Review of DOE Computational Science Graduate Fellowship

November 23, 2011

Submitted to the Office of Science,
US Department of Energy

Subcommittee of the Advanced Scientific Computing Advisory Committee

Marsha Berger	Courant Institute, NYU
Dona Crawford	Lawrence Livermore National Laboratory
Bruce Hendrickson	Sandia National Laboratories
Jeffrey Hittinger	Lawrence Livermore National Laboratory
Thomas Manteuffel	University of Colorado (Chair)
Steven Parker	NVIDIA
Vivek Sarkar	Rice University
William Tang	Princeton University

Introduction

In response to a charge letter form the Director of the Office of Science to the Chairman of the ASCAC, dated March 7, 2011, a Subcommittee of ASCAC was assembled to review the DOE Computational Science Graduate research Fellowship. This report to the full ASCAC is in response to that charge (see attached letter).

The Subcommittee was constituted in March of 2011 with the following members:

Marsha Berger (ASCAC)	Courant Institute, NYU
Dona Crawford	Lawrence Livermore National Laboratory
Bruce Hendrickson	Sandia National Laboratories
Jeffrey Hittinger	Lawrence Livermore National Laboratory
Thomas Manteuffel(ASCAC)	University of Colorado (Chair)
Steven Parker	NVIDIA
Vivek Sarkar(ASCAC)	Rice University
William Tang(ASCAC)	Princeton University

This report contains the findings and recommendations of the Subcommittee. The report was informed by data provided by ASCR and the Krell Institute, as well as interviews by a subset of the Subcommittee, (Dona Crawford, Jeffrey Hittinger and William Tang), with Barb Helland of DOE-SC and Thuc Hoang of NNSA, and separately with Jim Corones, Jeana Gingery, and Mary Ann Leung of the Krell Institute, conducted on July 22, 2011.

The specific charge to the Subcommittee includes the following:

1. Examine the effectiveness and impact of the CSGF, as compared to other educational programs.

2. Comment on the quality and breadth of the program over the past decade.

3. Address the participation of women and minorities in the program.

4. Address the projected need for trained computational scientists in the DOE laboratories and for continued US leadership in computational science.

The remainder of this report contains an Executive Summary with findings and recommendations followed by a more detailed discussion of the Subcommittee findings.

Executive Summary

The Computational Science Graduate Fellowship (CSGF) is a fellowship that supports selected graduate students for up to four years of study in pursuit of a Ph.D. at a U.S. University. The program is funded by the DOE Office of Science, through a grant to the Krell Institute. The grant is administered by ASCR and supports approximately 70 students. Complete information about the program can be found at http://www.krellinst.org/csgf/.

The following paragraphs contain the Subcommittee findings on the specific charges and provide recommendations. The remainder of the report provides data on which the findings were based and a more detailed discussion.

Projected Need

Supported by a number of recent reports, the Subcommittee has concluded that the need for well-trained computational scientists in government laboratories and in industry will far exceed the supply for the foreseeable future. This is especially true in the DOE laboratories. We conclude that the need for programs like the CSGF will increase over the next decade.

Recommendation: The Subcommittee recommends that ASCR along with the Office of Science continue to view growth of the computational science workforce as important to its mission and to the nation.

Effectiveness and Impact

Based on outcomes of the Fellowship alumni, the Subcommittee concludes that the DOE CSGF is an exceptionally effective program that has had a significant impact on the national Computational Science infrastructure. Fellows become active contributors in both the public and private sectors. Beyond the immediate impact of the alumni, the CSGF program is further strengthening the U.S. Computational Science community by engaging alumni in the CSGF program so that they have a hand in training the next generation of computational scientists.

As indication of direct benefit to the DOE, a large percentage of Fellows spend a portion of their early career in the DOE laboratories and an even larger portion continue interaction with the DOE laboratories as they pursue their careers in academia and industry.

In light of the effectiveness and impact of this program and in the context of the growing projected need, the Subcommittee has concluded that funding for this program is not only well spent, but that additional funding is needed and should be provided. The fact that qualified applicants far outnumber the awards indicates the program could grow significantly without impacting quality.

Recommendation: The Subcommittee recommends that the funding for the CSGF be put on a path to double over the next five years.

Quality and Breadth

The Subcommittee finds that the quality of the Fellows is exceptional. This is substantiated by the GPA and GRA scores of the applicants and by the outcomes of the awardees, who now occupy a wide range of prestigious positions in academia, industry, and government laboratories.

The Subcommittee also finds that the quality of the management is exceptional. The selection process is well planned and well executed. It is centered on screening and selection committees whose members are prominent computational scientists. The program is distinguished by a number of unique features, including a detailed Plan of Study that ensures the Fellow receives an interdisciplinary education, the Practicum that introduces the Fellow to the DOE Laboratories and to real world experiences, the CSGF Annual Conference that broadens the Fellows view of Computational Science and establishes a network of colleagues, and the alumni activities that support the Fellows' early careers.

The data suggest that the program covers a broad range of disciplines involved in Computational Science. However, Krell Institute has interpreted the mandate to imply support of computational science directed at specific applications in contrast to the more general development of the underlying, enabling sciences like Applied Mathematics and Computer Science. These enabling sciences are essential to the Computational Science endeavor, especially in light of the push to ultra-scale computing. The Subcommittee feels that the CSGF should be expanded to include enabling science when candidates in those sciences work in fields that are relevant to Computational Science.

