



QUADRO BASED ON THE KEPLER ARCHITECTURE

WP-06845-001_v01 | May 2013

White Paper - Alex Herrera



DOCUMENT CHANGE HISTORY

WP-06845-001_v01

Version	Date	Authors	Description of Change
01	May 14, 2013	AH, KH, SM	Initial Release

TABLE OF CONTENTS

About the Author	1
Workstations and Professional Graphics Still the Premier Tools for Professional Computing.....	2
Professional Graphics: The Cornerstone of Today's Workstation	4
Proven Performance, Optimized for Professional Applications	4
Workstation Graphics Taking on More of the Professional Workload	6
Setting the Workstation Platform Apart: Uptime Above all.....	6
Easy to Deploy and Manage.....	7
NVIDIA Quadro: Setting the Bar in Professional Graphics	8
Pushing Display Technology Beyond the Status Quo	8
Not Just for Graphics Anymore: CUDA lets You do More with Quadro	9
Quality and Reliability Go Hand-in-Hand	10
Easy to Deploy and Maintain	10
What the Kepler Architecture Brings to Quadro.....	12
Productivity: The Only Measure that Matters in Professional Computing.....	13

ABOUT THE AUTHOR

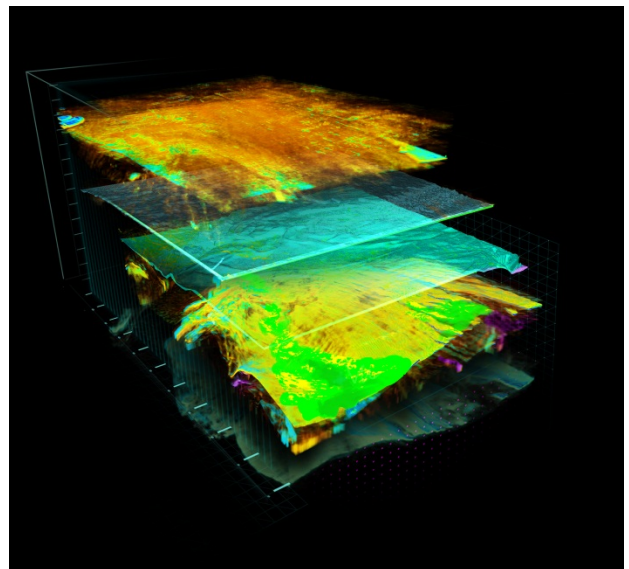
With more than 30 years of engineering, marketing and management experience in the semiconductor industry, Alex Herrera is now a consultant focusing on the technology and markets of high-performance graphics and workstations. Author of frequent articles covering both the business and technology of graphics, Alex is also responsible for the *JPR Workstation Report* series, published by Jon Peddie Research. He continues to advise companies competing in fields related to high-performance graphics and client-centric computing.

WORKSTATIONS AND PROFESSIONAL GRAPHICS STILL THE PREMIER TOOLS FOR PROFESSIONAL COMPUTING

The modern workstation bears little resemblance to the original systems that started powering businesses 30 years ago. Though the names on the outside have changed and the technology inside has advanced by leaps and bounds, in one important respect, the workstation hasn't really changed at all. It remains the one indispensable tool for high-demand computing professionals, the only machine purpose-built to tackle workflows in media and entertainment, manufacturing, geoscience, financial analysis, medicine and scientific research.

Today's workstation shares a common technology core with PCs, laptops, servers and supercomputers. And that's a good thing. Workstation builders now benefit from access to both best-of-breed silicon technology and economy of scale, a win-win combination that has delivered dramatically higher capability at far lower price points. Still, though it may share genes with its sister platforms, the workstation's DNA remains unique, reflecting professionals' particular computing priorities, both in the office and on the road: reliability, manageability, and application-specific performance.

Where consumers and corporate users may find last year's technology "good enough," professionals thirst for compute has proven insatiable, consuming every drop of incremental performance each new generation of technology and product



can deliver. In media and entertainment circles, last year's effects-filled blockbuster raises the bar on visuals for this year's feature, increasing the required computation per frame. In manufacturing, ever-pressured project schedules demand fewer and shorter iterations for design, modeling, engineering analysis, and styling. The explosion of volumetric data in medical and geoscience drives demand up exponentially. And every answer researchers and scientists find to today's questions leads to new, ever-more-complex questions to tackle tomorrow.

