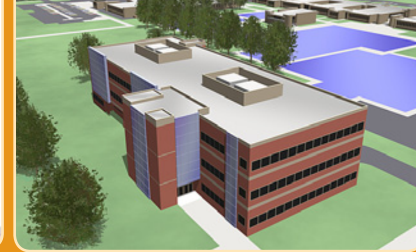
Computational Sciences & Mathematics

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by **Battelle** *Since 1965*



FROM BASIC RESEARCH TO DEPLOYABLE SOLUTIONS

The development of high-performance computers, the networks to connect them, and the scalable mathematics and software to run them are crucial for science-based solutions in energy, the environment, national security, and health. Computational Sciences & Mathematics at Pacific Northwest National Laboratory helps make this happen through innovative research to discover the mathematical algorithms and methods and the computer science and computational tools that enable researchers in many scientific disciplines to analyze, model, simulate, and predict complex phenomena.

The expertise of our experienced staff allows us to analyze and address the full range of needs, from basic research to deployable solutions. We focus on the following areas:

- *Data-intensive computing*. Our approach combines R&D in hybrid hardware architectures, adaptable software architectures, advanced user environments, and scalable analytic algorithms to provide end-users with real capabilities that enable scientific discovery and insight, decision support and control, and situational awareness and response. As part of the Advanced Simulation Capabilities for Environmental Modeling program, we are developing a user environment to make it easier for environmental scientists to use large amounts of data to create and simulate numerical models of groundwater flows. As part of the DOE Systems Biology Knowledge Base, we are investigating software integration and workflow approaches to enable biologists to work with a cloud-based, large-scale biological data repository.
- *Scientific data management*. We deliver solutions that facilitate data collection, processing, and storage through engineered systems and applied research. We focus on techniques to enable the management of peta-scale and larger scientific datasets. For example, we combine sensor data from a wide variety of heterogeneous instruments and derive Value-Added Products that reflect the measurements the scientists care most about; we develop and deploy workflows that operate in high-performance computing environments; and we have developed expertise in semantic data representations, metadata, and data provenance that has been applied to a variety of problems and programs.
- *High-performance computing*. Our approach merges science and technology by employing hardware that maximizes processor speed, memory and interconnect bandwidth, efficient use of secondary storage, and reliability; developing algorithms that are scalable, resource-efficient, load-balanced, and manages computational complexity and exploits space-time locality; and creating programming models, numerical libraries, communication libraries, compilers, and debuggers that support data decomposition, low communication overhead, and portability. One of our award-winning programming models, Global Arrays, enables researchers to more efficiently access global data, run bigger models, and simulate larger systems, resulting in a better understanding of the data and processes being evaluated.
- *Applied mathematics*. Scientific research and development (R&D) is a process of gaining fundamental understanding of physical, chemical, behavioral, and biological principles through computational modeling, experimentation, and data evaluation. As a leader in applied mathematics research, we develop and leverage mathematical models to quantify and control scientific uncertainty to further scientific discovery. PNNL's applied mathematics research portfolio includes: Mathematics for peta-scale data analysis from designed experiments; uncertainty quantification for systems of stochastic equations; and development of novel modeling approaches, analysis algorithms, and data evaluation methods for challenging multi-disciplinary research domains, including materials selection and design, human behavior, large-scale text analysis, and subsurface modeling.
- *Computational biology and bioinformatics*. We enable fundamental advances in the predictive modeling of biological networks and biomolecular systems through the application of high-performance computing. Our unique approaches involve the use of integrated genome-level biological data for network reconstruction and high-performance computing for large-scale, advanced data analytics. Our R&D efforts are focused on predictive modeling and simulation of biological systems and the development of data management, integration, and visualization software. Our interdisciplinary approach directly involves domain scientists to help design intuitive tools that precisely target the unique needs of the biological sciences.
- *Computational materials and informatics*. We develop advanced multi-scale computational methods aimed at understanding and predicting the macroscopic behavior and performance of materials under different (including extreme) environments. Our strength is in developing multi-physics modeling approaches that analyze a material's performance over multiple size and temporal scales and optimize this performance from a systems perspective. Materials informatics approaches are used to ascertain how materials properties vary with structure by creating a knowledge-based hierarchy-connecting theory, experimental data, and calculated data. The combination of direct simulation and derived materials knowledge techniques forms the basis of the materials discovery process.

Key Clients

DOE Office of Advanced Scientific Computing Research
DOE Office of Biological and Environmental Research
DOE Office of Energy Efficiency and Renewable Energy
U.S. Department of Homeland Security
U.S. Department of Defense
National Institutes of Health

Scientific Accomplishments

Computational Sciences and Mathematics' science and technology accomplishments and impacts have been significant and include the following:

Science: Researchers are elucidating the key functions of microbial communities found in the environment. They are developing unique computational biology methods to identify proteins and pathways active in microbial communities when there is no information on the genome/gene sequences of these microbes. Also, through CASS-MT, researchers are accelerating the frontier of high-performance computing and speeding solutions for bioenergy.

Energy: Researchers are gaining a greater understanding of electric grid failures through high-performance computing. Using PNNL's Cray XMT machine, researchers are conducting advanced power grid contingency analysis to help determine the impact of potential component failures. Research results show superior performance in selecting critical contingency cases from a large candidate set of possibilities. Without contingency selection, the number of cases would be so large as to exceed the capability of computational resources.

