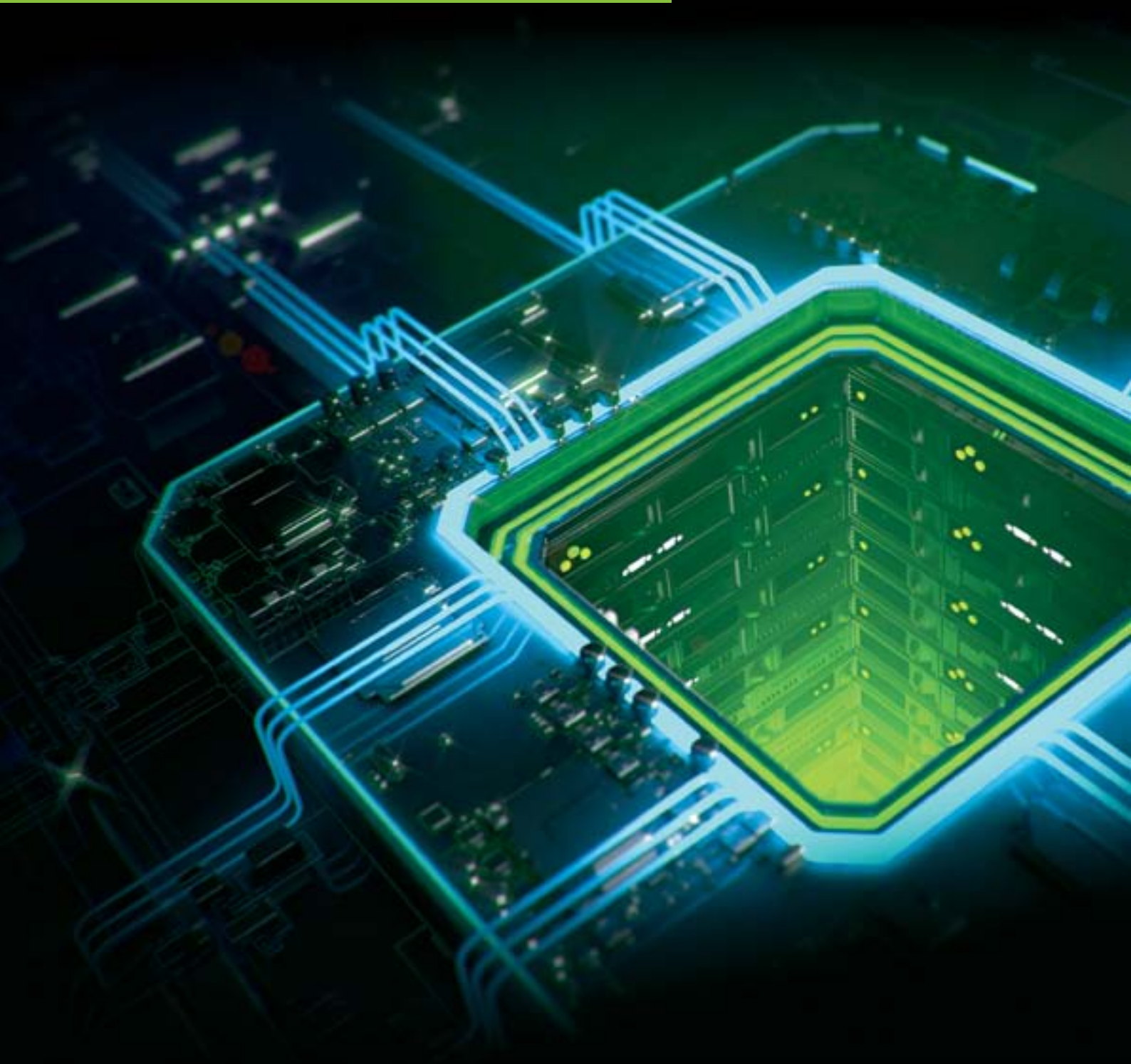
NVIDIA

REVOLUTIONIZING HIGH PERFORMANCE COMPUTING

NVIDIA® TESLA™



REVOLUTIONIZING HIGH PERFORMANCE COMPUTING NVIDIA® TESLA™

The need for computation in the high performance computing (HPC) Industry is increasing, as large and complex computational problems become commonplace across many industry segments. Traditional CPU technology, however, is no longer capable of scaling in performance sufficiently to address this demand.

The parallel processing capability of the GPU allows it to divide complex computing tasks into thousands of smaller tasks that can be run concurrently. This ability is enabling computational scientists and researchers to address some of the world's most challenging computational problems up to several orders of magnitude faster.

TABLE OF CONTENTS

NVIDIA® TESLA™ GPU COMPUTING

REVOLUTIONIZING HIGH PERFORMANCE COMPUTING

"We not only created the world's fastest computer, but also implemented a heterogeneous computing architecture incorporating CPU and GPU, this is a new innovation."

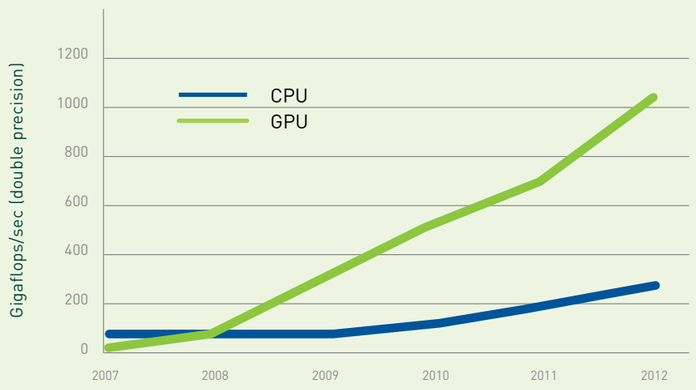
Premier Wen Jiabao
People's Republic of China

The high performance computing (HPC) industry's need for computation is increasing, as large and complex computational problems become commonplace across many industry segments. Traditional CPU technology, however, is no longer capable of scaling in performance sufficiently to address this demand.

The parallel processing capability of the GPU allows it to divide complex computing tasks into thousands of smaller tasks that can be run concurrently. This ability is enabling computational scientists and researchers to address some of the world's most challenging computational problems up to several orders of magnitude faster.

GPUs DELIVER SUPERIOR PERFORMANCE IN PARALLEL COMPUTATION.

GPUs deliver superior performance in parallel computation.



"The rise of GPU supercomputers on the Green500 signifies that heterogeneous systems, built with both GPUs and CPUs, deliver the highest performance and unprecedented energy efficiency,"

said Wu-chun Feng,
founder of the Green500 and
associate professor of
Computer Science at Virginia Tech.

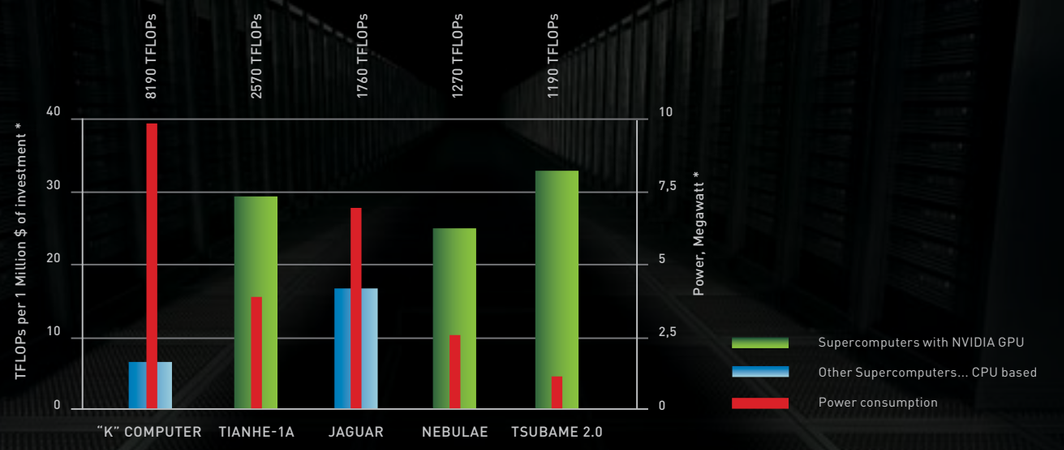
This advancement represents a dramatic shift in HPC. In addition to dramatic improvements in speed, GPUs also consume less power than conventional CPU-only clusters. GPUs deliver performance increases of 10x to 100x to solve problems in minutes instead of hours–while outpacing the performance of traditional computing with x86-based CPUs alone.

From climate modeling to advances in medical tomography, NVIDIA® Tesla™ GPUs are enabling a wide variety of segments in science and industry to progress in ways that were previously impractical, or even impossible, due to technological limitations.

Figure 1: Co-processing refers to the use of an accelerator, such as a GPU, to offload the CPU to increase computational efficiency.

THE WORLD'S TOP 5 SUPERCOMPUTERS

COST EFFECTIVENESS CHART: TFLOPS PER 1 MILLION $ OF INVESTMENT



* All data is taken from Top500 list and Media announcements open to public.

GPUs ARE REVOLUTIONIZING COMPUTING

WHY GPU COMPUTING?

With the ever-increasing demand for more computing performance, systems based on CPUs alone can no longer keep up. The CPU-only systems can only get faster by adding thousands of individual computers – this method consumes too much power and makes supercomputers very expensive. A different strategy is parallel computing, and the HPC industry is moving toward a hybrid computing model, where GPUs and CPUs work together to perform general purpose computing tasks.

As parallel processors, GPUs excel at tackling large amounts of similar data because the problem can be split into hundreds or thousands of pieces and calculated simultaneously.

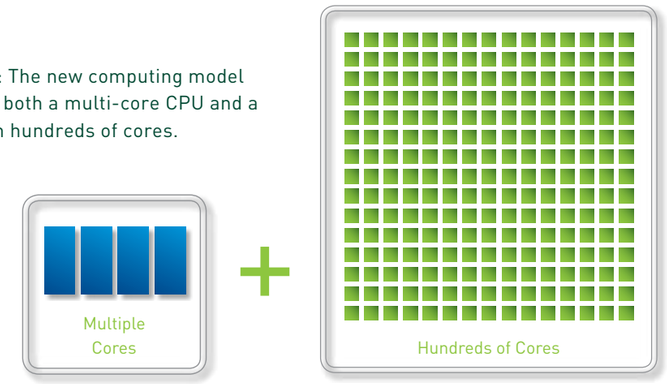
GPU SUPERCOMPUTING - GREEN HPC

GPUs significantly increase overall system efficiency as measured by performance per watt. "Top500" supercomputers based on heterogeneous architectures are, on average, almost three times more power-efficient than non-heterogeneous systems. This is also reflected on Green500 list – the icon of Eco-friendly supercomputing.

As sequential processors, CPUs are not designed for this type of computation, but they are adept at more serial based tasks such as running operating systems and organizing data. NVIDIA believes in applying the most relevant processor to the specific task in hand.

CORE COMPARISON BETWEEN A CPU AND A GPU

Figure 2: The new computing model includes both a multi-core CPU and a GPU with hundreds of cores.

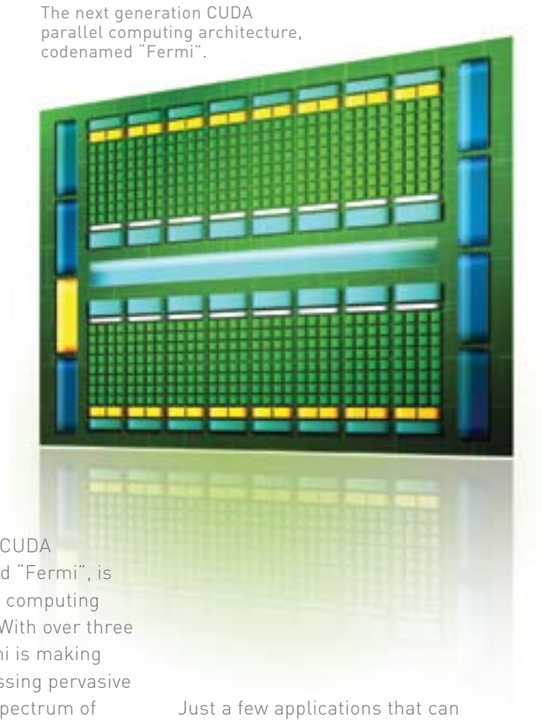


"I believe history will record Fermi as a significant milestone."

