SOLVING PREDICTION PROBLEMS IN EARTHQUAKE SYSTEM SCIENCE

Allocation: NSF/3.4 Mnh
PI: Thomas H. Jordan[1]
Collaborators: Scott Callaghan[1]; Robert Graves[2]; Kim Olsen[3]; Yifeng Cui[4]; Jun Zhou[4]; Efecan Poyraz[4]; Philip J. Maechling[1]; David Gill[1]; Kevin Milner[1]; Omar Padron, Jr.[5]; Gregory H. Bauer[5]; Timothy Bouvet[5]; William T. Kramer[5]; Gideon Juve[6]; Karan Vahi[6]; Ewa Deelman[6]; Feng Wang[7]

[1]Southern California Earthquake Center
[2]U.S. Geological Survey
[3]San Diego State University
[4]San Diego Supercomputer Center
[5]National Center for Supercomputing Applications
[6]Information Sciences Institute
[7]AIR Worldwide

SCIENTIFIC GOALS

Research by the Southern California Earthquake Center (SCEC) on Blue Waters is focused on the development of physics-based earthquake forecasting models. The U.S. Geological Survey (USGS), through its National Seismic Hazard Mapping Project (NSHMP), currently uses empirical Probabilistic Seismic Hazard Analysis (PSHA) for promoting seismic safety engineering and disaster preparedness across the United States, including California (http://earthquake.usgs.gov/hazards/). PSHA is the scientific basis for many engineering and social applications: performance-based design, seismic retrofitting, resilience engineering, insurance rate setting, disaster preparation and warning, emergency response, and public education.

Our goal is to develop physics-based models for the urban regions of California that are more accurate than the empirical NSHMP standard. A particular objective is to reduce the epistemic uncertainty in PSHA through more accurate deterministic simulations of earthquakes at seismic frequencies up to 1 Hz. Our ultimate goal is to extend physics-based PSHA across the full bandwidth needed for seismic building codes (i.e. up to 10 Hz). We are working with the USGS to integrate SCEC computational results into NSHMP products that the benefit the end-users of earthquake information.

ACCOMPLISHMENTS TO DATE

The SCEC team has combined the Uniform California Earthquake Rupture Forecast (UCERF), the official statewide model of earthquake source probabilities, with the CyberShake computational platform to produce urban seismic hazard models for the Los Angeles region at seismic frequencies up to 0.5 Hz (Fig. 1). UCERF is a series of fault-based models, released by the USGS, the California Geological Survey, and SCEC, that build time-dependent forecasts on time-independent rate models. The second version of the time-dependent model (UCERF2, 2008) has been implemented, and the third version, released last summer (UCERF3, 2014), is being adapted into the Blue Waters workflow.

CyberShake uses scientific workflow tools to automate the repeatable and reliable computation of large ensembles (millions) of deterministic earthquake simulations needed for physics-based PSHA (Graves et al., 2010). Each simulation calculates a component of horizontal ground motion from waves excited by a realistic pseudo-dynamic fault rupture and propagated through a realistic 3D model of Earth's anelastic crustal structure, which is what we mean by "physics-based." By carefully sampling the UCERF rupture variations, we can create probabilistic ground motion models (i.e. hazard curves) for specific sites within regions where the 3D variability of the seismic velocity structure is adequately known, such as the Los Angeles region. By implementing spacetime reversal symmetries (seismic reciprocity), we have been able to reduce the computational expense needed to generate the CyberShake ensembles by more than a factor of 1,000 relative to direct calculations of source-excited wavefields (Graves et al., 2011; Wang & Jordan, 2014).

The CyberShake PSHA Study 13.4, started in April 2013, used the UCERF2 earthquake rupture forecast and calculated hazard curves for 286 sites in Southern California at frequencies up to 0.5 Hz. Approximately 240 million horizontal-component seismograms were calculated for each model. These runs evaluated the consistency of alternative wave propagation codes and the impact of alternative community velocity models. This study, performed on NCSA's Blue Waters and TACC's Stampede using both CPUs and GPUs, was four times larger than the 2011 CyberShake study, but it was completed in approximately the same wall-clock time (~61 days). The CyberShake 13.4 study results have been used to calculate ground motions more rapidly and accurately for use in the California Integrated Seismic Network (CISN) ShakeAlert Earthquake Early Warning system (Böse et al., 2014) as well as for calculations in planned updates to model building codes. A new GPU-based strain Green tensor (SGT) solver was implemented that provided speedup by a factor of 110 in SGT calculations compared to the previous CPU-based code (Cui et al., 2013a).

