UNTETHER AI

tsunAImi® accelerator cards
tsn200

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 tsunAImi® 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 tsunAImi® 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 tsunAImi® 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...

imAIgine Software Development Kit

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

The imAIgine 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

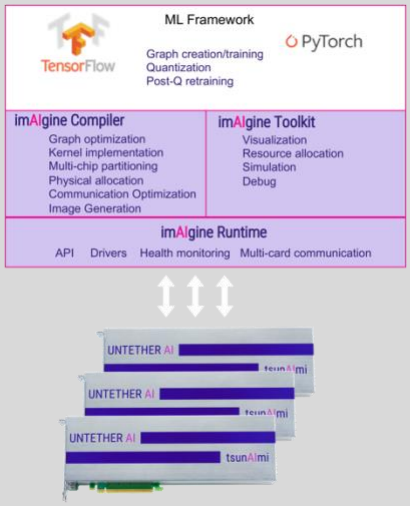
The imAIgine Toolkit

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

The imAIgine Runtime

- Provides C-based API for integration into your deep learning environment
- Monitor the health and temperature of the tsunAImi® 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	tsunAImi® 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	tsunAImi® 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	tsunAImi® 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

UNTETHER AI