Recommendation: The Subcommittee recommends that the focus of the CSGF program be expanded to include enabling sciences, either through modification of the current program mandate or through the introduction of separate programs.

Participation of Women and Minorities

The participation of women in the CSGF is significantly higher than in the general scientific community. The participation of minorities appears to be adequate, but the Subcommittee lacked data with which to compare this outcome. The Subcommittee believes that the diversity outreach activities of the Krell institute are commendable.

Recommendation: The Subcommittee commends the Krell Institute on its efforts to increase participation of women and minorities in the CSGF program and recommends that it continue these efforts.

Relation to other Educational Programs

The Subcommittee believes that the CSGF is unique in its focus on Computational Science. It provides features that other Graduate research Fellowships do not, such as the Plan of Study, the Practicum, the Annual CSGF Conference and efforts to keep alumni engaged. In this regard, the CSGF is an exceptional program that produces interdisciplinary scientists uniquely qualified to address current and future computational science challenges.

Recommendation: The Subcommittee concludes that the CSGF is a unique educational program with features the DOE can best provide and recommends that the DOE and ASCR continue stewardship of the CSGF program.

Report of the CSGF Subcommittee

Projected Need for Computational Scientists

The committee reviewed several recent reports that discuss the need for computational scientists, and combined them with our own perspectives from our home institutions.

As the U.S. Department of Energy works to meet its energy, environmental and national security missions, increasingly complex scientific and technological challenges must be addressed. New technologies will be required both for tapping the potential of new sources of energy and for effectively utilizing existing energy resources. Policy makers will need to understand quantitatively the impact of energy policies on the environment and be able to evaluate the risks associated with different strategies for waste storage and environmental cleanup. DOE must also ensure the safety and reliability of the nuclear stockpile while preventing the proliferation of nuclear materials (5). These challenges are detailed in the 2011 strategic plan for the U. S. Department of Energy (http://energy.gov/downloads/2011-strategic-plan).

The execution of DOE's strategic plan will require dramatic improvements in scientific understanding. Given the complexity of the systems of interest, experimental and theoretical approaches alone will not be able to provide sufficient understanding. These will need to be coupled with advanced computational modeling on state-of-the-art computers to begin to assess all the components of the systems over the entire range of scales or interest (1). Therefore, it is imperative that the DOE ensure an adequate supply of interdisciplinary scientists and engineers who can address these multi-faceted issues at the intersection of theory, experimentation and computation. The skills most urgently needed today include a combined understanding of a scientific discipline and computational science and/or computer science (4). Because HPC-based modeling and simulation has firmly established itself as the third branch of scientific inquiry, HPC should be as much a part of scientific education curricula as the two more established branches. Another strong recommendation is for government, academia, and industry to collaborate to increase the number of internship and fellowship opportunities (4). This sentiment was echoed in (3), "DOE, NSF and other agencies should consider creating fellowship programs to train graduate students and postdocs in HPC modeling and simulation, and expanding the Presidential Early Career Awards in Science and Engineering (PECASE) program in this area."

Making this even more imperative, the national security laboratories have found it increasingly difficult to attract and retain the most promising scientists and engineers of the next generation. Increased investments in the nuclear infrastructure and a highly skilled workforce are needed to ensure the long-term safety, security, and effectiveness of our nuclear arsenal. This includes developing and sustaining high quality scientific staff and supporting computational and experimental capabilities (2). The growing industrial demands for these skills further exacerbate the laboratories' recruiting and retention challenges.

From personal experience the committee found the need for computational scientists outpaces the availability of workers with appropriate skills. Industry, academia, national laboratories and government institutions are increasingly looking to computationally trained scientists to meet their current and future challenges. At one laboratory, there is an ongoing hiring requirement for 30 computational scientists. Based on the reports and committee experience, the committee finds CSGF could scale up the program by a factor of two and still not address the need. In addition, the current program could be expanded to more fully include the areas of Applied Mathematic and Computer Science that enable Computational Science.

1. Brown, D. L. (Ed.); *Applied Mathematics at the U.S. Department of Energy: Past, Present and a View to the Future*, - Report by an Independent Panel from the Applied Mathematics Research Community, May 2008. http://science.energy.gov/ascr/news-and-resources/program-documents/

2. *Nuclear Posture Review Report*, April 2010. http://www.defense.gov/npr/

3. *High Performance Computing and U.S. Manufacturing Roundtable White Paper*, from the High Performance Computing Initiative of the Council on Competitiveness, February 25, 2010. http://www.compete.org/publications/detail/1333/high-performance-computing-and-u.s.-manufacturing-roundtable/

4. Joseph, E; Conway, S; Wu, J; *IDC Special Study for DOE: HPC Talent and Skill Set Issues Impacting HPC Data Centers*, December 2010.

5. Oden, J. T. (Ed.); *Simulation-Based Engineering Science: Revolutionizing Engineering Science through Simulation* – Report of the National Science Foundation

Blue Ribbon Panel on Simulation-Based Engineering Science, February 2006. http://www.nsf.gov/pubs/reports/sbes_final_report.pdf

Effectiveness and Impact of the Program

In this section, we assess the CSGF in terms of the educational process, the outcomes for the fellows and the broader impact of the program. In terms of process, we consider the mechanisms by which the CSGF program manages the execution of the fellowship in order to ensure graduates succeed and become high-quality computational scientists. We also consider the quality of the program in terms of successful graduation rates and the types of opportunities offered to alumni of the program.

