Performance Modeling Codes for the QuakeSim Problem Solving Environment

Jay Parker[1], Andrea Donnellan[1], Gregory Lyzenga[1,2], John Rundle[3], and Terry Tullis[4]

[1] Jet Propulsion Laboratory/California Institute of Technology,
Pasadena, California
Jay.W.Parker@jpl.nasa.gov
http://www-aig.jpl.nasa.gov/public/dus/quakesim
[2] Department of Physics, Harvey Mudd College,
Claremont, California,
lyzenga@hmc.edu.
[3] Department of Physics,
University of California Davis,
Davis, California
rundle@physics.ucdavis.edu.
[4] Department of Geological Sciences
Brown University, Providence, Rhode Island
terry_tullis@brown.edu

Abstract. The QuakeSim Problem Solving Environment uses a web-services approach to unify and deploy diverse remote data sources and processing services within a browser environment. Here we focus on the high-performance crustal modelling applications that will be included in this set of remote but interoperable applications. PARK is a model for unstable slip on a single earthquake fault represented as discrete patches, able to cover a very wide range of temporal and spatial scales. GeoFEST simulates stress evolution, fault slip and visco-elastic processes in realistic materials. Virtual California simulates fault interaction to determine correlated patterns in the nonlinear complex system of an entire plate boundary region. Pattern recognition tools extract Karhunen-Loeve modes and Hidden Markov state models from physical and virtual data streams. Sequential code benchmarking demonstrates PARK computes 15,000 patches for 500 time steps in under 8 hours (SGI Origin 3000), GeoFEST computes 50,000 tetrahedral elements for 1000 steps in under 14 hours (Sun Workstation), and Virtual California computes 215 fault segments for 10,000 time steps in under 0.5 hours (Pentium III). QuakeSim goals for June 2004 are to deploy MPI parallel codes that compute 400,000 patches (PARK), 16,000,000 tetrahedra (GeoFEST) and 700 segments (Virtual California) in essentially the same wallclock time, incorporating powerful tools such as stress field multipoles and the ESTO/ PYRAMID mesh partitioning and refinement tools.

1 Introduction

The full objective over this three-year program (begun April 2002) is to produce a system to fully model earthquake-related data. Components of this system include:

1. A database system for handling both real and simulated data
2. Fully three-dimensional finite element code (FEM) with adaptive mesh generator capable of running on workstations and supercomputers for carrying out earthquake simulations
3. Inversion algorithms and assimilation codes for constraining the models and simulations with data
4. A collaborative portal (object broker) for allowing for seamless communication between codes, reference models, and data
5. Visualization codes for interpretation of data and models
6. Pattern recognizers capable of running on workstations and supercomputers for analyzing data and simulations

In order to develop a solid earth science framework for understanding and studying of active tectonic and earthquake processes, this task develops simulation and analysis tools to study the physics of earthquakes using state-of-the-art modeling, data manipulation, and pattern recognition technologies. We develop clearly defined accessible data formats and code protocols as inputs to the simulations. These are adapted to high-performance computers because the solid earth system is extremely complex and nonlinear resulting in computationally intensive problems with millions of unknowns. With these tools it will be possible to construct the more complex models and simulations necessary to develop hazard assessment systems critical for reducing future losses from major earthquakes. Use of multiple data types in a coherent modeling effort is illustrated in [1].

The system for unifying the data sources, modeling codes, pattern analysis and visualization is being constructed on a web-services model, and is described in a companion paper by Fox et al. and [2-6]. Some of the applications and their data linkages are shown in Figure 1. Within our web services framework, these applications may be running multiple instances on separate machines located anywhere on the internet. Data sources (as in [1], where GPS and inSAR data is used) and fault databases such as the one under development by [7-9] are also available through this portal system as distributed objects with clear interface definitions.

