



GOYA



AI Inference Card

Habana Labs GOYA is the industry's first commercially available deep learning inference processor product-line designed specifically to deliver superior performance, power efficiency and cost savings.

Habana Labs HL-100 PCIe cards incorporate a single GOYA Processor and are designed to accelerate various AI inferencing workloads, such as Image classification, object detection, natural language processing, text to speech, sentiment analysis, recommender systems and many others.

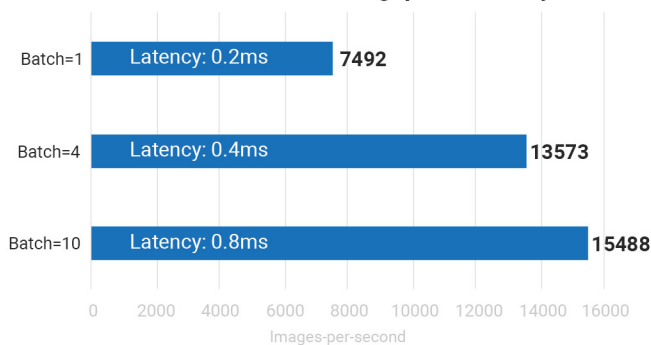
GOYA processor is a ground up design for neural network processing that incorporates a fully programmable TPC (Tensor Processing Core) along with its SynapseAI® software stack.



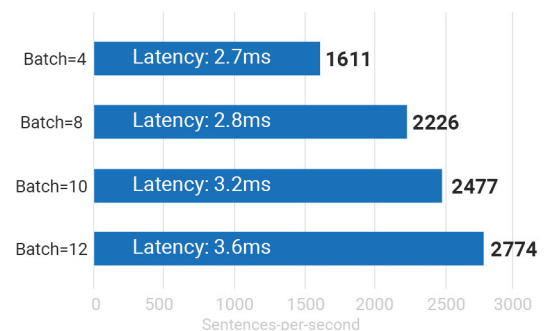
SPECIFICATIONS

PROCESSOR TECHNOLOGY	GOYA
SYSTEM INTERFACE	PCIe Gen 4.0 x 16 8-pin auxiliary power connector
FORM FACTORS & PRODUCT SKUS	HL-100: dual-slot, passive cooling full length, full height
MEMORY	16GB DDR4 memory ECC protected
THERMAL DESIGN POWER	TDP - 200W

GOYA ResNet50 Throughput & Latency



BERT – Base SQuAD Performance



Software Configuration: Ubuntu v-18.04, SynapseAI v-0.11.0-447

Hardware Configuration: Goya HL-100 PCIe card, Host: Xeon Gold 6152@2.10Ghz

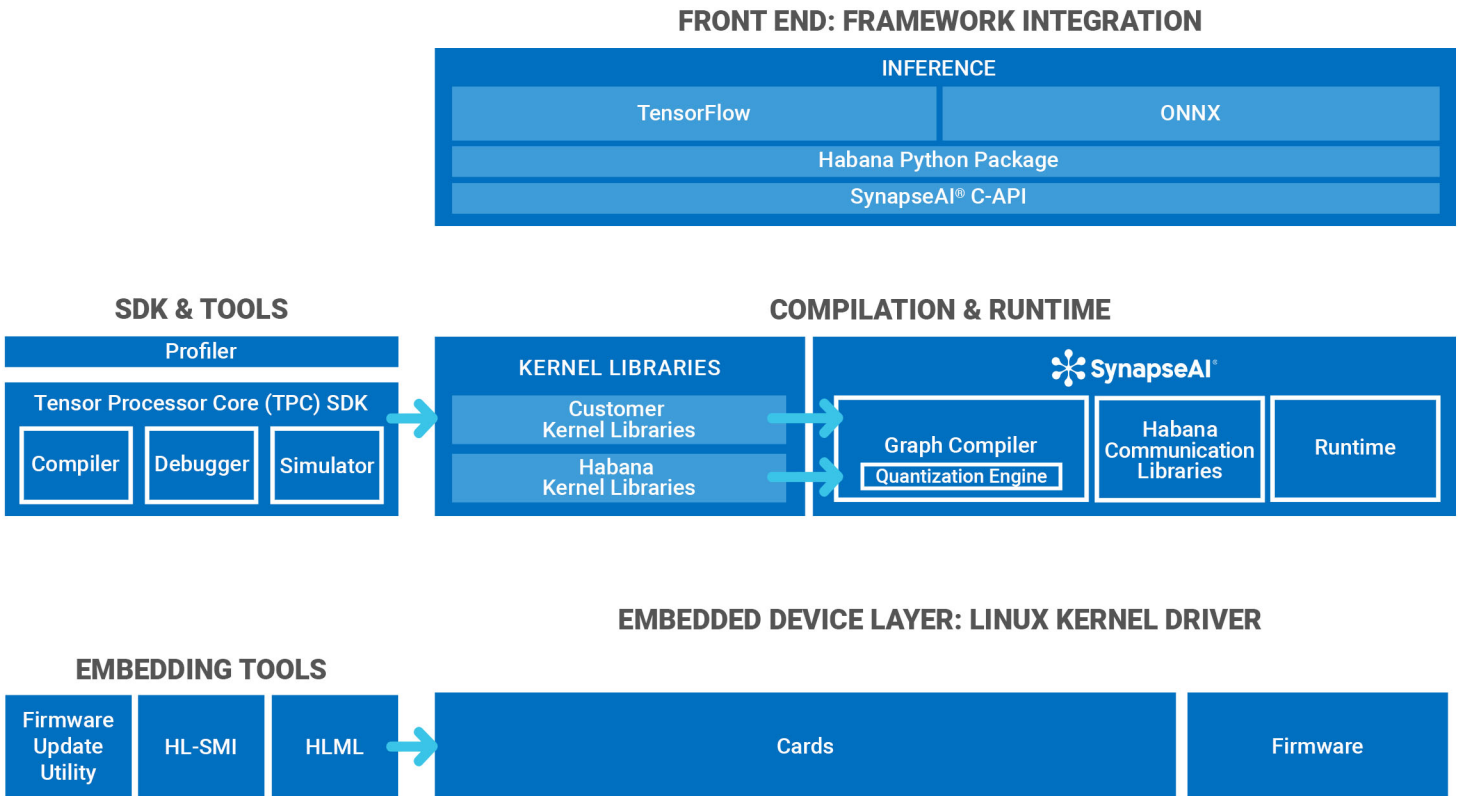
Fast Deployment with SynapseAI:

Habana Labs' SynapseAI® is a comprehensive inference software toolkit that simplifies the development and deployment of deep learning models. SynapseAI® provides inference network model compilation and runtime, eliminating the need of low level programming.

The GOYA platform is training-platform-agnostic: the deep neural network can be trained on any hardware platform. The SynapseAI® then compiles the trained model for use on the GOYA processor and the result is an optimized execution code in terms of accuracy, latency, throughput and efficiency for inferencing usage.

As part of its SW package, Habana provides an extensive set of TPC kernel libraries and opens its Tensor-Processing Core (TPC) for user programming, providing a complete TPC tool suite (debugger, simulator, compiler), which enables the user with flexibility to innovate and optimize to any unique end user requirements.

GOYA Software Stack



© 2020 Habana Labs Ltd. All rights reserved. Habana Labs, Habana, the Habana Labs logo, Gaudi, TPC and SynapseAI are trademarks or registered trademarks of Habana Labs Ltd. All other trademarks or registered trademarks and copyrights are the property of their respective owners.