EMBEDDED SOLUTIONS
NVIDIA® RTX MXM MODULES

NVIDIA RTX PERFORMANCE AND FEATURES IN AN MXM FORM FACTOR

NVIDIA® RTX® (Ampere™) MXM modules offer professional NVIDIA RTX performance, features, SDK and API support, exacting build standards, rigorous quality assurance, and broad ISV application compatibility.

Designed for the needs of embedded, ruggedized, or mobile system builders, these Ampere architecture-based products make the most advanced NVIDIA RTX real-time rendering and AI/DL/ML capabilities available to form factors unsuited to traditional PCI Express expansion cards. NVIDIA RTX MXM products offer superb graphics capabilities, outstanding FP32 compute capabilities, powerful Tensor Core TFLOPS for AI, and RT Cores that enable real-time physically-based photorealistic ray tracing. They tolerate wide ranging thermal or other environmental conditions, are ideal for blade or other deployments where high GPU density matters, offer reasonable power requirements, and feature flexible display output options.

From innovative edge AI inferencing or MV, transformational medical imaging, sophisticated signal processing, or synergistic breakthroughs enabled by high-performance graphics, AI, and simulation, NVIDIA RTX MXM solutions let you expand the boundaries of the possible.

THE PNY ADVANTAGE

From innovative edge AI inferencing or MV, transformational medical imaging, sophisticated signal processing, or synergistic breakthroughs enabled by high-performance graphics, AI, and simulation, NVIDIA RTX MXM solutions let you expand the boundaries of the possible.

For additional information or other product inquiries email **MXM@PNY.COM**.

PRODUCT FEATURES	NVIDIA RTX A4500	NVIDIA RTX A2000	NVIDIA RTX A1000	NVIDIA RTX A500
PNY Part Number	NRTXA4500-KIT	NRTXA2000-KIT	NRTXA1000-KIT	NRTXA500-KIT
GPU Architecture	NVIDIA Ampere	NVIDIA Ampere	NVIDIA Ampere	NVIDIA Ampere
Interface	MXM 3.1			
Form Factor	Standard MXM 3.1 Type B	Standard MXM 3.1 Type A		
Dimensions	82 x 105 x 4.8 mm	82 x 70 x 4.8 mm		
Peak FP32	17.8 TFLOPS	9.3 TFLOPS	7.4 TFLOPS	TBD
CUDA Cores	5120	2560	2048	2048
RT Cores	40	20	16	16
Tensor Cores	160	80	64	64
GPU Memory	8 GB | 16 GB	8 GB | 4 GB	4 GB	2 GB | 4 GB
Memory Type	GDDR6			
Memory Interface	256-bit	128-bit	128-bit	64-bit
Memory Bandwidth	384 GB/s	192 GB/s	192 GB/s	96 GB/s
Maximum Power	115W | 80W	60W | 35W		35W | 20W
Operating Temperature	0° C to 55° C | Relative Humidity 5% to 90%			
Storage Temperature	-40° C to 85° C			
Lifecycle Availability	Five Years			
Graphics APIs	DirectX 12 | Model 5.1 | OpenGL 4.6 | Vulkan 1.2			
Compute APIs	CUDA, CUDA-X AI, DirectCompute, OpenCL			
Operating Systems	Windows 11, 10 and Linux Drivers | 64-bit			

SUPPORT

- Pre- and post-sales technical support
- Dedicated NVIDIA Quadro Field Application Engineers
- U.S. based direct NVIDIA Quadro technical support hot line

PNY Technologies, Inc. 100 Jefferson Road, Parsippany, NJ 07054 | Tel 973-515-9700 | Fax 973-560-5590 | WWW.PNY.COM/PNYPRO