



Hailo-10H M.2 Key M ET Generative AI Acceleration Module

Immersive generative AI experience
on the edge

Key Features & Benefits

High performance processing of generative AI models including LLMs with minimal CPU / GPU load

Highest cost-efficiency (TOPS/\$) compared with existing solutions

Scalable, enabling simultaneous processing of multi-models

Powered by Hailo-10H AI accelerators with 40 Tera-Operations Per Second (TOPS)

Best-in-class power efficiency. Consumes less than 3.5W (typical)

Robust software suite supports state-of-the-art deep learning models & generative AI applications

Enables real-time, low latency, and high-efficiency AI inferencing on PCs, smart cars and other edge devices

Fast time to market using a standard M.2 form factor module

Technical Specifications

Form factor
M.2 Key M 2242 / 2280

Interface
PCIe Gen-3.0, 4-lanes (x4)

Supported
Microsoft Windows

ON-Module DDR
8GB LPDDR4

Supported host architectures
X86 or ARM aarch-64 based

Supported AI frameworks
TensorFlow, TensorFlow Lite,
Keras, PyTorch & ONNX

Hailo-10H M.2 Generative AI Acceleration Module

With M.2 form factor, the Hailo-10H based modules can be plugged into existing PCs and edge devices with an M.2 socket to execute deep neural network inferencing in real-time utilizing low power for a broad range of applications and market segments.

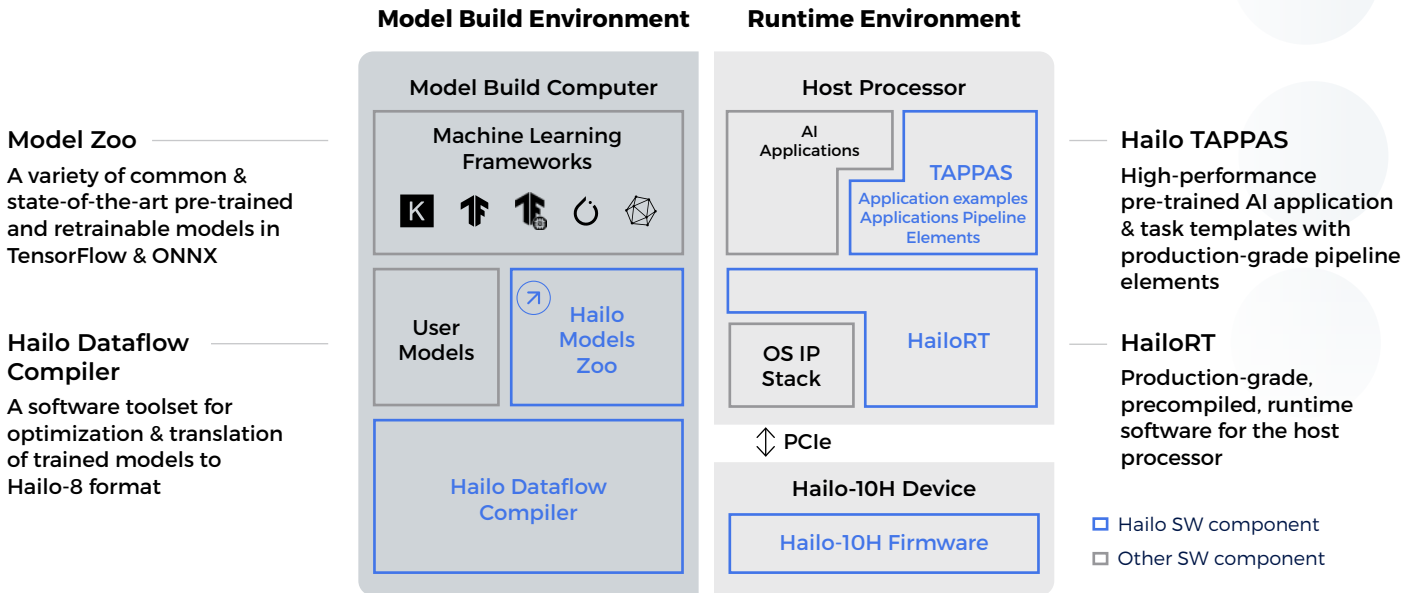
Featuring the Hailo-10H AI Processor

The Hailo-10H AI processor, delivering up to 40 tera-operations per second (TOPS), significantly outperforms all other edge AI processors. Its area and power efficiency are far superior to other leading solutions by an order of magnitude.

The Hailo-10H unique, powerful and scalable structure-driven dataflow architecture takes advantage of the core properties of neural networks. It enables edge devices to run deep learning applications at full scale more efficiently, effectively, and substantially than traditional solutions, while significantly lowering costs.

Comprehensive Software Suite

AI software suite which seamlessly integrates with existing deep learning development frameworks to allow smooth and easy integration in existing development ecosystems. The Hailo AI software suite includes:



Model Zoo
A variety of common & state-of-the-art pre-trained and retrainable models in TensorFlow & ONNX

Hailo Dataflow Compiler
A software toolset for optimization & translation of trained models to Hailo-8 format

Hailo TAPPAS
High-performance pre-trained AI application & task templates with production-grade pipeline elements

HailoRT
Production-grade, precompiled, runtime software for the host processor

Hailo Ecosystem

Hailo collaborates with partners globally to offer products for edge solutions – from small and fanless devices, to ruggedized appliances, embedded computing platforms and more.

To select your Hailo-based AI platform [click here](#)

System Usage

M.2 module connected to the various types of host processors via PCIe.