Answering that call, workstations continue to evolve, with each generation offering task-shrinking throughput improvements: more FLOPS, more bandwidth, and faster graphics. But workstation innovation has always gone well beyond simple, incremental improvements on the old model. Workstation-seeded technologies — like multi-graphics processing unit (GPU), multi-display, the Digital Video Pipeline, and GPU Compute — have transformed the way professionals arrange their workflows, not only shortening times for each task, but parallelizing tasks to yield dramatic increases in productivity.

PROFESSIONAL GRAPHICS: THE CORNERSTONE OF TODAY'S WORKSTATION

What ingredients go into a workstation today?

- ▶ One or two fast and reliable CPUs
- ▶ Large amounts of high-speed, error-correcting memory
- ▶ The most reliable, high-capacity storage, built on both traditional hard-disk and emerging solid-state technologies.

But more than any other, the component that is setting the workstation apart from all other platforms is the professional GPU. Workstation-caliber GPUs excel in precisely those attributes that workstation users care about most: reliability, manageability, and specialized performance for the applications that matter.

PROVEN PERFORMANCE, OPTIMIZED FOR PROFESSIONAL APPLICATIONS

When it comes to computer graphics, one size doesn't fit all. The fact is, professional use places special demands on graphics hardware. Demands that only workstation GPUs are engineered to meet.

Specifically engineered for professionals, NVIDIA Quadro® graphics render fast, reliably, and with the particular look applications like AutoCAD, 3Ds max, Catia and Solidworks expect. Hardware pipelines implement specialized graphics engines

optimized for shaders and rasterizers used in manufacturing and construction, and not games. Application-tuned drivers¹ contain carefully optimized code paths that expedite processing for the most common drawing operations. And for those professionals with the most demanding video and display requirements, NVIDIA® Quadro® graphics feature capabilities like SDI, genlock, frame lock and Stereoscopic 3D.

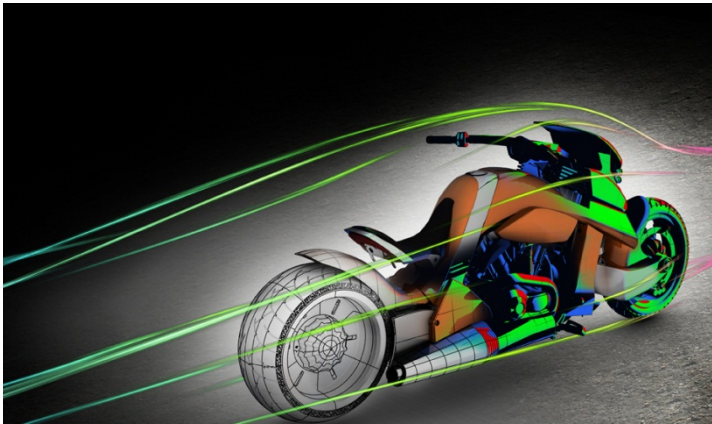
Table 1. Productivity and Reliability

	Quadro	Gaming	Integrated
Productivity			
Basic support for Business Applications	●	●	●
Tested, tuned and certified over 200 Professional Applications	●		
Maximus: Visualization + render/simulate on a single Workstation	●		
Higher bit depth/ color support (30-bit and VESA® DisplayPort™ (output) connectivity	●		
Productivity tools for Display Management (nView, Mosaic)	●		
Digital video pipeline for Professional Video Development	●		
Reliability			
Certified and available on Workstation Platforms	●		●
Drivers optimized for stability and Business Development	●		
Enterprise class and Applied Engineering support	●		
Built and tested by NVIDIA to single specification	●		
Extended availability + Business Class 3 year Warranty	●		

¹ For example, AutoCAD 3Ds max, and Solidworks

WORKSTATION GRAPHICS TAKING ON MORE OF THE PROFESSIONAL WORKLOAD

GPUs aren't just for graphics any more. The end result of a more-than-decade long pursuit to harness the power of GPUs for non-graphics tasks has arrived; GPU Compute Technology. The most powerful supercomputer on the planet now relies on NVIDIA



GPUs to achieve a staggering 27 Petaflops of throughput². But more than any other platform, it's the workstation that stands at this crossroads of rendering and GPU computation.

The convergence in technology makes for a compelling, cost-effective proposition: the same GPU

that renders models can now analyze and simulate them as well. Think core mechanical design tasks like finite element analysis (FEA) and computational fluid dynamics (CFD); volumetric data analysis and visualization for medical and geoscience applications; Monte Carlo simulations for financial asset valuation; and raytraced rendering, craved by product stylists and studios from Bollywood to Madison Avenue.

