Report of the 2014 NSF CyberBridges Workshop:

Developing the Next Generation of Cyberinfrastructure Faculty for Computational and Data-enabled Science and Engineering

June 2-3, 2014
Arlington, Virginia

Thomas Hacker
Purdue University

Suzanne Shontz
Mississippi State University

Table of Contents

1.0 Executive Summary

The third annual workshop for the NSF Division of Advanced Cyberinfrastructure (ACI) CAREER awardees was held June 2 – 3, 2014 in Arlington, Virginia. Twenty-seven attendees and five keynote speakers attended the workshop. The attendees, who were funded by nine divisions that support research related to the development and use of cyberinfrastructure, were selected from twenty-five institutions. Seven of the attendees of the workshop attended one of the previous workshops (i.e., either the 2012 or the 2013 workshop). Four of the attendees attended both the 2012 and 2013 workshops and returned for this year's workshop. Five keynote presentations were given by nationally- and internationally-recognized leaders in fields relevant to the use and development of cyberinfrastructure in science and engineering research. Each keynote presentation was followed by a discussion session with workshop attendees. The workshop provided a venue for CAREER awardees to interact and to develop new collaborations with leading researchers and other CAREER awardees.

Four broad themes emerged from the workshop presentations and discussions among attendees. First, **moving to exascale** poses a significant challenge for algorithms, system architectures, and the cyberinfrastructure science community. Not only will hard constraints such as scaling, data movement, and power limitations need to be addressed – the community of computer scientists, mathematicians, and experimental (domain) scientists will need to closely collaborate to fully exploit exascale capabilities. The second theme relates to the **growth in the influence and importance of data**. The analysis of the vast amount of data available from instruments, sensors, simulations, and social media outlets requires high performance computational resources, visualization, networks, and storage systems that are architected to eliminate performance bottlenecks and barriers to usability. The third theme, **sustaining the cyberinfrastructure ecosystem**, was focused on the challenges in developing technologies, improving usability, and fostering and promoting interdisciplinary work to continue the development and to promote the adoption of cyberinfrastructure. The fourth theme addressed **issues related to faculty life and interdisciplinary research**. Early career faculty described the need for awareness of and credit for non-traditional forms of scholarly output, such as software and data contributions. There also needs to be support and encouragement for interdisciplinary research. Ideas discussed at the workshop included improved mentoring for early career faculty, and the development of awards to spotlight the achievements of early and mid-career faculty. In terms of education, early career faculty are aware of the need for innovation in education. However, they are concerned that efforts invested in improving education will not be valued by research-focused senior colleagues.

2.0 Workshop Overview

In 2010, the NSF Office of Cyberinfrastructure (now the Division of Advanced Cyberinfrastructure) began making awards in NSF's Faculty Early Career Development (CAREER) program, to support investigators working on interdisciplinary research in cyberinfrastructure and the application of cyberinfrastructure to science and engineering research. We held the first workshop for Office of Cyberinfrastructure CAREER awardees on June 25-26, 2012, in Arlington, Virginia. At the time of the workshop, approximately 50 CAREER projects had been awarded to researchers funded or co-funded by OCI. The workshop attendees, who were funded by OCI and the NSF, BIO, CISE, ENG, EHR, and MPS directorates, were selected from 24 institutions. Five keynote presentations were given by nationally and internationally recognized leaders in fields relevant to the use and development of cyberinfrastructure in science and engineering research. Each keynote presentation was followed by a discussion session with workshop attendees. The workshop provided a venue for CAREER awardees to interact and to develop new collaborations with leading researchers and other CAREER awardees. As a result of this workshop, 55 potential new collaborations were identified by attendees. The workshop provided many opportunities for discussions among attendees and speakers. We received many positive comments and positive survey feedback from the attendees, and encouragement to propose a follow-on workshop in 2013.

To bring together the community of ACI CAREER awardees and to build upon the successes of the 2012 workshop, we held a second workshop on July 15-16, 2013 in Arlington, Virginia. . Attendees of the 2013 workshop, who were funded by thirteen divisions, were selected from twenty-nine institutions. Five keynote presentations were given by nationally and internationally recognized leaders in the areas of Computational and Data-enabled Science and Engineering, Scientific Visualization, High Performance Computing, Education, Grand Challenges and Interdisciplinary Research. The workshop provided a venue for CAREER awardees to interact and to develop new collaborations with other CAREER awardees and other leading researchers. Forty-six potential new collaborations were identified by the attendees. Interactions among attendees and speakers were also encouraged. We received many positive comments and positive survey feedback from the attendees and were encouraged to propose a follow-on workshop in 2014.

In 2014, we proposed and held a third workshop in order to bring together the community of ACI CAREER awardees and to build upon the successes of the 2012 and 2013 workshops. The goals of the workshop were to: (1) encourage networking and discussion among awardees; (2) provide a forum to facilitate the discovery of new synergies and connections among researchers from the community; and (3) provide inspiration and motivation for new research through a series of keynote presentations by leaders in the areas of Computational and Data-enabled Science and Engineering, Visualization, High Performance Computing, Education, Grand Challenges in Cyberinfrastructure and Interdisciplinary Research. Building the community of ACI CAREER awardees was an emphasis of the workshop. The workshop provided networking opportunities for attendees to seek and gain potential collaborators, and (as in prior years) included a poster session that allowed poster presenters to solicit additional interest from attendees.

2.1 Outcomes of the NSF CyberBridges Workshop

Similar to prior years of the workshop, attendees found that the workshop was useful, talks were interesting, the thematic areas included their areas of research and education, and there were sufficient opportunities for networking and collaboration.

Presentations and discussions at the workshop focused on several broad themes. The first addressed the work that will be needed to effectively use the tremendous capabilities of exascale systems as they become available over the coming decade. There are many difficult challenges to be addressed on the systems side (i.e., power management and managing data movement) as well as on the algorithmic side (multiscale algorithms that can adapt to the limitations of exascale systems.) Another area involved efforts needed to sustain research, development, and adoption of cyberinfrastructure. The problem areas discussed included technological issues, the need to address some needs in cyberinfrastructure today, and the need to increase the diversity of the community. The work needed to improve these areas included developing new methods for dealing with the increasing complexity of high performance computing systems, the need for a multidisciplinary approach for research and development in cyberinfrastructure that pairs computer scientists, mathematicians, and domain scientists, and the need to improve the usability of cyberinfrastructure. The third theme focused on issues related to faculty life and interdisciplinary research. Early career faculty attendees described the need for awareness and credit from senior colleagues for software and data contributions, especially for promotion and tenure. To aid this effort, the need for more early- and mid-career awards was discussed. Another topic discussed in this theme was the need for improved mentoring for early career researchers from professional societies and established leaders.

The thematic area of education and training was also addressed in the workshop. Broader training and education are needed in parallel computing to promote the effective use of multicore processors by the science community. Moreover, early career faculty are aware of the need for innovation in education, but are concerned that their efforts in this area would not be valued as much as research by senior colleagues. Attendees also discussed the need for a supportive environment that would allow them to engage in interdisciplinary research

The final thematic area of the workshop involved the growth in the influence and importance of data. The vast amount of data now available is driving the need to interconnect data resources with large scale computational and visualization systems. There is a need for new algorithms and techniques to more effectively exploit available data, and to match computational resources with these data.

2.2 Workshop Attendees

Twenty-seven attendees and five keynote speakers attended the workshop, from thirty-two institutions, including one international institution (Fig. 1). Many attendees were funded at least in part through the Division of Advanced Cyberinfrastructure, but several held awards partially funded through other divisions, including Computing and Communication Foundations, Materials Research, Chemistry, Physics, Biological Infrastructure, Information and Intelligent Systems, and Civil, Mechanical and Manufacturing Innovation (Fig. 2). Eleven attendees were funded entirely from outside ACI, and eight attendees were funded through multiple divisions. This year attendees' CAREER awards spanned nine divisions, which is somewhat less than the previous year but more than the first year that the workshop was held.

Figure 1. Map of the United States showing locations of the various attendee home institutions at the CyberBridges Workshop..

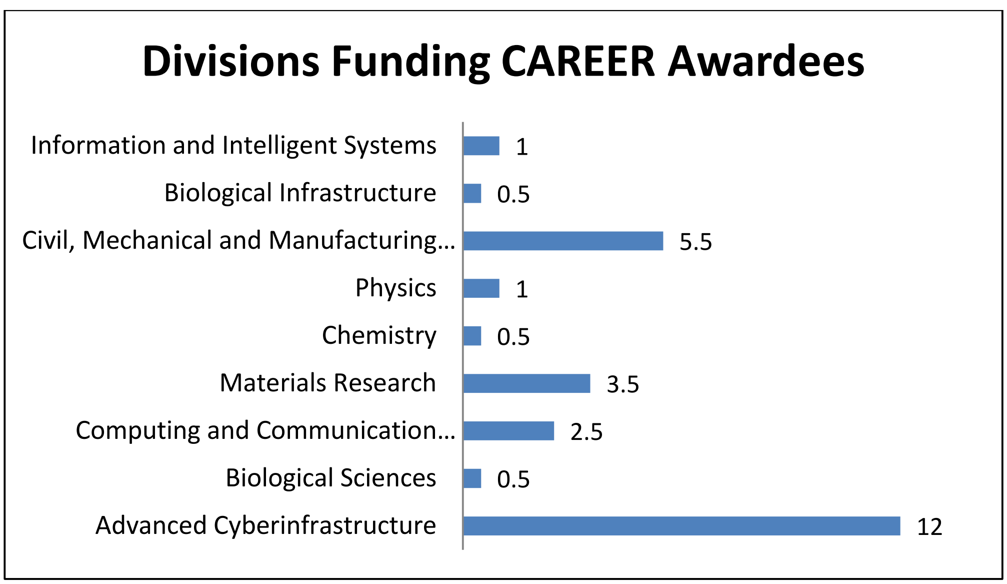


Figure 2. Distribution of the divisions funding the attendees of the CyberBridges Workshop. CAREER Awards funded through divisions are split evenly between the relevant offices.

2.3 Attendee Selection Process

Similar to prior years, faculty who received NSF CAREER Awards from ACI (including through co-funding with another NSF directorate or division) were invited to attend the workshop. The invitation process (in priority order) was to invite new (over the past year) CAREER awardees before prior awardees, followed by other ACI awardees before awardees from other NSF directorates.

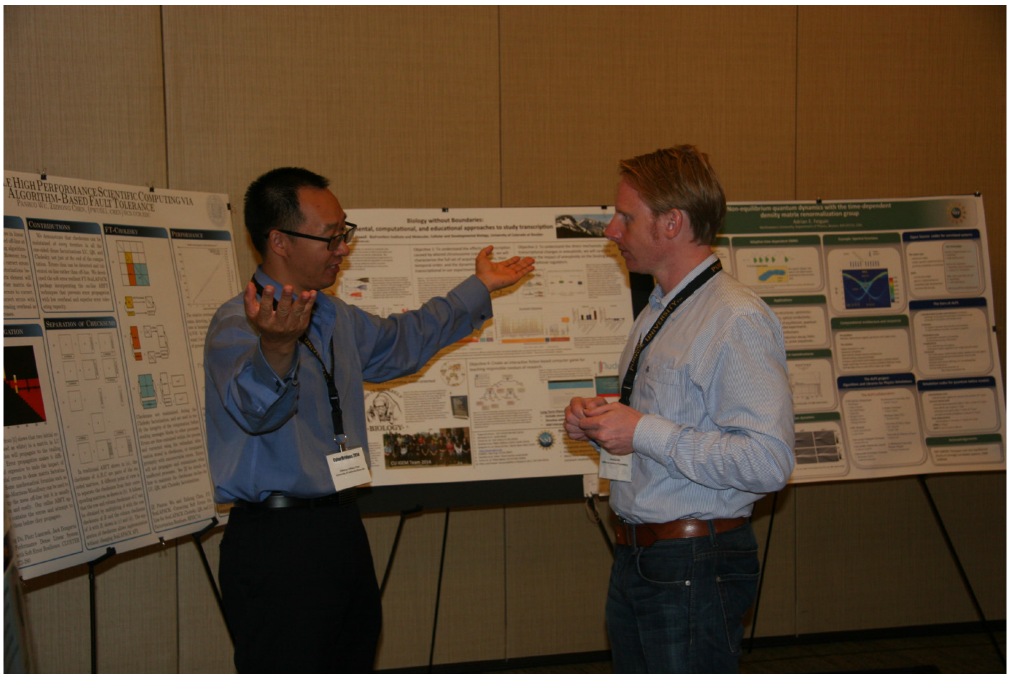


Figure 3. Twenty seven workshop participants attended talks and presented their work in a poster session. Shown here are two NSF CAREER Awardees discussing their work.

3.0 Workshop Themes

The workshop focused on five thematic areas that reflect the spectrum of research and education activities in which the Division of Advanced Cyberinfrastructure (ACI) is focused, and they encompass the types of computational and data-enabled science and engineering (CDS&E) in which ACI and many other NSF directorates are engaged.

1. <u>Computational and Data-enabled Science and Engineering</u>

The first thematic area of the workshop focused on *computational* and *data-enabled science and engineering*, which involves the development of algorithms and cyberinfrastructure necessary to perform large-scale simulations or to process and interpret data generated from experiments, simulations, models, and observations in science and engineering. Dr. David Keyes gave a keynote talk on the development of exascale computing algorithms for use in multi-physics simulations and led the ensuing discussion among workshop participants. Bulk synchronous processing (BSP) and the message passing interface (i.e., MPI) paradigm are currently used to parallelize algorithms; this strategy has served us well for the past two decades. However, exascale computing, is qualitatively different and requires new approaches to algorithm design. In particular, exascale algorithms should focus on: (1) reduced synchronization; (2) greater arithmetic intensity; (3) greater SIMD-style shared-memory concurrency; (4) built-in resilience to arithmetic faults and lost/delayed messages; (5) control of vertical data motion to minimize the energy used, and (6) variable precision. Algorithms should be designed with attention paid to high performance and high productivity on the "multis" including: (1) multi-scale; (2) multi-physics simulations in multi-dimensions; (3) multi-models; (4) multi-levels of refinement; (5) polyalgorithms in multiple precisions; (6) multi-protocol hybrid programming styles; (7) multi-core, massively multi-processor systems, and (8) a multi-disciplinary approach. Co-design should be employed for: (1) design of an effective combination of hardware and software, and (2) design of various algorithms for use in multi-physics simulations that minimize data motion. An example of an application area for which exascale numerical linear algebra algorithms are being developed is mirror adjustment for the European Extremely Large Telescope. Dr. Keyes also encouraged NSF CAREER Awardees to develop international collaborations; the NSF International Science and Engineering Sections funds international research activities of U.S. participants.

