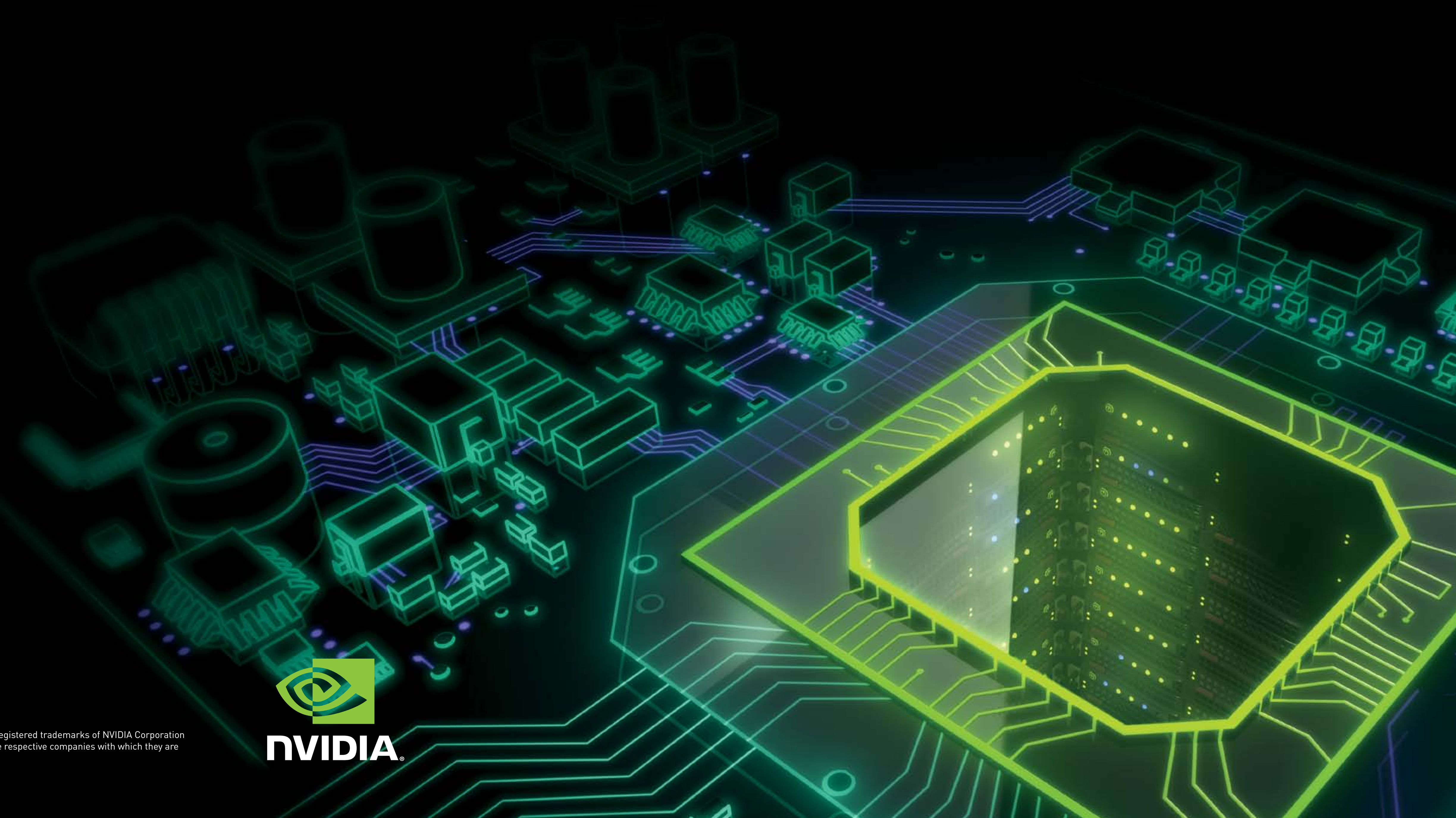
NVIDIA® TESLA® GPU COMPUTING
REVOLUTIONIZING HIGH PERFORMANCE COMPUTING



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

NVIDIA.

GPUS ARE REVOLUTIONIZING COMPUTING

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 Graphics Processing Unit (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.

> *"The convergence of new, fast GPUs optimized for computation as well as 3D graphics acceleration and industry-standard software development tools marks the real beginning of the GPU computing era."*

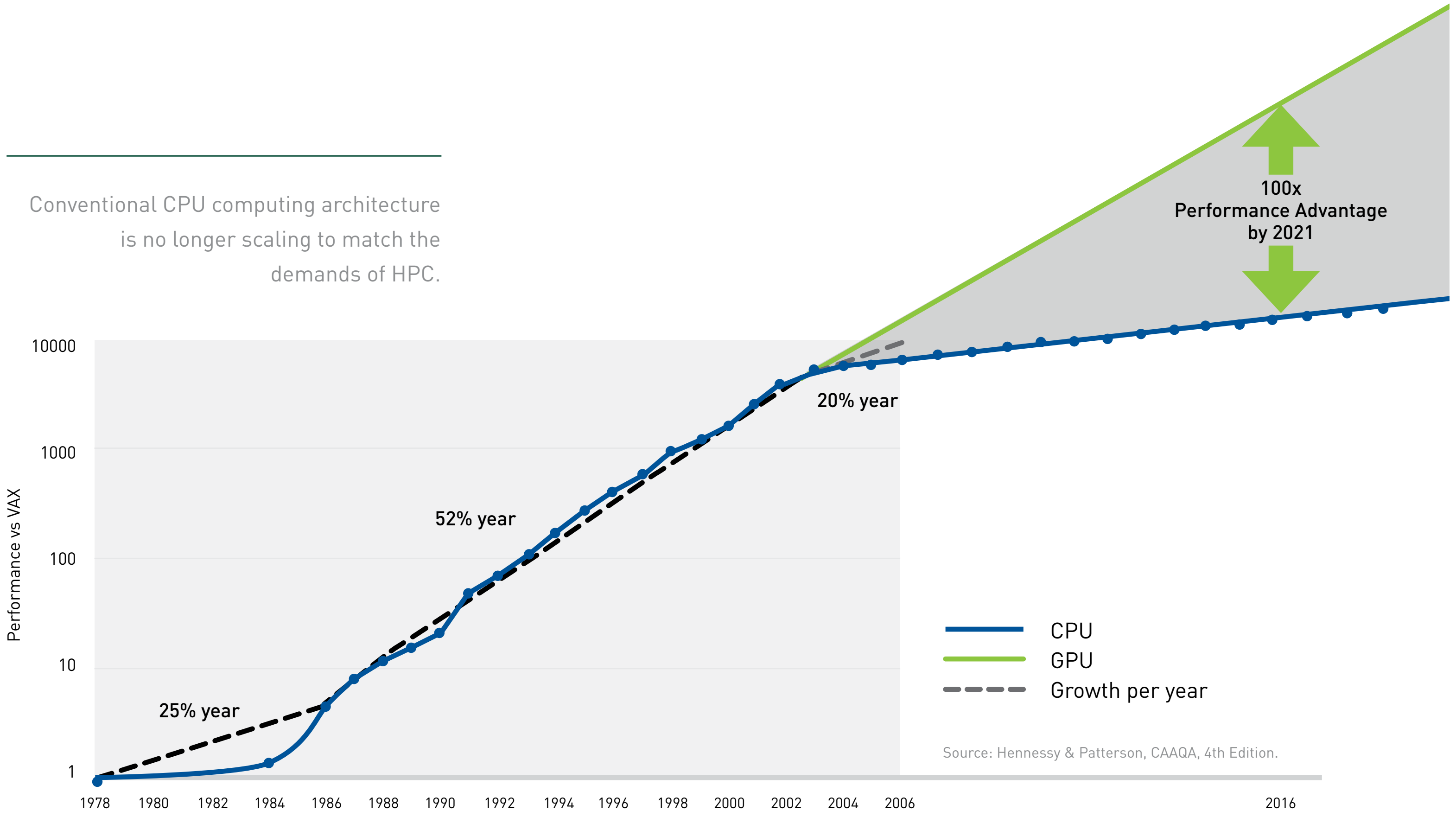
Nathan Brookwood
Principal Analyst & Co-Founder, Insight64

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

The use of GPUs for computation is a dramatic shift in HPC. GPUs deliver performance increases of 10x to 100x to solve problems in minutes instead of hours, outpacing the performance of traditional computing with x86-based CPUs alone. In addition, GPUs also deliver greater performance per watt of power consumed.