Health: Funding from the National Institutes of Health (NIH) is enabling researchers to explore the process involved in developing infections caused by the influenza and SARS viruses and the pathogenic bacteria *Yersinia* and *Salmonella*. Scientists at PNNL will provide modeling and analysis of the generated high-throughput data. The discoveries from the research could provide a foundation for more precise drug therapies and control strategies for a wide range of viral and bacterial pathogens.

Environment: Researchers are helping to improve modeling and understanding the role of clouds in global climate change by achieving an input/output (IO) bandwidth of 5 Gigabytes/sec for writing output from a global atmospheric model to shared files. The increase in IO bandwidth will allow researchers to run models at higher resolution, thus achieving higher accuracy and enabling simulations representing longer periods of time. Both are crucial to understanding future climate change.

National Security: In the fight against terrorism, analysts are bombarded with enormous volumes of data from documents, images, numbers, video, and audio. The National Visualization and Analytics Center™ at PNNL, established by the DHS, provides leadership and coordination among academia, industry, national laboratories, and government to create and deploy visual analytics technologies to help users analyze and interpret huge volumes of data.

Tools for Discovery—A Sampling

Global Arrays Toolkit dramatically simplifies writing code for supercomputers. The toolkit helps scientists to translate their ideas into highly efficient software that allows mathematical computations to run independently using subsets of processors of the supercomputer.

MeDICi: The PNNL-developed MeDICi is a middleware platform (computer software that connects software components or applications) that makes it easy to integrate separate codes into complex applications that operate as a data analysis pipeline. The MeDICi makes it easy to translate data as it moves from one application to another, turning a set of distributed heterogeneous components into an integrated pipeline.

STOMP: short for Subsurface Transport Over Multiple Phases, is a software tool that provides multidimensional modeling of subsurface flow and reactive transport phenomena. STOMP can accurately track and predict the subsurface migration of specific materials over time, and over multiple phases, for many applications.

Investing in the Future

Next-generation or extreme-scale computing is essential to enable breakthroughs in areas such as accurate weather and climate prediction, biotechnology, and new energy sources. To maintain our current leading scientific capabilities and build new ones to support client missions, the Laboratory is investing in the following activities:

- The eXtreme-Scale Computing Initiative: A Lab-wide initiative focused on developing an integrated research program with an interdisciplinary approach that brings together high-performance computer science, computational domain science, and applied mathematics to develop next-generation, extreme-scale modeling and simulation applications.

- Data-Intensive Computing Research:, Research that is concerned with capturing, managing, analyzing, and understanding data at volumes and rates that push the frontiers of current technologies. PNNL's approach to data-intensive computing is focused on three key research areas: hybrid hardware architectures, software architectures, and analytic algorithms.

State-of-the-Art Facilities and Equipment

PNNL is the steward of many resources for use in achieving the innovative research needed for scientific breakthroughs. These resources include:

ARM Climate Research Facility

The ARM Climate Research Facility (ACRF) including highly instrumented ground stations, a mobile facility, and an aerial vehicles program for studying cloud formation processes and their influence on climate change.

Center for Adaptive Supercomputing Software-Multithreaded Architectures (CASS-MT)

CASS-MT is dedicated to research on systems software, programming environments, and applications in a high-performance computing multithreaded architecture environment. The center offers the only Open-Science Cray XMT system, a one-of-a-kind supercomputer consisting of 64 multithreaded processors, 512 GB RAM, and a 4.3-TB Lustre parallel file system. Multithreaded processors allow multiple, simultaneous processing, helping researchers find solutions to the world's most complex challenges faster.

Chinook

The Chinook supercomputer has a peak performance of 160 teraflops. It currently has 37 terabytes of RAM, 840 terabytes of local disk, and 250 terabytes of a global shared file system. The system is connected via multiple 10-Gigabit Ethernet connections, allowing EMSL users to perform remote visualizations and transfer data to remote storage.

Computational Sciences Facility

Opened in 2009, the 75,600-square-foot Computational Sciences Facility houses computing capabilities in information and visual analytics as well as high-performance computing for extracting knowledge from massive amounts of heterogeneous data.

Electricity Infrastructure Operations Center (EIOC)

Through the EIOC, researchers are exploring ways to improve the reliability, lower costs, and lessen environmental impacts of the nation's electrical grid. The focus is on developing real-time tools and supporting their integration into operating systems.

Energy Smart Data Center (ESDC)

The ESDC employs state-of-the-art supercomputer technology, characteristic in size and energy consumption of supercomputers used in typical data centers, making it an ideal testbed for independent evaluation and validation of new technologies and equipment. The information collected can then be used by PNNL, along with its customers and partners, to design the most energy-efficient approaches to power and cooling.

At the ESDC, Andres Marquez is leading efforts to study data centers' power generation, conversion, and distribution as well as cooling challenges.

About Pacific Northwest National Laboratory

PNNL is a U.S. Department of Energy Office of Science research facility that delivers breakthroughs in science and science-based solutions for energy, environment, national security, and health. Battelle, based in Columbus, Ohio, has operated PNNL since 1965. PNNL has a long and successful track record of working with industry clients. PNNL is located in Richland, WA. Additional web resources are at http://www.pnl.gov.

For More Information, contact:

Dr. Moe Khaleel, Director
Computational Sciences
& Mathematics Division
Pacific Northwest National Laboratory
P.O. Box 999, K9-80
Richland, WA 99352
Phone: 509.375.2438
Email: moe.khaleel@pnl.gov

▸ www.pnl.gov

Pacific Northwest NATIONAL LABORATORY | U.S. DEPARTMENT OF ENERGY

Proudly Operated by Battelle *Since 1965*

PNNL-SA-70521 | February 2010
This document was printed on recycled paper.