

Remains

SPECIFICATIONS

Amplified Performance for Professionals

NVIDIA RTX A5000

PERFECTLY BALANCED. BLAZING PERFORMANCE.

The NVIDIA RTX[™] A5000 delivers the power, performance, capabilities, and reliability professionals need to bring their boldest ideas to life. Built on the NVIDIA Ampere architecture, the RTX A5000 combines 64 second-generation RT Cores, 256 third-generation Tensor Cores, and 8,192 CUDA[®] cores with 24 GB of graphics memory to supercharge rendering, AI, graphics, and compute tasks. Connect two RTX A5000s with NVIDIA NVLink¹ to scale memory and performance with multi-GPU configurations², allowing professionals to work with memory intensive tasks such as large models, ultra-high resolution rendering, and complex compute workloads. Support for NVIDIA virtual GPU software increases the versatility for enterprise deployments.

NVIDIA RTX professional graphics cards are certified with a broad range of professional applications, tested by leading independent software vendors (ISVs) and workstation manufacturers, and backed by a global team of support specialists. Get the peace of mind needed to focus on what matters with the premier visual computing solution for mission-critical business.

Features

- > PCI Express Gen 4
- > Four DisplayPort 1.4a connectors
- > AV1 decode support
- > DisplayPort with audio
- > 3D stereo support with stereo connector
- > NVIDIA GPUDirect[®] for Video support
- > NVIDIA virtual GPU (vGPU) software support
- > NVIDIA Quadro[®] Sync II³ compatibility
- > NVIDIA RTX Experience[™]
- > NVIDIA RTX Desktop Manager software
- > NVIDIA RTX IO support
- > HDCP 2.2 support
- > NVIDIA Mosaic⁴ technology

To learn more about the NVIDIA RTX A5000, visit www.pny.com/nvidia-rtx-a5000

1 NVIDIA NVLink sold separately. | 2 Connecting two RTX A5000 cards with NVLink to scale performance and memory capacity to 48GB is only possible if your application supports NVLink technology. Please contact your application provider to confirm their support for NVLink. | 3 Quadro Sync II card sold separately. | 4 Windows 10 and Linux. | 5 Peak rates based on GPU Boost Clock. | 6 Effective teraFLOPS (TFLOPS) using the new sparsity feature. | 7 Display ports are on by default for RTX A5000. Display ports are not active when using vGPU software. | 8 GPU supports DX 12.0 API, hardware feature level 12 + 1. | 9 Product is based on a published Khronos specification and is expected to pass the Khronos conformance testing process when available. Current conformance status can be found at **www.khronos.org/conformance**

© 2021 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, CUDA, GPUDirect, NVLink, Quadro, RTX Experience, and RTX 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. All other trademarks are property of their respective owners.