Dave Patterson, Director, Parallel Computing Research Laboratory, U.C. Berkeley Co-author of Computer Architecture: A Quantitative Approach

CUDA PARALLEL COMPUTING ARCHITECTURE

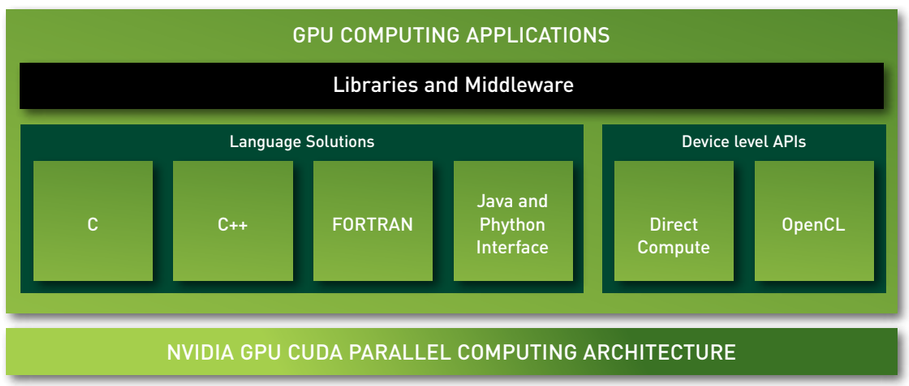
CUDA® is NVIDIA's parallel computing architecture and enables applications to run large, parallel workloads on NVIDIA GPUs. Applications that leverage the CUDA architecture can be developed in a variety of languages and APIs, including C, C++, Fortran, OpenCL, and DirectCompute. The CUDA architecture contains hundreds of cores capable of running many thousands of parallel threads, while the CUDA programming model lets programmers focus on parallelizing their algorithms and not the mechanics of the language.



The next generation CUDA parallel computing architecture, codenamed "Fermi".

The current generation CUDA architecture, codenamed "Fermi", is the most advanced GPU computing architecture ever built. With over three billion transistors, Fermi is making GPU and CPU co-processing pervasive by addressing the full-spectrum of computing applications. With support for C++, GPUs based on the Fermi architecture make parallel processing easier and accelerate performance on a wider array of applications than ever before.

Just a few applications that can experience significant performance benefits include ray tracing, finite element analysis, high-precision scientific computing, sparse linear algebra, sorting, and search algorithms.

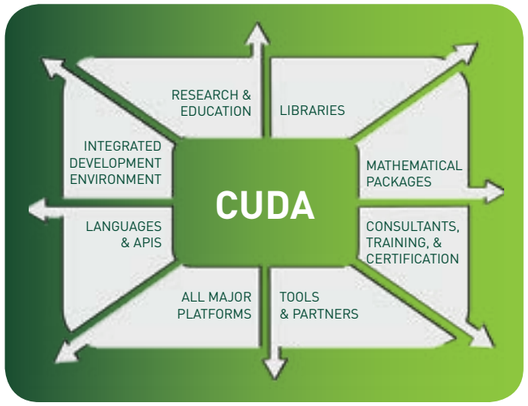


The CUDA parallel computing architecture, with a combination of hardware and software.

DEVELOPER ECOSYSTEM

In just a few years, an entire software ecosystem has developed around the CUDA architecture – from more than 400 universities worldwide teaching the CUDA programming model, to a wide range of libraries, compilers, and middleware that help users optimize applications for GPUs. This rich ecosystem has led to faster discovery and simulation in a wide range of fields including mathematics, life sciences, and manufacturing.

Locate a CUDA teaching center near you: research.nvidia.com/content/cuda-courses-map



Request a CUDA education course at: www.parallel-compute.com

DEVELOPERS TOOLS

Microsoft Visual Studio

NVIDIA PARALLEL NSIGHT DEVELOPMENT ENVIRONMENT FOR VISUAL STUDIO

NVIDIA Parallel Nsight software is the industry's first development environment for massively parallel computing integrated into Microsoft Visual Studio, the world's most popular development environment for Windows-based applications and services. It integrates CPU and GPU development, allowing developers to create optimal GPU-accelerated applications.

Parallel Nsight supports Microsoft Visual Studio 2010, advanced debugging and analysis capabilities, as well as Tesla 20-series GPUs.

For more information, visit: developer.nvidia.com/object/nsight.html.

NVIDIA Parallel Nsight software integrates CPU and GPU development, allowing developers to create optimal GPU-accelerated applications.

Request hands-on Parallel Nsight training: www.parallel-compute.com

GPU-ACCELERATED LIBRARIES

Take advantage of the massively parallel computing power of the GPU by using the GPU-accelerated versions of your existing libraries. Some examples include:

NVIDIA Math Libraries
A collection of GPU-accelerated libraries—including FFT, BLAS, sparse matrix operations, RNG, performance primitives for image/signal processing, and the Thrust C++ library of high-performance templated algorithms—that all deliver significant speedups when compared to CPU-only libraries. These highly optimized libraries are free of charge in the NVIDIA® CUDA® Toolkit available at www.nvidia.com/getcuda

EM Photonics CULA
A GPU-accelerated linear algebra (LA) library that dramatically improves the performance of sophisticated mathematics.

MAGMA
A collection of open-source, next-generation linear algebra libraries for heterogeneous GPU-based architectures supporting interfaces to current LA packages and standards (e.g. LAPACK and BLAS).

RogueWave IMSL
A comprehensive set of mathematical and statistical functions for Fortran applications to take advantage of GPU acceleration.

ACCELERATE YOUR CODE EASILY WITH OPENACC DIRECTIVES

GET 2X SPEED-UP IN 4 WEEKS OR LESS

"I have written micromagnetic codes (written in Fortran 90) to study the properties of two and three dimensional magnetic systems. The directives approach enabled me to port my existing code with ease to perform my computations on the GPU which resulted in a significant speedup (more than 20 times) of the computation."

Professor M. Amin Kayali,
University of Houston

Accelerate your code with directives and tap into the hundreds of computing cores in GPUs. With directives, you simply insert compiler hints into your code and the compiler will automatically map compute-intensive portions of your code to the GPU.

By starting with a free, 30-day trial of PGI directives today, you are working on the technology that is the foundation of the OpenACC directives standard.

OpenACC is:

- Easy: simply insert hints in your codebase
- Open: run the single codebase on either the CPU or GPU
- Powerful: tap into the power of GPUs within hours

Coming in Q1 2012, OpenACC is supported by industry parallel computing tool leaders: Cray, CAPS, and The Portland Group (PGI).

FOR A FREE 30 DAY TRIAL OF THE DIRECTIVES APPROACH REGISTER NOW!
www.nvidia.eu/openacc

DIRECTIVE-BASED SOLUTIONS

Directives allow you to quickly add GPU acceleration to the most performance critical sections of your application while maintaining portability. Directivebased solutions for C and Fortran available today include:

PGI ACCELERATOR™ COMPILERS

Similar to OpenMP, PGI Accelerator directives can be added to existing C99 or Fortran applications to achieve GPU acceleration on Linux, Mac OS X, and Windows.

CAPS HMPP

Directive-based compiler for hybrid computing

Based on C and FORTRAN directives, HMPP offers a high level abstraction for hybrid programming that fully leverages the computing power of stream processors without the complexity associated with GPU programming. HMPP compiler integrates powerful data-parallel back ends for NVIDIA CUDA and OpenCL that drastically reduce development time. The HMPP runtime ensures application deployment on multi-GPU systems.
www.caps-entreprise.com

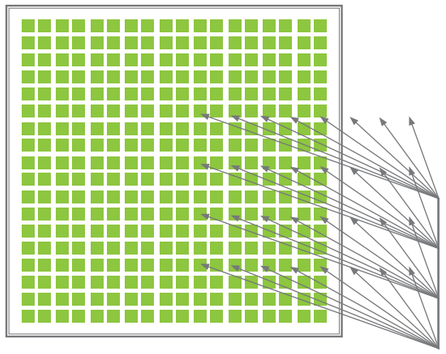
"The PGI compiler is now showing us just how powerful it is. On the software we are writing, it's at least 60 times faster on the NVIDIA card. We are very pleased and excited about the future uses. It's like owning a personal supercomputer."

Dr. Kerry Black,
University of Melbourne

OpenACC DIRECTIVES

The OpenACC Application Program Interface describes a collection of compiler directives to specify loops and regions of code in standard C, C++ and Fortran to be offloaded from a host CPU to an attached accelerator, providing portability across operating systems, host CPUs and accelerators.

The directives and programming model defined in this document allow programmers to create high-level host+accelerator programs without the need to explicitly initialize the accelerator, manage data or program transfers between the host and accelerator, or initiate accelerator startup and shutdown.



HINT

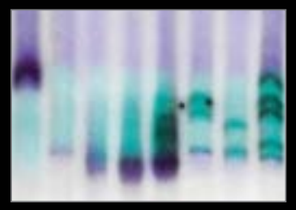
Add a simple compiler directive to provide a hint for the compiler to automatically parallelize your **for** loop and execute it on the GPU.

```
main() {
   double pi = 0.0f; long i;
   #pragma acc region
   for (i=0; i<N: i++)
   {
      double t= (double)((i+0.5)/N);
      pi +=4.0/(1.0+t*t);
   }
   printf("pi=%f\n",pi/N;
}
```

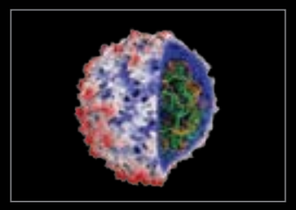
WATCH VIDEO*:
PGI ACCELERATOR, TECHNICAL PRESENTATION AT SC11

* USE YOUR PHONE, SMARTPHONE OR TABLET PC WITH QR READER SOFTWARE TO READ THE QR CODE.

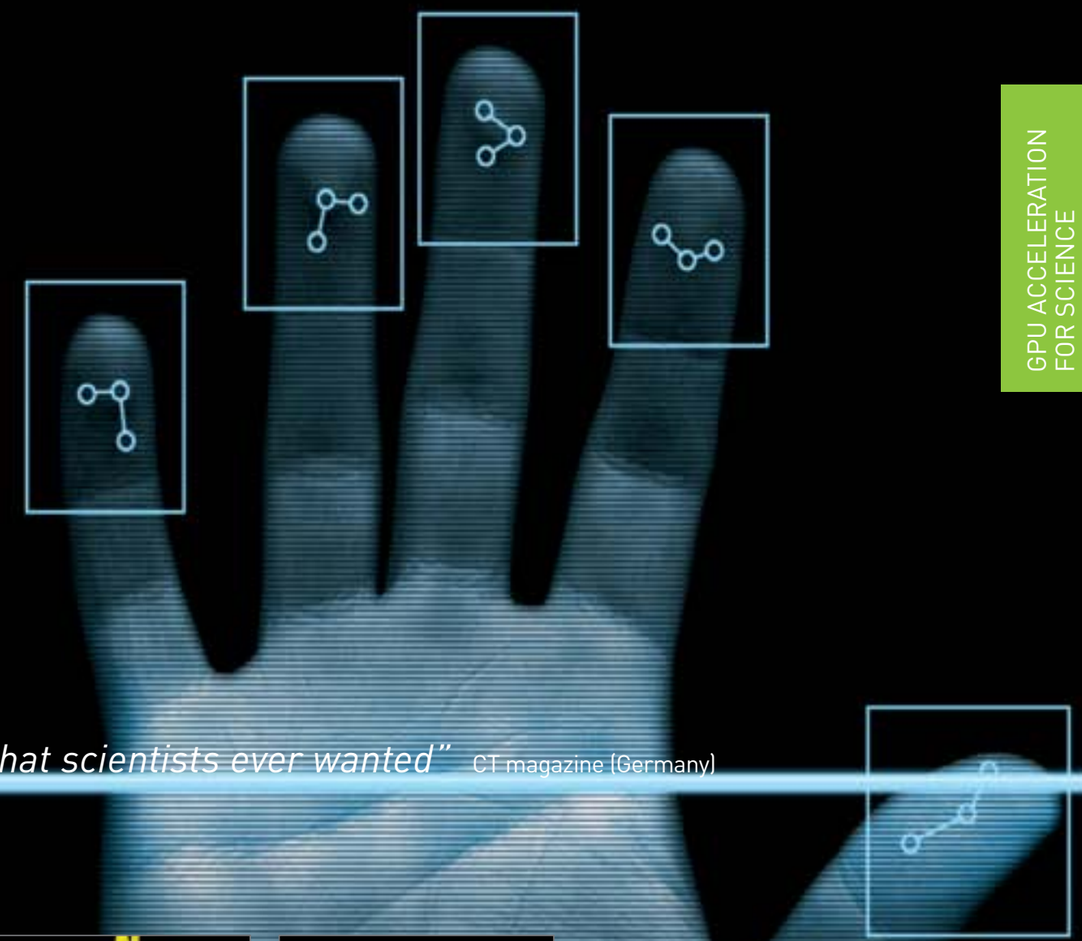
100X
Astrophysics
RIKEN



30X
Gene Sequencing
University of Maryland



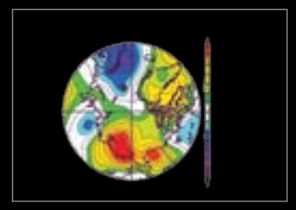
36x
Molecular Dynamics
University of Illinois,
Urbana-Champaign



