**NVIDIA Quadro RTX 6000  
The World’s First Ray Tracing GPU**

NVIDIA Quadro RTX 6000, powered by the NVIDIA Turing™ architecture and the NVIDIA RTX platform, brings the most significant advancement in computer graphics in over a decade to professional workflows. Designers and artists can now wield the power of hardware-accelerated ray tracing, deep learning, and advanced shading to dramatically boost productivity and create amazing content faster than ever before.

New RT Cores and Tensor Cores bring the power of realtime ray tracing and AI-enhanced workflows to millions of design and creative professionals. Combined with NVIDIA NVLink™ technology, RTX 6000 scales graphics memory and performance to drive the most demanding rendering, AI, and visual computing workloads. And the all-new VirtualLink® provides connectivity to next-generation, high-resolution VR HMDs to let you view your work in the most compelling virtual environments. Welcome to the future of professional visual computing.

|  |  |
| --- | --- |
| CUDA Cores | 4608 |
| NVIDIA RT Cores | 72 |
| NVIDIA Tensor Cores | 576 |
| RTX-OPS | 84T |
| Rays Cast | 10 Giga Rays/Sec |
| Peak Single Precision FP32 Performance | 16.3 TFLOPS |
| Peak Half Precision FP16 Performance | 32.6 TFLOPS |
| Peak Integer Operations INT8 Performance | 206.1 TOPS |
| Deep Learning TFLOPS | 130.5 Tensor TFLOPS |
| GPU Memory | 24 GB GDDR6 with ECC |
| Memory Bandwidth | 624 GB/Sec |
| System Interface | PCI Express 3.0 x16 |
| Display Connectors | DisplayPort 1.4 (4) + VirtualLink |

**Incredible Application Performance**

* Experience fast, interactive, professional application performance
* Latest NVIDIA Turing GPU architecture and ultra-fast graphics memory

**New RT Cores for Realtime Ray Tracing**

* 72 RT Cores accelerate photorealistic ray-traced rendering
* NVIDIA RTX technology brings realtime rendering to professionals

**Tensor Cores Accelerate AI Workflows**

* 576 Tensor Cores power AI development and training
* Incredible inferencing performance is ideal for “at the edge” deployment

**Next Generation VR Support**

* VirtualLink simplifies HMD cabling
* Advanced rendering and shading features for immersive VR

**Quadro RTX NVLink**

* Scales memory and performance for the largest visual computing workloads
* Faster than PCIe-based solutions and offers flexible configurations

**PERFORMANCE AND USEABILITY FEATURES**

**Revolutionary Realtime Ray Tracing Acceleration**

Incorporating new hardware-based ray tracing engines, Turing is the industry’s first graphics processor enabling real-time ray tracing. A single Quadro RTX 6000 board can render complex professional models with physically accurate shadows, reflections, and refractions to empower users with instant insight. Working in concert with applications leveraging APIs such as NVIDIA OptiX, Microsoft DXR and Vulcan ray tracing, systems based on Quadro RTX 6000 will power truly interactive design workflows to provide immediate feedback for unprecedented levels of productivity.

**RT Cores**

New dedicated hardware-based ray-tracing technology allows the GPU for the first time to real-time render film quality, photorealistic objects and environments with physically accurate shadows, reflections, and refractions. The real-time ray-tracing engine works with NVIDIA OptiX, Microsoft DXR, and Vulkan APIs to deliver a level of realism far beyond what is possible using traditional rendering techniques. RT cores accelerate the Bounding Volume Hierarchy (BVH) traversal and ray casting functions using low number of rays casted through a pixel.

**Enhanced Tensor Cores**

Purpose-built for deep learning matrix arithmetic at the heart of neural network training and inferencing functions, Turing GPUs include enhanced Tensor Cores that accelerate FP16 / FP32 matrix operations in addition to INT8 and two new INT4 and INT1 (binary) precision modes. Independent floating-point and integer data paths allow more efficient execution of workloads using a mix of computation and addressing calculations.

