



## News

[Press Releases](#)  
[Event Announcements](#)  
[CCT Weekly](#)  
[Grants and Funding](#)  
[Student News](#)  
[Archived News](#)

## GPGPUs, China Look to Build on Supercomputing Successes of 2010

(Source: [HPCwire](#))

Rather than go through a litany of predictions about the high performance computing space for 2011, it's probably more instructive to point to two clear trends established in 2010 that are set to change the face of HPC over the next 5 to 10 years: the rise of GPGPU supercomputing and the emergence of China as an HPC superpower.

As we [reported last year](#) GPU-powered petaflop machines and China supercomputing arose simultaneously over the last two years. The Asian nation has been most aggressive in taking advantage of GPGPUs to build big machines. This culminated in China grabbing the number one spot on the November 2010 TOP500 list with the 2.5 petaflops Tianhe-1A supercomputer, powered by NVIDIA Tesla GPUs.

The trends are just beginning though. As of today, only 11 of the top 500 systems are GPGPU-powered, representing just 2 percent of the total. But 3 of those are in the top 5, and 7 are in the top 100. A year before, there were just two GPU-equipped supercomputers on the entire list. The US and Europe, which developed much of the foundational hardware and software technology, have so far lagged in adopting GPGPU to build petascale machines.

As far as national presence, the US is still by far the dominant supercomputing power with 282 of the top 500 systems. China claims just 41 systems, which places it a distant second. But that's up from just 21 top machines in 2009. It's noteworthy that no other country in the list is accelerating its presence so rapidly.

The expectation is that both these trends will continue in 2011 and beyond. At the recent "Analyst Crossfire" forum held at SC10 in New Orleans (YouTube videos of which can be found in three parts: [here](#), [here](#), and [here](#)), the topics of GPGPU computing and China's emergence as a major HPC player were discussed at length. The four panelists -- Thomas Sterling, Professor of Computer Science at LSU, Jay Boisseau, Director of the Texas Advanced Computing Center (TACC), Peter ffolkes, VP Marketing Adaptive Computing; and Michael Wolfe, Sr. Compiler Engineer at PGI -- generally agreed that China and GPU supercomputing will solidify the gains made they made in 2010, but that the long-term prospects for GPGPUs are less assured.

TACC's Boisseau and Adaptive's ffolkes noted the GPGPU technology to build petascale supercomputers is available to everyone, but it was the Chinese that brought their resources to bear first. (Actually, the Japanese with their TSUBAME system were the true early adopters of GPGPU supercomputing, culminating in the number 2 ranked TSUBAME 2.0 machine in November 2010). The Chinese, though, moved purposefully to build a number of petascale or near-petascale machines, determining that GPGPU technology was the shortest path to that end. "I think they [the Chinese] are going to be a fixture in the TOP500 and especially even in the top 10 for several years -- and onward," said Boisseau.

The Chinese currently have 5 of their top 41 machines in the top 100 and are looking to build on that as they [build out their supercomputing infrastructure](#). "They are rapidly becoming a major political power in the world and HPC is part of that," said ffolkes.

LSU's Sterling noted that the Chinese are not simply buying their way into the upper echelons of supercomputing. The top Tianhe-1A machines is based on a home-grown system design, sporting a custom network interconnect and I/O processor developed from the silicon on up. "Unlike us, they have a long tradition of five year plans, and they stayed the course," said Sterling. And they will do so here as well. They have no doubt through the procedures and methods they've applied that they will be the leader in this field before the end of this decade." Indeed, China has stated it wants to be the first nation to field an exascale machine.

From Sterling's perspective the US is in deep crisis with respect to HPC technology research. "The Democrats don't understand the technology and the Republicans don't fund it unless it's for defense research," he said. "Blue Waters will come online shortly and will act as a placebo. We'll again be at the top of the list and we'll think we're doing fine, when in fact we are way behind in doing research for the future machines for the end of the decade."

Whether those machines will be GPGPU-based is less certain. Other accelerators like Intel's Many Integrated Core (MIC) processor won't debut for another year or so. And AMD's [Fusion \(CPU-GPU\) processors are just now making their way into client side of the ecosystem](#). Even [NVIDIA's roadmap for its discrete GPUs over the next couple of years](#) will likely produce something that is unrecognizable as a graphics processor of the last decade.

It's even possible that accelerators, GPUs or otherwise, will not figure prominently in the biggest machines. PGI's Wolfe noted that the only two publicly announced 10 petaflop systems -- the Power7-based Blue Waters in the US and the Sparc64 VIIIx-based Kei Soku Keisanki (aka the "K computer") in Japan -- will rely solely on CPUs, albeit very high-end ones.

Wolfe thinks Intel's MIC is a "fascinating architecture," which has the potential to unseat NVIDIA's current dominance in HPC acceleration arena. And AMD (ATI) GPUs have the raw performance advantage, he says. That could be the decisive factor once a more level playing field for GPGPU middleware is in place, which is what AMD is banking on with the open standard OpenCL API. That technology is even more attractive when seen against AMD's CPU-GPU Fusion roadmap, which will eventually wind its way into the server side of the business.

From ffolkes' perspective, the GPU will be just one of a number of accelerator architectures that drive future HPC machinery. Eventually though, he believes that accelerators will follow the same path as floating point units (FPUs), which used run free as discrete coprocessors before being integrated onto the CPU. Integrating accelerator logic into a standard processor will make the technology ubiquitous and essentially free, according to ffolkes. "And then the key becomes programmability," he said.

Sterling is even more circumspect about the GPGPU's longevity in HPC, at least in its current form. "With respect to the GPU, it's the flavor of the month," he said. "We've been here before with attached array processors."

Like others, Sterling believes heterogeneous architectures will be the model of the future, but the accelerator componentry will eventually be integrated on-chip, a la Fusion. That will solve much of the latency and bandwidth issues that currently limit performance on the PCI-connected discrete GPUs. It will also simplify the programming model.

"What we will develop is a uniform execution model for a non-uniform system architecture that scales," said Sterling. "That's when we will have arrived."

**Publish Date:**  
01-05-2011



Center for Computation & Technology

2003 Digital Media Center • Telephone: +1 225/578-5890 • Fax: +1 225/578-8957

© 2001–2025 Center for Computation & Technology • Official Web Page of Louisiana State University.