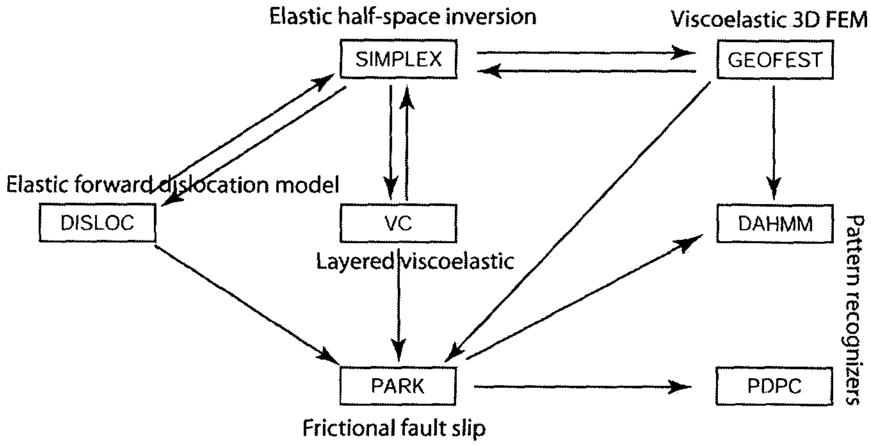

Figure 1. Linkages between programs that maybe run separately or coupled to other programs. Low-level movement of actual data occurs between programs. There is also communication of metadata and information about the files being transferred.

Three of these codes are targeted for improved performance, chiefly through design changes that make them efficient high-performance parallel codes. These codes are PARK, a boundary-element based code for studying unstable slip at the Parkfield segment of the San Andreas fault, Virtual California, which simulates the dynamic interaction of hundreds of fault segments comprising the active tectonics of California, and GeoFEST, a fully three-dimensional finite element code to model active tectonics and earthquake processes. Together with an adaptive mesh generator that constructs a mesh based on geometric and mechanical properties of the crustal structure, the GeoFEST system makes it possible to efficiently model time-dependent deformation of interacting fault systems embedded in a heterogeneous earth structure.

2 Scientific significance

The full interoperable system will allow users from many environments to discover and exploit a very wide range of applications, models, and physical measurements in distributed databases. This is seen as of crucial importance for gaining full advantage of massive new national programs for gathering new regional and global data sets, such as the Plate Boundary Observatory, Earthscope, and NASA orbiting inSAR missions.

PARK is a model for unstable slip on a single earthquake fault. Because it aims to capture the instability it is designed to represent the slip on a fault at many scales, and to capture the developing seismic slip details over an extraordinary range of time scales (subseconds to decades). Its simulation of the evolution of fault rupture is the most realistic of the tools considered here. When transformed into an efficient parallel simulation, it will be the tool of choice for researchers seeking to determine the nature and detectability of earthquake warning signals such as surface strains and patterns of microseismicity. This is the first earthquake simulation code to seek enhanced scalibility and speed by employing a multipole technique. The multipole experience gained here will also be transferrable to the Virtual California code and other boundary element simulations. The power of massive parallel computing is required for this problem in order to support many small slip patch elements in order to cover the nucleation scale that initiates the instability.

GeoFEST simulates stress evolution, fault slip and plastic/elastic processes in realistic materials. The product of such simuluations are synthetic observable time-dependent surface deformation on scales from days to decades. Diverse types of synthetic observations will enable a wide range of data assimilation and inversion techniques for ferreting out subsurface structure and stress history. In the short term, such a tool allows rigorous comparisons of competing models for interseismic stress evolution, and the sequential GeoFEST system is being used for this at JPL and UC Davis. Parallel implementation is required to go from local, single-event models to regional models that cover many earthquake events and cycles.

Virtual California simulates fault interaction to determine correlated patterns in the nonlinear complex system of an entire plate boundary region. The evolution of these patterns enables forecasts of future large events. The model produces synthetic seismicity and surface deformation, enabling an eventual data assimilation system for exploiting regional data collection. Capturing the nonlinear pattern dynamics of the fault system along a plate boundary implies the realization of a digital laboratory which allows understanding of the mechanisms behind the observations and patterns. Our technology development aims to produce and demonstrate a scalable cluster code. When that is deployed researchers will be able to create and verify patterns down to smaller spatial scales, which will enable cross-scale parameterization and validations which will in turn enable plate-boundary system analysis and greatly enhanced forcasts of large earthquakes.