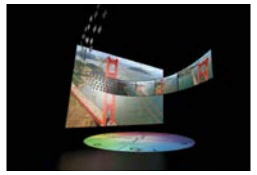
From climate modeling to medical tomography, NVIDIA® Tesla™ GPUs are enabling a wide variety of segments in science and industry to progress in ways that were previously impractical, or even impossible, due to technological limitations.

Conventional CPU computing architecture is no longer scaling to match the demands of HPC.



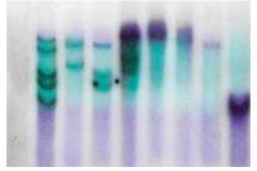
Source: Hennessy & Patterson, CAAQA, 4th Edition.

5X
Digital Content Creation
Adobe

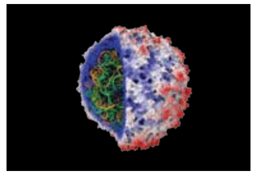


18X
Video Transcoding
Elemental Technologies

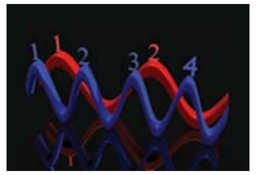
20X
3D Ultrasound
TechniScan



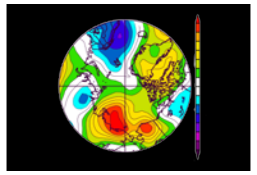
30X
Gene Sequencing
U of Maryland



36X
Molecular Dynamics
U of Illinois, Urbana-Champaign

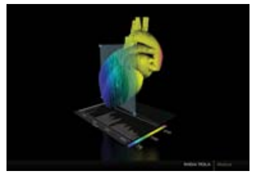


50X
MATLAB Computing
AccelerEyes

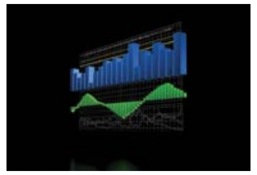


80X
Weather Modeling
Tokyo Institute of Technology

100X
Astrophysics
RIKEN



146X
Medical Imaging
U of Utah



149X
Financial Simulation
Oxford University

Tesla GPU computing is delivering transformative increases in performance for a wide range of HPC industry segments.

WHY GPU COMPUTING?

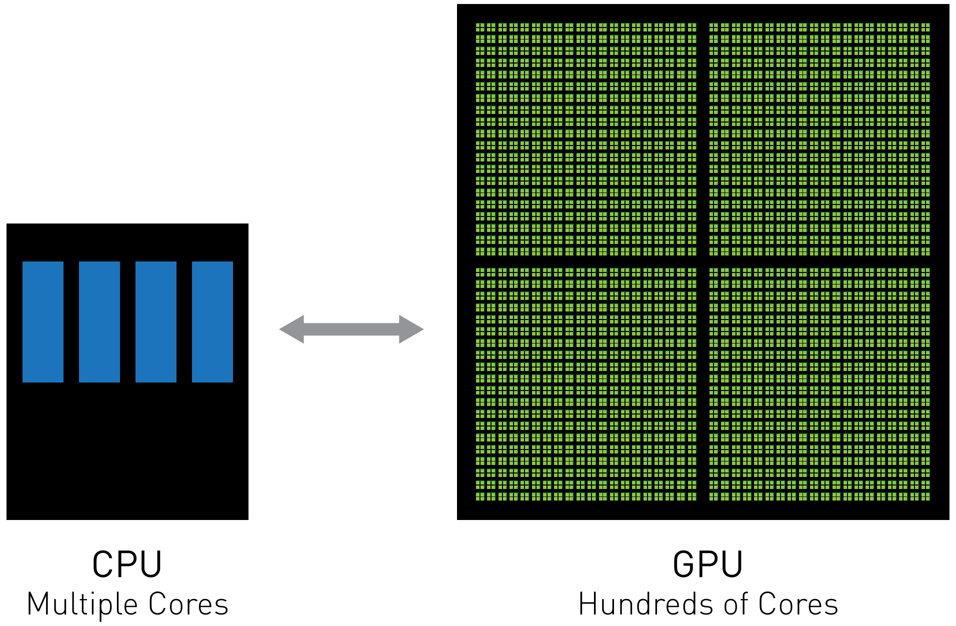
With the ever-increasing demand for more computing performance, the HPC industry is moving toward a hybrid computing model, where GPUs and CPUs work together to perform general purpose computing tasks.

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.

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's GPU solutions outpace others as they apply the most relevant processor to the specific task in hand.

NVIDIA TESLA
GPUS ARE REVOLUTIONIZING COMPUTING

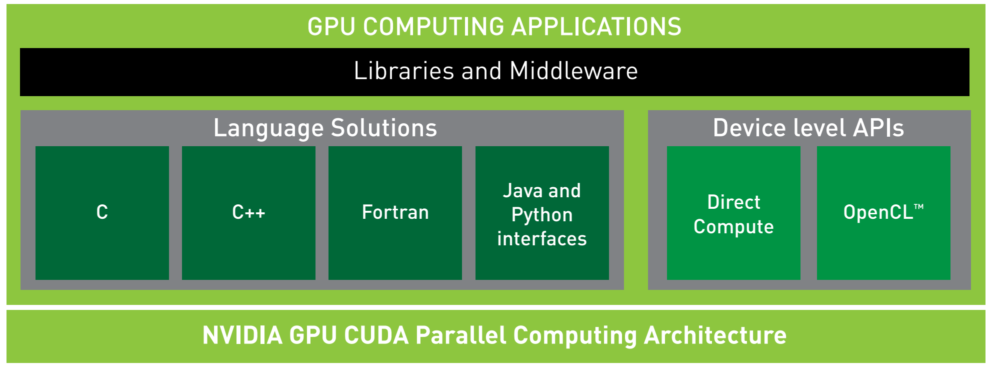
Core comparison between a CPU and a GPU.



PARALLEL ACCELERATION

Multi-core programming with x86 CPUs is difficult and often results in marginal performance gains when going from 1 core to 4 cores to 16 cores. Beyond 4 cores, memory bandwidth becomes the bottleneck to further performance increases.

To harness the parallel computing power of GPUs, programmers can simply modify the performance-critical portions of an application to take advantage of the hundreds of parallel cores in the GPU. The rest of the application remains the same, making the most efficient use of all cores in the system. Running a function on the GPU involves rewriting that function to expose its parallelism, then adding a few new function-calls to indicate which functions will run on the GPU or the CPU. With these modifications, the performance-critical portions of the application can now run significantly faster on the GPU.



