Written Testimony of
Dr. Kelly Gaither
Director of Visualization/Senior Research Scientist
Texas Advanced Computing Center
The University of Texas at Austin

Before the
House Science, Space and Technology Committee
Subcommittee on Research

Hearing on
"Applications for Information Technology Research & Development"

February 14, 2013

Thank you, Chairman Bucshon, Ranking Member Lipinski and members of the Subcommittee for the opportunity to testify before you today. In my capacity as Director of Visualization and Senior Research Scientist at the Texas Advanced Computing Center at The University of Texas at Austin, I have direct experience on the significant impact federal dollars have had on information technology research and development. I am speaking to you today as a researcher having worked at two interdisciplinary research centers: presently at the Texas Advanced Computing Center and, before that, at Mississippi State's National Science Foundation Engineering Research Center supporting high-fidelity physics simulations over complex geometries.

We often hear that we live in the information age. To be more specific, we live in the data age. We are inundated with data in all aspects of our life, both personal and professional. Data is merely a delivery vehicle for what we are truly interested in – knowledge and information. The process by which we uncover this information is what drives my research. In my laboratory, we are dedicated to developing new methods for finding information in what is often an enormous amount of data. The tremendous processing capacity present in our visual cortex makes visualization, or the process by which we transform data to visual imagery, a powerful means for ferreting out information. This transformation requires an understanding of algorithm design, computational geometry, perception, computational science, analytics and cognitive processes.

Over the past twenty years, I have been fortunate to research and develop methods for visualizing data of fluid flow over aircraft, emerging storm systems, biological processes and K-12 decision making to name a few. This work is always done in an interdisciplinary format bringing together researchers from a variety of backgrounds with the expectation that meshing our collective expertise will provide us greater opportunities for advancing the state of the art in science.

The Texas Advanced Computing Center

At the Texas Advanced Computing Center, our mission is to enable discoveries that advance science and society through the application of advanced computing technologies. To fulfill this mission, we identify, evaluate, deploy, and support powerful computing, visualization, and storage systems and software. Helping researchers and educators use these technologies effectively, and conducting research and development to make these technologies more powerful, more reliable, and easier to use is at the forefront of what we do. Thousands of researchers each year use the computing resources available at TACC to forecast weather and environmental disasters such as the BP oil spill, produce whole-Earth simulations of plate tectonics, and perform other research relevant to the public at large. The center is supported by the National Science Foundation, The University of Texas at Austin, The University of Texas System, and grants from other federal agencies. As a leading resource provider in the NSF XSEDE project, TACC is one of eleven centers across the country providing leadership-class computing resources to the national research community. At TACC, we support thousands of projects for thousands of researchers across all aspects of science and engineering. We partner with companies like Shell, Chevron, and BP to push the state of the art, providing beneficial advancements to their science and engineering process as well. Over the past eleven years, TACC has trained a multitude of professional staff and students who now work at companies like Google, Intel, and Microsoft. Additionally, TACC outreach programs have trained researchers and provided computational resources to over a hundred universities.

Science Only Advances if We Continue to Push the Envelope

As stated in the 2005 report to the president, "Computational Science: Ensuring America's Competitiveness," computational science has become the third pillar of 21st century science[1]. This third pillar complements theory and physical experimentation, allowing scientists to explore phenomenon that are too big, small, fast, or dangerous to investigate in the laboratory. Computational science has made significant progress in the last two decades, but has only been able to do so because of federal investments, interdisciplinary teamwork, and leveraging the successes of researchers before us.

As a computational science researcher, focusing on visualization, I have been supported by funding under the NITRD vision since I was a graduate student. This funding facilitated my education, my professional growth, and by extension, the students and staff that have trained under me over the years. Without this funding, I would have left academic research to pursue opportunities that would have allowed me to be self-sufficient.

One example of the interdisciplinary research that I have been funded to work on is visualization and data analysis of massive scale turbulent flow simulations to track and understand small-scale features [1]. Turbulence, the most common state of fluid motion in nature and engineering, is a Grand Challenge problem for the physical and computer sciences and has applications in aircraft and automobile design, energy, storm damage

---

[1] President's Information Technology Advisory Committee Report to the President,

and galaxy formation. Effectively simulating the wide range of non-linearly interacting three-dimensional fluctuations typical of applications requires the largest supercomputers in the world today. Understanding the geometric and dynamic descriptions of intense events in these simulations allows scientists to get closer to a more complete understanding of turbulent flows and more accurate models for engineering applications. The images shown in Figure 1 show the dramatic increase in data complexity as the size and scale grow from 512^3 to 1024^3 to 2048^3 to 4096^3. Our interdisciplinary team of researchers worked together to develop new methods for classifying and tracking these minute scale features over time.