Fellowship Management

During the fellowship, several mechanisms are used to ensure the effectiveness of training: the plan of study (POS), the practicum, and participation in the annual CSGF Conference. The POS submitted by the fellow as an applicant is treated as a contract. Any changes to the POS require an explanation and approval by the CSGF steering committee, which reviews the request thoroughly. Changes are required most frequently because of changes in the courses offered at the fellow's institution, and the steering committee ensures that appropriate replacements are identified that fulfill the required breadth of training in applied mathematics, science/engineering, and computer science.

The practicum deserves special mention. Every fellow is required to spend one summer working at a DOE laboratory with DOE research staff on a research project independent of the fellow's thesis research. The practicum is to occur before the end of the second year of the fellowship, and fellows are allowed (even encouraged) to return in subsequent summers or to have a second practicum at a different laboratory altogether. The practicum exposes the fellow to the difficult but exciting research problems of the laboratories and provides a great deal of hands-on experience. The CSGF Fellows are highly sought after, and the DOE laboratories actively engage in recruiting them. The CSGF practicum is unique; the NSF Graduate Fellowship, for instance, has no comparable requirement or opportunity.

Each year, the CSGF hosts a conference, typically in the Washington, D.C. area in which the fellows are expected to participate. Each fellow is required to present a poster or, if in their final year, a talk on their research. The conference also serves as an opportunity for fellows to see the breadth of research of their colleagues and to interact with DOE managers, CSGF screening, selection, and steering committee members, and DOE laboratory personnel.

The requirements placed on the fellows by the CSGF program are effective ways to ensure that the fellows receive an appropriate computational science education.

The CSGF Program actively manages the fellows' adherence to their POS, their practicum commitment, and their involvement in the annual conference, but this involvement in the fellows' professional development is not onerous.

Alumni

As discussed in more detail in the section on Quality and Breadth of the Program below, the program is producing successful outcomes. The ultimate measure of the success of the CSGF program, though, is the success of the program alumni. Self-reported data for 102 of the alumni from 2001-2009 were obtained from Krell. Of these, all but one (an Outward Bound instructor) have taken technical positions in academia, national labs or industry. Twenty-seven are or were employed by a DOE laboratory after completing their PhD. We break down the current employment data into different sectors: government laboratories (staff positions and postdoctoral positions at any government lab, not just DOE), academia (faculty, postdoctoral, and research scientist positions), industry, and other (non-technical positions). The results are shown in Figure 1. Approximately 32% of these alumni currently have positions at government laboratories, a positive consequence that suggests that the CSGF fellows received valuable mentoring from the laboratory scientists with whom they interacted. Further, 42% of the alumni have positions at academic institutions, and 24% in industry, leading to a a balanced distribution that demonstrates that the CSGF Program is meeting its goals of providing computational scientists in all sectors.

Fellows have moved on to postdoctoral and research staff positions at most of the DOE research laboratories and laboratories of other government agencies, including NRL, NIST, NASA, and the DOD. In fact, 47 former fellows are currently employed at the DOE laboratories while 22 other former Fellows have left employment at the DOE laboratories. Fellows have obtained academic positions at both domestic institutions, including Stanford, NYU, Princeton, Harvard, Cornell, UC Berkeley, Columbia, Caltech, Michigan, Texas, Wisconsin, and Illinois, and foreign institutions, including Oxford, Cambridge, KAUST[1], and the Institut Pasteur. Corporations hiring alumni include Microsoft, Google, Shell, Exxon Mobil, Seagate, Amyris Biotechnologies, Dataspora, British Petroliem, AREVA, Intellisis, and several companies in the financial industry. Several fellows have even started their own companies. Nine alumni have received highly-competitive postdoctoral fellowships, including the NSF Mathematical Sciences Postdoctoral Research Fellowship, the NSF International Research Postdoctoral Fellowship, the Chandra Postdoctoral Fellowship, the Computing Innovation Postdoctoral Fellowship, the Argonne National Laboratory Director's Postdoctoral Fellowship, the Wigner Fellowship, and the Wilkinson Fellowship.