Dr. Keyes posed the following questions to workshop attendees:

1. *Given that the starting requirements for impactful research in CDS&E often require a critical mass of collaborators from multiple disciplines, significant facilities, and a period of stable support during which to take some risk and invest in coding, what strategies can put a researcher into this enviable position?*

   Scribe: Sophya Garashschuk

   The breakout group discussed several strategies that can put a researcher into the enviable position of having the resources needed to do impactful research in CDS&E. First, grants that are awarded for longer periods of time (e.g., for 5 years, i.e., long enough to support a Ph.D. thesis project) would help one to be in this position. Second, the opportunity to engage national lab staff and researchers for a prolonged period of time would be helpful. One idea is to fund a team of experts in software/hardware with background in a related STEM discipline, like the Sustainable Software Innovation Institute for Computational Chemistry and Materials Modeling (S2I2C2M2), for example. Third, short courses by computer science experts who can help with specific projects of participants and assess changes made to the code (efficiency, including energy efficiency) would be helpful to CDS&E

researchers. Fourth, internships for graduate and undergraduate students from computer science to go into a STEM research lab, and for STEM grad students to go to computer science research labs and national labs for training and collaborative work would help to gain interdisciplinary knowledge which is needed for impactful CDS&E research. Fifth, interdisciplinary funding should be awarded for hard, risky problems that require scientific innovation in computer science relevant to a big STEM problem.

2. *Given that universities are increasingly joining other types of employers in seeking to own and exploit intellectual property generated by their researchers, how can researchers whose primary contributions are in scientific software benefit best from adoption of their software by others?*

   Scribe: Judy Qiu

   Software authors do not have the ownership of intellectual property while a university has the ownership in federal funded research projects. How can software authors benefit or get credit for software development?

   The group discussed three ways in which software authors can get benefit or credit for their software development efforts. First, they can obtain benefit through an open source project (i.e., credit) or commercial software (i.e., financial benefit). Second, through the use of electronic links to software in publications (e.g., by DOI), the users of the software can cite the software used in their publications; this can be done by working with the publisher's software. Third, university committees and administrators can evaluate software as products similar to how publications are evaluated when a candidate is evaluated for tenure and promotion.

3. *What can software developers fairly expect and seek in terms of attribution, professional reward, external support, and long-term satisfaction from adoption of their software?*

   Scribe: Haibin Ling

   Three groups were identified which could work to develop improved credit attribution models through for software contributions. First, in regards to academia, universities should study their IP policies governing the use of open source software and the credit given to software authors. In addition, universities should work to determine how to quantitatively evaluate software contributions in the evaluation of promotion and tenure. Second, journals could develop mechanisms for code development related to journal article publication; alternatively, publications *regarding the software could be accepted. Third, funding agencies, such as the NSF, should determine how to encourage data/code sharing. In addition, they should consider providing repositories with links to project results. It was noted that a database is easier to maintain in regards to requests for information on funded projects. Finally, a challenge that was identified is determining how developers should obtain credit when joining a software project later in its development; this issue is one that all three groups should address.*

4. *What actions or positions by professional societies (SIAM, ACM, APS, ASME, etc.) could best assist young CDS&E researchers in their career paths, e.g., by educating their superiors, helping recruit their students, or simply informing them about career-building opportunities?*

   Scribe: Jason McCormick

   The group identified numerous challenges currently facing young researchers. First, most of the high profile awards go to members of the community who are established and later in their careers. This does not always resonate with young researchers as the focus is often for work that has been done in the past and not on what is current and needed. Second, there needs to be a larger focus on career development. More workshops, panels, and conference sessions need to be focused toward graduate students and young researchers looking specifically at their needs. Third, due to the inability to involve young researchers in larger societal roles or the difficulty for a young researcher to establish themselves or get involved in these roles, a lack of continuity in leadership may arise as more established faculty retire. There needs to be more respect for young researchers' opinions and thoughts in regards to the direction of computational science. Fourth, there is a need for better mentorship through professional societies for young researchers. Fifth, professional societies need to promote a better understanding of the needs of the profession to young researchers and involve young researchers in developing these needs. Sixth, often, there are only one or two members in a given university department that are part of the computational science community. As a result, the professional societies need to provide a better link between those in the profession. They also need to promote the importance of multi-disciplinary work to help with tenure track faculty in the field since universities often do not know how to evaluate these researchers when it comes to tenure and promotion. Seventh, there is a question as to whether the computational science field is creating a post-doc purgatory due to the ratio of the number of Ph.D. students versus the number of open tenure track positions. Eighth, engineering students do not necessarily want to be involved in computational science. They often go into engineering to do more applied work. There is a lack of *understanding and promotion of the important role computational science plays in engineering. The problem may stem from how computational science is taught and a lack of understanding of the size of current research problems that are being faced and what is needed to address the size/scope of these problems. Students often have a difficult time transitioning from discrete to more complex problems.*

Numerous potential solutions were discussed by members of the breakout group. First, professional societies should provide more opportunities for inter-mixing of established faculty/researchers and young researchers through workshops and travel grants. Second, societies should establish a means of providing intentional mentoring between established faculty and leaders in the field and young researchers. The focus should be on developing young researchers into leaders so that they understand the types of service that are required of leaders and required for the field. Third, societies should better promote the big ideas and directions that are the focus of the computational science field. These need to be promoted to potential students to help with recruiting (e.g., to establish educational materials) and to academic administrations to assist them in evaluating tenure/promotion cases for young faculty in computational science. Fourth, societies should establish more young and mid-career awards and promote these researchers' accomplishments. Fifth, societies should push for curriculum changes that better meet stake holder needs in computational science/engineering (e.g., see the HPC university website which promotes a core knowledge base for computational science.) Sixth, societies should provide materials to students so that they have a better understanding of the process of going from their undergraduate degree to graduate school to their career (academic or industry.) There should be more intermediate opportunities for awards and better assurance of a career upon finishing (provide career advice.)

2. Visualization

In the area of Visualization, Dr. Tom DeFanti from the University of California, San Diego described a series of challenges in data, networks, and displays that he described as "great data, great networks, and great displays." Great data involves meeting challenges in big data that are impacting visualization today. These challenges include addressing performance bottlenecks in networking and storage. In the area of networking, the falling prices of network ports and increasing density has now made the costs of 10-Gb and 40-Gb network interfaces attractive for analyzing and visualizing datasets that are increasing in size over time with inexpensive and powerful high performance storage and networking. Near real-time data analytics and is becoming possible with the networking and storage technologies available today. He described advances and challenges in collecting and combining full 360-degree video panoramas from an array of video cameras. Dr. DeFanti described several systems under development (FIONA, a portable 4K display wall, and *CAVEcam*, and *Camelot*) that involve the collection and display of large volumes of high resolution image and video data. The questions that Dr. DeFanti presented to attendees for discussion focused challenges in the integration of visualization with data and computation, consideration of performing visualizations at exascale, and the problem of developing GPU applications for computation and visualization.

Figure 4: Dr. DeFanti's giving his keynote talk on Scientific Visualization which focused on his three themes of "great data, great networks, and great displays".

Dr. DeFanti posed the following question to workshop attendees:

1. *How can visualization be better integrated with data and computation?*

   Scribe: Onkar Sahni

   There were several points of discussion that arose when the group considered the question of the closer integration of visualization with data and computation. The first noted that datasets arising from science domain application areas often have a "spatial-temporal" form. Analysis of these data typically involve two steps: 1) the collection and analysis of data; and 2) visualization of the results. The group thought that a tighter integration of these steps with a direct link of visualization capabilities with the data source would be useful for the community. A possible approach to address this is to provide direct user interaction or feedback mechanisms within the visualization tool (e.g. computational steering) to aid the user. There was also discussion about "offline" vs. "online" visualization. With offline, data is delivered to the visualization system after computation. An online approach, which would be *more useful, would directly connect the live computation with the visualization system to facilitate user visualization of results and direct user control of the simulation or analysis. The group also felt that a tighter integration would also be helpful to support educational efforts in the domain sciences to support student learning.*

The group also discussed the need for "plug-and-play" tools to simplify the process of discovering and using visualization tools. Also, a catalog of publically available open source tools and libraries that can support integration between computation, data, and visualization would be helpful. Identification of the immediate pressing needs of the domain science community might be helpful to inform the efforts of visualization researchers and practitioners.

Finally, the visualization of high-dimensional data remains an unsolved and open question. A few success stories in this area would benefit the community.

In summary, to support the tighter integration of computation, data, and visualization, an approach suggested by the group is to create an abstraction of the visualization needs of the community that could inform development efforts to provide facilities within visualization systems to allow "on-the-fly" use of the system coupled with data and computation. Moreover, a set of reusable software libraries and examples of successful use of visualization would help to encourage adoption of visualization in research and education. Simpler "plug-and-play" tools would also help the adoption process. Open questions discussion the group were: how to discover an underlying theoretical model from extensive instrumental data; developing a way to provide visualization capabilities for "intermediate need" users with routine visualization needs; and the need for new ways of managing high-dimensional data (examples included uncertainly quantification).

2. *What are some of the important considerations for doing visualization at the exascale?*

   Scribe: Ioan Raicu

   The group discussed two scenarios in considering this question. The first reflects the needs today. Moving data from computation to visualization is a difficult problem, and the achievable bandwidth from persistent storage systems and network transport remains a problem. This situation forces researchers to keep their data locally on the computational platform, but makes is more difficult to perform computational steering. Moreover, since HPC systems today are primarily batch scheduled, there is little or no support for interactive computational steering.

The other scenario reflects needs in the future. There will be a need for more resilient storage technologies to avoid the high costs of data replication. Erasure codes were mentioned as one possible approach to address this problem. The other aspect that will affect the future is reaching limits in memory size needed to support in-situ analysis on the computational system. One possible approach to address this is the use of non-volatile high speed storage devices (such as SSDs) to facilitate visualization co-located with the computing and data.

3. *How should one go about developing GPU codes that make use of the GPU(s) for both?*

   Scribe: Kamesh Madduri

   The group discussion on how to develop applications to use GPU for both computation and visualization centered on two topics. In the area of architecture, there are some potential limitations of GPUs that may affect their suitability for both computation and visualization. The first is an open question of whether a GPU program would be more error-prone than CPUs, and potential concurrency bugs. The second is limited I/O bandwidth between GPU and CPU memory along with a limited amount of available GPU memory. Finally, another open question is if and when GPU and CPU architectures will converge or merge and what are the impacts of a merger on programming models. On one high end system, Blue Waters at NCSA, GPUs are used today for both computational and visualization tasks. The second area, programming, focuses on issues involve in programming models and use of GPUs. It is clear that GP-GPUs are useful for well-structured computational problems, such as dense linear algebra. There are, however, several open problems with GPUs in the area of programming. One issue limiting the adoption of GPUs is that programming GPUs for visualization tasks requires a set of low-level instructions, analogous to assembly language programming. Another issue is the need for double precision floating point number support for scientific applications, which are available only on higher-end GPUs. Additionally, the inherent computation required for visualization tasks such as near real-time results from rendering and ray tracing may themselves be computationally intensive, thereby reducing resources available for general computation.

4. *What are the major bottlenecks/challenges in the visualization field?*

   Scribe: Kamesh Madduri

   The group discussed several challenges in the visualization field which were centered on the way in which the research agenda is formulated in the field, the community of *researchers involved in visualization, and applications driving challenges in the field of visualization. In the first area, the group discussed a problem in which new visualization technologies seem to be developed in an "ad-hoc" manner for singular specific projects. There seems to be a need for a more comprehensive approach to articulate a research agenda for the visualization field that would be based on cross-cutting visualization challenges. Along these lines, another issue discussed involved the composition of the people working in visualization. Specifically, could visualization challenges be successfully addressed by teams of artists, computer scientists, engineers, and mathematicians? Finally, the group discussed the types of applications that are driving challenges in visualization. The types discussed involve the representation of numerical data and the need for visualization for large and complex data sets, as well as visualization in evidence-based medicine.*

In summary, for both questions 3 and 4, the group concluded it is possible to use GPU clusters for computationally-intensive tasks and visualization. Also, the programming efficiency (ease of programming), performance, and I/O bottlenecks present today in GPU technologies need to be addressed through the design of new algorithms. Finally, research challenges driving visualization are coming from the need to visualize novel data. An example of this involves genomics and personalized medicine.

3. High Performance Computing

Dr. Ed Seidel discussed the transformation of scientific computing and the effects of this transformation on science. High Performance Computing (HPC) has become an essential element of the ecosystem sustaining research today. He described an example from astrophysics focused on understanding black hole collision, and the Einstein open source astrophysics toolkit that is used to facilitate the collaboration of distributed communities of interest in a specific area of research. He described a growing need for this kind of high performance computing resources and software environment for many science and engineering disciplines (both in industry and academia) beyond physics. Dr. Seidel described some of the research projects in biology, astronomy, and chemistry that are using the Blue Waters system at NCSA. As a result of the increasing availability of big datasets from new instruments, such as the Sloan Digital Sky Survey and Large Synoptic Survey Telescope, there is a growing need to connect big data with large scale computational resources to serve these communities.

Dr. Seidel encouraged workshop attendees to seek to organize and communicate their needs to the National Science Foundation. Dr. Seidel mentioned the needs of the vast majority of the community who rely on data sharing as a form of communication who do not have specialized "big data" needs. Open data sharing will help to speed innovation in the future – the materials genome project is one example of this emerging area. The grand challenges the community will need to address involve the integration of computing, big data, networks, cloud, and instruments. An example of this is the integration of XSEDE with Blue Waters and high speed national networks to support projects such as LSST and NEON. The National Data Service Consortium (NDS) is an effort to provide a data service that links publishers, universities, supercomputing centers, and data-oriented research projects. Dr. Seidel described several ways in which universities and supercomputer centers are working to respond to these challenges.

Figure 5: Workshop Co-Chair, Dr. Shontz talking with Dr. Seidel after his presentation.

The questions posed to the attendees by Dr. Seidel were:

1. *What are the new application demands in emerging data intensive high end computing and high performance computing?*

   Scribe: Linwei Wang

   The group discussed the continuing need for new approaches that can automatically select the specific parts of the data to be analyzed and the element of the computation needed for analysis. There is a need for the close integration of data management, simulation, and visualization for analysis for domain science problems. The need to save software along with curated data was discussed, as well as the architectural issues with scaling when using a combination of GPUs and CPUs.

2. *What are the most important research problems in creating the next generation of cyberinfrastructure?*

    Scribe: Hamed Hatami-Marbini

    The group described several broad areas with important research problems that could benefit from the use of cyberinfrastructure. The first area was disaster management that integrates social networking to understand the behavior of populations during a disaster. The second was the need for city and state level traffic simulation based on modeling the behavior of drivers. Other areas included quantum chemistry (linking experimentalists with computational scientists), life sciences (data analytics), and civil and mechanical engineering (modeling fluid flow in a porous media and computational fluid dynamics).

