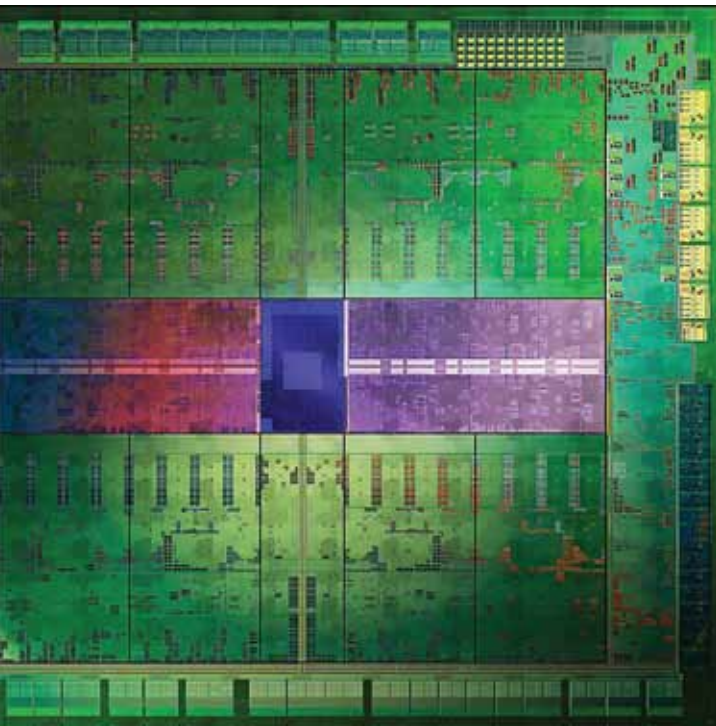


Software Accelerates Computing Time for Complex Math

NASA Technology

The 19th century mathematician Carl Friedrich Gauss, the astronomer par excellence of his day, called mathematics “the queen of the sciences.” If Gauss were alive now, he would likely marvel at the kinds of scientific breakthroughs and insights NASA scientists are now achieving through a number-crunching technology that was not possible during his time.

Our advantage in the modern age is the supercomputer: a souped-up, multiprocessor machine capable of churning out calculations at the rate of billions, trillions, or even quadrillions of operations per second. (The technical term used for these calculations is floating-point operations per second, or FLOPS.)



NVIDIA's GeForce GTX 680 consists of 3.54 billion transistors.

With nearly 100 active missions, NASA has no shortage of data, and the Agency is always looking for ways to better process and manage the results of its scientific endeavors. NASA has even built one of the world's most powerful supercomputers at Ames Research Center, Pleiades, which allows researchers to comb through vast swaths of data and model everything from the interaction of atoms to the formation of entire galaxies.

While supercomputers are the thoroughbreds of the processing world, not all scientists have access to such machines, which can cost hundreds of millions of dollars to build and millions more per year in maintenance and electric bills. For some, their only recourse is a desktop or laptop, whose FLOPS performance is often massively lower than its supercomputer counterpart. The same calculation that takes a supercomputer a day or two to solve could very well take a standard PC over a month to complete.

One company caught NASA's attention by finding a way to connect ordinary scientists and ordinary machines with extraordinary processing power. Doing so would give the Agency access to a new technology, allowing researchers to complete some projects locally on their PCs rather than calling on the Pleiades supercomputer to do the same job remotely.

Technology Transfer

If you've played a video game lately on a console or a computer, you've probably noticed how lifelike and smooth the graphics appear. The industry has come a long way since the days of Pong, and a crucial moment in its development was brought about by the invention in the late 1990s of what's called a graphics processing unit (GPU)—an electronic chip that can accelerate the processing of a massive number of computations at an astonishing speed.

GPU accelerators are critical to creating the realism of today's video games, as they not only perform the vector calculations needed to accelerate the rendering


of millions of triangles, which in combination comprise the images seen on a screen, but also do so at a rate of 60 times per second. The result is real-time action. But beyond ushering in a new era of electronic entertainment, the new technology also meant there was now a faster way of solving scientific problems that utilized parallel computing, a kind of super computing which, like video game graphics rendering, functions by solving a great many calculations simultaneously.

