TESLA S1070 DATASHEET

NVIDIA TESLA 1U COMPUTING SYSTEM

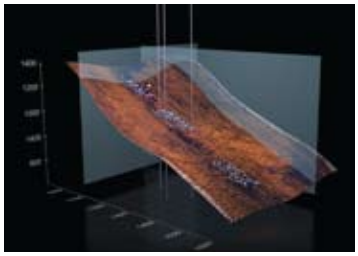
SOLVE TOMORROW'S PROBLEMS TODAY

HIGH PERFORMANCE PETASCALE COMPUTING SYSTEMS

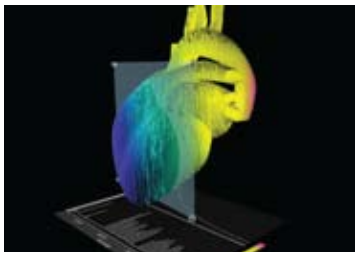
NVIDIA® Tesla™ S1070 computing system delivers the world's first teraflop processor, combining breakthrough performance with energy efficiency. A cluster with as few as 250 Tesla systems would have a peak theoretical performance of a petaflop. This system raises the bar for many-core computing in a heterogeneous environment that mixes multi-core CPUs and many-core GPUs for optimized performance. The combination of performance and energy efficiency enables scientists, engineers, and business users to tackle larger problems with the most advanced algorithms. Tesla S1070 delivers incredible performance per watt and can upgrade the performance of your data centers without requiring infrastructure changes for power or cooling and without massive increases in your energy bills.

FEEDING THE RELENTLESS DEMAND FOR HPC PERFORMANCE

The NVIDIA Tesla S1070 computing system, with almost 1,000 cores in a 1U chassis, delivers four teraflops of peak performance—double the performance of the previous Tesla generation. This new Tesla system includes 16 GB of ultra-fast memory for maximum performance with larger data sets. With Tesla's massively-parallel, many-core architecture, users can tackle next-generation computing problems today. Greater computing power fuels the pace of scientific advances and Tesla creates a discontinuity in accessibility of floating point horsepower.



SEISMIC EXPLORATION



LIFE SCIENCES

MOLECULAR DYNAMICS

MANY-CORE ARCHITECTURE DELIVERS OPTIMUM SCALING ACROSS HPC APPLICATIONS

Demand for computing performance in science and industry has far outpaced the ability of traditional CPU to keep up, even with the recent shift to multi-core CPUs. Many-core computing is the architectural answer to this problem, delivering hundreds of cores in a single processor compared to multi-core designs with only four, six, or eight. The availability of processors with hundreds of cores creates a discontinuity in computing because 1U systems with nearly one thousand cores are practical for the first time. This is practical now because the computing cores in a GPU were designed to be part of a massively-parallel system rather than being designed like a traditional CPU core. Dedicated ultra-fast memory for each Tesla processor also improves scalability as total memory bandwidth expands linearly when more GPUs are added to the solution.

HIGH-EFFICIENCY COMPUTING PLATFORM FOR ENERGY-CONSCIOUS ORGANIZATIONS

Unlike any other solution available in the HPC space, the Tesla S1070 delivers four teraflops in a 1U chassis with a typical energy footprint of only 700 watts. This "high density computing" allows data center managers to deliver more performance for their users without new demands on the electrical and thermal capabilities of their existing data centers. Tesla S1070 creates the foundation for new green computing initiatives, saving the world while saving you money.

CUDA™ TECHNOLOGY UNLOCKS THE POWER OF TESLA MANY-CORE PROCESSORS

The CUDA C compiler simplifies many-core programming by enabling code development in a high-level language and optimizing code to run on systems without knowledge of how many cores are in the hardware. CUDA applications automatically take advantage of more cores or fewer cores in a system, so they can scale from entry-level notebook GPUs to high-end GPUs in technical workstations and further into racks of GPUs in data centers. This allows developers to "code once" and deploy on a range of systems, as well as scale forward in time as future GPUs deliver more performance per watt and more cores per processor. The benefit for software users is the opportunity to boost computing performance simply by adding GPUs or using their existing GPUs in new ways.

FEATURES AND BENEFITS

TERAFLOP PROCESSORS	Delivers up to four teraflops of performance in a 1U rack-mount system giving Tesla products unmatched performance per unit of volume.
MASSIVELY-PARALLEL, MANY-CORE ARCHITECTURE	240 processor cores per processor with the ability to execute thousands of concurrent threads.
IEEE 754 FLOATING POINT PRECISION	Ensures your results meet industry standard precision.
DOUBLE-PRECISION MATH	Meets the precision requirements of your most demanding applications.
ASYNCHRONOUS TRANSFER	Turbocharges system performance by executing data transfers, even when the computing cores are busy.
SYSTEM MONITORING FEATURES	Simple management and monitoring post-installation helps your IT staff manage systems with ease. Remote capabilities as well as status lights on the front and rear of the unit ensure your staff can see the status whether they are on the other side of the rack or the other side of the world.
DUAL GEN2 PCIe CABLE CONNECTIONS	Maximizes bandwidth between the host system and the Tesla processors, with up to 12.8 GB/s peak transfer rates.
GEN2 PCIe CABLE WITH SMALL-FORM-FACTOR (SFF) HOST ADAPTER CARD	Enables Tesla systems to work with virtually any PCIe-compliant host system with an open PCIe slot (x8 or x16).

TECHNICAL SPECIFICATIONS

HOST REQUIREMENTS FOR TESLA 1U SYSTEMS

- Host system with PCI Express x8 or PCI Express x16
- PCIe Gen2 for best results
- Linux (64-bit and 32-bit)
  - Red Hat Enterprise Linux 4 and 5
  - SUSE 10.3
- NVIDIA recommends using systems tested for compatibility with Tesla 1U systems. For a complete list, visit http://www.nvidia.com/object/tesla_compatible_platforms.html.

TESLA 1U ARCHITECTURE

- Four teraflops of computing performance in a 1U configuration
- Massively-parallel, many-core architecture
- 960 scalar processor cores (240 per GPU)
- Ultra-fast memory access with 408 GB/sec total bandwidth (102 GB/sec peak bandwidth per GPU)
- 4x 512-bit GDDR3 memory interface (512-bit interface per GPU)
- Integer, IEEE 754 single-precision, and IEEE 754 double-precision floating point operations
- Hardware Thread Execution Manager enables thousands of concurrent threads per GPU
- Parallel shared memory enables processor cores to share data in local caches

SCALABLE SOLUTIONS

- Scalable from one to thousands of GPUs
- Available with one or two PCIe connections per system

SOFTWARE DEVELOPMENT TOOLS

- C language compiler, debugger, profiler, and emulation mode for debugging
- Standard numerical libraries for FFT (Fast Fourier Transform), BLAS (Basic Linear Algebra Subroutines), and CuDPP (CUDA Data Parallel Primitives)

PHYSICAL SPECIFICATIONS

- Chassis dimensions: 1.73" Height x 17.5" Width x 28.5" Deep
- Typical Power: 700 Watts

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

NVIDIA