    In summary, many fields could benefit from the use of high performance computing and cyberinfrastructure, especially in the analysis of large datasets generated by many science domains today. There is an unmet need for an increase in user-friendly cyberinfrastructure platforms and additional training to help researchers learn to use these platforms. There is a need for improved algorithms to better use platforms for faster computational simulations.

3. *What are the major challenges to scale major HPC applications/algorithms to millions and billions of cores and possible ways to overcome some of these challenges?*

    How do we create interdisciplinary teams?

    Scribe: Tim Mueller

    To address the question of major scaling challenges, the group discussed three major challenges impacting scaling: resilience, complexity, and efficiency. In the area of resilience, given the scale of the largest systems and the relatively low reliability of these systems, scalable algorithms that can tolerate failures are needed. More testing is needed to validate and develop these new algorithmic approaches. In the area of complexity and efficiency, sensitivity to system architecture (such as memory bottlenecks and heterogeneous platforms) is needed to determine where to run applications and to tune these applications for the environments where they operate.

To address the second question of the development of interdisciplinary teams, the group discussed the need for NSF support for "grand challenge funding", the improvement of usability (user efficiency), and the development of partner teams of computer scientists with experimental scientists to develop new techniques and technologies that could use new hardware.

4. Education

The fourth thematic area of the workshop focused on *education.* Dr. Sushil Prasad described the IEEE Technical Committee on Parallel Processing (TCPP) curriculum initiative. The goal of this initiative is to revise the core curriculum in computer science at the undergraduate level to include teaching on parallel and distributed computing in existing courses. Curriculum revision is needed due to a change in the computing landscape. Since multi-cores are mass marketed and general purpose GPUs are in laptops and handheld devices, everyone needs to learn about parallel and distributed computing and HPC, not just graduate students or those specializing in HPC. The curriculum revision will be useful for students, educators, universities, and industry.

TCPP has proposed a core curriculum for computer science and computer engineering graduates with the goal that every individual computer science and computer engineering graduate must be at the proposed level in parallel and distributed computing by the time he/she graduates as a result of his/her required coursework. Existing courses, such as CS1, CS2, systems, and data structures and algorithms were identified as courses in which parallel and distributed computing topics from architecture, programming, algorithms, and cross-cutting ideas could be introduced. The TCPP curriculum has been adopted and is being evaluated by early adopters in the United States and abroad.

Another outcome of the TCPP initiative is the formation of the Center for Parallel and Distributed Computing Curriculum Development and Educational Resources (CDER). The center has put together pedagogical and instructional materials for teaching parallel and distributed computing topics. In addition, CDER's efforts include a textbook project on how to integrate parallel and distributed computing into core courses at the undergraduate level. NSF CAREER Awardees are encouraged to use the curriculum in their classes as relevant and to contribute to the CDER book project.

Dr. Prasad also provided career advice to NSF CAREER Awardees. His advice included several techniques for integrating research with teaching. He also emphasized the importance of quality teaching and research and of gaining visibility and recognition through IEEE senior memberships and service on both technical program committees and editorial boards.

Dr. Prasad posed the following questions:

1. *What is the role of a teacher-scholar? How does one integrate research with undergraduate education?*

   Scribe: Thomas Wies

   Best practices for involving undergraduate students in research were identified. It was noted that undergraduate students should be involved in research early on in their studies, i.e., before they become seniors so that there is enough time for them to do research. It is possible to identify suitable students for research projects when they are still in high school (e.g., through university level filtering systems, such as Fulbright scholarship.) Freshmen and sophomore students can be involved in small projects first to teach the basics at the beginning. They can be involved in larger projects in their third and fourth years.

   There are several ways to get students involved in research. Students need to be exposed to the tools used in research; such discussion can be incorporated into the classroom. In addition, introductory courses should be used to present big ideas of the discipline that help connect the dots for non-majors in the classes. This gives students an incentive to continue with advanced classes in the major and encourages a broader, cross-disciplinary transfer of knowledge. Programming exercises can be used to excite students about research questions. Cloud providers, such as Amazon, can be engaged so that students can do HPC research involving medium-sized experiments. In addition, graduate students can be involved in mentoring undergraduate students; this can be a rewarding experience for graduate students and increases their chance of finding a job later.

   Undergraduate students can be paid for their research, e.g., through the use of NSF REU supplements to existing grants. Students can also become co-authors on publications and can attend conferences to present their research. In addition, this improves their resumes and makes them more competitive in graduate school applications.

2. *How important is it for faculty at research universities to innovate in education? Should they publish on new courses they've developed, new models of teaching, curricular guidance, etc.?*

   Scribe: Bill Xu

It is very important for faculty at research universities to innovate in education, and faculty should be encouraged to publish such innovations. However, junior faculty, senior faculty, university administrators, and the NSF all have different perspectives on this issue.

Junior faculty are often concerned that putting too much effort into teaching innovation might jeopardize their tenure cases, as promotion and tenure criteria emphasize research. Allocation of sufficient time for these activities is thus inherently risky. However, junior faculty realize that accomplishments from teaching, including training and recruiting good graduate and undergraduate students are important. Publishing in educational journals is important, but typically are not viewed the same as important research publications. Teaching innovation can help junior faculty add credentials when seeking NSF funding.

One efficient and effective strategy (or best practice) to implement innovative teaching in classroom is to partner with experts in areas such as cognitive science and education. The experts can be used to formulate the educational hypothesis, develop the contrast group, and determine the experimental design. The class instructor implements the experiment using the classroom as the testbed. This reduces the amount of effort required by faculty member. The scientific data can lead to publication in an educational journal or conference without cost too much time from the faculty member.

Senior faculty are in the enviable position of being free to explore new pedagogical methods. A fundamental question is if an innovative teaching method is also an effective teaching method (i.e., chalk board versus multimedia).

University administrators sometimes advocate for innovative teaching models and use them as selling points to secure funding from state and other agencies. Faculty are then asked to implement the innovative teaching methods. There can be a disconnection between the ideology and implementation as faculty might not be completely informed or motivated to deliver the new educational method.

The NSF promotes innovative education, as it is part of its mission within the area of education and workforce training. NSF encourages faculty to incorporate the most recent innovation into course syllabi and to train the new generation of work force.

3. *How should one attract students into HPC, parallel/distributed computing, and computational science courses and research?*

Scribe: Zizhong (Jeffrey) Chen

There are several ways in which one can attract students into HPC, parallel/distributed computing, and computational sciences courses and research. In regards to courses, faculty can make flyers advertising such courses to be sent to prospective students. Case studies can be used in class in order to teach students HPC at the undergraduate level. Another idea is to explain to students in class that computers are typically not just a single core today; this should help interest them in HPC research projects. In regards to recruitment of undergrad researchers, it is helpful to e-mail prospective students information on research projects. Recruiting students who are in a 5-year B.S./M.S. program is a particularly good idea, as they might have studied HPC during their fifth year and may also have more time for research. Undergraduate students working on research projects can be paid with start-up funds or REU funding. Interested students who are working on research can be encouraged to attend a local SIAM chapter meeting or an HPC conference. Initially, they could attendance the conference, whereas later on, they could give a presentation on their research.

4. *Can training students at the interface of distinct disciplines result in jack of all trades but master of none?*

Scribe: Adrian Feiguin

Training students at the interface of multiple disciplines can result in them being the jack of all trades but master of none; however, this is not necessarily the case. Learning multiple skills and conducting interdisciplinary research could be a valuable asset for students. However, this depends on the context and has to be done with guidance and has to be done wisely. For instance, research groups tend to hire specialists or "technicians" to aid in the computational aspects of a project. In some cases the individual does not necessarily have to understand the science. For instance, an undergraduate student could carry out a repetitive process streamlined in a very well established pipeline ("run scripts"), with little training. But this is not necessarily interdisciplinary research or experience. It could turn out to be a valuable experience for an undergraduate student, but definitely not for a graduate student. Students should learn and understand the science. At the same time, students can learn a lot of science, yet not gain practical skill and "not know how to do anything". Practical skills and training are necessary. It is important for a student to both master all trades but be the master of one trade; students need to have an area of specialization. This is especially true for students interested in a career in academia.

5. Grand Challenges in Cyberinfrastructure and Interdisciplinary Research

Dr. Thom Dunning presented on "Grand Challenges in Simulation Science and Computing". In this presentation, he described the computer as one of the most powerful instruments today for science. Computational simulation and data science is essential today to advance discovery in a many areas of science and engineering. He described examples in Astronomy and Chemistry that represent new needs in simulation and data science. Dr. Dunning described the cycle of research that includes theory, models, algorithms, simulation software, validation, and finally all of these feeding back into changes in theory. He described how computational simulation accuracy and fidelity increased from the 1970's until today due to advances in theory coupled with new computational techniques and increasing computing power. These new capabilities allowed chemists to increase the complexity of the models they could simulate. This has led to the ability of computational simulations to produce results as accurate as results obtained from physical experiments in a laboratory. On a philosophical level, Dr. Dunning described how the changes in research capabilities brought about from advances in computational simulation is leading to a new "Age of Constructivism," in which knowledge of theory and how systems work can be used to build up usable simulation models than can simulate real-world physical systems. He described three examples of where this has led to advances in real-world applications: internal combustion engines, climate models, and aircraft design. Dr. Dunning then described several grand challenges in computing technology that impact the capabilities of computing platforms used for simulation. These challenges include heat dissipation in computer chips, limits in single core performance and the increasing number of cores in systems, non-uniform memory access, and the growing complexity of high performance computing system in both scale and the use of novel components (such as GPUs).

The questions he posed to the group were:

1. *How can we best foster multidisciplinary and interdisciplinary research and education in our universities?*

Merged with Question 2

2. *What is the basic knowledge needed by future generations of computational and data scientists and engineers and how can we best deliver this knowledge?*

Scribe: Kamesh Madduri

The group discussed these two questions that were posed by Dr. Dunning. Questions three and four were not addressed by workshop attendees. The first area of discussion centered on articulating the needs and motivation for increasing education efforts and to best foster interdisciplinary efforts. There is a clear need for the next cohort of scientists and engineers to be trained to manage and exploit sensor data becoming available today from infrastructure and manufacturing. Although the curriculum for science and engineering continues to evolve and change, room needs to be made for at least a minimum amount of training in core data science as well as computational science. Both engineers as well as their management need to understand the use and impact of computational simulation. In terms of the approach to develop curriculum and content, several approaches were discussed. First, a possible approach is to provide introductory interdisciplinary survey courses. Additionally, adequate education and training is needed to ensure competence in methods rather than knowledge. An example of this mentioned in the group was how the educational needs of students in CDS&E in relational databases (i.e., creating, populating, and using DBMS system) could not be adequately with database courses currently taught in computer science departments.. Another need discussed by the group is for a shorter course sequence to provide additional education for practitioners working in the field today. Two possible approaches to enhance education discussed by the group were to promote research internships at universities and national laboratories, and to create undergraduate summer internships with domain science faculty with student from applied mathematics and computer science.

Figure 6: Dr. Thom Dunning giving his keynote talk on Grand Challenges in Cyberinfrastructure and Interdisciplinary Research.

The barriers that impact these efforts were discussed by the group. On an institutional level, there was concern about institutional inertia to change at the departments, faculties, and universities. A specific example is the difficulties new faculty encounter when they seek to engage in interdisciplinary work within their institution. These faculty need strong support from the senior colleagues and department heads to engage in this type of research. Exemplar institutions that support interdisciplinary research were discussed in the group: UC Merced and KAUST were mentioned. Barriers to change also come from students. The first discussed by the group was that students were very savvy as computer users, but were not comfortable with programming and software development. The group also *discussed the level of discomfort many computer science students have with mathematics, and their preferences to deal with pseudocode.*

3. *What is needed to enable computational modeling and simulation and data-intensive science to take full advantage of the current and next generation of computing technologies?*

4. *How can we reform the merit review system to ensure fairness in the funding of multidisciplinary and/or interdisciplinary proposals and in the hiring and promotion of faculty involved in these projects, especially young faculty?*

These two questions were not addressed by attendees.

Figure 7: NSF CAREER Awardees network during a break

4.0 Invited Speakers and Panelists from the National Science Foundation

Dr. Farnam Jahanian discussed numerous opportunities for computational and data-intensive research related to cyberinfrastructure within the ACI in the Directorate of Computer and Information Science and Engineering at the NSF. In particular, he indicated that health-care, well-being, security, transportation, energy, among others, were areas of national importance that connected to cyberinfrastructure research. Dr. Jahanian also identified five areas for investment by the NSF and the nation including data-driven discovery, expanding the limits of computation, melding of the cyber and physical world, cybersecurity, and universal connectivity. The NSF is also interested in increasing the number of women and minorities in computer science and the STEM fields, in general. However, the issue is one of inclusion as opposed to diversity, as the United States is diverse.

Ms. Irene Qualters described the importance of multidisciplinary research in order to advance in science and to create the right kind of cyberinfrastructure to position the United States for the future. She described several multidisciplinary research programs within ACI including the Cyberinfrastructure Framework for the 21st Century Initiative (CIF 21), the Computational and Data-enabled Science and Engineering program, and the Exploiting Parallelism and Scalability program. ACI also supports research in networking and security, computational infrastructure, and workforce development.

Figure 8: NSF CAREER Awardees network with Ms. Irene Qualters after her invited talk.

Also of importance to ACI is to increase in the number of women and minorities participating in cyberinfrastructure research; this is needed for increased innovation as a nation and broadening of participation. Another challenge facing the cyberinfrastructure community is the computing crisis brought about by demand that has outstripped computational resources; community discussion is needed on this.

Figure 9. An NSF CAREER Awardee networking with Dr. Almadena Chtchelkanova

Dr. Almadena Chtchelkanova held a question and answer session along with a discussion with the group focused on providing information for interacting with program directors and learning about funding opportunities, including supplemental funding programs at the NSF.

Dr. A.J. Meir described issues of importance in the Division of Mathematical Sciences (DMS) at the NSF including reproducibility of science, replicability of the results obtained from software, community-building, and sharing of data and other resources. He also discussed the DMS viewpoint on computational mathematics and research in this area. He indicated that computational mathematics includes floating-point arithmetic, integer calculations, and symbolic computing. Research in this area includes the development of new algorithms, mathematical analysis of the algorithms, and new mathematics driving the algorithms.

Dr. Nigel Sharp discussed some of the pressing problems in astronomy and astrophysics that could benefit from the involvement of researchers with expertise in cyberinfrastructure and high performance computing. These problems include creating large-scale simulations of objects within the universe, multi-scale simulation of supernovae, and quantum gravity in astrophysics. He also described issues involved in archiving data, extracting knowledge from real-time data, and automated analytics for astronomy data. A final topic he addressed was the need to involve human based analysis based on volunteers to review datasets. One example of this that he described is the Galaxy Zoo project.

