Computational Infrastructure for Geodynamics (CIG)

NSF Award Number EAR-0426271

Five-year Strategic Plan: Sept. 1, 2005 to Aug. 31, 2010

Submitted July 26, 2005

Table of Contents

1. An Overview of CIG

The Computational Infrastructure for Geodynamics (CIG) develops, supports, and disseminates community-accessible software for the geoscience community from model developers to end-users. The software is being developed for problems ranging widely from mantle and core dynamics, crustal and earthquake dynamics, magma migration, seismology, and related topics. With a high level of community participation, CIG leverages the state-of-the-art in scientific computing into a suite of open-source tools and codes. The infrastructure now under development consists of:

- a coordinated effort to develop reusable, well documented and open-source geodynamics software;
- the basic building blocks – an infrastructure layer – of software by which state-of-the-art modeling codes can be quickly assembled;
- extension of existing software frameworks to interlink multiple codes and data through a superstructure layer;
- strategic partnerships with the larger world of computational science and geoinformatics;
- specialized training and workshops for both the geodynamics and larger Earth science communities.

CIG is building a small core team[1] of dedicated software architects and engineers whose work is guided by scientific objectives (see Section 3) formulated by the scientific community. The Software Development Team (SDT) provides software services to the community in terms of programming, documentation, training, and support. Guidance for the programmers comes from a Science Steering Committee (SSC) and an Executive Committee (EC) whose emphasis is to identify and balance common needs across disciplines. On Sept 1st, or shortly thereafter, the SDT will relocate to the Center for Innovative Technologies (CIT[2]), a new facility in Pasadena.

Since the official start date of the project, Sept. 1, 2004, the CIG staff, committee members, and members of the community have been diligently putting the project into place. In late 2004, the Member Representatives elected the first Executive Committee (EC) consisting of:

- Mark A. Richards, (Chairman, Oct., 2007), University of California, Berkeley

[1] Matt Knepley (0.5FTE at Argonne National Laboratory) an expert on Solvers with a PhD in computer science; since May 2nd, Leif Strand a skilled and experienced programmer with BS in computer science; since June 27th, Walter Landry an experienced computational scientist with a PhD in astronomy; starting Sept. 1st, Luis Armendariz, an entry level software engineer with a BS in computational mathematics. We have budgeted for one additional software engineer starting Sept. 1, 2005.

- Marc Spiegelman (Vice Chairman , Oct., 2006), Columbia University
- Bill Appelbe (At-Large, Oct., 2007), Victorian Partnership for Advanced Computing, Melbourne, Australia
- Carl Gable (At-Large, Oct., 2005), Los Alamos National Laboratory

The EC are joined by the Director (M. Gurnis) and Chief Software Architect (M. Aivazis) in an *ex-officio* capacity.

In May and June, 2005 the Member Representatives voted in the first Science Steering Committee made up of the following individuals:

- Brad Aagaard (June 2008), USGS [Earthquake dynamics]
- Wolfgang Bangerth (June 2008), Texas A&M [Computational science]
- Roger Buck (June 2006) Lamont-Doherty Earth Observatory of Columbia University [Crustal dynamics]
- Omar Ghattas (June 2007), Carnegie Mellon University [Computational science]
- Peter Olson (June 2007), John Hopkins [Geodynamo]
- Marc Parmentier (June 2006), Brown University [Magma dynamics]
- Jeroen Tromp (June 2007), Caltech [Computational seismology]
- Shijie Zhong (June 2008), University of Colorado at Boulder [Mantle convection]

This summer, CIG has been sponsoring or co-sponsoring four workshops in which the organizers and participants have been evaluating the current state of software in their respective fields and preparing preliminary plans for CIG software development over the short and long term. The four workshops include CIG/IRIS Workshop on Computational Seismology in Stevenson, WA (June 8, 2005), Geodynamic Modeling of Tectonics Processes in Breckenridge, CO (June 10-12, 2005), the CIG Mantle Convection Workshop in Boulder, CO (June 19 to 24), and the SCEC/CIG Workshop on Finite Element Models for Fault Systems and Tectonic Studies in Los Alamos (July 11-14). All four groups have already responded with posting reports to CIG on the web site[2].

The SSC reviewed these reports and discussed software priorities for the Project Year that starts September 1, 2005. The Director then distilled the reports, taking into account SSC comments, into a set of tasks that was then passed to the EC. Because of the shortness in time between the election of the SSC, the workshop dates, and the need to negotiate the Strategic Plan with the National Science Foundation, the Director and members of the Executive Committee worked more closely on this Strategic Plan than we expect in the steady-state operation of the CIG committee structure.

2. An overview of the strategic plan

[2] Shortcuts to all four of the reports can be found from the CIG home page, http:geodynamics.org

CIG strategy is centered on four interrelated goals: Development of a software repository and web site, development of a software framework, facilitating the use and access to the TeraGrid, organization and facilitation of community participation, and the training of software users.

2.1 Repository & Web site. The software repository and attendant web site are central to CIG's objectives of facilitating collaboration and sharing of validated open-source models and reuseable components. Moreover, the repository is critical to bring modern software engineering practices to our community and CIG's software development team. Our goal is to have a single repository[3] for developer use that manages multiple developers working potentially concurrently on modular software components shared through the repository.

Two key issues need to be addressed at the outset to ensure that the software repository is efficacious:

- *Installation ease* – downloading and *building* of CIG software by end-users
- *Robustness* of CIG software (freedom from bugs and glitches)

Installation ease: We pay careful attention to the means CIG software is downloaded and installed on user computers (the build procedure). This is a critical goal because nothing annoys and distracts users more than software that is difficult to download and install correctly. Software that is dependent on multiple levels of component libraries is particularly susceptible to build problems. In year one, a critical goal for CIG has been getting an effective "build" procedure working. Now that CIG has several members of the SDT in place, we expect that much of this will be completed by Sept 1st. The build procedure will evolve as a key principle in all CIG software development.