GPU COMPUTING APPLICATIONS

Libraries and Middleware

Language Solutions: C, C++, Fortran, Java and Python interfaces

Device level APIs: Direct Compute, OpenCL™

NVIDIA GPU CUDA Parallel Computing Architecture

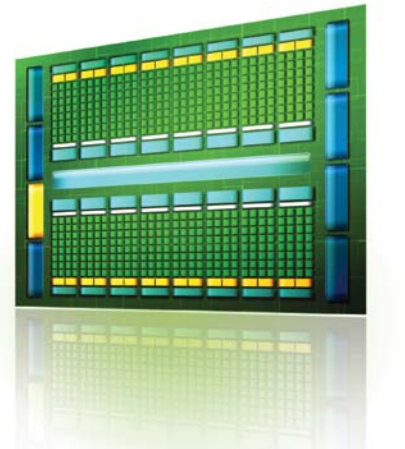
Developers use industry standard languages and tools to leverage the massively parallel CUDA architecture.

CUDA PARALLEL COMPUTING ARCHITECTURE

CUDA™ is NVIDIA's parallel computing architecture. 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.

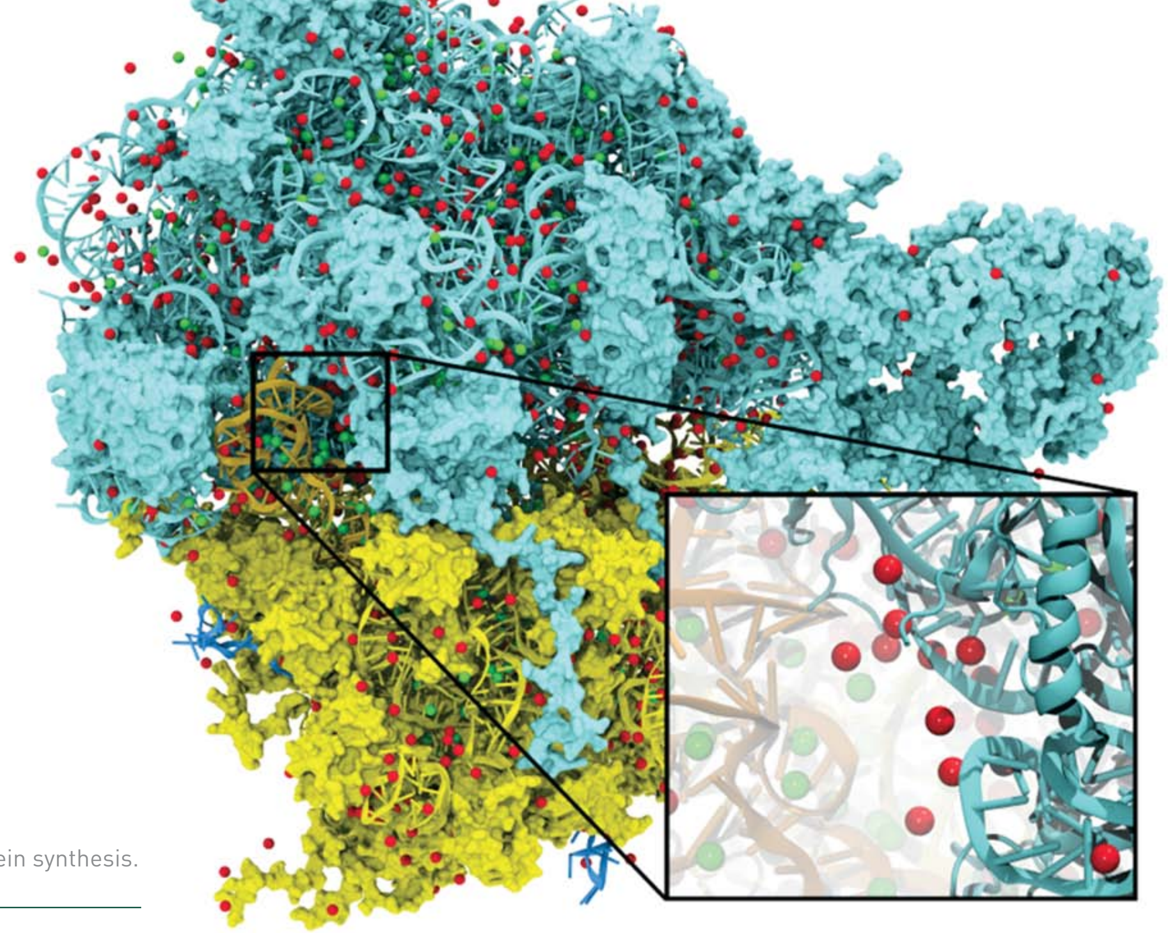
The latest generation CUDA architecture, codenamed "Fermi", is the most advanced GPU computing architecture ever built. With over three billion transistors, Fermi is making GPU and CPU co-processing pervasive by addressing the full-spectrum of computing applications. With support for C++, GPUs based 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 next generation CUDA parallel computing architecture, codenamed "Fermi".

> *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

NVIDIA TESLA
GPUS ARE REVOLUTIONIZING COMPUTING



Ribosome for protein synthesis.

University of Illinois: Accelerated molecular modeling enables rapid response to H1N1

WHO BENEFITS FROM GPU COMPUTING?

Computational scientists and researchers who are using GPUs to accelerate their applications are seeing results in days instead of months, even minutes instead of days.

CHALLENGE A first step in mitigating a global pandemic, like H1N1, requires quickly developing drugs to effectively treat a virus that is new and likely to evolve. This requires a compute-intensive process to determine how, in the case of H1N1, mutations of the flu virus protein could disrupt the binding pathway of the vaccine Tamiflu, rendering it potentially ineffective.

This determination involved a daunting simulation of a 35,000-atom system, something a group of University of Illinois, Urbana-Champaign scientists, led by John Stone, decided to tackle in a new way using GPUs.

Conducting this kind of simulation on a CPU would take more than a month to calculate...and that would only amount to a single simulation, not the multiple simulations that constitute a complete study.

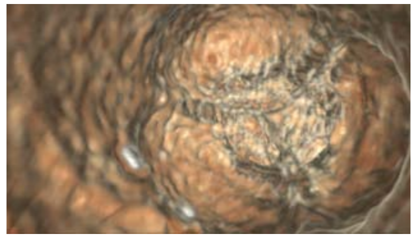
SOLUTION Stone and his team turned to the NVIDIA CUDA parallel processing architecture running on Tesla GPUs to perform their molecular modeling calculations and simulate the drug resistance of H1N1 mutations. Thanks to GPU technology, the scientists could efficiently run multiple simulations and achieve potentially life-saving results faster.

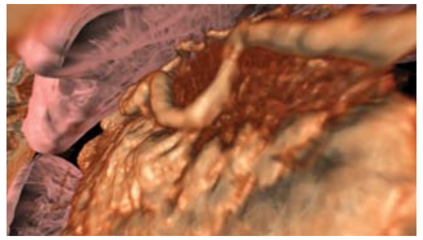
IMPACT The GPU-accelerated calculation was completed in just over an hour. The almost thousand-fold improvement in performance available through GPU computing and advanced algorithms empowered the scientists to perform "emergency computing" to study biological problems of extreme relevance and share their results with the medical research community.

This speed and performance increase not only enabled researchers to fulfill their original goal—testing Tamiflu's efficacy in treating H1N1 and its mutations—but it also bought them time to make other important discoveries. Further calculations showed that genetic mutations which render the swine or avian flu resistant to Tamiflu had actually disrupted the "binding funnel," providing new understanding about a fundamental mechanism behind drug resistance.

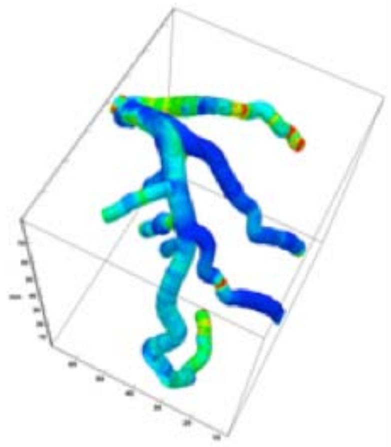
In the midst of the H1N1 pandemic, the use of improved algorithms based on CUDA and running on Tesla GPUs made it possible to produce actionable results about the efficacy of Tamiflu during a single afternoon. This would have taken weeks or months of computing to produce the same results using conventional approaches.

