Computational Infrastructure for Geodynamics:
An Interim Report

Introduction

Research on global-scale dynamical processes in the Earth and planets is increasingly reliant on sophisticated, large-scale computational models. Improved understanding of fundamental physical processes such as mantle convection and core dynamos in the terrestrial planets and satellites, fluid dynamics of the giant planets, crustal and lithospheric deformation on Earth, and global seismic wave propagation, are heavily dependent on better numerical modeling. Surprisingly, the rate-limiting factor for progress in these areas is not just computing hardware, as was once the case. For example, the computer hardware advances projected to occur in the next decade will likely make it feasible to realistically model mantle convection over a time period equivalent to the whole of Earth history. However, much more problematic is the fact that advances in software -- the codes we use to model global dynamics and the tools to interpret our results -- are not keeping pace with the dramatic improvements in the hardware.

The premise of this report is that the common challenges faced by future computations in geodynamics and related fields are best met in the environment of a broadly-based computational center, dedicated to modeling the fundamental processes that shape the solid Earth and planets. Specifically, we argue here that the time has come to develop a community framework to serve the combined computational goals of the Earth and planetary dynamics, global seismology, and crustal dynamics communities.

It is important to emphasize that the unified approach we are proposing here represents a genuinely radical departure from our current research environment. In the US and abroad, modeling tools in geophysics are usually developed and maintained by individual investigators, or small groups of investigators working in the same institution. But it is becoming increasingly difficult for any single individual, or even a small group, to keep up with sweeping advances in computing hardware, parallel processing software, and numerical modeling methodology. We have learned that our traditional or "heroic" approach to code development and maintenance is deficient in a number of respects, as described below.

To remedy this deficiency, we propose to establish the community organization necessary to develop and maintain computational infrastructure for geophysical modeling, broadly defined. This infrastructure will consist of hardware and technical support, senior-level numerical analysts and computational specialists, and a governing structure dedicated to modern software design, code validation, and user-friendly graphical interface development for the larger community of Earth and planetary scientists.

An important first step was to hold an international meeting for geophysicists and computing experts in July 2002, at Lake Tahoe, CA. This document summarizes the findings of that workshop. From the consensus developed at the workshop, our plan is to write an infrastructure proposal, with broad community input and support. We envision a virtual institute of scientists and software engineers, guided by a governing board, but also including a physical center hosting both a core software design team and necessary hardware support. The institute will require a budget of about $2 million per year, with the majority of the budget devoted to software engineering, and perhaps about 25% of the budget available for hardware acquisition and support. We will propose that NSF take the lead role in funding our initiative, but expect additional support from NASA, DOE, and other partners in the computer hardware and software industry.

Science Goals

I. Mantle Dynamics

Mantle convection remains the most intellectually challenging dynamical problem in geophysics. While it makes the most stringent demands, it also represents the greatest opportunity for a unified approach. Many of the basic questions remain unanswered, or, at best, answered ambiguously. Progress on fundamental questions, such as the dynamic origin of plates in the Earth, the hemispheric dichotomy and Tharsis bulge on Mars, layering, stratification and geochemical reservoirs, the thermal history of the Earth, the interpretation of seismic tomography, the source of volcanic hotspots, and the connections with long-term climate change, all require an interdisciplinary approach. To help answer these questions, numerical models of mantle convection must assimilate new information from a wide range of disciplines, including seismology, geochemistry, mineral and rock physics, geodesy, and tectonics.

The technical challenges here are substantial. Mantle convection is characterized by strongly variable (i.e., stress, temperature, and pressure-dependent) viscosity. In the lithosphere there are critical processes such as fracture and shear zone deformation (strain localization) that are physically distinct from the viscous flow deeper in the mantle, and occur on fundamentally different (smaller) length scales. In addition, the mantle is chemically heterogeneous, is replete with silicate melts and volatiles, and has numerous pressure- and temperature-induced structural (phase) changes that affect its dynamics.

Equally important, we have learned that it is not practical to model mantle convection using a wholly theoretical or "first-principles" approach, even if we had the hardware resources at our disposal. Two decades of experience has taught us that "first-principle" mantle convection models cannot replicate the unique history of the Earth, starting from arbitrary initial conditions. Accordingly, mantle convection models need to be more closely integrated into the "real world" through direct data assimilation and by direct testing against a variety of observations. Data assimilation and model comparison in this field is not easy. Although some of the data is quantitative (such as geoid, heat flow, present-day and past plate motions), other equally useful data are only partly quantified (such as structural, chemical, and rheological data, and observations drawn from the geologic record), requiring a sophisticated interface for use in convection models. Development of such interfaces is only beginning.