GPU ACCELERATION
FOR SCIENCE
AND RESEARCH

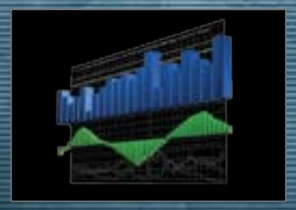
"NVIDIA's Tesla Fermi card includes almost everything what scientists ever wanted" CT magazine (Germany)

50x
MATLAB Computing,
AccelerEyes



80x
Weather Modeling
Tokyo Institute of
Technology

146x
Medical Imaging
U of Utah



149x
Financial Simulation
Oxford University

MATLAB ACCELERATIONS ON TESLA GPUs

MATLAB PERFORMANCE WITH TESLA

MathWorks®

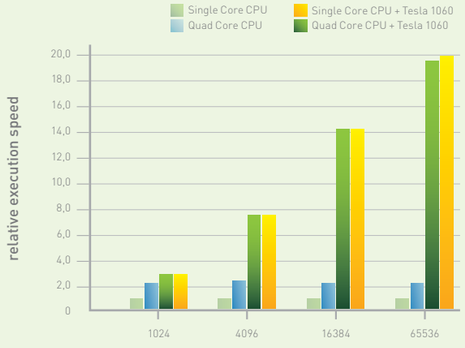
NVIDIA and MathWorks have collaborated to deliver the power of GPU computing for MATLAB users. Available through the latest release of MATLAB 2010b, NVIDIA GPU acceleration enables faster results for users of the Parallel Computing Toolbox and MATLAB Distributed Computing Server. MATLAB supports NVIDIA® CUDA™-enabled GPUs with compute capability version 1.3 or higher, such as Tesla™ 10-series and 20-series GPUs. MATLAB CUDA support provides the base for GPU-accelerated MATLAB operations and lets you integrate your existing CUDA kernels into MATLAB applications.

The latest release of Parallel Computing Toolbox and MATLAB Distributed Computing Server takes advantage of the CUDA parallel computing architecture to provide users the ability to

- Manipulate data on NVIDIA GPUs
- Perform GPU accelerated MATLAB operations
- Integrate users own CUDA kernels into MATLAB applications
- Compute across multiple NVIDIA GPUs by running multiple MATLAB workers with Parallel Computing Toolbox on the desktop and MATLAB Distributed Computing Server on a compute cluster

RELATIVE PERFORMANCE, POINT-IN-POLYGON DEMO

Compared to Single Core CPU Baseline



Core 2 Quad Q6600 2.4 GHz, 6 GB RAM, Windows 7, 64-bit, Tesla C1060, single precision operations.

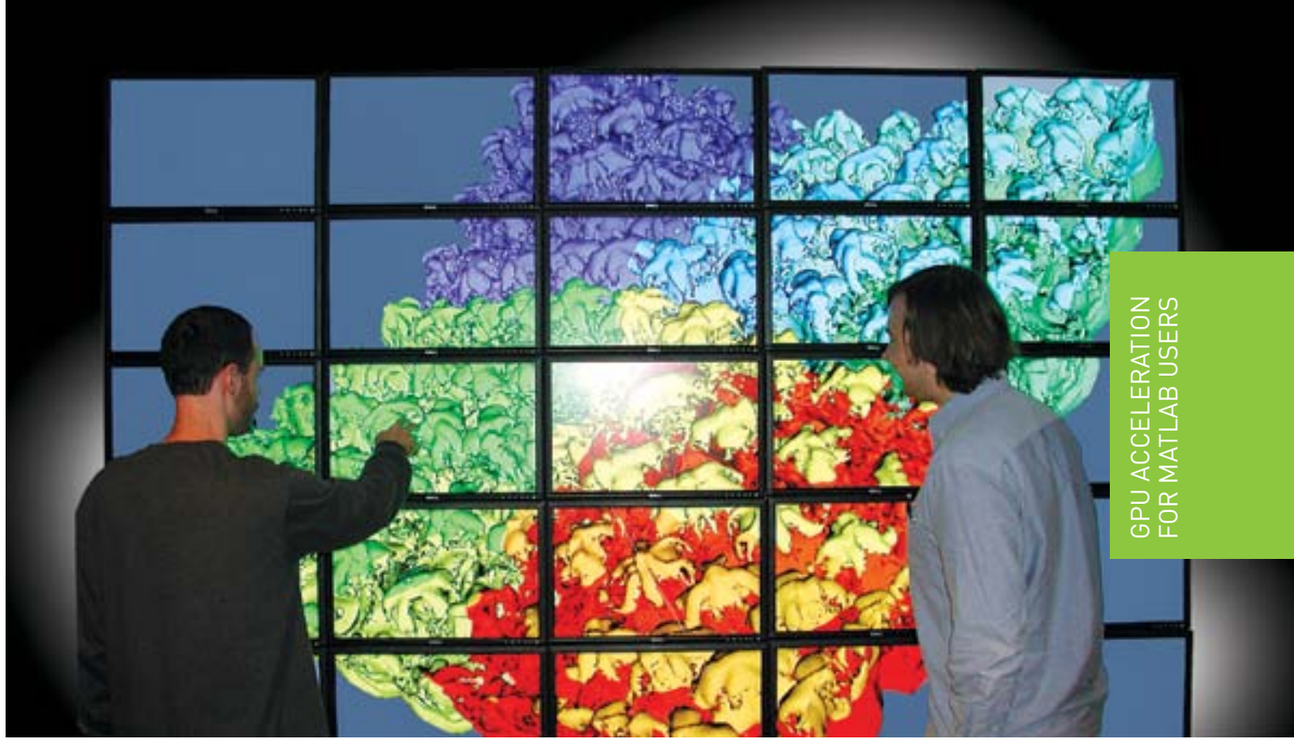
http://www.mathworks.com/products/distriben/demos.html?file=/products/demos/distribtb/MapDemo/MapDemo.html

TESLA BENEFITS

Highest Computational Performance
- High-speed double precision operations
- Large dedicated memory
- High-speed bi-directional PCIe communication
- NVIDIA GPUDirect™ with InfiniBand

Most Reliable
- ECC memory
- Rigorous stress testing

Best Supported
- Professional support network
- OEM system integration
- Long-term product lifecycle
- 3 year warranty
- Cluster & system management tools (server products)
- Windows remote desktop support



GPU ACCELERATION FOR MATLAB USERS

Image - courtesy of the Maryland CPU-GPU Cluster team.

ACCELEREYES

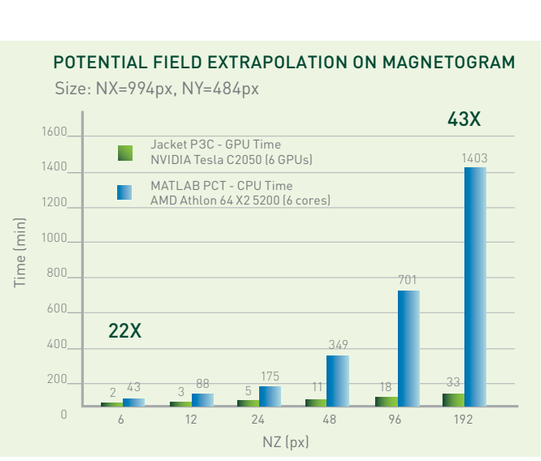
GPU COMPUTING IN MATLAB WITH ACCELEREYES JACKET

Jacket includes many key features to deliver results on full applications:

- Over 500 functions, including math, signal processing, image processing, and statistics
- Specialized FOR loops to run many iterations simultaneously
- An optimized runtime to optimize memory bandwidth and kernel configurations
- Integrate users' own CUDA kernels into MATLAB via the Jacket SDK
- Compute across multiple NVIDIA GPUs via Jacket MGL and HPC

With Jacket programming, the MATLAB community can enjoy GPU-acceleration with an easy, high-level interface.

POTENTIAL FIELD EXTRAPOLATION ON MAGNETOGRAM
Size: NX=994px, NY=484px



RECOMMENDED TESLA & QUADRO CONFIGURATIONS

High-End Workstation
- Two Tesla C2070/ C2075 GPUs
- Quadro 2000
- Two quad-core CPUs
- 24 GB system memory

Mid-Range Workstation
- Tesla C2070/C2075 GPUs
- Quadro 400
- Quad-core CPU
- 12 GB system memory

Entry Workstation
- Tesla C2070/C2075 GPUs
- Quadro 400
- Single quad-core CPU
- 6 GB system memory

JOIN THE RENDERING REVOLUTION

AND MAXIMIZE YOUR SPEED TO PHOTOREALISM WITH IRAY AND NVIDIA FERMI CLASS GPUs

The iray workflow is further enhanced with massive acceleration from NVIDIA Graphics Processing Units (GPUs) based on the NVIDIA® CUDA™ architecture. While iray produces identical images on either CPUs or GPUs, 3ds Max users will enjoy up to 6x faster results over a quad-core CPU[2] when using an NVIDIA® Quadro® 5000 or NVIDIA® Tesla™ C2050. Designers looking for the fastest iray results can further boost their speed by adding additional NVIDIA GPUs to their system.

The massive processing power and large frame buffers of the new Fermi-based Quadro and Tesla GPUs dramatically decrease time to photorealistic results.

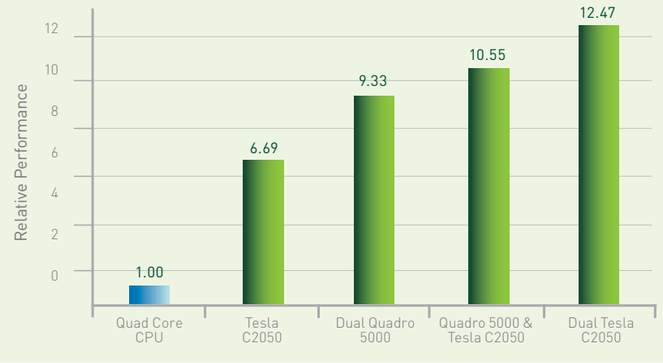
[1] 3ds Max 2011 subscription content required
[2] 3ds Max 2011 64-bit on Win 7 64-bit with 8GB of system memory using a Quadro 5000 or Tesla C2050 vs. an Intel® Q9300 quad core processor
[3] Benchmark testing conducted on Intel Core 2 Quad (Q9300) @ 2.5GHz with 8gb system memory with Windows 7 64-bit. Average results recorded across several 3ds Max scenes

RECOMMENDED CONFIGURATIONS

IRAY SPECIALIST
Quadro 5000 + Tesla C2050
> The visualization specialist wanting to increase the performance of iray
> Delivers up to 10x iray speed up over a quad core CPU

IRAY EXPERT
Quadro 5000 + (3) Tesla C2050
> The visualization expert needing the fastest possible iray results
> Delivers a blazing +20x iray speed up over a quad core CPU

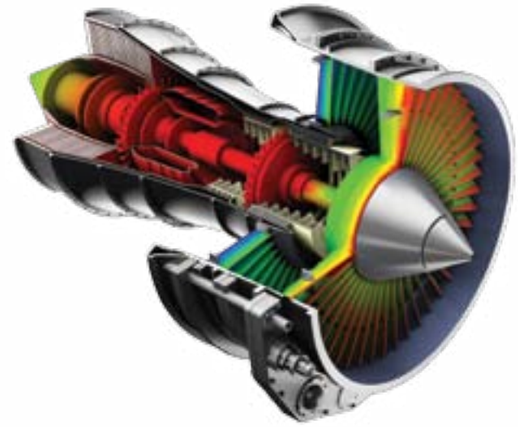
3DS MAX IRAY RENDERING BENCHMARK[3]



WATCH VIDEO*:

VISUALISE AS NEVER BEFORE WITH NVIDIA MAXIMUS. NVIDIA'S CEO JEN-HSUN HUANG'S KEYNOTE AT SC11

* USE YOUR PHONE, SMARTPHONE OR TABLET PC WITH QR READER SOFTWARE TO READ THE QR CODE.



