NIST NSCI Seminar Series

NIST, along with IARPA, is one of the two government Foundational R&D Agencies engaged in the National Strategic Computing Initiative announced by President Obama last August. As such, the agency has decided to organize a bi-monthly seminar program beginning May 17, 2016 and ending around mid-February of 2017. The talks are an hour long and begin at 1pm eastern time. The seminar series is intended to cover a broad array of topics designed to increase the knowledge base of NIST staff in advanced computing and data science as well as covering specific areas where NIST is likely to play a large role such as materials for future computational platforms, measurement and testing science, algorithms and applied mathematics, neuromorphic and quantum computing.

Outsiders are welcome to attend but need to be badged to enter the NIST campus. Please contact Barry Schneider (bis@nist.gov) if you wish to attend. The list of speakers is not complete, but I include those who have already confirmed that they will come.

- Marc Manheimer IARPA 6/7/2016 Bldg 221/B145

Title***: Progress in Superconducting Computing***

Since the 1950s, there have been several efforts to develop a computer based on superconducting logic. Each was abandoned due to technical limitations and to overwhelming competition from the semiconductor industry. The end of Moore's Law related improvements for CMOS coupled with the requirement to develop computers at the exascale has motivated the search for viable post-Moore technologies, suggesting another look at superconducting technology. The talk will discuss the IARPA Cryogenic Computing Technology program to develop a prototype superconducting computer, and related developments around the world.

- Erik DeBenedictis SNL 6/21/2016 Bldg 221/B145

Title: ***The Sensible Machine Grand Challenge***

Erik P. DeBenedictis, Sandia National Lab

Abstract

The Sensible Machine concept started with an OSTP RFI response by Stan Williams and grew over six months into a IEEE-led community activity advocating the advancement of computing in the direction of learning machines. The talk will discuss some of this history to motivate a technical discussion.

The talk will explain emerging ideas that could facilitate the Sensible Machine direction, with this talk concentrating on the architectural end of the technology stack rather than materials and physics. Through Moore's Law, the semiconductor industry focused on line width reduction for microprocessors and memory. However, a Sensible Machine should have logic and memory integrated down to the device level for speed and energy efficiency. While numerical computing will remain important, a Sensible Machine will make extraordinary use of computational primitives for learning and should be optimized for those. Sensible Machines could possibly drive a new growth path for the computer industry, yet the advancement may be in applications becoming increasingly sophisticated essentially due to the computer auto-programming software rather than just more efficient Boolean logic. There is also a path to improving traditional supercomputing through learning and neural networks.

- Steve Koester U of Minnesota 7/19/2016 Portrait Room/101

Dr. Koester received his Ph.D. in 1995 from the University of California, Santa Barbara. From 1997 to 2010 he was a research staff member at the IBM T. J. Watson Research Center and performed research on a wide variety of electronic and optoelectronic devices, with an emphasis on those utilizing the Si/SiGe material system. From 2006-2010 he served as manager of Exploratory Technology at IBM Research where his team investigated advanced devices and integration concepts for use in future generations of microprocessor technology. Since 2010, he has been a Professor of Electrical & Computer Engineering at the University of Minnesota where his research focuses on novel electronic, photonic and sensing device concepts with an emphasis on graphene and other 2D materials. Dr. Koester has authored or co-authored over 200 technical publications, conference presentations, and book chapters, and holds 65 United States patents. He is an associate director for the SRC/DARPA-funded center for spintronic materials interfaces and novel architectures (C-SPIN) and an associate editor for *IEEE Electron Device Letters*.

