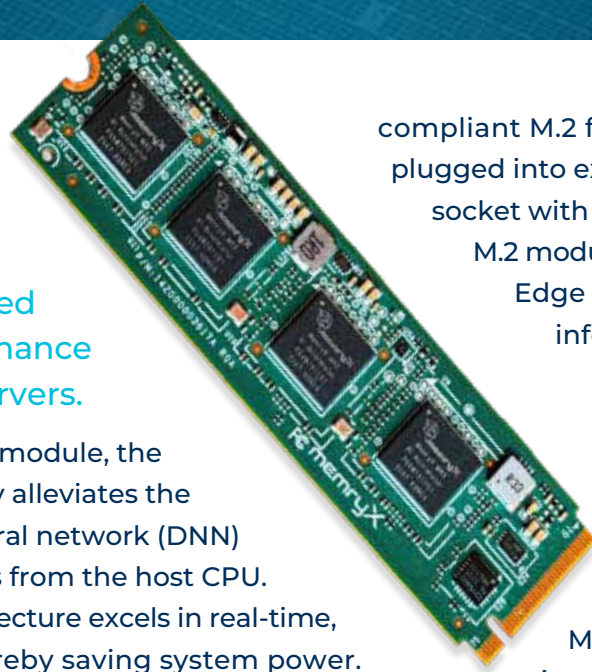


# MX3 M.2 AI Accelerator Module

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

- Supports all common frameworks
- Dataflow architecture for ultra-low latency
- Advanced power management features
- 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
- 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)	
Electrical	Input Voltage	3.3V +/- 5%	
	Power	Typical: 8W Maximum: 14W (M.2 limit)	
	Interface	PCIe Gen 3, 2x2-lanes USB Gen 3 x2	8 GT/s per PCIe lane
System	IC	4 MemryX MX3's	
	Frameworks	ONNX, PyTorch, TensorFlow, Keras, TensorFlow Lite	
	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	
	Certification	CE / FCC Class A	
	Operating Temperature	0°C - 70°C Commercial -40°C - 85°C Industrial	
Environmental	Operating Humidity	10% ~80% RH non-condensing	

## Ordering Information

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

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

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



MemryX, Inc. is an AI semiconductor startup headquartered in Ann Arbor, MI USA with branches in Taipei and Hsinchu, Taiwan. We develop a highly scalable and innovated AI accelerator that offers high performance, low power, and customer ease of implementation for embedded Edge AI vision-based applications and real-time processing.

Information contained is subject to change without notice. MemryX, Inc, logo, MX3, and other designated brands included herein are registered trademarks in the US and other countries

© 2024 MemryX Inc., All Rights Reserved