HOW BLUE WATERS PROJECT STAFF HELPED

SCEC scientists and technical staff have worked closely with the Blue Waters staff to achieve a series of breakthroughs in computational earthquake physics. This technical collaboration involved development of GPU codes, integration of SCEC software into the Blue Waters environment, and optimization of the CyberShake runtime performance. It enabled SCEC to shorten the overall time by a factor of four, from eight weeks to two weeks. During the Blue Waters early access period, we used SCEC development and optimization of parallel, GPU-enabled earthquake wave propagation software supported by NCSA/University of Illinois at Urbana-Champaign Enhanced Intellectual Services for Petascale Performance (NEIS-P2). The Blue Waters and SCEC groups significantly improved performance by optimizing application job placement within the torus, through I/O improvements, and by tuning the number and type of simultaneous workflow jobs. This work included development of topology-aware placement that improved performance by 30% for 4,096 node jobs (131,072 core equivalents).

SCEC's CyberShake workflow is heterogeneous, first running a small number (hundreds) of large parallel strain Green's tensor calculations, then a large number (hundreds of millions) of loosely coupled serial tasks that synthesize seismograms from individual earthquakes. The Blue Waters technical group helped the SCEC team by integrating the Blue Waters scheduler with the GRAM-based job submission system running at SCEC. In this way, an SCEC server could submit tens of thousands of jobs to the Blue Waters scheduler entirely automatically using scientific workflow tools. Blue Waters staff assisted SCEC in batching many tasks together to reduce the number of jobs submitted to the Blue Waters queue and tuned the Blue Waters Lustre file system to improve scaling and congestion protection control for millions of tasks. Blue Waters staff also adjusted the job scheduling with extended reservations to supply a baseline number of nodes that allowed SCEC to complete the CyberShake computational phase in two weeks.

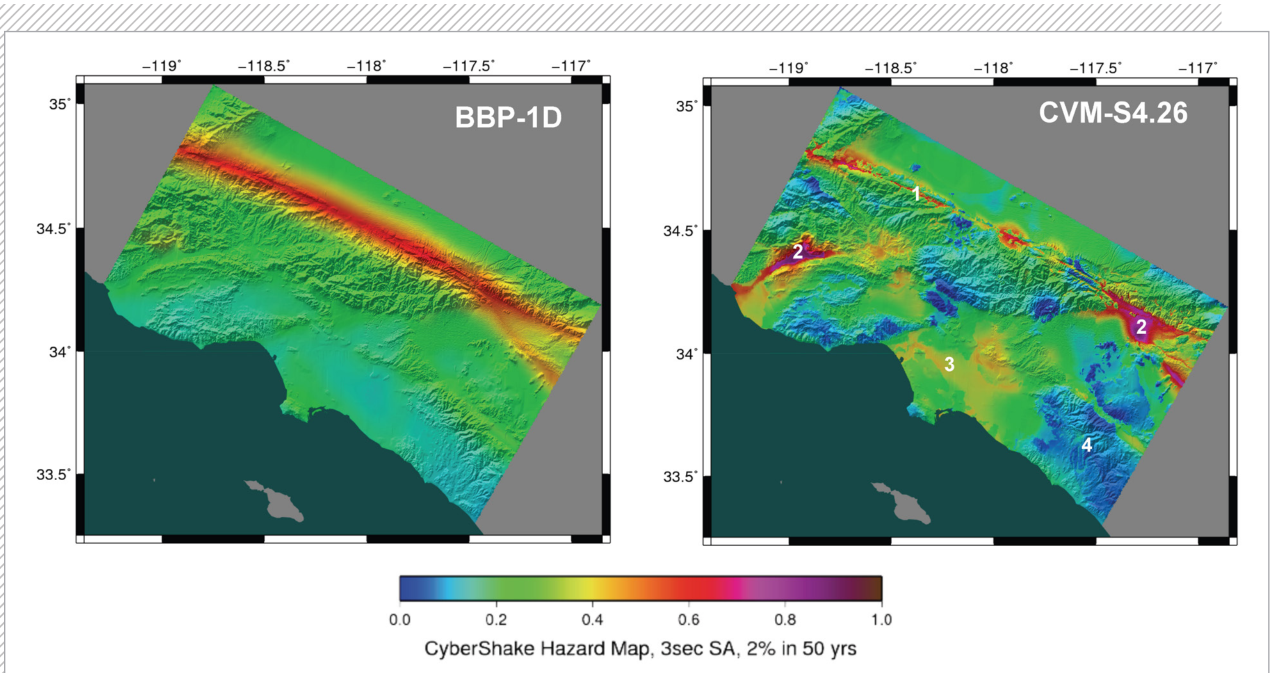

Figure 1. Two CyberShake hazard models for the Los Angeles region calculated on Blue Waters using a simple 1D earth model (left) and a more realistic 3D earth model (right). Seismic hazard estimates produced using the 3D earth model show lower near-fault intensities due to 3D scattering, much higher intensities in near-fault basins, higher intensities in the Los Angeles basins, and lower intensities in hard-rock areas.