Single Bulk: VCNRTXA5000-SB Education: VCNRTXA5000-EDU Bulk: VCNRTXA5000-BLKGPU memory24 GB GDDR6Memory interface384-bitMemory bandwidth768 GB/sError-correcting code (ECC)YesNVIDIA Ampere architecture- based CUDA Cores8,192NVIDIA third-generation Tensor Cores256NVIDIA second-generation RT Cores64Single-precision performance22.2 TFLOPS5RT Core performance22.2 TFLOPS5NVIDIA NVLinkLow profile bridges connect two NVIDIA NVLinkNVIDIA NVLink112.5 GB/s (bidirectional)System interfacePCI Express 4.0 x16Power consumptionActiveForm factor4.4 Y 10.5" L, dual stot, full heightDisplay connectors4.x DisplayPort 1.4a7Max simultaneous displaysX x096 x 2160 @120 Hz, x x5120 x 2880 @ 60 Hz, 27 K80 x 4320 @	PNY Part Number	Retail: VCNRTXA5000-PB
Bulk: VCNRTXA5000-BLKGPU memory24 GB GDDR6Memory interface384-bitMemory bandwidth768 GB/sError-correcting code [ECC]YesNVIDIA Ampere architecture- based CUDA Cores8,192NVIDIA third-generation Tensor Cores64NVIDIA second-generation RT Core performance64.2Single-precision performance54.2 TFLOPS*Tensor performance54.2 TFLOPS*NVIDIA NVLinkLow profile bridges connect two NVIDIA RTX A5000 GPUs*NVIDIA NVLink112.5 GB/s (bidirectional)System interfacePCI Express 4.0 x16Power consumptionTotal board power: 230 WThermal solutionActiveForm factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4 x 4096 x 2160 @ 120 Hz, x x 5120 x 2880 @ 60 Hz, 2 x 7680 x 4320 @ 60 Hz		
GPU memory24 GB GDDR6Memory interface384-bitMemory bandwidth768 GB/sError-correcting code (ECC)YesNVIDIA Ampere architecture- based CUDA Cores8,192NVIDIA third-generation Tensor Cores256NVIDIA second-generation RT Core performance64RT Core performance27.8 TFLOPS*RT Core performance22.2 TFLOPS*NVIDIA NVLinkLow profile bridges connect two NVIDIA NVLinkSystem interfacePCI Express 4.0 x16Power consumptionTotal board power: 230 WThermal solutionActiveForm factor4.4" H x 10.5" L, dual stot, full heightDisplay connectors4 x 4096 x 2160 @ 120 Hz, x 7580 x 4320 @ 60 Hz, 2x 7680 x 4		
Memory interface384-bitMemory bandwidth768 GB/sError-correcting code [ECC]YesNVIDIA Ampere architecture- based CUDA Cores8,192NVIDIA third-generation Tensor Cores256NVIDIA second-generation RT Cores64Single-precision performance27.8 TFLOPS*RT Core performance24.2 TFLOPS*NVIDIA NVLinkLow profile bridges connect two NVIDIA RTX A5000 GPUs'NVIDIA NVLink112.5 GB/s (bidirectional)System interfacePCI Express 4.0 x16Power consumptionTotal board power: 230 WThermal solutionActiveForm factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4 x 4096 x 2160 @ 120 Hz, x x 7680 x 4320 @ 60 Hz, z x 7		Bulk: VCNRTXA5000-BLK
Memory bandwidth768 GB/sError-correcting code (ECC)YesNVIDIA Ampere architecture- based CUDA Cores8,192NVIDIA third-generation Tensor Cores256NVIDIA second-generation RT Cores64Single-precision performance27.8 TFLOPS*RT Core performance22.2 TFLOPS*NVIDIA NVLinkLow profile bridges connect two NVIDIA RTX A5000 GPUs*NVIDIA NVLink112.5 GB/s (bidirectional)System interfacePCI Express 4.0 x16Power consumptionTotal board power: 230 WThermal solutionActiveForm factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4x 4096 x 2160 @ 120 Hz, x x 7680 x 4320 @ 60 Hz, z x 7680 x 4320 @ 60 Hz, <br< td=""><td>GPU memory</td><td>24 GB GDDR6</td></br<>	GPU memory	24 GB GDDR6
Error-correcting code [ECC]YesNVIDIA Ampere architecture- based CUDA Cores8,192NVIDIA third-generation Tensor Cores256NVIDIA second-generation RT Cores64Single-precision performance27.8 TFLOPS*RT Core performance22.2 TFLOPS*NVIDIA NVLinkLow profile bridges connect two NVIDIA RTX A5000 GPUs*NVIDIA NVLink112.5 GB/s Ibidirectional]System interfacePCI Express 4.0 x16Power consumptionTotal board power: 230 WThermal solutionActiveForm factor4x DisplayPort 1.4a*Max simultaneous displays4x 4096 x 2160 @ 120 Hz, x 5760 x 4320 @ 60 Hz, 2x 7680 x 4320 @ 60 H	Memory interface	384-bit