Robustness: A key problem that faces any dynamic software repository is ensuring that "nothing breaks" despite frequent dynamic changes needed to meet the evolving scientific goals of the community. CIG uses state-of the art software engineering technology – agile computing[4] to minimize the risks of software development for continuously evolving requirements. In particular, the repository uses unit and regression testing. *Unit testing* requires developers to construct test cases for each new piece of functionality as the actual code is written. The result is a suite of unit tests that can repeatedly validate a large fraction of the software in the repository. Further, as software defects are detected and repaired, CIG developers will construct further unit tests that verify the absence of the flaw and integrate these new unit tests into the test suite. As the software repository grows, there is a need for *regression testing* – automating rerunning the entire test suite against the entire software contents in the repository on all supported platforms (operating systems and processors that the repository has been ported to and is maintained on) the current code base on a frequent (daily) basis against the test suite. Nightly regression testing generates an electronic report that contains the build and test

---

[3] We have been using the CACR (Center for Advanced Computer Research at Caltech) CVS (Concurrent Version System) repository for CIG. We have selected Subversion (SVN) as the CIG repository software and by Aug. 15, 2005 all CIG software will be available from svn.geodynamics.org.

[4] http://www.agilealliance.org

failures (including and the platforms on which they occurred). Regression testing allows the SDT to rapidly identify when a change in a repository component or platform has caused an error or inconsistency. Regression testing allows users of the repository to have confidence in the robustness of the repository. Regression test has already been in use by CIG participants for many years. Regression testing will be in place for some of CIG's software by Sept 1st and will then be updated based on current practices in subsequent years.

The web site will be used for both software downloads[5] as well as a means by which codes, including CIG software, can be benchmarked and compared to existing benchmark results (the later point is needed for the community to participate in our system of software verification and validation). This is essential for increasing the overall quality of our science especially as it moves into the realm of complex, multi-physics and multi-scale simulations. Our intermediate term goal is for CIG to develop a standard procedure by which geophysical codes can be compared. This will mean the establishment of benchmarks that are sensible mathematically (verification) and geophysically (validation). A common and permanent repository will be created for benchmark results. CIG will collaborate with individual communities in order to develop this procedure: CIG staff will adapt and refine this procedure as it gains experience. On our web site, we have posted the benchmarks used by the short time-scale tectonics community as well as specification of the benchmarks for the verification of spherical mantle convection codes. We will soon start working with the computational seismology community and proceed on to magma-migration, the geodynamo, and long term crustal dynamics, not necessarily in this order, as described in Section 4.

2.2 The software framework. The development of a software framework for geophysical modeling and its upkeep are central to CIG's strategy. The developing framework consists of software on two levels:

- An *infrastructure layer* of software by which community and state-of-the-art modeling codes will be assembled

- A *superstructure layer* of software by which multiple codes and data can be interlinked through the extension of existing software frameworks

CIG's resources are limited and entirely new software will only be generated when needed. CIG will rely on lower level components developed in more science-neutral arenas, including some DoE supported projects, and higher-level frameworks to interlink software as much as possible. The workshops and the reports described above were used in a preliminary design of components for the software framework. Many of the community's short term goals for CIG involve migrating existing software into our repository, developing suites of test examples & user manuals for this software, and then reengineering relevant pieces of this software within the context of a superstructure

[5] As of this date, CIG software available on the web includes CitcomS.py Version 2.0 (a finite element code to solve thermal convection problems relevant to earth's mantle) and PyLith (a regineered version of the popular Tecton finite element program for problems in crustal dynamics).

framework. Most long-term goals involve writing new software with common components. We have been able to identify software development that both satisfy short-term goals while laying the groundwork for a system of software with common components. Here, we briefly provide an overview of such components and development tasks with details in Section 4c, below.

Computational seismology: A central goal for computational seismology is the establishment of automated and on-demand simulations (e.g. seismic wave propagation and synthetic seismograms) through a seismology *science portal*. A science portal is a web site that launches a simulation on a remote machine using data gathered from web sites and databases and then returns the results to the user. There are several components of portals, some of which are beyond the CIG scope, such as seismic databases. In order for a science portal to operate effectively, there is the need to select between multiple simulation methods and/or codes with uniform specifications. Consequently, the short-term goal of this community is migrating their major methods for simulating seismic waves into an accessible repository and then reengineering such codes to work with a superstructure framework, in this case Pyre, the science neutral Python framework. This will allow the seismic codes to be initiated in a standard fashion. We have not identified low-level components from seismic codes as reusable within an infrastructure framework, but continued development of the superstructure and control of simulations in the framework via the web is a generic CIG need. Moreover, we have identified the movement of data and results between geodynamic simulations and seismic wave propagation as a community need for the EarthScope and CSEDI program areas, for example. Science portals will be developed as collaborations with other organizations and CIG is nurturing such relationships with IRIS, GEON, and CACR.

Short time-scale tectonics: For deformation over short time-scales, several long-term goals have been identified, especially enhancement of software for the simulation of multiple earthquake cycles with sufficient resolution to capture the buildup of strain in the crust, strain release in propagating ruptures that radiate seismic waves, and postseismic relaxation of the crust. Infrastructure is needed to couple crustal dynamics software to other models of Earth processes, as well as allow data assimilation into crustal dynamics software. The availability of such software would promote integration of EarthScope data into crustal dynamics modeling. We have identified tasks that allow us to build common components from two widely used finite element codes, EqSim and a reengineered version of Tecton. Two members of the community, Brad Aagaard (USGS) and Charles Williams (RPI) have been working with CIG to integrate their two codes into the Pyre framework. A significant amount of commonality was identified between the codes and since then Aagard and Williams have coordinated their development with a plan to merge their codes into a single suite of modules, PyLith. In addition, in year one CIG added Python bindings to the Portable, Extensible Toolkit for Scientific Computation (PETSc)[6]. This significantly reduces the amount of code required to write solvers in the superstructure Pyre framework using PETSc. Consequently, PyLith uses the PETSc solvers directly from the framework, an important step toward our vision of a framework with common components.