[1] King Abdullah University of Science and Technology

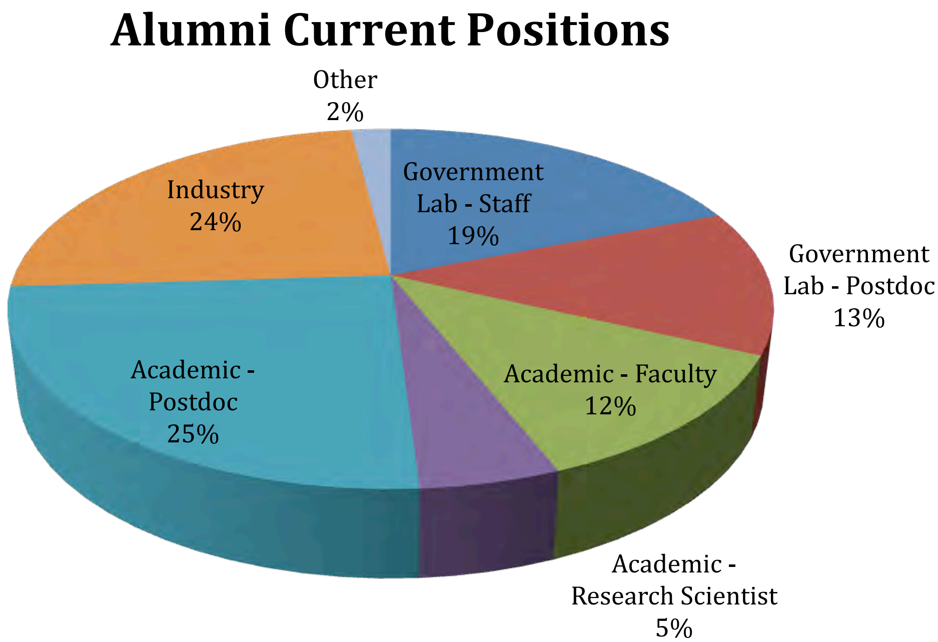

Figure 1: Breakdown of current employment areas for fellows 2001-2009.

The success of the CSGF alumni is clear, and this translates directly into the impact of the CSGF program. The national interest is served well by CSGF alumni who become active contributors in both the public and private sectors. Alumni lead cutting edge research in industry and government laboratories. In addition, a large percentage of Alumni enter academic careers and train future computational scientists. Beyond the immediate impact of the alumni, the CSGF program is further strengthening the U.S. computational science community by engaging alumni involvement in the CSGF program so that they have a hand in training the next generation of computational scientists.

Quality and Breadth of the Program

To assess the quality and breadth of the program, we considered the CSGF recruitment, application, and selection processes. We reviewed data on formal processes and policies, people involved, and outcomes. Data was obtained from the Krell Institute, and interviews were conducted with the NNSA and DOE-SC Program Managers and the Krell Institute Staff.

Based on our investigation, we find that the CSGF Program is of very high quality and very good breadth. In the following subsections, we discuss specifics that led us to this conclusion.

Application Process and Applicants

One way to ensure a high-quality program with a breadth of areas represented is to ensure a diverse pool of well-qualified applicants. Achieving this requires outreach to the scientific community to ensure that the best and brightest future computational scientists are aware of the program and are encouraged to apply. The process also requires a well-designed application that produces the input required for an effective selection process. We obtained information from Krell on both of these topics as well as statistics on the applicant pools.

Outreach

We received detailed information of the outreach and recruitment activities for 2009 and 2010. It is clear that the program pursues many different avenues to increase awareness of the program. Physical recruitment materials include fliers, posters, and the annual DEIXIS publication that highlights the research of fellows and alumni of the program. These are distributed annually at Supercomputing, a handful of university graduate job fairs (e.g., UT Austin, VT, UC Berkeley), and at a variety of conferences including SciDAC[2], SIAM CSE[3] and PP[4], WEPAN[5], NAMEPA[6], Tapia[7], Teragrid[8], NAFA[9], GHC[10], Sigma Xi[11], SACNAS[12], SWE[13], USA Science & Engineering Festival, and student symposiums at the NNSA laboratories. Over 11,000 posters are also mailed each year to organizations and individuals with connections to the CSGF program, for example, fellows, alumni, university coordinators, advisers, and DOE laboratory contacts.

Electronic communications are also used extensively. Announcements are emailed to many of these same individuals (the mailing lists for 2009 and 2010 were approximately 2500 and 4000, respectively) and organizations as well as email lists, such as those for WEPAN, SIAM, and NAFA. Both social and professional networking sites are used to announce the program; on Facebook this includes the Computational Science Education, CSGF Network and WEPAN groups, and on LinkedIn, this includes the CSGF Network, Computational Scientists & Engineers, High Performance & Supercomputing, SciDAC, SIAM and Systers groups. Online advertising is also purchased with the American Institute of Physics and the Materials Research Society.

Finally, advertisement is purchased in print media that coincides with the application period in the November-December time frame. This avenue appears to have expanded in 2010. Advertisements were placed in *Computing in Science and Engineering*, *SIAM News*, *Diversity Careers*, *Materials Research Science Bulletin*,

---

[2] DOE Scientific Discovery through Advanced Computing
[3] Society for Industrial and Applied Mathematics Computational Science and Engineering
[4] Society for Industrial and Applied Mathematics Parallel Processing for Scientific Computing
[5] Women in Engineering Program Advocates Network
[6] National Association Minority Engineering Program Advocates
[7] Richard Tapia Celebration of Diversity in Computing
[8] NSF Teragrid
[9] National Association of Fellowship Advisers
[10] Grace Hopper Celebration of Women in Computing
[11] The Scientific research Society
[12] Society for Advancement of Chicanos and Native Americans in Science
[13] Society of Women Engineers

Association of Women in Science, and the *Association of Women in Mathematics Newsletter*.

On the whole, the outreach activities appear to cast a wide net that targets the broad computational science community, with a particular focus on attracting under-represented candidates. The choice of organizations and distribution mechanisms is appropriate to ensure a potential candidate pool from a wide range of application areas within the greater computational science field.