II. Crustal Dynamics

Crustal dynamics involves physical and chemical processes at scales ranging from a single fault to the entire crust. Major research topics include the physics of earthquakes, mechanics of individual faults and the interaction among fault systems, seismic cycles, damage mechanics, magmatic processes and physical volcanology, postglacial and postseismic rebound, and mountain building. Crustal dynamics is especially challenging by virtue of the wide range in spatial and temporal scales, and also because it involves the interaction between fluid mechanical and solid mechanical parts. Most of the phenomena listed above involve overlapping physical processes, often operating on distinctly different time and length scales, and subject to radically different constitutive laws.

In terms of societal importance, the most challenging, aspect of crustal dynamics is the general problem of earthquake physics. Modeling fault dynamics promises new breakthroughs in earthquake hazard mitigation by uniting various earthquake observations (data assimilation) into comprehensive treatments at the regional (fault zone) level. This integration must cross time scales of seconds (fault rupture) to years (viscoelastic relaxation) to geologic time (development of fault systems), and spatial scales of hundreds of kilometers (plate boundary faults) to centimeters (fault gouge zones). This grand challenge will require sophisticated treatments of fault zone rheology, dynamic mesh regridding, and the harnessing of massive computational resources in order to realize the full potential for addressing earthquake prediction and mitigation issues. Embedding such models in the general framework for whole-Earth dynamics (mantle convection, plate tectonics) will allow for the development of "fundamental" earthquake models, in which the underlying forces on faults are accounted for in a natural and self-consistent manner.

Computational models will contribute fundamentally to the intellectual foundations in areas related to crustal dynamics. Computational models for melt generation, transport, and eruption are just now reaching the level of sophistication necessary to shed light on long-standing petrological, volcanological, and geochemical problems that have defied simple theoretical treatment. In order to delineate the relationship between solid-Earth dynamics and climate change, far more realistic models of the Earth's response to glacial loading cycles are needed, to include complicated effects such as the contrasts in rheological properties beneath continents and oceans. Treating such problems within a single modeling framework can remove the artificial separation of crustal and deep-Earth studies that exists today in geodynamics.

III. Core Dynamos and Giant Planets

The past decade has witnessed a major breakthrough in understanding Earth's core -- the development of the first successful 3D numerical dynamo models. At least in this subject, geodynamicists have been fortunate. Numerical dynamo models exhibit the most basic features of the geomagnetic field, such as geocentric axial dipolar average magnetic fields, magnetic secular variation, and occasional polarity reversals, despite the fact that the models are very far from realistic in terms of their physical properties. We find ourselves in the enviable position of knowing *a priori* that numerical models are capable of still better representation of the geodynamo, and, furthermore, that they hold the prospect of understanding the wide variety of other dynamos found throughout the solar system. Even so, there is much room for progress. For example, we do not understand how dynamos work in the environment of the Earth's rapid rotation. We do not understand the effect of turbulence in the core, and perhaps most surprisingly, we do not know how dynamos work in highly conductive liquid metals. Furthermore, although we have a few particular numerical examples to guide us, we do not yet understand the underlying cause of polarity reversals, the controls on magnetic field morphology of the other planets, and how mantle structure and history influences the geodynamo.

Planetary dynamos are common phenomena throughout our solar system. Among the terrestrial planets, Earth and Mercury have active dynamos, and Mars and the Moon likely had dynamos in their past. Only Venus seems to be without one, for reasons that are still not understood. Dynamo action is a key indicator of planetary evolution, and a full understanding of why some terrestrial planets have active dynamos while others do not is among the most perplexing questions in planetary science.

The internal dynamics of the giant planets, both within and beyond our own solar system, is also an area of very active research. Recent spacecraft missions have provided a wealth of new observations of the four gas planets in our solar system that beg for dynamical explanations. In addition, astronomers have now detected more than sixty extra-solar giant planets, and this number will undoubtedly continue to grow. Interior dynamics of giant planets is a field closely allied with the study of the cores of the terrestrial planets, because the effects of planetary rotation and magnetic fields are dominant in each environment.

