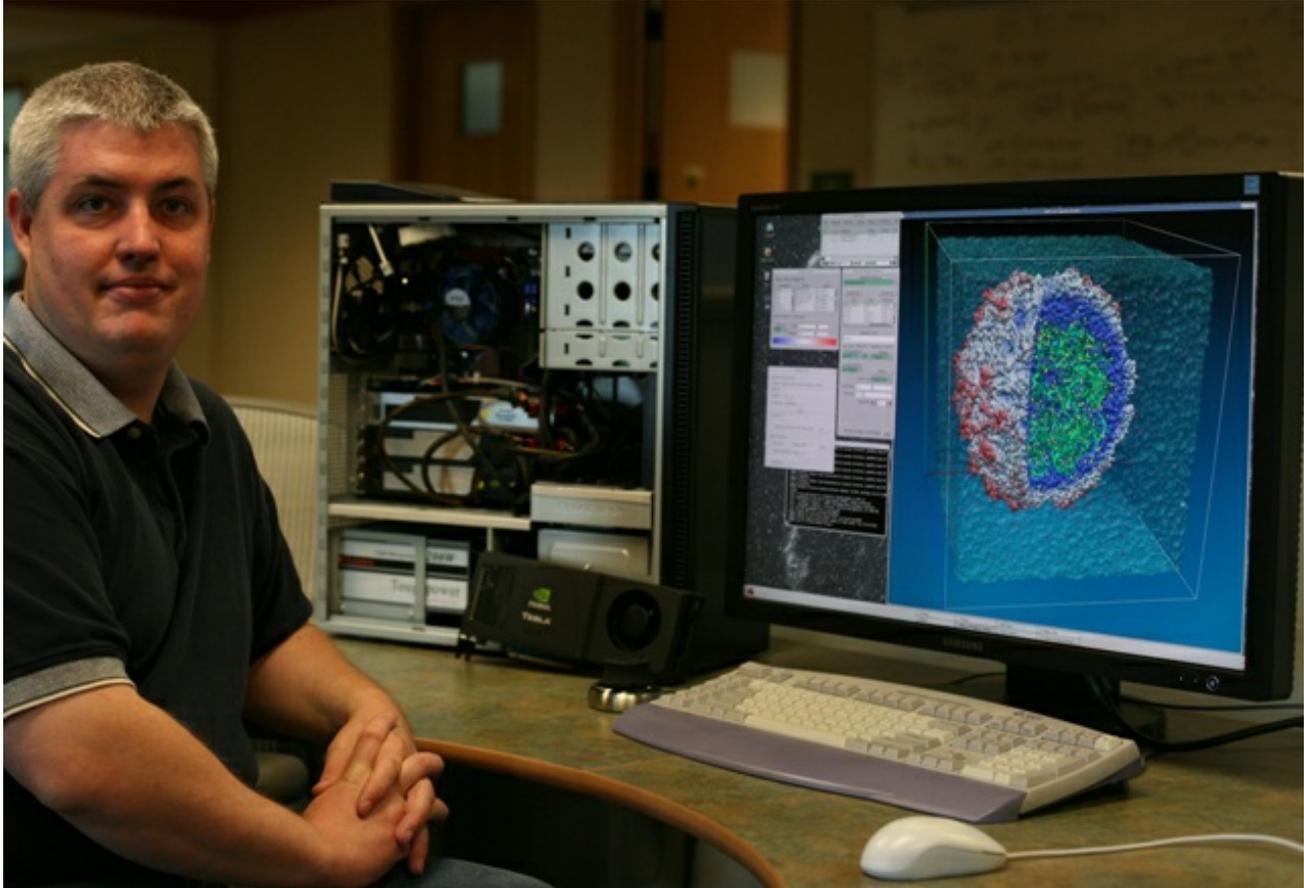# Personal Supercomputers Promise Teraflops on Your Desk

By Priya
Ganapati

About a year ago John Stone, a senior research programmer at the University of Illinois, and his colleagues found a way to bypass the long waits for computer time at the National Center for Supercomputing Applications.

Stone's team got "personal supercomputers," compact machines with a stack of graphics processors that together pack quite a punch and can be used to run complex simulations.