WHY THE RESEARCH MATTERS

SCEC's CyberShake scientific contributions to PSHA have the potential to change standard practices in earthquake engineering and emergency management. Specifically, this work is helping to reduce the total uncertainty in long-term hazard models, which has important practical consequences for the seismic provisions in building codes, especially for critical-facility operators. Models used in PSHA contain two types of uncertainty: aleatory variability that describes the intrinsic randomness of the earthquake-generating system, and epistemic uncertainty that characterizes our lack of knowledge about the system. In physics-based PSHA, we can reduce the overall uncertainty through an iterative modeling process that involves two sequential steps: (1) introduce new model components that translate aleatory variability into epistemic uncertainty, and (2) assimilate new data into these representations that reduce the epistemic uncertainty. The most important effects are 3D geological heterogeneities in Earth's crust, which scatter seismic wavefields and cause local amplifications in strong ground motions that can exceed an order of magnitude (Fig. 1). In empirical PSHA, an "ergodic assumption" is made that treats most of this variability as aleatory. In physics-based PSHA, crustal structure is represented by a 3D seismic velocity model—a community velocity model (CVM) in SCEC lingo—and the ground motions are modeled for specified earthquake sources using an anelastic wave propagation (AWP) code, which accounts for 3D effects. As new earthquakes are recorded and other data are gathered (e.g., from oil and gas exploration), the CVM is modified to fit the observed seismic waveforms, which reduces the epistemic uncertainties in the earthquake simulations.

WHY BLUE WATERS

We use Blue Waters to perform large-scale, complex scientific computations involving thousands of large CPU and GPU parallel jobs, hundreds of millions of short-running serial CPU tasks, and hundreds of terabytes of temporary files. These calculations are beyond the scale of available academic HPC systems, and in the past they required multiple months of time to complete using NSF Track-2 systems. Using the well-balanced capabilities of Blue Waters CPUs, GPUs, disks, and system software, together with scientific workflow tools, SCEC's research staff can now complete CyberShake calculations in weeks rather than months. This enables SCEC scientists to improve methodology more rapidly as we work towards CyberShake calculations at the scale and resolution required by engineering users of seismic hazard information.

PRE-PETASCALE PREPARATION

The CyberShake scientific and computational methodology has been under development since 2007. We have continued to optimize and enhance both the seismic research codes and the middleware to perform simulations on cutting-edge systems. CyberShake calculations have been run on a sequence of NSF computers including NCSA's Mercury, TACC's Ranger, NICS's Kraken, and TACC's Stampede.

We made a number of improvements to prepare CyberShake for execution on Blue Waters. Very high-performance parallel software, called AWP-ODC, was integrated into the CyberShake platform to perform the large parallel CyberShake calculations. A GPU version of AWP-ODC was developed and optimized for use on Blue Waters GPU nodes. SCEC staff wrote a new version of the seismogram synthesis code which reduced the I/O requirements by more than 80%. Additionally, SCEC worked with the Pegasus-WMS middleware developers to create an MPI wrapper for the seismogram synthesis jobs called Pegasus-MPI-Cluster. With this software tool, workflows consisting of many loosely coupled, serial seismogram synthesis tasks look like a single, large MPI job to the Blue Waters scheduler, and can execute in the Blue Waters environment without modifications to the scientific codes. These improvements prepared SCEC to execute CyberShake on Blue Waters, calculating physics-based seismic hazard maps four times faster than previously possible.

LOOKING FORWARD TO THE NEXT TRACK-1 SYSTEM

Seismic hazard analysis (SHA) is the scientific basis for many engineering and social applications for performance-based design, seismic retrofitting, resilience engineering, insurance rate setting, disaster preparation, emergency response, and public education. All of these applications require a probabilistic form of seismic hazard analysis (PSHA) to express the deep uncertainties in the prediction of future seismic shaking. As currently applied, PSHA is largely empirical, based on parametric representations of fault rupture rates and ground motions that are adjusted to fit the available data. The data are often very limited, especially for large-magnitude earthquakes. For example, no major earthquake (M>7) has occurred on the San Andreas Fault in California during the post-1906 instrumental era. Consequently, the forecasting uncertainty of current PSHA models, such as the U.S. National Seismic Hazard Mapping Project (NSHMP), is very high. Reducing the uncertainty in PSHA through more accurate deterministic simulations of earthquakes at higher frequencies has many societal benefits, ranging from better safety designs to saving the costs associated with overdesign.