Many of the first order questions about interior dynamics in large gas planets remain unanswered. Examples of very basic phenomena that are poorly understood include the origin of zonal winds and oval vortices on Jupiter and Saturn and their connection to convection in the interior, the pressure and temperature of metallic phase transitions in hydrogen and their control on convection and magnetic field generation, the large misalignment between rotation and dipole field axes on Uranus and Neptune, and MHD induction effects in the interiors of the Galilean satellites.

The codes now being developed and applied to the geodynamo can be adapted to study convection and dynamo action in the other planets, including the gas giants. There are several effects that need to be incorporated into terrestrial planet dynamo models for application to gas planets, such as large compressibility, phase transformations, pressure- and temperature-dependent conductivity, and compositional differentiation. But there is the prospect that understanding the variety dynamos in our solar system and beyond can serve as the foundation for a truly physically-based comparative planetology.

IV. Computational Global Seismology

Global seismology is a relatively mature discipline of the Earth sciences. From a seismological perspective, the Earth is spherically symmetric to a good first-order approximation. Based upon 1-D Earth models seismologists now predict the arrival times of seismic waves, which travel for several tens of minutes from an earthquake to a station, to within several seconds. Over the past twenty years, seismologists have successfully used asymptotic methods, such as ray theory, to construct spectacular images of the Earth's interior, from the global scale down to detailed images of subduction zones.

Because of the success of simple spherically symmetric Earth models in conjunction with perturbation theory, global seismologists have been reluctant to embrace fully numerical approaches. At this time however, there is a pressing need for more accurate forward modeling techniques. We know that lateral variations in shear-wave speed exceed the five percent level in the upper- and lowermost mantle. Ultra-low velocity zones at the core-mantle boundary have been reported, lateral variations in short-period surface-wave phase speed can be in excess of ten percent, and anisotropy is invoked to explain shear-wave splitting and the directional dependence of oceanic Rayleigh-wave and continental Pn propagation. The crust exhibits strong lateral variations and varies in thickness by an order of magnitude from about seven kilometers underneath the oceans to more than seventy kilometers underneath Tibet. Clearly, linear perturbation theory is inappropriate in this context, and yet all `crustal corrections' rely on the validity of this approach. Seismologists are barely able to simulate wave propagation in general anisotropic models, and inversions for anisotropic Earth structure are in their infancy.

The challenge in global seismology is to calculate highly accurate, fully 3-D synthetics. One of the difficulties lies in the boundary conditions: one needs to resolve a slow thin crust with highly variable thickness. Another challenge is posed by the fact that dispersion and anisotropy are important in global seismology, and thus the mesh or grid should not introduce significant numerical dispersion or anisotropy. Attenuation also plays an important role in global seismology, which requires the introduction of memory into the system and is numerically expensive. There are sharp fluid-solid discontinuities at the inner-core and core-mantle boundaries and the ocean floor which need to be accommodated. Finally, at long periods, the effects of self-gravitation and rotation become relevant.

On a PC cluster computer one can now simulate fully 3-D global seismic wave propagation at periods greater than 18 seconds. On the Earth Simulator in Japan, today's fastest computer, it is possible to model 3.5 second waves. To put these capabilities into perspective, a typical normal-mode summation code that calculates semi-analytical synthetic seismograms for 1-D Earth models is accurate to 6 seconds. So on the Earth Simulator we are able to simulate global seismic wave propagation in fully 3-D Earth models at periods shorter than current seismological practice for 1-D spherically symmetric models!

Today, the Global Seismographic Network contains more than 120 stations, the first permanent ocean-bottom seismometer has been installed, and in the near future US Array and the Advanced National Seismic System may blanket the contiguous United States with hundreds of broadband instruments. These advances in instrumentation must be matched by similar advances in theory and computation. Now is the time for seismologists to join efforts by making 1-D and 3-D codes widely available through the proposed geodynamics framework.

V. Common Research Themes

We have identified a number of themes that cross-cut the individual disciplines in our coalition. Most of our problems span huge ranges of scales in space and in time. Most include time-dependent or nonlinear governing equations for the underlying physical processes. Most involve heterogeneous data types. Some of these data are quantitative and vast (EarthScope, InSAR, etc), others are semiquantitative and sparse (paleomagnetism, paleoseismology, etc), while still others are qualitative (geology). All of our disciplines need spherical representations for models, model results, and data, and all need better tools to interface data and models. All models require compatible and interchangeable software tools, or modules, for grid generation and dynamic re-gridding, a rich menu of numerical solution methods (finite elements, spectral elements, finite difference, finite volume, etc.), efficient solvers, domain decomposition and parallelization techniques, graphical interfaces, and sophisticated graphical options for effective visualization of enormous quantities of model output and data. Modern software engineering is now so advanced that all these needs can be met within a single community model framework. The benefit extends far beyond the practical goals of model development, though, since such a framework would have great intellectual impact in unifying all aspect of global geodynamics and planetary dynamics.

