Coral

M.2 Accelerator Datasheet

Version 1.0

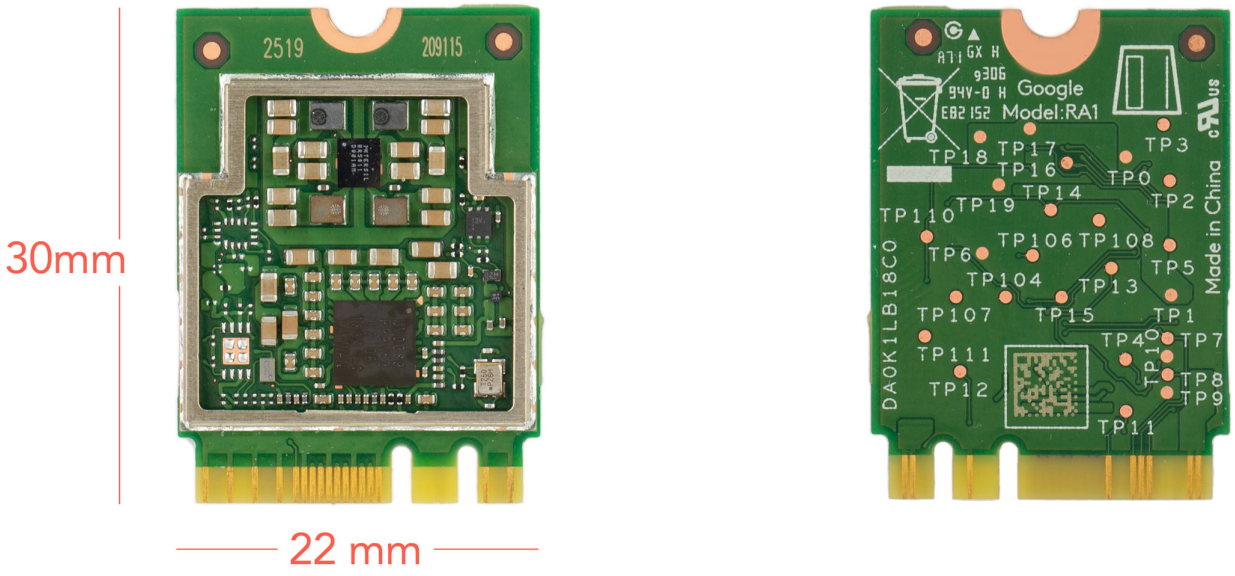

Table of contents

Features

- Google Edge TPU ML accelerator
- Available in two M.2 form factors:
  - M.2-2230-A-E-S3 (A+E key)
  - M.2-2280-B-M-S3 (B+M key)
- Supports Debian Linux and other variants on host CPU

Photo shows A+E key form factor with shield can removed

Overview

The Coral M.2 Accelerator is an M.2 module that brings the Edge TPU coprocessor to existing systems and products.

The Edge TPU is a small ASIC designed by Google that provides high performance ML inferencing with low power requirements: it's capable of performing 4 trillion operations (tera-operations) per second (TOPS), using 0.5 watts for each TOPS (2 TOPS per watt). For example, it can execute state-of-the-art mobile vision models such as MobileNet v2 at 400 FPS, in a power efficient manner. This on-device processing reduces latency, increases data privacy, and removes the need for constant high-bandwidth connectivity.

The M.2 Accelerator is a dual-key M.2 card (either A+E or B+M keys), designed to fit any compatible M.2 slot. This form-factor enables easy integration into ARM and x86 platforms so you can add local ML acceleration to products such as embedded platforms, mini-PCs, and industrial gateways.

Requirements

The Coral M.2 Accelerator must be connected to a host computer with the following specifications:

- Any Linux computer with a compatible M.2 module slot
  - Debian 6.0 or higher, or any derivative thereof (such as Ubuntu 10.0+)
  - System architecture of either x86-64 or ARM32/64 with ARMv8 instruction set

For software required on the host, see the **software and operation section**.

Specifications

The design of the M.2 Accelerator adheres to the PCI-SIG's specification for the PCI Express M.2. For in-depth mechanical details, refer to that specification.

