Nvidia Reveals Maxwell Successor With Tesla P100

Written by Marco Attard
08 April 2016

Nvidia announces Pascal-- the successor to Maxwell GPU architecture-- with the Tesla P100, a graphics card aimed at High-Performance Computing (HPC) and artificial intelligence applications based on the new GP100 GPU.

The Tesla P100 is the first full-size Nvidia GPU built using the TSMC 16nm FinFET process (previously the company was using a 28nm process). It is also the first to use 2nd generation High Bandwidth Memory (HBM2), allowing for what the Nvidia promises is a big performance boost.

How big? Nvidia claims P100 performance clocks at 21.2 teraflops of half-precision (FP16) floating point performance, 10.6 teraflops of single precision (FP32), and 5.3 teraflops (1/2 rate) of double precision. In comparison, the Titan X and Tesla M40 offer 7 teraflops of single precision floating point performance. As for memory, the P100 features 720GB/s bandwidth (via 4096-bit memory bus) and 16GB capacity.

One has to point out the Tesla P100 does not carry a fully enabled version of the Pascal architecture. As mentioned earlier, while based on the GP100 GPU it has 56 of the 60 streaming multiprocessors (SM) enabled.

Nvidia Reveals Maxwell Successor With Tesla P100

Written by Marco Attard
08 April 2016

Also launched is a product powered by the Tesla P100-- the DGX-1, a supercomputer Nvidia says can "meet the unlimited computing demands of artificial intelligence." Built as a turnkey system, it carries x8 Tesla GP100 cards to provide 170 teraflops of half-precision performance from 28672 CUDA cores, together with x2 16-core Intel Xeon E5-2698 v3 2.3GHz CPUs, 512GB DDR4 RAM, x4 1.92TB SSD RAID, dual 10GbE LAN, and x4 InfiniBand EDR network ports.

The result, according to the company, delivers as much throughput as 250 x86 servers.

"Artificial intelligence is the most far-reaching technological advancement in our lifetime," Nvidia adds. "It changes every industry, every company, everything. It will open up markets to benefit everyone. Data scientists and AI researchers today spend far too much time on home-brewed high performance computing solutions. The DGX-1 is easy to deploy and was created for one purpose-- to unlock the powers of superhuman capabilities and apply them to problems that were once unsolvable."

The DG-X should hit general availability come June 2016, while the Tesla GP100 will be available to server makers by early 2017.

Go [Introducing the New Nvidia Pascal Architecture](#)