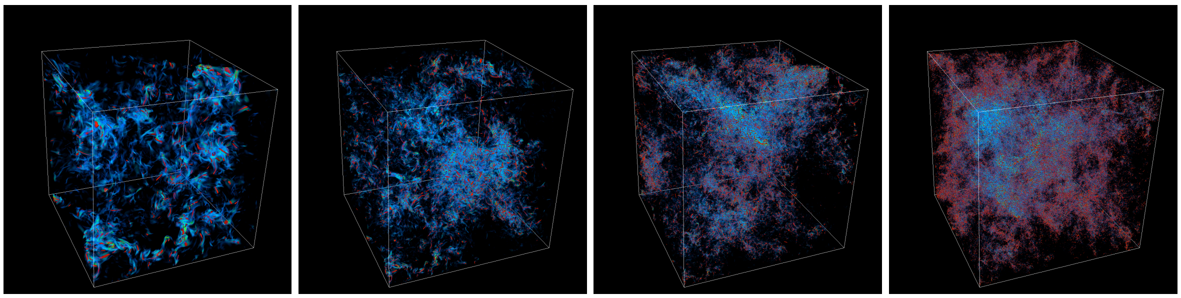

Figure 1: Visualizations showing increasing levels of complexity in the data as the size and scale of the turbulent flow increases.

I am also the principal investigator for the largest hardware-accelerated visualization cluster in the world, capable of producing visualizations of data on an unprecedented scale. This Longhorn team is also an interdisciplinary effort bringing together researchers from the University of Texas at Austin, the Scientific Computing and Imaging Institute (SCI) at the University of Utah, the Purdue Regional Visual Analytics Center (RVAC), the Data Analysis Services Group at the National Center for Atmospheric Research (NCAR), the University of California Davis (UC Davis) and the Southeastern Universities Research Association (SURA) [2]. As a result of this project, we have enabled more than 800 active projects on the system representing 619 individual researchers conducting large-scale computational science. This resource has facilitated researchers across the nation from fields of science ranging from computational physics and chemistry to linguistics and social science. The Longhorn team worked in collaboration with Dell Computers to architect a commodity based system that could be easily replicated at other universities on a multitude of scales. As a result of this funding, we have trained hundreds of people at institutions across the nation with a targeted effort to conduct training at Minority Serving Institutions.

I am also tasked with setting research direction and maintaining a visualization laboratory that serves the University of Texas at Austin population. This 2900 square foot laboratory is home to one of the largest tiled displays in the world, with a peak count of 328 million pixels. In this laboratory, we research and develop tools and interaction mechanisms for

---

[2] Gaither, K. (PI), "*Enabling Transformational Science and Engineering Through Integrated Collaborative Visualization and Data Analysis for the National User Community*," NSF XD VIS ($7,000,000) August 2009 – July 2012.

next generation visualization environments that can be replicated at a variety of scales across a multitude of institutions. We also designed and constructed a large-format touch table that has 32 point touch capability at 5mm resolution. This touch table behaves much like an oversized iPad providing a natural mechanism for learning and rapidly testing new research concepts. Additionally, the visualization laboratory has had more than 18,000 people come through the doors, many of which are K-12 students. While we work diligently to provide a production visualization laboratory to the UT Austin constituency, we also research and develop displays and interaction mechanisms for visualizing next generation science. To this end, we recently submitted a proposal to research a new paradigm for data exploration through the seamless integration of multiple spatially aware visualization systems that form a visualization ecosystem. Recent advances in human-centered computing have introduced computing interfaces that enable the user to interact with, and manipulate compute devices and data displays in a more intuitive manner. Specifically, recent trends in mobile computing, cloud computing, and ubiquitous computing have simplified human-computer interactions with everyday devices. The trends and advancements in these areas have laid the foundation for a new abstraction that will bridge a gap in human-computer interaction. This abstraction, shown in Figure 2, will provide a seamless environment in which individuals can navigate data unimpeded by the physical constraints and boundaries set by compute devices and displays, and will provide a reactive environment that evolves with user defined interactions.

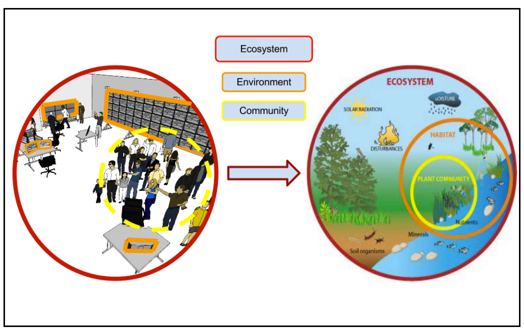

Figure 2: This next generation visualization laboratory is an analog of the biological ecosystem that seeks to develop an interactive, reflective, and reactive computing environment for visualization.

The funding for these projects and others has enabled me to have a national impact by developing leading programs in visualization and training for future computational scientists. The investment in computational science in general, and high end computing specifically, have provided much needed resources for me and for a large number of researchers working on national open science problems. These resources and technologies require a significant investment. Without them, we as researchers would be bound by investments at the institutional scale, significantly inhibiting our ability to fully exploit intellectual meritorious research at full scale.