SETTING THE WORKSTATION PLATFORM APART: UPTIME ABOVE ALL

While some users might be able to sacrifice a certain degree of reliability, Workstation Professionals have no such luxury. When it comes to running mission-critical software, 24/7 availability is not negotiable. The reason's clear for a municipal command operation, for example, where public safety is on the line. If a hardware failure shuts down a municipal transit control center at rush hour, the best-case scenario would be late trains and snarled traffic.

But availability matters in all professional computing circles, and reliability has always been a hallmark of workstation design. It needs to be, because one untimely hardware failure can mean the difference between a bottom line that's inked in black or one in red.

² Oak Ridge National Laboratory's Titan, as ranked in www.top500.org (November, 2012)

Should the real-time digital video pipeline in a broadcast studio go down, there is no time for a retake. The damage is done. A hardware glitch knocking an oil rig off line can mean hundreds of thousands of dollars hacked off the bottom line for every hour it sits idle.

While hardware failures are costly enough, the impact of downtime can pale in comparison to the impact of undetected computational errors. The potential consequences of poor data integrity in medical imaging applications go without saying. And a single bit error propagating its way through a real-time pricing model could devastate a financial institution or trading exchange.

EASY TO DEPLOY AND MANAGE

What ultimately matters to a business isn't the cost to acquire an asset, it's the total cost incurred over that asset's lifetime. The total cost of ownership (TCO) includes not only the asset's original invoice but all man-hours spent deploying and maintaining it post-purchase.

With an estimated "80% of unplanned outages due to ill-planned changes made by administrators³, the expenses due to both excessive maintenance and unavailable systems can quickly inflate an IT asset's TCO to a sobering five times the initial purchase price. [1]

Clearly, minimizing expenses over the long run demands more than robust hardware alone. Easing deployment, reducing administrative overhead, and monitoring status, are goals that require a comprehensive set of management tools. Tools only available with workstation-caliber GPUs.

³ The IT Process Institute's Visible Ops Handbook

[1] Gartner, Inc.

NVIDIA QUADRO: SETTING THE BAR IN PROFESSIONAL GRAPHICS

From its inception, NVIDIA's Quadro line of Professional Graphics products has not only delivered more of the visual performance and features users have always demanded from their workstations, it's introduced technologies that have re-shaped the way Workstation Professionals get their jobs done.

PUSHING DISPLAY TECHNOLOGY BEYOND THE STATUS QUO

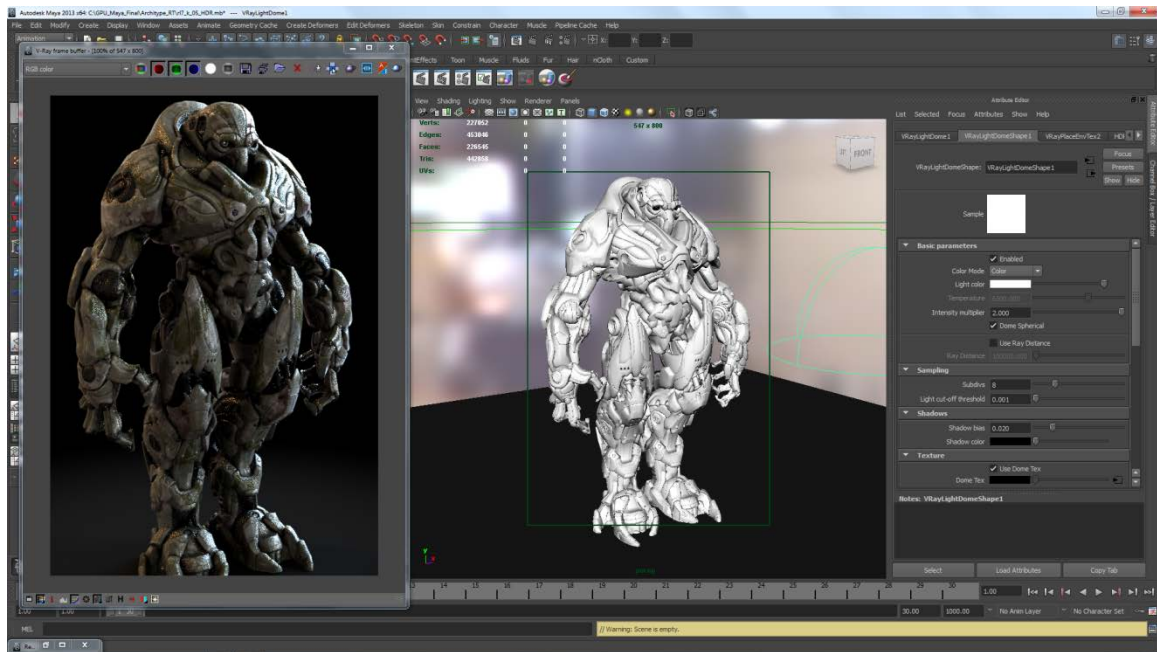
Few technology advancements can match the productivity boost of multi-display technology. Now driving four independent displays from a single card, Quadro lets professionals manage more tasks at a time. NVIDIA's nView® Desktop Management software allows customized, multi-display environments tailored to the most particular professional's preferences and workflows.