Title: ***Spintronics as an enabler for a new paradigm in computational Technology***

Since the emergence of "Moore's Law" scaling for semiconductor transistors 50 years ago, it has been recognized that the exponential increase of transistor density would ultimately come to an end. While this end-point has been extended much farther than originally thought, the limits of scaling are indeed upon us, and improvements in computational speed and efficiency are now mainly being achieved through novel architectures and new physics, rather than through shrinking device size. Looking toward the future, new physics of device operation will play and even more important role. For this reason, spintronics emerges as a promising platform for future computational systems, since it takes advantage of very different physics of device operation that can enable both improved energy efficiency as well as novel computing paradigms. In this talk, I will describe recent advances in spintronic materials and devices, and will show how new device concepts can be realized based upon physical phenomena such as the spin Hall effect, exchange bias interactions and magnetoelectric effects. I will also compare and contrast various novel spintronic architectures both for Boolean and neuromorphic computing platforms. Finally, throughout this talk, I will provide an overview of the ongoing research in C-SPIN, a vertically integrated research center funded by SRC and DARPA, which has the mission of advancing fundamental knowledge of spintronic materials, devices and novel architectures.

- Ian Foster U of Chicago 8/2/2016 Red Auditorium/101

Title:***Accelerating discovery via science services***

We have made much progress over the past decade toward harnessing thecollective power of IT resources distributed across the globe. In big-science projects in high-energy physics, astronomy, and climate,thousands work daily within virtual computing systems with global scope. But we now face a far greater challenge: Exploding data volumes and powerful simulation tools mean that many more—ultimately most?--researchers will soon require capabilities not so different from those used by such big-science teams. How are we to meet these needs? Must every lab be filled with computers and every researcher become an IT specialist? Perhaps the solution is rather to move research IT out of the lab entirely: to develop suites of science services to which researchers can dispatch mundane but time-consuming tasks, and thus to achieveeconomies of scale and reduce cognitive load. I explore the past, current, and potential future of large-scale outsourcing and automation for science, and suggest opportunities and challenges for today's researchers. I use examples from Globus and other projects to demonstrate what can be achieved.

- Ian Young Intel 8/16/2016 LR B/101

Tile: **Benchmaking of Beyond CMOS Logic**

- Jim Demmel UC Berkeley 9/6/2016 LR B/101

Title: Communication-Avoiding Algorithms for Linear Algebra and Beyond

Algorithms have two costs: arithmetic and communication, i.e. moving data between levels of a memory hierarchy or processors over a network. Communication costs (measured in time or energy per operation) already greatly exceed arithmetic costs, and the gap is growing over time following technological trends. Thus our goal is to design algorithms that minimize communication. We present algorithms that attain provable lower bounds on communication, and show large speedups compared to their conventional counterparts. These algorithms are for direct and iterative linear algebra, for dense and sparse matrices, direct n-body simulations, and some machine learning algorithms. Several of these algorithms exhibit perfect strong scaling, in both time and energy: run time (resp. energy) for a fixed problem size drops proportionally to the number of processors p (resp. is independent of p). Since for some emerging non-volatile memory technologies writes are much more expensive than reads, we also present write-avoiding algorithms, that also do asymptotically fewer writes than reads. Finally, we describe extensions to algorithms involving very general loop nests and array accesses, in a way that could be incorporated into compilers.

- Stan Williams HP 9/20/2016 Red Auditorium/101

Title: **Electrical and Physical Characterization of Nano- and Non-Linear Devices for Future Computing**

