**Comparing GPUs and CPUs**

http://www.hpcwire.com/hpcwire/2011-03-29/comparing\_gpus\_and\_cpus.html

Tiffany Trader, Associate Editor, HPCwire

A feature article at the TeraGrid website takes look at the most common building blocks of today's supercomputers, the ubiquitous CPUs and GPUs. Interest in GPUs was already high when China's Tianhe-1A supercomputer achieved a number one TOP500 ranking using the power of the graphics chips. With that success, many in the HPC community are wondering what GPU computing can do for them. Author Jan A. Zverina addresses this question and examines how GPU computing fits into the overall HPC landscape.

While for some applications GPUs can offer 20x performance increases or more over CPUs, that doesn't mean they are always the right choice. Peter Varhol, a contributing editor for the online magazine *Desktop Engineering* (DE), sums up the challenge:

*"The GPU remains a specialized processor, and its performance in graphics computation belies a host of difficulties to perform true general-purpose computing. The processors themselves require rewriting any software; they have rudimentary programming tools, as well as limits in programming languages and features."*

In addition to the programming challenge, there's also the communications bottleneck, getting data to and from the GPU. Ross Walker, an assistant research professor with the San Diego Supercomputer Center (SDSC) at UC San Diego, a TeraGrid partner, explains the dilemma thusly:

*"The use of GPUs speeds up a single node considerably, sometimes more than 30 fold. But if at the same time we don't develop a 30-fold higher bandwidth and 30- fold lower latency interconnect, scaling will always be limited across clusters of GPUs."*

When it comes to industry standards, CPUs have the advantage of being well-supported, while GPU systems must rely mostly on proprietary software systems. Varhol explains that it's difficult for software vendors to support multiple platforms, so until GPUs are more widely-accepted, there may be problems with porting software. There is some industry support, namely from the primary graphics chip vendors themselves, NVIDIA and AMD. Developers looking to program GPUs can choose between NVIDIA's proprietary CUDA parallel computing architecture and AMD's OpenCL (Open Computing Language) programming standard.

The general consensus seems to be that with the proper resources and training, GPUs are worth the trouble. "Essentially, if the effort has been made to port the code to GPUs then the performance improvement over CPU systems can be phenomenal," Walker explains.

One area where GPUs really shine is data analysis, where they may net speedups of 200x. This is significant since the amount of post-computational scientific data is growing quickly.

Zverina makes sure to point out that GPUs won't completely replace CPUs since "GPUs still require CPUs to access data from disk, or to exchange data between compute nodes in a multi-node cluster." A so-called GPU supercomputer is really a GPU-CPU hybrid system.

Overall, it seems the researchers at TeraGrid are cautiously optimistic about the potential of GPU computing, but, As Zverina writes, "some researchers say more needs to be done to attract, train, and support developers for good GPU code, especially as TeraGrid transitions to the eXtreme Digital (XD) program this year."