

Public release date: 14-Jul-2008

[[Print Article](#) | [E-mail Article](#) | [Close Window](#)]



Contact: Mary Beckman
mary.beckman@pnl.gov
509-375-3688
[DOE/Pacific Northwest National Laboratory](#)

Multithreaded supercomputer seeks software for data-intensive computing

New collaboration to develop software for advanced supercomputers

RICHLAND, Wash. -- The newest breed of supercomputers have hardware set up not just for speed, but also to better tackle large networks of seemingly random data. And now, a multi-institutional group of researchers has been awarded \$4.0 million to develop software for these supercomputers. Applications include anywhere complex webs of information can be found: from internet security and power grid stability to complex biological networks.

The difference between the new breed and traditional supercomputers is how they access data, a difference that significantly increases computing power. But old software won't run on the new hardware any more than a PC program will run on a Mac. So, the Department of Defense provided the funding this month to seed the Center for Adaptive Supercomputing Software, a joint project between the Department of Energy's Pacific Northwest National Laboratory and Cray, Inc, in Seattle.



The Center for Advanced Supercomputing Software will take advantage of the multithreaded processors in the Cray XMT.

[Click here for more information.](#)

"The system will allow much faster analysis of complex problems, like understanding and predicting how the power grid behaves -- one of the most complex engineering systems ever built," said Moe Khaleel, director of Computational Sciences and Mathematics at PNNL, which is leading the project.

Other researchers in the software collaboration hail from Sandia National Laboratories, Georgia Institute of Technology, Washington State University and the University of Delaware.

These new machines are built with so-called "multithreaded processors" that enable multiple, simultaneous processing compared with the linear and slower approach of conventional systems. The Center will focus on applications for the multithreaded Cray XMT, one of which Cray delivered to PNNL in September 2007 (<http://www.pnl.gov/news/release.asp?id=271>).

"Traditional supercomputers are not well suited for certain kinds of data analysis, so we want to explore this advanced architecture," said PNNL computational scientist Daniel Chavarría.

In previously published work, PNNL computational scientist Jarek Nieplocha used a predecessor of the Cray XMT to run typical software programs that help operators keep the power grid running smoothly. Adapted to the advanced hardware, these programs ran 10 times faster on the multithreaded machine. "That was the best speed ever reported. We're getting closer to being able to track the grid in real time," said Nieplocha.

In biology, another complex web is woven by genes (or their protein products) working together inside people's cells. "We have discovered genes implicated in breast cancer using a massively multithreaded algorithm on the Cray XMT," said Georgia Tech computational scientist and engineer David A. Bader. "It's like finding a needle in a haystack. The algorithm

searches for genes whose removal quickly causes networks and pathways in the cell to breakdown."

The processors and computer memory in the advanced computers interact in a novel way. In traditional supercomputers, each processing chip gets a dollop of memory to use for its computations. To perform a calculation, the chip dips into the memory, does its work, then accesses the memory again for its next calculation, like an elephant dipping its trunk into a bag of peanuts and eating them one at a time. Each processor-memory unit is linked together over a network, and performance improvements come with more and faster processors and sleek network connections.

The Cray XMT multithreaded system lumps all the memory together, and the processors freely access the much larger memory pool. But like an elephant with many trunks, each processor has multiple threads: it dips into memory with one thread, and while that thread is performing the calculation at hand, another thread goes into the memory, and another.

By the time all the threads have dipped, the original thread has finished its calculation and is ready for another trip to the memory bank. A many-trunked elephant would have a distinct speed advantage plowing through a bag of peanuts over its hungrier zoo-mate, just as a multithreaded system does.

"The processors are doing useful work all the time, so the computer can be faster," said Chavarría. Each Cray XMT processor has 128 hardware threads with which to access the shared memory.

Conceptually, this advantage translates into the machines being able to handle complex, random networks of data. Mainstream machines split up the data, assigning parcels of data to individual processing units. For example, a supercomputer trying to model how a community of microbes behaves would subdivide the community spatially.

The computer would then analyze what goes on within each subdivision, but it couldn't reach across other subdivisions to find out what happened to the microbe that wandered off to the other side of its habitat. Multithreaded machines, however, can examine the whole space at once, essentially assigning each thread to a microbe.

"If all of your microbes move to the other side of the territory, it doesn't matter, because the threads still have access," said Chavarría.

Another advantage multithreaded machines have over mainstream computers is in power consumption. Although the Cray has not yet been tested, other multithreaded machines have shown reduced energy usage compared to traditional architectures.

###

Additional contacts:

Stefany Wilson, stefany@cc.gatech.edu, 404-894-7253
Christopher Miller, cmiller@sandia.gov, 505-844-0587

The Computational Science & Engineering division at the Georgia Institute of Technology (CSE) was established in 2005 to strengthen and better reflect the critical role that computation plays in the science and engineering disciplines. CSE supports interdisciplinary research and education in computer science and applied mathematics. The Georgia Tech CSE program is designed to innovate and create new expertise, technologies, and practitioners in areas including high performance and grid computing, modeling and simulation, and data analysis and mining. Visit CSE at <http://cse.cc.gatech.edu/>.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin

company, for the U.S. Department of Energy's National Nuclear Security Administration. With main facilities in Albuquerque, N.M., and Livermore, Calif., Sandia has major R&D responsibilities in national security, energy and environmental technologies, and economic competitiveness.

Pacific Northwest National Laboratory is a Department of Energy Office of Science national laboratory where interdisciplinary teams advance science and technology and deliver solutions to America's most intractable problems in energy, national security and the environment. PNNL employs 4,000 staff, has a \$855 million annual budget, and has been managed by Ohio-based Battelle since the lab's inception in 1965. Visit PNNL at <http://www.pnl.gov/news>

[[Print Article](#) | [E-mail Article](#) | [Close Window](#)]

