MX3 M.2 AI Accelerator

Quick Start Guide

Rev 1.0

Revision History

Date	Version	Revision
2024.10/3	1.0	Initial release.

Table of Contents

1.0 Introduction

Thank you for our interest in MemryX's M.2 AI Accelerator Module ("Module") consisting of 4 custom MX3 AI processing ICs. The Module can perform AI inferencing out-of-the-box on hundreds of AI common models, offloading AI processing tasks from the Host's processor.

SYSTEM	
AI Processor	MemryX MX3 (x4)
Host Processor Support	ARM, x86, RISC-V
ELECTRICAL	
Input Voltage	3.3V +/- 5%
Interface	PCIe Gen 3, 2 x 2-lanes
MECHANICAL	
Form Factor	NGFF M.2-2280-D5-M, Socket 3
Dimensions	3.15" x 0.87" (22 x 80 mm)
ENVIRONMENTAL	
Operating Temperature	0°-70° C
COMPLIANCE	
Certification	CE / FCC Class A, RoHS

Table 1. M.2 MODULE Key Specifications

This guide, along with the information provided at https://developer.memryx.com/docs/ , is intended to instruct the user on how to install, set up, and activate the Module in a PC, Server, or Laptop (Host Platform).

2.0 Package Contents

The Module comes in a recyclable cardboard box. Package contents include:

1. M.2 AI Accelerator Module (x1)
2. Heatsink (x1)
3. Bottom Bracket (x1)
4. Heatsink Mounting Screws (x4)
5. M.2 Mounting Screws (x2)

3.0 Host Requirements

The Module is designed to plug into any available Host platform M.2 2280 M-key MVME socket. Space permitting, it is highly recommended to assemble the Module with the provided custom heatsink for improved thermal dissipation and reliability.

In addition, MemryX's Software Development Tools runs on the OS and CPU combinations as shown in the Table below. MemryX's SDK consisting of Compiler, Simulator, and Accl/Benchmark tools are required to take a user's trained model and port it to the Module. The software Driver and RT Tool must be installed on the Platform used for inferencing. Note that the Platform used for compiling/simulating/and deploying the model to the Module can be different than the platform used for running the AI model while inferencing.

Component	Linux x86_64	Linux ARM 64	Windows (native)	Windows (WSL)
Tools				
└── Compiler	✅	✅	❌	✅
└── Simulator	✅	❌	❌	✅
└── Accl/Benchmark	✅	✅	❌	❌
Driver & C++ Runtime	✅	✅	✅	❌

MemryX's Driver and Run Time tool are required to be deployed into the Host processor that will be used for inferencing.

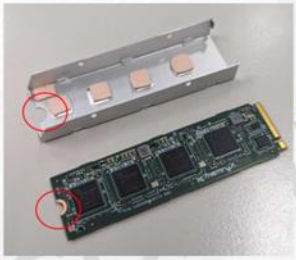
4.0 Assembly of Heatsink

Due to their compact size, and typical heat dissipation, M.2 Modules can get hot. It is strongly recommended that, space permitting, the supplied heatsink and bottom bracket be used. Assembly of the Module, heatsink and bracket must occur prior to installing into the Host Platform.

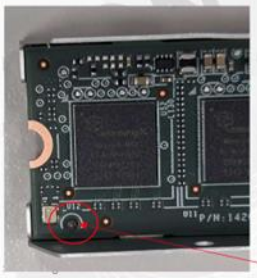
Step1: Remove blue film on mount base

Step2: Install PCBA

The cut on same side



Golden finger side need to through the hook first.