Increasingly, the benefits of more expansive screen space are extending beyond the desktop into large-scale informational and collaborative spaces, like immersive virtual environments, commercial and government command/control centers, and digital signage. Integrating features like bezel compensation, tear-free synchronization and Stereoscopic 3D, Quadro's comprehensive NVIDIA® Mosaic technology makes it all possible, supporting environments from a few desktop monitors to an array of 4K projectors, overlap-calibrated to create a seamless, ultra-resolution display.

NOT JUST FOR GRAPHICS ANYMORE: CUDA LETS YOU DO MORE WITH QUADRO

With the introduction of NVIDIA's CUDA® technology, NVIDIA unleashed the power of programmable GPUs for non-graphics tasks, creating a new and viable computing paradigm. One that's since infiltrated platforms from workstations to supercomputers alike. A comprehensive software development platform for GPU Compute, CUDA turns every Quadro into an equal-opportunity processor, accelerating multiple facets of the professional's workflow, not just rendering.

Leading ISVs⁴, representing every corner of professional computing, natively support CUDA in their tools, to dramatically shorten execution times for a range of compute-intensive tasks in engineering, science, animation, digital video and finance. And now NVIDIA's Maximus® technology extends that support into an efficient multi-GPU architecture, bringing true concurrent, parallel processing of the most common stages of professional workflows, for example, design and rendering.



⁴ ISVs supporting both CUDA and Maximus include (among others): Dassault, ANSYS, Paradigm, Autodesk, Adobe, eyoen, PTC, Solidworks, Quantel, Abaqus, Adobe MATLAB, and Blackmagic.

QUALITY AND RELIABILITY GO HAND-IN-HAND

Quadro's design philosophy emphasizes attention to quality in every facet of development:

- ▶ Quality in engineering
- ▶ Manufacturing, and materials
- ▶ Testing and support

The heart and soul of Quadro products, Quadro GPUs undergo thorough chip qualification and testing, across voltage, frequency, temperature, and via life-test simulation, to virtually eliminate failures in the field.

Quadro boards cut no corners in pursuit of product integrity, with 30 µin gold-plated, PCI Express card fingers, for example. Quadro boards pass the most stringent durability tests, including tolerance to shock and vibration, while undergoing extensive thermal qualification, to ensure the hardware can stand up to harsh environments.

Along with nearly 9,000 hours of Quadro specific testing for each driver, NVIDIA also teams up with leading ISVs, including Autodesk and Dassault Systemés, to put Quadro hardware and drivers through a battery of application-specific regression tests ensuring each new configuration runs the same programs every other certified workstation before it did, accurately and reliably. ISV certified Quadro solutions eliminate guesswork in configuring drivers, operating system versions, application versions, and graphics cards. IT managers get the peace of mind knowing that the entire enterprise is running the same combination of hardware and software that suppliers have validated in their own labs, a combination they stand behind.

EASY TO DEPLOY AND MAINTAIN

Deploying, configuring and maintaining a single workstation isn't particularly burdensome. But now consider not one, but ten, a hundred, or even a thousand workstation seats, some of which need to be configured similarly and some differently, but all updated synchronously. Then keep in mind that one mistake or omission in procedure can mean a failure that leaves one user — or perhaps an entire staff — idle for the day. The need for an effective means to manage a sea of systems and users is paramount.

Supporting Windows 7 and beyond, and built upon the standard Windows Management Instrumentation (WMI) framework, NVIDIA's Display and GPU Management toolkit provides effective means to easily and remotely deploy, configure, update and monitor Quadro across the enterprise, keeping costs down and systems operational. An IT manager can create a fresh control panel profile or download a newly certified Quadro driver, then broadcast to all workstations in the enterprise, whether they number one or a thousand. Or if only one machine has an issue, he can reset or adjust its settings quickly and remotely.