> “The benefits of GPU computing can be replicated in other research areas as well. All of this work is made speedier and more efficient thanks to GPU technology, which for us means quicker results as well as dollars and energy saved.
>
> John Stone
> Sr. Research Programmer
> University of Illinois at Urbana-Champaign





Harvard University: Finding Hidden Heart Problems Faster



320 detector-row CT has enabled single heart beat coronary imaging so that the entire coronary contrast opacification can be evaluated at a single time point. The full 3D course of the arteries, in turn, allows researchers to simulate the blood flowing through it by using computational fluid flow simulations, and subsequently compute the endothelial shear stress.

CHALLENGE Heart attacks, the leading cause of death worldwide, are caused when plaque, that has built up on artery walls, dislodges and blocks the flow of blood to the heart. Up to 80% of heart attacks are caused by plaque that is not detectable by conventional medical imaging. Even viewing the 20% that is detectable requires invasive endoscopic procedures, which involve running several feet of tubing into the patient in an effort to take pictures of arterial plaque.

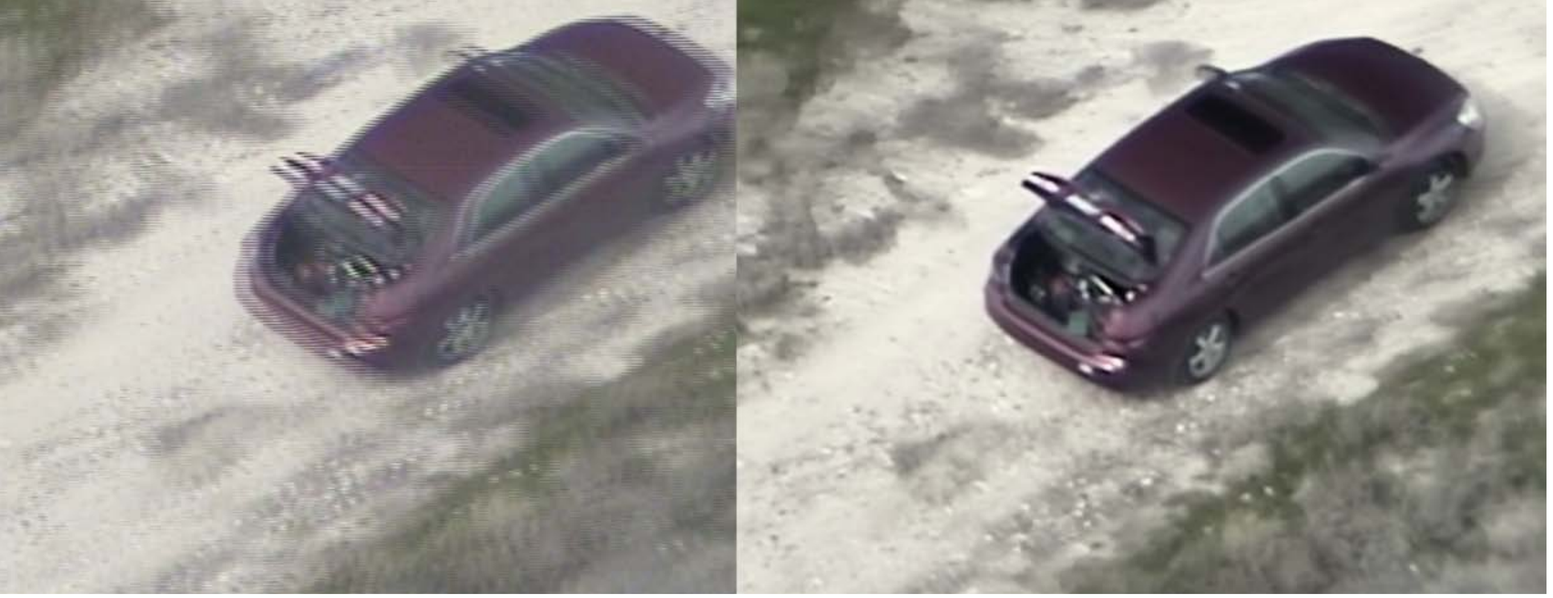
This level of uncertainty with regard to the exact location of potentially deadly plaque poses a significant challenge for cardiologists. Historically, it has been a guessing game for heart specialists to determine if and where to place arterial stents in patients with blockages. Knowing the location of the plaque could greatly improve patient care and save lives.

SOLUTION A team of researchers, including doctors at Harvard Medical School and Brigham & Women's Hospital in Boston, Massachusetts, have discovered a non-invasive way to find dangerous plaque in a patient's arteries. Tapping into the computational power of GPUs, they can create a highly individualized model of blood flow within a patient in a study called hemodynamics.

The buildup of plaque is highly correlated to the shape—or geometry—of a patient's arterial structure. Bends in an artery tend to be areas where dangerous plaque is especially concentrated.

Using imaging devices like a CT scan, scientists are able to create a model of a patient's circulatory system. From there, an advanced fluid dynamics simulation of the blood flow through the patient's arteries can be conducted on a computer to identify areas of reduced endothelial sheer stress on the arterial wall. A complex simulation like this one requires billions of fluid elements to be modeled as they pass through an artery system. An area of reduced sheer stress indicates that plaque has formed on the interior artery walls, preventing the bloodstream from making contact with the inner wall. The overall output of the simulation provides doctors with an atherosclerotic risk map. The map provides cardiologists with the location of hidden plaque and can serve as an indicator as to where stents may eventually need to be placed—and all of this knowledge is gained without invasive imaging techniques or exploratory surgery.

IMPACT GPUs provide 20x more computational power and an order of magnitude more performance per dollar to the application of image reconstruction and blood flow simulation, finally making such advanced simulation techniques practical at the clinical level. Without GPUs, the amount of computing equipment—in terms of size and expense—would render a hemodynamics approach unusable. Because it can detect dangerous arterial plaque earlier than any other method, it is expected that this breakthrough could save numerous lives when it is approved for deployment in hospitals and research centers.



MotionDSP: The increasing importance of the GPU in the Armed Forces

"Super-resolution" algorithms allow MotionDSP to reconstruct video with better and cleaner detail, increased resolution and reduced noise.

CHALLENGE Unmanned Aerial Vehicles (UAVs) represent the latest in high-tech weaponry deployed to strengthen and improve the military's capabilities. But with new technologies come new challenges, such as capturing actionable intelligence while flying at speeds upwards of 140 mph, 10 miles above the earth.

One key feature of the UAV is that it is capable of providing a real-time stream of detailed images taken with multiple cameras on the vehicle simultaneously. The challenge is that the resulting images need to be rendered, stabilized and enhanced in real-time and across vast distances in order to be useful.

Once they have been processed, the images can give infantry critical information about potential challenges ahead—the end goal being to ensure the safety and protection of military personnel in the field.

Using CPUs alone, this process is very time consuming and does not allow information to be viewed in real-time. As a result, military action could be based on potentially outdated intelligence data and inaccurate guides.

SOLUTION MotionDSP, a software company based in San Mateo, California, has developed "super-resolution" algorithms that allow it to reconstruct video with better and cleaner detail, increased resolution and reduced noise. All of which are ideal for the live streaming of video from the cameras attached to a UAV.