R. Stanley Williams

Senior Fellow

Hewlett Packard Labs

With the end of Moore's Law in sight, there is a great deal of angst in the information technology community over how computing can keep pace now that data is being generated and accumulated at an exponential rate. One solution is to perform exponentially more computation per unit of energy expended in a computer. This may very well require the exploitation of nonlinear dynamical systems to encode and process information in unconventional ways. Both nanoscale structures and neurons can display pathologically nonlinear responses such as chaos to a small stimulus, and in many ways the former can be used to emulate the latter. After a brief introduction to a couple of nonlinear electronic devices, i.e. passive or synaptic memristors and locally active or axonic memristors, I will describe the electronic and physical characterization tools and techniques that we have developed to characterize these systems. Standard electronic test and measurement systems are largely incapable of providing the appropriate time and/or frequency dependent information required to quantitatively characterize and model memristors. We have built flexible high-speed systems that enable us to watch the electronic switching in highly nonlinear systems in real-time from 10's of picoseconds to minutes. This type of data is critical to construct compact models for both the switching and the reliability of dynamical devices. We base our models, as much as possible, on the actual physical mechanisms that occur inside the devices as they operate. For this purpose, we have worked with both the Advanced Light Source at LBNL and the Stanford Synchrotron Radiation Laboratory to utilize the technique of Scanning Transmission X-ray Microscopy to examine functioning memristors *in situ* and *in operando* under controlled temperature and electrical bias conditions. We have imaged the structure and chemical composition of different conductance states of devices in order to determine what and how atoms move inside solid state devices as they are switching electrically under an external bias, whether that switching occurs via a phase transition or through drift, diffusion and thermophoresis of atomic species like oxygen. The electrical and mechanistic information come together in the compact device models that we supply to circuit architects so that they can faithfully and predictively simulate a wide range of circuits before they commit to a design that will be fabricated.

- Thom Dunning U of Washington 10/4/2016 Portrait Room/101

Title:***Leading-edge Computers and The Extraordinary Research They Enable***

Thomas H. Dunning Jr.

Northwest Institute for Advanced Computing, Pacific Northwest National Laboratory & University of Washington; Department of Chemistry, University of Washington, Seattle, Washington 98195; and Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Abstract

A new generation of supercomputers—*petascale computers*—is providing scientists and engineers with the ability to simulate a broad range of natural and engineered systems with unprecedented fidelity. Just as important, in this increasingly data-rich world, these new computers also allow researchers to manage, integrate and analyze unprecedented quantities of data, seeking connections, patterns and knowledge. The impact of this new computing capability will be profound, affecting science, engineering *and* society.

In 2013, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, with funding from the National Science Foundation, deployed a computing system that can *sustain* one quadrillion calculations per second on a broad range of science and engineering applications as well as manage and analyze petabytes of data. This computer, *Blue Waters*, has been configured to solve the most compute-, memory- and data-intensive problems in science and engineering. It has tens of thousands of chips (CPUs & GPUs), more than a petabyte and a half of memory, tens of petabytes of disk storage, and hundreds of petabytes of archival storage.

But, computer technology continues to move forward with the U. S. Department of Energy planning to deploy computer systems with *peak performances* of 100s of PFLOPs next year and 1,000s of PFLOPs in the early 2020s. However, this increase in performance can only be achieved with significant changes in the underlying computing technologies. This presents both and opportunity and a challenge for computational scientists and engineers. The presentation will describe these leading-edge computing systems, illustrate the role that they play or will play in a few areas of research, and describe the challenges facing the development of exascale modeling and simulation.

- Barry Schneider ACM/NIST 10/18/2016

Title: ***CyberScience and CyberInfrastructure: A New Approach to Discovery in Science and Engineering***

Virtually every scientific discipline has been impacted by the revolution of the digital age. Modeling and simulation and the extraction of useful information from increasingly large data sets have become the sine qua non for progress in science and engineering (S&E). Taking a broad view, it is obvious that without computation fields such as cosmology, general relativity, materials science(MD,MGI), nuclear, plasma, condensed matter and high energy physics (QCD) would be severely and negatively impacted. Other fields, such as atomic and molecular physics rely on computation to produce accurate atomic data on fundamental atomic and molecular properties as well as collision and photonic cross sections required for many practical applications. This is not only important in catalyzing new discovery but also for technological advancement. In addition, the size of the data sets that aerated being generated from instruments such as telescopes, particle accelerators and sensors are becoming so large thatit is impossible to extract any really meaningful results without large-scale computation and data analysis. I use the term cyberscience to denote this new approach to discovery in S&E. However, it is also the case that without accompanying developments in cyberinfrastucture (hardware, software, networking) this revolution would not have been possible.
In this talk I will provide an overview of some of the interesting developments that have been happening in CS and CI over the past few decades and speculate a bit on where things might be headed in the next decade as we face the challenges of Moores's law beginning to fail.