Pattern analysis methods are another tool type we are developing. Hidden Markov methods are one method incorporated into this framework [10]. Another method [11,12] bins many decades of seismic activity on a gridded representation of Califorinia. Eigensystem analysis reveals clusters of correlated space and time activity, which have been subjected to a phase dynamics forecast. When this method attains parallel speedup we will produce better forecasts and enable speedy tests of earthquake correlations and predictions. This will be due to the ability to use much smaller geographic cell sizes and so forecast the frequent magnitude 4 earthquakes, not just the rare magnitude 6 events.

3 Technology Accomplishments

The development of the three performance codes is at an early stage. Our planned milestones (including code parallel performance goals) are posted at the web site http://www-aig.jpl.nasa.gov/public/dus/quakesim/milestones.html.

The Milestone E code baseline benchmark marks the beginning of moving the three codes, PARK, geoFEST, and Virtual California to high-performance parallel computers, enabling the simulation of interacting fault systems leading to earthquakes. The scientific significance and performance results have been reported in http://www-aig.jpl.nasa.gov/public/dus/quakesim/MilestoneE_Baselines.pdf and are summarized in the following table:

Table 1. Summary of performance codes sequential computing time on baseline problem with resolution, time steps indicated

Code	Resolution/ Time Steps	Platform (processors)	Wallclock time
PARK	15,000 patches/ 500 steps	SGI Origin 3000 (1)	7h53m
geoFEST	50,000 elements/ 1000 steps	Sun workstation (1)	13h43m
Virtual California	215 segments/ 10,000 steps	Pentium III (1)	5m38s (stress GF) 15m 48s (steps)

Performance for PARK on a single processor can handle a fault region with 15,000 fault elements over 500 time steps in just under 8 hours. But the problem of determining possible earthquake precursors requires a finer sampling. This project is committed to demonstrating the computation of 400,000 elements on 1024 processors, enabling 50,000 time steps in 40 hours by June, 2004. Such performance will be attained by combining efficient parallel problem decomposition and by exploiting an NlogN multipole technique [13] to reduce the currently N^2 element interaction compuations. This multipole method has been demonstrated in parallel astrophysics simulations with considerable success.

Performance for Virtual California [14] has been demonstrated at 215 interacting fault segments for 10,000 time steps in 20 minutes on a 1 GHz Linux workstation. This type of simulation allows determination of correlated patterns of activity that can be used to forecast seismic hazard, with relevance for earthquakes of Magnitude 6. Finer sampling may allow verification of such forecasts by including much smaller, more frequent earthquakes. An initial parallel demonstration appears in [15] The aims of this task are to validate parallel scaling of the Virtual California code, and evaluate the potential for the multipole technique pioneered in the PARK demonstration.

GeoFEST [16] is a direct implementation of stress-displacement volumetric finite elements. The mechanics include elasticity and visco-elastic deformation. Current workstation capacity can solve for the deformation and stress evolution due to a single-fault rupture, such as the Northridge event, using ~100,000 finite element equations. We seek to analyze the modes of interaction of the entire Southern California system of interacting faults, covering a portion of the crust ~1000 km on a side. Such a simulation would require ~5M equations to determine first-order effects, and certainly higher density for faithful representation of the time-dependent stress field. Current techniques require running such a model through thousands of time steps to attain a stable background stress field and assess the patterns of fault interactions. These considerations motivate tailoring the code toward hundreds of processors to attain solutions in reasonable turnaround time.

Our team has created a meshing tool with mesh generation and adaptation capabilities, while the ESS project has developed a parallel adaptive meshing library (the Pyramid library) for supporting parallel dynamic adaptive meshing of unstructured meshes. Techniques of parallel adaptive meshing, combined with local error estimates, can be effectively used to compute an initial state of the displacement and stress fields, and to significantly reduce the computational load of computing the evolving stress field in the finite element modeling.