"A new feature in ANSYS Mechanical leverages graphics processing units to significantly lower solution times for large analysis problem sizes."

By Jeff Beisheim,
Senior Software Developer, ANSYS, Inc.

ANSYS

SPEED UP ANSYS SIMULATIONS WITH A GPU

With ANSYS® Mechanical™ 14.0 and NVIDIA® Professional GPUs, you can:

- Improve product quality with 2x more design simulations
- Accelerate time-to-market by reducing engineering cycles
- Develop high fidelity models with practical solution times

How much more could you accomplish if simulation times could be reduced from one day to just a few hours? As an engineer, you depend on ANSYS Mechanical to design high quality products efficiently. To get the most out of ANSYS Mechanical 14.0, simply upgrade your Quadro GPU or add a Tesla GPU to your workstation, or configure a server with Tesla GPUs, and instantly unlock the highest levels of ANSYS simulation performance.

ACTIVATE A GPU WITH ANSYS HPC PACK

To unlock the GPU feature in ANSYS Mechanical 14.0, you must have an ANSYS HPC Pack license, the same scheme also required for going parallel for greater than 2 CPU cores.

For academic license users, the GPU capability is included with the base ANSYS Academic license that provides access to ANSYS Mechanical and no add-on Academic HPC licenses are required.

FUTURE DIRECTIONS

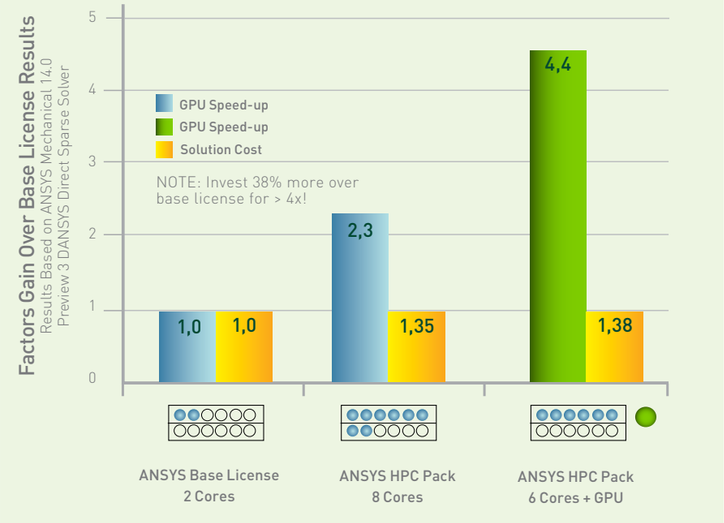
As GPU computing trends evolve, ANSYS will continue to enhance its offerings as necessary for a variety of simulation products. Certainly, performance improvements will continue as GPUs become computationally more powerful and extend their functionality to other areas of ANSYS software.

RECOMMENDED CONFIGURATIONS

Workstation
- Tesla C2075 + Quadro 2000

or
- Quadro 6000
- 48 GB System Memory

Server
- 2x Tesla M2090
- Dual-socket CPU
- 128 GB System Memory



Results from HP Z800 Workstation, 2 x Xeon X5670 2.93GHz 48GB memory, CentOS 5.4 x64; Tesla C2075, CUDA 4.0.17

LBULTRA PLUG-IN FOR RTT DELTAGEN

LBULTRA DEVELOPED BY FLUIDYNA

FAST FLOW SIMULATIONS DIRECTLY WITHIN RTT DELTAGEN

The flow simulation software LBultra works particularly fast on graphics processing units (GPUs). As a plug-in prototype, it is tightly integrated into the high-end 3D visualization software RTT DeltaGen by RTT. As a consequence, flow simulation computing can be proceeded directly within RTT DeltaGen.

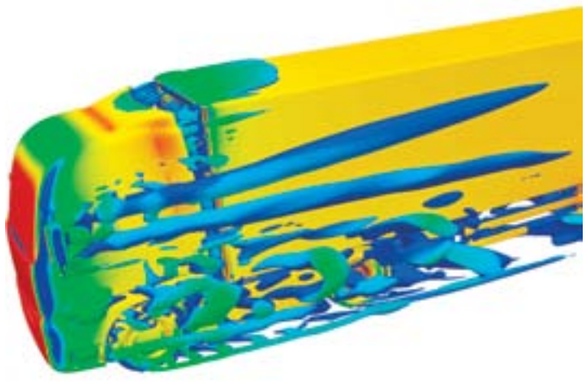
Due to this coupling system, the designer can do a direct flow simulation of their latest design draft – enabling the designer to do a parallel check of aerodynamic features of the vehicle's design draft.

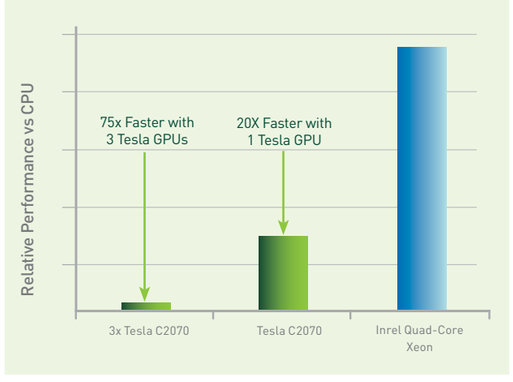
IMPROVED OPTIMUM PERFORMANCE THROUGH EARLY INVOLVEMENT OF AERODYNAMICS INTO DESIGN

First of all, a certain scenario is selected, such as analyzing a spoiler or an outside mirror. Next, various simulation parameters and boundary conditions such as flow rates or resolution levels are set, which also influence the calculation time and result's accuracy.

After determining the overall simulation, the geometry of the design is handed over to LBultra and the simulation is started. While the simulation is running, data is being visualized in realtime in RTT DeltaGen.

In addition, there is an opportunity of exporting simulation results. These results may then, for instance, be further analyzed in more detail by experts using highly-capable fluid mechanics specialist programs and tools.





GPUs ACCELERATE CFD AND STRUCTURAL MECHANICS

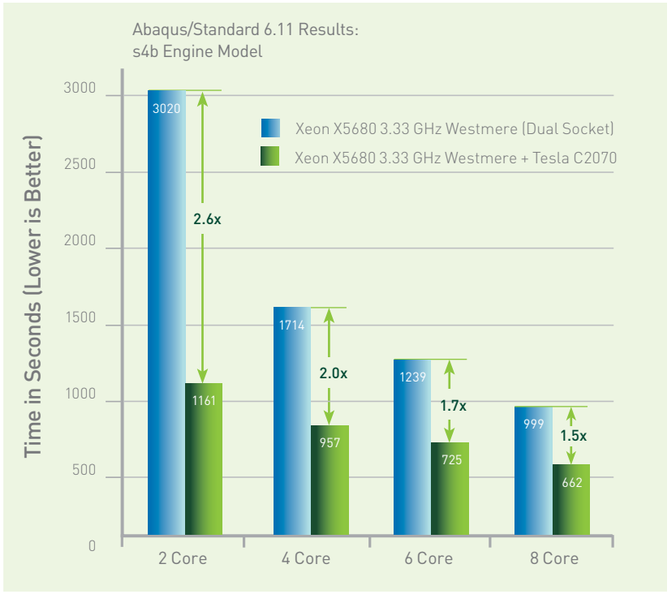
SIMULIA Abaqus / Standard

As products get more complex, the task of innovating with more confidence has been ever increasingly difficult for product engineers. Engineers rely on Abaqus to understand behavior of complex assembly or of new materials.

Recent performance studies conducted together with SIMULIA engineers demonstrated that the Tesla C2070 GPU acceleration of an Intel Xeon 5680 (Westmere) CPU with 2 cores greater than 2.6x and for 4 cores greater than 2x.

Tesla and Quadro GPU computing products are designed to deliver the highest computational performance with the most reliable numerical accuracy, and are available and supported by the world's leading professional system manufacturers.

SIMULIA



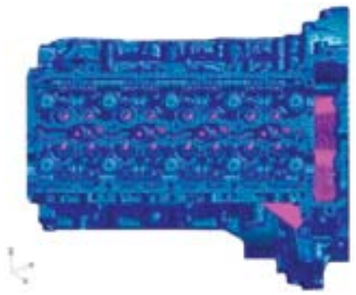
RECOMMENDED TESLA CONFIGURATIONS

Workstation
- Tesla C2070
- Dual-socket Quad-core CPU
- 48 GB System Memory

Server
- 2x Tesla M2090
- Dual-socket Quad-core CPU
- 128 GB System Memory

MSC NASTRAN 2012

5X PERFORMANCE BOOST WITH SINGLE GPU OVER SINGLE CORE, >1.5X WITH 2 GPUS OVER 8 CORE



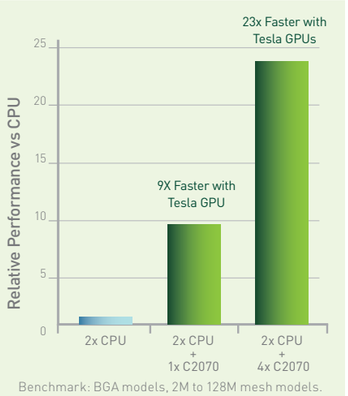
- Nastran direct equation solver is GPU accelerated
  - Real, symmetric sparse direct factorization
  - Handles very large fronts with minimal use of pinned host memory
  - Impacts SOL101, SOL103, SOL400 that are dominated by MSCLDL factorization times
  - More of Nastran (SOL108, SOL111) will be moved to GPU in stages

- Support of multi-GPU and for both Linux and Windows
  - With DMP> 1, multiple fronts are factorized concurrently on multiple GPUs; 1 GPU per matrix domain
  - NVIDIA GPUs include Tesla 20-series and Quadro 6000
  - CUDA 4.0 and above

20X FASTER SIMULATIONS WITH GPUs

DESIGN SUPERIOR PRODUCTS WITH CST MICROWAVE STUDIO

What can product engineers achieve if a single simulation run-time reduced from 48 hours to 3 hours? CST Microwave Studio is one of the most widely used electromagnetic simulation software and some of the largest customers in the world today are leveraging GPUs to introduce their products to market faster and with more confidence in the fidelity of the product design.



Benchmark: BGA models, 2M to 128M mesh models.
CST MWS, transient solver.
CPU: 2x Intel Xeon X5620.
Single GPU run only on 50M mesh model.

RECOMMENDED TESLA CONFIGURATIONS

Workstation
- 4x Tesla C2070
- Dual-socket Quad-core CPU
- 48 GB System Memory

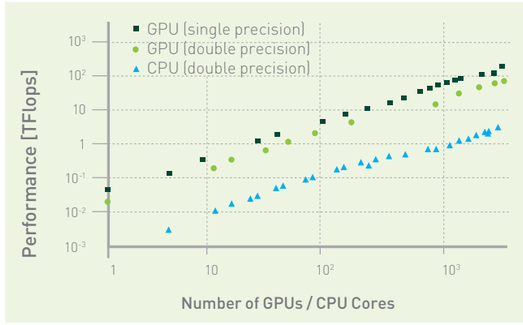
Server
- 4x Tesla M2090
- Dual-socket Quad-core CPU
- 48 GB System Memory

NVIDIA GPUs ACCELERATE WEATHER PREDICTION

ASUCA (WEATHER MODELING) JAPAN'S TERASCALE WEATHER SIMULATION

Regional weather forecasting demands fast simulation over fine-grained grids, resulting in extremely memory-bottlenecked computation. ASUCA is the first high- resolution weather prediction model ported fully to CUDA.

ASUCA is a next-generation, production weather code developed by the Japan Meteorological Agency, similar to WRF in the underlying physics (non-hydrostatic model).