Computational Infrastructure for Geodynamics - Why now?

Rapid advances in computer power are driving the individual geodynamics code developers out of business. "Personalized" codes, products of a programming style established in the early days of research computing, have many shortcomings in today's environment. They are subject to single-point-of-failure breakdowns, are poorly documented and validated, and lack user-friendly interfaces, graphics, and data management tools. The effect of these shortcomings is to make the codes difficult to use, and almost impossible to maintain and improve outside of the small research groups where they originated. As investigators change their research focus, these personalized codes are often not maintained, and typically fall into disuse, without a suitable replacement to fill their role. The net result is either a loss of capability for the discipline (if they are not superceded by another code), or alternatively, a huge waste of effort, as another code writer "re-invents" what has been done before. A further problem with personalized codes is that the frontier for scientifically-meaningful calculations in geodynamics has been pushed so far forward that very few individuals have the time or the resources to write a code today that will be considered state-of-the-art in a few years' time.

These difficulties are most clearly seen in the plight of graduate students using and developing computational tools for geodynamics modeling. It has become almost impossible, within the normal duration of graduate study, to develop, test, and document a code, and then to apply the code in solving important geophysical problems. This has spawned a very discouraging atmosphere for graduate study in geodynamics -- the burden of modern code development limits the more rewarding activities of scientific exploration that motivate students in Earth and planetary sciences. Yet much of NSF and NASA funded research in geodynamics is still held together by graduate student "heroes" who struggle to develop computational tools in an increasingly complex and rapidly evolving software and hardware environment. This is no longer a viable mode of operation for research in geodynamics.

The consensus of Lake Tahoe workshop is that our community is ready to change its mode of operation by organizing its efforts into a computational center. We anticipate that a host of benefits will derive from this organization. Indeed, we feel the global computational geodynamics framework we intend to develop will change the entire culture of education and research in geophysics. Community supported modeling tools will quickly come into standard use in graduate and undergraduate education, linking researchers in academia, industry, and government laboratories, and communicating the excitement of our science to the general public. The resulting "democratization" of geodynamic modeling will serve large and small institutions alike, and will go a long way toward achieving true integration of education and research in the geosciences. In addition, the expected advances in software will inevitably lead to a more process-oriented approach within each discipline, where models are used to assimilate large data sets from seismology, geochemistry, mineral physics, geodesy, geomagnetism, and tectonics.

Needs, Results, and Risks

The workshop identified important needs of our proposed center and desirable results of its activities, as well as potential risks attending its formation.

Needs

The first need is for an organizational structure that naturally and efficiently brings together the talents of scientists and software professionals. During the Lake Tahoe meeting a number of organizational prototypes were considered, such NCAR (National Center for Atmospheric Research) and IRIS (Incorporated Research Institutions for Seismology). There was a clear consensus at the meeting that an IRIS-type model was preferable, with key support personnel and facilities centrally located, but with an effectively *virtual institute* of participating institutions and investigators steering and developing the science.

A central software design/engineering group should maintain the modeling framework, monitor and control source code development, facilitate code validation, maintain effective code documentation, develop graphical user interfaces and graphics software. There should be a centrally managed hardware facility, although other hardware (e.g., Beowulf class clusters) may be maintained at participating institutions. The central site should also maintain user/visitor facilities.

Support for broad collaboration, including international collaboration, is essential to the success of the Center. This support should include travel, communications, and hardware support for investigators in participating institutions comprising the scientific working groups (see more detail below). It is envisioned that the working groups will be the primary developers of scientific applications codes. However, the bulk of the Center's resources will likely be devoted to supporting the software development team and (to a lesser degree) the hardware requirements of the whole enterprise.