Dr. Rudi Eigenmann discussed efforts within the NSF to support computational and data-enabled science and engineering with high performance computing systems such as Blue Waters, XSEDE, and campus-level supercomputing systems. He described ways in which attendees could get involved in cyberinfrastructure efforts both on the application and systems side to help increase the productivity of the domain scientists and to help them solve their problems. He encouraged attendees to consider the research issues involved in developing the next generation of cyberinfrastructure.

Dr. Daniel Katz described a vision within ACI focused on "Software as Infrastructure", in which NSF will seek to promote software as infrastructure for the research community. As a part of this effort, NSF will nurture a software "ecosystem" that will bring together computer scientists and domain scientists to provide new capabilities to facilitate transformative forms of research. He described ACI software cluster programs (Exploiting Parallelism and Scalability (XPS), Computational and Data-Enabled Science & Engineering (CDS&E), and Software Infrastructure for Sustained Innovation (SI2). Dr. Katz described a new ACI-SBE Dear Colleague Letter (NSF 14-059) that seeks to encourage the development of new models for establishing metrics and ways for establishing credit for novel forms of data and software sharing. He also discussed the issue of career paths of non-tenure track researchers in the near term, and issues involved in supporting the reproducibility of simulations and data analytics in science.

Program Director Panel Breakout Notes

1. What is computational mathematics? What is computational mathematics research?

Scribe: Onkar Sahni

The group discussed this question, and decided that computational mathematics involves finding new efficient algorithms along with use of mathematical approaches and tools focused on solving forward problems and inverse problems as well as optimization problems. These approaches include algorithms and solution methods suitable for current as well as emerging computer architectures.

2. How can the NSF help to nurture and build the community of cyberinfrastructure researchers?

Scribe: Christopher Rozell

The approach discussed by the group focused on community and mechanisms to help nurture and build the community of cyberinfrastructure researchers. In the area of community, approaches discussed included focused workshops and conferences similar to the XSEDE conference that would be focused on interdisciplinary research; the development of meaningful "metrics for success" that could be used for academic review; the development of software tracks within established domain science conferences; increased access to national science high performance computing resources; and increased industry involvement to spur the development over a shorted timeline of impactful projects. In the area of mechanisms, the topics discussed were: increased program funding and more frequent solicitations as well as increased cross-program interdisciplinary funding opportunities.

Figure 10. Workshop Co-Chair Dr. Thomas Hacker participates in a breakout session led by Dr. Tom DeFanti

3. What are the most important research problems in creating the next generation of cyberinfrastructure?

Scribes: Tim Mueller and Thomas Hacker

The group discussed the need for a new approach for systems level design in several areas: fundamental research in workflows to automate complex tasks that use a distributed cyberinfrastructure; addressing the growing complexity of cyberinfrastructure systems to ensure that the systems are easy to use, reliable, and maintainable; developing new approaches to motivate researchers to develop new software tools and other materials needed to improve the usability and adoption of cyberinfrastructure systems.

The group identified some of the major challenges that would need to be addressed. The first is the need for an infrastructure that is designed to be user-friendly built on some of the design *principles for usability that can be found today in commercial systems. The group noted that an abundance of resources are available on a campus level that are suitable for embarrassingly parallel work. Another need is for workflow systems that could link major systems and automate tasks that could be accessed through a unified programming paradigm, such as Swift or a Pi-calculus based language. Another major area of discussion was in economics and system design, specifically, the question of how cyberinfrastructure could be designed from the start to not be excessively complex and difficult to learn and adopt. The group noted that good system design needs to be considered at the start of the cyberinfrastructure project to ensure that usability and user efficiency is a central element and philosophy of the CI system.*

5.0 Attendee Feedback Survey

To collect attendee feedback, we conducted an anonymous Qualtrics survey that was emailed to workshop attendees at the end of the workshop. The purpose of this survey was to collect feedback from attendees about the workshop, and to ask for suggestions for improvement for the workshop. We received 20 responses to the survey (see the Appendix A for the complete text of the survey and responses).

Table 1. Participant responses to the workshop survey. Summary responses are reported as the mean and standard deviation (S.D.) of individual responses. Survey participants answered questions on a Likert Scale with the following numeric assignment. Strongly Disagree (SD = 1), Disagree (D = 2), Neither Agree nor Disagree (AD = 3), Agree (A = 4), and Strongly Agree (SA = 5).

Question	Mean	S.D.
The five focus areas of the workshop (Grand Challenges, Data, Visualization, Education, and High Performance Computing) included my area of research and education. If not, please provide a comment describing a new broad area you would like to see covered in a follow-on workshop.	4.0	1.1
The disciplinary areas of workshop attendees were sufficiently broad to facilitate interdisciplinary engagement.	4.1	1.2
The workshop format (keynote talks followed by discussion) was useful and engaging.	3.9	1.2
The talks were relative, informative, and interesting.	4.0	1.1
The poster session was useful and engaging.	3.7	1.3
There were sufficient opportunities for networking and collaboration.	4.0	1.1
The hotel accommodations, meeting space, and meals were adequate.	4.5	0.9
The workshop was helpful in learning more about the NSF and available funding opportunities at NSF.	4.4	0.9
The workshop should include CAREER awardees beyond ACI.	3.8	1.0

The workshop should include attendees from outside the NSF CAREER program.	3.5	0.7

As in prior years, participants generally agreed that the thematic areas of the workshop included their areas of research and education, and that the disciplinary areas of the various attendees were broad enough to facilitate engagement. Participants strongly agreed that the workshop was helpful in learning more about the funding opportunities available at NSF.

Participants agreed that the workshop format was useful, talks were interesting, and that there were opportunities for networking and collaboration, but several participants felt that these could be improved upon. One participant suggested shortening the discussion time to allow more time for networking, and two attendees suggested including introductions or lightning talks to facilitate networking.

Compared with last year, attendees agreed less strongly that the poster session was useful and engaging. While most attendees thought the poster session was useful, a significant fraction felt that the format of the poster session (all attendees simultaneously presenting posters) actually inhibited networking, as it reduced traffic to individual posters. Compared with last year, participants had more concerns about the poster session.

As in previous years, participants agreed that the workshop should include CAREER awardees beyond ACI, but with limited agreement that the workshop should include attendees outside the CAREER program. Participants also agreed that the length and size of the workshop were about right, and that the workshop should be held annually.

For open response questions, in terms of new broad areas of topics attendees would be interested in for future workshops, there was interest in increasing attendance of CAREER PIs from application areas (such as biology and physics), and a desire for information on funding options outside of NSF. The other open question related to suggested changes or improvements. Attendee suggestions including eliminating the poster session, assigning people and scribes within each breakout group, and increasing social time. The responses to each open response question are in Appendix A.

Compared with last year, survey results for this year showed that participants reported significantly more opportunities to learn about the NSF, with 11 out of 20 respondents strongly agreeing that they found the workshop helpful in learning more about available funding opportunities.

5.1 Comparing survey results for the 2012, 2013, and 2014 NSF CyberBridges workshops

To provide a basis for comparing survey results with prior years, Figure 12 shows a comparison of responses for the past three years of the workshop based on a Likert scale with the assignment Strongly Disagree (SD = 1), Disagree (D = 2), Neither Agree nor Disagree (AD = 3), Agree (A = 4), and Strongly Agree (SA = 5).

The lightest bar at the top of each triplet for each question is the response to the 2014 survey, the middle bar is for the 2013 survey, and the bottom dark bar represents the 2012 survey. The computed standard deviation for each year is shown as an error bar for each result.

The first observation from this figure is that the results are fairly consistent across all three years. There is no statistically significant difference among the questions for each year. The upward trends that can be seen from this figure are: (1) increasing interest in including attendees from outside the NSF CAREER program; (2) increased preference for the meeting space in 2012 and 2014; and (3) increased perceived helpfulness in learning about NSF opportunities. These results indicate that in coming years, it may worthwhile considering adding attendees from outside the NSF CAREER program somehow. A possible approach to increase the number of attendees from outside the NSF CAREER program is to invite senior leaders in each of the fields with the goal of increasing interactions between new faculty and senior leaders, as well as to increase interest in senior leaders in attending and participating in the workshop.

The declining trends are: (1) sufficient opportunities for networking and collaboration; (2) satisfaction with the poster session; (3) workshop format; (4) breadth of disciplinary areas in the workshop; and (4) focus areas of the workshop included attendees' areas of research and education. The decreasing satisfaction with the poster session and networking opportunities could possibly be addressed in future workshops by eliminating the poster session and adding more time for discussion. We have considered lightning talks by attendees, but the overall time that would be needed (over four hours) would likely detract from the overall workshop, and reduce time for networking and interaction. Future workshops should consider changing the focus area of the workshop, or adding new additional areas (with an invited speaker who is a leader in the area) to increase the breadth of the focus areas and to possibly replace the time spent in the poster session.

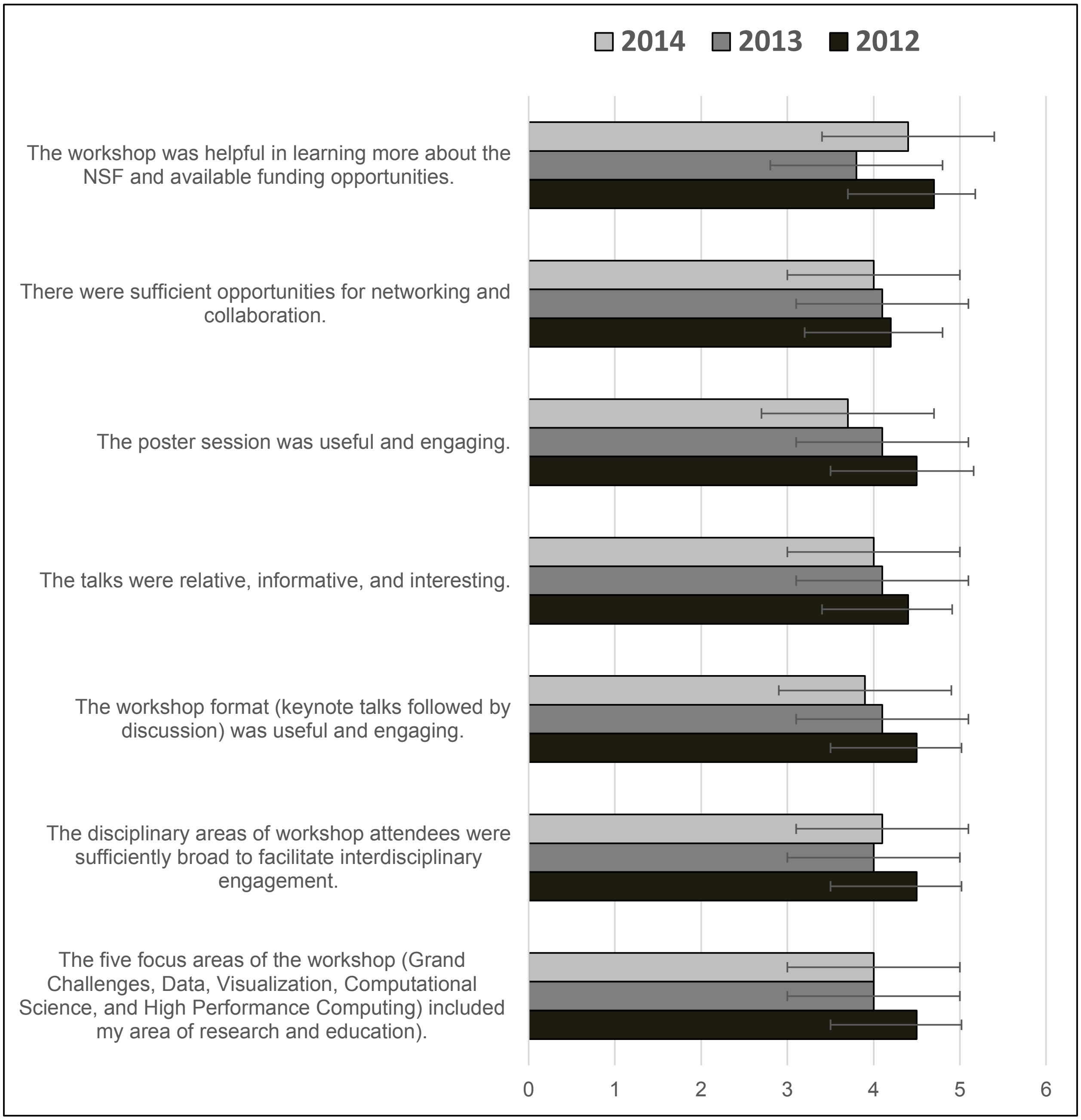


Figure 12. Comparison of survey responses from 2012, 2013, and 2014 CyberBridges workshops.

6.0 Observations from the Workshop

Four broad themes emerged from the workshop presentations and discussion.

The first, **moving to exascale**, involved the challenges that will need to be met by the community as new exascale systems become available in the coming decade. Exascale systems will require new algorithmic approaches to address fundamental problems in working across multiple problem scales that are also able to adapt to the capabilities and inherent limitations (such as power) of exascale systems. New approaches are needed to tolerate component failures and as well as automatically adapt to the architectural characteristics and bottlenecks of these systems. New algorithmic approaches will need to be addressed through joint community efforts based on mutually beneficial partnerships between computing scientists and experimental (domain) scientists to develop the new technologies and approaches needed to fully exploit the capabilities of exascale systems.

The second theme, **sustaining the cyberinfrastructure ecosystem**, focused on the challenges and efforts needed to continue to develop and promote the adoption of cyberinfrastructure in support of discovery in science and engineering. Cyberinfrastructure is a core foundation that is powering computational and data-enabled science and engineering, and continued growth in the capabilities of cyberinfrastructure are necessary to power new discoveries in science. There are technological challenges that need to be addressed to continue this growth, including the management of power and heat, current barriers limiting the growth of single core performance, the associated proliferation of cores within a system, and the increasing complexity of high performance computing systems that seek to accommodate these constraints as well as to integrate new architectural elements (such as GPUs). The NSF Program Director panel discussed several themes related to efforts in the use of and research in cyberinfrastructure. Several speakers mentioned that a multidisciplinary approach is needed for research and development in cyberinfrastructure, as well as the need to increase diversity in the community. Other specific challenges discussed during the workshop included the need to improve usability of cyberinfrastructure platforms, and the need for additional training on the use of these systems.

The research agenda for cyberinfrastructure science for the next generation of cyberinfrastructure will need to address several gaps: the need for better workflow systems that can automate complex tasks across a distributed cyberinfrastructure; attention is needed on the problem of improving usability - new approaches are needed to motivate the research community to address the usability of cyberinfrastructure.

The third theme involves **issues related to faculty life and interdisciplinary research**.

Much of the discussion during the workshop related to issues affecting early career faculty and their efforts in interdisciplinary research. In terms of faculty life, attendees described the need for awareness and credit from senior colleagues for software and data contributions made by early career faculty, especially for promotion and tenure. The NSF may be able to increase awareness of the value of this through continued encouragement of software code and data sharing. To aid this effort, the need was discussed for more early- and mid-career awards to spotlight the achievements of early- and mid-career faculty.

