

Coral

Mini PCIe Accelerator Datasheet

Version 1.2

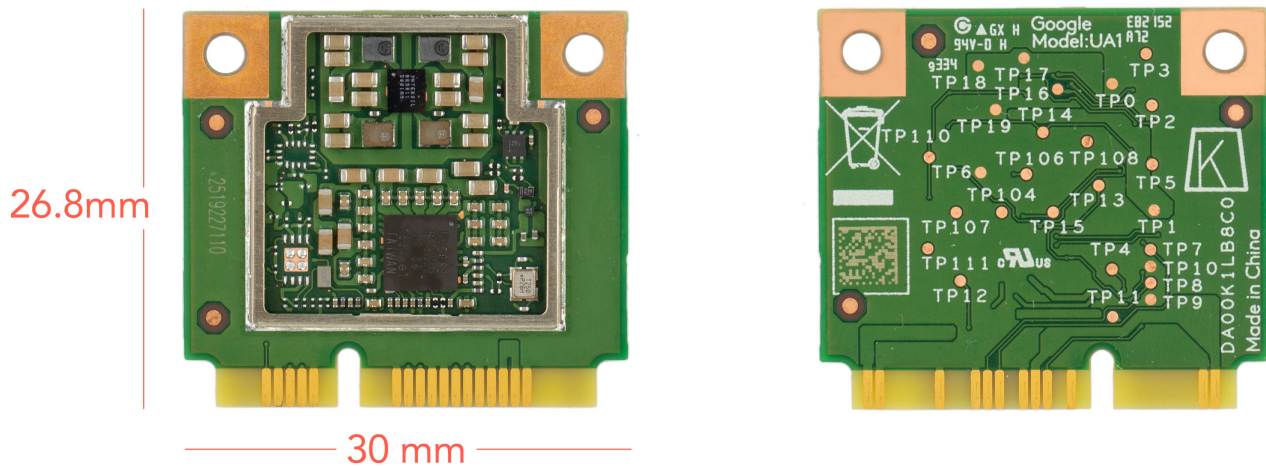


Photo shows the card with shield can removed

Version 1.2 (December 2019)

Copyright 2020 Google LLC. All rights reserved.

Features

- Google Edge TPU ML accelerator
- Standard Half-Mini PCIe card
- Supports Debian Linux and other variants on host CPU

Overview

The Coral Mini PCIe Accelerator is a PCIe 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 almost 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 Mini PCIe Accelerator is a half-size Mini PCIe card designed to fit in any standard Mini PCIe 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.

Table of contents

- Requirements
- Specifications
- Dimensions
- Power specifications
- Thermal limit and operating frequency
- Connector pinout
- Software and operation
- Document revisions

Requirements

The Coral Mini PCIe Accelerator must be connected to a host computer with the following specifications:

- Any Linux computer with a compatible Mini PCIe 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 Mini PCIe Accelerator adheres to the PCI-SIG's electromechanical specification for the PCI Express Mini Card. For in-depth mechanical details, refer to that specification.

Table 1. Mini PCIe Accelerator technical specs

Physical specifications	
Dimensions	30.00 x 26.80 x 2.55 mm
Weight	3.6 g
Host interface	
Hardware interface	Half-Mini PCIe card
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)

Physical specifications	
Dimensions	30.00 x 26.80 x 2.55 mm
Weight	3.6 g
Host interface	
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

- PCB width: 30.00 mm ± 0.15 mm
- PCB height: 26.80 mm ± 0.15 mm
- PCB thickness: 1.00 mm ± 0.05 mm
- Top-side component height: 1.55 mm ± 0.10 mm
- Bottom-side component height: 0 mm

For in-depth mechanical specs, refer to the PCI Express Mini Card Electromechanical Specification.