"Now instead of taking a couple of days and waiting in a queue, we can do the calculations locally," says Stone. "We can do more and better science."

Personal supercomputers are available in many flavors, both as clusters of CPU and graphics processing units (GPUs). But it is GPU computing that is gaining in popularity for its ability to offer researchers easy and quick access to raw computing power. That's opening up a new market for makers of GPUs, such as Nvidia and AMD, which have traditionally focused on high-end video cards for gamers and graphics pros.

True supercomputers, the rock stars of computing, are capable of millions of calculations per second. But they can be extremely expensive — the fastest supercomputer of 2008, IBM's RoadRunner, costs $120 million — and access to them is limited. That's why smaller versions, no bigger than a typical desktop PC, are becoming a hit among researchers who want access to massive processing power along with the convenience of having a machine at their own desk.

"Personal supercomputers that can run off a 110 volt wall circuit allow for a significant amount of performance at a very reasonable price," says John Fruehe, director of business development for serve and workstation at AMD. Companies such as Nvidia and AMD make the graphics chips that personal supercomputer resellers assemble into personalized configurations for customers like Stone.

Demand for these personal supercomputers grew at an average of 20 percent every year between 2003 and 2008, says research firm IDC. Since Nvidia introduced its Tesla personal supercomputer less than a year ago, the company has sold more than 5,000 machines.

"Earlier when people talked about supercomputers, they meant giant Crays and IBMs," says Jie Wu, research manager for technical computing at IDC. "Now it is more about having smaller clusters."

Today, most U.S. researchers at universities who need access to a supercomputer have to submit a proposal to the National Science Foundation, which funds a number of supercomputer centers. If the proposal is approved, the researcher gets access to an account for a certain number of CPU hours at one of the major supercomputing centers at the universities of San Diego, Illinois or Pittsburgh, among others.

"Its like waiting in line at the post office to send a message," says Stone. "Now you would rather send a text message from your computer rather than wait in line at the post office to do it. That way it is much more time efficient."

Personal supercomputers may not be as powerful as the mighty mainframes, but they are still leagues above their desktop cousins. For instance, a four-GPU Tesla personal supercomputer from Nvidia can offer 4 teraflops of parallel supercomputing performance with 960 cores and two Intel Xeon 5500 Series Nehalem processors. That's just a fraction of the IBM RoadRunner's 1 petaflop speed, but it's enough for most researchers to get the job done.

For researchers, this means the ability to run calculations faster than they can with a traditional desktop PC. "Sometimes researchers have to wait for six to eight hours before they can have the results from their tests," says Sumit Gupta, senior product manager at Nvidia. "Now the wait time for some has come down to about 20 minutes."

It also means that research projects that typically would have never get off the ground because they are deemed too costly and too resource and time intensive now get the green light. "The cost of making a mistake is much lower and a lot less intimidating," says Stone.

The shift away from large supercomputers to smaller versions has also made research more cost effective for organizations. Stone, who works in a group that develops software used by scientists to simulate and visualize biomolecular structures, says his lab has 19 personal supercomputers shared by 30 researchers. "If we had what we wanted, we would run everything locally because it is better," says Stone. "But the science we do is more powerful than what we can afford."

The personal supercomputing idea has also gained momentum thanks to the emergence of programming languages designed especially for GPU-based machines. Nvidia has been trying to educate programmers and build support for CUDA, the C language programming environment created specifically for parallel programming the company's GPUs. Meanwhile, AMD has declared its support for OpenCL (open computing language) this year. OpenCL is an industry standard programming language. Nvidia says it also works with developers to support OpenCL.

Stone says the rise of programming environments for high performance machines have certainly made them more popular. And while portable powerhouses can do a lot, there is still place for the large mainframe supercomputers. "There are still the big tasks for which we need access to the larger supercomputers," says Stone. "But it doesn't have to be for every thing."

Photo: John Stone sits next to a personal supercomputer- a quad-core Linux PC with 8GB of memory and 3 GPUs (one NVIDIA Quadro FX 5800, and two NVIDIA Tesla C1060) each with 4GB of GPU memory/ Kirby Vandivort