also employ exascale computation to interpret and analyze exascale data anticipated from experiments in the near future.

Without these tools, scientific results would remain hidden in the data generated by these simulations. These tools will enable researchers to extract knowledge and share key findings.

Center for Data-Driven Discovery

A centerpiece of Brookhaven's computing initiative is the Lab's Center for Data-Driven Discovery (C3D), currently led by CSI Director Kerstin Kleese van Dam. C3D is driving the integration of intellectual, programmatic, and data/computational infrastructure, thereby accelerating and expanding discovery and focusing the tools and capabilities across the entire Laboratory into a single scientific resource.

The Center will draw on the leadership in scientific data and management analysis systems already in place at Brookhaven Lab and coordinate contributions from programs across the Laboratory to address four core missions:

- Provide a focal point for leading data science research
- Develop, test, and provide real-time data processing cyber-infrastructure for next-generation projects including new beamlines at NSLS-II and the proposed Electron-Ion Collider (EIC)
- Educate the next generation of expert data scientists
- Translate research advances into tools and expertise that lead to measurable scientific progress and improved industrial competitiveness, driving industrial collaborations with Laboratory users such as GE, Pfizer, and their research partners.

Hundreds of scientists from Brookhaven and thousands of facility users from universities, industry, and other laboratories around the country and throughout the world will benefit from the capabilities developed by C3D personnel to analyze, simulate, and predict complex phenomena important to many areas of science.

The missions of C3D and the overall CSI are well aligned with the broad missions and goals of many agencies and industries, especially those of DOE's Office of Science and its Advanced Scientific Computing Research (ASCR) program.

Strategic Partnerships

Computational scientists at Brookhaven will also seek to establish partnerships with key players in academia and industry (e.g. Stony Brook University's Institute for Advanced Computational Science, Rensselaer Polytechnic Institute, Oak Ridge National Laboratory, IBM, and Intel). One existing example of a successful partnership is the collaboration of Brookhaven Lab's high-energy and nuclear physics research groups with IBM that led to the development of the BlueGene supercomputing architecture now used on the world's most powerful commercially available supercomputers.

Likewise, the future development of capabilities for advanced big data analysis will have a tremendous impact on the competitiveness of scientific and business endeavors by paving the way to gain new insights ahead of competitors through faster and more efficient routes to discovery. The presence of C3D at Brookhaven Lab, and its collaborative work with New York State partners, will therefore make New York State research and industry more competitive by developing the expertise, tools, and infrastructure to be at the forefront of transformative data science.

Industrial partners may be particularly interested in how to interface big-data experimental problems (such as those that will be explored at NSLS-II, or in the fields of high-energy and nuclear physics) with high-performance computing using advanced network technologies. "Computing-system-on-a-chip" technologies, like that used by the IBM BlueGene series, open the door to customizing high-performance network interface cards and application program interfaces.

In addition, the development of asynchronous data access and transports based on remote direct memory access techniques and improvements in quality of service for network traffic could significantly lower the energy footprint for data processing while enhancing processing performance.

C3D together with planned improvements to Brookhaven National Laboratory office space and infrastructure (the Discovery Park project) will be a leading example on how to develop and attract new businesses that translate leading scientific research into widespread commercial success.

Contacts for more information

Kerstin Kleese van Dam
Director, Computational Science Initiative
phone: 631 344-6019
email: kleese@bnl.gov

Ada Palatnik
Human Resources Liaison for CSI
phone: 631 344-5875
email: apalatnik@bnl.gov

U.S. DEPARTMENT OF ENERGY

BROOKHAVEN NATIONAL LABORATORY

11/12/15

Brookhaven National Laboratory

Computational Science Initiative

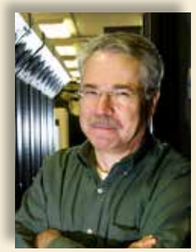
Leveraging Expertise and Investments to Translate "Big Data" into Discovery and Innovation

Computational Science Initiative

Leveraging Expertise and Investments to Translate "Big Data" into Discovery and Innovation

Building on its capabilities in data-intensive computing and computational science, the U.S. Department of Energy's (DOE) Brookhaven National Laboratory is embarking upon a major new Computational Science Initiative.

Kerstin Kleese van Dam, director of Brookhaven Lab's Computational Sciences Initiative and interim director of the Center for Data-Driven Discovery (C3D)



Brookhaven physicist Michael Ernst, head of the BNL Scientific Data and Computing Center, which includes the RHIC & ATLAS Computing Facility

Barbara Chapman, professor of Applied Mathematics & Statistics and Computer Science at Stony Brook University and head of the Brookhaven Lab Computer Science and Mathematics research team

Robert Harrison, chief scientist of Brookhaven Lab's Computational Science Initiative and director of Stony Brook University's Institute for Advanced Computational Science

The Computational Science Initiative (CSI) integrates computer science, mathematics, computational science, and domain science expertise and investments across Brookhaven Lab—including the flagship facilities that attract thousands of scientific users each year—to build the Laboratory's leadership in tackling the "big data" challenges at the frontiers of scientific discovery.