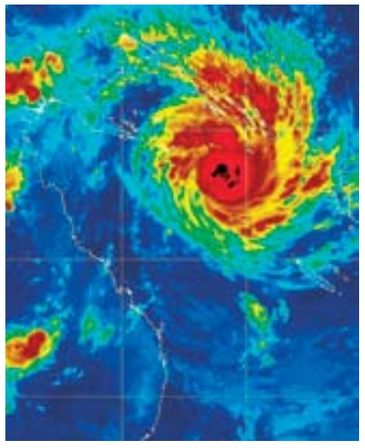


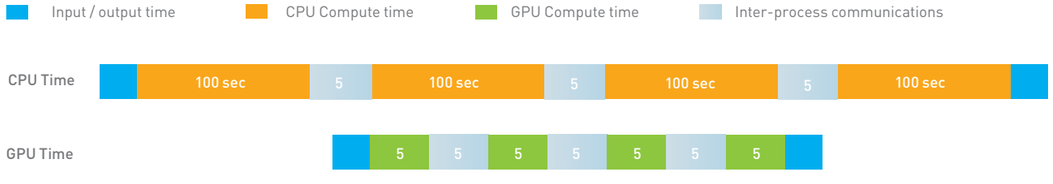
NOAA NIM EXPLORING GPU COMPUTING TO REFINE WEATHER FORECASTING

Earth System Research Lab in the National Oceanic & Atmospheric Administration (NOAA) of the United States has developed a next generation global model to more accurately and efficiently forecast weather. The Non-hydrostatic Icosahedral Model (NIM) uses the icosahedral horizontal grid and is designed to run on thousands of processors including Tesla GPUs.

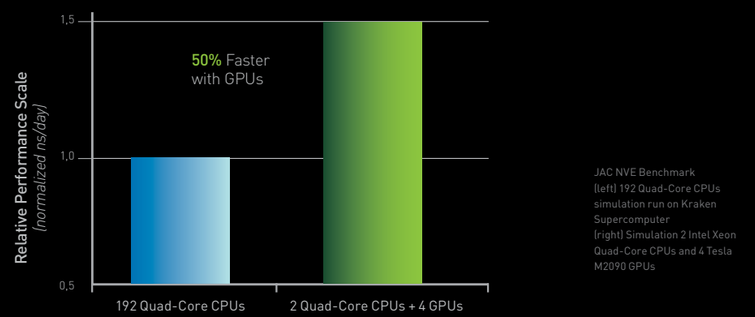
NIM Dynamics package has been ported over to CUDA for single GPU implementation. NOAA is actively working on the code to run parallel jobs on multiple GPUs.

The benchmark below shows ~20x improvement in computation time by using GPUs compared to a 1 core Nehalem Intel CPU. Due to significant computational speedup with GPUs, inter-process communication now becomes the bottleneck in simulation.





source: http://www.esrl.noaa.gov/research/events/espc/8sep2010/Govett_ESRL_GPU.pdf



JAC NVE Benchmark
(left) 192 Quad-Core CPUs simulation run on Kraken Supercomputer
(right) Simulation 2 Intel Xeon Quad-Core CPUs and 4 Tesla M2090 GPUs

16 CPUs
0.36 ns/day
7,737 kJ

Processors
10x Faster
4x Energy Savings

16 CPUs + 24 GPUs
3.44 ns/day
1,142 kJ

"Molecular Dynamics Simulation of a Biomolecule with High Speed, Low Power and Accuracy Using GPU-Accelerated TSUBAME2.0 Supercomputer", Du, Udagawa, Endo, Sekijima

TESLA™ BIO WORKBENCH:
AMBER ON GPUs

ENABLING NEW SCIENCE

TESLA™ BIO WORKBENCH

The NVIDIA Tesla Bio Workbench enables biophysicists and computational chemists to push the boundaries of life sciences research. It turns a standard PC into a "computational laboratory" capable of running complex bioscience codes, in fields such as drug discovery and DNA sequencing, more than 10-20 times faster through the use of NVIDIA Tesla GPUs.

It consists of bioscience applications; a community site for downloading, discussing, and viewing the results of these applications; and GPU-based platforms.

Complex molecular simulations that had been only possible using supercomputing resources can now be run on an individual workstation, optimizing the scientific workflow and accelerating the pace of research. These simulations can also be scaled up to GPU-based clusters of servers to simulate large molecules and systems that would have otherwise required a supercomputer.

Applications that are accelerated on GPUs include:

- **Molecular Dynamics & Quantum Chemistry**
  AMBER, GROMACS, GAMESS, HOOMD, LAMMPS, NAMD, TeraChem (Quantum Chemistry), VMD
- **Bio Informatics**
  CUDA-BLASTP, CUDA-EC, CUDA-MEME, CUDASW++ (Smith-Waterman), GPU-HMMER, MUMmerGPU

For more information, visit:
www.nvidia.com/bio_workbench

WATCH VIDEO*:
THE COMPUTATIONAL SCIENCE BEHIND SHAMPOO

* USE YOUR PHONE, SMARTPHONE OR TABLET PC WITH QR READER SOFTWARE TO READ THE QR CODE.

LEVERAGE SUPERCOMPUTER-LIKE PERFORMANCE FOR YOUR AMBER RESEARCH WITH TESLA GPUs

AMBER

Researchers today are solving the world's most challenging and important problems. From cancer research to drugs for AIDS, computational research is bottlenecked by simulation cycles per day. More simulations mean faster time to discovery. To tackle these difficult challenges, researchers frequently rely on national supercomputers for computer simulations of their models.

GPUs offer every researcher supercomputer-like performance in their own office. Benchmarks have shown four Tesla M2090 GPUs significantly outperforming the existing world record on CPU-only supercomputers.

TRY THE TESLA SIMCLUSTER

A ready to use solution
for science and research.
www.nvidia.eu/cluster

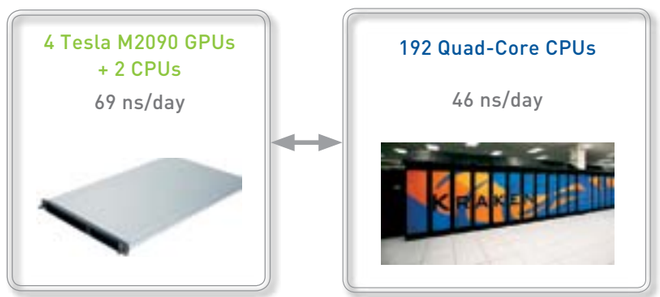
RECOMMENDED HARDWARE CONFIGURATION

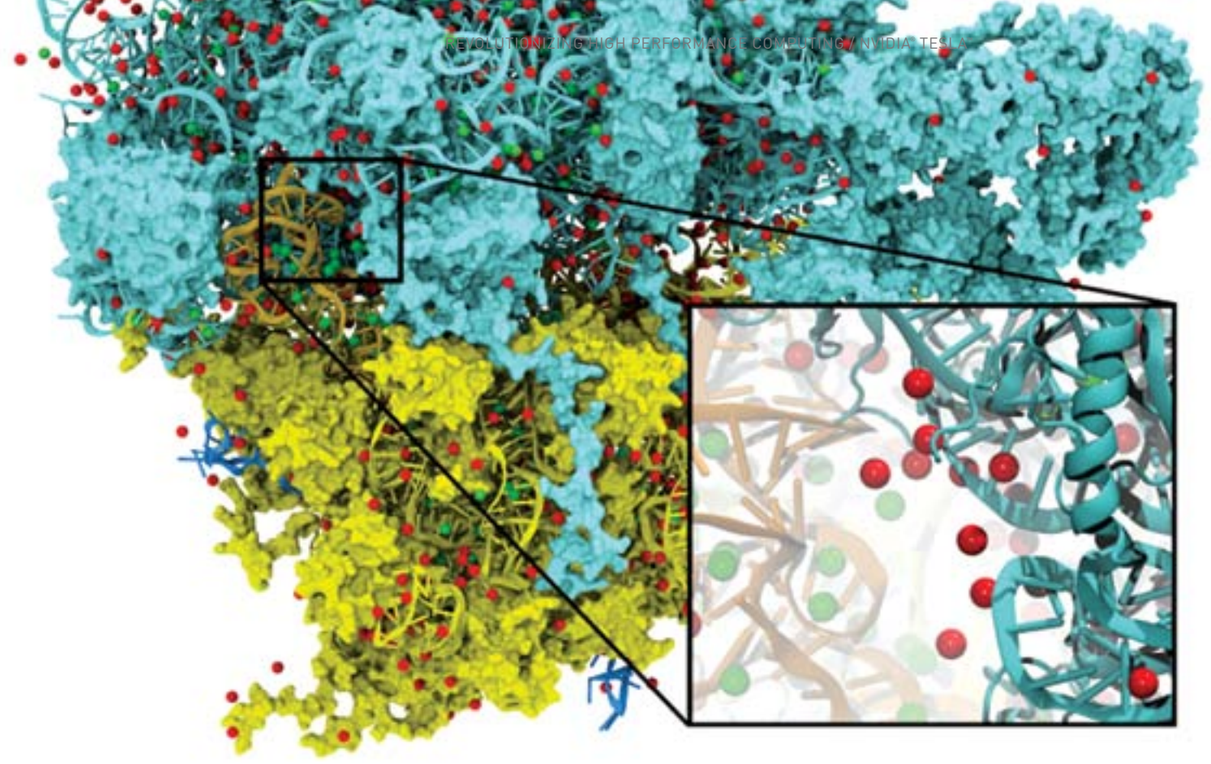
Workstation

- 4xTesla C2070
- Dual-socket Quad-core CPU
- 24 GB System Memory Server
- Up to 8x Tesla M2090s in cluster
- Dual-socket Quad-core CPU per node
- 128 GB System Memory

Server

- 8x Tesla M2090
- Dual-socket Quad-core CPU
- 128 GB System Memory





CUDA-BLASTP

CUDA-BLASTP is designed to accelerate NCBI BLAST for scanning protein sequence databases by taking advantage of the massively parallel CUDA architecture of NVIDIA Tesla GPUs. CUDA-BLASTP also has a utility to convert FASTA format database into files readable by CUDA-BLASTP.

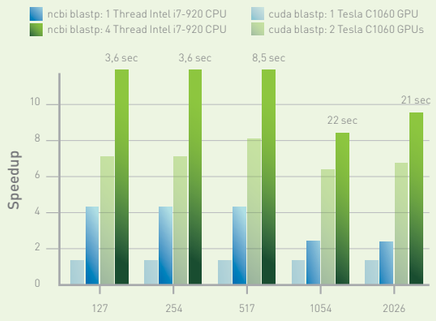
GPU-HMMER

GPU-HMMER is a bioinformatics software that does protein sequence alignment using profile HMMs by taking advantage of the massively parallel CUDA architecture of NVIDIA Tesla GPUs. GPU-HMMER is 60-100x faster than HMMER (2.0).

CUDA-BLASTP running on a workstation with two Tesla C1060 GPUs is 10x faster than NCBI BLAST (2.2.22) running on an Intel i7-920 CPU. This cuts compute time from minutes of CPUs to seconds using GPUs.

CUDA-BLASTP vs NCBI BLASTP Speedups

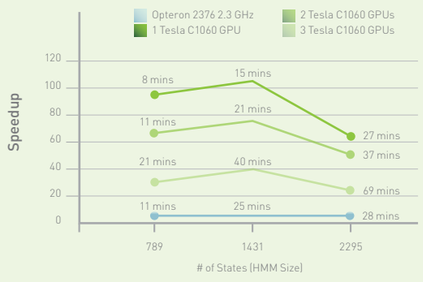
Data courtesy of Nanyang Technological University, Singapore



GPU-HMMER accelerates the hmmsearch tool using GPUs and gets speed-ups ranging from 60-100x. GPU-HMMER can take advantage of multiple Tesla GPUs in a workstation to reduce the search from hours on a CPU to minutes using a GPU.

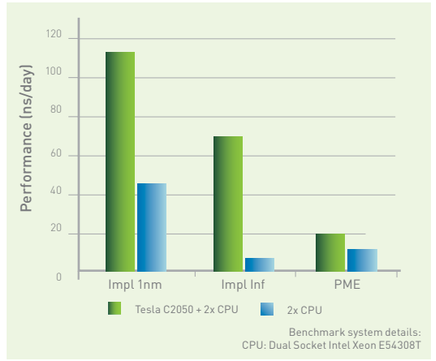
HMMER: 60-100X FASTER

Data courtesy of Nanyang Technological University, Singapore



GROMACS

GROMACS is a molecular dynamics package designed primarily for simulation of biochemical molecules like proteins, lipids, and nucleic acids that have a lot complicated bonded interactions. The CUDA port of GROMACS enabling GPU acceleration supports Particle-Mesh-Ewald (PME), arbitrary forms of non-bonded interactions, and implicit solvent Generalized Born methods.