**Mesh Shading**

Provides a compute-based geometry pipeline to speed processing and culling for geometrically complex models and scenes to improve performance by up to 2x.

**Motion Adaptive Shading**

Offers more granular control over how GPU horsepower is distributed (i.e. more cycles applied on the detailed areas of a scene and fewer on the less detailed areas) to increase performance and at the same image quality, or product similar image quality with a 50% reduction in the time required to generate shaded pixels.

**Variable Rate Shading (VRS)**

More control over pixel shading rate; efficient for effects like motion, blur, foveated shading. This capability enables shading and geometry samples to process at different rates for more efficient execution.

**Texture Space Shading**

Decouples shading from screen space, improving shading efficiency and reuse.

**Multi-View Rendering (MVR)**

Extension of Single Pass Stereo rendering multiple views in a single pass with unique view origin positions or view directions.

**High Speed GDDR6 Memory**

Built with 24GB of state-of-the-art GDDR6 memory delivering 55% greater throughput than the previous-generation GDDR5x technology, Quadro RTX 6000 is well-equipped to handle latency-sensitive applications handling large datasets.

**Error Correcting Code (ECC) on Graphics Memory**

Meet strict data integrity requirements for mission critical applications with uncompromised computing accuracy and reliability for workstations.

**Encode/Decode Engine Enhancements**

H.264 encode bitrate/quality improvements to modulate performance to reduce bitrate (e.g. ~6% at the same performance level, ~11% with reduced performance).

Vidmem output for encoded bitstream and motion vectors to accelerate post-processing for end-to-end GPU workflows.

Optical flow at 4 x 4 granularity for improved video frame interpolation/extrapolation, object tracking, motion segmentation, action recognition, etc.

I-frame only decode for H.264 and HEVC to allow inferencing with I-frame only to speed use cases such as video surveillance.

VP9 10-/12-bit decode to enable VP9 HBD content decoding.

**TURING VR**

Ray tracing and variable rate shading enhance visual quality, while Multi-view rendering provides wider field of view, support for next-generation HMDs and displays. The RT Cores provide for accurate acoustic simulations and VirtualLink makes VR setup easier than ever with a single cable connection.

**IMAGE QUALITY**

**Full-Scene Antialiasing (FSAA)**

Dramatically reduce visual aliasing artifacts or “jaggies” with up to 64x FSAA (128x with SLI Mode) for unparalleled image quality and highly realistic scenes.

**32K Texture and Render Processing**

Texture from and render to 32K x 32K surfaces to support applications that demand the highest resolution and quality image processing.

**DISPLAY FEATURES**

**NVIDIA Quadro Mosaic™ Technology**

Transparently scale the desktop and applications across up to 8 GPUs and 32 displays from a single workstation while delivering full performance and image quality.

**DisplayPort 1.4**

Support up to four 8K (7680 x 4320) monitors at 60 Hz, four 5K (5120 x 2880) displays at 60 Hz, or four 4K (3840 x 2160 or 4096 x 2160) displays at up to 120 Hz. Quadro RTX 6000 supports HDR over DisplayPort 1.4 (SMPTE 2084/2086. BT. 2020) with 4K at 60 Hz 10b/12b HEVC decode, 4K at 60 Hz 10b HEVC Encode.

**NVIDIA nView Advanced Desktop Software**

Gain unprecedented end-user control of the desktop experience for increased productivity in single large display or multi-display environments.

**NVIDIA Quadro Sync II**

Synchronize the display and image output of up to 32 displays from 8 GPUs (connected through two Quadro Sync II boards) in a single system, reducing the number of systems needed to create an advanced video visualization environment.

**OpenGL Quadro Buffered Stereo Support**

Provides a smooth and immersive 3D Stereo experience for professional applications.

**Professional 3D Stereo Synchronization**

Robust control of stereo effects through a dedicated connection to directly synchronize 3D stereo hardware to a Quadro graphics card.