[6] See http://www-unix.mcs.anl.gov/petsc/petsc-2/

Mantle convection: For mantle convection, several long-term goals have been identified, especially development of compressible spherical convection codes, which is particularly important for a better integration of mineral physics and seismology with deep-earth dynamics. The community also identified development of methods to couple small-scale physics, especially lithospheric deformation and melting/melt migration, into large-scale mantle flow models, as well as modular tools for grid refinement, element types, and solvers. In order to move in the direction of a compressible formulation of a code for spherical mantle convection, CIG will convene a small workshop in early 2006 to bring together software engineers and interested geophysicists. Preliminary task have been formulated to provide the software components needed for the specific recommendations that will emerge from the workshop. First, a new compressible solver will be written for CitcomS, the publicly available convection code for flow within a spherical geometry. A reengineered version in which a set of components is controlled by Pyre is already available from CIG, CitcomS.py Version 2.0. Second, we will collaborate with the Victorian Partnership for Advanced Computing (VPAC), a not-for-profit company in Melbourne Australia[7], on the expansion of their infrastructure framework, StGermain. StGermain is an infrastructure framework for building meshes, advecting particles, calling solvers, orchestrating I/O, and related functions; two geophysics codes have been developed with StGermain[8]. We will collaborate with VPAC on expanding the features within StGermain and then build a spherical mantle convection solver with these common components (as with the other geophysical codes). The use of StGermain would simultaneously solve the need for particle advection capabilities in convection codes. Moreover, this would allow different solvers to be easily swapped in and out of the code since the framework is built upon PETSc. StGermain codes can also be orchestrated with the Pyre superstructure framework allowing code coupling, a feature specified by nearly all of the CIG communities.

Long-time scale tectonics: The wide availability of a 3-D implementation of the ALE/PIC method[9] and implicit methods that include adaptive mesh refinement (AMR) are important long-term goals for lithospheric deformation/long-time scale tectonics. AMR is desirable for efficiency and is a necessity for solving large-scale problems with high spatial resolution in regions of strain localization and fluid/rock interactions. As a short-term task, CIG will work an existing 2-D ALE code, migrate the software into the repository, develop a suite of test examples, and develop a users manual. It is not clear if

---

[7] CIG and VPAC are currently working on a Memorandum of Understanding (MOU) that will define the nature of the collaboration. The MOU will be a non-binding legal agreement between Caltech and VPAC that will define the open source (public) ownership of the software, the joint maintenance of the software repository, and the requirements expected in contracts between the respective organizations.

[8] The geophysics codes were developed through collaboration of geophysics groups with VPAC engineers. One of these is SNARK, developed by the geophysics group at Monash University in Australia and is a hybrid between a PIC method and an implicit Finite Element Method. It is suitable for long-time scale crustal dynamics and mantle convection. The second of these is SNAC developed by the GeoFramework group at Caltech. This code uses the FLAC method in a 3-D domain and is used for long term crustal dynamics with visco-elastoplastic rheologies.

[9] The ALE (Augmented Lagrangian-Eulerian) and PIC (particle-in-cell) methods both use a combination of Lagrangian particles and Eularian meshes.

there is any need to bring this code into the superstructure framework as it will be used as a stand alone code while giving the CIG engineers software and working examples to pattern a 2-D/3-D ALE code after. We will then collaborate with VPAC on the development of a 3-D ALE code using common components in the StGermain framework. Much of this functionality exists from the SNARK application. A benefit of this approach is that the ALE method would be able to use different equation solvers available through PETSc. For AMR, we see considerable value in implementing deal.II, the finite element differential equations analysis library, as components in the infrastructure framework. Wolfgang Bangerth (Texas A&M), the author of the deal.II library, and Luc Lavier (UTIG) have submitted a proposal to the NSF CMG program and would collaborate with CIG to make the products of their development available through our infrastructure frameworks.

Cross disciplinary projects: Through our choice of tools and tasks, we are attempting to develop software that simultaneously satisfies multiple needs in the geosciences. The overlap between the needs of mantle convection modeling and long-time scale tectonics, on the one hand, and between short- and long-time scale tectonics, on the other, are important examples where synergies exist. First, by building components in the StGermain framework in collaboration with VPAC, we will develop components that will be used for both long-time scale tectonics and mantle convection. For example, the ALE code would not just operate for a Cartesian geometry but for a spherical geometry as well and this would give us the ability to easily couple mantle convection with crustal deformation. Second, the use of implicit finite element methods with the PyLith software will be able to be used by both the short-time and long-time scale communities, although it is emerging from only one. In addition, the long-term goal of the tectonics community is AMR methods, but PyLith, being developed by the short-time scale group, will likely just as easily incorporate the AMR.

2.3 Facilitating the use and access to the TeraGrid. CIG doesn't maintain its own major computational hardware, but we are starting to develop a relationship with the NSF TeraGrid so as to provide a resource for both the community and CIG staff. Perhaps most critical is that CIG software is being developed so that it works well on any platform from desktop Linux & Mac OSX platforms to Beowulfs clusters and large supercomputers such as those available on the TeraGrid. Since many members of the community from smaller departments and universities do not have ready access to parallel computers, availability of CIG software on the TeraGrid provides an important resource for this part of the community. CIG currently has a moderate allocation from TeraGrid and the SSC will develop a strategy for a larger allocation this coming year. CIG needs for TG resources include: workshops, software testing, allocation to the geophysics community for production runs using CIG software, and, in collaboration with other groups, development of several science gateways (e.g. web portals) that utilize CIG developed and maintained codes.

