Computational challenges in seismology

Jacobo Bielak and Omar Ghattas

Computational Mechanics Laboratory Department of Civil and Environmental Engineering Carnegie Mellon University, Pittsburgh, PA 15213, USA (e-mail: Jacobo_Bielak@love.cml.cs.cmu.edu; phone: 412 268 2958; fax: 412 268 7813).

General

Seismology and engineering seismology have been dramatically affected by the emergence of high-performance computing. A recent report by the National Research Council's Committee on Seismology (NRC, 1996) is entirely devoted to surveying the status and potential of high-performance computing activities in this field. Here we summarize several applications and focus on one particular problem, the modeling and forecasting of complex earthquake ground motion in large basins, in order to illustrate the computational resources that are required for tackling successfully this type of problem.

Computational challenges in seismology span a wide range of disciplines and often have significant economic and social implications, such as the mitigation of seismic hazards, treaty verification for nuclear weapons, and increased discovery of economically recoverable petroleum resources. In the realm of seismic hazard mitigation alone it is well to recall that despite continued progress in building code development in the US and abroad, the 1994 Northridge Earthquake, the costliest natural disaster in the US, resulted in 58 deaths and caused on the order of 30 billion dollars of damage. The 1995 Hyogoken-Nambee Earthquake had a magnitude 6.8, similar to that of the Northridge Earthquake, but its source of energy release was much closer to the Earth's surface and in closer proximity to a highly populated area, the city of Kobe and surrounding region in the Osaka Bay. The toll from the earthquake exceeded 5000 dead and 25,000 injured. More than 200,000 houses, about 10% of the total stock within the affected area, were damaged, including more than 80,000 collapsed. Estimated losses exceed 200 billion dollars. One critical first step that remains far from being fully understood in the process of designing new facilities and retrofitting existing structures is to be able to forecast the earthquake ground motion to which a structure will be exposed during its lifetime. Until such forecasting can be done reliably complete success in the design process will not be fully achieved.

Computational problems in seismology concern the recording, analysis, and simulation of seismic waves in the Earth. At present the major challenges fall into two categories. First, there are problems for which the underlying physics and mathematics are adequately understood but are nevertheless intractable because of computational limitations. For example, seismologists engaged in petroleum exploration and development know how to construct accurate three-dimensional images of subsurface structures using prestack depth migration techniques. There would be tangible economic returns from exploiting this well-understood methodology, but current supercomputers are not fast enough to make the computations feasible except for limited regions of the crust. The problem of computing the ground motion of large sedimentary basins during earthquakes is another example of an important problem in which the physical and mathematical formulations are well-understood. But three-dimensional calculations that involve realistic models and cover the full range of frequencies of interest to structural engineers are not yet possible. We elaborate on this problem in the next section.

Second, there are research problems in which high-performance computing resources are required to advance scientific understanding of seismic wave generation and propagation. Modeling of earthquake rupture processes in the lithosphere represent an important example of this type of problem. Although it is widely recognized that earthquakes originate from slip along geologic faults in the Earth's crust, there is little understanding of the processes that control the distribution and occurrence of seismicity. This is a complex issue that spans a huge range of time scales and distances. With current supercomputers, only systems of very limited size can be investigated and only with rather rough approximations to the governing physics. One of the greatest challenges for this work is the need to treat the details of fracture on a minute scale, together with the large-scale dynamics of slip along the entire fault. For simulations over an entire earthquake cycle (about 100 years for portions of the San Andreas Fault), the time scales must account for both the slow pace of stress accumulation (tens of years) and the rapid nature of fracturing (fractions of a second).

Another example in which seismic wave propagation calculations play an important role is in the verification of the Comprehensive Test Ban Treaty because seismic recording networks provide the primary technology for detecting nuclear explosions on a global basis. Because the recorded signal includes both source and path effects, the challenge is to identify clandestine explosions by comparing computational models of seismic wave propagation through the Earth with observed seismograms. These simulations require powerful supercomputers and need to be performed for a wide range of geologic terrains and seismic sources. Results of the simulations are then used to discriminate suspected nuclear explosions from the many natural and man-made events that produce measurable seismograms. In order to differentiate between such events will require the use of more realistic models and the need to acquire the deep understanding of the physical processes that produce small changes in recorded signals. More powerful computers than those available today will be required toward these ends.

Ground motion modelling in large basins

Modeling and forecasting earthquake ground motion in large basins is a highly challenging and complex task. The complexity arises from several sources. First, multiple spatial scales characterize the basin response: the shortest wavelengths are measured in tens of meters, whereas the longest measure in kilometers, and basin dimensions are on the order of tens of kilometers. Second, temporal scales vary from the hundredths of a second necessary to resolve the highest frequencies of the earthquake source up to a couple of minutes of shaking within the basin. Third, many basins have highly irregular geometry. Fourth, the soils' material properties are highly heterogeneous. Fifth, strong earthquakes give rise to nonlinear material behavior. And sixth, geology and source parameters are only indirectly observable, and thus introduce uncertainty into the modeling process.

