

MX3 M.2 AI Accelerator Module

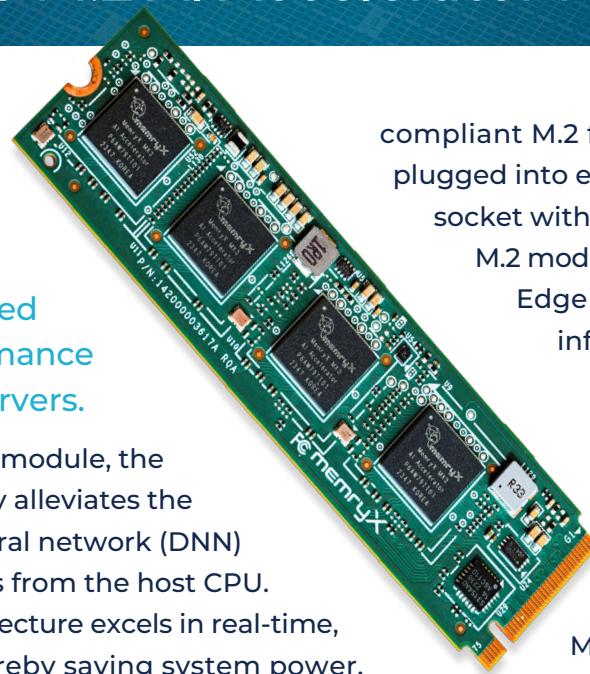
The MemryX MX3 M.2 AI Accelerator Module is a powerful, energy-efficient solution designed to facilitate high-performance AI inference for edge servers.

Engineered as a companion module, the MX3 M.2 Module significantly alleviates the processing load of deep neural network (DNN) computer vision (CV) models from the host CPU. Its distinctive dataflow architecture excels in real-time, low latency inferencing thereby saving system power.

The MX3 M.2 2280 Module is compatible with the industry

compliant M.2 form factor and can be plugged into existing systems with an M.2 socket with no hardware changes. Each M.2 module contains four MemryX MX3 Edge AI Accelerators to execute inferencing in real-time for a broad range of applications and market segments.

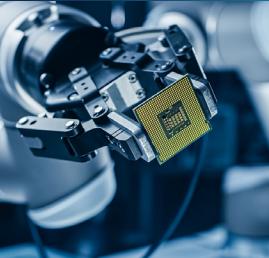
Designed to meet the demands of modern AI applications, the MemryX M.2 AI Accelerator Module sets new benchmarks in performance, power efficiency, and adaptability.



Key Benefits

<ul style="list-style-type: none"> • Supports all common frameworks • Dataflow architecture for ultra-low latency • Advanced power management features 	<ul style="list-style-type: none"> • Up to 80 million weight parameters • On-chip storage of model parameters and matrix operators eliminates the need for external DRAM • 2/4-lane PCIe Gen3 for up to 4GB/s bandwidth 	<ul style="list-style-type: none"> • Supports multiple concurrent models • Floating-point activations for high accuracy • End user models can be deployed as-is without quantization, pruning, compression, or retraining
---	--	--

Markets

Industrial 4.0 & Robotics	Automotive	IoT	Metaverse	Smart Vision Systems	Computing Devices
					

Technical Specifications

Mechanical	Form Factor	NGFF M.2 2280 M Key Socket 3	
	Dimensions	3.15" x 0.87" (22 x 80 mm)	
	Input Voltage	3.3V +/- 5%	
Electrical	Power	Typical: 8W Maximum: 14W (M.2 limit)	
	Interface	PCIe Gen 3, 2x2-lanes USB Gen 3 x2	8 GT/s per PCIe lane
	IC	4 MemryX MX3's	
	Frameworks	ONNX, PyTorch, TensorFlow, Keras, TensorFlow Lite	
System	Weight parameters	80 million, 4-bit 40 million, 8-bit	
	Host Processor	Any including ARM, x86, and RISC-V	
	OS Support	Ubuntu 18.04, 64-bit Windows, 64-bit Android	
	SDK	MemryX Developer Hub	
Environmental	Certification	CE / FCC Class A	
	Operating Temperature	-40°C to 85°C Industrial	
	Operating Humidity	10% ~80% RH non-condensing	

Ordering Information

Part #	Description
MX3-2280-M-4	4-chip M.2 module, 22x80 mm, M-Key
MX3-2280-M-2	2-chip M.2 module, 22x80 mm, M-Key

Model as publicly available with no pruning, compression, or re-training. Performance is AI core and can vary based on host interface.

For More information, go to MemryX.com

1600 Huron Parkway, 2nd Floor, Ann Arbor, MI 48109, USA