Another issue involved the need for more and better mentoring for early career researchers from professional societies as well as from established leaders. Attendees noted the risk of a leadership gap within the community as the current generation of leaders retire. The final topic related to faculty life was the need for longer term grants (sufficiently long to support a Ph.D. dissertation project), and the need for support for interdisciplinary projects that involve "hard and risky problems" that require fundamental discovery in both computer science and domain sciences. One item of note was there was a limited awareness among early career faculty attendees of the availability of cross-agency NSF funding opportunities that can supplement existing grants.

Another broad area involved **education and training**. The growth in multicore architecture has accentuated the need for broad education and training in parallel computing for computer science students as well as for the broader community of scientists and engineers. In terms of innovative teaching, early career faculty are aware of the need for innovation in education. However, they are concerned that these efforts will not be valued as much a research by senior colleagues focused on research who will evaluate their promotion and tenure cases. Undergraduates could and should be involved in research early in their studies as a component of an innovative approach to education. Finally, attendees discussed the need for students to become the master of at least one discipline, but students also need to develop a broad set of skills to be functional in future interdisciplinary environments.

The final area of discussed within this theme involves **interdisciplinary research and education**. Early career faculty interesting in pursuing interdisciplinary research need to be sustained by a culture and an ecosystem that encourages and promotes interdisciplinary research. Some possible approaches were discussed that included the development of workshops and conferences focused on interdisciplinary research, the development of software focused tracks in domain science conferences, and the involvement of industry to promote interdisciplinary development within timeframes shorter than those typically seen in academic research and development.

The final broad theme of the workshop focused on the **growth in the influence and importance of data**. The increasing availability of large data sets from instruments, simulations, and sensors is driving the need to connect data resources and datasets with large scale computational resources and visualization systems. The grand challenges that will need to be addressed involve the integration of computing, visualization systems, networks, data, and instruments. In terms of visualization, one of the major challenges affecting visualization today is the need to overcome performance bottlenecks in storage systems and networking that limit the use and analysis of large volumes of data from storage systems. These bottlenecks will also impact the use of future exascale systems.

One area of discussions focused on the need for new algorithms and techniques to automatically select and match data to be analyzed with computational resources needed for analysis. Attendees discussed the many possibilities for the use of data oriented applications built on the analysis of large datasets, such as the analysis of social networking data to understand how population may react to a natural disaster. A corresponding need to curate and preserve data was discussed for long-term use.

The intersection of visualization with computation and data was another area of discussion. Tighter integration of visualization systems with storage and computational systems is needed to foster the development and to improve the usability of interactive applications (such as computational steering). Researchers today must manage visualization data separately from the visualization system on a computational platform. There is a need for more "plug and play" tools for the user community. GPUs are one potential element that could be used, although GPUs have inherent limitations that impede use in terms of architecture and programming models. The visualization of high-dimensional data is an open problem, and the need for the visualization and representation of numerical data for large and complex data set is one of the drivers of this problem. In terms of a research and development agenda for visualization, attendees noted that new visualization technologies are often developed as "ad-hoc" solutions for single projects. The need for a more comprehensive and better organized approach was identified with a team approach that should include artists, computer scientist, and domain researchers.

7.0 Lessons Learned from the Workshop

We have conducted the CyberBridges workshop for the past three years, and have identified several "lessons learned" to benefit future workshops. One major challenge was to manage the balance of time between speaker presentations and discussions. We found that by adding breaks and providing adequate time for discussion, the workshop could stay comfortably on schedule while still providing attendees adequate time for discussion. We also found that using a priority based scheme for inviting attendees worked well. Based on declining satisfaction with the poster session, a new approach is needed that will allow attendees to describe and discuss their work with others in a more interactive format. Finally, as Drs. Hacker and Shontz are nearing the of their CAREER awards, we are planning to transition leadership of the workshop to new CAREER awardees in the coming year.

Acknowledgements

Our efforts in planning and conducting the CyberBridges workshop and the development of this report were supported by Natasha Nikolaidis (Purdue), Linda White (Mississippi State), Clifton Abbott (Mississippi State), Dianna Janus (Mississippi State), Lynn Grimes (Purdue), Debbie Miethke (Purdue), Natalie Eaves (Mississippi State), Michelle Latham (Mississippi State), Cathy

Chandler (Mississippi State), and Marilda O'Bryant (Mississippi State). Dr. Linwei Wang from RIT provided helpful comments and editing for the report. This workshop was supported by National Science Foundation Award ACI-1430620.

Appendix A. Detailed Survey Results

The complete text and responses of the survey sent out to attendees are described below.

In the first section, survey participants answered questions on a Likert Scale with the following numeric assignment. Strongly Disagree (SD = 1), Disagree (D = 2), Neither Agree nor Disagree (AD = 3), Agree (A = 4), and Strongly Agree (SA = 5).

1. The five focus areas of the workshop (Grand Challenges, Data, Visualization, Education, and High Performance Computing) included my area of research and education.
   Results: Mean Value: 4.0. Responses: (1) SD, (2) D, (0) AD, (10) A, (7) SA

2. The disciplinary areas of workshop attendees were sufficiently broad to facilitate interdisciplinary engagement.
   Results: Mean Value: 4.1. Responses (1) SD, (2) D, (0) AD, (8) A, (9) SA

3. The workshop format (keynote talks followed by discussion) was useful and engaging.
   Results: Mean Value: 3.9. Responses (1) SD, (2) D, (3) AD, (7) A, (7) SA

4. The talks were relative, informative, and interesting.
   Results: Mean Value: 4.0. Responses (1) SD, (1) D, (3) AD, (7) A, (8) SA

5. The poster session was useful and engaging.
   Results: Mean Value: 3.7. Reponses (2) SD, (2) D, (2) AD, (8) A, (6) SA

6. There were sufficient opportunities for networking and collaboration.
   Results: Mean Value: 4.0. Responses (1) SD, (1) D, (2) AD, (10) A, (6) SA

7. The hotel accommodations, meeting space, and meals were adequate.
   Results: Mean Value: 4.5. Responses (1) SD, (0) D, (0) AD, (6) A, (13) SA

8. The workshop was helpful in learning more about the NSF and available funding opportunities at NSF.
   Results: Mean Value: 4.4. Responses (1) SD, (0) D, (0) AD, (8) A, (11) SA

9. The workshop should include CAREER awardees beyond ACI.
   Results: Mean Value: 3.8. Responses (1) SD, (0) D, (5) AD, (10) A, (4) SA

10. The workshop should include attendees from outside the NSF CAREER program.
    Results: Mean Value: 3.5. Responses (0) SD, (1) D, (9) AD, (9) A, (1) SA

Next, the survey asked participants to rate the following questions with a response from the options Too short/too few (1 = S); About right (2 = AR); and Too long/too many (3 = L).

11. How would you rate the following: Length of the workshop
    Results: Mean Value: 2.1 Responses (0) S, (19) AR, (1) L

12. How would you rate the following: Number of attendees
    Results: Mean Value: 1.9. Responses (2) S, (18) AR, (0) L

The next questions asked about the frequency and cost of the workshop, and about the respondent's source of CAREER funding.

13. Would you be interested in attending the workshop in the future without full travel reimbursement?
    Results: Yes 50%, No 50%

14. How frequently should the workshop be held?
    Twice a year: 5%, Annually 70%, Every other year 25%

15. Is a component of your CAREER award funded from ACI?
    Results: Yes 45%, No 50%, one non-answer

The final questions provided an open form to allow participants to provide written feedback:

1. Are there any new broad areas or topics that you would like to see covered in a follow on workshop?
    a. Alternative funding options from NSF, career advice for young investigators
    b. Complex systems specific implementation problems, and more applied presentation topics
    c. More specific research areas
    d. It would be useful to have more CAREER PIs from application areas attend –it would be nice to have more computational physicists, computational biologists, etc.
    e. Work out session/short course to teach a particular and common simulation tool
    f. Exascale-related issues (some of it was covered in this workshop)
    g. Business and economics
    h. Big data
    i. How cyberinfrastructure can be leveraged to develop new technologies
    j. Algorithmic aspects of data analysis. When people talk about "big data", they are often talking about how to move it/handle it/ but not the statistical foundations of

   how to extract information from it. This is a critical component of cyberinfrastucture that often gets shortchanged.
   k. I would like to see a panel which focused on other agencies (e.g., NIH and DoE) in addition to NSF.

2. What changes or improvements could we make to the workshop in the future?
   Responses:
   a. Discussions were too long – can get better engagement by shortening. Another option for getting better discussion engagement is to assign groups rather than self-assign
   b. More guided discussions; better questions (redundancy in questions)
   c. More interesting speakers
   d. Assign scribes and leaders for each breakout session, difficult to hear with four groups talking in one room, provide abstracts before soliciting questions from attendees
   e. Let people buy beer and wine at dinners and poster session. Have short 2-minute presentations about the posters by attending PIs. Have a round of introductions of everybody at the start of the workshop
   f. Mutual introduction of each other
   g. Send questions for breakout or discussions in advance
   h. Maybe focus on a chosen domain specific area every year?
   i. A little bit more time for social with other awardees
   j. Wireless network needs to work in a better way
   k. Eliminate the mandatory poster session. It actually made it harder to network, and the time and money spent preparing the poster could have been better spent on many other things
   l. Poster session was strange in that essentially all of the audience was essentially presenting posters themselves. There just weren't enough people looking at posters to make it worth the time and trouble to make one.
   m. It's distracting when NSF CAREER participants don't participate in the breakout session and leave the room and talk with each other or leave the workshop when we're not on break. There needs to be a mechanism for making the NSF CAREER participants participate in the breakout session or forfeit their reimbursements.

3. Please provide suggestions for keynote speakers for follow-on workshops. [Suzanne]
   a. A method in which follow-up questions could be asked (not specific to speakers but to organizers)
   b. Mike Norman (SDSC), Tom Sterling, somebody from DOE
   c. Geoffrey Fox

d. Bring in experts in computer systems and big data areas
e. Gabrielle Allen, Qiang Du, Juan Meza, and ask this year's keynote speakers who they would recommend.

4. Please provide suggestions for community building activities outside of the workshop.
   a. Social Activities, start-up seed funds to have collaborative basic research/visits (collaborative work between attendees)
   b. It would be nice to have a mailing list or online forum for this kind of thing.
   c. summer institute on computation modeling, HPC, etc.
   d. CAREER Awardee/Cyberbridge participant meetings/dinner at SC conference? And or other meetings?
   e. workshops held adjunct with large conferences such as IPDPS
   f. Webinar where one NSF CAREER participant discusses his/her research to the group; write a white paper/position paper on topics of relevance to the group (e.g., open-source software, interdisciplinary solicitations) and present it to NSF.

Appendix B. Speaker and Attendee Biographies and Photos

Biographies were current as of June 2014, at the time of the workshop

Keynote Speakers and Invited Speakers from the National Science Foundation

Farnam Jahanian, Division Director, Directorate for Computer Science and Engineering, National Science Foundation

Farnam Jahanian leads the National Science Foundation Directorate for Computer and Information Science and Engineering (CISE). He guides CISE in its mission to uphold the Nation's leadership in scientific discovery and engineering innovation through its support of fundamental research in computer and information science and engineering and of transformative advances in cyberinfrastructure. Dr. Jahanian is on leave from the University of Michigan, where he holds the Edward S. Davidson Collegiate Professorship and served as Chair for Computer Science and Engineering from 2007 – 2011 and as Director of the Software Systems Laboratory from 1997 – 2000. His research on Internet infrastructure security formed the basis for the Internet security company Arbor Networks, which he co-founded in 2001 and where he served as Chairman until its acquisition in 2010. He has testified before Congress on a broad range of topics, including cybersecurity and Big Data. Dr. Jahanian holds a master's degree and a Ph.D. in Computer Science from the University of Texas at Austin. He is a Fellow of ACM, IEEE and AAAS.

Irene Qualters, Division Director, Directorate of Advanced Cyberinfrastructure, National Science Foundation

Irene M. Qualters is currently Division Director of Advanced Cyberinfrastructure (ACI) at the National Science Foundation (NSF). ACI is responsible for programs with a total annual budget in FY2013 of over $200 million. These programs support the acquisition, development, and provisioning of state-of-the-art cyberinfrastructure resources, tools, and services essential to the conduct of 21st century science and engineering research and education. ACI is also responsible for the NSF-wide vision, strategy, planning and coordination for research cyberinfrastructure. She joined NSF as a Program Director in December 2009, participating in multidisciplinary, interagency and international activities as well as overseeing several major computational projects within the division's portfolio, including the Blue Waters project at NCSA/UIUC and the Stampede project at TACC/UT at Austin. Irene has a Master's degree in Computer Science. Prior to beginning her NSF responsibilities, she had a distinguished 30-year career in industry, with executive leadership positions for research and development organizations within the technology sector. During her twenty years at Cray Research, in increasingly larger leadership roles, she participated in the development of the first commercially successful vectorizing compiler, the first multiprocessor version of Unix and Cray's landmark massively parallel computer, the T3E. Subsequently, for six years, as Vice President, she led the Research Information Systems for Merck Research Labs (MRL). She is an expert in parallel computer system architectures and in a wide variety of software from scientific applications to compilers to file systems and operating systems.

Thomas Defanti, Research Scientist, Qualcomm Institute University of California, San Diego, Distinguished Professor Emeritus, Computer Science, University of Illinois at Chicago

Thomas A. DeFanti, PhD, is a research scientist at the Qualcomm Institute, part of the California Institute for Telecommunications and Information Technology, University of California, San Diego. At the University of Illinois at Chicago, DeFanti is a distinguished professor emeritus in the department of Computer Science. Before retiring in 2004, he led the UIC Electronic Visualization Laboratory with colleague Dan Sandin for 31 years. Currently, he is principal investigator of the NSF International Research Network Connections Program TransLight/StarLight. DeFanti is an internationally recognized expert in computer graphics since the early 1970s. DeFanti has amassed a number of credits, including: use of EVL hardware and software for the computer animation produced for the 1977 "Star Wars" movie; contributor and co-editor of the 1987 National Science Foundation-sponsored report "Visualization in Scientific Computing;" recipient of the 1988 ACM Outstanding Contribution Award; appointed an ACM Fellow in 1994; and appointed one of several USA technical advisors to the G7 GIBN activity in 1995. He also shares recognition along with EVL director emeritus Dan Sandin for conceiving the CAVE virtual reality theater in 1991. He, Dan, and longtime colleague Greg Dawe continue to build novel virtual reality devices for various clients worldwide. Striving for a more than a decade to connect high-resolution visualization and virtual reality devices over long distances, DeFanti has lead state, national and international teams to build the most advanced production-quality networks available to scientists, with major NSF funding. He is a founding member of GLIF, the Global Lambda Integrated Facility, a global group that manages international switched wavelength networks for research and education. In the USA, DeFanti established the 10 Gigabit Ethernet CAVEwave research network between EVL/StarLight, Seattle/Pacific Northwest GigaPop, and UCSD for OptIPuter and other national/international research uses, which was a model for future high-end science and engineering collaboration infrastructure.