We are therefore in an excellent position to evaluate the ESS Adaptive Mesh Refinement (AMR) library with our finite element code, and to combine the strengths of both meshing tools into an integrated adaptive meshing library for a scalable parallel computing environment. The integrated adaptive meshing library will support the entire process of an unstructured parallel application, including initial mesh generation, dynamic parallel mesh adaptation with efficient mesh quality control, and dynamic load balancing. This will be a valuable tool for improving the computational efficiency of our finite element modeling.

Current GeoFEST performance computes 1000 time steps, and 50,000 volumetric finite elements with an unstructured mesh with large range of element sizes, in just under 13 hours on a 400 Mhz Sun Solaris computer. We plan to demonstrate problems with a complex embedded system of faults with 16,000,000 finite elements with the same number of time steps and solution time on 880 processors. We will use PYRAMID for parallel decomposition and adaptive mesh refinement, and a parallel iterative Conjugate Gradient method for the sparse matrix solution.

4 Status and Plans

In this fractional first year we have

7. Agreed upon a software engineering plan,
7. Selected a review board of regarded earthquake and computational science experts,
8. Established a public web page area (http://www-aig.jpl.nasa.gov/public/dus/quakesim/) to allow documentation and access to our products,
9. Demonstrated and documented the serial performance of three important codes believed to be scalable to high performance computing.
10. Established requirements and design policy for a framework that will make community earthquake codes interoperable and deliverable to the scientific community.

In the coming year we will develop the three main codes for efficient parallel performance. Virtual California and geoFEST will be adapted for Riva and Parvox parallel rendering. Also, geoFEST will be linked with the PYRAMID mesh refinement library. We will develop the interoperability framework and tie in a fault database, mesh generator and the geoFEST code and demonstrate the prototype framework with a combined simulation.

This program differs from many past efforts in that the source code for the three performance codes is publicly available through the web site listed (geoFEST is subject to a no-cost license arrangement to US citizens through the Open Channel Foundation). The applications will also be available through the QuakeSim interoperability framework, including a supported ability to invoke the applications through the web-based portal for large problem solutions for authorized investigators. These modes of access will be supported as the applications are extended to efficient parallel software.

Acknowlegments: This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

1. Donnellan, A., J. Parker, and G. Peltzer, Combined GPS and InSAR models of postseismic deformation from the Northridge earthquake, *Pure and Appl. Geophys.*, 2261–2270, 2002.
2. Donnellan, Andrea, Geoffrey Fox, John Rundle, Dennis McLeod, Terry Tullis, Lisa Grant, Jay Parker, Marlon Pierce, Gregory Lyzenga, Anne Chen, John Lou, The Solid Earth Research Virtual Observatory: A web-based system for modeling multi-scale earthquake processes, *EOS Trans, AGU, 83*(47), Fall Meeting Suppl., Abstract NG528-08, Invited, 2002.
3. Fox, Geoffrey, Hasan Bulut, Kangseok Kim, Sung-Hoon Ko, Sangmi Lee, Sangyoon Oh, Shrideep Pallickara, Xiaohong Qiu, Ahmet Uyar, Minjun Wang, Wenjun Wu, Collaborative Web Services and Peer-to-Peer Grids to be presented at 2003 Collaborative Technologies Symposium (CTS'03). http://grids.ucs.indiana.edu/ptliupages/publications/foxwmc03keynote.pdf
4. Fox, Geoffrey, Hasan Bulut, Kangseok Kim, Sung-Hoon Ko, Sangmi Lee, Sangyoon Oh, Xi Rao, Shrideep Pallickara, Quinlin Pei, Marlon Pierce, Ahmet Uyar, Wenjun Wu, Choonhan Youn, Dennis Gannon, and Aleksander Slominski, "An Architecture for e-Science and its Implications" in Proceedings of the 2002 International Symposium on *Performance Evaluation of Computer and Telecommunications Systems*, edited by Mohammed S.Obaidat, Franco Davoli, Ibrahim Onyuksel and Raffaele Bolla, Society for Modeling and Simulation International, pp 14-24 (2002). http://grids.ucs.indiana.edu/ptliupages/publications/spectsescience.pdf
5. Fox, Geoffrey, Sung-Hoon Ko, Marlon Pierce, Ozgur Balsoy, Jake Kim, Sangmi Lee, Kangseok Kim, Sangyoon Oh, Xi Rao, Mustafa Varank, Hasan Bulut, Gurhan Gunduz, Xiaohong Qiu, Shrideep Pallickara, Ahmet Uyar, Choonhan Youn, *Grid Services for Earthquake Science, Concurrency and Computation: Practice and Experience in ACES Special Issue*, 14, 371-393, 2002. http://aspen.ucs.indiana.edu/gemmauisummer2001/resources/gemandit7.doc