We have already been able to utilize TG resources for workshops. For example, during the recent CIG Mantle Convection Workshop in Boulder, workshop participants were given accounts on the TeraGrid and they were able to run tutorial examples of the mantle convection code CitcomS.py that CIG maintains and distributes on the web. We will continue to develop expertise in this area and future improvements will include the follow:

- On our Web site, we will develop a set of pages so that members of the community can transparently get accounts to TG under CIG allocations. We will also provide instructions on accessing these resources.
- In our User Manuals, we will develop chapters and test examples that illustrate how to run the codes on the TG.
- For some of our software and for some TG platforms, we will develop the software so that it is pre-installed on the supercomputer for use by any user.

The allocations will give the CIG community and staff a concrete estimate of how our software performs and scales on TG resources. When CIG works on the build procedure, it is essential that we develop and exercise the software on as many platforms as possible and the TeraGrid provides us with a useful set of resources where we can accomplish this.

There are probably two significant impediments to the use of national supercomputer facilities by many segments of the scientific community. First, is the availability of software that effectively uses massively parallel supercomputers. Clearly, the fact that CIG software is designed for such purposes and is preinstalled on such platforms should largely overcome these problems for the greater geodynamics community. Second, is the need to write a proposal that documents the level of supercomputer resources needed and that the software effectively uses and scales on the platform in question. CIG will provide the necessary results showing the performance and scaling of our software on shared national resources. Such performance information will be a natural product of CIG software development. Consequently, we are likely to see increased usage of TeraGrid facilities by a broader range of geoscientists.

2.4 Organizing community participation. Centrally important for CIG, is the guidance from the scientific community on what this infrastructure should accomplish for their evolving research needs. This is accomplished through community oversight of CIG, committees, and workshops. Two key principles guide CIG's interaction with the community:

- *Openness*: All CIG reviews, meeting minutes, and other documents are openly available to the community in a timely fashion, for review and public discussion, unless this conflicts with individual or institutional privacy rights (such as salary level or personal information). Openness minimizes the risk of actual or perceived bias or conflicts of interest within CIG.
- *Interaction*: All CIG committees and workshops should have an open and balanced representation of both the scientific and software communities. Interaction minimizes the risk of balkanizing the community CIG serves.

Much of the organization of CIG for the community is now in place. The ByLaws were formally adopted by the Member Representatives and the Executive Committee and Science Steering Committees have likewise been put into place (see Section 1). The EC is now responding to requests given in the workshop reports to put into place several discipline-specific working groups. Workshops and committee meetings will continue to be the mechanism by which we layout a software system that delivers the core functionality of geodynamics in an open and extensible fashion. As described already, CIG has been sponsoring or co-sponsoring four workshops this summer with the principal goal of increasing community participation in CIG.

The Web site is an important tool for community participation. It is not just used as a means to distribute software and documentation, as described above, but it is also used for members to communicate. CIG currently uses a technology called Plone[10] that allows users to edit the web pages directly. For example, each of the workshop reports has a comment section at the end and registered users of the web site can add comments and content to both their own individual area as well as special areas for each working group. In addition, CIG maintains List-Servs for both the entire community (cig-all@geodynamics,org) , each committee, and each subject area. These list servs are cross-linked so that a user can easily navigate from the Plone editable areas to the List-servs.

2.5 User training. A key objective of CIG is the widespread adoption of CIG-developed software by the general Earth science community and, in particular, by geophysicists. Both significant training and comprehensive documentation will be required. Potential users will want to know how to use CIG software as well as gain an understanding of the underlying algorithms and implementation details. Based on current research directions, sophisticated users who require a deeper understanding of the workings of the software will likely have different backgrounds and interests. One class of user consists of traditional computational scientists want to use several different computational components (such as a solver and a mesher) to create working codes that incorporate algorithmic innovations. Another class of user consists of scientists who attempt to string together several components of CIG software with non-CIG data analysis tools. In addition, there will be a class of less-technically-demanding users who will want to use CIG codes in a standard manner (e.g., for pedagogical purposes or using standard codes on new input data sets).

Several mechanisms will be used to train CIG users: (A) small, focused workshops; (B) visits to the CIG site to work with CIG staff and other users; and (C) the production of training manuals and a web site. CIG will use these mechanisms to provide training for both the expert and non-expert user.

We plan to have our first training workshop this year and then subsequently hold several training sessions per year. However, a technical writer was recruited in year one and started working on the development of user manuals and training material.

[10] Plone is a user-friendly and powerful open source Content Management System, see http://plone.org/.

3. Annual Program Plan

CIG does not carry out science directly, but provides the infrastructure to modeling specialists and other earth scientists who use modeling tools, thus facilitating the science funded elsewhere. CIG provides the infrastructure to allow several areas of computational geophysics to move forward and solve previously intractable problems. In this plan, we briefly describe the specific areas of science that we will likely facilitate in the coming year.

An expanding area of research is the study of crustal deformation on short-time scales in tectonically active areas opened by the wide (and growing) availability of dense, continuous GPS networks. The research is quickly expanding as the PBO component of EarthScope is deployed, as data flows from existing dense continuous geodetic networks in Southern California, Japan, & Tawain, and through earth deformation imaged through the InSAR method. Forefront research is focused on the nature of the strain and stress field before and after earthquakes, silent "earthquakes", the interaction between active faults in a plate boundary zone, and magmatic inflation and deflation. Computational models play a fundamental role in all of these studies. CIG will be working with a group of researchers to coordinate development and validation of 3D quasi-static, finite-element codes for modeling crustal deformation; develop deformation models with observed topography, fault geometries, rheological properties, geologic slip rates, geodetic motions, and earthquake histories; and use these models to infer fault slip, rheologic structure, and fault interactions through stress transfer.