Application

The CSGF application is based on an online form that is submitted through the CSGF website by the candidate. There are two main sections to this form. The first records the biographical information that would commonly be found on a resume or CV including the candidate's previous education; anticipated graduate program, adviser, and research topic; awards and honors; publications; and employment, including experiences at federal or private laboratories. The second portion of the form focuses on the applicant's research and graduate education plans. It is comprised of three research statements, a Program of Study (POS), and an additional comments section. Three letters of reference and official Graduate Record Examination (GRE) scores are added to form a complete application. From the perspective of quality and breadth, the most important aspects of the application are the research statements and POS.

The research statements address three topics: the field of interest, the program of study, and high-performance computation and research. Specifically, the applicant is asked to address:

1. Their specific research interest and how computational science will spur advances;
2. Why they selected the courses in their POS and how these courses will impact their research plans; and
3. How the use of high performance computing (HPC) would advance their research beyond what is possible on a modest-sized cluster.

Each topic is preceded with a short motivation and explanation, e.g., for HPC, a list of common reasons for migrating to HPC is provided, and the response is not to exceed 300 words.

In the POS section, the applicant provides an accounting of previous and future coursework categorized into Science/Engineering, Mathematics and Statistics, and Computer Science. The credit hours, academic level (graduate or undergraduate), term/year, and grade (if completed) for each course is indicated. It is a requirement of the program that a fellow must take two courses in the categories outside of their primary research area, so for example, an engineering student would be required to take two Mathematics or Statistics and two Computer Science courses beyond their planned engineering courses. For each course listed, the applicant provides a course description. The graduate adviser then approves the plan.

The application addresses the key areas by which the candidates are evaluated within the goals of the CSGF program. Other than providing some basic background material on the research plan topics, the research plan does not impose a particular view of computational science on the applicants and allows them to make the case for how their research fits within the scope of the program. The additional comments section provides another opportunity for candidates in less traditional computational science areas to convince the reviewers. The POS explicitly ensures that the applicant understands and can construct a schedule of coursework that reflects the broad nature of computational science. Finally, in addition to the subjective material by which reviewers can evaluate the quality of a candidate, the application provides a sufficient amount of quantitative information on the candidate's previous performance.

Finally, we note that the DOE program managers specifically commented on the impressive software infrastructure the Krell Institute has developed to handle the CSGF application process (and fellowship management). Comments were made that this is far superior to that used by other national programs, e.g. NSF. Certainly, our task of reviewing the program was enabled substantially by output generated from these databases, and Krell is justifiably proud of this system.

Quantitative Results

The pool of Fellowship applicants has grown from 159 in 2002 to 628 in 2011. This nearly four-fold increase reflects the growing interest in computational science among students and within universities. It presumably also reflects an expectation of a robust job market for computational scientists. In that time, the number of new Fellows has declined from 25 to 17. Expressed in percentages, 15.7% of the applicants in 2002 received a Fellowship, compared to only 2.7% of the applicants in 2011.

A breakdown of the applicants by the self-reported field of study is provided in Table 1. Over the decade, Biology and Bioengineering have increased their share of the applicants from roughly 8% to 17%. Similarly, the Physical Sciences have increased from 17% to 21%, albeit not monotonically. The share of engineering applicants has remained relatively steady at about 40%, with some fluctuation. Only the Computer Science and Mathematics area has seen an overall relative decline from 26% to 18%.

Table 1: Breakdown of CSGF applicants by major field of study for years 2002-2011.

Area	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Bio & Bioeng	13	40	43	49	66	72	64	60	80	108
Math & CS	41	90	76	74	108	71	69	76	113	115
Engineering	65	113	116	133	149	150	152	125	194	243

Physical Sci	27	58	74	71	69	92	76	73	119	134
Social Sci	1	1	1	1	3	1	1	1	3	1
Did not report	12	16	12	10	15	10	9	14	21	27
Total	*159*	*318*	*322*	*338*	*410*	*396*	*371*	*349*	*530*	*628*

The quantitative metrics from the applications for the 2002-2011 fellowship years are shown in Table 2. Note that these are the averages of *all* applicants, not the CSGF recipients. Clearly, the fellowship attracts a high-quality applicant pool if these standard measures are used as an indication.

Table 2: CSGF applicants' quantitative measures for 2002-2011.

Year	Average UGPA	Average Percentile GRE Verbal	Average Percentile GRE Quantitative
2002	3.51	74	87
2003	3.60	75	85
2004	3.62	75	82
2005	3.59	73	83
2006	3.61	75	82
2007	3.68	75	85
2008	3.64	78	87
2009	3.60	79	86
2010	3.59	77	84
2011	3.64	77	85

This quantitative data indicates that CSGF program has been successful in reaching the broader computational science community. The data is consistent with an improving awareness of the program through effective outreach and also correlates well with the increased emphasis in scientific computing in the biological sciences in the last decade.

We remark that the number of applicants has grown significantly over the past decade while the number of awards has been fairly stable. In 2011, 628 applicants resulted in 17 awards. This yields a success rate of 2.7%, which is painfully low. There are two conclusions that can be drawn from this. The first is that there is a large body of highly qualified students interested in Computational Science. The second is that the program could be grown significantly without compromising the quality of the selection process or the quality of awards.

Selection Process And Participants

Prior to our meetings with program managers and representatives of Krell, one concern of this subcommittee was that the criteria by which fellows are selected was not clearly documented. We were provided with very detailed documentation on the *process* and the *participants*, but could not find explicit program definitions of *criteria* that could be used to evaluate the non-quantitative aspects of the application. At most, the CSGF application contains a very general definition:

> "Computational science" involves the innovative and essential use of high-performance computation, and/or the development of high-performance computational technologies, to advance knowledge or capabilities in a scientific or engineering discipline.