DeFanti has also been active in the ACM SIGGRAPH organization and in the ACM/IEEE Supercomputing (SC) conferences. Current and past activities include: secretary of SIGGRAPH (1977-1981); co-chair of the SIGGRAPH 79 conference; chair of the 11,000-member SIGGRAPH organization (1981-1985); co-chair of the 1998, 2000, 2002, and 2005 iGrid conferences, and editor with UIC's Dana Plepys of the "SIGGRAPH Video Review" video publication, from 1979 to 2012.

He is profiled in Wikipedia at http://en.wikipedia.org/wiki/Thomas_A._DeFanti, at http://www.evl.uic.edu/tom/, and http://www.calit2.net/people/staff_detail.php?id=67

Edward Seidel, Director, National Center for Supercomputing Applications, Founder Professor of Physics, Professor of Astronomy, University of Illinois at Urbana-Champaign

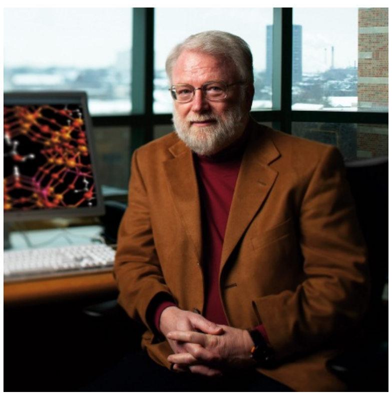
NCSA director Edward Seidel is a distinguished researcher in high-performance computing and relativity and astrophysics with an outstanding track record as an administrator. In addition to leading NCSA, he is also a Founder Professor in the University of Illinois Department of Physics and a professor in the Department of Astronomy. His previous leadership roles include serving as the senior vice president for research and innovation at the Skolkovo Institute of Science and Technology in Moscow, directing the Office of Cyberinfrastructure and serving as assistant director for Mathematical and Physical Sciences at the U.S. National Science Foundation, and leading the Center for Computation & Technology at Louisiana State University. He also led the numerical relativity group at the Max Planck Institute for Gravitational Physics (Albert Einstein Institute) in Germany. Seidel is a fellow of the American Physical Society and of the American Association for the Advancement of Science, as well as a member of the Institute of Electrical and Electronics Engineers and the Society for Industrial and Applied Mathematics. His research has been recognized by a number of awards, including the 2006 IEEE Sidney Fernbach Award, the 2001 Gordon Bell Prize, and 1998 Heinz-Billing-Award. He earned a master's degree in physics at the University of Pennsylvania in 1983 and a doctorate in relativistic astrophysics at Yale University in 1988. http://www.ncsa.illinois.edu/

David Keyes, Professor of Applied Mathematics and Computational Science, Director of the Extreme Computing Research Center, King Abdullah University of Science and Technology

David Keyes is Professor of Applied Mathematics and Computational Science, having served as the Dean of the Division of Computer, Electrical, and Mathematical Sciences and Engineering (CEMSE) at KAUST for its first 3.5 years. Also an Adjunct Professor and former Fu Foundation Chair Professor in Applied Physics and Applied Mathematics at Columbia University, and an affiliate of several laboratories of the U.S. Department of Energy, Keyes graduated in Aerospace and Mechanical Sciences from Princeton in 1978 and earned a doctorate in Applied Mathematics from Harvard in 1984. He works at the algorithmic interface between parallel computing and the numerical analysis of partial differential equations, with a focus on scalable solvers for emerging extreme architectures that require drastic reductions in communication and synchronization. Keyes was awarded an NSF Presidential Young Investigator Award in 1989. For his algorithmic influence in scientific simulation, Keyes has been recognized as a Fellow of the Society for Industrial and Applied Mathematics (SIAM) and of the American Mathematical Society (AMS), with the Sidney Fernbach Award of the IEEE Computer Society, and with ACM's Gordon Bell Prize. Author or editor of more than a dozen federal agency reports and member of several federal advisory committees on computational science and engineering and high performance computing, in 2011, Keyes received the SIAM Prize for Distinguished Service to the Profession. http://cec.kaust.edu.sa

Sushil Prasad, Professor of Computer Science, Director of the Distributed and Mobile Systems Lab, Georgia State University

Sushil K. Prasad (BTech'85 IIT Kharagpur, MS'86 Washington State, Pullman; PhD'90 Central Florida, Orlando - all in Computer Science/Engineering) is a Professor of Computer Science at Georgia State University (GSU) and Director of Distributed and Mobile Systems (DiMoS) Lab. Sushil has been honored as an ACM Distinguished Scientist in Fall 2013 for his research on parallel data structures and applications. He was the elected chair of IEEE Technical Committee on Parallel Processing for two terms (2007-11), and received its highest honors in 2012 - IEEE TCPP Outstanding Service Award. Currently, he is leading the NSF/IEEE-TCPP curriculum initiative on parallel and distributed computing, in coordination with ACM/IEEE CS 2013 curriculum taskforce, with a vision to ensure that all computer science and engineering graduates are well-prepared in parallelism through their core courses in this era of multi- and many-cores desktops and handhelds. http://www.cs.gsu.edu/prasad/



Thom Dunning, Co-director of the Northwest Institute for Advanced Computing, University of Washington, and Professor Emeritus, Department of Chemistry, University of Illinois Urbana-Champaign

Thom H. Dunning, Jr., Ph.D., is currently the co-director of the Northwest Institute for Advanced Computing and Professor Emeritus in the Department of Chemistry at the University of Illinois at Urbana-Champaign.

Dr. Dunning is the former director of the National Center for Supercomputing Applications. As NCSA director, Dr. Dunning was responsible for the development and deployment of the cyberinfrastructure needed by the nation's academic research and education community, including one of the world's most powerful supercomputers—*Blue Waters*.

Before joining UIUC, Dr. Dunning was the founding director of the Joint Institute for Computational Sciences at the University of Tennessee and Oak Ridge National Laboratory and vice president for supercomputing and networking for the University of North Carolina System. Earlier, he served as Assistant Director for Scientific Simulation in the Office of Science at the U.S. Department of Energy, director of PNNL's Environmental Molecular Sciences Laboratory, head of the Theoretical and Computational Chemistry Group at Argonne National Laboratory, and staff member at Los Alamos National Laboratory.

Dr. Dunning is a theoretical and computational chemist who has authored more than 150 scientific publications on topics ranging from techniques for molecular calculations to computational studies of high power lasers, combustion chemistry, and the chemistry of the main group elements. Six of his papers are "citation classics" with more than 1,000 citations each. Dr. Dunning is a Fellow of the American Physical Society, the American Association for the Advancement of Science, and the American Chemical Society. He received DOE's E.O. Lawrence Award in 1997 and the ACS's Computers in Chemical and Pharmaceutical Research Award in 2011.

Dr. Dunning obtained his bachelor's degree in Chemistry in 1965 from the Missouri University of Science & Technology and his doctorate in Chemistry from the California Institute of Technology in 1970. http://www.niac-uw.org/people/thom-h-dunning-jr/

Program Director Panelists

A.J. Meir, Program Director, Computational Mathematics, National Science Foundation

A.J. Meir is a professor of mathematics in the Department of Mathematics and Statistics at Auburn University. During the 2013–2014 academic year he is on assignment at the National Science Foundation in the Division of Mathematical Sciences where he is a Program Director (Rotator) in the Computational Mathematics Program. A. J. Meir holds a B.Sc. in Aeronautical Engineering from the Technion – Israel Institute of Technology and a Ph.D. in Mathematics from Carnegie Mellon University. His primary interests are numerical and computational mathematics: Numerical P.D.E., Numerical Analysis, Computational Science, and Modeling and Simulation of Complex Coupled Phenomena. http://wp.auburn.edu/ajm/ http://www.nsf.gov/staff/staff_bio.jsp?lan=ajmeir

Nigel Sharp, Program Director, Large Synoptic Survey Telescope Project, Division of Astronomical Sciences, National Science Foundation

Nigel Sharp is the Program Director for the Large Synoptic Survey Telescope project, in the Division of Astronomical Sciences (AST) in the Directorate for Mathematical and Physical Sciences (MPS) of the National Science Foundation (NSF). He has additional programmatic responsibilities for awards both individual and project, mostly in theory, computation, data areas, and cross-NSF (inter-divisional) programs. After three degrees in physics, mathematics, and astrophysics at the University of Cambridge (the real one), Nigel moved to Texas and had a varied career there, in Australia and Arizona, in astronomy theory and observation, including instrumentation (vacuum, cryogenic, real-time control, and user interface) and telescope management (maintenance, upgrades, user service). His cyber-work has included supercomputer access and numerical methods consulting, and systems management, networking and security at an NSF FFRDC. After all that, it made sense to join NSF and work on the federal funding of science, which he has done for almost twelve years.

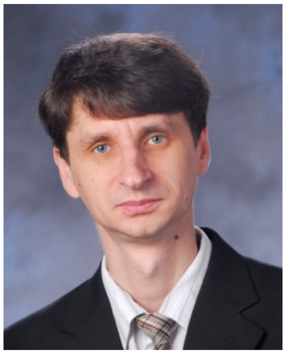
Rudolf Eigenmann, Program Director, Directorate of Computer and Information Science and Engineering, National Science Foundation

Rudolf Eigenmann is a Program Director in NSF's Computer and Information Science and Engineering Directorate, Division of Advanced Cyberinfrastructure. He currently serves as the program officer for the XSEDE project. Dr. Eigenmann is on leave from Purdue University, where he is a Professor in the School of Electrical and Computer Engineering. His research interests include optimizing compilers, programming methodologies and tools, performance evaluation for high-performance computers, and cyberinfrastructure. He has published his work in over 150 papers in international conferences, journals, and workshops. He received his Ph.D. in Electrical Engineering/Computer Science in 1988 from ETH Zurich, Switzerland.
https://engineering.purdue.edu/~eigenman/

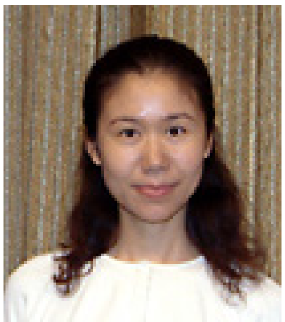
Almadena Y. Chtchelkanova, Program Director, Directorate of Computer and Information Science and Engineering, National Science Foundation

Dr. Almadena Chtchelkanova is currently a Program Director at the Directorate for Computer and Information Science and Engineering at the National Science Foundation. Dr. Chtchelkanova is in charge of the area of High Performance Computing research. She is a member of inter-agency working groups on High End Computing (HECIWG), Interagency Modeling and Analysis Group (IMAG). She is involved in cross-divisional and cross-organizational programs such as Earthcube, Big Data, *SI2*: Software Institutes for Sustained Innovations, and others. Her current portfolio includes over 250 awards totaling over $30,000,000.
In 2010 she served for 2 months as an US Department of State Embassy Fellow in Ankara, Turkey.
Before joining NSF in 2005 Dr. Chtchelkanova worked for Strategic Analysis, Inc. as a Senior Scientist providing technical and programmatic support and oversight to Defense Advanced Research Program Agency (DARPA) programs such as Spintronics, Quantum Information Science and Technology (QuIST), Virtual Integrated Prototyping, Molecular Observation and Imaging, and others.
Dr. Chtchelkanova spent four years working at the Laboratory for Computational Physics and Fluid Dynamics at the Naval Research Laboratory located in Washington, DC. Dr. Chtchelkanova has considerable experience in the area of High Performance Computing (HPC) applications. She developed and implemented portable, scalable, parallel adaptive mesh generation algorithms for computational fluid dynamics, weather forecast, combustion and contaminant transport.
Dr. Chtchelkanova holds an MA degree from the Department of Computer Sciences at the University of Texas at Austin (1996) Ph.D. degree in physics from Lomonosov State University in Moscow, Russia.

Workshop Attendees



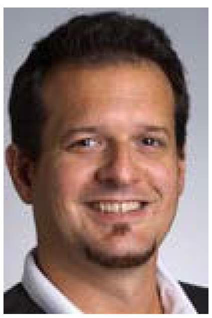
Alexander Alexeev is an Assistant Professor in the George W. Woodruff School of Mechanical Engineering at the Georgia Institute of Technology. He obtained his Ph.D. degree in Mechanical Engineering in 2003 from the Technion - Israel Institute of Technology. He joined Georgia Tech in 2008 after finishing his postdoctoral studies at TU Darmstadt in Germany and at the University of Pittsburgh. He uses modeling and simulations to solve problems at the intersection of engineering, medicine, and biology. His research interests include mesoscale modeling of complex fluids, soft and active materials, interfacial phenomena, and microfluidics. http://cfms.gatech.edu/



Qing (Cindy) Chang is an Assistant Professor in the Department of Mechanical Engineering at SUNY Stony Brook University. She is interested in advanced manufacturing modeling and real-time adaptive control of dynamic manufacturing systems. Her NSF Career project investigates collaborative sensing information processing, and an intelligent online control for battery manufacturing. She obtained her PhD at University of Michigan, and worked at General Motors R&D Center before joining the faculty at Stony Brook University. **http://me.eng.sunysb.edu/~sms/**

Zizhong (Jeffrey) Chen is an Assistant Professor of Computer Science at the University of California, Riverside. He received his Ph.D in computer science from the University of Tennessee, Knoxville under the supervision of Professor Jack Dongarra. He is interested in high performance computing, numerical algorithms and software, and large scale computer simulations. He has worked closely with Jack Dongarra for many years on a variety of high performance computing (HPC) projects including Sca/LAPACK, LAPACK for Clusters, Sparse Linear Algebra Libraries and Algorithms, and FT-MPI, and developed a strong expertise in designing highly scalable HPC software and using the start-of-the-art HPC platforms (e.g., Kraken, Stampede, and Jaguar/Titan). His research has been supported by National Science Foundation, Department of Energy, CMG Reservoir Simulation Foundation, Nvidia, and Microsoft Corporation. He received a Best Paper Award from the International Supercomputing Conference in 2004, an Outstanding Faculty Award from the Colorado School of Mines in 2010, and a CAREER Award from National Science Foundation in 2012. http://www.cs.ucr.edu/~chen/

Robin Dowell is an Assistant professor in Molecular, Cellular, and Developmental Biology and Computer Science at the University of Colorado in Boulder. She is a member of the BioFrontiers Institute and the Linda Crnic Institute for Down Syndrome. She received her D.Sc. in Biomedical Engineering, a M.S. in Computer Science, a B.S. in Genetics, and a second B.S. in Computer Engineering. Robin's research leverages computational and experimental approaches to better understand how common types of genomic variations impact transcriptional regulation. Our work focuses on comparing genome variations between individuals within a species to understand and predict their unique transcriptional response to perturbations. Our work centers around two distinct questions: understanding the impact of aneuploidy on adaptation and transcription; and dissecting how genomic variation within a population impacts transcriptional regulation and response. On the educational front, I am focused on developing educational initiatives that transcend disciplinary boundaries to provide hands-on research experiences. These efforts have resulted in a new graduate program (IQ Biology), a campus undergraduate iGEM team, and new approaches to the teaching of responsible conduct of research (RCR).