CIG will directly impact this research by providing assistance to this community in the inter comparison of different modeling codes, essential for the verification and validation of the methods used in modeling. We will work toward the adoption of standard file formats (for all areas of computational geophysics) and this will allow this group to easily compare results. We will work closely to reengineer one of their community codes with parallel equation solvers. The adoption of standard I/O and parallel solvers will allow the community to more easily use the finite element codes and run progressively larger problems on Beowulf clusters and the TeraGrid. This year we would expect explicit scientific results, probably for problems tailored to Southern California earthquakes and deformation within Pacific-rim subduction zones, including Cascadia.

We will also be working closely with the mantle convection community. Current research in this area is motivated by global seismic tomographic inversions (facilitated by GNS), questions emerging from interdisciplinary research in the CSEDI program, questions emerging from interdisciplinary research in the MARGINS program (Especially the Subduction Factory initiative), and regional studies of the western U.S. now gearing up in anticipation of the USArray component of EarthScope. Some of the most challenging computational issues arise for convection in multicomponent convection, such as thermo-chemical convection. A related problem will be in the integration of melting and chemical transport into convection codes. There is substantial interest in coupling models of mantle and crustal dynamics. CIG will directly impact this field through the adoption of common file formats, establishing benchmarks, and then developing a common procedure to compare different methods and codes. In the last ten years, studies of mantle convection have dramatically expanded into more complex physical (rheological, multiphase, geometrical) systems and there is a need for the community to validate and verify results. The CIG staff will also be working with several mantle convection codes to use common components and this will allow more users to have access to validated mantle convection software by the end of the year starting Sept. 1, 2005. We expect that the software will be used to study lower mantle dynamics, including the nature of the complex region at the base of the mantle, superplumes, the layering of mantle convection, the fixity and/or drift of hot-spot plumes, the interaction of the continental lithosphere with mantle convection, the origin of mantle convection, and subduction zones, including the dynamics of the mantle wedge.

An important aspect of our adoption of standard interfaces for codes and common file formats is that we are actively planning to coordinate this effort with observational initiatives. Obviously, much of the standardization will happen internal to CIG (through the workshops in crustal deformation and mantle convection and through CIG software engineers). However, we plan to develop with EarthScope and/or IRIS a standard format for the storage of both tomography models (3-D) and geodynamic models (3-D plus time). This simple standardization should vastly expand interdisciplinary research. Some obvious benefits would be in mantle convection that is being actively investigated observationally by seismologists and theoretically by modelers. For studies of superplumes (a subject of substantial interest) one could rapidly get answers to questions such as: What is the flow field predicted by a particular seismic inversion? What seismic waveforms are predicted for a specific realization of a thermo-chemical superplume? Seismic tomographic inversions from the USArray deployment could be fed directly into a flow model for mantle flow and crustal deformation. Much of the infrastructure development for this will likely be completed in year one with scientific results emerging in year two.

4. Annual Goals and Milestones:

4.a All goals and milestones for Year 1 (Sept. 1, 2005 – Aug. 31, 2006)

I. Software Repository
- a. Maintain software available on web
  - i. CitcomS.py
  - ii. PyLith
- b. Adoption of version control, regression testing, and uniform build procedure
- c. Community accepted benchmarks
  - i. Mantle Convection Community
  - ii. Fault-interactions on earthquake time-scales.
  - iii. Seismic wave propagation

II. Computational Framework
- a. Specification of APIs for geophysical codes & adoption of standard formats for storing data (such as netCDF or HDF5)
- b. Frozen codes with common interfaces
  - i. 1-D seismic codes (3)
  - ii. SpecFEM3D
  - iii. Sopale (2-D ALE)
  - iv. 3-D Cartesian Citcom
  - v. STAG3D (?)
  - vi. Terra
  - vii. MC Post prcecessing
  - viii. Okada Dislocation
- c. New components at Infrastructure level
  - i. 3-D ALE (StGermain)
  - ii. New compressible solver for CitcomS.py
  - iii. Compressible spherical convection code (StGermain)
  - iv. Provide guidance for melt migration code in framework

III. Organizing Community Participation
- a. Joint meeting of EC and SSC, November, 2005, Pasadena, CA
- b. Small workshop on compressible mantle convection codes, Pasadena, CA.
- c. Finite element modeling of fault interactions workshop (Co-sponsorship with SCEC, summer 2006)
- d. Workshop on geodynamo codes (Summer, 2006 in Europe, find a partner).
- e. Co-sponsor the GeoMod pre-conference workshop on benchmarking crustal dynamics codes (summer, 2006, with Europeans).

IV. User Training
- a. User manuals

i. Maintain and expand manuals
- -CitcomS.py
- -PyLith

ii. Start new manuals
- -1-D Seismic codes
- -SpecFEM3D
- -Sopale
- -Cartesian 3-D Citcom
- -Short manuals for Okada code and STD MC post processing.

b. Start development of material for training sessions

4.b Software Repository Years 1-5

Year 1 Sept. 1, 2005 – Aug. 31, 2006	Year 2 Sept. 1, 2006 – Aug. 31, 2007	Year 3 Sept. 1, 2007 – Aug. 31, 2008	Year 4 Sept. 1, 2008 – Aug. 31, 2009	Year 5 Sept. 1, 2009 – Aug. 31, 2010
• Expand & Maintain donated software available on web -CitcomS.py 2.0 -PyLith 1.0 • Start distribution of new CIG generated Software • Maintenance and update of version control, regression testing, and uniform build procedure • Community accepted benchmarks: - Complete for Mantle Convection Community - Complete for fault-interactions group -Initiate for Seismic wave propagation	• Maintain distribution of donated software available on web • Maintenance and update of version control, regression testing, and uniform build procedure • Community accepted benchmarks: - Continue to work on seismic wave propagation - Initiate for geodynamo problem - Initiate for long term crustal dynamics	• Maintain distribution of existing and new software on web • Maintenance and update of version control, regression testing, and uniform build procedure • Community accepted benchmarks - Complete for geodynamo - Complete for long term crustal dynamics - Initiate for magma migration	• Maintain distribution of existing and new software on web • Maintenance and update of version control, regression testing, and uniform build procedure • Community accepted benchmarks - Complete for magma migration	• Maintain distribution of existing and new software on web • Maintenance and update of version control, regression testing, and uniform build procedure • Continued posting and revision of all community benchmarks