Several questions were raised, for example: How does one weigh the application of existing technologies to advance science versus the development of new methods and techniques in this pursuit? Is there any preference for simulation over more data-centric scientific discovery, i.e., data mining and informatics? Is the goal to encourage non-computational scientists to enter into the field of scientific computing or to reward outstanding applicants already in the field? This concern about criteria used to evaluate the applicants was an important issue we sought to address in the interviews.

We discovered that this situation is by design in order to allow the program to adapt with the evolving field of computational science. The concern of Krell is that to delineate explicit criteria would lead to a "check-the-box" process that, in time, would cause the program to ossify, reducing both its quality and breadth. Instead, Krell has produced a more organic process that sets few boundaries and relies on carefully chosen screening, selection, and steering committee members to maintain the program and to allow it to evolve with computational science. Questions of the type we raised were acknowledged to be absolutely critical and are annually the source of much debate on the selection committee. Because of this ongoing discussion, which is a much more difficult and time-consuming approach to selecting fellows, it is believed that the process produces a stronger and broader final group of recipients. According to Krell, the ambiguity in the task allows one "non-traditional" or "high-risk" candidate to be awarded a fellowship every two to three years.

The current screening and selection process itself is well documented in the brochure "DOE Computational Science Graduate Fellowship Review Process;" we will only briefly outline the process. The success of the process is fundamentally dependent on the quality of the participants reviewing the applications, so we instead devote more discussion to the screening and selection of committee members.

Screening and Selection Process

The current review process has two main phases: screening and selection. All applications are read completely by at least two reviewers. The screening phase removes inappropriate or incomplete applications and pares down the bulk of the applications. The selection phase is an intensive multi-step, multi-team process that further reduces the results of the screening phase down to the final set of recipients.

The screening phase begins with a pre-screening review by Krell staff. Applications are divided into three groups: incomplete, phony, or seriously flawed applications that are removed from consideration; top tier applications that are sent directly on to the selection phase for review; and the remainder that are sent to the screening committee for review. Top tier applications are selected through a quantitative formula that considers undergraduate GPA and GRE scores and that has been validated against the applications and results of previous years. The screening committee is divided into teams of two. Each team thoroughly reviews approximately thirty applications and typically passes forward seven or eight consensus candidates. The resulting application pool is the union of the top tier candidates and the applications selected by each screening team.

The selection process starts with this reduced pool of applications. Calibration of the selection committee is done using previous successful and unsuccessful applications. The selection process is divided into four rounds. In the first round, the twelve-member selection committee is divided into four teams (A, B, C, and D), each with three members. Each team receives one quarter of the applications, and each team member is assigned two-thirds of these applications to review. Each team then puts forward roughly half of their pool. In round two, Teams A and B swap the applications put forward in the first round; similarly, C and D swap applications. In round three, the two merged Teams AB and CD put forward roughly half of their candidates, and each application is assigned a reviewer who will lead the discussion of that application. Finally, in round four, the entire selection committee reviews each remaining proposal as a committee, and applications that should be offered fellowships are selected as well as a second pool of candidates who the committee feels are qualified should more funding be made available or should a selectee decline the fellowship. The final three rounds of the selection process are conducted over an intensive three-day meeting, and the DOE Program Managers usually attend the final day of discussions.

The review process is well designed to balance the need to handle a large pool of applications while giving each application a fair and thorough review. The process has built-in checks and balances, and it appears that Krell reviews the process across program-years to ensure that it functions as intended.

Screening and Selection Committees

As indicated earlier, the quality and breadth of the selected fellows is dependent on the quality and breadth of the reviewers. Krell indicated that they are very careful in who they ask to serve on the screening and selection committees. We were provided with a great deal of information on the history and composition of the screening and selection committees, including short biographical sketches for most years since 2005.

All reviewers have Ph.D.'s in technically relevant fields. The selection committee is mostly comprised of internationally-known experts in computational science, and at least one alumnus/a of the program is typically represented on the selection committee. Between 2001 and 2010, fifty-seven individuals have served on the selection committee; committee members tend to serve three or more successive years, which brings some amount of continuity to the process while allowing an influx of new reviewers and perspectives. The screening committee, in contrast, has grown since 2004 from a handful of experts to a large group mostly comprised of alumni of the program. A total of thirty-nine individuals have served on the screening committee between 2004 and 2010. A key criterion for the selection of screening committee members is that they have a thorough knowledge of the CSGF program, so it is natural to draft alumni and current and past selection committee members to serve. Thus, in terms of qualifications, the reviewers represent highly-regarded computational scientists.

We synthesized the data provided from Krell into a format similar to that for the applicants to reconstruct distributions of reviewers by general field of expertise and whether they worked in academia, industry, or government (mostly government laboratories). We note that this is an inexact process, since some reviewers had joint appointments and reviewers do not necessarily work in the field in which they received their degrees.

Table 3: Breakdown of CSGF selection committee by major field of expertise for years 2001-2010.

