuclear Sci, Wellington), Donna Eberhart-Phillips (Otago Univ), Gary Fuis, (USGS-Menlo Park), Stuart Henrys (Inst. Geol. Nuclear Sci, Wellington), Tom Henyey (USC), Naoshi Hirata (U. Tokyo), George Jiracek (SDSU), Luc Lavier (UTIG), Vadim Levin (Rutgers Univ), Kirk McIntosh (UTIG), Anne Meltzer (Lehigh Univ), Jeffrey Park (Yale Univ), Hiroshi Sato (U. Tokyo), Martha Savage (Victoria University of Wellington), Tim Stern (Victoria University of Wellington), Uri ten Brink (USGS-Woods Hole), Phil Wannamaker (Utah), Francis Wu (SUNY/Binghamton).

**Graduate and Post Doctoral Advisors**

- **Graduate Advisor:** George Thompson, Stanford University
- **Post-doctoral Sponsor:** Thomas Henyey, Univ. Southern California, Thomas McEvilly, U.C. Berkeley

**Thesis Advisor and Post-graduate Scholar Sponsor**


Stefan Kleffmann, Martin Scherwath (Victoria University of Wellington, Wellington, New Zealand).

**Philip Maechling Vita:**

**Biographical Sketch: Philip Maechling**
Professional Preparation:
BS: Applied Physics, Xavier University, Cincinnati, Ohio, 1983

Professional Appointments:
2002-Present: IT Architect, Southern California Earthquake Center, University of Southern California, Los Angeles, CA
2001-2002: Lead Software Engineer, Kaiser Permanente, Pasadena, CA
2000-2001: Lead Software Engineer, Cysive, Irvine, CA
1993-2000: Earthquake Monitoring Software Development Coordinator, Seismological Lab, California Institute of Technology, Pasadena, CA
1991-1993: Senior Software Engineer, System and Software Development Division, TRW, Redondo Beach, CA
1986-1991: Senior Software Engineer, Ground Systems Group, Hughes Aircraft Company, Fullerton, CA
1983-1986: Software Engineer, Standard Communications, Carson, CA

Professional Associations & Memberships:
2002-present: Member, Advanced National Seismic System (ANSS) Evolutionary Architecture Working Group
1993-present American Geophysical Union
1993-present Seismological Society of America

Publications Related to the Proposed Project:


Collaborators within the last 48 months:
Thomas Jordan (University of Southern California, Southern California Earthquake Center), Jean-Bernard Minster (Scripps Institution of Oceanography), Carl Kesselman (Information Sciences Institute), Reagan Moore (San Diego Supercomputer Center), Marcio Faerman (San Diego Supercomputer Center), John McRaney (University of Southern California), Ewa Deelman (Information Sciences Institute), Yolanda Gil (Information Sciences Institute), Kim Olsen (San Diego State University), Steve Day (San Diego State University), Jacobo Bielak (Carnegie Mellon University), David O’Hallaron (Carnegie Mellon University)

Others Working on Project
This project utilizes significant resources currently funded through the SCEC/CME project. The following participants on the SCEC/CME project have agreed to assist with the execution of this project.

Edward Field (US Geological Survey), who led the development of the OpenSHA seismic Hazard Analysis software, will assist with the integration of the simulation results into an accurate seismic hazard map. Edward is a SCEC/CME collaborator but he is not funded through the project because he is currently a Federal Employee.

Steve Day (San Diego State University) will assist with the conversion of the Frankel-2002 ERF source descriptions into double-couple point source descriptions appropriate for AWM-Olsen codes.

Reagan Moore (SDSC) and Marcio Faerman (SDSC) will assist with the registration of the simulation results into the SRB and the metadata management effort.

Ewa Deelman and Carl Kesselman (USC Information Sciences Institute) will assist with the SCEC grid-based workflow tools and run-time data management.

Philip Maechling, David Okaya, and Vipin Gupta (USC) will conduct the computations and perform the post processing into seismic hazard maps.

Gaurang Mehta (ISI) and Sridhar Gullapalli (ISI) will monitor the grid-based workflow and job submission system.

David Okaya and Phil Maechling (USC) will monitor the analysis efforts.

Thomas H. Jordan (USC) will provide scientific interpretation of the results.