4.c Software Framework Years 1-5

Year 1 Sept. 1, 2005 – Aug. 31, 2006	Year 2 Sept. 1, 2006 – Aug. 31, 2007	Year 3 Sept. 1, 2007 – Aug. 31, 2008	Year 4 Sept. 1, 2008 – Aug. 31, 2009	Year 5 Sept. 1, 2009 – Aug. 31, 2010
• Reengineering: • Frozen codes with common interfaces - 1-D synthetic seismogram codes (3) - SpecFEM3D -Sopale (2-D ALE code) -3-D Cartesian Citcom -MC Post processing -STAG3D -Okada displocation • Infrastructure Layer: • Specification of APIs for geophysical codes & adoption of standard formats for storing data (such as netCDF or HDF5) • 3-D ALE with StGermain • New Compressible solver for CitcomS.py • Compressible MC Spherical with StGermain • PyLith	•Reengineering: • Frozen codes with common interfaces - geodynamo code to be donated - Additional 1-D synthetic seismogram codes (2) - Magma Migration code • Infrastructure Layer: • 3-D ALE with StGermain • Compressible MC Spherical with StGermain • Superstructure Layer: • Link existing solvers or codes with superstructure layers as proposed by members • Continue to generalize Pyre coupers and exchangers for geophysical code	• Infrastructure Layer: • Develop new solvers with libraries as proposed by members •Project Specific Layer •Develop post-processing of solver-linkage modules specific by members • Superstructure Layer: • Link new solvers with superstructure layers as proposed by members	• Infrastructure Layer: • Develop new solvers with libraries as proposed by members •Project Specific Layer •Develop post-processing of solver-linkage modules specific by members • Superstructure Layer: • Link new solvers with superstructure layers as proposed by members	• Infrastructure Layer: • Develop new solvers with libraries as proposed by members •Project Specific Layer •Develop post-processing of solver-linkage modules specific by members • Superstructure Layer: • Link new solvers with superstructure layers as proposed by members

4.d Organizing Community Participation Years1-5

Year 1 Sept. 1, 2005 – Aug. 31, 2006	Year 2 Sept. 1, 2006 – Aug. 31, 2007	Year 3 Sept. 1, 2007 – Aug. 31, 2008	Year 4 Sept. 1, 2008 – Aug. 31, 2009	Year 5 Sept. 1, 2009 – Aug. 31, 2010
• Joint meeting of EC and SSC, Pasadena, CA • Compressible mantle convection workshop in Pasadena, CA. • Finite element modeling of fault interactions workshop (Co-sponsorship with SCEC) • Long-term crustal dynamics workshop (co-sponsor with "GeoMod" European group) • Geodynamo community workshop (find international partner) • Future Directions of coupling codes workshop (co-sponsor with GeoFramework)	• 1 EC meeting • 1 Science Steering Committee • Sponsor or co-sponsor about 3 workshops as proposed by members	• 1 EC meeting • 1 Science Steering Committee • Sponsor or co-sponsor about 3 workshops as proposed by members	• 1 EC meeting • 1 Science Steering Committee • Sponsor or co-sponsor about 3 workshops as proposed by members	• 1 EC meeting • 1 Science Steering Committee • Sponsor or co-sponsor about 3 workshops as proposed by members

4.e User Training Years 1-5

Year 1 Sept. 1, 2005 – Aug. 31, 2006	Year 2 Sept. 1, 2006 – Aug. 31, 2007	Year 3 Sept. 1, 2007– Aug. 31, 2008	Year 4 Sept. 1, 2008 – Aug. 31, 2009	Year 5 Sept. 1, 2009 – Aug. 31, 2010
•Applying the Pyre Framework (co-sponsor with GeoFramework) • Training session on use of reengineered FE code for fault group (with SCEC) • Start user manuals - CIG software repository - Reengineered codes that use PETSc • Start development of material for training sessions	•Sponsor visitations of students, post-docs and faculty to CIG site • Training session on use of reengineered FE code for mantle convection (hold in conjunction with topical science conference) • Start user manuals - Finish for CIG software repository - Finish for reengineered codes that use PETSc • Material for training sessions - Finish for Reegineered codes with PETSc	•Develop training material and web site for new CIG software •Sponsor visitations of students, post-docs and faculty to CIG site •Hold specialized training session at CIG for newly developed software •Hold specialized training session at other site in conjunction with topical conference, preferable one that is sponsored by another group.	•Develop training material and web site for new CIG software •Sponsor visitations of students, post-docs and faculty to CIG site •Hold specialized training session at CIG for newly developed software •Hold specialized training session at other site in conjunction with topical conference, preferable one that is sponsored by another group.	•Develop training material and web site for new CIG software •Sponsor visitations of students, post-docs and faculty to CIG site •Hold specialized training session at CIG for newly developed software •Hold specialized training session at other site in conjunction with topical conference, preferable one that is sponsored by another group.