Area	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Bio & Bioeng	0	0	0	0	1	0	0	0	0	0
Math & CS	5	5	4	3	4	5	4	4	4	6
Engineering	0	0	2	4	3	4	5	4	3	2
Physical Sci	5	4	6	5	4	3	3	4	5	4
Total	*10*	*9*	*12*	*12*	*12*	*12*	*12*	*12*	*12*	*12*

Table 4: Breakdown of CSGF selection committee by organization for years 2001-2010.

Area	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Government	2	2	5	5	4	6	5	6	8	10
Industry	1	1	0	1	1	1	0	0	0	0
Academia	7	6	7	6	7	5	7	6	4	2
Total	*10*	*9*	*12*	*12*	*12*	*12*	*12*	*12*	*12*	*12*

The selection committee data in Table 3 and Table 4 demonstrate the inclusion of reviewers from engineering over the time period and a fairly even division between Math/CS, Engineering, and Physical Sciences. The selection committee has gradually evolved from being dominated by academics to being dominated by laboratory and other government representatives. Representation from Biology and industry has been sporadic at best, and this is perhaps a cause for concern.

Table 5: Breakdown of CSGF screening committee by major field of expertise for years 2004-2010.

Area	2004	2005	2006	2007	2008	2009	2010
Bio & Bioeng	0	0	0	0	2	2	4
Math & CS	1	2	2	2	6	3	6
Engineering	0	1	2	2	1	5	12
Physical Sci	0	2	4	2	5	6	6
Total	*1*	*5*	*8*	*6*	*14*	*16*	*28*

Table 6: Breakdown of CSGF screening committee by organization for years 2004-2010.

Area	2004	2005	2006	2007	2008	2009	2010
Government	0	1	4	3	6	9	16
Industry	0	0	0	0	1	1	3
Academia	1	4	4	3	7	6	9
Total	*1*	*5*	*8*	*6*	*14*	*16*	*28*

The screening committee data is presented in Table 5 and Table 6. The growth of the committee is clear, no doubt motivated by the significant increase in applicants between 2002 and 2004. The most recent data reflects a distribution similar to the distribution of applicants across the fields. Similar to the selection committee, the participation of reviewers from the laboratories and other government organizations has grown to dominate the committee; many members of the screening committee in recent years are alumni who currently work at DOE laboratories.

The data indicate that the screening and selection committees represent the necessary breadth for the program. Nevertheless, consideration should be given to the inclusion of more perspectives from the Biosciences and from industry. The reasons for and effects of the recent trend towards heavy involvement of laboratory personnel should be studied further.

Fellows

A breakdown of the awardees by the self-reported field of study for years 2002 through 2011 is provided in Table 7. Aside from the early domination in the engineering field, the breakdown demonstrates fellowships awarded in all areas in proportions similar to the mix of the applicants. Each area shows some periods of increase and decline, which could be due to the strength and number of applicants and/or the make-up of the committees. In no case is there a general trend that suggests that any one field is systematically being eliminated from consideration.

Quantitative metrics of the awardees are presented in Table 8. We see, as expected, that the fellows have higher average undergraduate GPA and GRE scores than the applicant pools for the same years. Between 2002 and 2011, the GPAs have fluctuated between 3.72 and 3.92, and the quantitative GRE scores have remained fairly steady around the 90th percentile. The verbal GRE scores have shown an overall improvement over the decade to values that rival the quantitative GRE scores. These (admittedly relative) metrics demonstrate that the CSGF program is awarding fellowships to high-quality candidates.

Table 7: Breakdown of CSGF awardees by major field of expertise for years 2002-2011.

Area	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Bio & Bioeng	3	1	1	3	3	5	6	3	2	3
Math & CS	3	3	3	1	5	1	2	2	5	3
Engineering	15	8	4	7	3	5	6	6	6	4
Physical Sci	3	3	6	4	8	4	4	5	7	7
Social Sci	0	0	1	0	0	0	0	0	0	0
Did not report	1	1	0	0	1	1	0	0	0	0
Total	*25*	*16*	*15*	*15*	*20*	*16*	*18*	*16*	*20*	*17*

Table 8: CSGF awardee quantitative measures for 2002-2011.

Year	Average UGPA	Average Percentile GRE Verbal	Average Percentile GRE Quantitative
2002	3.72	77	90
2003	3.86	86	90
2004	3.90	83	88
2005	3.73	80	88
2006	3.92	85	89
2007	3.87	86	89
2008	3.80	86	91
2009	3.86	85	92
2010	3.81	87	91
2011	3.88	90	90

Another way to measure of the quality of the fellows is to consider the universities to which they were admitted for their graduate studies and their success at these institutions. Figure 2 provides this data for the years for which data was provided (2001-2009). Only institutions with more than one fellow in this period are explicitly identified. Fellows attended forty-four different universities that represent most of the top research universities in the country. We have limited reporting on graduation statistics for the years 2007 and forward; most of these fellows are still in school. However, if we consider the years 2001-2005, 92% of the 95 reported fellows have received their Ph.D. This number increases to 96% if we include those who expect to receive their degree this year and remove the one deceased fellow. If we expand the years of consideration to 2001-2006, we again find a 96% graduation rate if we include the seven additional fellows who expect to receive their degrees by the end of this year. These are very good success rates.

Clearly, the fellows represent an elite group of graduate students. Their fields of study span the range of computational science subjects. That so many fellows attend top research institutions indicates that the CSGF program is selecting from among the very top graduate students in the country.