NVIDIA Ampere architecture- based CUDA Cores8,192NVIDIA third-generation Tensor Cores256NVIDIA second-generation RT Cores64Single-precision performance27.8 TFLOPS*RT Core performance54.2 TFLOPS*Tensor performance222.2 TFLOPS*NVIDIA NVLinkLow profile bridges connect two NVIDIA RTX A5000 GPUs*NVIDIA NVLink bandwidth112.5 GB/s (bidirectional)System interfacePCI Express 4.0 x16Power consumptionTotal board power: 230 WThermal solutionActiveForm factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4 x DisplayPort 1.4a"Max simultaneous displays54.200 GPU 20 Hz, 4x 5120 x 2880 @ 60 Hz, 2x 7680 x 4320 @ 60 HzPower connector1x 8-pin PCleEncode/decode engines1x encode, 2x decode [+AV1 decode]VGPU software support?NVIDIA VPC/vApps, NVIDIA RTX Virtual WorkstationvGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsCUDA 11.6, DirectCompute,	Memory bandwidth	768 GB/s
based CUDA CoresINVIDIA third-generation Tensor Cores256NVIDIA second-generation RT Cores64Single-precision performance27.8 TFLOPS*RT Core performance22.2 TFLOPS*Tensor performance222.2 TFLOPS*NVIDIA NVLinkLow profile bridges connect two NVIDIA RTX A5000 GPUs*NVIDIA NVLink bandwidth112.5 GB/s Ibidirectional)System interfacePCI Express 4.0 x16Power consumptionTotal board power: 230 WThermal solutionActiveForm factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4 x DisplayPort 1.4a"Max simultaneous displays5420 % 60 Hz, x 257 680 x 4320 % 60 Hz, x 257 680 x 4320 % 60 HzPower connector1x 8-pin PCleEncode/decode engines1x encode, 2x decode [+AV1 decode]VGPU software support?NVIDIA vPC/vApps, NVIDIA RTX virtual WorkstationvGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsCUDA 11.6, DirectCompute,	Error-correcting code (ECC)	Yes
Tensor Cores64NVIDIA second-generation RT Cores64Single-precision performance27.8 TFLOPS5RT Core performance222.2 TFLOPS5Tensor performance222.2 TFLOPS5NVIDIA NVLinkLow profile bridges connect two NVIDIA RTX A5000 GPUs1NVIDIA NVLink bandwidth112.5 GB/s (bidirectional)System interfacePCI Express 4.0 x16Power consumptionTotal board power: 230 WThermal solutionActiveForm factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4 x DisplayPort 1.4a7Max simultaneous displays4x 4096 x 2160 @ 120 Hz, 4x 5120 x 2880 @ 60 Hz, 2x 7680 x 4320 @ 60 HzPower connector1x 8-pin PCleEncode/decode engines1x encode, 2x decode (+AV1 decode)VGPU software support7NVIDIA VC/vApps, NVIDIA RTX Virtual WorkstationvGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsCUDA 11.6, DirectCompute,		8,192
RT CoresStrandSingle-precision performance27.8 TFLOPS*RT Core performance54.2 TFLOPS*Tensor performance222.2 TFLOPS*NVIDIA NVLinkLow profile bridges connect two NVIDIA RTX A5000 GPUS*NVIDIA NVLink bandwidth112.5 GB/s (bidirectional)System interfacePCI Express 4.0 x16Power consumptionTotal board power: 230 WThermal solutionActiveForm factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4 x DisplayPort 1.4a"Max simultaneous displaysXx 4096 x 2160 @ 120 Hz, X 5720 x 2880 @ 60 Hz, 2 x 7680 x 4320 @ 60 HzPower connector1x 8-pin PCleEncode/decode engines1x encode, 2x decode I+AV1 decodelVGPU software support?NVIDIA vPC/vApps, NVIDIA RTX Virtual WorkstationvGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsCUDA 11.6, DirectCompute,		256
RT Core performance54.2 TFLOPS*Tensor performance222.2 TFLOPS*NVIDIA NVLinkLow profile bridges connect two NVIDIA RTX A5000 GPUs*NVIDIA NVLink bandwidth112.5 GB/s (bidirectional)System interfacePCI Express 4.0 x16Power consumptionTotal board power: 230 WThermal solutionActiveForm factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4 x DisplayPort 1.4a"Max simultaneous displays4x 4096 x 2160 @ 120 Hz, 4 x 5120 x 2880 @ 60 Hz, 2 x 7680 x 4320 @ 60 HzPower connector1x 8-pin PCleEncode/decode engines1x encode, 2x decode (+AV1 decode)VGPU software support?NVIDIA vPC/vApps, NVIDIA RTX Virtual WorkstationvGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsDirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68", Vulkan 1.3"		64