5. Allocation of resources by goal

Sept. 1, 2005 – Aug. 31, 2006

Software Engineer #1 (WL,Numerical Analyst/Modeler) [Total 1 FTE]

- I (repository): 0.1 FTE
- II.a (uniform data): 0.075 FTE
- II.b.iii (Sopale reengineering): 0.2 FTE
- II.c.i (3D ALE): 0.5 FTE
- III.b (Compressible Convec. Wkshp): 0.025 FTE
- III.e (GeoMod Wkshp): 0.05 FTE
- IV.a.ii (Sopale Manual): 0.05 FTE

Software Engineer #2 (LS, Software Integration) [Total 1 FTE]

- I (repository & benchmarks): 0.1 FTE
- II.a (uniform data): 0.1 FTE
- II.b (1-D seismic codes): 0.3 FTE
- II.c (SpecFEM3D): 0.4 FTE
- IV.a.ii (Seismic manuals): 0.1 FTE

Software Engineer #3 (LA, Software Integration) [Total 1 FTE]

- I (repository & benchmarks): 0.2 FTE
- II.a (uniform data): 0.1 FTE
- II.b.iv-vi (convection codes): 0.25 FTE
- II.c.ii (Compressible Solver CitcomS.py): 0.2 FTE
- II.c.iii (new Compressible Spherical MC): 0.3 FTE
- IV.a.ii (MC manuals): 0.05 FTE

Software Engineer #4 (Software Integration) [Total 1 FTE]

- I (repository & benchmarks): 0.1 FTE
- II.b.vii (MC Post process): 0.2 FTE
- II.b.viii (Okada): 0.2 FTE
- II.c.iv (Melt Migration): 0.4 FTE
- II.c (PyLith): 0.1 FTE
- IV.a.ii (Okada & MC manuals): 0.1 FTE

ANL Subcontract [Total 0.5 FTE]

- II.c.ii (Compressible Solver CitcomS): 0.1 FTE
- II.b (PyLith): 0.3 FTE
- II.b (melt Migration): 0.05 FTE
- III.a (Short term Wkshp): 0.05 FTE

VPAC Subcontract [Total 1 FTE]

II.c.i (3D ALE): 0.5 FTE
II.c.ii (Compressible Solver CitcomS): 0.5 FTE

Director [Total 0.12 FTE paid by CIG]
Center Management 0.12 FTE

Chief Software Architect [Total 0.12 FTE]
II (framework): 0.12 FTE

Secretary [Total 1 FTE]
III (community workshops): 0.3 FTE
Paperwork for Management: 0. 5 FTE
General Web Management: 0.2 FTE

Technical Writer/Web master [Total 1 FTE]
I.a (software on web): 0.05 FTE
I.c. (benchmarks on web): 0.05 FTE
IV.a (User Manuals): 0.7 FTE
General Web Management: 0.2FTE

Supplies and Expenses [$65K]
I, II, III, IV

Travel [$28K]
I, II, III, IV

Participant Costs [$91K]
III, IV

6. Membership

6.1 CIG Members and Member representatives:

Argonne National Laboratory (MSC)
California Institute of Technology
Colorado State University
Columbia University
Harvard University
Johns Hopkins University
Lawrence Livermore National Laboratory
Los Alamos National Laboratory (ES)
Massachusetts Institute of Technology
Oregon State University
Pennsylvania State University
Princeton University
Purdue University
Rensselaer Polytechnic Institute
State University of New York at Stony Brook
U.S. Geological Survey (Menlo Park)
University of California, Berkeley
University of California, Davis
University of California, Los Angeles
University of California, San Diego
University of Colorado
University of Maine
University of Maryland
University of Michigan
University of Minnesota
University of Oregon
University of Southern California
University of Texas at Austin
University of Washington
Washington University
Woods Hole Oceanographic Institution

6.2 CIG Foreign Affiliates and representatives:

Australian National University
Monash University
University of Sydney
Victorian Partnership for Advanced Computing

6.3 Strategy for keeping members informed

Member representatives and individuals within the larger CIG community (including those at member institutions) will be kept informed through several different means.

1. Through e-mail. CIG maintains several list servers through the CIG website including several for the main committees (e.g. Executive Committe, Science Steering Committee) as well as for working groups and general information (e.g. cig-all@geodynamics.org).

2. Through the http://Geodynamics.org web site. The upcoming CIG calendar of events is posted and continuously revised. Nearly all CIG documents, including proposals submitted to CIG, the annual revision of the CIG Strategic Plan, By-Laws, etc., are posted of this site. The Web site is the principal means for standard software downloads, sharing of community benchmark, specifications of standards, and distribution of user & training manuals.

3. The annual CIG Business meeting. This meeting will be open to all and will be a forum for open discussions of the working of CIG, including past and upcoming activities & the Strategic Plan. In year one, this meeting will be held in conjunction with the Fall AGU meeting in San Francisco. Depending on the success of this meeting, this may be the venue for subsequent years.

4. CIG sponsored and co-sponsored workshops and training sessions. The current status of CIG will be presented at these workshops and we expect that CIG members will attend such workshops.

7. Five Year Management Plan

CIG will need the expertise, vision, and guidance of the community if it is to remain a nimble and evolving organization. Consequently, we are adopting a *community-centric* management structure that draws upon features of successful NSF-supported community infrastructure projects in the Earth sciences. The management plan, outlined here, has been codified in a set of by-laws available on our web site (http://Geodynamics.org).

7.a. Institutional Membership and Executive and Science Standing Committees.

CIG is an institutionally-based organization governed by an Executive Committee. The structure of CIG recognizes member institutions, which are educational and not-for-profit organizations with a sustained commitment to CIG objectives, and a number of foreign affiliate members. The Member Institutions will change over time because CIG is an *open organization*, available to any institution seeking to collaborate on the development of open-source software for computational geodynamics and related disciplines.