**Ultra High Resolution Desktop Support**

Get more Mosaic topology choices using high-resolution display devices with a 32K max desktop size.

**VirtualLink**

VirtualLink is an industry standard alternate mode of USB Type-C that delivers 4 display lanes of HBR3 DisplayPort, USB 3.1 Gen2 SuperSpeed data, and up to 27 W of power to next generation VR HMDs (Head Mounted Displays).

**SOFTWARE SUPPORT**

**NVIDIA RTX Platform for Realtime Ray Tracing**

Software framework that makes realtime ray tracing possible, portable, and presentable.

Provides interoperability between rasterization, ray tracing, compute and AI/Deep Learning.

New Turing ray tracing acceleration in OptiX, DXR and Vulkan.

NVIDIA MDL, now open source, and support for Pixar’s Universal Scene Description (USD) promote portability and consistency.

**NVIDIA NGX: AI for Visual Applications**

Provides a neural graphics framework for image processing.

Brings AI-based features to end user applications to enhance graphics, photos, imaging, and video processing.

NGX features utilize Tensor Cores to maximize performance and efficiency.

Allows NVIDIA to bring advances in AI features to developers who place them into applications with NVIDIA pre-trained neutral networks.

**NVIDIA NGX Technology**

AI InPainting allows the removal of existing content from images and replaces it with realistic computer-generated alternatives.

AI Up-Res increases the resolution of an image or video by 2x, 4x, or 8x using AI to create new pixels by interpreting the image and intelligently placing data in the new image.

DLSS (Deep Learning Super Sample) removes jagged lines to smooth them, producing a higher quality image than by using other techniques

AI Slow-Motion inserts interpolated frames into a video stream to provide smooth, slow-motion video.

**Software Optimized for AI**

Deep learning frameworks such as Caffe2, MXNet, CNTK, TensorFlow, and others deliver dramatically faster training times and higher multi-node training performance. GPU accelerated libraries such as cuDNN, cuBLAS, and TensorRT deliver higher performance for both deep learning inference and High Performance Computing (HPC) applications.

**NVIDIA CUDA Parallel Computing Platform**

Natively execute standard programming languages like C/C++ and Fortran, and APIs such as CUDA, OpenCL, OpenACC and Direct Compute to accelerate techniques such as ray tracing, video and image processing, or finite element analysis or computational fluid dynamics (among others).

**Unified Memory**

A single, seamless 49-bit virtual address space allows for the transparent migration of data between the full allocation of CPU and GPU memory.

**NVIDIA GPUDirect for Video**

GPUDirect for Video speeds communication between the GPU and video I/O devices by avoiding unnecessary system memory copies and CPU overhead.

**MULTI-GPU TECHNOLOGY SUPPORT**

**NVIDIA NVLink High Speed Interface**

Connect a pair of Quadro RTX 6000 boards with NVLink to essentially double the amount of GPU memory available, double CUDA, RT and Tensor Core counts, and scale application performance by enabling GPU-to-GPU data transfers at rates up to 100 GB/Sec.

**3D GRAPHICS ARCHITECTURE**

* Scalable geometry architecture
* Hardware tessellation engine
* NVIDIA GigaThread™ engine with 7 async copy engines
* Shader Model 5.1 (OpenGL 4.5 and DirectX 12)
* Up to 32K x 32K texture and render processing
* Transparent multisampling and super sampling
* 16x angle independent anisotropic filtering
* 32-bit per-component floating point texture filtering and blending
* 64x full scene antialiasing (FSAA)/128x FSAA in SLI Mode
* Decode acceleration for MPEG-2, MPEG-4 Part 2 Advanced Simple Profile, H.264, HEVC, MVC, VC1, DivX (version 3.11 and later), and Flash (10.1 and later)
* Dedicated H.264 & HEVC Encoder
* Bly-ray dual-stream hardware acceleration (supporting HD picture-in picture playback)
* NVIDIA GPU boost automatically improves GPU engine throughput to maximize application performance