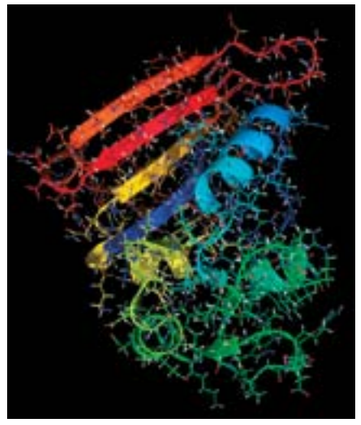
RECOMMENDED HARDWARE CONFIGURATION

Workstation

- 1xTesla C2070
- Dual-socket Quad-core CPU
- 12 GB System Memory

CUDA-MEME

CUDA-MEME is a motif discovery software based on MEME (version 3.5.4). It accelerates MEME by taking advantage of the massively parallel CUDA architecture of NVIDIA Tesla GPUs. It supports the OOPS and ZOOPS models in MEME.



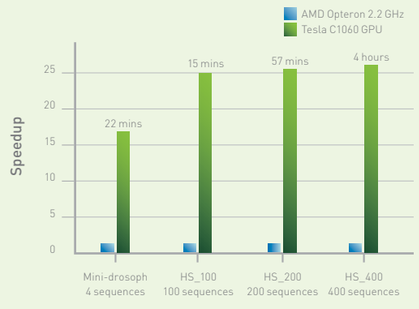
CUDA-EC

CUDA-EC is a fast parallel sequence error correction tool for short reads. It corrects sequencing errors in high-throughput short-read (HTSR) data and accelerates HTSR by taking advantage of the massively parallel CUDA architecture of NVIDIA Tesla GPUs. Error correction is a preprocessing step for many DNA fragment assembly tools and is very useful for the new high-throughput sequencing machines.

CUDA-MEME running on one Tesla C1060 GPU is up to 23x faster than MEME running on a x86 CPU. This cuts compute time from hours on CPUs to minutes using GPUs. The data in the chart below are for the OOPS (one occurrence per sequence) model for 4 datasets.

CUDA-MEME VS MEME SPEEDUPS

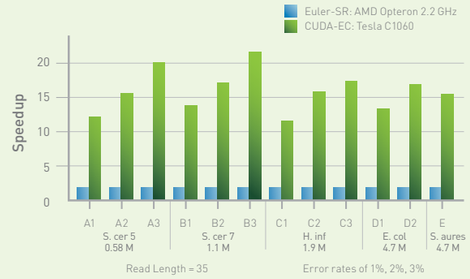
Data courtesy of Nanyang Technological University, Singapore



CUDA-EC running on one Tesla C1060 GPU is up to 20x faster than Euler-SR running on a x86 CPU. This cuts compute time from minutes on CPUs to seconds using GPUs. The data in the chart below are error rates of 1%, 2%, and 3% denoted by A1, A2, A3 and so on for 5 different reference genomes.

CUDA-EC vs Euler-SR Speedups

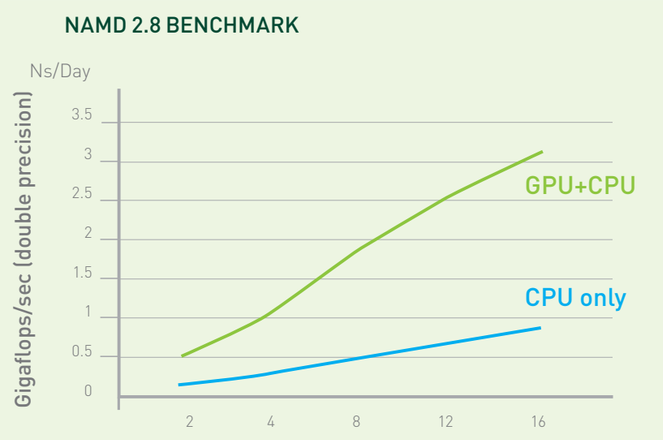
Data courtesy of Nanyang Technological University, Singapore



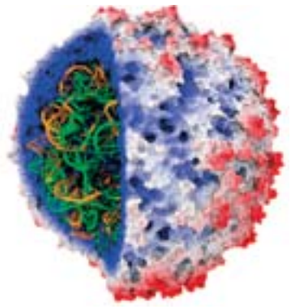
NAMD

The Team at University of Illinois at Urbana-Champaign (UIUC) has been enabling CUDA-acceleration on NAMD since 2007, and the results are simply stunning. NAMD users are experiencing tremendous speed-ups in their research using Tesla GPU. Benchmark (see below) shows that 4 GPU server nodes out-perform 16 CPU server nodes. It also shows GPUs scale-out better than CPUs with more nodes.

Scientists and researchers equipped with powerful GPU accelerators have reached new discoveries which were impossible to find before. See how other computational researchers are experiencing supercomputer-like performance in a small cluster, and take your research to new heights.



NAMD 2.8 B1, STMV Benchmark
A Node is Dual-Socket, Quad-core x5650 with 2 Tesla M2070 GPUs
NAMD courtesy of Theoretical and Computational Bio-physics Group, UIUC

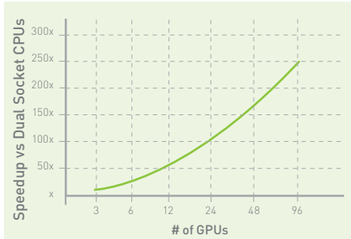


LAMMPS

LAMMPS is a classical molecular dynamics package written to run well on parallel machines and is maintained and distributed by Sandia National Laboratories in the USA. It is a free, open-source code.

LAMMPS has potentials for soft materials (biomolecules, polymers) and solid-state materials (metals, semiconductors) and coarse-grained or mesoscopic systems.

The CUDA version of LAMMPS is accelerated by moving the force calculations to the GPU.



Benchmark system details:
Each node has 2 CPUs and 3 GPUs. CPU is Intel Six-Core Xeon
Speedup measured with Tesla M2070 GPU vs without GPU
Source: http://users.nccs.gov/~wb8/gpu/kid_single.htm

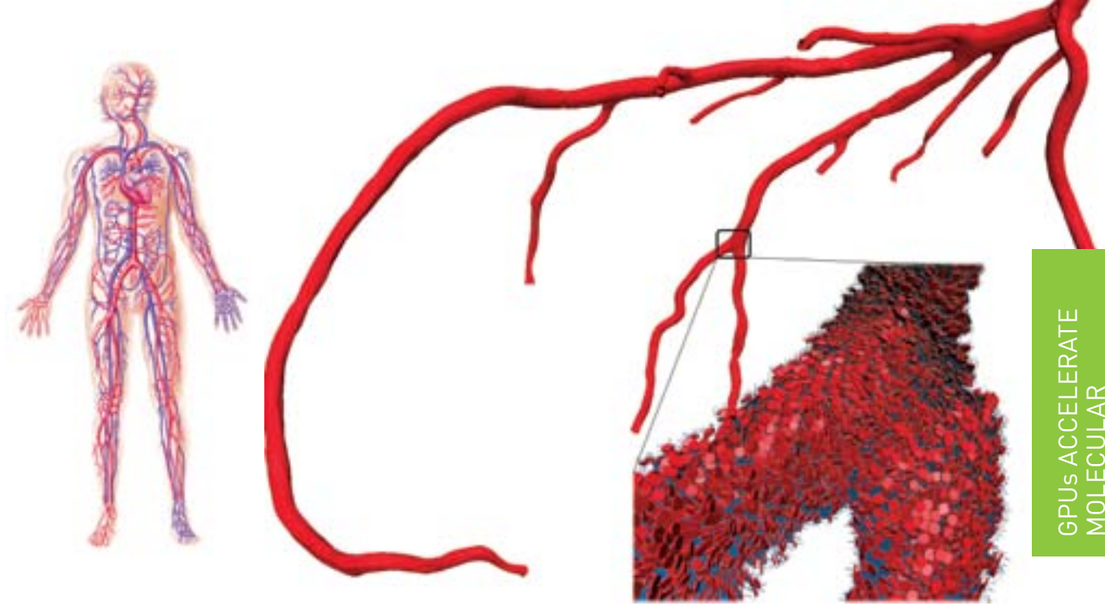
RECOMMENDED HARDWARE CONFIGURATION

Workstation

- 4xTesla C2070
- Dual-socket Quad-core CPU
- 24 GB System Memory

Server

- 2x-4x Tesla M2090 per node
- Dual-socket Quad-core CPU per node
- 128 GB System Memory per node



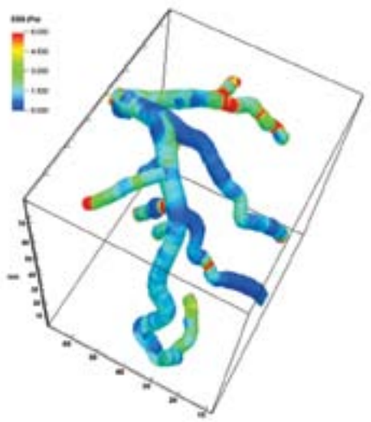
GPUs ACCELERATE MOLECULAR DYNAMICS AND MEDICAL IMAGING

NVIDIA'S TECHNOLOGY SET TO SPAWN CORONARY CAT SCANNER OF THE FUTURE

THE CHALLENGE

Many deaths from heart failure are caused by problems in the coronary arteries and by pathologies that are visible, to varying degrees, using traditional methodologies.

The study of blood flows involves reconstructing the precise geometry of the network of arteries that surrounds the heart, which can be mapped using images deriving from high-resolution computerised tomography. To map these sections a machine of similar design to that of a CAT scanner, but with much higher resolution is being used. The higher the number of sections mapped, the higher the fidelity of the model.



THE SOLUTION

The practice is incredibly complicated. First - the need to define what is coronary artery and what is not, within images that are far from easy to interpret. This done, specially developed software is used to attempt to interpret the coronary map and then define the blood flow, thus putting the ambitious project into practice.

This obviously takes an enormous toll in terms of processing power, given that the result needs to be produced in real time. The interpolation of the images and subsequent creation of the map and the fluid dynamics model requires processing power of between 10 and 20 Teraflops, which is possible thanks to the high power of NVIDIA TESLA GPUs. GPUs have already been successfully applied to supercomputing, but now they are being used for medical applications too, because they offer higher analytical capacity than traditional diagnostic systems.

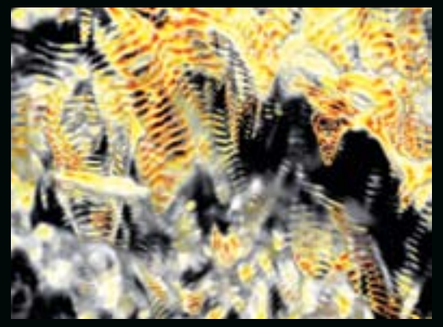
THE IMPACT

The system is still at the testing stage and it will be some time before it is widely used in national health service hospitals. However, the developments that it may spawn in the medical/diagnostic arena, thanks to the processing power of the GPUs, are not hard to imagine.

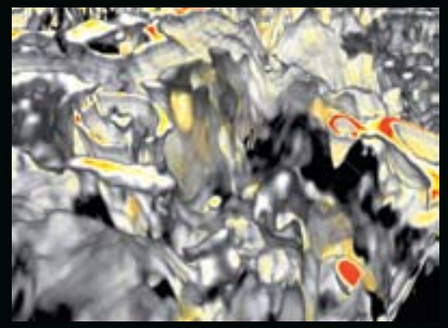
During the testing phase, a serious anomaly was detected in a member of the research staff, who showed no outward signs of ill health but had volunteered to take part in the initial experiments out of pure curiosity. An emergency operation prevented the worst from happening, thus pointing the way to extremely promising scenarios in the field of preventive medicine. A plaque of cholesterol was on the verge of entirely obstructing one of the coronary arteries, but the problem was detected thanks to the 3D view of the anomalous blood flow.

This use of 3D graphics in diagnostic applications will save many human lives in the future, considering that, in the experimentation stage alone, it saved the life of a member of the research project team.