The Executive Committee is the primary decision-making body of CIG; it will meet at least twice per year to approve the annual science plan, management plan, and budget, and to deal with major business items, including the election of a Nominating Committee. With the Director, the Executive Committee will handle the day-to-day decision-making responsibilities through its regular meetings, teleconferences, and electronic mail. The Executive Committee will have seven members. It will have four voting members: the Chairman, the vice Chairman, and two members at-large. These members will be elected by representatives of member institutions for staggered three-year terms. The three nonvoting members are the Director, the Chief Software Architect, and the Chairman of the Science Steering Committee. The Executive Committee will have the authority to approve proposal submissions and contractual arrangements for CIG. The Executive Committee believes that having an odd number of voting members is prudent than an even and so will be proposing an amendment to the CIG ByLaws to increase the number of voting members from four to five.

CIG has a Science Steering Committee that consists of eight elected members including a chairperson. The committee has a balance of expertise in both geoscience and computational sciences and provides guidance within all of the sub-disciplines of computational geodynamics. Their principal duties are to assess the competing objectives and needs of all the sub-disciplines covered by CIG, provides initial assessment of proposals submitted to CIG, and will revise the Five Year Strategic Plan. Recommendations from the SSC are passed on the EC.

7.b. Administration.

The Director is the Chief Executive Officer of the organization and bears ultimate responsibility for its programs and budget. The Director's responsibilities include: (a) devising a fair and effective process for the development of the Strategic Plan, based on proposals or work plans such as those submitted to the Executive Committee by the

Science Steering Committee, and overseeing the plan's implementation, (b) acting as P.I. on proposals submitted by the core CIG facility, retaining final authority to make and implement decisions on grants awarded to the core facility and contracts, (c) ensuring that funds are properly allocated to various CIG activities, (d) overseeing the preparation of technical reports. The CIG ByLaws do not yet stipulate the term of the Director and so a discussion item at our future Business meetings will devising a mechanism for the orderly transition to subsequent Directors.

The Chief Software Architect (CSA) will serve as a non-voting member of the Executive Committee. His role is to provide advice and perspective to the Executive Committee on the overall composition, integration, and balance between software development activities of the organization. He provides frequent assessments of our software, identifies new opportunities in both computational science and methods for software development, and provides evaluations of prospective members of the Software Development Team. The Executive Committee retains the authority to appoint the CSA.

7.c. Formulating CIG Priorities and Management of its Resources.

Concepts and plans for CIG activities will come directly from the community, member institutions, working groups and their elected committees. Ideas and plans will move from members to the Science Steering Committee and finally to the Executive Committee. As part of the development of the Strategic Plan , the SSC formulates a prioritized list of tasks for software development for the coming year, how these tasks are both inter-related and related to the broader needs of the community, and then transmits this as a recommendation to the Executive Committee. On at least a yearly basis, the Executive Committee will allocate resources to specific software development tasks. Following this allocation of resources, the EC will periodically appoint small committees to interface directly with the software development team (SDT).

It is expected that members of the SSC will be fully engaged in a dialog with the user community and active users of CIG software. Besides the constant dialog that such committee members would naturally have with the community, CIG will have a formal process for bringing new ideas up from the community. On a continual basis, users from Member Institutions will be able to submit one-page proposals for new CIG software development tasks. These proposals can be submitted at any time and are posted on the web for the community to read and evaluate. There will be a comments page where members of the user community can add scientific comments and evaluation. Periodically, but at least once per year, the SSC will evaluate these proposals in light of other information obtained from the community, formulate a prioritized list of tasks, and then submit it to the Executive Committee.

At its disposal, the Executive Committee will have resources to respond to the evolving community needs expressed through these task lists, including the Software Development Team and funds for contracts. However, the Executive Committee will also put into place two mechanisms for generating new resources and funds for CIG.

- *Augmented funding*. CIG will agree to develop additional software upon receipt of augmented funding. For example, a PI at a Member Institution may submit a science proposal to a federal agency in which the proposed work is either wholly or in part dependent upon software not yet available. This software would presumably be more specialized than the highest priority and core CIG tasks, but still encompassed within the mission of CIG and needs of the community. Following submission of a one-page proposal as described above, the Executive Committee will determine whether or not CIG can develop this software. If CIG can develop the software, the EC will detail the resources and funding required on a form for attachment to the PI's proposal. If the proposal successfully passes through peer review and the federal agency agrees to fund the project with augmentation to CIG funding, we will develop the software.
- *Collaborative proposals*. CIG has a specialized staff with skills in software development, numerical analysis, information technology, and related fields, skills not readily accessible within the geoscience community. We believe that members of the community will formulate collaborative research projects with SDT members. If such collaborative projects are judged to be of high merit for CIG by the EC, CIG will develop collaborative proposals. We expect one target of opportunity to be federal programs that require collaboration between scientists from both information technology and the domain sciences, such as the geosciences. It would be expected that such projects would provide funding for both external PIs and members of the SDT.

Software developed through either of these two mechanisms will be open source and made available to the community without restriction, like all CIG software. During its first two years of operation CIG must by necessity focus on a core set of objectives, and would most likely be unable to respond to proposals through these two mechanisms of expansion and funding. However, these two approaches would likely play an increasingly large role within CIG after its formative period.

8. Annual CIG allocations and expenditures

Category	Amount Allocated in NSF Award Year 2	Expected Expenditure Year 2
Senior Personnel	$37,800	$45,126
5 Other Professional (Technical)	$396,550	$411,700
Secretarial	$36,050	$37,000
Total Salaries	$469,680	$493,826
Fringe Benefits	$124,465	$128,395
Total Salaries & Fringe	$594,145	$622,221
Travel	$14,020	$27,625
Participant costs	$90,700	$90,700
Material & Supplies	$83,950	$65,000
Consultant Services	$100,000	$00
Subaward	$126,074	$226,074
Total Other Direct	$310,024	$291,074
Indirect Costs	$491,111	$468,380
Total Direct & Indirect	$1,500,000	$1,500,000

On July 5, 2005 we estimated the balance for year 1 that would occur on September 1, 2005, to be approximately $100,000.

9. Additional funding

none