



Key Features

500 TOPS
Inference
Performance

Low-Profile
PCIe card

40W Typical
Power

204 MB of on-
chip SRAM

20,000 fps
ResNet-50 V1.5

Low Latency,
batch=1

Overview

The tsunAI^{mi} tsn200 accelerator card for neural network inference is bringing supercomputing to the edge in a small and power efficient form factor, with over 500 TOPS of INT8 performance in a half-height, half-length, low profile PCIe card. The power efficiency of the onboard runAI200™ device enables a 75W thermal design point (TDP), with typical application power of 40W. This makes tsunAI^{mi} tsn200 the most efficient edge computing card in its class, powered entirely by the PCIe edge connector. The x16 PCI-Express Gen4 interface supports up to 32 GB/s of bandwidth.

Applications

The tsunAI^{mi} tsn200 accelerator cards are designed for high performance, power sensitive edge applications like video analytics and signals processing. The low-profile form factor enables easy installation into small form factor computers and edge servers alike.

Markets	Application	Networks
Vision	Classification, object detection, semantic segmentation	ResNets, YOLO, SSD, Unets
High Performance Compute	Radar, signal processing, deep packet inspection	BLAS, FFTs, TCNs
Computational Storage	Analytics acceleration, processing in memory	CNNs, RNNs, MLPs, etc...

imAIne Software Development Kit

The imAIne SDK gives developers powerful automated tools and supporting software to quickly go from pilot model to production. It is organized into three parts.

The imAIne Compiler

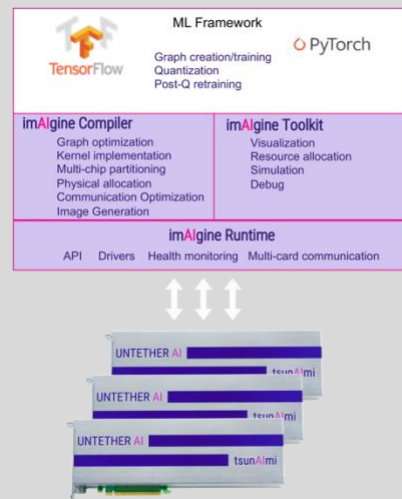
- Import TensorFlow, PyTorch, or ONNX graphs directly
- Automated quantizer and extracts performance without sacrificing accuracy
- Specify performance levels, silicon utilization, and power consumption targets

The imAIne Toolkit

- Evaluate functionality and performance using the extensive profiling and simulation tools

The imAIne Runtime

- Provides C-based API for integration into your deep learning environment
- Monitor the health and temperature of the tsunAI^{mi} acceleration cards to ensure proper operation and prevent thermal damage



Familiar frameworks

Quantization and layer optimization done in familiar ML framework

Automated graph lowering

Optimization and allocation algorithms

Extensive feedback

Resource allocations, congestions, cycle-accurate simulation

Easily integrated runtime

Hardware abstraction, communication, and monitoring

Product Specification

Specification	tsunAlmi® tsn200 accelerator card
Form factor	half height, half length, low profile
PCIe Interface	X16 PCIe Gen4
Clock Frequency	Variable, up to 840 MHz
Memory	204MB on-chip SRAM

Thermal Specification

Parameter	tsunAlmi® tsn200 accelerator card
Total board power	TDP 75W, typical application power ~40W
Cooling	Passive (active available in engineer samples only)
runAI200 maximum operating temperature	85°C Junction

Environmental

Parameter	tsunAlmi® tsn200 accelerator card
Operating temperature	0°C to 55°C
Storage temperature	-40°C to 75°C
Operating humidity	5% to 90% relative humidity
Storage humidity	5% to 95% relative humidity

Power Connector

75W via PCIe edge connector

Notice

THE INFORMATION DISCLOSED TO YOU HEREIN (THE "MATERIALS") IS PROVIDED SOLELY FOR THE SELECTION AND USE OF UNTETHER AI'S PRODUCTS. TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, MATERIALS ARE MADE AVAILABLE "AS IS". UNTETHER AI MAKES NO REPRESENTATIONS OR WARRANTIES, WHATSOEVER WITH RESPECT TO THE MATERIALS OR THE PRODUCTS, INCLUDING BUT NOT LIMITED TO REPRESENTATIONS OR WARRANTIES OF MERCHANTABILITY; SECURITY; RELIABILITY; ACCURACY; QUALITY; INTEGRATION; FITNESS FOR A PARTICULAR PURPOSE, NON-INFRINGEMENT, OR TITLE; THAT THE INFORMATION PROVIDED IN THIS MATERIAL IS SUITABLE FOR ANY PURPOSE; NOR THAT THE IMPLEMENTATION OF SUCH INFORMATION WILL NOT INFRINGE ANY THIRD PARTY PATENTS, COPYRIGHTS, TRADEMARKS, OR OTHER RIGHTS. WITHOUT LIMITING THE GENERALITY OF THE FOREGOING, UNTETHER AI EXPRESSLY DISCLAIMS ANY REPRESENTATION, CONDITION, OR WARRANTY THAT ANY INFORMATION PROVIDED TO YOU HEREUNDER, CAN OR SHOULD BE RELIED UPON BY YOU FOR ANY PURPOSE WHATSOEVER. UNTETHER AI DISCLAIMS ANY AND ALL LIABILITY RELATED TO THIS MATERIAL AND WILL NOT BE LIABLE FOR ANY LOSSES OR DAMAGE CAUSED BY RELIANCE ON THE INFORMATION IN THIS MATERIAL.

No license, either expressed or implied, is granted for any intellectual property rights of Untether AI or any third party through the information in this Material. Untether AI shall not be liable (whether in contract or tort, including negligence, or under any other theory of liability) for any loss or damage of any kind or nature related to, arising under, or in connection with, the Materials (including your use of the Materials), including for any direct, indirect, special, incidental, or consequential loss or damage (including loss of data, profits, goodwill, or any type of loss or damage suffered as a result of any action brought by a third party) even if such damage or loss was reasonably foreseeable or Untether AI had been advised of the possibility of the same. Untether AI assumes no obligation to correct any errors contained in the Materials or to notify you of updates to the Materials or to any products. You may not reproduce, modify, distribute, or publicly display the Materials without Untether AI's prior written consent. You should obtain the latest relevant Material before placing orders and should verify that such information is current and complete. All orders are subject to Untether AI's contract which outlines any applicable terms and conditions for a product.

Trademarks

Untether AI, tsunAlmi, runAI, imAIgine SDK are trademarks and/or registered trademarks of Untether AI Corporation in the U.S and other countries. Other company names may be trademarks of the respective companies.

Copyright

© 2023 Untether AI Corporation. All rights reserved.

UNTETHER AI