Table 1. M.2 Accelerator technical specs

Physical specifications	
Dimensions	A+E key: 22 x 30 x 2.3 mm B+M key: 22 x 80 x 2.3 mm
Weight	A+E key: 3.1 g B+M key: 5.8 g
Host interface	
Hardware interface	M.2 A+E key (M.2-2230-A-E-S3) or M.2 B+M key (M.2-2280-B-M-S3)
Serial interface	PCIe Gen2 x1
Operating voltage	
DC supply	3.3V +/- 10 %
Environmental reliability	
Temperature [1]	-40 ~ 85° C (storage) -20 ~ 70° C (operating)
Relative humidity	0 ~ 100% (non-condensing)
Mechanical reliability	
Op-shock	100 G, 11ms (persistent) 1000 G, 0.5 ms (stress) 1000 G, 1.0 ms (stress)
Op-vibe (random)	0.5 Grms, 5 - 500 Hz (persistent) 3 Grms, 5 - 800 Hz (stress)
Op-vibe (sinusoidal)	0.5 Grms, 5 - 500 Hz (persistent) 3 Grms, 5 - 800 Hz (stress)
Compliance	
Countries[2]	Unit shipped as component. Certification/compliance to be done by customer.
ESD [3]	1kV HBM, 250V CDM

*[1] Operational temperature range depends on the **power consumption** and **thermal management** in your system.*
[2] We can provide certification example to demonstrate that a reasonably designed system meets certification requirements.
[3] Always handle in static safe environment.

Dimensions

A+E key dimensions

- Total size: 22 x 30 x 2.3 mm
- PCB size: 22 x 30 x 0.8 mm (M.2 TYPE 2230-S3-A-E)
- Top-side component height (max): 1.5 mm (S3 class)
- Bottom-side component height: 0 mm

For in-depth mechanical specs, refer to the PCI Express M.2 Specification.

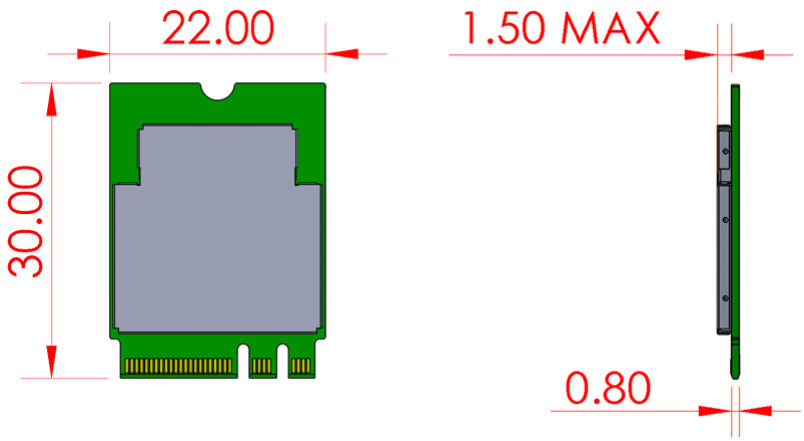

Figure 1. M.2 A+E key dimensions (in millimeters)

B+M key dimensions

- Total size: 22 x 80 x 2.3 mm
- PCB size: 22 x 80 x 0.8 mm (M.2 TYPE 2280-S3-B-M)
- Top-side component height (max): 1.5 mm (S3 class)
- Bottom-side component height: 0 mm

For in-depth mechanical specs, refer to the PCI Express M.2 Specification.

