The Importance of High-Performance Computing in Earthquake System Science Research

Scott Callaghan and Philip Maechling

Simulation is often referred to as the “third pillar of science”. In complement with theory and experiment, numerical simulation is able to probe phenomena and domains which are difficult or impossible to investigate directly. This is certainly true in earthquake system science, where direct measurements are limited to the near-surface, laboratories are restricted to small-scale experiments, and large events which constitute the primary hazard to the built environment are rare. Many elements of the earthquake problem rely on large-scale numerical modeling and simulation, often referred to as high performance computing (HPC). This includes both physical investigations such as rupture dynamics, wave propagation, and tomography, and applications such as multi-cycle earthquake simulators, seismic hazard analysis, and operational earthquake forecasting. We will use HPC as an umbrella term for all types of large-scale simulation requiring multi-core computing, including but not limited to high throughput computing, cloud computing, and big data. In this white paper, we describe how HPC will help future NSF earthquake research centers perform groundbreaking research.

The Impact of High Performance Computing

HPC has led to many key developments in earthquake system science in the past two decades. The 2004 TeraShake simulations dramatically illustrated waveguide and basin amplification effects of a southern San Andreas event for the Los Angeles area. Simulations of scenario earthquakes have been used to drive earthquake preparedness through exercises and public awareness, such as ShakeOut (2008) in Southern California, HayWired (2018) in the San Francisco Bay Area, and Rose Canyon (2020) in San Diego. Simulated annealing methods were used in the UCERF3 grand inversion to develop state-of-the-art earthquake rupture forecasts (2014, 2015). Machine learning algorithms have been very successful at improving catalog completeness at low magnitudes. Tomographic inversions, impossible without access to large computational and storage resources, have improved the accuracy of velocity models, which in turn enhances the predictive power of ground motion simulations.

As supercomputers become more powerful, modelers are able to integrate more complex physics, increase spatial and temporal resolution, and simulate higher frequency ground motion, which can bring the models into better agreement with observations. For example, in the past several years modelers at the Southern California Earthquake Center (SCEC) have enhanced their wave propagation simulations with frequency-dependent attenuation, heterogeneity in velocity and fault models, and topography, and are currently validating the addition of nonlinear source and site effects. These improvements have resulted in ground motion models which better match observations at frequencies up to 5 Hz, of interest to building engineers.

This, in turn, makes simulation products more useful for end users and decision-makers. For example, CyberShake, SCEC’s physics-based seismic hazard analysis platform, is able to more fully capture basin amplification and directivity effects than empirical ground motion prediction equations. As a result, CyberShake results are planned for inclusion in the City of Los Angeles building codes, showing their potential to reduce seismic risk and improve community resilience.

HPC can and should continue to play an important role in earth science research. To that end, we outline some recommendations below.

Successful HPC Program Elements

Integrated support for HPC in a future earthquake research center is needed to continue to improve our earth structure and earthquake process models. We have identified several elements for a successful HPC-intensive research program:

- *Access to HPC resources.* Medium- and large-scale university, NSF, and DOE HPC and cloud resources are typically allocated through a competitive proposal process. A future center must have access to dedicated computing resources or be able to compete successfully for these shared open-science resources.
- *Scalable software.* In order to take best advantage of large-scale HPC resources, researchers must have software which is capable of running simultaneously on a large number of cores. This includes tightly coupled parallel codes, embarrassingly parallel codes, and suites of ensemble simulations.
- *Middleware and integration.* HPC codes are frequently just one part of a software ecosystem which includes middleware such as workflow tools, containers, and libraries. Codes from multiple research groups often must be integrated to generate more impactful results. For example, a velocity mesh generator, a wave propagation code, a risk model, and a visualization tool might need to be integrated to produce imagery of risk estimates from an event. A major contribution of a future center is the ability to bring together multiple domain codes and software tools to create scientific products using the best available end-to-end science.
- *Reproducibility.* The scientific method requires results to be reproducible. In an HPC context, reproducibility relies on capturing sufficient software and data, metadata, and provenance to enable simulations to be repeated.
- *Verification.* A test-based process for confirming that code is working correctly is required. This is especially important for HPC codes, as they may be run on a variety of types of hardware with diverse software stacks. Verification is essential for end users to trust simulation results.
- *Validation*. Simulations are only useful insofar as they model the earth. Validation exercises which compare models to observation, and update the models appropriately, must be a key part of a future research center's HPC efforts. End users must be convinced that the models are accurate before simulation-based data products can have a role in influencing their decision making.
- *Data management.* While seismology has well-established observational data management practices, simulation data management practices are often project-specific. Large simulations may produce terabytes to petabytes of data. For this data to be maximally useful to the community, it should be archived, searchable, and delivered upon request, and phased out or replaced with newer results as the science advances. Reproducibility, verification, and validation all also rely on well-managed data. We recommend useful data products are assigned DOIs. A future center should have a cogent plan for managing both observational and simulation-based data products, considering the needs of both modelers and end users.

HPC Software Development

To deliver on the above elements, we recommend dedicated HPC software developers be part of a future earthquake research center. A team of HPC developers gives the center the ability to integrate models and methods and to perform large research calculations. Through our experiences at SCEC, we have found that a collaboration which involves both HPC software developers and research scientists working on large-scale simulations is very effective. Each brings something to the table: research scientists can improve the accuracy of the simulations by adding model complexity, and software developers can optimize and integrate software to run efficiently at large scales. The HPC world is a quickly-changing one, with new hardware platforms and software tools emerging frequently. Advances in hardware mean that new physics and larger simulations become computationally tractable, but an active software development team is required for a future center to take advantage of new opportunities.

We recognize that the funding of software development staff is a particularly challenging issue for a research center. We recommend that a future earthquake research center should include dedicated funding for HPC software development. One potential avenue for this would be to have the NSF CSSI program participate in funding research computing for a future center. The specific HPC skills needed will depend on the research focus of the center. We recommend avoiding an arrangement in which funding for HPC software development is limited to the duration of a specific research center project. In this case, if the project funding is disrupted, it becomes difficult to maintain essential research codes, computational capabilities, and allocations on large systems.

Looking Ahead

One of the expectations for a future earthquake research center is to accomplish goals beyond the reach of individual research teams. High performance computing is a critical part of modern scientific collaborative research. The ability to integrate and run high performance scientific software opens up advanced research capabilities for a center. Future earthquake research centers may develop advanced modeling capabilities, release scientific software to the research community, and run large community models and simulations. For a future earthquake research center to reap the benefits of a productive computational research program, access to HPC resources, scalable codes, software developers, and adequate funding must be integrated into the organization and structure of the center.