Mustang-T100-T5

Feature

- 5 x Coral Edge TPU™ ML accelerator
- 20 TOPS peak performance (int8)
- Host interface PCIe Gen2 x4
- Low-profile PCIe form factor
- Support Multiple card
- Approximate 15W
- RoHS compliants

Specifications

Model Name	Mustang-T100-T5
Main Chip	Five Coral Edge TPU™ Accelerator Module
Operating Systems	Linux: 64-bit version of Debian 10 or Ubuntu 16.04 (or newer)
	Windows: 64-bit version of Windows 10
Dataplane Interface	PCI Express Gen2 x4
Power Consumption	Approximate 15W
Operating Temperature	-20°C~55°C
Cooling Solution	Active
Dimensions	Standard half-height, half-length, single-slot PCIe card
Dip Switch/LED indicator	Identify card number
Support Framwork	Tensorflow Lite
Precision	INT8

Ordering Information

Part No.	Description
Mustang-T100-T5-R10	TPU Accelerator Card with 5 x Coral edge TPU, PCIe Gen2 x4 interface, RoHS

Packing List

1 x Full height bracket
1 x QIG

Scalable Infrastructure, Support Multiple Cards

You can install up to eight Mustang-T100 AI accelerator cards in one system to support additional AI workload and expand the AI computing capabilities in any requirement.

x8

Multitasking or Pipelining, Select Your Inferencing Mode

For numerous AI applications at the edge, clients can select from two different modes to run your inferencing project depending on their needs.

Multitasking function to run each model in parallel

If you need to run multiple models, you can assign each model to a specific Edge TPU and run them in parallel at the same time for extreme computing performance.
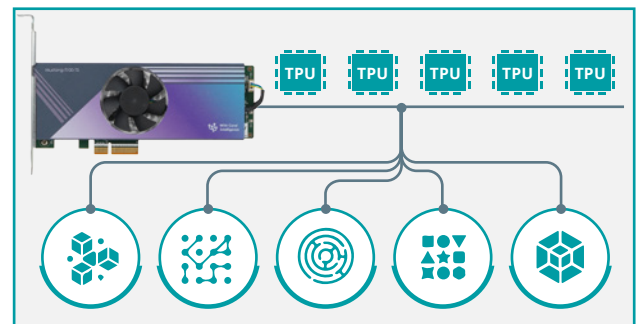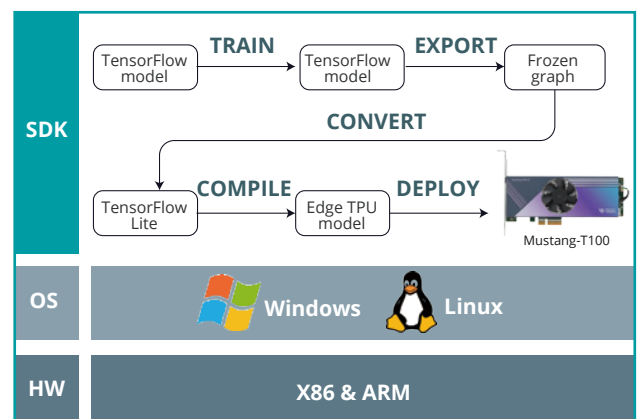
Model pipelining to get faster throughput and low latency

For other scenarios that require very fast throughput or large models, pipelining your model allows you to execute different segments of the same model on different Edge TPUs. This can improve throughput for high-speed applications and can reduce total latency for large models.
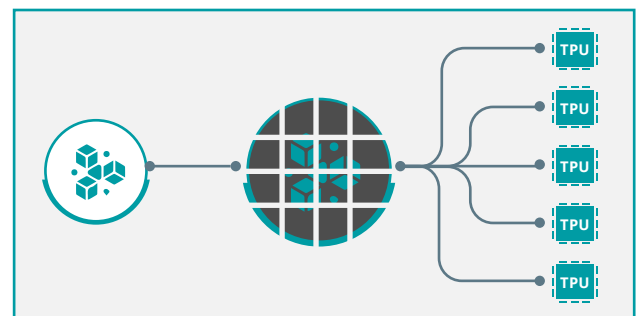
High Compatibility From The Start

To support diverse needs from IT or AI developers, the Mustang-T100 can be implemented in various operating systems, such as Linux, X86, ARM, or small FF to accelerate and maximum edge AI performance. More, combined with TensorFlow Lite, no need to build ML training models from the ground up. TensorFlow Lite models can be compiled models to run on the Edge TPU completely.