- Ed Seidel NCSA 11/1/2016 Bldg 221/B145

Title: ***Trends in Scientific Computing and Data***

Abstract:

Supercomputing has reached a level of maturity and capability where many areas of science and engineering are not only advancing rapidly due to computing power, they cannot progress without it. Detailed simulations of complex astrophysical phenomena, HIV, earthquake events, and industrial engineering processes are being done, leading to major scientific breakthroughs or new products that cannot be achieved any other way. These simulations typically require larger and larger teams, with more and more complex software environments to support them, as well as real world data. But as experiments and observation systems are now generating unprecedented amounts of data, which also must be analyzed via large-scale computation and compared with simulation, a new type of highly integrated environment must be developed where computing, experiment, and data services will need to be developed together. I will illustrate examples from NCSA's Blue Waters supercomputer, and from major data-intensive projects including the Large Synoptic Survey Telescope, and give thoughts on what will be needed going forward.

Bio:

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, and before that he was a senior research scientist leading the numerical relativity group at NCSA.

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 bachelor's degree in mathematics and physics from the College of William and Mary, a master's degree in physics at the University of Pennsylvania in 1983, and a doctorate in relativistic astrophysics at Yale University in 1988.

- John Martinis Google 11/15/2016 Boulder

John Martinis

Google & Department of Physics, University of California Santa Barbara

Title: ***Roadmap for building a quantum computer***

Abstract

I will overview the basic strategy and roadmap for the quantum-AI project at Google, which has the goal of building a useful quantum computer. For hardware, the key metric is building scalable qubits with 2-qubit gate errors below 0.1-0.2% [J.M.Martinis, NPJQI 1, 15005 (2015)]. For software, I will describe a new "quantum-supremacy" test that can demonstrate the exponential power of a quantum processor by checking its output with a classical computer, which is intractable for even the world's most advanced classical supercomputer beyond 42-50 qubits. We are working to perform this experiment in the next 2 years.

- Jack Dongarra University of Tennessee/ORNL 12/6/2016 221/B145

Title:*An Overview of High Performance Computing and Benchmark Changes for the Future.*

In this talk we examine how high performance computing has changed over the last 10-years and look toward the future in terms of trends. These changes have had and will continue to have a major impact on our numerical scientific software. A new generation of software libraries and algorithms are needed for the effective and reliable use of (wide area) dynamic, distributed and parallel environments. Some of the software and algorithm challenges have already been encountered, such as management of communication and memory hierarchies through a combination of compile--time and run--time techniques, but the increased scale of computation, depth of memory hierarchies, range of latencies, and increased run--time environment variability will make these problems much harder.

- Bob Wisnieff IBM 12/20/2016 101/Lecture Room B

Bob Wisnieff

IBM T. J. Watson Research Center

Title: ***New Technologies for Improved Computer Performance***

Tremendous progress has been made in building computer systems with higher performance of the last several decades, however the need to build systems with even higher performance exists in a number of key strategic areas. CMOS scaling will continue to be exploited to increase the areal density of devices and to reduce the power per operation beyond the 10nm node. For some applications like training neuromorphic systems and the modeling of quantum systems even more performance is required. Some of the increased performance can be obtained through re-architecting the system, using lower precision computations, using GPU like architectures with more inter thread communication bandwidth, and using FPGA like programmability. Beyond this device level innovation is required. We will review some of the progress that has been made in developing the materials and processes to enable performance increases at the device level.

- Gert Cauwenrberghs UCSD 1/17/2017 101/Lecture Room B

Title: ***Not Available***

- To Be Detremined 2/7/2017