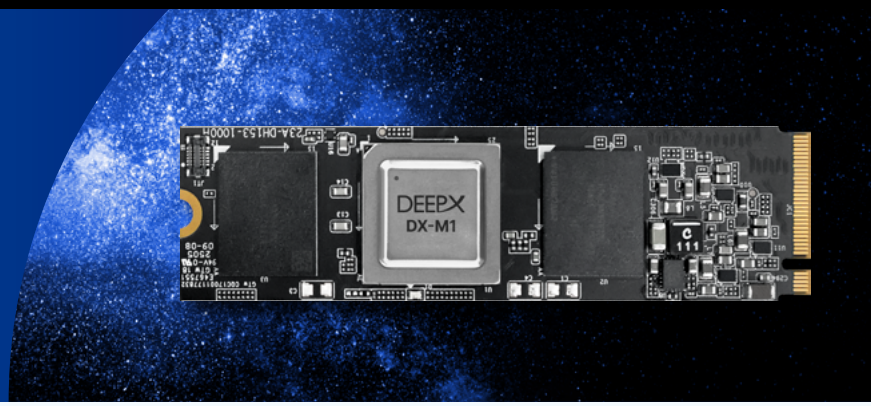
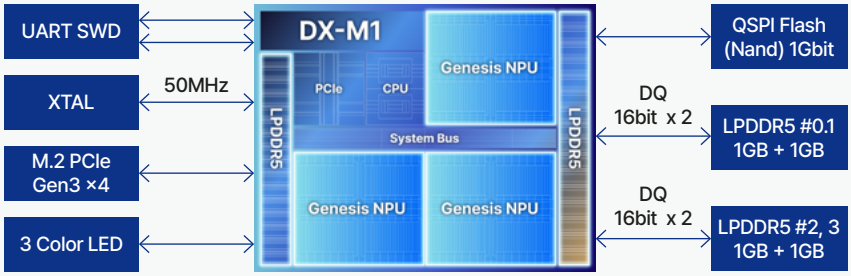
DX-M1 M.2 LPDDR5×2 for Every AIoT

The DEEPX DX-M1 M.2 module brings server-grade AI inference directly to edge devices. Delivering 25 TOPS of performance at just 2-5W, the module achieves 20x better performance efficiency (FPS/W) than GPGPUs while maintaining GPU-level AI accuracy.



"Integrating powerful AI vision processing and essential core features into a single chip, DEEPX drives innovative edge AI solutions for diverse intelligent systems."

DX-M1 M.2 LPDD5Rx2: Functional Block Diagram



Specifications

Feature	AI Accelerator	Details
Processor	INT8 Performance	25 TOPS
Signal Interface	PCI Express	PCIe Gen.3 ×4 / Bandwidth: 4GB/s *Compatible to PCIE x1
Power	Power Consumption	2W min., 5W max. for DEEPX supported models
Operating	Temperature	-25 ~ 85°C (Throttling) -25 ~ 65°C (Non_Throttling)
Environment	Humidity	40 °C @ 85% relative humidity (non-condensing)
Thermal Solution	Cooling	Heatsink (Option)
Physical	Form Factor	M.2 2280 (Key M)
	Dimensions	22mm x 80mm x 4.1mm
	Power Range	3.3V ± 5%
Software Support	Windows	Windows 11, 10 64 bit
	Linux	Ubuntu 22.04, 20.04 LTS Support Yocto Project and Docker
	Framework	Support TensorFlow, TensorFlow Lite, ONNX, Keras, PyTorch by Dataflow complier converted
System Support	CPU Platform	x86, ARM Based Architecture

DEEPX

Key Features

- Type: AI Accelerator
- Form Factor: M.2 M Key (22 × 80 mm)
- Interface: PCIe Gen.3 ×4
- Memory: 4GB LPDDR5, QSPI 1Gbit NAND Flash
- Host HW: x86, ARM Based Architecture

Support DXNN® SDK

DXNN® SDK is a comprehensive SW development environment for deploying AI on DEEPX NPUs. It integrates tools for compiling, optimizing, simulating, and inferring the latest AI models, such as YOLO, ViT, and VLMs. And it provides an optimized, ready-to-use environment as the DX-All Suite package to support fast and efficient AI development.

Target Applications

- 3D Sensing & Stereo Cameras
- AI NVR (Network Video Recorder)
- AI CCTV
- Robotics
- Automotive
- Video Conferencing Cameras
- CMS (Camera Monitoring System)
- Autonomous Robotic Platforms
- Drones
- AR/VR
- SBC (Single Board Computer)
- ADAS/AD

DEEPX HQ
5F, 20 Pangyoyeok-ro 241beon-gil, Seongnam-si, Gyeonggi-do, South Korea

USA
1735 Technology Drive Suite 740.San Jose, CA U.S.A

Taiwan
No. 66-10, Yucheng St., Nangang Dist., Taipei City 115012, Taiwan (R.O.C.)

Europe
Tel-Aviv Israel

Buy Now! - DX TechBridge Program

Contact Sales - 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 Traning

TensorFlow | TensorFlow Lite | Keras | XGBoost | PyTorch

↓

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.

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-Model Zoo
DEEPX Pre-compiled Models

or

Compiled User Model
Optimized Binary for DEEPX NPU

→

2) AI Model Runtime Environment

AI Model Input Sources

DX-APP
SDK Usage Examples

User App
Your Custom Application

AI App Dev Tool

DX-STREAM™

gstreamer

Ecosystem

aws AWS IoT Greengrass | WNDRVR | NetworkOptix

CVEDIA | VCA | Baidu 百度

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™	DX-M1M	DX-H1	DX-V3	DX-M2™
AI Accelerator	AI Accelerator	AI Accelerator	AI Vision SoC	GenAI Accelerator

DEEPX HQ
5F, 20 Pangyoyeok-ro 241beon-gil, Seongnam-si, Gyeonggi-do, South Korea

USA
1735 Technology Drive Suite 740.San Jose, CA U.S.A

Taiwan
No. 66-10, Yucheng St., Nangang Dist., Taipei City 115012, Taiwan (R.O.C.)

Europe
Tel-Aviv Israel

Buy Now! - DX TechBridge Program

Contact Sales - sales@deepx.ai