MotionDSP's product, Ikena ISR, leverages NVIDIA's CUDA parallel computing architecture allowing it to render, stabilize and enhance live video faster and more accurately than its competitors. Ikena ISR features computationally intense, advanced motion-tracking algorithms that provide the basis for sophisticated image stabilization and super-resolution video reconstruction. Perhaps most importantly, it can all be run on off-the-shelf Windows laptops and servers.

Using NVIDIA Tesla GPUs, MotionDSP's customers, which include a variety of military-funded research groups, are making UAVs safer and more reliable while reducing deployment costs, improving simulation accuracy and dramatically boosting performance.

IMPACT Using only CPUs to execute the kind of sophisticated video post-processing algorithms required for effective reconnaissance would result in up to six hours of processing for each one hour of video—not a viable solution when real-time results are critical. In contrast, Tesla GPUs enable MotionDSP's Ikena ISR software to process any live video source in real-time with less than 200ms of latency. Moreover, instead of requiring expensive CPU-clustered computing systems to complete the work, Ikena can perform at full capacity on a standard workstation small enough to fit inside military vehicles.

> "With the GPU, we're bringing higher quality video to all ISR platforms, including smaller UAVs, by being smarter about how we utilize COTS PC technology. Our technology makes the impossible possible, and this is making our military safer and better prepared."
>
> Sean Varah
> Chief Executive Officer, MotionDSP

Merlin International is one of the fastest growing providers of information technology solutions in the United States; their Collaborative Video Delivery offering – which includes Ikena – helps support the defense and intelligence missions of the US Federal Government.

"MotionDSP's use of GPU technology has greatly enhanced the capabilities of its Ikena software, enabling it to deliver real-time super-resolution analysis of intelligence video — something that simply was not possible before," said John Trauth, President of Merlin International. "Integrating this technology into our Collaborative Video Delivery solution can enable our government customers to quickly and easily access the data they need for effective intelligence, surveillance and reconnaissance (ISR) – this saves lives and significantly increases mission success rates."

Bloomberg: GPUs increase accuracy and reduce processing time for bond pricing

CHALLENGE Getting a mortgage and buying a home is a complex financial transaction, and for lenders, the competitive pricing and management of that mortgage is an even greater challenge. Transactions involving thousands of mortgages at once are a routine occurrence in financial markets, spurred by banks that wish to sell off loans to get their money back sooner.

Known as collateralized debt obligations (CDO) and collateralized mortgage obligations (CMO), baskets of thousands of loans are publicly traded financial instruments. For the banks and institutional investors who buy and sell these baskets, timely pricing updates are essential because of fast-changing market conditions.

Bloomberg, one of the world's leading financial services organizations, prices CDO/CMO baskets for its customers by running powerful algorithms that model the risks and determine the price.

Financial engineering is integral to today's buying and selling decisions.

> *One of the challenges Bloomberg always faces is that we have very large scale. We're serving all the financial and business community and there are a lot of different instruments and models people want calculated."*
>
> Shawn Edwards
> CTO, Bloomberg

This technique requires calculating huge amounts of data, from interest rate volatility to the payment behavior of individual borrowers. These data-intensive calculations can take hours to run with a CPU-based computing grid. Time is money, and Bloomberg wanted a new solution that would allow them to get pricing updates to their customers faster.

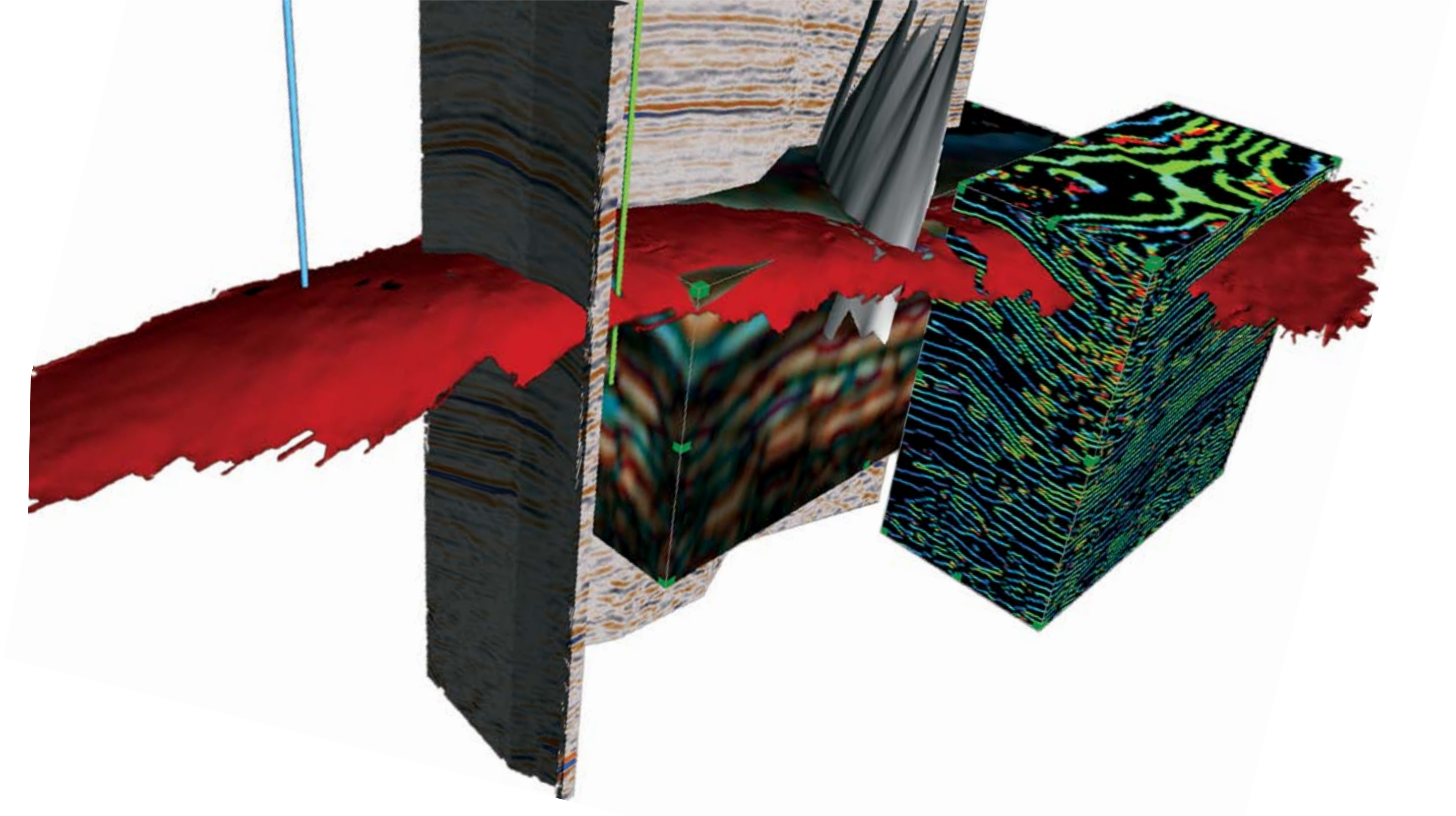
SOLUTION Bloomberg implemented an NVIDIA Tesla GPU computing solution in their datacenter. By porting their application to run on the NVIDIA CUDA parallel processing architecture to harness the power of GPUs, Bloomberg received dramatic improvements across the board. Large calculations that had previously taken up to two hours can now be completed in two minutes. Smaller runs that had taken 20 minutes can now be performed in just seconds.

In addition, the capital outlay for the new GPU-based solution was one-tenth the cost of an upgraded CPU solution, and further savings are being realized due to the GPU's efficient power and cooling needs.

IMPACT As Bloomberg customers make CDO/CMO 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.