6. Fox, Geoffrey, Hasan Bulut, Kangseok Kim, Sung-Hoon Ko, Sangmi Lee, Sangyoon Oh, Xi Rao, Shrideep Pallickara, Quinlin Pei, Marlon Pierce, Ahmet Uyar, Wenjun Wu, Choonhan Youn, Dennis Gannon, and Aleksander Slominski, "An Architecture for e-Science and its Implications" in Proceedings of the 2002 International Symposium on *Performance Evaluation of Computer and Telecommunications Systems*, edited by Mohammed S.Obaidat, Franco Davoli, Ibrahim Onyuksel and Raffaele Bolla, Society for Modeling and Simulation International, pp 14-24 (2002). http://aspen.ucs.indiana.edu/gemmauisummer2001/resources/gemandit7.doc
7. Gould, M., Grant, L., Donnellan, A., and D. McLeod. The GEM Fault Database: A Preliminary report on design and approach. *Proceedings and Abstracts*, SCEC Annual Meeting, p.75-76, 2002.
8. Grant, L. B. and Gould, M. M. Paleoseismic and geologic data for earthquake simulation. Computational Science, Data Assimilation, and Information Technology for Understanding Earthquake Physics and Dynamics, 3rd ACES International Workshop, Maui, Hawaii, USA, p.30, 2002.
9. Grant, L. B. and M. M. Gould. Assimilation of paleoseismic data for earthquake simulation. Submitted to Pure and Applied Geophysics, 2002.
10. Granat, R., and A. Donnellan, A hidden Markov model tool for geophysical data exploration, *Pure and Appl. Geophys*, 2271–2284, 2002.
11. Rundle, JB, W Klein, KF Tiampo, Andrea Donnellan, and GC Fox, Detection and Analysis of Space-Time Patterns in Complex Driven Threshold Systems, paper submitted to the 2003 International Conference on Computational Science, Melbourne, AU June 2-4, 2003.
12. Tiampo, KF, JB Rundle, S. McGinnis and W. Klein, Pattern dynamics and forecast methods in seismically active regions, *Pure and Appl. Geophys.*, 159, 2429-2467, 2002.
13. Tullis, T. E., J. Salmon, N. Kato, and M. Warren, The application of fast multipole methods to increas the efficiency of a single fault numerical earthquake model, *Eos. Trans. Am. Geophys, Union, Fall Meeting Suppl.*, 80, F924 1999.
14. Rundle, JB, PB Rundle, W Klein, J Martins, KF Tiampo, A Donnellan and LH Kellogg, GEM plate boundary simulations for the Plate Boundary Observatory: Understanding the physics of earthquakes on complex fault systems, *Pure and Appl. Geophys.*, 159, 2357-2381 2002.
15. Tiampo, KF, JB Rundle, S. Gross and S. McGinnis, Parallelization of a large-scale computational earthquake simulation program, *Concurrency & Computation, Practice and Experience*, 14, 531-550, ACES Special Issue, 2002.
16. Parker, Jay, Gregory Lyzenga, Jin-Fa Lee, John Lou, and Andrea Donnellan, Technologies for larger crustal studies by finite element simulation, Southern California Earthquake Center Annual Meeting, Oxnard, CA, September 8-11 2002.