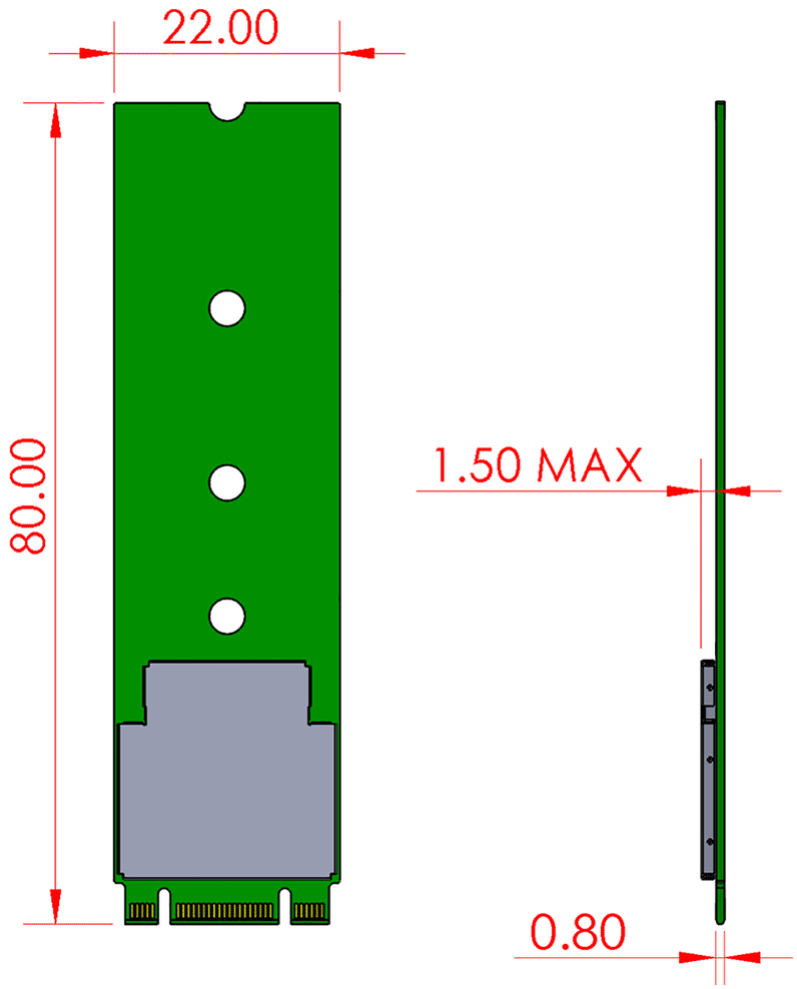

Figure 2. M.2 B+M key dimensions (in millimeters)

Power specifications

The M.2 Accelerator is powered by 3.3V as shown in the **connector pinout**. Typical power consumption depends on the model architecture and operating parameters, but some sample power consumption is shown in table 2. See the below section for information about the different **operating frequencies**.

Table 2. M.2 Accelerator typical power consumption

	Low operating frequency	Nominal operating frequency	Max operating frequency
MobileNet v2	0.6 W (7.1 ms @ 141 fps)	0.9 W (3.9 ms @ 256 fps)	1.4 W (2.4 ms @ 416 fps)
Inception v3	0.5 W (58.7 ms @ 17 fps)	0.6 W (51.7 ms @ 19.3 fps)	0.7 W (48.2 ms @ 20.7 fps)

Thermal limit and operating frequency

The thermal resistance and max allowed temperature of the Edge TPU stack-up is as follows:

- **Thermal resistance (junction to top of shield can)**: 11 °C/W
- **Maximum Edge TPU junction temperature**: 100 °C

The M.2 Accelerator does not include a thermal solution to dissipate heat from the system. In order to sustain maximum performance from the Edge TPU, it's important that you design your system so the Edge TPU operates well below the maximum Edge TPU temperature. If the Edge TPU gets too hot, it slowly reduces the operating frequency and may reset to avoid permanent damage.

The PCIe driver includes a power throttling mechanism (also known as dynamic frequency scaling) and an emergency shutdown mechanism, based on temperature readings from the Edge TPU. By default, this system checks the Edge TPU die temperature every 5 seconds and responds as follows:

- If the Edge TPU is below 85°C, continue at the "maximum" operating frequency.
- If the Edge TPU reaches 85°C, reduce the operating frequency 50% (from "maximum" to "normal").
- If the Edge TPU reaches 90°C, reduce the operating frequency another 50% (from "normal" to "low").
- If the Edge TPU reaches 95°C, reduce the operating frequency yet another 50% (from "low" to "lowest").
- If the Edge TPU reaches 100°C, reset the Edge TPU.

By reducing the operating frequency, the Edge TPU's inferencing speed becomes slower, but it also consumes less power and hopefully avoids reaching the hardware reset threshold.

As long as the Edge TPU does not reset and the Edge TPU temperature returns to lower levels, the system restores the operating frequency in the reverse manner—ultimately returning to the maximum operating frequency when the Edge TPU is below 85°C.