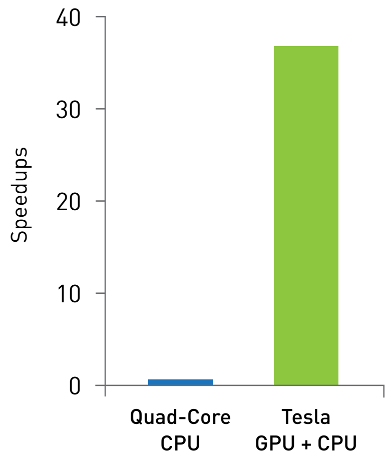
GPU ACCELERATION

Large calculations that had previously taken up to two hours can now be completed in two minutes. Smaller runs that had taken 20 minutes can now be performed in just seconds.



> CUDA-based GPUs power large computation tasks and interactive computational workflows that we could not hope to implement effectively otherwise."
>
> Steve Purves
> ffA Technical Director

ffA: Accelerating 3D seismic interpretation



The latest benchmark results using Tesla GPUs have produced performance improvements of up to 37x versus high-end workstations with two quad core CPUs.

CHALLENGE In the search for oil and gas, the geological information provided by seismic images of the earth is vital. By interpreting the data produced by seismic imaging surveys, geoscientists can identify the likely presence of hydrocarbon reserves and understand how to extract resources most effectively. Today, sophisticated visualization systems and computational analysis tools are used to streamline what was previously a subjective and labor intensive process.

Today, geoscientists must process increasing amounts of data as dwindling reserves require them to pinpoint smaller, more complex reservoirs with greater speed and accuracy.

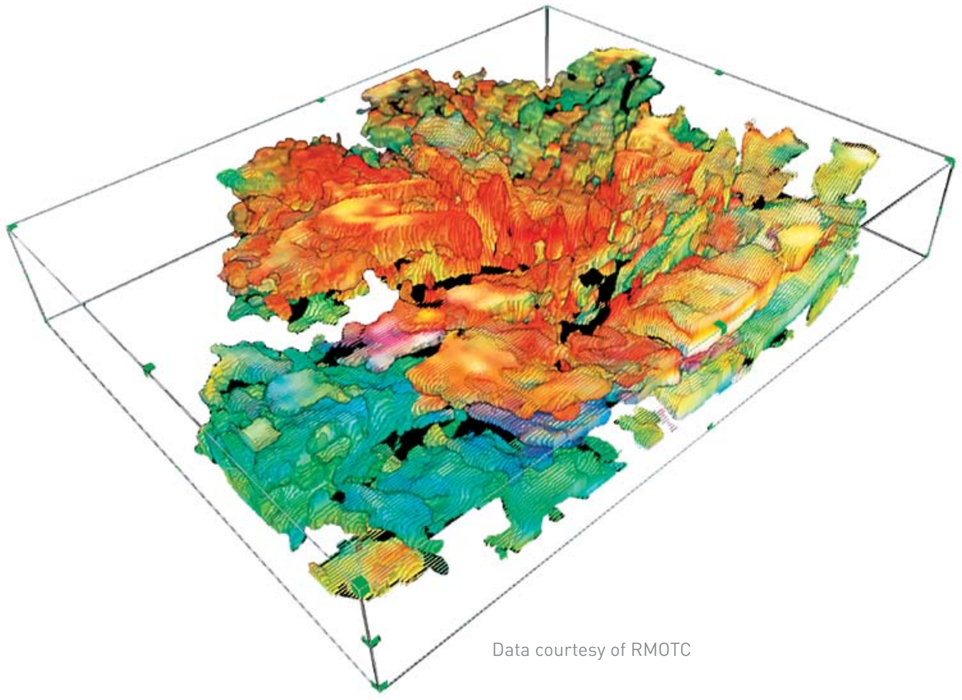
SOLUTION UK-based company ffA provides world leading 3D seismic analysis software and services to the global oil and gas industry. Its software tools extract detailed information from 3D seismic data, providing a greater understanding of complex 3D geology, improving productivity and reducing uncertainty within the interpretation process. The sophisticated tools are compute-intensive so it can take hours, or even days, to produce results on conventional high performance workstations.

With the recent release of its 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.

This step change in performance significantly increases the amount of data that geoscientists can analyze in a given timeframe. Plus, it allows them to fully exploit the information derived from 3D seismic surveys to improve subsurface understanding and reduce risk in oil and gas exploration and exploitation.

IMPACT NVIDIA CUDA is allowing ffA to provide scalable high performance computation for seismic data on hardware platforms equipped with one or more NVIDIA Tesla and Quadro GPUs. The latest benchmark results using Tesla GPUs have produced performance improvements of up to 37x versus high-end workstations with two quad core CPUs.

"Access to high performance, high quality 3D computational tools on a workstation platform drastically improves the productivity curve in 3D seismic analysis and seismic interpretation, giving our users a real edge in oil and gas exploration and de-risking field development."

