Datasheet

NVIDIA RTX PRO 5000 Blackwell GPUs

Powering the next era of AI.

Transform Workflows With Next-Level Workstation Performance

As AI continues to advance at an incredible pace, industries face mounting pressure to harness its transformative power and adopt tools capable of handling generative AI, real-time simulation, and hyperrealistic rendering. Enterprises need solutions that combine breakthrough performance, scalability, and versatility to tackle the rise of increasingly complex workloads—from training domain-specific AI models to rendering billion-polygon engineering designs or simulating real-world physics with pixel-perfect accuracy.

Unlock next-level AI performance and neural rendering capabilities with NVIDIA RTX PRO™ 5000 Blackwell GPUs. Built on the NVIDIA Blackwell architecture and equipped with 48 or 72 GB of ultra-fast GDDR7 memory, it accelerates everything from AI development, large language model (LLM) inference, and generative AI workflows, along with high-fidelity simulations, video production, and complex 3D modeling from your desktop. With twice the memory of the last generation, you can work with big data, make complex 3D models, and run AI-enhanced multi-app workflows without interruption.

With transformative AI performance gains over the previous generation, you can run complex generative AI workflows more efficiently using the new FP4 data format, which reduces memory demands. Accelerate graphics workloads with dramatic speed boosts to design hyperdetailed vehicles, immersive VR environments, and lifelike renders faster than ever. Streamline video production pipelines with up to three encode and three decode engines that support 4:2:2 color formats and multi-stream workflows.

With the RTX PRO 5000, seamlessly run local AI assistants, craft photorealistic visuals with neural rendering, and optimize precision-critical simulations for engineering and scientific research, all with unmatched stability and speed.

Key Features

- Enhanced streaming multiprocessors (SMs) built for neural shaders
- Fifth-gen Tensor Cores support FP4 precision and DLSS 4 Multi Frame Generation
- Fourth-gen Ray Tracing Cores built for detailed geometry
- 48 or 72 GB of GDDR7 memory
- 1.3 TB/s of memory bandwidth
- Ninth-gen NVENC and sixth-gen NVDEC with 4:2:2 support
- PCIe Gen 5
- Four DisplayPort 2.1b connectors
- Multi-instance GPU (MIG) support
- AI Management Processor

Breakthrough Innovations

The NVIDIA Blackwell architecture combines breakthrough AI, ray tracing, and neural rendering technology, with massive performance and memory improvements to drive cutting-edge professional creative, design, and engineering workflows and power the next decade of innovation.

Blackwell Streaming Multiprocessor: The new SM features increased processing throughput and new neural shaders that integrate neural networks inside of programmable shaders to drive the next decade of AI-augmented graphics innovations.

Fifth-Gen Tensor Cores: Deliver up to 3x the performance of the previous generation and support for FP4 precision for faster AI model processing times with reduced memory usage, enabling local fine-tuning of LLMs and generative AI.

Fourth-Gen Ray Tracing Cores: Double the ray-triangle intersection rate of the previous generation to create photoreal, physically accurate scenes and immersive 3D designs with RTX Mega Geometry, which enables up to 100x more ray-traced triangles.

Next-Gen Video Engines: Enhance video conferencing, video production, and streaming workflows with real-time AI processing. Ninth-generation NVENC and sixth-generation NVDEC engines provide support for 4:2:2 encoding and decoding to explore a new realm of high-resolution video workflows.

GDDR7 Memory: New and improved GDDR7 memory significantly boosts bandwidth and capacity, empowering your applications to run faster and work with larger, more complex datasets. With 48 or 72 GB of GPU memory and 1.3 TB/s of bandwidth, tackle massive 3D and AI projects, fine-tune AI models locally, explore large-scale VR environments, and drive larger multi-app workflows.

DLSS 4: Multi Frame Generation ensures ultra-smooth frame pacing for lifelike simulations. Experience up to 3x faster frame rates and stunning image quality in supported game engines and 3D rendering applications for smoother, more responsive performance.

PCIe Gen 5: Support for PCIe Gen 5 provides double the bandwidth of PCIe Gen 4, improving data-transfer speeds from CPU memory and unlocking faster performance for data-intensive tasks like AI, data science, and 3D modeling.

DisplayPort 2.1: Achieve unparalleled visual clarity and performance, driving high-resolution displays at up to 8K at 240 Hz and 16K at 60 Hz. Increased bandwidth enables seamless multi-monitor setups, ideal for multitasking and collaboration, while HDR and higher color depth support ensures superior color accuracy for precision work, such as video editing, 3D design, and live broadcasting.

Universal MIG: Split a single RTX PRO 5000 into multiple isolated instances, each with its own resources, allowing for concurrent execution of multiple workloads, optimized GPU utilization, and secure isolation of different applications or users.

Enterprise Reliability

Designed for professionals who demand the best, NVIDIA RTX PRO solutions deliver unparalleled performance, reliability, and support. Every GPU is rigorously tested for a wide range of design, engineering, and AI workflows and continually optimized through enterprise drivers. With extensive ISV certifications, robust IT management tools, and enterprise-grade support, RTX PRO workstations are the trusted choice for enterprise and mission-critical work.

Specifications

	NVIDIA RTX PRO 5000 Blackwell	NVIDIA RTX PRO 5000 72GB Blackwell
GPU Architecture	NVIDIA Blackwell	
NVIDIA® CUDA® Cores	14,080	
Tensor Cores	Fifth generation	
Ray Tracing Cores	Fourth generation	
AI TOPS	2,064 AI TOPS[1,2]	
Single-Precision Performance	65 TFLOPS[1]	
RT Core Performance	196 TFLOPS[1]	
GPU Memory	48 GB GDDR7 with ECC	72 GB GDDR7 with ECC
Memory Interface	512-bit	
Memory Bandwidth	1,344 GB/s	
System Interface	PCIe 5.0 x16	
Display Connectors	4x DisplayPort 2.1b	
Max Simultaneous Displays	>4x 4,096 x 2,160 @ 120 Hz >4x 5,120 x 2,880 @ 60 Hz >2x 7,680 x 4,320 @ 60 Hz	
Video Engines	3x NVENC (ninth gen) 3x NVDEC (sixth gen)	
MIG Support	>Up to 2x 24 GB >Up to 1x 48 GB	>Up to 2x 36 GB >Up to 1x 72 GB
Power Consumption	Total board power: 300 W	
Power Connector	1x PCIe CEM5 16-pin	
Thermal Solution	Active	
Form Factor	4.4” x 10.5” L, dual slot, full height	
Graphics APIs	DirectX 12, Shader Model 6.6, OpenGL 4.6[3], Vulkan 1.3[3]	
Compute APIs	CUDA 12.8, OpenCL 3.0, DirectCompute	

Ready to Get Started?

To learn more, visit: nvidia.com/rtx-pro-5000

1. Peak rates based on GPU Boost Clock.
2. Effective FP4 TOPS with sparsity.
3. Product is based on a published Krhonos specification and is expected to pass the Khronos conformation testing process when available. Current conformance status can be found at www.khronos.org/conformance

NVIDIA