Connector pinout

A+E key pinout

Table 3. M.2 Accelerator A+E key pinout

Top side pins		Bottom side pins	
Signal	Pin	Pin	Signal
GND	75	74	3.3V
NC	73	72	3.3V
NC	71	70	NC
GND	69	68	NC
NC	67	66	NC
NC	65	64	NC
GND	63	62	NC
NC	61	60	NC
NC	59	58	NC
GND	57	56	NC
NC	55	54	NC
CLKREQ0# (3.3V)	53	52	PERST0# (3.3V)
GND	51	50	NC
REFCLKn0	49	48	NC
REFCLKp0	47	46	NC
GND	45	44	NC
PETn0	43	42	NC
PETp0	41	40	NC
GND	39	38	NC
PERn0	37	36	NC

Top side pins		Bottom side pins	
Signal	Pin	Pin	Signal
PERp0	35	34	NC
GND	33	32	NC
Key E slot	31	30	Key E slot
Key E slot	29	28	Key E slot
Key E slot	27	26	Key E slot
Key E slot	25	24	Key E slot
NC	23	22	NC
NC	21	20	NC
NC	19	18	GND
NC	17	16	NC
Key A slot	15	14	Key A slot
Key A slot	13	12	Key A slot
Key A slot	11	10	Key A slot
Key A slot	9	8	Key A slot
GND	7	6	NC
NC	5	4	3.3V
NC	3	2	3.3V
GND	1		

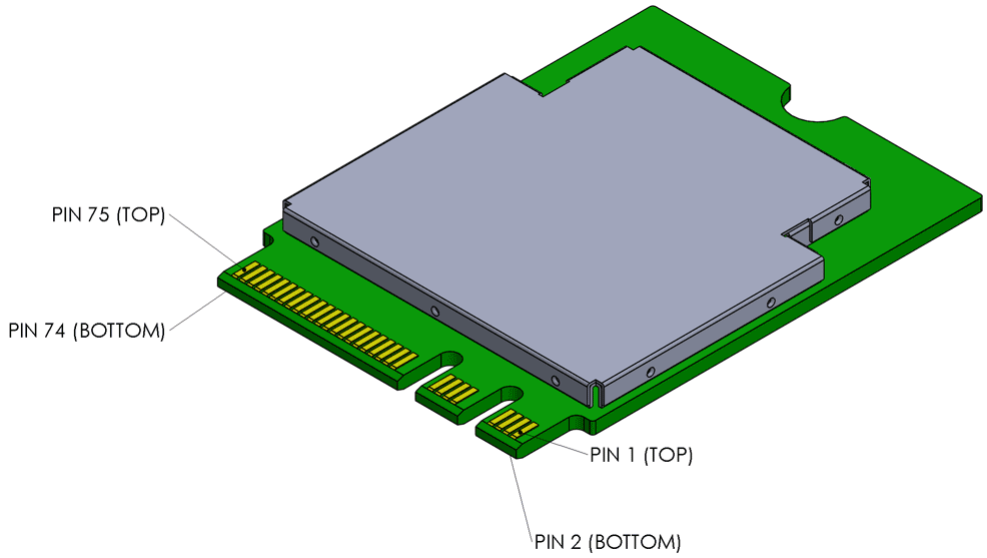

Figure 3. M.2 A+E key pin positions

B+M key pinout

Table 4. M.2 Accelerator B+M key pinout

Top side pins		Bottom side pins	
Signal	Pin	Pin	Signal
GND	75	74	3.3V
GND	73	72	3.3V
GND	71	70	3.3V
NC	69	68	NC
NC	67	66	Key M slot
Key M slot	65	64	Key M slot
Key M slot	63	62	Key M slot
Key M slot	61	60	Key M slot
Key M slot	59	58	NC
GND	57	56	NC
REFCLKp0	55	54	NC
REFCLKn0	53	52	CLKREQ0# (3.3V)
GND	51	50	PERST0# (3.3V)

Top side pins		Bottom side pins	
Signal	Pin	Pin	Signal
PERp0	49	48	NC
PERn0	47	46	NC
GND	45	44	NC
PETp0	43	42	NC
PETn0	41	40	NC
GND	39	38	NC
NC	37	36	NC
NC	35	34	NC
GND	33	32	NC
NC	31	30	NC
NC	29	28	NC
GND	27	26	NC
NC	25	24	NC
NC	23	22	NC
GND	21	20	NC
Key B slot	19	18	Key B slot
Key B slot	17	16	Key B slot
Key B slot	15	14	Key B slot
Key B slot	13	12	Key B slot
NC	11	10	NC
NC	9	8	NC
NC	7	6	NC
NC	5	4	3.3V
GND	3	2	3.3V
GND	1		