The timely analysis and interpretation of data is essential to addressing the nation's grand challenges in science, national security, and industry. CSI is investigating, in particular, novel approaches to harnessing the power of large data volumes and data rates to drive discovery. To this end, its activities are split into enabling and translational capabilities.

Enabling Capabilities

- Computer Science and Mathematics research into novel methodologies in support of large-scale, multi-modal, and streaming data analysis
- The BNL Scientific Data and Computing Center, housing the latest systems in high-performance and data-intensive computing, data storage, and networking, offering everything from novel research platforms to highly reliable production services

Translational Capabilities

- The Computational Science Laboratory, a collaborative space for the development of advanced algorithms and their characterization and optimization, also brings together computer scientists, mathematicians, and leading computational scientists to develop next-generation numerical simulation models
- The Center for Data Driven Discovery (C3D), a multi-disciplinary center for the development, deployment, and operation of data-intensive discovery services for science, national security, and industry

The CSI philosophy is a multi-disciplinary and collaborative approach to scientific research and development, with research targeted at and informed by the key challenges observed in close interactions with our clients in science, national security agencies, and industry. Our success is measured in equal parts by the advancement we can bring to computer science and mathematics, as well as by the transformational impact we have on our clients' mission space.

The CSI brings together under one umbrella the expertise that fosters cross-disciplinary collaboration and makes optimal use of existing technologies, while also leading the development of new tools and methods that will benefit science both within and beyond the Laboratory. Key partners include nearby universities such as Columbia, Cornell, New York University, Stony Brook, and Yale, as well as IBM Research.

Computer Science and Mathematics

The Computer Science and Mathematics research team is led by Barbara Chapman, a professor of Applied Mathematics & Statistics and Computer Science at nearby Stony Brook University. Its primary research subject is the development of advanced methods that enable large scale, multi-modal and streaming data analysis. The group takes a broad co-design approach, investigating novel techniques from the network and architecture level up to the application and workflow layers. The team's research is informed by the data-handling needs of Brookhaven Lab's large-scale scientific research facilities, such as the National Synchrotron Light Source II (NSLS-II), the Center for Functional Nanomaterials (CFN), and the Relativistic Heavy Ion Collider (RHIC), as well as the Laboratory's involvement in research at the European Large Hadron Collider (LHC) and other key projects such as the DOE's Atmospheric Radiation Measurement (ARM) Program and the Systems Biology Knowledge Base.

BNL Scientific Data and Computing Center

Brookhaven Lab has a strong history in the successful operation of large-scale computational science, data management, and analysis infrastructure. This expertise and tools developed at the Lab have been key factors in the success of the scientific programs at RHIC, NSLS-II, and CFN—all DOE Office of Science User Facilities—and also in biological, atmospheric, and energy systems science.

These capabilities have also been a crucial part of the Lab's participation in international research collaborations like the ATLAS experiment at the LHC, and will help make the case for building a future electron ion collider (EIC) at Brookhaven.

One example of Brookhaven's computing expertise is the RHIC & ATLAS Computing Facility (RACF). Formed in 1997 to support experiments at RHIC, Brookhaven's flagship particle collider for nuclear physics research, the RACF is now at the center of a global computing network connecting more than 2,500 researchers around the world with the data from RHIC and the ATLAS experiment at the Large Hadron Collider in Europe.

This world-class center houses an ever-expanding farm of computing cores (50,000 as of 2015), receiving data from the millions of particle collisions that take place each second at RHIC, along with petabytes of data generated by the LHC's ATLAS experiment—storing, processing, and distributing that data to and running analysis jobs for collaborators around the nation and the world.

The success of this distributed approach to data-intensive computing, combined with new approaches for handling data-rich simulations, has helped establish the U.S. as a leader in high-capacity computing, thereby enhancing international competitiveness. RACF will serve as a model for computing and data investigations under the new initiative, and as such will form the core of the new BNL Scientific Data and Computing Center, led by Brookhaven physicist Michael Ernst. The new center will also house Brookhaven Lab's institutional computing system, a range of data-intensive computing systems, as well as computing and data services operated for other third party clients.

Computational Science Laboratory

The Computational Science Laboratory, a new collaborative laboratory for advanced algorithm development and optimization led by Brookhaven computer scientist Nicholas D'Imperio, will bring together expertise in high-performance computing, math, and domain science. It will specifically address the challenge of developing new tools and techniques to deliver on the promise of exascale science—the ability to compute at a rate of 10^{18} (a billion billion) floating point operations per second (exaFLOPS). These computation rates are necessary to handle the copious amount of data created by computational models and simulations. The Computational Science Laboratory will

(continued on back)

Brookhaven's CSI pushes the evolution of equipment to meet the data needs of modern science.

Handling expanding data needs requires expertise across disciplines, from basic electronics to novel algorithms development.

Brookhaven's expertise in data handling, analysis, and curation continues to foster partnerships with academia and industry for the benefit of science and society.