Desirable Results

A key development in recent years is the advent of modeling frameworks (see below), which are effectively tool kits for scientific computing, generally developed around modern parallel computing. Perhaps the most important material goal of the Center will be the development, or adaptation, of a modeling framework suitable for modeling all the classes of problems in computational geodynamics described above. Although we may never achieve a true “point-and-click” code framework, a primary goal is to reduce code development time from months and years to days or weeks. Such a goal has been demonstrated to be realistic in one recent framework developed for geophysical applications, the SNARK Project in Australia.

Additional properties of the modeling framework should include transparent open-source software, common data and i/o formats, consistent modular programming structures, self-documentation and validation methodology, and ease of community use via graphical user interfaces, where “community” refers ideally to all Earth and planetary scientists, including graduate and undergraduate students.

Risks

The Center should strive to implement software standards that minimize the risks inherent in the development of major modeling codes. Some of the risks identified at the workshop are: inadequate communication between computational professionals and scientists, undetected inconsistencies in requirements and designs, insufficient testing, brittle architectures, overwhelming complexity, *ad hoc* code management, uncontrolled change propagation, insufficient automation, and failure to recognize or acknowledge risks.

Computational Frameworks

Computational frameworks are intended to standardize code development in such fashion as to minimize the risks listed above. A successful framework adheres to a standard format for data and to official language standards, uses modules to achieve generality, allows for extension to meet future applications and encourages users to contribute components in an open source environment. Computational frameworks are now in widespread use in Atmospheric Science (particularly in climate models), and are coming into use in the Ocean Sciences. They are not yet widely used in the Solid Earth Sciences, although there are several pilot projects underway that were described by workshop participants.

There are many advantages to adopting a framework approach. Frameworks permit exchange of codes (interoperability). They promote the reuse of standardized software while preserving efficiency. They are an efficient way to deal with changes in computer architecture. They present the programmer and the scientists with a unified and well-defined task. They allow for shared costs of the housekeeping aspects of software development. They provide greater institutional continuity to model development than piecemeal approaches.

We heard presentations on several frameworks, including the NCAR CCM framework, the ESMF in development for a joint ocean-atmosphere model, and the SNARK framework in Australia. Although we were generally convinced that a framework approach is necessary, we did not endorse any one of these three options at the workshop.

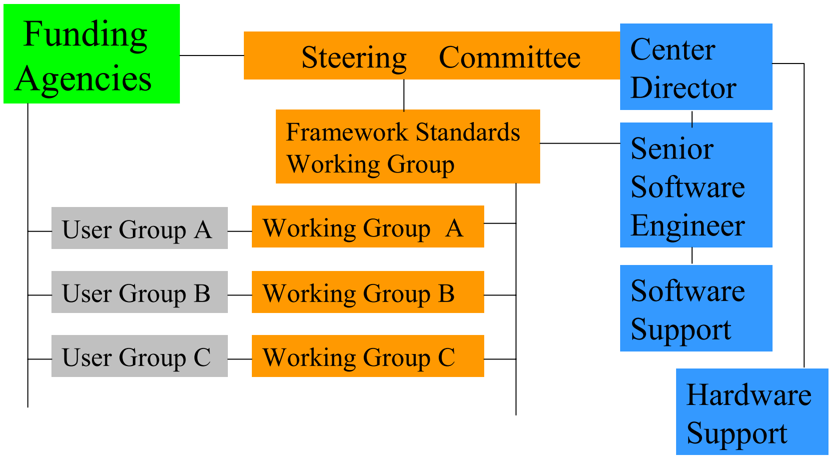
Organizational Structure

A potentially acceptable organizational chart is shown in Figure 1, reflecting a combination of centralized and distributed activities. The Center Director (most likely a university geophysics professor), a central software design and hardware support team (perhaps initially 3-5 personnel, including a senior software design engineer), and dedicated cluster-type parallel computer and graphics facilities would be located at or near a university with a distinguished geophysics research program. Additional software support personnel might also be flexibly located at other participating institutions in response to code development needs. These salaried personnel would provide expertise in numerical modeling, parallel computation, modeling frameworks, graphical interfaces, data/output management and representation, and code validation/documentation. Support for both short-term and long-term visitors to the Center is also needed, and these visitors may include students, post-docs, university faculty, senior scientists, and public outreach specialists. (There is a strong sentiment, however, that the central software development team should not employ post-docs, as the team's tasks are of a largely sustained-support nature.)

