

Nvidia pushes ARM supercomputing

reader comments

48 with 35 posters participating

Share this story

Graphics chip maker Nvidia is best known for consumer computing, vying with AMD's Radeon line for framerates and eye candy. But the venerable giant hasn't ignored the rise of GPU-powered applications that have little or nothing to do with gaming. In the early 2000s, UNC researcher Mark Harris began work popularizing the term "GPGPU," referencing the use of Graphics Processing Units for non-graphics-related tasks. But most of us didn't really become aware of the non-graphics-related possibilities until GPU-powered bitcoin-mining code was released in 2010, and shortly thereafter, strange boxes packed nearly solid with high-end gaming cards started popping up everywhere.

From digital currencies to supercomputing

The Association for Computing Machinery grants one or more \$10,000 Gordon Bell Prize every year to a research team that has made a break-out achievement in performance, scale, or time-to-solution on challenging science and engineering problems. Five of the six entrants in 2018—including both winning teams, Oak Ridge National Laboratory and Lawrence Berkeley National Laboratory—used Nvidia GPUs in their

supercomputing arrays; the Lawrence Berkeley team included six people from Nvidia itself.

Enlarge / The impressive part about the segmentation masks overlaid on this map projection has nothing to do with antialiasing—it's the 300+ petaflops needed to analyze an entire planet's worth of atmospheric data in order to produce it.

In March of this year, Nvidia acquired Mellanox, makers of the high-performance network interconnect technology InfiniBand. (InfiniBand is frequently used as an alternative to Ethernet for massively high-speed connections between storage and compute stacks in enterprise, with real throughput up to 100Gbps.) This is the same technology the LBNL/Nvidia team used in 2018 to win a Gordon Bell Prize (with a project on deep learning for climate analytics).

The acquisition sent a clear signal (which Nvidia also spelled out clearly for anyone who wasn't paying attention) that the company was serious about the supercomputing space and not merely looking for optics to advance its place in the consumer market.

Moving toward a more-open future

This strong history of research and acquisition underscores the importance of the move Nvidia announced Monday morning at the International Supercomputing Conference in Frankfurt. The company is making its full stack of supercomputing hardware and software available for ARM-powered high-performance computers, and it expects to complete the project by the end of 2019. In a Reuters interview, Nvidia VP of accelerated computing Ian Buck described the move as a “heavy lift” technically, requested by HPC researchers in Europe and Japan.

Most people know ARM best for power-efficient, relatively low-performance (compared to traditional x86-64 builds by Intel and AMD) systems-on-chip used in smartphones, tablets, and novelty devices like the Raspberry Pi. At first blush, this makes ARM an odd choice for supercomputing. However, there’s much more to HPC than individually beefy CPUs. On the technical side of things, data-center-scale computing generally relies as much or more on massive parallelism as per-thread performance. The typical Arm SOC’s focus on power efficiency means that much less power draw and cooling is necessary, allowing more of them to be crammed into a data center. This means a potentially lower cost, lower footprint, and higher reliability for the same amount of computer.

But the licensing is potentially even more important—where Intel, IBM, and AMD architectures are closed and proprietary, ARM’s are wide open. Unlike the x86-64 CPU manufacturers, ARM doesn’t make chips itself—it simply licenses its technology out to a wide array of manufacturers who then build actual SOC’s with it.

Enlarge / Pinebook is moving past its original \$99 tinkerer's laptop to this upcoming \$199 daily driver, the magnesium-alloy-bodied Pinebook Pro. It would be impossible for such a small company to compete head-to-head on price with Chromebooks on x86 hardware.

This open-architecture hardware design appeals to a wide array of technologists, including developers wanting to accelerate design cycles, security wonks worried about the hardware equivalent of a Ken Thompson hack buried in a closed CPU design and manufacture process, and innovators trying to bring down the cost barrier of entry-level computing.

Hopefully, Nvidia's move to support ARM in HPC will trickle down to support for more prosaic devices as well, meaning cheaper, more powerful, and friendlier devices in the consumer space.

<https://arstechnica.com/?p=1523321>