The pin will go through the hole

Step3: Remove blue film on heatsink and then install onto PCBA

Logo need to toward golden finger

Step4: Fasten screws x4

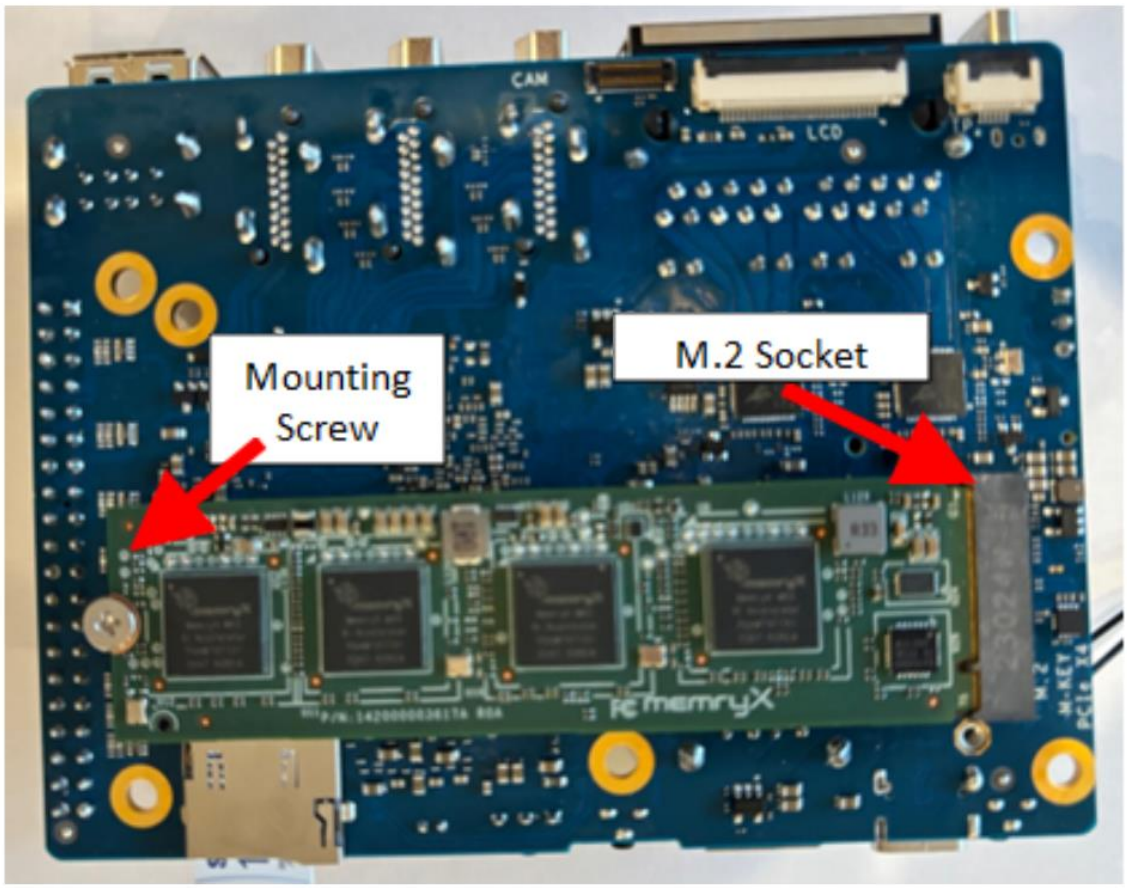
Screws x2 each side

Heatsink align the hook of mount base

5.0 Hardware Installation

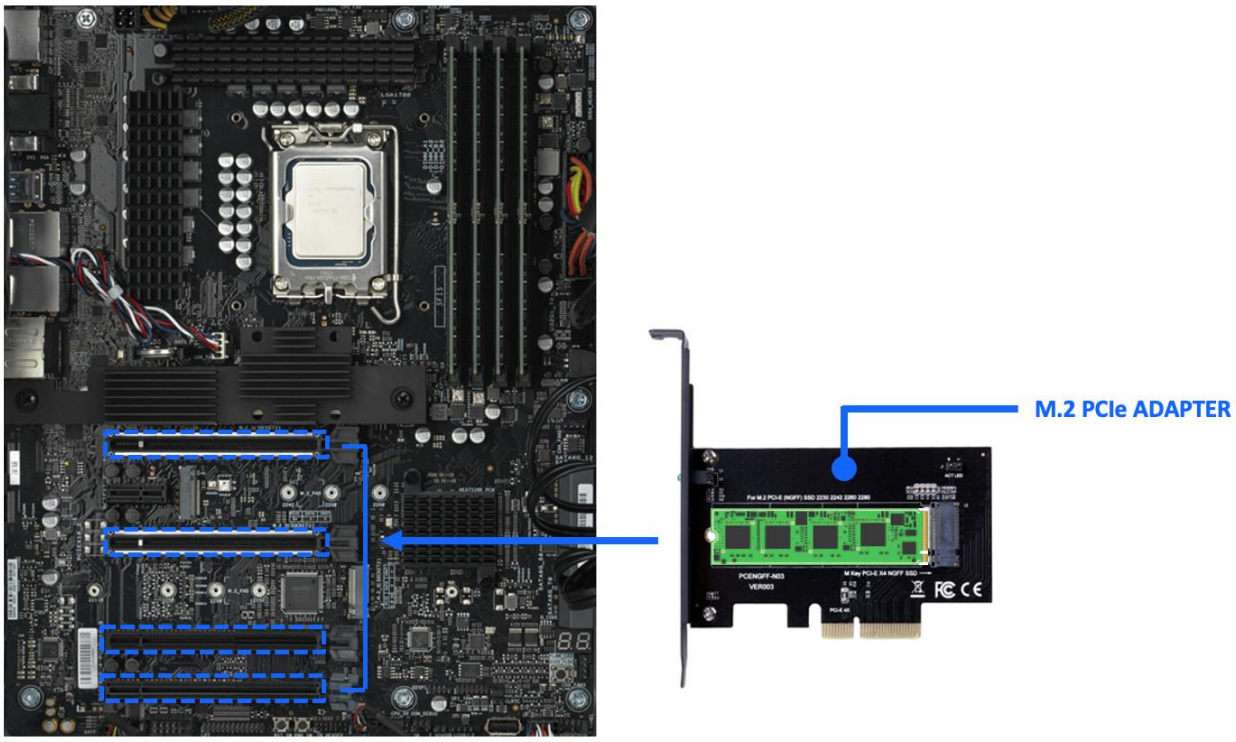
5.1 Host Platform with M.2 Socket

The figure below shows the M.2 Module without heatsink installed onto an Orange Pi. It is important to ensure the “gold-finger” traces visible on one side of the Module are fully inserted into the M.2 Socket for proper contact. On the opposite side away, from the M.2 Socket, a mounting screw is also required to keep the Module from dislodging from the M.2 Socket.



5.1 Host Platform with PCIe Socket

For systems without an M.2 socket but having a PCIE x8 or x 16 socket, an M.2 PCIe adapter card can be used as shown below.



6.0 Detect Module

Now that the Module has been installed, boot the system, and use this command to verify the module is recognized:

```
$ lspci -d -d 1fe9:
 0000:01:00.0 Processing accelerators: Device 1fe9:0100
```

7.0 SDK and Driver Installation

Now that the Module has been installed into the Host Platform, it is time to install the SDK and the driver.

SDK and tool install instructions are here:
https://developer.memryx.com/docs/get_started/install_tools.html

Driver installation instructions are here:
https://developer.memryx.com/docs/get_started/install_driver.html

8.0 Communication Verification

To verify the inferencing Host Platform and Module are communicating, using Terminal perform one of the following checks depending on the Host OS installed.

8.1 Linux:

Linux Windows

```
ls /dev/memx*
```

should return:

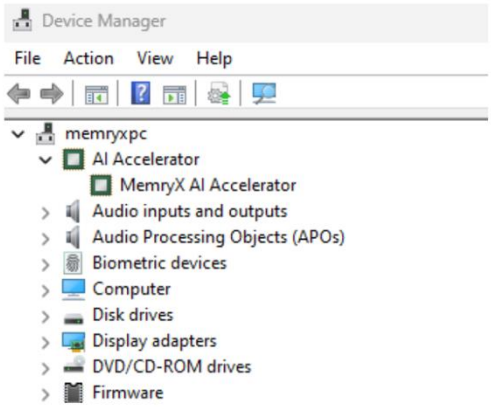
```
/dev/memx0
```

8.2 Windows:

Linux **Windows**

Open Device Manager.

At the top of the list there should be an entry for "MemryX AI Accelerator" with a green icon.



9.0 Hello MXA!

Now that the Module and Host are communicating, we can run an AI model on and view the Module performance (FPS and latency).