A Scientific Steering Committee would consist of perhaps 8-10 Earth/planetary scientists and computational scientists/software engineers, drawn from various academic institutions, scientific laboratories, and, perhaps, industry. The Steering Committee would maintain primary responsibility for achieving the stated goals of the Center in a timely fashion, liaison with funding agencies, and selecting and coordinating the activities of a number of scientific working groups developing application codes within the overall modeling framework.

The *Framework Standards* working group, consisting of both scientists and software engineers, would be key to the success of the entire enterprise. A number of scientific working groups would provide great flexibility in addressing various scientific problems, and could be constructed with differing degrees of focus (core dynamo models, plate rheologies in mantle convection simulations, etc.). This model of distributed working groups cemented by common software standards, engineering, and support has worked well in the climate modeling community as they have developed the new Community Climate Systems Model (CCSM) within NASA's new Earth Systems Modeling Framework (ESMF). There is every expectation that the working groups will involve international collaborators, as indicated by both the global nature of the science and by foreign participation at the Lake Tahoe meeting.

Proposed Organizational Structure



Funding

We expect to make an initial proposal to NSF to start the Center, but anticipate that additional support may flow from NASA, especially if the ESMF is adopted as the modeling framework. A support level of order \$1.5-2 million per year is needed, with perhaps 50% of the funding devoted to professional software engineering support, 30% devoted to hardware support and upgrades, and the remaining 20% devoted to supporting scientific collaborations.

Proposed Center Locations

Several proposed locations for the center were discussed at the Workshop. The following locations received endorsements: Boulder CO (possibly in conjunction with NCAR); Northern CA (involving UC campuses at Davis, Berkeley, and Santa Cruz); Southern CA (possibly at UCLA and involving but not at Cal Tech); Purdue University (West Lafeyette IN). We did not make a decision on this matter at the workshop.

Proposal Writing Team

* Mark Richards (UC Berkeley)/Peter Olson (Johns-Hopkins) [coordination]
* Jeroen Tromp (Caltech) [seismology]
* Jeremy Bloxham (Harvard([core dynamics]
* Mike Gurnis (Caltech)/Louise Kellogg (UC Davis) [mantle dynamics]
* Gerald Schubert (UCLA) [planetary dynamics]
* Marc Spiegelman (Lamont/Columbia) [melts/chemistry]
* Tom Clune (NASA Goddard) [software engineering/frameworks]

Workshop Participants

Peter Olson, convenor (Johns-Hopkins)
Mark Richards, convenor (UC Berkeley)
Bill Appelbe (VPAC, Australia)
John Baumgardner (Los Alamos)
Bruce Buffett (Univ. of British Columbia)
Lawrence Buja (NCAR, Boulder)
Hans-Peter Bunge (Princeton)
Tom Clune (NASA Goddard)
Chris Ding (Lawrence Berkeley Lab.)
Adam Dziewonski (Harvard)
Gary Glatzmaier (UC Santa Cruz)
David Gubbins (Leeds, UK)
Mike Gurnis (Caltech)
Garrett Ito (Univ. Hawaii)
Louise Kellogg (UC Davis)
Scott King (Purdue)
Masaru Kono (Okayama, Japan)
Weijia Kuang (NASA Goddard)
Adrian Lenardic (Rice Univ.)
Philippe Lognonne (IPGP, France)
Hiroaki Matsui (RIST, Japan)
Louis Moresi (Monash Univ., Australia)
Robin Reichlin (NSF, Geophysics)
Barbara Romanowicz (UC Berkeley)
Gerald Schubert (UCLA)
Slava Solomatov (New Mexico State)
Marc Spiegelman (LDEO, Columbia)
David Stegman (UC Berkeley)
Ikura Sumita (Kanazawa Japan)
Paul Tackley (UCLA)
Jeroen Tromp (Caltech)
Peter van Keken (Univ. Michigan)
Woo-Sun Yang (Lawrence Berkeley Lab.)
John Vidale (UCLA)
David Yuen (Univ. Minnesota)
Shijie Zhong (Univ. Colorado)

Questions asked of participants prior to the workshop

1. What are the most important objectives of a computational geodynamics center?

2. What problems in geodynamics should the center focus on?

3. What balance between software development and hardware facilities do you think is appropriate for the center?

4. Do you favor a centralized or geographically-distributed (virtual) center? What location(s) do you suggest for the center?

5. What sort of governing structure should we adopt to oversee the center?

6. What areas of technical expertise are essential to have at the center?

7. What numerical methods do you foresee as being the most important for modeling in your area of specialization?