Georgia Tech engaged in $100 million next-generation high-performance computing initiative
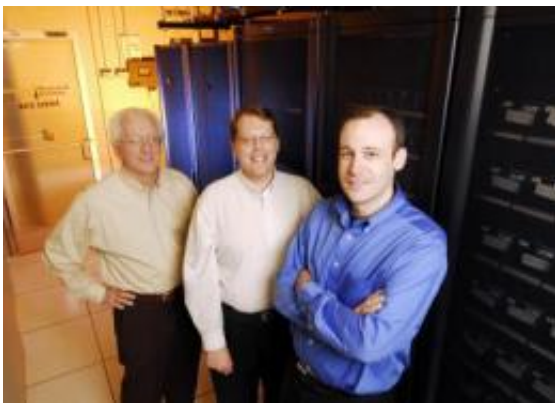
November 8 2010

Researchers Mark Richards, David Bader and Dan Campbell (left-to-right) pose in the Advanced Computing Technology Lab operated by the Georgia Tech Research Institute. Credit: Gary Meek

Imagine that one of the world's most powerful high performance computers could be packed into a single rack just 24 inches wide and powered by a fraction of the electricity consumed by comparable current machines. That would allow an unprecedented amount of computing power to be installed on aircraft, carried onto the battlefield for commanders – and made available to researchers everywhere.

Putting this computing power into a small and energy-efficient package, and making it reliable and easier to program, are among the goals of the

new DARPA Ubiquitous High Performance Computing (UHPC) initiative. Georgia Tech researchers from three different units are supporting key components of this $100 million challenge, which will require development of revolutionary approaches not bound by existing computing paradigms.

If UHPC meets its ambitious eight-year goals, the new approaches and technologies it develops could redefine the way that computing systems are envisioned, designed and used.

"The opportunity we have is to go far beyond the current product roadmaps," said David Bader, a professor in Georgia Tech's School of Computational Science and Engineering. "We really have the opportunity to change the industry and to design our applications with new computing architectures. For the first time in the history of computing, we will be able to work with a clean slate."

To attain the program's ambitious goals, DARPA funded four groups – led by NVIDIA Corp., Intel Corp., the Massachusetts Institute of Technology and Sandia National Laboratories – to develop UHPC prototypes. A fifth group, led by the Georgia Tech Research Institute (GTRI), will develop applications, benchmarking and metrics that will be used to drive UHPC system design considerations and support performance analysis of the developing system designs.

"Our team is developing a set of five difficult problems of a size and scope that the machines they are talking about should be able to accomplish," said Dan Campbell, a GTRI principal research engineer who is co-principal investigator of the benchmarking initiative. "Our challenge is picking the right problems and specifying them at the right level of abstraction to allow innovation and properly represent what the DoD will need in 2018."

The five problems highlight the unique computing needs of the U.S. military:

- Analysis of the vast streams of data originating with widespread sensor systems, unmanned aerial vehicles and new generations of radar systems. The data will be analyzed for nuggets of useful information in ways that are not possible today.

- A dynamic graph challenge, in which many entities interact to create a problem of "connecting the dots." That could mean analyzing relationships in social media to find possible adversaries, or understanding network traffic for cyber-security challenges.

- The decision tree, comparable to a chess game in which many possible interconnected options, each with complex implications, must be analyzed quickly. This could help field commanders or corporate CEOs make better decisions.

- Materials shock and hydrodynamics issues, challenges important to improving future generations of materials.

- Molecular dynamics simulations, which use high-performance computers to understand interactions between very large systems, such as protein folding.

"We need to be able to take in a lot more data and understand it a lot more thoroughly than we can now," said Mark Richards, a principal research engineer in the Georgia Tech School of Electrical and Computer Engineering and co-principal investigator of the benchmarking team. "That might allow us to find adversaries we can't find now because we're unable to tease that information out of the data

flow."

While the benefits of making such computing power widely available are obvious, how these machines will be designed, built and reliably operated is not.

"Meeting these very ambitious program goals will pose significant technical challenges," said Bader, who leads application development on the NVIDIA team and is part of the benchmarking group. "The technology roadmaps in such areas as interconnection networks, microprocessor design and technology fabrication will be pushed to their limits."

Meeting power limitations of just 57 kilowatts per rack – the amount of electricity produced by a portable military generator – may be the toughest among them. The fastest computer currently in operation requires seven megawatts of power.

"Reducing the power consumption means less energy per computation," noted Richards. "But as we lower the device voltage, we get closer to the physical noise. That will allow more errors due to the physics of the devices, and all kinds of things will have to be done to address that."

And the entire machine will have to fit into a 24-inch wide, 78-inch high and 40-inch deep cabinet.

But the physical implementation of the machines is just one part of the challenge, Bader noted. How people will work with them poses a perhaps more difficult challenge because it will require thinking about computers in a new way.

"Over the past 20 or 30 years, we've taken a single computing design and kept tweaking it through advances like miniaturizing parts," he said. "But

we really haven't changed the global nature of how the machine works. To meet DARPA's power efficiency goals, we really will need to change the way we program the machine."

That also affects the humans who interact with these highly-parallel machines, which could have as many as a half-million separate threads operating at the same time. DARPA's initial goal is to build machines capable of petaflop speed – a trillion operations per second – which could lead into the next generation of exascale computers a thousand times more capable.

"We will need to find new ways of thinking about computers that will make it feasible for humans to comprehend what is going on inside," Campbell said. "It's a huge programming challenge."

To encourage collaboration in solving these complex problems, DARPA has embraced the idea of open innovation. It expects the organizations to work together on common critical topics, creating a collaborative environment to address the system challenges. New technology generated by the program – believed to be today's largest DoD computing research initiative – is likely to move quickly into industry.

"There is certainly an expectation among the companies that what they are doing in this project is going to change how we do mainstream computing," Bader said. "The technology transfer implications are certainly obvious."

Provided by Georgia Institute of Technology