Tensor performance222.2 TFLOPS*NVIDIA NVLinkLow profile bridges connect two NVIDIA RTX A5000 GPUs*NVIDIA NVLink bandwidth112.5 GB/s (bidirectional)System interfacePCI Express 4.0 x16Power consumptionTotal board power: 230 WThermal solutionActiveForm factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4x DisplayPort 1.4a"Max simultaneous displays4x 4096 x 2160 @ 120 Hz, 4x 5120 x 2880 @ 60 Hz, 2x 7680 x 4320 @ 60 HzPower connector1x 8-pin PCIeEncode/decode engines1x encode, 2x decode (+AV1 decode)VGPU software support?NVIDIA vPC/vApps, NVIDIA RTX Virtual WorkstationvGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsDirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68", Vulkan 1.3"	Single-precision performance	27.8 TFL0PS⁵
NVIDIA NVLinkLow profile bridges connect two NVIDIA RTX A5000 GPUs1NVIDIA NVLink bandwidth112.5 GB/s (bidirectional)System interfacePCI Express 4.0 x16Power consumptionTotal board power: 230 WThermal solutionActiveForm factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4x DisplayPort 1.4a"Max simultaneous displays4x 4096 x 2160 @ 120 Hz, 4x 5120 x 2880 @ 60 Hz, 2x 7680 x 4320 @ 60 HzPower connector1x 8-pin PCleEncode/decode engines1x encode, 2x decode (+AV1 decode)VR readyYesvGPU software support?NVIDIA vPC/vApps, NVIDIA RTX Virtual WorkstationvGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsDirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68", Vulkan 1.3"	RT Core performance	54.2 TFLOPS ⁵
NVIDIA RTX A5000 GPUs1NVIDIA RTX A5000 GPUs1NVIDIA NVLink bandwidth112.5 GB/s (bidirectional)System interfacePower consumptionTotal board power: 230 WThermal solutionActiveForm factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4x 4096 x 2160 @ 120 Hz, 4x 5120 x 2880 @ 60 Hz, 2x 7680 x 4320 @ 60 HzPower connector1x 8-pin PCleEncode/decode enginesVGPU software support2VGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsCompute APIsCupy Licensing Cupy Cupy And	Tensor performance	222.2 TFL0PS ⁶
System interfacePCI Express 4.0 x16Power consumptionTotal board power: 230 WThermal solutionActiveForm factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4x DisplayPort 1.4a?Max simultaneous displays4x 4096 x 2160 @ 120 Hz, 4x 5120 x 2880 @ 60 Hz, 2x 7680 x 4320 @ 60 HzPower connector1x 8-pin PCleEncode/decode engines1x encode, 2x decode [+AV1 decode]VR readyYesvGPU software support?NVIDIA vPC/vApps, NVIDIA RTX Virtual WorkstationvGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsDirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68", Vulkan 1.3"Compute APIsCUDA 11.6, DirectCompute,	NVIDIA NVLink	
Power consumptionTotal board power: 230 WThermal solutionActiveForm factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4x DisplayPort 1.4a?Max simultaneous displays4x 4096 x 2160 @ 120 Hz, 4 x 5120 x 2880 @ 60 Hz, 2 x 7680 x 4320 @ 60 HzPower connector1x 8-pin PCleEncode/decode engines1x encode, 2x decode (+AV1 decode)VR readyYesvGPU software support?NVIDIA vPC/vApps, NVIDIA RTX Virtual WorkstationvGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsDirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68", Vulkan 1.3"Compute APIsCUDA 11.6, DirectCompute,	NVIDIA NVLink bandwidth	112.5 GB/s (bidirectional)