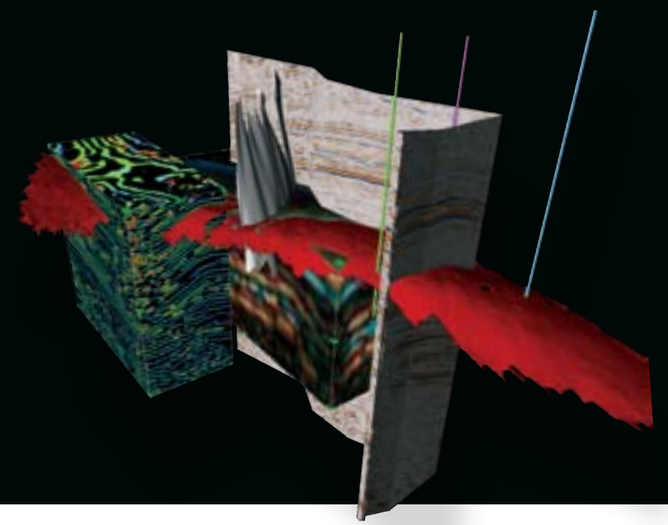
BEFORE



AFTER



GPU enhanced voxel rendering providing, improved visualization of karsts in Barnett Shale



OIL & GAS: Paradigm VoxelGeo

REDEFINING INTERPRETATION WITH NVIDIA TESLA GPUs

THE CHALLENGE

Every energy company has a goal to make the best of their large drilling budgets. Drilling wells in the right place, and with accurate knowledge of the associated risks, really impacts the bottom line. When improved accuracy leads to more pay zone and less uncertainty about the production process, the ROI improves significantly.

Paradigm software aims at giving customers as much information as possible out of an existing seismic dataset, where some contractors may be advising the customer to shoot new data for better resolution. Tangible evidence for this is the ability to drill more productive feet in a well (GeoSteering, accurate models, fine-tuned depth conversion, better characterization leading to drilling in sweet spots).

HSE is a critical topic, where nothing can be left out. Avoiding overpressure zones, identifying fracture areas that could affect drilling all these factors should be taken into consideration, carefully processed and visualized with maximum detail. Working efficiently to reduce project turnaround without compromising quality has a definite impact on customer performance. It also implies that the data volumes are growing and with the users' need to see more detail comes increasingly compute intensive data processing.

THE SOLUTION

The only adequate solution would be a re-write of the most compute-intensive parts of the rendering engine using CUDA programming on NVIDIA GPUs to harness the compute power of massively parallel processors. The GPU-enabled solution should work effectively with both Windows and Linux. A minimal code-porting effort would provide a sizeable improvement in the overall efficiency shortening the overall project time as a result.

THE IMPACT

GPU rendering delivers 10x to 40x acceleration, increasing the massive data crunching efficiency by an order of magnitude. Users get a much more realistic rendering in 3D Taking voxel visualization to the next level. Porting the key processes to NVIDIA GPU architecture has brought about a new rendering process with significantly improved spatial perception and real-time computation of key attributes. Further integration of the CUDA – enabled code to the interpretation suite has produced a notable improvement in display quality, making it easier to perceive depth and volume.

As a result VoxelGeo remains the reference voxel product and the product of choice for industry leaders.

ffA

CHANGING THE WAY 3D SEISMIC INTERPRETATION IS CARRIED OUT

THE CHALLENGE

The data produced by seismic imaging surveys is crucial for identifying the presence of hydrocarbon reserves and understanding how to extract it.Today, geoscientists must process increasing amounts of data as dwindling reserves require them to pinpoint smaller, more complex reservoirs with greater speed and accuracy.

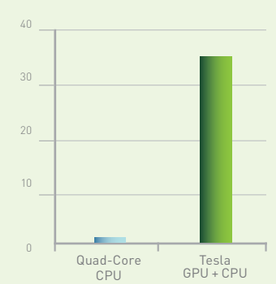
THE SOLUTION

UK-based company ffA provides world leading 3D seismic analysis software and services to the global oil and gas industry. The sophisticated tools providing a greater understanding of complex 3D geology are extremely compute-intensive. With the recently released CUDA enabled 3D seismic analysis application, ffA users routinely achieve over an order of magnitude speed-up compared with performance on high end multi-core CPUs.

THE IMPACT

NVIDIA CUDA is allowing ffA to provide scalable high performance computation for seismic data on hardware platforms equipped with one or more NVIDIA Quadro FX and NVIDIA Tesla GPUs significantly increasing the amount of data that can be analyzed in a given timeframe and allowing them to improve subsurface understanding and reduce risk in oil and gas exploration and exploitation.

STREAMLING WITH GPUs FOR BETTER RESULTS



The latest benchmarking results using Tesla GPUs have produced performance improvements of up to 37x versus high end dual quad core CPU workstations.

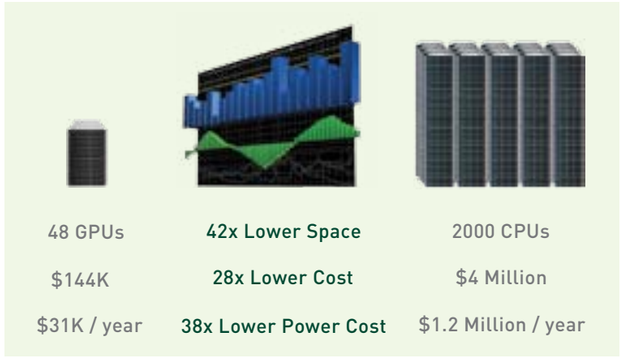
WATCH VIDEO*:
FFA AND NVIDIA GPUs REVOLUTIONIZE THE SEARCH FOR NEW OIL AND GAS FIELDS

* USE YOUR PHONE, SMARTPHONE OR TABLET PC WITH QR READER SOFTWARE TO READ THE QR CODE.

GPU COMPUTING IN FINANCE – CASE STUDIES

BLOOMBERG: GPUs INCREASE ACCURACY AND REDUCE PROCESSING TIME FOR BOND PRICING

Bloomberg implemented an NVIDIA Tesla GPU computing solution in their datacenter. By porting their application to run on the NVIDIA CUDA parallel processing architecture Bloomberg received dramatic improvements across the board. As Bloomberg customers make crucial buying and selling decisions, they now have access to the best and most current pricing information, giving them a serious competitive trading advantage in a market where timing is everything.



48 GPUs	42x Lower Space	2000 CPUs
$144K	28x Lower Cost	$4 Million
$31K / year	38x Lower Power Cost	$1.2 Million / year

NVIDIA TESLA GPUs USED BY J.P. MORGAN TO RUN RISK CALCULATIONS IN MINUTES, NOT HOURS

THE CHALLENGE

Risk management is a huge and increasingly costly focus for the financial services industry. A cornerstone of J.P. Morgan's cost-reduction plan to cut the cost of risk calculation involves accelerating its risk library. It is imperative to reduce the total cost of ownership of J.P. Morgan's risk-management platform and create a leap forward in the speed with which client requests can be serviced.

THE SOLUTION

J.P. Morgan's Equity Derivatives Group added NVIDIA® Tesla M2070 GPUs to its datacenters. More than half the equity derivative-focused risk computations run by the bank have been moved to running on hybrid GPU/CPU-based systems, NVIDIA Tesla GPUs were deployed in multiple data centers across the bank's global offices. J.P. Morgan was able to seamlessly share the GPUs between tens of global applications.

THE IMPACT

Utilizing GPUs has accelerated application performance by 40X and delivered over 80 percent savings, enabling greener data centers that deliver higher performance for the same power. For J.P. Morgan, this is game-changing technology, enabling the bank to calculate risk across a range of products in a matter of minutes rather than overnight. Tesla GPUs give J.P. Morgan a significant market advantage.

DRIVING THE HPC REVOLUTION ON EUROPE'S LARGEST HYBRID RESEARCH CLUSTERS

BSC-CNS TRIPLES ITS CALCULATION CAPACITY

The Barcelona Supercomputing Center – now has a new cluster with graphical accelerators which will be used to consolidate its research in programming models, tool development and application porting. The new BSC-CNS cluster is a special-purpose supercomputer that offers an optimum peak performance for specific applications or programs.

Its advantage, compared to general-purpose machines, is its speed and low energy consumption. Mateo Valero, BSC-CNS Director, says, "We currently have some of the world's best programming models so we are in an optimal position to provide some specific applications the most efficient usage of the new system, offering a significant increase in their performance".

About BSC

The mission of Barcelona Supercomputing Center is to research, develop and manage information technology in order to facilitate scientific progress. With this objective, there has been taken special dedication to research areas such as Computer Sciences, Life Sciences, Earth Sciences and Computational Applications in Science and Engineering.

GPU EXTENSION OF THE CURIE SUPERCOMPUTER AT GENCI

About GENCI

GENCI, Grand Equipement National de Calcul Intensif, is owned for 49 % by the French State, for 20 % by CEA, 20 % by CNRS, 10 % by French Universities and 1% by INRIA. GENCI's role is to set the strategic direction and to make France's most important investments in High-Performance Computing (HPC), in particular as the custodian of France's commitment to PRACE.

As part of its CURIE supercomputer that is currently being installed by Bull, GENCI has acquired 16 bullx blade chassis, housing a total of 144 B505 accelerator blades, i.e. 288 NVIDIA Tesla M2090 GPUs (147,456 GPU cores). The 1.6-petaflops CURIE system will be entirely dedicated to the use of the European PRACE partnership (the Partnership for Advanced Computing in Europe).

TESLA GPU COMPUTING SOLUTIONS

The Tesla 20-series GPU computing solutions are designed from the ground up for high-performance computing and are based on NVIDIA's latest CUDA GPU architecture, code named "Fermi". It delivers many "must have" features for HPC including ECC memory for uncompromised accuracy and scalability, C++ support, and 7x double precision performance compared to CPUs. Compared to the typical quad-core CPUs, Tesla 20-series GPU computing products can deliver equivalent performance at 1/10th the cost and 1/20th the power consumption.

• SUPERIOR PERFORMANCE

Highest double precision floating point performance and large on-board memory to support large HPC data sets

• HIGHLY RELIABLE

Uncompromised data reliability through ECC protection and stress tested for zero error tolerance

• DESIGNED FOR HPC

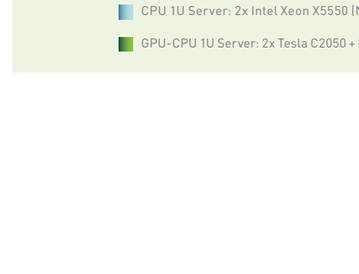
For more information on Tesla GPU computing products and applications, visit www.nvidia.eu/tesla.

TESLA DATA CENTER PRODUCTS

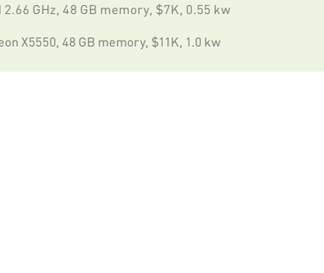
Available from OEMs and certified resellers, Tesla GPU computing products are designed to supercharge your computing cluster.

TESLA WORKSTATION PRODUCTS

Designed to deliver cluster-level performance on a workstation, the NVIDIA Tesla GPU Computing Processors fuel the transition to parallel computing while making personal supercomputing possible — right at your desk.



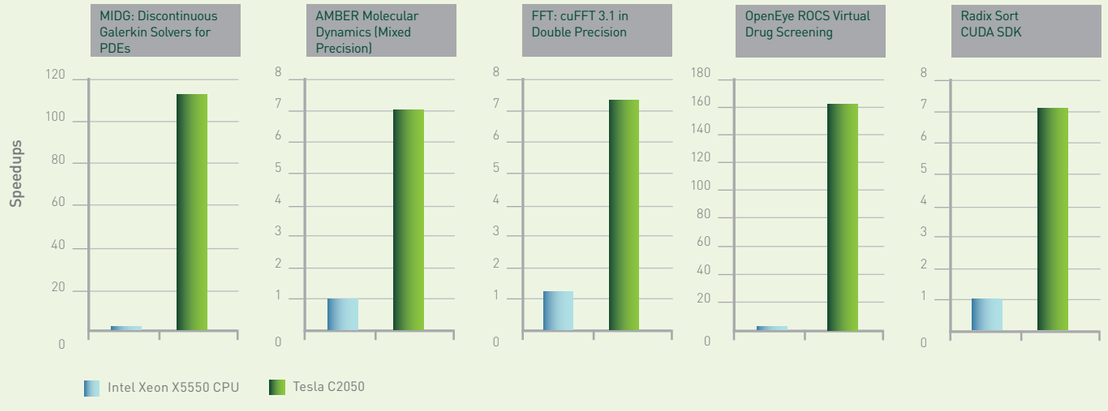
Tesla M2070/M2075/M2090 GPU Computing Modules enables the use of GPUs and CPUs together in an individual server node or blade form factor.