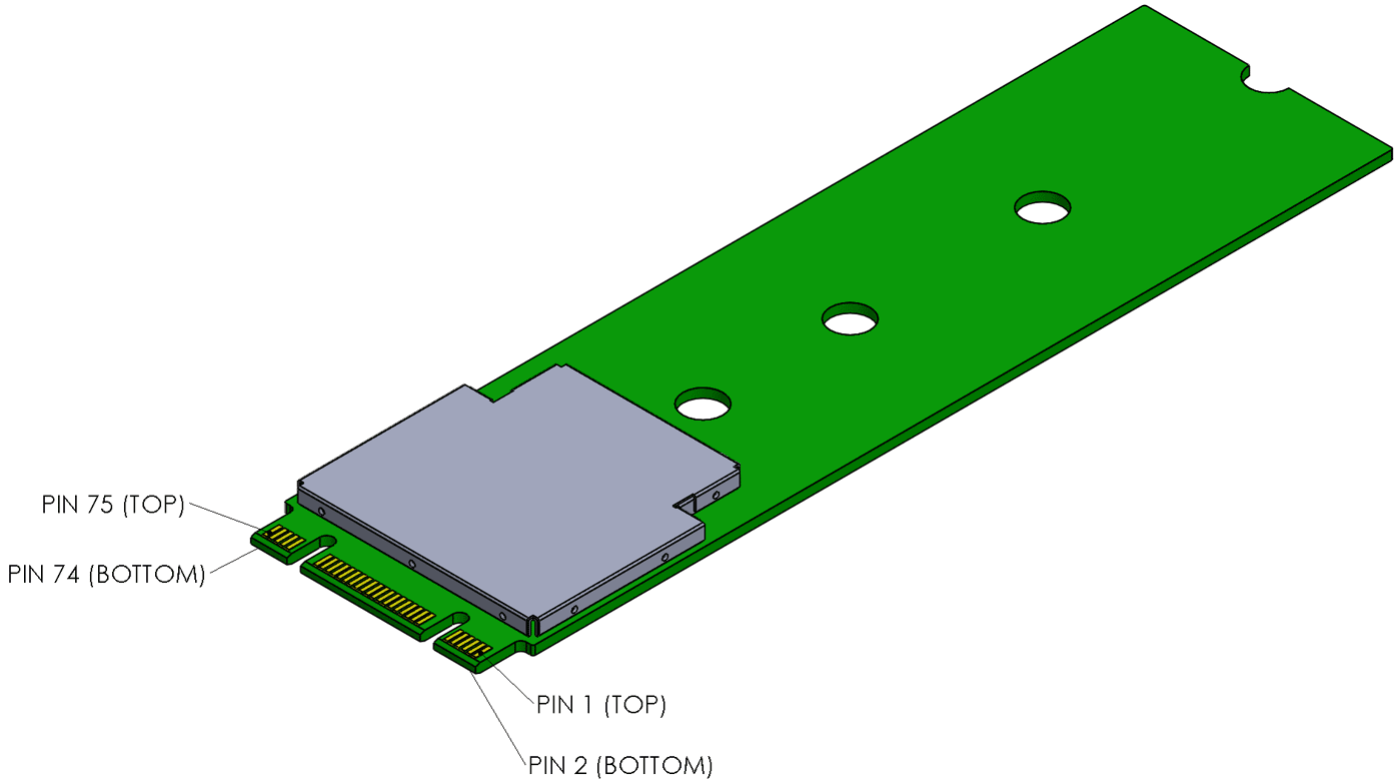

Figure 4. M.2 B+M key pin positions

Software and operation

The host system must be running Debian Linux 6.0 or higher, or any derivative thereof, and have the Edge TPU runtime and API library installed.

The PCIe kernel driver is already upstreamed to kernel.org with source, since version 4.19. For earlier versions, dkms driver is available via gasket-dkms deb package at `https://packages.cloud.google.com/apt coral-edgetpu-stable main`.

To learn how to create models and run inferences the Edge TPU, read **TensorFlow models on the Edge TPU**.