Thermal solutionActiveForm factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4 x DisplayPort 1.4a"Max simultaneous displays4x 4096 x 2160 @ 120 Hz, 4x 5120 x 2880 @ 60 Hz, 2x 7680 x 4320 @ 60 HzPower connector1x 8-pin PCleEncode/decode engines1x encode, 2x decode [+AV1 decode]VR readyYesvGPU software support?NVIDIA vPC/vApps, NVIDIA RTX Virtual WorkstationvGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsDirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68", Vulkan 1.3"Compute APIsCUDA 11.6, DirectCompute,	System interface	PCI Express 4.0 x16
Form factor4.4" H x 10.5" L, dual slot, full heightDisplay connectors4x DisplayPort 1.4a"Max simultaneous displays4x 4096 x 2160 @ 120 Hz, 4x 5120 x 2880 @ 60 Hz, 2x 7680 x 4320 @ 60 HzPower connector1x 8-pin PCleEncode/decode engines1x encode, 2x decode [+AV1 decode]VR readyYesvGPU software support?NVIDIA vPC/vApps, NVIDIA RTX Virtual WorkstationvGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsDirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68", Vulkan 1.3"Compute APIsCUDA 11.6, DirectCompute,	Power consumption	Total board power: 230 W
dual slot, full heightDisplay connectors4x DisplayPort 1.4a7Max simultaneous displays4x 4096 x 2160 @ 120 Hz, 4x 5120 x 2880 @ 60 Hz, 2x 7680 x 4320 @ 60 HzPower connector1x 8-pin PCleEncode/decode engines1x encode, 2x decode (+AV1 decode)VR readyYesvGPU software support?NVIDIA vPC/vApps, NVIDIA RTX Virtual WorkstationvGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsDirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68°, Vulkan 1.3°Compute APIsCUDA 11.6, DirectCompute,	Thermal solution	Active
Max simultaneous displays4x 4096 x 2160 @ 120 Hz, 4x 5120 x 2880 @ 60 Hz, 2x 7680 x 4320 @ 60 HzPower connector1x 8-pin PCleEncode/decode engines1x encode, 2x decode (+AV1 decode)VR readyYesvGPU software support?NVIDIA vPC/vApps, NVIDIA RTX Virtual WorkstationvGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsDirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68°, Vulkan 1.3°Compute APIsCUDA 11.6, DirectCompute,	Form factor	
4x 5120 x 2880 @ 60 Hz, 2x 7680 x 4320 @ 60 HzPower connector1x 8-pin PCleEncode/decode engines1x encode, 2x decode (+AV1 decode)VR readyYesvGPU software support?NVIDIA vPC/vApps, NVIDIA RTX Virtual WorkstationvGPU profiles supportedSee the Virtual GPU Licensing GuideGraphics APIsDirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68°, Vulkan 1.3°Compute APIsCUDA 11.6, DirectCompute,	Display connectors	4x DisplayPort 1.4a ⁷
Encode/decode engines 1x encode, 2x decode [+AV1 decode] VR ready Yes vGPU software support? NVIDIA vPC/vApps, NVIDIA RTX Virtual Workstation vGPU profiles supported See the Virtual GPU Licensing Guide Graphics APIs DirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68°, Vulkan 1.3° Compute APIs CUDA 11.6, DirectCompute,	Max simultaneous displays	4x 5120 x 2880 @ 60 Hz,
VR ready Yes vGPU software support ⁷ NVIDIA vPC/vApps, NVIDIA RTX Virtual Workstation vGPU profiles supported See the Virtual GPU Licensing Guide Graphics APIs DirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68°, Vulkan 1.3° Compute APIs CUDA 11.6, DirectCompute,	Power connector	1x 8-pin PCle
vGPU software support? NVIDIA vPC/vApps, NVIDIA RTX Virtual Workstation vGPU profiles supported See the Virtual GPU Licensing Guide Graphics APIs DirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68°, Vulkan 1.3° Compute APIs CUDA 11.6, DirectCompute,	Encode/decode engines	1x encode, 2x decode (+AV1 decode)
Virtual Workstation vGPU profiles supported See the Virtual GPU Licensing Guide Graphics APIs DirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68°, Vulkan 1.3° Compute APIs CUDA 11.6, DirectCompute,	VR ready	Yes
Graphics APIs DirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.68°, Vulkan 1.3° Compute APIs CUDA 11.6, DirectCompute,	vGPU software support ⁷	
6.6, OpenGL 4.68°, Vulkan 1.3° Compute APIs CUDA 11.6, DirectCompute,	vGPU profiles supported	See the Virtual GPU Licensing Guide
	Graphics APIs	
	Compute APIs	