Current capabilities

The Quake group at CMU has been working since 1993 on modeling earthquakes in large basins on parallel supercomputers (see www.cs.cmu.edu/~quake and Bao el al, 1998). Over the past five years we have used such distributed memory machines as the Thinking Machines CM-2, Intel Paragon XPS, TMC CM-5, and Cray T3D and T3E for the most computationally demanding portion of the modeling. The basic computational steps include: (1) generating a mesh that resolves the elastic wavelengths in the soil; (2) partitioning the mesh into subdomains that will be mapped to processors of a target parallel machine; (3) discretizing the governing elastic wave propagation equations by finite elements; (4) integrating these equations in time by an explicit finite difference scheme; and (5) visualizing and postprocessing the resulting ground motion. The main computational effort is in steps 3 and 4, and that is why these are done on parallel machines. However, as our problem size has increased, the remaining (currently sequential) steps are becoming bottlenecks.

Currently, our largest simulation stems from modeling the 1994 Northridge Earthquake in the San Fernando Valley. This simulation is characterized by: 100 million tetrahedral finite elements, 17 million grid points, 50 million resulting ODEs, 6000 time steps, 24Gb main memory, 256 Cray T3E processors, and 4 hours of runtime to simulate the wave propagation. The preprocessing computations (generating and partitioning the mesh, which is done just once for each basin model) are carried out on a large memory sequential machine (actually one processor of a DEC 8400 with 8Gb memory), and require several days of wall clock time. Actually much of this time was spent swapping, since the meshing phase required 10Gb of memory. Parallelizing the meshing component is a very difficult problem, but will be essential for scaling up to larger problems.

We note that these computations are based on unstructured mesh algorithms that we have developed over the last 5 years. Unstructured meshes introduce significant complexities on parallel machines. However, by adapting the cell size to the wavelength of propagating waves, they lead to a tremendous reduction in the number of grid points whenever the wavelengths vary significantly throughout the domain, as is the case of earthquake ground motion in basins. For example, the Northridge simulation described above would have required 3 billion grid points had a regular grid been used. Furthermore, due to the Courant condition, there would have been 5 times as many time steps as with the unstructured mesh.

Changes in level of science

There have been significant changes in the level of science that we have been able to do over the past 5 years. This is due to three reasons: first, the hardware we are using is about a factor of 50 more powerful (speed and memory) than what was available to us 5 years ago; second, the unstructured adaptive algorithms that we have developed have resulted in another factor of 50 reduction in computing time relative to available regular-grid methods; third, we have made use of much more accurate and detailed geological models of Southern California basins developed by seismologists with whom we collaborate. These three significant advances have enabled us to create detailed and realistic simulations of earthquakes in the San Fernando Valley. Results of these and other simulations by us and others show that large peak motions occur at great distances from an earthquake's epicenter because of resonance effects that amplify the local ground motions in the sedimentary basin. Moreover, the high spatial and temporal resolution of our simulations has enabled us to explain the extreme spatial variation of ground shaking that occurs across the valley in terms of peaks and nodal lines of modal valley response. This means that a small change in the structure's location can result in a dramatic change in response. In addition, since the response also varies rapidly with frequency, it follows that two slightly different structures located essentially at the same site can experience substantially different responses.

Future objectives

While providing much useful information, current earthquake simulations (ours and others') in many cases are not capable of adequately reproducing observed seismograms. The likely reason for this is that these models are based on a number of restrictive assumptions, made largely to reduce the computational effort. Our current Southern California model is restricted to (1) the San Fernando Valley; (2) a highest resolved frequency of 1Hz; (3) a softest soil of 200m/s shear velocity; (3) linear constitutive model; (4) simple inversion-based source; (5) non-inversion based geological model; (6) kinematic sources; and (7) deterministic material parameters. Relaxing these restrictions significantly increases the computational stakes. In particular: (1) extending our model to the entire Greater Los Angeles Basin will increase our problem size by a factor of 10; (2) increasing the highest resolved frequency to 4Hz (a value that is desirable for structural engineering purposes) implies a 64-fold increase in size (since size scales like the cube of resolved frequency); (3) moving to a constitutive law that is capable of representing the nonlinear, poroelastoplastic behavior of soil will result in at least two orders of magnitude increase in computational complexity; (4) solution of the inverse problem to determine basin and source parameters necessitates repeated solutions of the forward problem, perhaps one to two orders of magnitude; (5) probabilistic modeling of seismic events also requires repeated simulations for all possible earthquake scenarios in a particular basin, say of the order of 10-50 times. Our current computations require gigaflop computing. Enhancing our codes with these capabilities will push us into the realm of teraflop/petaflop computing. Significant work remains in developing algorithms for these more complex models that can scale to petaflop computers. As difficult as this is, these enhancements are essential if we are to generate more realistic simulations of past earthquakes, and thus forecast more confidently the ground motion due to future events. Because of the critical role that ground motion plays in infrastructure design, the accelerated availability of teraflop computers will have a direct impact on public safety and welfare.

References

[1] National Research Council (1996), *High-Performance Computing in Seismology, Committee on Seismology,* National Academy Press, Washington, D.C. 1996.

[2] H. Bao, J. Bielak, O. Ghattas, L.F. Kallivokas, D.R. O'Hallaron, J.R. Shewchuk, and J. Xu (1998), *"Large-scale simulation of elastic wave propagation in heterogeneous media on parallel computers,"* Comput. Methods Appl. Mech. Engrg. 152, 85-102.