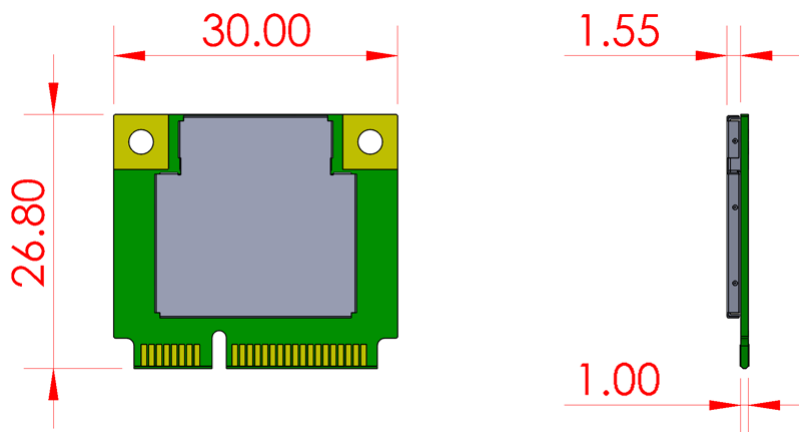


Figure 1. Mini PCIe Accelerator dimensions (in millimeters)

Power specifications

- DC supply: 3.3 V (see [connector pinout](#))
- Max power consumption: 4 W (host should limit power)

Typical power consumption depends on the model architecture and operating parameters, but some sample power consumption is shown in table 2 (based on different [operating frequencies](#)).

Table 2. Mini PCIe 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 Mini PCIe 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

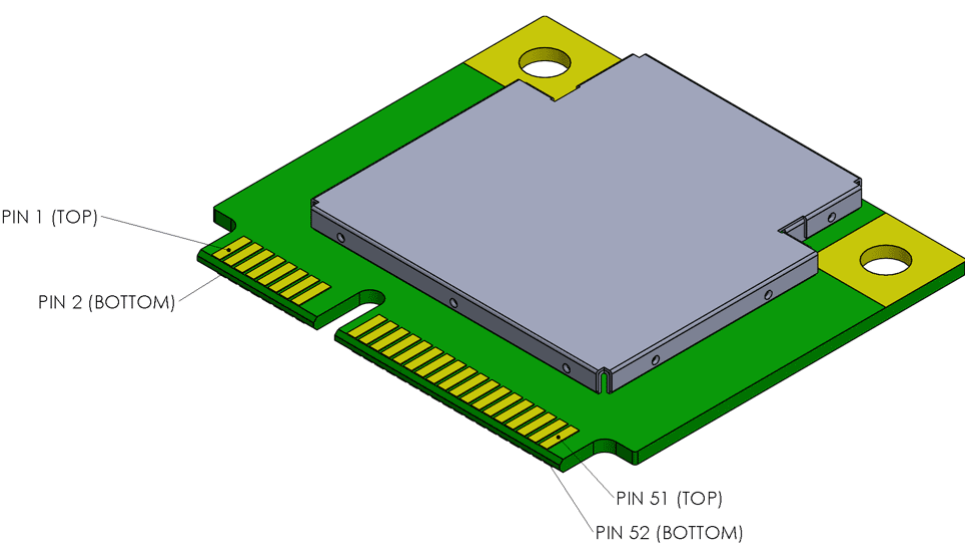


Figure 2. Mini PCIe Accelerator pin positions

Table 3. Mini PCIe Accelerator connector pinout

Top side pins		Bottom side pins	
Signal	Pin	Pin	Signal
NC	1	2	3.3V
NC	3	4	GND

Version 1.2 (December 2019)

Top side pins		Bottom side pins	
Signal	Pin	Pin	Signal
NC	5	6	NC
CLKREQ# (3.3V)	7	8	NC
GND	9	10	NC
REFCLK-	11	12	NC
REFCLK+	13	14	NC
GND	15	16	NC
Key slot			
NC	17	18	GND
NC	19	20	NC
GND	21	22	PERST# (3.3V)
PERn0	23	24	3.3V
PERp0	25	26	GND
GND	27	28	NC
GND	29	30	NC
PETn0	31	32	NC
PETp0	33	34	GND
GND	35	36	NC
GND	37	38	NC
3.3V	39	40	GND
3.3V	41	42	NC
GND	43	44	NC
NC	45	46	NC
NC	47	48	NC
NC	49	50	GND
NC	51	52	3.3V

Version 1.2 (December 2019)

Copyright 2020 Google LLC. All rights reserved.

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](#).

Document revisions

Table 4. History of changes to this document

Version	Changes
1.2 (December 2019)	Revised dimensions and added tolerances
1.1 (October 2019)	Added max power consumption
1.0 (August 2019)	Initial release