**NVIDIA CUDA PARALLEL PROCESSING ARCHITECTURE**

* New RT (Ray Tracing) Core per SM
* Turing SM Architecture (streaming multi-processor design that delivers greater processing efficiency)
* Dynamic Parallelism (GPU dynamically spawns new threads without going back to the CPU)
* Mixed-precision (1-, 4-, 8-, 16-, 32- and 64-bit) computing
* API support includes: CUDA C, CUDA C++, DirectCompute 5.0, OpenCL, Java, Python, and Fortran
* Error correction codes (ECC) on graphics memory
* Configurable up to 96 KB of RAM (dedicated shared memory size per SM)

**ADVANCED DISPLAY FEATURES**

* Support for any combination of four connected displays
* Four DisplayPort 1.4 outputs (supporting resolutions such as 3840 x 2160 at 120 Hz, 5120 x 2880 at 60 Hz and 7680 x 4320 at 60Hz)
* DisplayPort to VGA, DisplayPort to DVI (single-link and dual-link) and DisplayPort to HDMI cables (resolution support based on dongle specifications)
* HDR support over DisplayPort 1.4 (SMPTE 2084/2086, BT. 2020) (4K at 60Hz 10b/12b HEVC Decode, 4K at 60Hz 10b HEVC Encode)
* HDCP 2.2 support over DisplayPort and HDMI connectors
* 12-bit internal display pipeline (hardware support for 12-bit scanout on supported panels, applications and connection)
* NVIDIA 3D Vision™ technology, 3D DLP, Interleaved, and other 3D stereo format support
* Full OpenGL quad buffered stereo support
* Underscan/overscan compensation and hardware scaling
* NVIDIA nView multi-display technology
* Support for large-scale, ultra-high resolution visualization using the NVIDIA SVS platform which includes NVIDIA Mosaic, NVIDIA Sync and NVIDIA Warp/Blend technologies

**DISPLAYPORT AND HDMI DIGITAL AUDIO**

* Support for the following audio modes: Dolby Digital (AC3), DTS 5.1, Multi-channel (7.1) LPCM, Dolby Digital Plus (DD+), and MPEG-2/MPEG-4 AAC
* DisplayPort Data rates of 48 KHz
* HDMI Digital Audio Data rates of 44.1 KHz, 48 KHz, 88.2 KHz, 96 KHz, 176 KHz, and 192 KHz
* Word sizes of 16-bit, 20-bit, and 24-bit

**AVAILABLE ACCESSORIES**

* **RTX6KNVLNKX16S2RKIT** provides an NVLink connector for the RTX 6000 suitable for standard PCIe slot spacing motherboards, effectively fusing two physical boards into one logical entity with 9216 CUDA Cores, 1152 Tensor Cores, 144 RT Cores, and 48 GB of GDDR6 memory, with a bandwidth of 100 GB/Sec. Order PN RTX6KNVLNKX16S2RKIT when ultimate performance and capabilities are required with two RTX 6000 boards (application support required).
* **RTX6KNVLINKX16S3RKIT** offers an NVLink connector for the RTX 6000 copatible with motherboards featuring wider PCIe slot spacing. All other features and benefits are identical to the standard slot spacing version.
* **DP-HDMI-THREE-PCK** Connect the Quadro RTX 5000 to HDMI displays at resolutions up to 4K with PNY Part Number DP-HDMI-THREE-PCK. The four included DisplayPort to HDMI adapters are recommended by NVIDIA, provide outstanding image quality, and are built to professional standards.
* **DP-HDMI-SINGLE-PCK** Connect the Quadro RTX 5000 to an HDMI display at resolutions up to 4K with PNY Part Number DP-HDMI-SINGLE-PCK. The included DisplayPort to HDMI adapter is recommended by NVIDIA, provides outstanding image quality, and is built to professional standards.