Please follow these instructions: https://developer.memryx.com/docs/get_started/hello_mxa.html

10.0 Thermal

As with most all M.2 modules, heat dissipation is of utmost concern. Use of one or more fans (i.e. convection cooling) and heatsink are recommended for optimal thermal dissipation. For fan-less systems it is recommended to provide a low thermal resistance conductive path for heat to flow from the top of the MX3 ICs to the outside of the system. Please refer to MemryX's M.2 Thermal Design Guide for further information.

The table below lists typical Thermal Dissipated Power (TDP) for high, medium, and low power AI models and the airflow and heatsink requirements to operate the Module at various ambient temperatures. Note that even a minimum airflow can significantly improve heat dissipation. For fan-less systems every situation is different as will be the optimum thermal solution (e.g. heat pipe).

Cases	1	2	3	4
Condition	Worst	Normal	Low Power	Low Power
System TDP	9.27W	7.7W	7.7W	5.15W
Ambient Temp	70C	70C	50C	25C
Heatsink	Yes	Yes	Yes	No
Airflow Requirement (Min)	3 CFM	1 CFM	1 CFM	1 CFM

10.1 Polling MX3 IC temperature

The user can query the internal chip temperature of each of the four MX3 IC's using the following terminal command. For maintain high performance and reliability, the chip temperature should not exceed 85°C.

```
cat /sys/memx0/temperature
```

11.0 Next Steps

Now that the M.2 Module and Host are connected, refer to MemryX's Developer Hub for examples of running the compiler, simulator, and the M.2 module.

Please refer to the following URL for tutorials and examples to become familiar with MemryX's extensive tools and capabilities: https://developer.memryx.com/docs/tutorials/tutorials.html

12.0 Contact Us

If you have any questions or encounter any issues, please reach out to us at www.memryx.com or contact your sales representative. Additionally, we welcome feedback you may have on usability, performance, or overall experience using these samples.