SDK

TensorFlow model → **TRAIN** → TensorFlow model → **EXPORT** → Frozen graph

CONVERT

TensorFlow Lite → **COMPILE** → Edge TPU model → **DEPLOY** → Mustang-T100

OS — Windows   Linux

HW — X86 & ARM

Accelerator Module datasheet

- Coral Edge TPU™ ML accelerator: 4 TOPS peak performance (int8) / 2 TOPS per watt
- Integrated power management
- PCIe Gen2 x1 or USB 2.0 interface
- Surface-mounted (LGA) module
- Size: 15.0 x 10.0 x 1.5 mm
- **Support ARM (Linux) & X86(Windows & Linux)**
- Weight: 0.67 g
- Operating temp: -20°C~55°C
- RoHS compliant
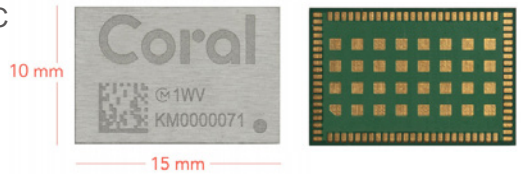
Introduction

Coral is a hardware and software platform for building intelligent devices with fast neural network inferencing.

At the heart of our devices is the Edge TPU coprocessor. This is a small ASIC built by Google that's specially-designed to execute state-of-the-art neural networks at high speed, with a low power cost.

The Edge TPU is capable of performing 4 trillion operations (tera-operations) per second (TOPS), using 0.5 watts for each TOPS (2 TOPS per watt).

Mustang−T100−T5 Block Diagram
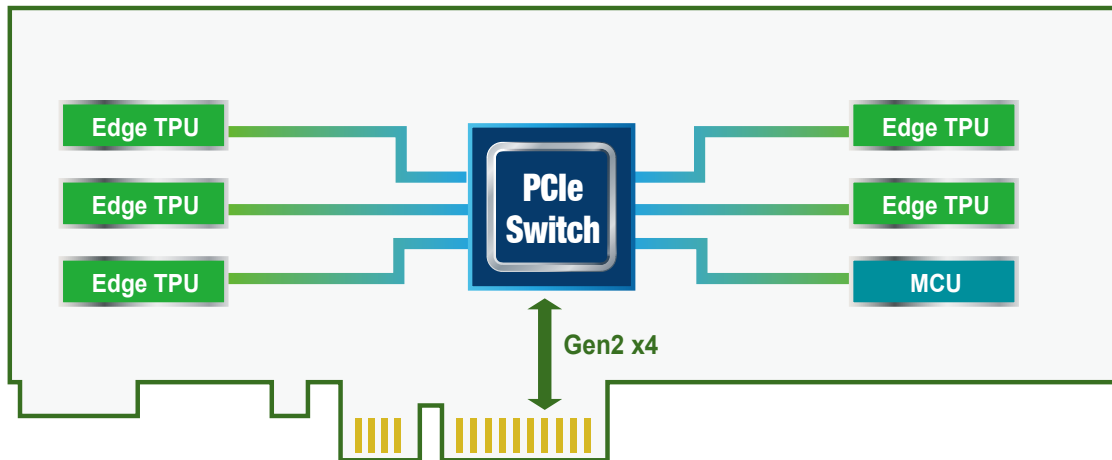
Mustang−T100−T5 Dimensions (Unit: mm)