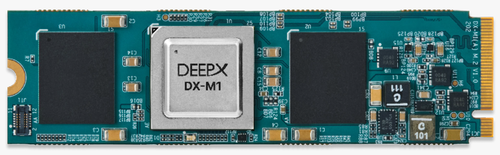
DEEPX DX-M1 AI Accelerator

DX-M1 sets a new benchmark in energy-efficient AI computing, delivering **25 TOPS (200 eTOPS) at under 5W**, significantly outperforming competitors in performance-per-watt.

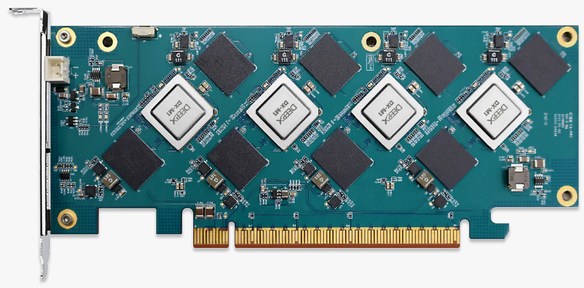
DX-M1 Chip-Based Application Modules



DX-M1 M.2 LPDDR5×2

Type: AI Accelerator
AI Perfermance: 25 TOPS / 3~5W
Form Factor: M.2 M Key (22 × 80 mm)
Interface: PCIe Gen.3 ×4
Memory: 4GB LPDDR5, QSPI 1Gbit NAND

CES INNOVATION AWARDS 2024 HONOREE
Robotics Technology



DX-H1 PCIe Card Quattro

Type: AI Accelerator
AI Perfermance: 100 TOPS / 20W
Form Factor: PCIe Card (167 × 66.4mm)
Interface: PCIe Gen.3 (4×4×4×4 Bifurcation)
Memory: 16GB LPDDR5, QSPI 1Gbit NAND

CES INNOVATION AWARDS 2024 HONOREE
Computer HW & Components

Key Features

Outstanding Power Efficiency
Exceptional AI performance with reduced power consumption, enabling longer battery life for your products and greater energy savings for data centers. (Butter-Proof)

Dedicated DRAM Design
Engineered for simultaneous multi-model processing, our dedicated DRAM ensures smooth, high-performance AI execution—even when running several models at once.

Universal CPU Compatibility
Designed for seamless integration with any host CPU architecture, enabling faster deployment, simplified development, and full compatibility of existing infrastructure.

Leading Cost Efficiency
Minimizes reliance on costly on-chip SRAM, delivering high performance with a dramatically lower bill of materials—bringing powerful AI accessible across more devices.

AI Performance

AI Model	Input Resolution	Accuracy (mAP) FP32 GPU	Accuracy (mAP) *IQ8 DX-M1	FPS DX-M1	NPU Power (W) DX-M1	FPS/W DX-M1
YOLOv5m	640×640	45.08	45.07	236	3.16	74.75
YOLOv7e6	640×640	55.58	55.45	19.49	2.50	7.81
YOLOv8x	640×640	53.63	53.12	49.03	3.17	15.48
YOLOv8l	640×640	52.57	52.01	93	3.60	25.32
YOLOv8m	640×640	50.11	49.56	135	2.76	48.32
YOLOv9c	640×640	52.86	52.36	47.89	2.83	16.95

DEEPX Headquarters
5F, 20, Pangyoyeok-ro 241beon-gil, Bundang-gu,
Seongnam-si, Gyeonggi-do, 13494, Republic of Korea

Apply - DX TechBridge Kit
Sales Support - sales@deepx.ai

DXNN® SDK

DXNN® (DEEPX Neural Network) SDK streamlines AI deployment on **DEEPX NPUs** by integrating version-aligned tools for compilation, optimization, simulation, and inference. For **efficient development**, it's offered as the **DX-AS (All Suite), a fully integrated and optimized package.**

How It Works: 4-Step AI Deployment with DXNN® SDK

Step 1

Prepare Your ONNX Model

Convert a pre-trained AI model from frameworks such as TensorFlow, PyTorch, or scikit-learn into the ONNX format, which enables model interoperability.

AI Model Training

TensorFlow | TensorFlow Lite | PyTorch | Keras | XGBoost | Darknet

↓

ONNX
User Must Convert to ONNX Before Input

Step 2

Convert Model for DX-NPU

Use the DX-COM compiler to convert the ONNX model into a format optimized for DEEPX NPU hardware, ensuring efficient execution on the target NPU.

ONNX
User Must Convert to ONNX Before Input

↓

DX-COM™

↓

DEEPX AI Model

Step 3

Real-time Operating DX-RT

Deploy the optimized model using the DX-RT runtime API. Input data is fed into the model, and inference results are produced in real-time on-device.

DEEPX AI Model | Input Tensor | Output Tensor

DX-RT™ API

DEEPX AI Products

Step 4

Run Your Application

Apply NPU-generated AI inference results to applications like Image & Video Analytics, Object Detection, Robotics and Automation, and Smart Homes and Factories.

DEEPX AI Products Applications

Image & Video Analytics | Object Detection | Robotics and Automation | Smart Homes and Factories

DXNN® Full Stack Architecture

DXNN® Full Stack Architecture streamlines AI model deployment onto DEEPX products using its two-stage AI Model Compile and Runtime Environments.

1) AI Model Compile Environment

AI Model Input Sources

TensorFlow | TensorFlow Lite | PyTorch | Keras | XGBoost

↓

ONNX User Model
User Must Convert to ONNX Before Input

↓

AI Model Compiler

- **DX-COM®** NPU Model Compiler
- **DX-TRON** Graphical Model Viewer
- **DX-SIM** Model Execution Simulator

↓

- **DX-Model Zoo** DEEPX Pre-compiled Models
- **or**
- **Compiled User Model** Optimized Binary for DEEPX NPU

2) AI Model Runtime Environment

AI Application

- **DX-APP** SDK Usage Examples
- **User App** Your Custom Application

AI App Dev Tool

- DX-STREAM™
- gstreamer

Ecosystem

- AWS IoT Greengrass
- WNDRVR
- NetworkOptix
- CVEDIA
- VCA Applied Intelligence
- NotaAI

DX-RT™

- C++-based DX-RT™ API
- Python-based DX-RT™ API
- Core Runtime Library
- Device Drivers (Windows or Linux)

↕ PCIe

DEEPX AI Products

- **DX-M1™** AI Accelerator
- **DX-H1** AI Accelerator
- **DX-V3** AI Vision SoC
- **DX-M2™** GenAI Accelerator