Adrian Feiguin joined Northeastern University as Assistant Professor in 2012, after spending 3 years as Assistant Professor at the University of Wyoming. His field of expertise is computational condensed matter, focusing on theoretical and computational aspects of low dimensional strongly interacting quantum systems. This physics is realized under extreme conditions, such as very low temperatures, high pressure, or high magnetic fields, and low spatial dimensions, and it is mostly governed by the collective behavior of the electrons inside a solid. http://www.northeastern.edu/afeiguin/

Sophya Garashchuk is an associate professor of physical and theoretical chemistry at the University of South Carolina. She received the NSF CAREER award in 2011. The goal of Sophya's research is to understand the role of quantum-mechanical effects of nuclear motion in reactive systems in condensed phase, to which end an approximate quantum trajectory dynamics method is being developed in her group. For generality and scalability to molecular systems of a few hundred atoms, implementation of the quantum trajectory dynamics is based on semi-empirical electronic structure calculations and 'mean-field' approximations to the quantum force. http://www.chem.sc.edu/people/facultyStaffDetails.asp?SID=83

Thomas Hacker is an Associate Professor of Computer and Information Technology at Purdue University and Visiting Professor in the Department of Electrical Engineering and Computer Science at the University of Stavanger in Norway. Dr. Hacker's research interests center around high- performance computing and networking on the operating system and middleware layers. Recently his research has focused on cloud computing, cyberinfrastructure, scientific workflows, and data-oriented infrastructure. Dr. Hacker is also co-leader for Information Technology for the Network for Earthquake Engineering Simulation (NEES), which brings together researchers from fourteen universities across the country to share innovations in earthquake research and engineering. http://www2.tech.purdue.edu/cpt/SelfStudy/CPTFacultyVitas/FacultyStaff/DisplayStaffMember.asp?member=tjhacke

Samer Hani Hamdar is an assistant professor at the George Washington University (GWU) where he is the director of the Traffic and Networks Research Laboratory (TNRL). Dr. Hamdar is an affiliate faculty member at the GWU Center for Intelligent System Research (CISR) and the National Crash Analysis Center (NCAC). He holds a BE Degree from the American University of Beirut, a MS Degree from the University of Maryland, College Park and a PhD Degree from the Northwestern University – all in Civil and Environmental Engineering. Dr. Hamdar worked on different projects covering different transportation areas. These projects include two recent National Science Foundation (NSF) Projects titled "Collaborative Research: New Methods for Measuring, Evaluating and Predicting the Safety Impact of Road Infrastructure Systems on Driver Behavior; and "CAREER: Collision Prediction and Vehicle Control Using an Episode-Based Modeling Framework". His primary research interests include Driver Behavior Modeling, Traffic Flow Theory, Intelligent Transportation Systems, Pedestrian Behavior Modeling, Transportation Safety, Evacuation Modeling and Disaster Management. He has an international research background having participated in projects in Germany, Saudi Arabia and the USA. He is a member of the Transportation Research Board (TRB) Traffic Flow Theory and Characteristics Committee (AHB45) and the Task Force on Emergency Evacuations (ANB80T). http://hamdar.seas.gwu.edu

Hamed Hatami-Marbini is an Assistant Professor in the School of Mechanical and Aerospace Engineering at Oklahoma State University. He received his PhD in Mechanical Engineering from Rensselaer Polytechnic Institute in 2009 and was a postdoctoral fellow at Stanford University from 2009 to 2011. His research interests are Solid Mechanics and Multiscale Computational Analysis of Materials, Micro and Nano Mechanics of Materials, Biomechanics and Biomaterials, Mechanics of Natural Composite Materials, and Mechanics of Random Media. http://www.mae.okstate.edu/node/107

Shantenu Jha is an Assistant Professor in ECE at Rutgers University. Before moving to Rutgers, he was the lead for Cyberinfrastructure Research and Development at the CCT at Louisiana State University. His research interests lie at the triple point of Cyberinfrastructure R&D, "theoretical" Applied Computing and Computational Science. Building upon his background in computing and biomolecular sciences, he hopes to play a part in the upcoming revolution at the interface of computing and health-science — global health and "personalized" medicine. Shantenu is the PI of RADICAL (http://radical.rutgers.edu) and RADICAL-Cybertools (http://radical-cybertools.github.com) which are a suite of standards-driven and abstractions-based tools used to support large-scale science and engineering applications, including on most major Production Distributed Cyberinfrastructure -- such as US NSF's XSEDE and the European Grid Infrastructure. In addition he is desiging MIDAS: Middleware for Data-intensive Analytics and Science. Away from work, Jha tries middle-distance running and biking, tends to indulge in random musings as an economics-junky (e.g., is currently musing on the lessons that the architects of the the global financial system can learn from the architecture of global distributed computing systems), and tries to use his copious amounts of free time with a conscience.

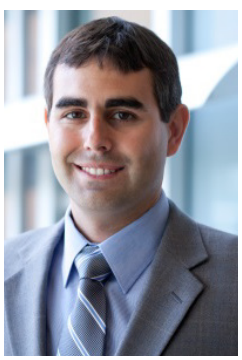
Kapil Khandelwal is an Assistant Professor in the Department of Civil & Environmental & Earth Sciences at the University of Notre Dame. He received BS in Civil Engineering from IIT-Roorkee (India), MS in Structural Engineering From IIT-Delhi (India) and Ph.D. in Civil Engineering from the University of Michigan, Ann Arbor. His research interested includes: computational solid mechanics (FEM), gradient elasticity/plasticity, computational fracture mechanics, topology optimization, and progressive collapse of structural systems. http://ceees.nd.edu/profiles/kkhandelwal

Emmanouil (Manos) Kioupakis is an Assistant Professor of Materials Science and Engineering at the University of Michigan, Ann Arbor, MI. He received his undergraduate degree in Physics from the University of Crete, Greece, in 2001 and the Ph.D. degree in Physics from the University of California, Berkeley, CA, in 2008, under the direction of Prof. Steven G. Louie. From 2008 to 2011 he was a Postdoctoral Scholar at the University of California, Santa Barbara under the direction of Prof. Chris G. Van de Walle. His research interests include first-principles computational studies of quantum processes in electronic materials. He received the NSF CAREER Award in 2013 and the Jon R. and Beverly S. Holt Award for Excellence in Teaching in 2014. http://www-personal.umich.edu/~kioup/UM/Welcome.html

Haibin Ling is an assistant professor of computer and information sciences at Temple University. He got his B.S. in mathematics and MS in computer science from Peking University, and his PhD in computer science from University of Maryland, College Park. Before joining Temple, he also spend one year at University of California Los Angeles as a postdoc. He had worked for Microsoft Research Asia and Siemens Corporate Research as well. Dr. Ling's research interests include computer vision, medical image analysis, human computer interaction, and machine learning. He received the Best Student Paper Award at the ACM Symposium on User Interface Software and Technology (UIST) in 2003, and the NSF CAREER Award in 2014. He has served as an Area Chair of CVPR 2014 and as a Guest Co-Editor for the Pattern Recognition Special Issue on Discriminative Feature Learning from Big Data for Visual Recognition. http://www.dabi.temple.edu/~hbling/

Laurence Loewe is an Assistant Professor at the University of Wisconsin- Madison. He investigates questions in the new field of evolutionary systems biology, which merges systems biology and population genetics. To enable this, his group is developing two major tools. The first, Evolvix, is a new programming language that makes it easy for biologists to describe the systems they study in mathematically rigorous form and link them to real data. The second, Evolution@home, is a globally distributed computing system that is being redesigned for analyzing the flood of simulation data generated by Evolvix models. His group works 'in silico' on diverse topics like circadian clocks, antibiotic resistance evolution, the population genetics of harmful mutations and species extinction. He is interested in bridging the gap between simple analytically understandable mathematical models and biological reality by building rigorous simulation models to answer various evolutionary questions.
http://evolution.ws/people/loewe http://evolvix.org

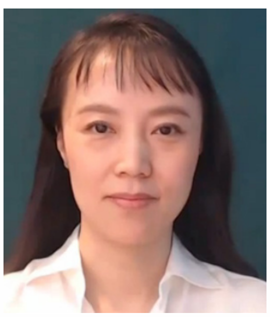
Kamesh Madduri is an assistant professor in the Computer Science and Engineering department at The Pennsylvania State University. He received his PhD in Computer Science from Georgia Institute of Technology's College of Computing in 2008, and was previously a Luis W. Alvarez postdoctoral fellow at Lawrence Berkeley National Laboratory. His research interests include high-performance computing, parallel graph algorithms, and massive scientific data analysis. He is a recipient of the NSF CAREER award (2013), a co-recipient of the best paper award at the 42nd International Conference on Parallel Processing (2013), and was awarded the first Junior Scientist prize from the SIAM Activity group on Supercomputing (2010). He is a member of IEEE, ACM, and SIAM. http://www.cse.psu.edu/~madduri/



Jason P. McCormick is an Assistant Professor in the Department of Civil and Environmental Engineering at the University of Michigan. Dr. McCormick's primary interests are in the areas of extreme load mitigation, structural response reduction through innovative passive control systems, and the performance of steel structures and components under seismic loads. This work includes characterizing and optimizing the properties of non-traditional civil engineering materials through experiments at multiple scales, the study of hollow structural sections through large-scale testing and high fidelity finite element models, and the evaluation of aging effects on the performance of steel bridge systems. He currently serves on the American Institute of Steel Construction (AISC) Task Committee 6 on Connections and the Subcommittee on Hollow Structural Section Connections. He is the recipient of the 2010 AISC Milek (Faculty) Fellowship, 2010 Chi Epsilon Great Lakes District "James M. Robbins" Excellence in Teaching Award and 2014 NSF CAREER Award. http://www.cee.umich.edu/Jason_McCormick

Dr. Tim Mueller is an assistant professor at Johns Hopkins University. His primary area of research is the computational design and discovery of new materials, with a focus on materials for energy storage and conversion. To accurately predict the properties of materials his research group uses a variety of methods including *ab-initio* calculations and materials informatics. Prior to joining the faculty of Johns Hopkins he co-founded a battery company, Pellion Technologies, at which he managed a high-throughput computational search for advanced battery materials. He has a Ph.D. in Materials Science and Engineering from MIT and an A.B. in Applied Mathematics from Harvard University. http://muellergroup.jhu.edu

Christian D. Ott, Professor of Theoretical Astrophysics in the TAPIR group at Caltech, works on simulations stellar explosions and mergers of neutron star binaries. He received his Diploma in Physics from the University of Heidelberg in 2003 and his PhD from the Albert Einstein Institute and the University of Potsdam in 2007. He is also a member of the LIGO Scientific Collaboration that aims to make the first direct detection of gravitational waves within this decade. http://www.tapir.caltech.edu/~cott



Dr. Judy Qiu is an assistant professor of Computer Science in the School of Informatics and Computing at Indiana University and an assistant director of the school's Digital Science Center. Her research interests are parallel and distributed systems, cloud computing, and high-performance computing. Qiu leads the SALSA project, involving professional staff and Ph.D. students from the School of Informatics and Computing. SALSA focuses on data-intensive computing at the intersection of cloud and multicore technologies with an emphasis on scientific data analysis applications by using MapReduce and traditional parallel computing approaches. Her research has been funded by NSF, NIH, Microsoft, Google and Indiana University. She is a recipient of NSF CAREER Award in 2012 and Indiana University Trustees Award for Teaching Excellence in 2013-2014. http://www.cs.indiana.edu/~xqiu/

Dr. Ioan Raicu is an assistant professor in the Department of Computer Science (CS) at Illinois Institute of Technology (IIT), as well as a guest research faculty in the Math and Computer Science Division (MCS) at Argonne National Laboratory (ANL). He is also the founder (2011) and director of the Data- Intensive Distributed Systems Laboratory (DataSys) at IIT. He has received the prestigious NSF CAREER award (2011 - 2015) for his innovative work on distributed file systems for exascale computing. He was a NSF/CRA Computation Innovation Fellow at Northwestern University in 2009 - 2010, and obtained his Ph.D. in Computer Science from University of Chicago under the guidance of Dr. Ian Foster in March 2009. He is a 3-year award winner of the GSRP Fellowship from NASA Ames Research Center. His research work and interests are in the general area of distributed systems. His work focuses on a relatively new paradigm of Many-Task Computing (MTC), which aims to bridge the gap between two predominant paradigms from distributed systems, High- Throughput Computing (HTC) and High-Performance Computing (HPC). His work has focused on defining and exploring both the theory and practical aspects of realizing MTC across a wide range of large-scale distributed systems. He is particularly interested in resource management in large scale distributed systems with a focus on many-task computing, data intensive computing, cloud computing, grid computing, and many-core computing. Over the past decade, he has co-authored 86 peer reviewed articles, book chapters, books, theses, and dissertations, which received over 3250 citations, with a H- index of 22. His work has been funded by the NASA Ames Research Center, DOE Office of Advanced Scientific Computing Research, the NSF/CRA CIFellows program, and the NSF CAREER program. He has also founded and chaired several workshops, such as ACM Workshop on Many-Task Computing on Grids and Supercomputers (MTAGS), the IEEE Int. Workshop on Data-Intensive Computing in the Clouds (DataCloud), and the ACM Workshop on Scientific Cloud Computing (ScienceCloud). He is on the editorial board of the IEEE Transaction on Cloud Computing (TCC), the Springer Journal of Cloud Computing Advances, Systems and Applications (JoCCASA), as well as a guest editor for the IEEE Transactions on Parallel and Distributed Systems (TPDS), the Scientific Programming Journal (SPJ), and the Journal of Grid Computing (JoGC). He has been leadership roles in several high profile conferences, such as HPDC, CCGrid, Grid, eScience, and ICAC. He is a member of the IEEE and ACM. More information can be found at http://www.cs.iit.edu/~iraicu/, http://datasys.cs.iit.edu/, and http://www.linkedin.com/in/ioanraicu. http://www.cs.iit.edu/~iraicu/

Christopher J. Rozell received a B.S.E. degree in Computer Engineering and a B.F.A. degree in Music (Performing Arts Technology) in 2000 from the University of Michigan. He attended graduate school at Rice University, receiving the M.S. and Ph.D. degrees in Electrical Engineering in 2002 and 2007, respectively. Following graduate school he joined the Redwood Center for Theoretical Neuroscience at the University of California, Berkeley as a postdoctoral scholar. In 2008 Dr. Rozell joined the faculty at the Georgia Institute of Technology where he is currently an Assistant Professor and holds the Demetrius T. Paris Junior Professorship in Electrical and Computer Engineering.