Tesla C2070/C2075 GPU Computing Processor delivers the power of a cluster in the form factor of a workstation.

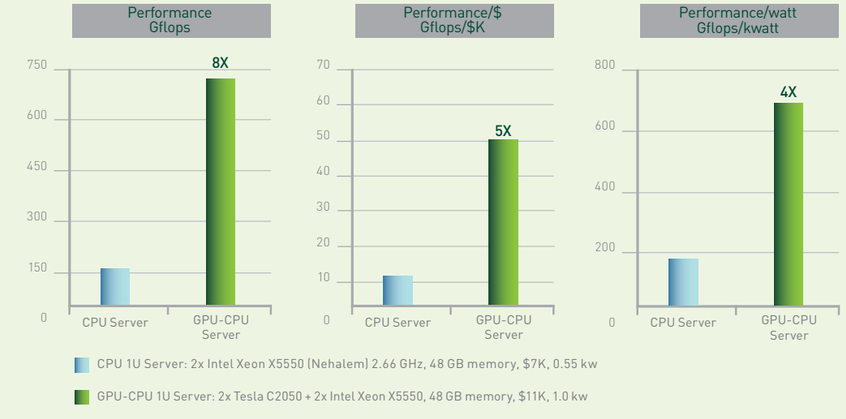
Highest Performance, Highest Efficiency

Workstations powered by Tesla GPUs outperform conventional CPU-only solution in life scince applications.



Highest Performance, Highest Efficiency

GPU-CPU server solutions deliver up to 8x higher Linpack performance.



TESLA C-CLASS GPU COMPUTING PROCESSORS
THE POWER INSIDE PERSONAL MULTI-TERAFLOPS SYSTEMS

NVIDIA CUDA TECHNOLOGY UNLOCKS THE POWER OF TESLA'S MANY CORES

The CUDA C programming environment simplifies many-core programming and enhances performance by offloading computationally-intensive activities from the CPU to the GPU. It enables developers to utilise NVIDIA GPUs to solve the most complex computation intensive challenges such as protein docking, molecular dynamics, financial analysis, fluid dynamics, structural analysis and many others.

The NVIDIA® Tesla™ Personal Supercomputer is based on the revolutionary NVIDIA® CUDA™ parallel computing architecture and powered by up to thousands of parallel processing cores.

YOUR OWN SUPERCOMPUTER

Get nearly 4 teraflops of compute capability and the ability to perform computations 250 times faster than a multi-CPU core PC or workstation.

NVIDIA CUDA UNLOCKS THE POWER OF GPU PARALLEL COMPUTING

The CUDA parallel computing architecture enables developers to utilise C or FORTRAN programming with NVIDIA GPUs to run the most complex computationally intensive applications. CUDA is easy to learn and has become widely adopted by thousands of application developers worldwide to accelerate the most performance demanding applications.

ACCESSIBLE TO EVERYONE

Available from OEMs and resellers worldwide, the Tesla Personal Supercomputer operates quietly and plugs into a standard power strip so you can take advantage of cluster level performance anytime you want, right from your desk.

PETASCALE COMPUTING WITH TERAFLOP PROCESSORS

The NVIDIA Tesla computing card enables the transition to energy efficient parallel computing power by bringing the performance of a small cluster to a workstation. With hundreds of processor cores and a standard C compiler that simplifies application development, Tesla cards scale to solve the world's most important computing challenges more quickly and accurately.

TESLA PERSONAL SUPERCOMPUTER

Cluster Performance on your Desktop	The performance of a cluster in a desktop system. Four Tesla GPU computing processors deliver nearly 4 teraflops of performance.
Designed for Office Use	Plugs into a standard office power socket and quiet enough for use at your desk.
Massively Parallel Many Core GPU architecture	448 parallel processor cores per GPU that can execute thousands of concurrent threads.
High-Speed Memory per GPU	Dedicated compute memory enables larger datasets to be stored locally for each processor minimizing data movement around the system.
IEEE 754 Floating Point Precision (single -precision and double -precision)	Provides results that are consistent across platforms and meet industry standards.

	Tesla C2070	Tesla C2075*
Architecture	Tesla 20-series GPU	
Number of Cores	448	
Caches	64 KB L1 cache + Shared Memory / 32 cores, 768 KB L2 cache	
FP Peak Performance	1,03 TFlops (single) 515 GFlops (double)	
FP Application Efficiency (Tesla C1060 reference)	1,5 - 2 (single) 3 - 4 (double)	
GPU Memory	6 GB GDDR5 ECC 5.25 GB with ECC on	
Memory Bandwidth	144 GB/s ECC off 115 GB/s ECC on	
Video Output	DVI-I	
System I/O	PCIe x16 Gen2 (bi-directional async. transfer)	
Positioning	Best price/performance solutions for double precision codes and when ECC memory required	
Power Consumption	225W	200W

* Extended ECC reporting and Dynamic power scaling capability available.

	Tesla M2070	Tesla M2075*	Tesla M2090
Architecture	**Tesla 20-series GPU**		
Number of Cores	448		512
Caches	64 KB L1 cache + Shared Memory / 32 cores, 768 KB L2 cache		
FP Peak Performance	1030 GFlops (single) 515 GFlops (double)		1331 GFlops (single) 665 GFlops (double)
FP Application Efficiency (Tesla C1060 reference)	1.5 - 2 (single) 3 - 4 (double)		
GPU Memory	6 GB GDDR5 ECC 5,25 GB with ECC on		6 GB GDDR5 ECC 5,25 GB with ECC on
Memory Bandwidth	144 GB/s ECC off 115 GB/s ECC on		
Positioning	Best price/performance solutions for double precision codes and when ECC memory required		
Power Consumption	225W	200W	225W

* Extended ECC reporting and Dynamic power scaling capability available.

TESLA M-CLASS GPU COMPUTING MODULES

DESIGNED FOR DATACENTER INTEGRATION

Tesla M-class GPU Computing Modules enable the seamless integration of GPU computing with host systems for high-performance computing and large data center, scale-out deployments.

SYSTEM MONITORING FEATURES

Tesla Modules deliver all the standard benefits of GPU computing while enabling maximum reliability and tight integration with system monitoring and management tools. This gives data centre IT staff much greater choice in how they deploy GPUs, offering a wide variety of rack-mount and blade systems plus the remote monitoring and remote management capabilities they need.

ENDORSED BY THE HPC OEMS

The Tesla Modules are only available in OEM systems that have been specifically designed and qualified along with NVIDIA engineers. They are designed for maximum reliability; their passive heatsink design eliminates moving parts and cables. Tesla Modules have been selected by the most important HPC vendors world-wide.

Tesla GPU's high performance makes them ideal for seismic processing, biochemistry simulations, weather and climate modeling, signal processing, computational finance, CAE, CFD, and data analystics.

The Tesla 20-series GPU Computing Processors are the first to deliver greater than 10X the double precision horsepower of a quad-core x86 CPU and the first GPUs to deliver ECC memory. Based on the Fermi architecture, these GPUs feature up to 665 gigaflops of double precision performance, 1 teraflops of single precision performance, ECC memory error protection, and L1 and L2 caches.

TRY THE TESLA SIMCLUSTER

A ready to use solution for science and research.
www.nvidia.eu/cluster

NVIDIA TESLA AND CUDA LINKS

NEWS AND SUCCESS STORIES

NVIDIA GPU Computing on Twitter
http://twitter.com/gpucomputing

CUDA weekly newsletter
www.nvidia.com/object/cuda_week_in_review_newsletter.html

News and articles
www.nvidia.co.uk/page/tesla-articles.html

Tesla Video on YouTube
www.youtube.com/nvidiatesla

Success stories
www.nvidia.co.uk/page/tesla_testimonials.html

DEVELOPMENT TOOLS

CUDA Zone
www.nvidia.co.uk/cuda

CUDA in Action
www.nvidia.co.uk/object/cuda_in_action.html

CUDA Books
www.nvidia.co.uk/object/cuda_books.html

CUDA Training & Consulting
www.nvidia.co.uk/page/cuda_consultants.html

Order CUDA course
www.parallel-compute.com

Software Development Tools
www.nvidia.co.uk/object/tesla_software_uk.html

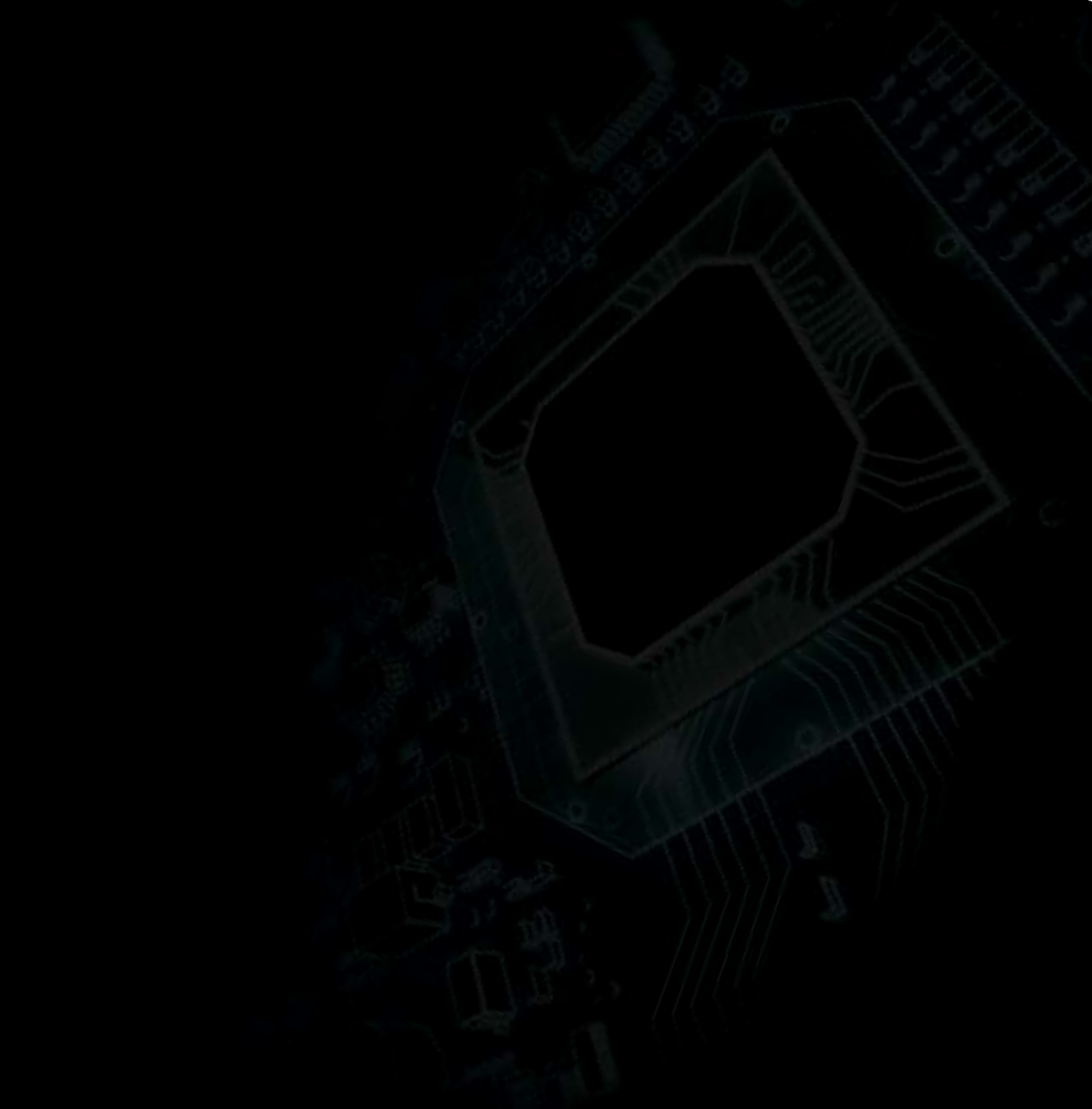
HARDWARE

NVIDIA High Performance Computing
www.nvidia.eu/tesla

Personal Supercomputers with Tesla
www.nvidia.co.uk/psc

Tesla Data Center Solutions
www.nvidia.co.uk/page/preconfigured_clusters.html

Tesla Products Technical Descriptions
www.nvidia.co.uk/page/tesla_product_literature.html



NVIDIA