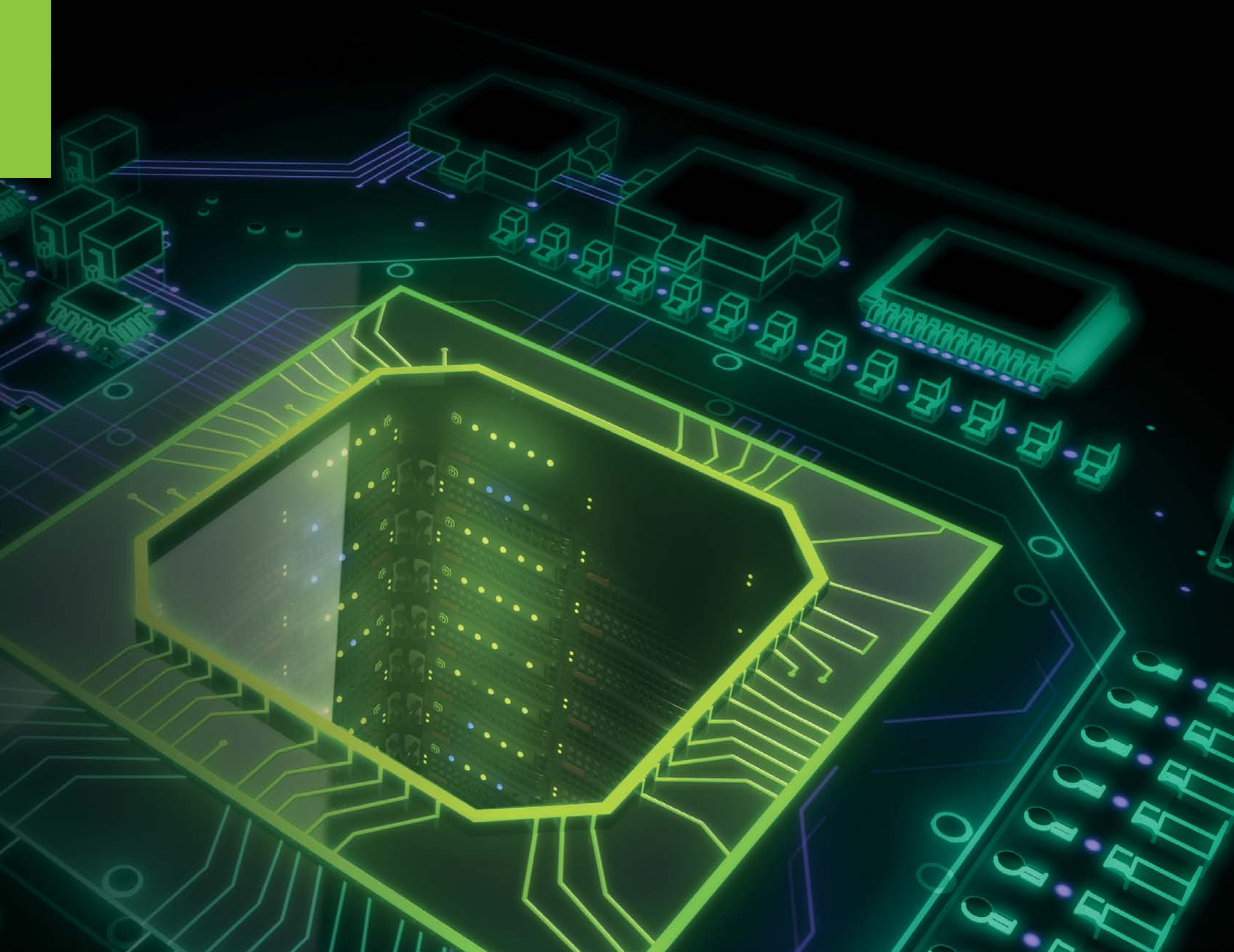


Data courtesy of RMOTC

THE SHIFT TO PARALLEL COMPUTING AND THE PATH TO EXASCALE

Approximately every 10 years, the world of supercomputing experiences a fundamental shift in computing architectures. It was around 10 years ago when cluster-based computing largely superceded vector-based computing as the de facto standard for large-scale computing installations, and this shift saw the supercomputing industry move beyond the petaFLOP performance barrier. With the next performance target being exascale-class computing, it is time for a new shift in computing architectures—the move to parallel computing.

The shift is already underway.



> “Future computing architectures will be hybrid systems with parallel-core GPUs working in tandem with multi-core CPUs.”
>
> Jack Dongarra
> Distinguished Professor at University of Tennessee

TSUBAME 1.2, by Tokyo Institute of Technology, was the first Tesla GPU based hybrid cluster to appear on the Top500 list.

In November 2008, Tokyo Institute of Technology became the first supercomputing center to enter the Top500 with a GPU-based hybrid system—a system that uses GPUs and CPUs together to deliver transformative increases in performance without breaking the bank with regards to energy consumption. The system, called TSUBAME 1.2, entered the list at number 24.

Fast forward to June 2010 and hybrid systems have started to make appearances even higher up the list. "Nebulae", a system installed at the Shenzhen Supercomputing Center in China, equipped with 4640 Tesla 20-series GPUs, made its entry into the list at number 2, just one spot behind Oak Ridge National Lab's "Jaguar", the fastest supercomputer in the world.

What is even more impressive than the overall performance of Nebulae, is how little power it consumes. While Jaguar delivers 1.77 petaFLOPs, it consumes more than 7 megawatts of power to do so.

To put that into context, 7 megawatts is enough energy to power 15,000 homes. In contrast, Nebulae delivers 1.27 petaFLOPs, yet it does this within a power budget of just 2.55 megawatts. That makes it twice as power-efficient as Jaguar. This difference in computational throughput and power is owed to the massively parallel architecture of GPUs, where hundreds of cores work together within a single processor, delivering unprecedented compute density for next generation supercomputers.

Another very notable entry into this year's Top500 was the Chinese Academy of Sciences (CAS). The "Mole 8.5" supercomputer at CAS uses 2200 Tesla 20-series GPUs to deliver 207 teraFLOPS, which puts it at number 19 in the Top500.

CAS is one of the world's most dynamic research and educational facilities. Prof. Wei Ge, Professor of Chemical Engineering at the Institute of Process Engineering at CAS, introduced GPU computing to the Beijing facility in 2007 to help them with discrete particle and molecular dynamics simulations. Since then, parallel computing has enabled the advancement of research in dozens of other areas, including: real-time simulations of industrial facilities, the design and optimization of multi-phase and turbulent industrial reactors using computational fluid dynamics, the optimization of secondary and tertiary oil recovery using multi-scale simulation of porous materials, the simulation of nano- and micro-flow in chemical and bio-chemical processes, and much more.

These computational problems represent a tiny fraction of the entire landscape of computational challenges that we face today, and these problems are not getting any smaller. The sheer quantity of data that many scientists, engineers and researchers must analyze is increasing exponentially and supercomputing centers are over-subscribed as demand is outpacing the supply of computational resources. If we are to maintain our rate of innovation and discovery, we must take computational performance to a level where it is 1000 times faster than what it is today. The GPU is a transformative force in supercomputing and represents the only viable strategy to successfully build exascale systems that are affordable to build and efficient to operate.

"Five years from now, the bulk of serious HPC is going to be done with some kind of accelerated heterogeneous architecture." said Steve Scott, CTO of Cray Inc.

"Nebulae", powered by 4640 Tesla 20-series GPUs, is one of the fastest supercomputers in the world.

NVIDIA TESLA
WORLD'S FIRST COMPUTATIONAL GPU

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 the double precision performance of CPUs. Compared to 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.

TESLA GPU COMPUTING SOLUTIONS

NVIDIA Tesla products are designed for high-performance computing, and offers exclusive computing features.

Superior Performance

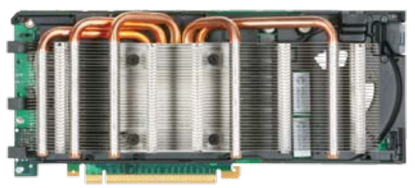
- Highest double precision floating point performance
- Large HPC data sets supported by larger on-board memory
- Faster communication with InfiniBand using NVIDIA GPUDirect™

Highly Reliable

- ECC protection for uncompromised data reliability
- Stress tested for zero error tolerance
- Manufactured by NVIDIA
- Enterprise-level support that includes a three-year warranty

Designed for HPC

- Integrated by leading OEMs into workstations, servers and blades



Tesla M2050/M2070 GPU Computing Module enables the use of GPUs and CPUs together in an individual server node or blade form factor.

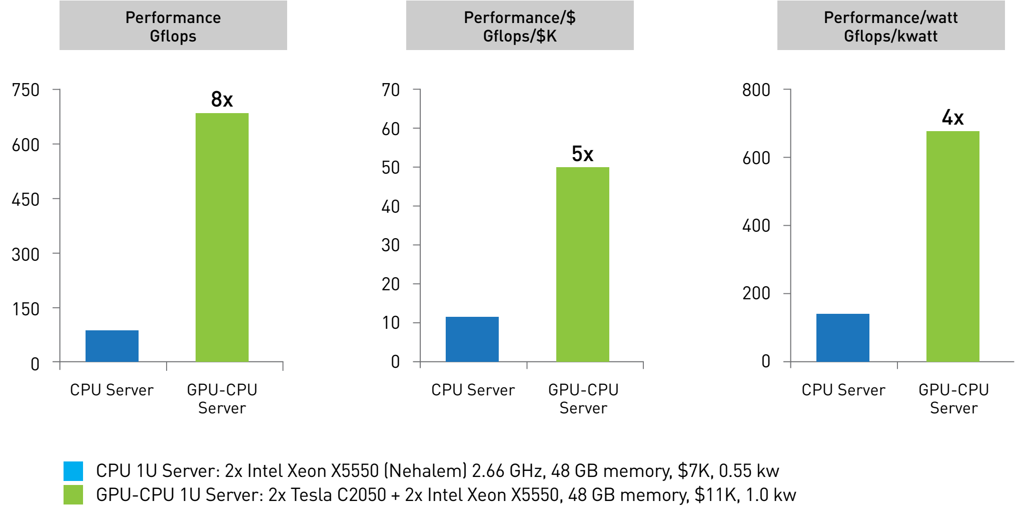
Tesla S2050 GPU Computing System is a 1U system powered by 4 Tesla GPUs and connects to a CPU server.

Tesla C2050/C2070 GPU Computing Processor delivers the power of a cluster in the form factor of a workstation.

