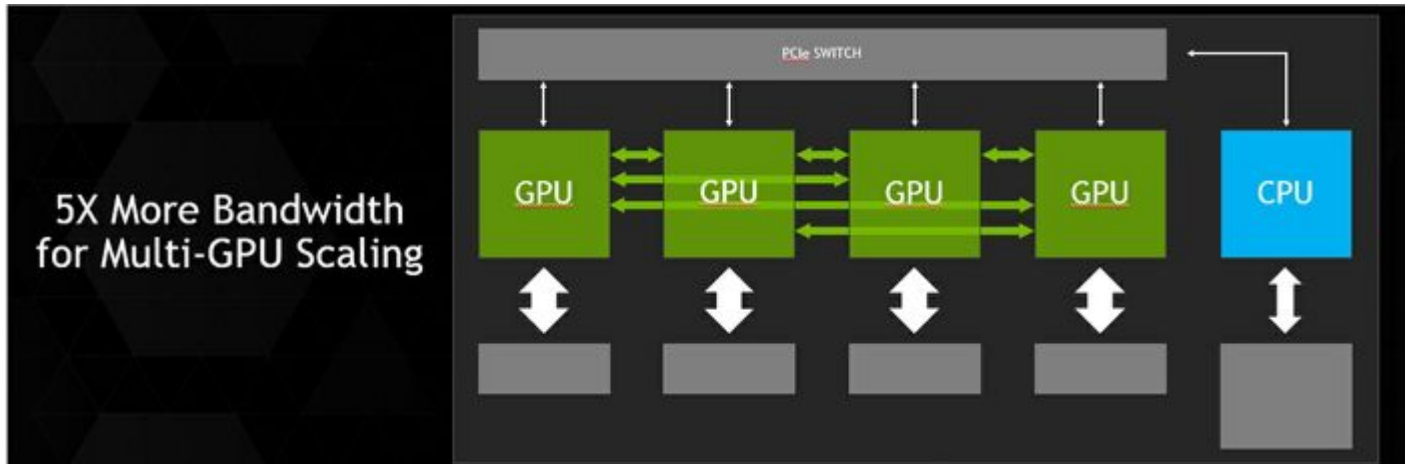


# NVIDIA Launches World's First High-Speed GPU Interconnect, Helping Pave the Way to Exascale Computing

## NVLink Enables Tightly Integrated CPUs and GPUs

SAN JOSE, CA -- GTC -- NVIDIA today announced that it plans to integrate a high-speed interconnect, called NVIDIA® NVLink™, into its future GPUs, enabling GPUs and CPUs to share data five to 12 times faster than they can today. This will eliminate a longstanding bottleneck and help pave the way for a new generation of exascale supercomputers that are 50-100 times faster than today's most powerful systems.

NVIDIA will add NVLink technology into its Pascal GPU architecture -- expected to be introduced in 2016 -- following this year's new [NVIDIA Maxwell compute architecture](#). The new interconnect was co-developed with IBM, which is incorporating it in future versions of its POWER CPUs.



"NVLink technology unlocks the GPU's full potential by dramatically improving data movement between the CPU and GPU, minimizing the time that the GPU has to wait for data to be processed," said Brian Kelleher, senior vice president of GPU Engineering at NVIDIA.

"NVLink enables fast data exchange between CPU and GPU, thereby improving data throughput through the computing system and overcoming a key bottleneck for accelerated computing today," said Bradley McCredie, vice president and IBM Fellow at IBM. "NVLink makes it easier for developers to modify high-performance and data analytics applications to take advantage of accelerated CPU-GPU systems. We think this technology represents another significant contribution to our OpenPOWER ecosystem."

With NVLink technology tightly coupling IBM POWER CPUs with [NVIDIA Tesla® GPUs](#), the POWER data center ecosystem will be able to fully leverage GPU acceleration for a diverse set of applications, such as high performance computing, data analytics and machine learning.

### Advantages Over PCI Express 3.0

Today's GPUs are connected to x86-based CPUs through the [PCI Express \(PCIe\)](#) interface, which limits the GPU's ability to access the CPU memory system and is four- to five-times slower than typical CPU memory systems. PCIe is an even greater bottleneck between the GPU and IBM POWER CPUs, which have more bandwidth than x86 CPUs. As the NVLink interface will match the bandwidth of typical CPU memory systems, it will enable GPUs to access CPU memory at its full bandwidth.

This high-bandwidth interconnect will dramatically improve accelerated software application performance. Because of memory system differences -- GPUs have fast but small memories, and CPUs have large but slow memories -- accelerated computing applications typically move data from the network or disk storage to CPU memory, and then copy the data to GPU memory before it can be crunched by the GPU. With NVLink, the data moves between the CPU memory and GPU memory at much faster speeds, making [GPU-accelerated applications](#) run much faster.

### Unified Memory Feature

Faster data movement, coupled with another feature known as [Unified Memory](#), will simplify GPU accelerator programming. Unified Memory allows the programmer to treat the CPU and GPU memories as one block of memory. The programmer can operate on the data without worrying about whether it resides in the CPU's or GPU's memory.

Although future NVIDIA GPUs will continue to support PCIe, NVLink technology will be used for connecting GPUs to NVLink-enabled CPUs as well as providing high-bandwidth connections directly between multiple GPUs. Also, despite its very high bandwidth, NVLink is substantially more energy efficient per bit transferred than PCIe.

NVIDIA has designed a module to house GPUs based on the Pascal architecture with NVLink. This new GPU module is one-third the size of the standard PCIe boards used for GPUs today. Connectors at the bottom of the Pascal module enable it to be plugged into the motherboard, improving system design and signal integrity.

NVLink high-speed interconnect will enable the tightly coupled systems that present a path to highly energy-efficient and scalable exascale supercomputers, running at 1,000 petaflops ( $1 \times 10^{18}$  floating point operations per second), or 50 to 100 times faster than today's fastest systems.

To Keep Current on NVIDIA:

- Like NVIDIA on [Facebook](#).

- Connect with NVIDIA on [LinkedIn](#).
- Use #GTC14 on [Twitter](#).
- View NVIDIA videos on [YouTube](#).
- Keep up with the [NVIDIA Blog](#).
- Use the Pulse news reader to subscribe to the NVIDIA Daily News feed.

## About NVIDIA

Since 1993, [NVIDIA](#) ( NASDAQ : NVDA ) has pioneered the art and science of [visual computing](#). The company's technologies are transforming a world of displays into a world of interactive discovery — for everyone from gamers to scientists, and consumers to enterprise customers. More information at <http://nvidianews.nvidia.com/> and <http://blogs.nvidia.com/>.

---

© 2014 NVIDIA Corporation. All rights reserved. NVIDIA and the NVIDIA logo 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.

## Media Contacts

George Millington  
+1 408 562 7226  
[gmillington@nvidia.com](mailto:gmillington@nvidia.com)