

# Hailo-8 M.2 Al Acceleration Modules

Delivering data center class performance to edge devices



### **Key Features & Benefits**

Powered by 26 Tera-Operations Per Second (TOPS) Hailo-8 Al Processor

Best-in-class power efficiency

Enabling real-time, low latency, and high-efficiency Al inferencing on edge devices

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

Scalable, enabling simultaneous processing of multi-streams & multi-models

Robust software suite supports state-of-the-art deep learning models & applications out-of-the-box Supporting extended temperature range of -40°C to 85°C

Fast time to market using a standard form factor module, with key M, key B+M & key A+E

→ Optional extensions for Key M & Key B+M

## **Technical Specifications**

Form factor options M.2 key M, key B+M, key A+E

#### Dimensions

- → Key M, B+M: 22×42 mm with breakable extensions to 22×60 mm & 22×80 mm
- → Key A+E, 22×30 mm

#### Interface

PCIe Gen-3.0, 2-lanes (4-lanes in key M)

Supported host architectures
X86 or ARM based

#### Supported OS Linux, Windows

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

### Hailo-8 M.2 AI Acceleration Modules

Compatible with M.2 form factor, the Hailo-8 based modules can be plugged into an existing edge device with an M.2 socket to execute deep neural network inferencing in real-time utilizing low power for a broad range of market segments.



M.2 Key M (2242/2260/2280)



M.2 Key B+M (2242/2260/2280)



M.2 Key A+E (2230)

### Featuring the Hailo-8 AI Processor

The Hailo-8 AI processor, delivering up to 26 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.

Hailo-8 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**

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

#### **Model Build Environment Runtime Environment Model Build Computer Host Processor** Machine Learning Applications **TAPPAS** Frameworks **TAPPAS** A set of full application 作 Application examples, implementing examples TensorFlow TensorFlowlite pipeline elements and K $\dot{O}$ ⇎ pre-trained AI tasks PyTorch **HailoRT HailoRT** Hailo User Model Production-grade, Models precompiled, runtime 700 software for the host processor ↑ PCIe Hailo Dataflow Hailo-8 Device Compiler Hailo SW component Hailo-8 Firmware Other SW component

#### Model Zoo

Over 100 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 Ecosystem**

Hailo collaborates with leading partners to seamlessly integrate their products with ours, creating easy-to-use, dependable, and highquality solutions that adhere to the industry's most rigorous standards. Our extensive network of trusted partners covers various regions and market segments, delivering innovative solutions to our customers.

To view our extensive network click here

### System Usage

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



#### **Part Numbers**

Hailo-8 2280 M key M.2 module Extended Temperature: HM218B1C2FAE Hailo-8 2242 M key M.2 module Extended Temperature: HM218B1C2HAE Hailo-8 2230 A+E key M.2 module Extended Temperature: HM218B1C2KAE Hailo-8 2280 B+M key M.2 module Extended Temperature: HM218B1C2LAE

www.hailo.ai

Click here for more information & product inquiry