He can monitor systems' health, querying graphics and display states; or go even further to detect hardware status that can predict failures before they happen, for example setting an alert if the GPU's temperature or fan speed exceeds a defined threshold. Quadro's attention to detail extends beyond the hardware to include full-featured software designed to minimize the time spent configuring GPUs and maximize the time spent using them.

WHAT THE KEPLER ARCHITECTURE BRINGS TO QUADRO

Now leveraging NVIDIA's latest Kepler™ GPU architecture, Quadro benefits from its most compelling, professional geared graphics architecture to date. To shape Kepler, NVIDIA architects took the best of the previous, workstation-proven Fermi architecture and improved on it for better performance and lower power consumption: for example, by implementing better task scheduling, more power-efficient organization of chip resources, and inclusion of new GPU Boost and Dynamic Parallelism technologies.

During operation, GPU Boost dynamically quantifies available headroom in thermal power, and then dials up chip frequencies and voltages accordingly, trading that headroom for increased performance. With Dynamic Parallelism, the GPU handles nested programming loops on its own, without CPU intervention, eliminating inter-processor communication overhead and driving up throughput.

Kepler's improved multi-display support means a single Quadro can supply up to four independent display streams. In pursuit of richer visuals, Kepler's "bindless" texture support delivers a virtually unlimited number of simultaneous textures, allowing modelers and animators more layers and effects with less overhead. And Kepler's new FXAA and TXAA anti-aliasing modes take another big step forward to photo-realistic image quality in real time.

Add it all up, and Kepler delivers a 3x improvement in performance/Watt with uncompromised image fidelity, yielding NVIDIA's most complete and highest-performing professional GPU to date.

PRODUCTIVITY: THE ONLY MEASURE THAT MATTERS IN PROFESSIONAL COMPUTING

Workstations — and the GPUs that drive them — aren't toys. They prove their worth not by metrics like frame rates or polygons/sec but by one measure only: how much faster can work get done with it than without it? With professional workflows perpetually compute-constrained, each new generation of workstation technology has the potential to reduce bottlenecks, shorten development cycles, and improve product quality.

With a long history of both evolutionary improvement and revolutionary innovation, NVIDIA's Quadro line of professional graphics solutions is time-proven and application-tested to deliver on that potential. Now tapping into Kepler technology, the latest round of Quadro products are NVIDIA's most compelling tools for businesses intent on improving that most important of metrics: the bottom line.

Notice

The information provided in this specification is believed to be accurate and reliable as of the date provided. However, NVIDIA Corporation ("NVIDIA") does not give any representations or warranties, expressed or implied, as to the accuracy or completeness of such information. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This publication supersedes and replaces all other specifications for the product that may have been previously supplied.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and other changes to this specification, at any time and/or to discontinue any product or service without notice. Customer should obtain the latest relevant specification before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer. NVIDIA hereby expressly objects to applying any customer general terms and conditions with regard to the purchase of the NVIDIA product referenced in this specification.

NVIDIA products are not designed, authorized or warranted to be suitable for use in medical, military, aircraft, space or life support equipment, nor in applications where failure or malfunction of the NVIDIA product can reasonably be expected to result in personal injury, death or property or environmental damage. NVIDIA accepts no liability for inclusion and/or use of NVIDIA products in such equipment or applications and therefore such inclusion and/or use is at customer's own risk.

NVIDIA makes no representation or warranty that products based on these specifications will be suitable for any specified use without further testing or modification. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer's sole responsibility to ensure the product is suitable and fit for the application planned by customer and to do the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer's product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this specification. NVIDIA does not accept any liability related to any default, damage, costs or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this specification, or (ii) customer product designs.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this specification. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA. Reproduction of information in this specification is permissible only if reproduction is approved by NVIDIA in writing, is reproduced without alteration, and is accompanied by all associated conditions, limitations, and notices.

ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, "MATERIALS") ARE BEING PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA's aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the NVIDIA terms and conditions of sale for the product.

VESA DisplayPort

DisplayPort and DisplayPort Compliance Logo, DisplayPort Compliance Logo for Dual-mode Sources, and DisplayPort Compliance Logo for Active Cables are trademarks owned by the Video Electronics Standards Association in the United States and other countries.

Trademarks

NVIDIA, the NVIDIA logo, CUDA, Kepler, Maximus, nView, and Quadro are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

© 2013 NVIDIA Corporation. All rights reserved.