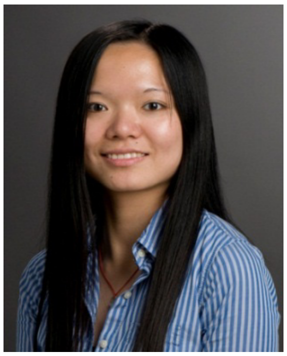
His research interests live at the intersection of signal processing, machine learning and computational neuroscience. Specifically, his lab uses tools from modern data analysis to improve our understanding of neural systems and insight from modern neuroscience to build more effective computational systems, with applications ranging from biotechnology to remote sensing. His research lab is affiliated with both the Center for Signal and Information Processing and the Laboratory for Neuroengineering. Dr. Rozell received the National Science Foundation CAREER Award in 2014, and previously was the recipient of the Texas Instruments Distinguished Graduate Fellowship at Rice University. In addition to his research activity, Dr. Rozell was awarded the CETL/BP Junior Faculty Teaching Excellence Award at Georgia Tech in 2013. http://users.ece.gatech.edu/~crozell/

Dr. Onkar Sahni is currently an Assistant Professor in the Department of Mechanical, Aerospace and Nuclear Engineering at Rensselaer. He joined Rensselaer in 2011, after working as research scientist/engineer at the Center for Predictive Engineering and Computational Science (PECOS) at the University of Texas-Austin. His research is focused on simulation-based predictive tools for coupled fluid flow problems involving turbulence and flow control. His research puts emphasis on adaptive and high-order methods, uncertainty quantification techniques and extreme-scale computing. http://www.scorec.rpi.edu/~sahni/ http://faculty.rpi.edu/node/1123

Suzanne Shontz is an Assistant Professor in the Department of Mathematics and Statistics at Mississippi State University. She is also affiliated with the Center for Computational Sciences, the Department of Computer Science and Engineering, and the Graduate Program in Computational Engineering at Mississippi State. In August, she will join the Department of Electrical Engineering and Computer Science at The University of Kansas (KU) as an Associate Professor. At KU, she will also be affiliated with the Information and Telecommunication Technology Center and the Graduate Program in Bioengineering. Suzanne's research is in parallel scientific computing and focuses on the development of meshing and numerical optimization algorithms and their applications to medicine, image processing, electronic circuits, acoustics, and materials. Suzanne is the recipient of a 2011 NSF PECASE Award from President Obama for her research in computational and data-enabled science and engineering. She also received a 2011 NSF CAREER Award for her research on parallel dynamic meshing algorithms, software, and theory for simulation-assisted medical interventions. Along with Thomas Hacker of Purdue University, she is a Co-Chair of the 2012, 2013 and 2014 NSF CyberBridges Workshops. Suzanne chaired the 2010 International Meshing Roundtable, the premier conference in unstructured mesh generation, and has served on numerous program committees for international conferences in computational and data-enabled science and engineering. She is also an Associate Editor of the De Gruyter Open Book Series in Medicine. http://sshontz.math.msstate.edu

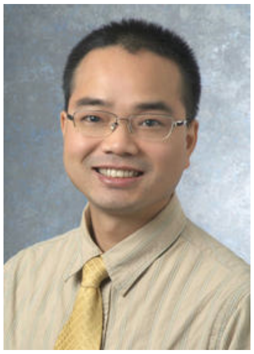
Dr. Jun Wang is an Associate Professor in Department of Electrical Engineering and Computer Science at the University of Central Florida, Orlando, FL, USA. He received his Ph.D. in Computer Science and Engineering from University of Cincinnati in 2002. He is the recipient of National Science Foundation Early Career Award 2009 and Department of Energy Early Career Principal Investigator Award 2005. He has authored over 80 publications in premier journals such as IEEE Transactions on Computers, IEEE Transactions on Parallel and Distributed Systems, and leading HPC and systems conferences such as HPDC, EuroSys, ICS, Middleware, FAST. He has conducted extensive research in the areas of Computer Systems and High Performance Computing. His specific research interests include massive storage and file System in local, distributed and parallel systems environment. His group has secured more than three million dollars federal research fundings in thelast five years, including a 55-million NASA project. He has graduated 6 Ph.D. students who upon their graduations were employed by major

US IT corporations (e.g., Google, Microsoft, etc). He currently serves on the editorial board for the IEEE Transactions on Parallel and Distributed Systems since 2012. He is chairing the 10th IEEE International Conference on Networking, Architecture, and Storage network, and has co-chaired the IEEE ScaleCom 2012, 1st International Workshop on Storage and I/O Virtualization, Performance, Energy, Evaluation and Dependability (SPEED 2008) held together with HPCA.
http://eecs.ucf.edu/faculty_template/wang/index.php
http://www.eecs.ucf.edu/~jwang/



Linwei Wang is an Assistant Professor in the PhD Program of Computing and Information Sciences at the Rochester Institute of Technology in Rochester, NY. Her research interests center around data-driven modeling, statistical inference, and simulation-based optimization, with application to computational physiology and personalized biomedicine. She currently directs the Computational Biomedical Lab in RIT, with a recent research focus on personalized modeling of *in-vivo* cardiovascular systems using noninvasive biomedical data, and its application to improve patient care in cardiac arrhythmia and other heart diseases. Her NSF CAREER project investigates the integration of physics-based modeling into data-driven inference and learning methods. Dr. Wang obtained her bachelor degree in Optic-Electrical Engineering from Zhejiang University (China) in 2005, her master degree in Electronic and Computer Engineering from Hong Kong University of Science and Technology in 2007, and her PhD in Computing and Information Sciences from RIT prior to joining the faculty of RIT in 2009. http://phd.gccis.rit.edu/linweiwang/

Thomas Wies is an Assistant Professor in the Computer Science Department of New York University. He holds a Masters degree in Computer Science from Saarland University, Germany (2005) and received his doctorate in Computer Science from the University of Freiburg, Germany in 2009.
Before joining NYU in 2011, Dr. Wies held post-doctoral positions at École Polytechnique Fédérale de Lausanne, Switzerland and at the Institute of Science and Technology Austria. His research interests are in formal methods, verification, and programming languages. In particular, he is interested in the theory and development of tools that increase software productivity and assist software engineers in building reliable software. This includes tools for automated verification, automated error detection, and automated debugging. http://cs.nyu.edu/wies



Dr. Xiong (Bill) Yu is an associate professor at the Department of Civil Engineering, Case Western Reserve University. He also holds courtesy appointments in the Department of Electrical Engineering and Computer Science, the Department of Materials Science and Engineering, CWRU. Dr. Yu received his Ph.D. degree from Purdue University, B.S. and M.S. degrees from Tsinghua University, China. His research interest is in the broad area of civil engineering with emphasis on embracing innovative sensors and materials to improve sustainability and intelligence of the civil infrastructure system. He is the PI of over 20 research projects sponsored by various agencies and private industry with total value exceeding $3 million. Dr. Yu is a member of ASCE, ISSMGE, IEEE, ASME, SPIE, ASTM and TRB. He serves on SHRP and NCHRP project panels. He is a member of editorial board of three ASCE and ASTM journals. Dr. Yu is a recipient of the NSF CAREER award in 2009. He has published over 150 papers. http://filer.case.edu/xxy21/Index.html

Appendix C. Poster Session

The poster session provided a forum for attendees to present work from their CAREER projects and to discuss and share their work with other researchers to foster connections and potentially new collaborations. All of the workshop participants presented posters, which are listed in Table 1.

Table 1. Posters presented at the Cyberbridges 2014 conference.

Name	Institution	Poster Title
Alexander Alexeev	Georgia Institute of Technology	Mesoscale computational model of soft polymer networks
Qing (Cindy) Chang	SUNY at Stonybrook	Collaborative modeling for distributed sensing and real-time intelligent control to improve battery manufacturing
Zizhong (Jeffrey) Chen	University of California, Riverside	Dependable high performance scientific computing at the extreme scale via algorithm based fault tolerance
Robin Dowell	University of Colorado, Boulder	Biology without boundaries; experimental, computational, and educational approaches to studying transcription
Adrian Feiguin	Northeastern University	Non-equilibrium quantum dynamics in strongly correlated systems
Sophya Garashchuk	University of South Carolina	Exploration of quantum effects on reactivity of large molecular systems with the quantum trajectory/electronic structure dynamics
Thomas J. Hacker	Purdue University	Understanding system faults and improving job reliability for large-scale HPC systems
Samer H. Hamdar	George Washington University	Vehicle to vehicle and vehicle to infrastructure communication logics using an episode based modeling framework
Haamed Hatami-Marbini	Oklahoma State University	Investigation of mechanical properties of random fiber networks
Shantenu Ja	Rutgers University	Abstractions for distributed dynamic data-intensive (D3) science on NSF distributed cyberinfrastructure
Kapil Khandelwal	University of Notre Dame	Multiscale topology optimization: Elasticity with microstructures
Emmanouil (Manos) Kioupakis	University of Michigan, Ann Arbor	Predictive modeling of electronic materials
Haibin Ling	Temple University	High-order spectral analysis for groupwise correspondence: theory, algorithms, and applications

Laurence Loewe	University of Wisconsin-Madison	Maximizing the expressivity of the Evolvix model description language
Kamesh Madduri	Pennsylvania State University	Algorithmic and software foundations for large-scale graph analysis
Jason McCormick	University of Michigan, Ann Arbor	Controlling seismic and wind response with enhanced energy dissipation from innovative materials
Tim Mueller	Johns Hopkins University	Predicting the surface structure of crystalline materials
Christian Ott	California Institute Of Technology	New paradigms for massive astrophysical computation in the petascale and exascale era
Judy Qiu	Indiana University	Generalizing MapReduce as a unified cloud And HPC routine
Ioan Raicu	Illinois Institute of Technology	Poster title not given
Christopher Rozell	Georgia Institute of Technology	Exploiting low-dimensional structure in data for more effective, efficient and interactive machine intelligence
Onkar Sahni	Rensselaer Polytechnic Institute	Abstractions and algorithms for efficient adaptive analysis of stochastic PDEs
Suzanne Shontz	Mississippi State University	Parallel dynamic meshing techniques for simulation-assisted medical interventions
Jun Wang	University of Central Florida	SLAM: scalable locality-aware middleware for I/O in scientific analysis and visualization
Linwei Wang	Rochester Institute of Technology	CAREER: integrating physical models into data-driven interference
Thomas Wies	New York University	Program abstractions for automated debugging
Xiong (Bill) Yu	Case Western Reserve University	Cyberinfrastructure for human factors reduction to improve transportation safety

Appendix D. Collaboration Activities at the Workshop

Before the workshop, attendees were asked to provide a list of areas of interest in which they were seeking collaborators. These were included in the workshop program to help attendees identify potential collaborators. The areas of interest for potential collaborations identified by NSF CAREER Awardees are summarized in Table 2. Attendees were encouraged to discuss potential collaborations with each other and to network with each other throughout the workshop.

Table 2. Sample of Areas of Interest. Attendees listed potential areas of collaboration.

Researcher	Areas of Interest for Collaboration
Alexeev, Alexander	High performance computing, computational fluid dynamics, mesoscale and particle based computational methods
Chang, Qing (Cindy)	Data-driven modeling, large-scale systems, real-time control and decision making, energy management
Chen, Ziahong (Jeffrey)	Biology, chemistry, earth sciences, economics and business, environmental sciences, mechanical engineering, medicine, physics, any other field that uses high performance computing knowledge
Dowell, Robin	Visualization, large scale datasets on transcription, responsible conduct of research training and education
Feiguin, Adrian	Quantum information, quantum chemistry
Garashchuk, Sophya	Interested in collaboration with computer scientists to implement electronic structure calculations on-the-fly practical for thousands of electrons using CPU/GPU and, more generally, in the area of the Materials Genome Initiative
Hacker, Thomas	Reliability, high performance computing
Hani Hamdar, Samer	Communication and infrastructure (electricity and roadway) network vulnerabilities during extreme conditions (external natural and man-made hazards). Security and resiliency of vehicle to vehicle and vehicle to infrastructure communication for an efficient (safe and reliable) autonomous driving.
Hatami-Marbini, Hamed	Experimental validation studies for computational models, design of (bio-inspired) materials based on numerical simulations, multiscale modelling, parallel programming, and biomechanics.
Jha, Shantenu	Climate and polar science, biomedical and biomolecular science
Khandelwal, Kapil	Multiscale mechanics, large scale optimization, GPU/parallel algorithms and optimization, uncertainty quantification.
Kioupakis, Emmanouil (Manos)	Materials theory, condensed matter theory, electronic structure calculations

Ling, Haibin	Computer vision, including visual recognition, visual surveillance, visual summarization, etc. Biomedical image analysis, including anatomic structure detection and segmentation, image-guided diagnosis. Privacy protection, especially visual privacy protection
Loewe, Laurence	I look for advice on the best possible choices of hash-functions for various purposes to enable the general data structures needed for Evolvix development. I welcome potential collaborators with cutting edge simulation engines worth integrating into Evolvix by developing language elements that make it easy to drive these simulation engines by building interesting biological models for them
Madduri, Kamesh	Graph analytics and its applications in bioinformatics, intelligence, online social data analysis, scientific computing; HPC education for non-CS students; complementary expertise to target NSF's BIGDATA, XPS, CDS&E solicitations
McCormick, Jason P.	Finite element modeling including contact, crack propagation and weld behavior, mixed computational and experimental approaches, 3D visualization for research and teaching, educational research
Mueller, Tim	Experimental synthesis and characterization of materials for electrochemical energy storage and conversion; Experimental synthesis and characterization of material surfaces
Ott, Christian D.	Massively-parallel computation in heterogeneous CPU+accelerator environments.
Qiu, Judy	Applications that need Big Data system support
Raicu, Ioan	Data-intensive computing applications (requiring either POSIX or NoSQL interfaces), communities interested in data provenance, large-scale workflow-based applications
Rozelle, Christopher	Application areas requiring data analysis
Sahni, Onkar	Automatic differentiation, source transformation and embedding techniques,

	adaptive software for high-dimensional spaces, energy monitoring tools (for PDE assembly and linear solver codes)
Shontz, Suzanne	Applications involving dynamic meshing, scientific visualization, GPU computing
Wang, Jun	Big Data applications, data-intensive high performance computing
Wang, Linwei	Multi-physics, multi-scale modeling and simulation, statistical inference & optimization, machine learning, Signal and image analysis, high performance computing, scientific visualization
Wies, Thomas	Programming languages, static and dynamic program analysis, automated debugging: fault localization, program repair, software verification and reliability, concurrency, automated reasoning.
Yu, Xiong (Bill)	Wireless network, system integration, data and algorithm, etc