With access to a continuing sequence of Track-1 level systems, SCEC is striving to extend the upper frequency limit of physics-based PSHA from 0.5 Hz in current models to 2 Hz in the 3-5 year timeframe and 5 Hz in the 5-10 year timeframe. This will enable damage predictions to be made for most of the buildings in the Los Angeles area (current predictions are limited to only tall buildings). The long-term goal is to extend physics-based PSHA across the full bandwidth needed for seismic building codes (i.e. up to 10 Hz) and to extend the geographic area of CyberShake to include all of California.

Additionally, physics-based PSHA results can be used in operational earthquake forecasting (OEF), quantifying the change in earthquake probability due to recent observational information. OEF has the potential to help communities improve their preparation for potentially destructive earthquakes, and the CyberShake results can be used to perform short-term physics-based OEF.

COMMUNITY IMPACT

In February 2014, at our request, NCSA made a policy change regarding workflow management software that allowed us to use the heterogeneous architecture of Blue Waters much more effectively, reducing the total CyberShake makespan to 342 hours (~14 days). Hazard maps for 2% probability of exceedance in 50 years from the CyberShake 14.2 study are shown in Fig. 1. A comparison of the more accurate 3D model with the 1D model illustrates the importance of complex geological structures in governing the amplitudes of strong ground motions.

This success illustrates how domain scientists working with supercomputer managers and engineers can achieve better vertical integration of HPC systems and improve their efficiency and effectiveness. In particular, the use of scientific workflow tools on Blue Waters has helped to increase the scale of the calculations by two orders of magnitude over the last five years without increased personnel.

If the CyberShake method demonstrably improves on current PSHA methods, it may impact PSHA users including scientific, commercial, and governmental agencies like the USGS. For seismologists, CyberShake provides new information about the physics of earthquake ground motions, the interaction of fault geometry, 3D earth structure, ground motion attenuation, and rupture directivity. For governmental agencies responsible for reporting seismic hazard information to the public, CyberShake represents a new source of information that may contribute to their understanding of seismic hazards, which they may use to improve the information they report to the public. For building engineers, CyberShake represents an extension of existing seismic hazard information that may reduce some of the uncertainties in current methods, which are based on empirical ground motion attenuation models.

PUBLICATIONS

Cui, Y., K.B. Olsen, J. Zhou, P. Small, A. Chourasia, S.M. Day, P.J. Maechling, T.H. Jordan. (2012). Development and optimizations of a SCEC community anelastic wave propagation platform for multicore systems and GPU-based accelerators. *Seismol. Res. Lett.*, 83:2 (2012), p. 396, doi:10.1785/gssrl.83.2.231.

Cui, Y., E. Poyraz, K. B. Olsen, J. Zhou, K. Withers, S. Callaghan, J. Larkin, C. Guest, D. Choi, A. Chourasia, Z. Shi, S. M. Day, P. J. Maechling, and T. H. Jordan, Physics-based seismic hazard analysis on petascale heterogeneous supercomputers. *The International Conference for High Performance Computing, Networking, Storage and Analysis*, Denver, Colo., November 17-22, 2013, doi:10.1145/2503210.2503300.

Cui, Y., E. Poyraz, J. Zhou, S. Callaghan, P. Maechling, T. H. Jordan, L. Shih, and P. Chen, Accelerating CyberShake Calculations on the XE6/XK7 Platform of Blue Waters. *2013 Extreme Scaling Workshop*, Boulder, Colo., August 15-16, 2013, doi:10.1109/XSW.2013.6.

Roten, D., and K. B. Olsen, 3D Simulations of M 7 Earthquakes on the Wasatch Fault, Utah, Part II: Broadband (0–10 Hz) Ground Motions and Nonlinear Soil Behavior. *Bull. Seismol. Soc. Am.*, 102:5 (2012), pp. 2008-2030, doi:10.1785/0120110286.

Shi, Z., and S. M. Day, Rupture dynamics and ground motion from 3D rough-fault simulations. *J. Geophys. Res. Solid Earth*, 118:3 (2013), pp. 1122-1141, doi:10.1002/jgrb.50094.

Taborda, R., and J. Bielak, Ground-Motion Simulation and Validation of the 2008 Chino Hills, California, Earthquake. *Bull. Seismol. Soc. Am.*, 103:1 (2013), pp. 131-156, doi:10.1785/0120110325.

Wang, F., and T. H. Jordan, Comparison of probabilistic seismic hazard models using averaging-based factorization. *Bull. Seismol. Soc. Am.*, (2014), doi: 10.1785/0120130263.

This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.