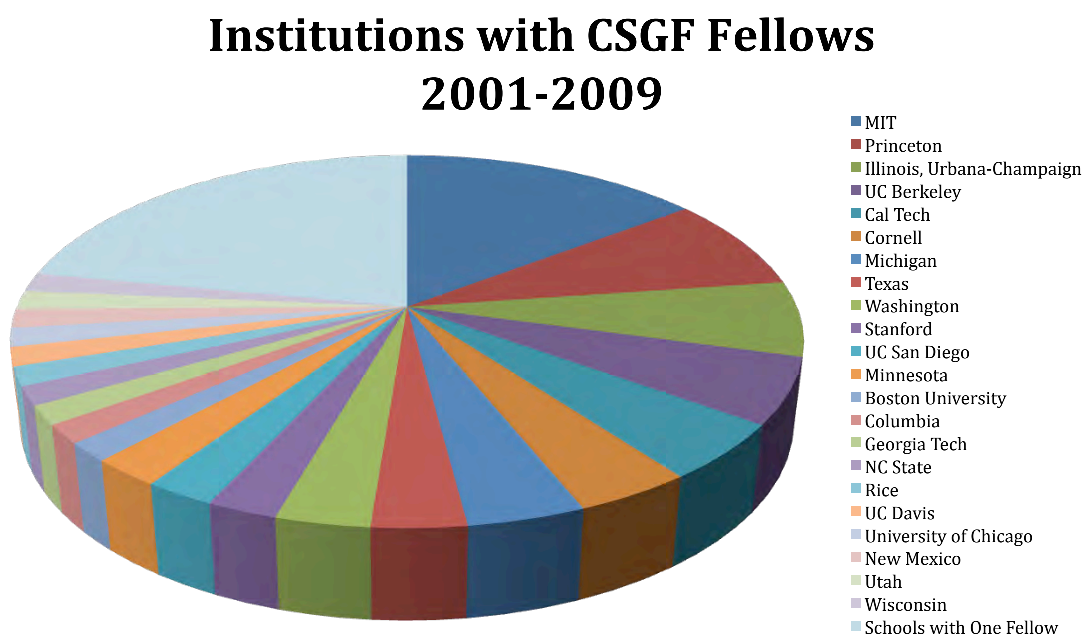

Figure 2: Distribution of fellows across universities for years 2001-2009.

Participation of Women and Minorities

In this section, we address the participation of women and minorities in the CSGF program. The data provided by ASCR indicate that over the past decade 26% of the applicants (who reported) were female while 29% of the awardees (who reported) were female. These numbers are significantly higher than the percentage of women generally found in scientific disciplines. For example, the Society for Industrial and

Applied Mathematics (SIAM) has approximately 15% female members. The AMS reports that only 11% of mathematics Ph.D.s are awarded to women.

The data also show that 26% of the applicants (who responded) claim minority status while 19% of recipients are minority. It is difficult to analyze this difference because the proposals are reviewed blindly, as described in the section on Quality and Breadth.

The Krell institute has a very active outreach program as detailed in the Outreach subsection of the section describing Quality and Breadth above. Activities specifically targeting women and minorities include Krell Staff, Fellows and Alumni attending diversity related conferences. Often booths are staffed and on occasion, plenary talks are given. Sample activities include a presence at the following conferences:

- Women in Engineering Program Advocates Network (WEPAN),
- Association Minority Engineering Program Advocate (NAMEPA),
- Society of Women Engineers,
- Society for the Advancement of Chicanos and Native Americans in Science (SACNAS) ,
- Grace Hopper Conference Celebrating Women in Computing,
- Richard Tapia Celebration of Diversity in Computing Conference,
- SCxy Broader Engagement Program.

In addition, printed materials are sent to Diversity Careers, Association of Women in Science, and Association of Women in Mathematics. Email announcements are sent to a variety of organizations such as Women in Engineering WEPAN, Systers (women in computing organization), and to individuals identified at diversity related conferences such as SACNAS.

The Subcommittee feels that these efforts are commendable and no doubt lead to a higher participation of women and minorities than otherwise.

Other Educational Programs

The Subcommittee is not aware of any similar educational programs specifically targeting Computational Science. The National Science Foundation has a Graduate Research Fellowship in all Divisions. Any of these might make an award to a proposal that is computational in nature. For these we have no data. Two organizations that are closely associated with computational science are the Division of Mathematical Sciences (DMS) and the Computing and Information Sciences (CISE). The following table indicates the number of GRFs awarded by them over the past five years.

Year	2006	2007	2008	2009	2010
CISE	53	52	82	102	110
DMS	28	23	63	62	80

Again, no data was available of the number of fellowships focused on Computational Science. This Subcommittee believes that only a small fraction of these awards address computational issues. For example, in DMS Applied and Computational Mathematics are two of eleven programs.

Other programs that grant graduate research fellowships that might include Computational Science include:

- DOD National Defense Science and Engineering Graduate Fellowship
- NASA Graduate Student Research Programs
- EPA STAR Graduate Fellowship
- USDA National Needs Graduate Fellowship
- NIH NRSA for Individual Pre-doctoral Fellowships

However, none of these programs provide the features that make the DOE CSGF especially effective, namely, the Program of Study, the Practicum, the CSGF Annual Conference and the sustained connection with the CSGF Alumni.

The Subcommittee believes that the DOE CSGF is the only educational program specifically focused on Computational Science.