



NVIDIA Pioneers New Standard for High Performance Computing With Tesla GPUs Built on Kepler Architecture

New Architecture Triples Energy Efficiency, Makes GPU-Accelerated Computing Systems Easier to Program

SAN JOSE, CA -- **GPU Technology Conference** - NVIDIA today unveiled a new family of Tesla® GPUs based on the revolutionary [NVIDIA® Kepler™ GPU computing architecture](#), which makes GPU-accelerated computing easier and more accessible for a broader range of high performance computing (HPC) scientific and technical applications.

The new NVIDIA Tesla K10 and K20 GPUs are computing accelerators built to handle the most complex HPC problems in the world. Designed with an intense focus on high performance and extreme power efficiency, Kepler is three times as efficient as its predecessor, the NVIDIA Fermi™ architecture, which itself established a new standard for parallel computing when introduced two years ago.

"Fermi was a major step forward in computing," said Bill Dally, chief scientist and senior vice president of research at NVIDIA. "It established GPU-accelerated computing in the top tier of high performance computing and attracted hundreds of thousands of developers to the GPU computing platform. Kepler will be equally disruptive, establishing GPUs broadly into technical computing, due to their ease of use, broad applicability and efficiency."

The Tesla K10 and K20 GPUs were introduced at the GPU Technology Conference (GTC), as part of a series of announcements from NVIDIA, all of which can be accessed in the [GTC online press room](#).

NVIDIA developed a set of innovative architectural technologies that make the Kepler GPUs high performing and highly energy efficient, as well as more applicable to a wider set of developers and applications. Among the major innovations are:

- **SMX Streaming Multiprocessor** -- The basic building block of every GPU, the SMX streaming multiprocessor was redesigned from the ground up for high performance and energy efficiency. It delivers up to three times more performance per watt than the Fermi streaming multiprocessor, making it possible to build a supercomputer that delivers one petaflop of computing performance in just 10 server racks. SMX's energy efficiency was achieved by increasing its number of CUDA® architecture cores by four times, while reducing the clock speed of each core, power-gating parts of the GPU when idle and maximizing the GPU area devoted to parallel-processing cores instead of control logic.
- **Dynamic Parallelism** -- This capability enables GPU threads to dynamically spawn new threads, allowing the GPU to adapt dynamically to the data. It greatly simplifies parallel programming, enabling GPU acceleration of a broader set of popular algorithms, such as adaptive mesh refinement, fast multipole methods and multigrid methods.
- **Hyper-Q** -- This enables multiple CPU cores to simultaneously use the CUDA architecture cores on a single Kepler GPU. This dramatically increases GPU utilization, slashing CPU idle times and advancing programmability. Hyper-Q is ideal for cluster applications that use MPI.

"We designed Kepler with an eye towards three things: performance, efficiency and accessibility," said Jonah Alben, senior vice president of GPU Engineering and principal architect of Kepler at NVIDIA. "It represents an important milestone in GPU-accelerated computing and should foster the next wave of breakthroughs in computational research."

NVIDIA Tesla K10 and K20 GPUs

The NVIDIA Tesla K10 GPU delivers the world's highest throughput for signal, image and seismic processing applications. Optimized for customers in oil and gas exploration and the defense industry, a single Tesla K10 accelerator board features two GK104 Kepler GPUs that deliver an aggregate performance of 4.58 teraflops of peak single-precision floating point and 320 GB per second memory bandwidth.

The NVIDIA Tesla K20 GPU is the new flagship of the Tesla GPU product family, designed for the most computationally intensive HPC environments. Expected to be the world's highest-performance, most energy-efficient GPU, the Tesla K20 is planned to be available in the fourth quarter of 2012.

The Tesla K20 is based on the GK110 Kepler GPU. This GPU delivers three times more double precision compared to Fermi architecture-based Tesla products and it supports the Hyper-Q and dynamic parallelism capabilities. The GK110 GPU is expected to be incorporated into the new Titan supercomputer at the [Oak Ridge National Laboratory](#) in Tennessee and the Blue Waters system at the [National Center for Supercomputing Applications](#) at the [University of Illinois at Urbana-Champaign](#).

"In the two years since Fermi was launched, hybrid computing has become a widely adopted way to achieve higher performance for a number of critical HPC applications," said Earl C. Joseph, program vice president of High-Performance

Computing at IDC. "Over the next two years, we expect that GPUs will be increasingly used to provide higher performance on many applications."

Preview of CUDA 5 Parallel Programming Platform

In addition to the Kepler architecture, NVIDIA today released a preview of the CUDA 5 parallel programming platform. Available to more than 20,000 members of NVIDIA's GPU Computing Registered Developer program, the platform will enable developers to begin exploring ways to take advantage of the new Kepler GPUs, including dynamic parallelism.

The CUDA 5 parallel programming model is planned to be widely available in the third quarter of 2012. Developers can get access to the preview release by signing up for the GPU Computing Registered Developer program on the [CUDA website](#).

About NVIDIA Tesla GPUs

NVIDIA Tesla GPUs are massively parallel accelerators based on the NVIDIA CUDA parallel computing platform. Tesla GPUs are designed from the ground up for power-efficient, high performance computing, computational science and supercomputing, delivering dramatically higher application acceleration for a range of scientific and commercial applications than a CPU-only approach. Today, Tesla GPUs power three of the world's top five supercomputers.

More information about NVIDIA Tesla GPUs is available at the [Tesla website](#). Follow us on [Twitter](#) (@NVIDIATesla).

About GTC

The [GPU Technology Conference](#) advances global awareness of GPU computing and visualization, and their importance to the future of science and innovation. View the latest news from NVIDIA and its partners in the [GTC press room](#).

About NVIDIA

[NVIDIA](#) (NASDAQ: NVDA) awakened the world to computer graphics when it invented the [GPU](#) in 1999. Today, its [processors](#) power a broad range of products from [smartphones](#) to [supercomputers](#). NVIDIA's [mobile processors](#) are used in [cell phones](#), [tablets](#) and [auto infotainment systems](#). [PC gamers](#) rely on GPUs to enjoy spectacularly immersive worlds. Professionals use them to create [3D graphics](#) and visual effects in movies and to design everything from golf clubs to jumbo jets. And researchers utilize GPUs to advance the frontiers of science with [high performance computing](#). The company has more than 4,500 patents issued, allowed or filed, including ones covering ideas essential to modern computing. For more information, see [www.nvidia.com](#).

Certain statements in this press release including, but not limited to statements as to: the availability, impact and benefits of the NVIDIA Tesla K10 and K20 GPUs; the availability, impact and benefits of the CUDA 5 parallel programming platform; and the effects of the company's patents on modern computing are forward-looking statements that are subject to risks and uncertainties that could cause results to be materially different than expectations. Important factors that could cause actual results to differ materially include: global economic conditions; our reliance on third parties to manufacture, assemble, package and test our products; the impact of technological development and competition; development of new products and technologies or enhancements to our existing product and technologies; market acceptance of our products or our partners products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of our products or technologies when integrated into systems; as well as other factors detailed from time to time in the reports NVIDIA files with the Securities and Exchange Commission, or SEC, including its Form 10-K for the fiscal period ended January 29, 2012. Copies of reports filed with the SEC are posted on the company's website and are available from NVIDIA without charge. These forward-looking statements are not guarantees of future performance and speak only as of the date hereof, and, except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances.

(C) 2012 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, CUDA, Fermi, Kepler, and Tesla are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated. Features, pricing, availability, and specifications are subject to change without notice.

Ken Brown
Corporate Communications
+1-408-486-2626
kebrown@nvidia.com