Science can only advance with better, more capable tools – be they software, hardware or some blend therein. Building on prior knowledge and capabilities allows us to leverage the innovations of researchers that have come before us. Without new scientific capabilities, we hamstring research and inhibit our ability to move forward. We are a country of innovators and this innovation must be fostered with significant investment and patience. Building on the work of others fundamentally depends on us having access to new capabilities that prior researchers did not have. The NITRD program gives us those resources and funding for resources at a scale that individual institutions would not be able to afford otherwise.

Significant Research Directions NITRD Should be Pursuing

First let me commend the efforts to create national programs with increasing focus on data. There are many fields of science that depend on our ability to quickly process vast amounts of data. Currently, data growth rates far outstrip our ability to process and analyze it. However, increasing the focus on data intensive computing should not be done to the exclusion of funding research and development in modeling and simulation. It is imperative that we, as a national IT strategy, strive to build a balanced portfolio of funding opportunities that focus on all aspects of next generation science.

We can see evidence of a shift in the nation's High End Computing strategy. The National Science Foundation is now funding one or two high-end resources every other year versus the one or two a year they funded in the past. This decrease in HEC funding is not limited to the resources themselves, but extends to many of the underlying scientific applications and crucial software tools as well. This dip in resource provisioning is at odds with the increased need for high end computing technologies in open science research.

Additionally, there has always been a lack of persistent funding for visualization. It is generally agreed that visualization is a powerful means of synthesizing data, but many equate visualization with vision and take it for granted. However, the transformation from data to visual imagery is rooted in scientific principles that include concepts in human perception, numerical linear algebra and cognition. Blending these concepts in a reproducible, concise visual context allows us to create meaningful information from vast amounts of data. While it is generally recognized that visualization is a crucial part of the science pipeline today, it will only be more crucial in the big data era. We need students and researchers well versed in numerical algorithms, data mining, statistics, and computer science, but there is nothing as efficient as a great visualization to communicate science.

As complexity increases, so does the need for visualizations that act as both a communications mechanism and a teaching tool. Persistent funding is needed for visualization – including scientific visualization, information visualization and visual analytics. At TACC, we are committed to researching and developing visualization tools and providing visualization resources to both local and national user communities. At present, there is very little funding for this relative to the need in the open science community.

Ensuring a Persistent Pipeline to Meet the Nation's IT Needs

I graduated from high school in one of the poorest states in the nation at a time when young women were not encouraged to pursue higher education, let alone major in computer science. Fortunately, I was encouraged by my family to make some strategic decisions regarding career choices, and I have had the opportunity to work with a number of mentors in all aspects of computational science. I have been paid by federally funded computational research dollars since I was roughly 24 years old. I am a by-product of long-term federal funding for basic and applied research. Without this funding, I would not have had the opportunities to participate in many of the interdisciplinary projects that focus on solving some of society's most challenging issues – hurricane prediction, imaging the human body at the subcellular level, and designing safer, more fuel efficient vehicles to name a few. There is no substitute for immersing students in all aspects of computational research throughout their educational process. This immersion, however, requires persistent commitments from funding agencies and educators alike. We need opportunities to educate students in interdisciplinary research and provide invaluable hands-on experience working with teams of researchers. Funding for research programs and graduate students to work on these programs is crucial to making fundamental advances. There is no substitute for more research funding. However, in national IT at the leading edge, we also need a curriculum change to carve out a home for students to thrive. The overwhelming business need in the dot.com era shifted the attention to focus specifically on those needs, a necessary shift at that time. We are missing a thriving focus on research and development that is not driven by quarterly profit bottom line. There needs to be much more investment in curriculum development for people to work on the science and data intensive applications for large-scale problems. To put it succinctly, there must be curriculum, there must be funding, and there must be exciting opportunities for our students to stay in the field.

Summary of Testimony

I would like to reiterate my appreciation for allowing me to speak to you today about the impact that the NITRD program has had in my research and subsequently for the general public. In closing, I would like to summarize the key points I have spoken about today:

1. **Make significant continued investments in the NITRD program.** As a taxpayer, I recognize that these are tight economic times. However, as a researcher I know that investments in research will keep us at the forefront of innovation. We must not shortchange our commitment and vision to continue the successes of those that have come before us.
2. **Maintain a balanced portfolio of NITRD funding opportunities for researchers in computational science.** We must find a way to fund additional investments, not to the exclusion of existing funding streams. It is a combination of efforts that is most likely to be fruitful.
3. **Provide exciting opportunities to entice students to stay in computational science.** We must deal head on with the brain drain that our universities are experiencing in undergraduate and post-graduate education. While, there is no magic bullet that will solve this problem, it seems clear that a new approach is warranted. This new approach requires an investment in both curriculum development and student research to provide exciting opportunities for future generations of scientists.

[1] K. P. Gaither, H. Childs, K. Schulz, C. Harrison, B. Barth, D. Donzis, and P. Yeung. Using Visualization and Data Analysis to Understand Critical Structures

in Massive Time Varying Turbulent Flow Simulations. IEEE Computer Graphics and Applications, 32(4):34-45, 2012.