One of the first companies to recognize the potential applications for GPUs other than delivering video game graphics was Newark, Delaware-based EM Photonics Inc., a company that specializes in high-performance computing software. In the mid-2000s they began to develop code designed not for graphics rendering but for solving complex algorithms used in the modeling of antennas and optical devices. But programming GPUs at first, according to EM Photonics CEO Eric Kelmelis, was not easy.

“The hardware was initially designed for rendering graphics,” he says, “so you made it think it was rendering graphics. But in truth, it was doing a scientific computing operation for you—running an equation.”

That all changed in 2006, when NVIDIA, the company that invented GPUs, released the CUDA parallel computing platform and programming model to make developing software for the powerful processor chip more user friendly.

With the added ease of use provided by the CUDA platform, EM Photonics set its sights on completing an ambitious, first-of-its-kind project: programming a family of GPU-accelerated linear algebra libraries, including an implementation of the de facto industry standard LAPACK. These solvers often have to deal with an enormous amount of data, and traditional versions require supercomputers in order to run in a timely manner. Moving these tools to GPU accelerators was the kind of innovation that could benefit scientists who use laptops or desktops to run these solvers.



To accomplish its goal, the company applied for Small Business Innovation Research (SBIR) funding, which Ames granted to them. NASA researchers, like other scientists, are constantly running linear algebra equations to accomplish mission objectives. Says NASA computer scientist Creon Levitt, who sat on the evaluation committee, “There was an obvious utility in having this kind of software, and nobody else was doing it. EM Photonics had the appropriate background in related technologies, so it seemed quite likely that they could pull it off.”

It is clear by now that he was right. In 2007 the company’s programmers cracked their knuckles and got to work. In August 2009 the CULA Dense package was commercially released.

Benefits

Running CULA Dense on a regular computer can be compared to accessing higher gears on a car—gears that you never knew existed. That’s because, before CULA Dense arrived, a scientist’s computer would run LAPACK solvers on its central processing unit (CPU). While CPUs are more adept at solving sequential problems, or problems that each require step-by-step processes, they are not as fast and efficient as GPUs when programmed for parallel computing, especially when it involves using localized data.

EM Photonics, in creating CULA Dense, removed that formidable programming barrier, providing a simple and accessible tool for solving these types of problems that any scientist without computer expertise could use.

According to Henry Jin, a researcher in the supercomputing division at NASA Ames, the difference in aggregate calculating power between the two chips is staggering. “A modern CPU can give you about 20 gigaflops at its peak,” he says. “A single GPU accelerator can easily give you up to one teraflop, so that’s a thousand

gigaflops. So from the FLOPS point of view, there is a big advantage with GPUs.”

Put another way, it means that CULA Dense can solve parallel calculations, on average, 6 to 10 times faster than CPU-based LAPACK applications. In some cases, processing times are reduced by more than 100-fold. The same projects that used to take weeks to complete now take days; those that took days are now processed in hours. Whether it’s modeling the interactions between distant galaxies or simulating a model fighter jet landing on an aircraft carrier, performing complex algorithms on a personal computer has never been faster.

In the 3 years since the software has been on the market, CULA Dense has acquired more than 12,000 users working in government agencies, the private sector, and academic institutions all over the world. Even Titan, the fastest supercomputer in the world as of November 2012, which is housed at Oak Ridge National Laboratory in Tennessee, runs the application to increase its already astronomical computing speed.

Kelmelis notes that, as a result of the product’s success, both revenues and the number of company employees are

up by 10 percent. And through a separate SBIR contract with Ames, EM Photonics more recently developed and commercialized another library of linear algebra solvers called CULA Sparse, which provides scientists with a further assortment of mathematical tools that have access to GPU processing power.

According to NVIDIA’s general manager of GPU computing software, Ian Buck, “The success of solvers like CULA demonstrates the broad applicability of GPUs to address a range of scientific challenges. Today there are hundreds of CUDA-based applications in use around the world to enable new breakthroughs in everything from brain tumor and HIV/AIDS research, to the search for cleaner, renewable energy.”

Regarding the company’s collaboration with NASA, Kelmelis says, “It helped us launch into a whole new product area. In the past we were very special-purpose-application focused, but CULA has allowed us to reach a much broader audience and deliver on some very advanced technology.” ♦

A portrait of global aerosols at a 10-kilometer resolution, simulated by the Goddard Earth Observing System Model (GEOS-5) on NASA’s Discover supercomputer. The GEOS-5 is capable of simulating worldwide weather at resolutions of 10 to 3.5 kilometers.