**SPECIFICIATIONS**

|  |  |
| --- | --- |
| **CUDA Cores** | 4608 |
| **NVIDIA RT Cores** | 72 |
| **NVIDIA Tensor Cores** | 576 |
| **GPU Memory** | 24 GB GDDR6 with ECC |
| **RTX-OPS** | 84T |
| **Rays Cast** | 10 Giga Rays/Sec |
| **Peak Single Precision FP32 Performance** | 16.3 TFLOPS |
| **Peak Half Precision FP16 Performance** | 32.6 TFLOPS |
| **Peak Half Precision INT8 Performance** | 206.1 TOPS |
| **Deep Learning TFLOPS**[**1**](https://www.pny.com/nvidia-quadro-rtx-6000#1) | 130.5 Tensor TFLOPS |
| **Multi-GPU Scalability** | NVLink |
| **NVLink Bandwidth** | 100 GB/Sec |
| **Memory Bandwidth** | 624 GB/s |
| **System Interface** | PCI Express 3.0 x16 |
| **Maximum Power Consumption** | [295 W](http://www.step.sister.loves.making.brother.pny.com/File%20Library/Support/PNY%20Products/User%20Guides%20and%20Tutorials/Quadro/Quadro-Power-Guidelines.pdf) |
| **Energy Star Enabling** | Yes |
| **Thermal Solution** | Ultra-quiet active fansink |
| **Form Factor** | 4.4” H x 10.5” L, Dual Slot |
| **Display Connectors** | DisplayPort 1.4 (4) + VirtualLink |
| **DVI-D Single-Link Connector** | Yes, via included adapter |
| **HDMI Support** | Yes, via included adapter |
| **Number of Displays Supported** | 4 |
| **Maximum DP 1.4 Resolution** | HDR 7680 x 4320 at 60 Hz |
| **5K Display Support** | HDR 5120 x 2880 at 60 Hz |
| **4K Display Support** | HDR 4096 x 2160 or 3840 x 2160 at 120 Hz |
| **Maximum DVI-D DL Resolution** | 2560 x 1600 at 60 Hz via 3rd party adapter |
| **Maximum DVI-D SL Resolution** | 1920 x 1200 at 60 Hz via included adapter |
| **HDCP Support** | Yes |
| **Professional 3D Support** | Via optional Stereo Connector Bracket |
| **Quadro Sync II Compatible** | Yes (Frame Lock and Genlock) |
| **NVIDIA GPU Direct Compatible** | Yes |
| **Graph APIs** | Shader Model 5.1, OpenGL 4.5, DirectX 12.0, Vulkan 1.0 |
| **Compute APIs** | CUDA, DirectCompute, OpenCL |
| **NVIEW** | Yes |
| **NVIDIA MOSAIC** | Yes |
| **Warranty** | 3 Years |

1. FP16 matrix multiply with FP16 and FP32 accumulate

**WARRANTY & SUPPORT**

PNY provides unsurpassed service and commitment to its professional graphics customers offering: 3-year warranty, free pre- and post-sales support, dedicated Quadro Field Application engineers and direct tech support hot lines. In addition, PNY delivers a complete solution including the appropriate adapters, cables, brackets, software installation disc and documentation to ensure a quick and successful install.

* [3 Year Limited Warranty](https://www.pny.com/File%20Library/Support/PNY%20Products/Warranties/Quadro%20Graphics/NVIDIA-Quadro-Warranty.pdf)

**SUPPORTED PLATFORMS**

* Microsoft Windows 10 (64-bit)
* Microsoft Windows 8.1 and 8 (64-bit)
* Microsoft Windows 7 (64-bit)
* Linux – Full OpenGL implementation, complete with NVIDIA and ARB extensions (64-bit)

**MINIMUM SYSTEM HARDWARE REQUIREMENTS**

* Intel Core i5, i7, Xeon or later
* AMD Phenom or Opteron-class or later
* PCIe x16 Gen 3 (preferred) expansion slot
* 8GB of system memory (more preferred)
* Internet connection (for driver installation)
* DisplayPort (preferred), HDMI, or DVI compatible display