# NVIDIA

TESLA™ M2050 / M2070
GPU COMPUTING MODULE
## SUPERCOMPUTING AT 1/10TH THE COST

Based on the next-generation CUDA™ architecture codenamed "Fermi", the Tesla™ M2050 and M2070 Computing Modules enable seamless integration of GPU computing with host systems for high-performance computing and large data center, scale-out deployments.

The 20-series Tesla GPUs are the first to deliver greater than 10X the double-precision horsepower of a quad-core x86 CPU and the first to deliver ECC memory. The Tesla M2050 and M2070 modules deliver all of the standard benefits of GPU computing while enabling maximum reliability and tight integration with system monitoring and management tools. This gives data center IT staff much greater choice in how they deploy GPUs, with a wide variety of rack-mount and blade systems and with remote monitoring and remote management capabilities they need.

Compared to CPU-only systems, servers with Tesla 20-series GPU Computing Modules deliver supercomputing power at 1/10th the cost and 1/20th the power consumption while providing the highest compute density.
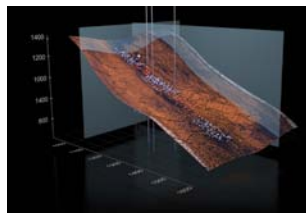
OIL & GAS
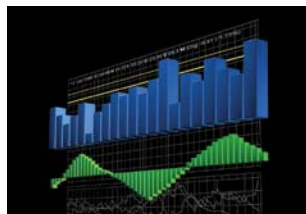
FINANCE

SCIENCE

TECHNICAL SPECIFICATIONS

FORM FACTOR
> 9.75" PCIe x16 form factor

 # OF TESLA GPUS
> 1

DOUBLE PRECISION FLOATING POINT PERFORMANCE (PEAK)
> 515 Gflops

SINGLE PRECISION FLOATING POINT PERFORMANCE (PEAK)
> 1.03 Tflops

TOTAL DEDICATED MEMORY*
> Tesla M2050: 3GB GDDR5
> Tesla M2070: 6GB GDDR5

MEMORY SPEED
> 1.55 GHz

MEMORY INTERFACE
> 384-bit

MEMORY BANDWIDTH
> 148 GB/sec

POWER CONSUMPTION
> 225W TDP

SYSTEM INTERFACE
> PCIe x16 Gen2

THERMAL SOLUTION
> Passive heatsink cooled by host system airflow

SOFTWARE DEVELOPMENT TOOLS
> CUDA C/C++/Fortran, OpenCL, DirectCompute Toolkits. NVIDIA Parallel Nsight™ for Visual Studio

*Note: With ECC on, a portion of the dedicated memory is used for ECC bits, so the available user memory is reduced by 12.5%. (e.g. 3 GB total memory yields 2.625 GB of user available memory.)

TESLA M2050 / M2070 GPU COMPUTING MODULE

Features	Benefits
448 CUDA CORES	Delivers up to 515 Gigaflops of double-precision peak performance in each GPU, enabling servers from leading OEMs to deliver a Teraflop or more of double-precision performance per 1 RU of space. Single precision peak performance is over one Teraflop per GPU.
ECC MEMORY	Meets a critical requirement for computing accuracy and reliability in datacenters and supercomputing centers. Offers protection of data in memory to enhance data integrity and reliability for applications. Register files, L1/L2 caches, shared memory, and DRAM all are ECC protected.
UP TO 6GB OF GDDR5 MEMORY PER GPU	Maximizes performance and reduces data transfers by keeping larger data sets in local memory that is attached directly to the GPU.
SYSTEM MONITORING FEATURES	Integrates the GPU subsystem with the host system's monitoring and management capabilities. This means IT staff can manage all of the critical components of the computing system through a common management interface such as IPMI or OEM-proprietary tools.
DESIGNED FOR MAXIMUM RELIABILITY	Passive heatsink design eliminates moving parts and cables.
NVIDIA PARALLEL DATACACHE™	Accelerates algorithms such as physics solvers, ray-tracing, and sparse matrix multiplication where data addresses are not known beforehand. This includes a configurable L1 cache per Streaming Multiprocessor block and a unified L2 cache for all of the processor cores.
NVIDIA GIGATHREAD™ ENGINE	Maximizes the throughput by faster context switching that is 10X faster than previous architecture, concurrent kernel execution, and improved thread block scheduling.
ASYNCHRONOUS TRANSFER	Turbocharges system performance by transferring data over the PCIe bus while the computing cores are crunching other data. Even applications with heavy data-transfer requirements, such as seismic processing, can maximize the computing efficiency by transferring data to local memory before it is needed.
CUDA PROGRAMMING ENVIRONMENT WITH BROAD SUPPORT OF PROGRAMMING LANGUAGES AND APIs	Choose C, C++, OpenCL, DirectCompute, or Fortran to express application parallelism and take advantage of the innovative "Fermi" architecture.
HIGH SPEED, PCIE GEN 2.0 DATA TRANSFER	Maximizes bandwidth between the host system and the Tesla processors. Enables Tesla systems to work with virtually any PCIe-compliant host system with an open PCIe slot (x8 or x16).

DRIVERS AND DOWNLOADS

> Tesla M2050/M2070 is supported under
  - Linux 32-bit and 64-bit
  - Windows Server 2003 and 2008
> Vertical Solutions and Software page:
  www.nvidia.com/object/vertical_solutions.html
> Software
  - Drivers — NVIDIA recommends that users get drivers for M-series products from their System OEM to ensure that driver is qualified by the OEM on their system.
  - Tools — Software development tools are available at
    www.nvidia.com/object/ tesla_software.html

SUPPORT

> HARDWARE SUPPORT
  For RMA requests, replacements and warranty issues regarding your NVIDIA based product, please contact the OEM that you purchased it from.
> KNOWLEDGEBASE
  Our knowledgebase is available online 24x7x365 and contains answers to the most common questions and issues: **www.nvidia.custhelp.com/cgi-bin/nvidia.cfg/php/enduser/std_alp.php**
> USER FORUMS
  Discuss Tesla products, talk about CUDA development, and share interesting issues, tips and solutions with your fellow NVIDIA Tesla users on the CUDA discussion forums:
  www.forums.nvidia.com

To learn more about NVIDIA Tesla, go to **www.nvidia.com/tesla**

NVIDIA.