TESLA DATA CENTER PRODUCTS

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

Highest Performance, Highest Efficiency

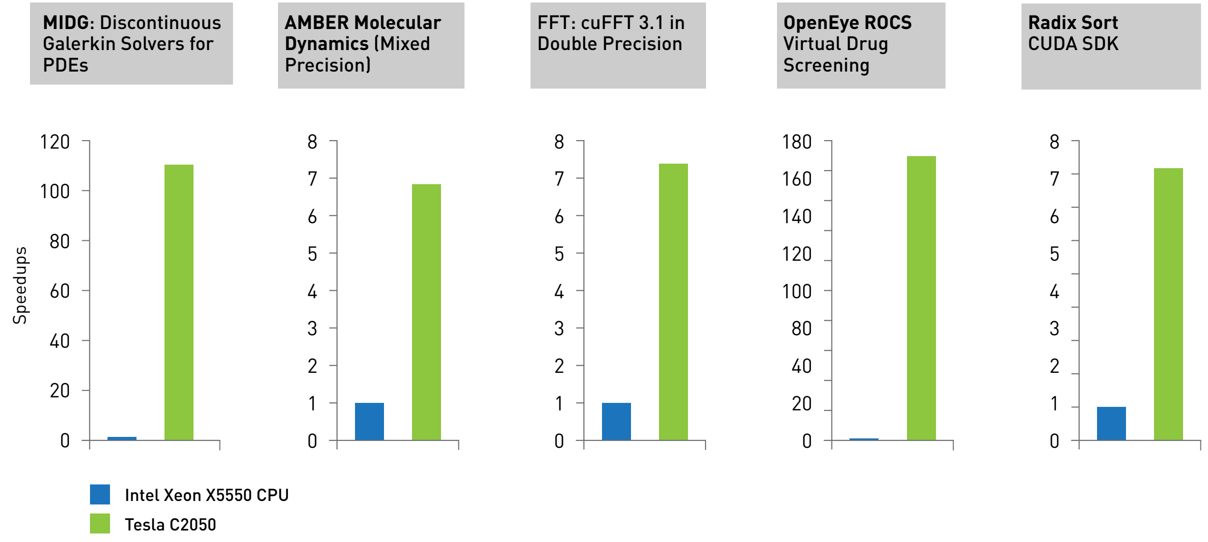


GPU-CPU server solutions deliver up to 8x higher Linpack performance.

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.

Highest Performance, Highest Efficiency



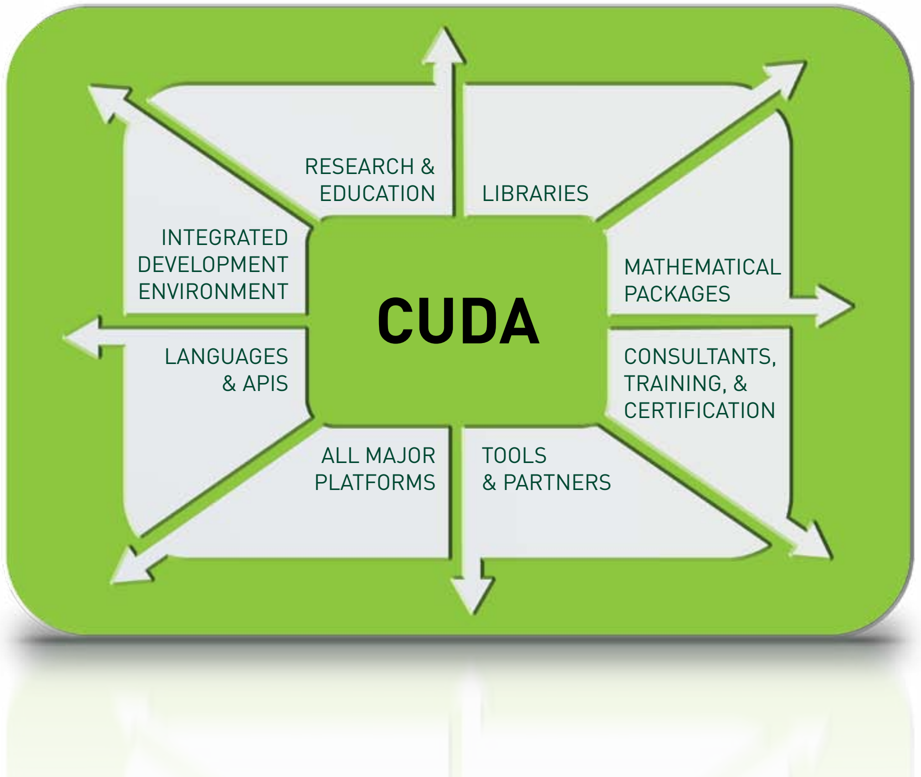
Workstations powered by Tesla GPUs outperform conventional CPU-only solutions in life science applications.

DEVELOPER ECOSYSTEM AND WORLDWIDE EDUCATION

In just a few years, an entire software ecosystem has developed around the CUDA architecture—from more than 350 universities worldwide teaching the CUDA programming model, to a wide range of libraries, compilers and middleware that help users optimize applications for GPUs.

A rich ecosystem of software applications, libraries, programming language solutions, and service providers support the CUDA parallel computing architecture.

The NVIDIA GPU Computing Ecosystem



NVIDIA's CUDA architecture has the industry's most robust language and API support for GPU computing developers, including C, C++, OpenCL, DirectCompute, and Fortran. NVIDIA Parallel Nsight™, a fully integrated development environment for Microsoft Visual Studio is also available. Used by more than six million developers worldwide, Visual Studio is one of the world's most popular development environments for Windows-based applications and services. Adding functionality specifically for GPU computing developers, Parallel Nsight makes the power of the GPU more accessible than ever before.

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

In addition to the CUDA C development tools, math libraries, and hundreds of code samples in the NVIDIA GPU computing SDK, there is also a rich ecosystem of solutions:

Libraries and Middleware Solutions

- Acceleware FDTD libraries
- CUBLAS, complete BLAS library*
- CUFFT, high-performance FFT routines*
- CUSP
- EM Photonics CULA Tools, heterogeneous LAPACK implementation
- NVIDIA OptiX and other AXEs*
- NVIDIA Performance Primitives for image and video processing **www.nvidia.com/npp***
- Thrust

Compilers and Language Solutions

- CAPS HMPP
- NVIDIA CUDA C Compiler (NVCC), supporting both CUDA C and CUDA C++*
- Par4All
- PGI CUDA Fortran
- PGI Accelerator Compilers for C and Fortran
- PyCUDA

GPU Debugging Tools

- Allinea DDT
- Fixstars Eclipse plug-in
- NVIDIA cuda-gdb*
- NVIDIA Parallel Nsight for Visual Studio
- TotalView Debugger

GPU Performance Analysis Tools

- NVIDIA Visual Profiler*
- NVIDIA Parallel Nsight for Visual Studio
- TAU CUDA
- Vampir

Cluster and Grid Management Solutions

- Bright Cluster Manager
- NVIDIA system management interface (nvidia-smi)
- Platform Computing

Math Packages

- Jacket by AccelerEyes
- Mathematica 8 by Wolfram
- MATLAB Distributed Computing Server (MDCS) by Mathworks
- MATLAB Parallel Computing Toolbox (PCT) by Mathworks

Consulting and Training

Consulting and training services are available to support you in porting applications and learning about developing with CUDA.

For more information, visit **www.nvidia.com/object/cuda_consultants.html.**

Education and Certification

- CUDA Certification www.nvidia.com/certification
- CUDA Center of Excellence research.nvidia.com
- CUDA Research Centers research.nvidia.com/
- CUDA Teaching Centers research.nvidia.com/
- CUDA and GPU computing books

* Available with the latest CUDA toolkit at **www.nvidia.com/getcuda**

For more information about the CUDA Certification Program, visit **www.nvidia.com/certification**.