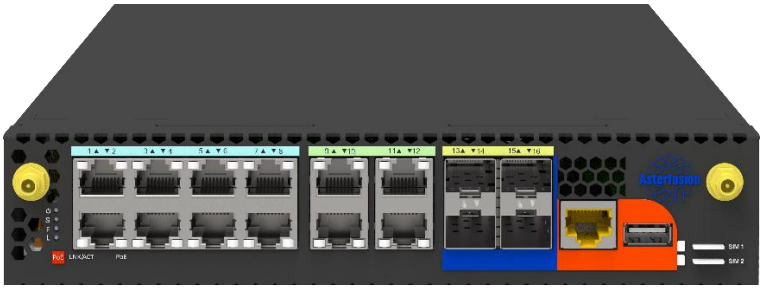
Overview

The Asterfusion ET2500 is an AI-ready edge network platform designed for modern deployments, including routing, security, MEC, and edge compute applications. Powered by Marvell OCTEON 10 CN102XX processor featuring 8-core ARM 64-bit Neoverse N2 architecture, the platform combines high-performance computing, hardware-accelerated packet processing with up to 58Gbps networking throughput, inline security acceleration, and optional AI inference capabilities in a compact low-power appliance.

The platform supports major Linux distributions including Ubuntu, Debian, CentOS, and OpenWRT, enabling customers to deploy containerized edge applications and open-source software such as Nginx, iptables, UFW, and other edge services on a single device. Running Asterfusion's native AsterNOS-VPP router software, the ET2500 delivers enterprise-grade routing, secure VPN, firewall, and edge networking services optimized through VPP and DPDK acceleration.



ET2500

An optional AI acceleration module delivering up to 160 TOPS of compute performance and 24 GB of onboard memory further extends the ET2500 with intelligent edge capabilities. The ET2500 supports deployment of modern AI frameworks and large language models (LLMs) in the 30B parameter class directly at the network edge, enabling use cases such as intelligent operations, automated troubleshooting, anomaly detection, and private AI assistants. By combining networking, security, edge compute, and AI acceleration within a single platform, the ET2500 enables tighter integration between AI workloads and network services while reducing hardware complexity, power consumption, and operational overhead.

Hardware Highlight

- Compact half-width 1U edge network appliance
- 8 x 2.5GHz ARM64 Neoverse N2 Core
- 16GB pluggable DDR5 SO-DIMM, up to 48G
- 4 x 10GE, 4 x 2.5GE and 8 x 1GE, optional 8 x 1GE PoE+, 4 x 2.5GE PoE++ @150W budget
- True inline crypto engine
- Optional AI accelerator module with 160TOPS INT8 computing performance
- Optional M.2 SSD up to 4TB
- 2 pluggable modules with M.2 form, extending support 5G/LTE, WiFi6E/7, BlueTooth5.3, GNSS, etc.
- Optional PTP/SyncE module with 20ns accuracy and OCXO holdover
- 50 Gbps intelligent data processing for routing, firewall, IPSec and SSL/TLS
- <60 Watt with FULL configuration and workload (w/o PoE)

Application Scenarios

Based on the open hardware-software decoupled architecture, the ET2500 combines a rich array of open-source software for control plane with hardware-optimized data plane, and can connect to SSDs, 5G/LTE, WiFi6E/7, and other devices via M.2 and USB interfaces, thereby addressing diverse application scenarios. Here are some typical scenarios that can be used individually or in combination:

- **Router/VPN/Firewall:** AsterNOS-VPP (SONiC-VPP)
  - Hardware-optimized vector packet technology and DPDK accelerate data plane forwarding, delivering up to 58Gbps forwarding performance.
  - Multi-WAN Routing across Ethernet and 5G/LTE uplinks with traffic steering and failover.
  - Hierarchical QoS policies enable precise traffic for users, services, and applications.
  - Hardware-accelerated VPN with encryption/decryption engine supports up to 50Gbps throughput.
  - Advanced policy-based traffic control with GeoIP/Geosite-aware filtering.
- **Edge ARM Server：** Standard Linux distribution with VPP/DPDK acceleration
  - Powered by an 8-core 2.5GHz ARM64 Neoverse N2 processor optimized for networking and edge infrastructure workloads.
  - Open Linux platform supporting standard software ecosystems, containers, and cloud-native applications.

- Optimized for open software frameworks including DPDK, VPP, Nginx, featuring a hardware-based true inline crypto engine to boost cryptographic application performance.
- Delivers up to 58Gbps accelerated networking performance for packet processing.
- Maintains high energy efficiency, keeping total system power consumption below 60W even under full-workload operation.

■ **AI Inference for Edge Networking:** AI Framework+ Inference Engine +LLM

- Combines edge networking, security acceleration, and AI inference capabilities in a compact and power-efficient platform.
- Integrated AI module delivers up to 160 TOPS AI computing performance with 24GB LPDDR5 memory for edge AI workloads and local LLM inference.
- Compatible with popular AI frameworks and runtimes including PyTorch, llama.cpp, Ollama, vLLM, and TensorFlow Lite.
- Supports quantized AI models such as Llama 3 8B, GPT-OSS-20B-A3B, Qwen3 7B–30B, and other GGUF-based edge inference models.
- Running Qwen3-30B with 2048-token input delivers approximately 37ms TTFT (Time to First Token) and 30 tokens/s generation throughput, with support for up to 32K context length.

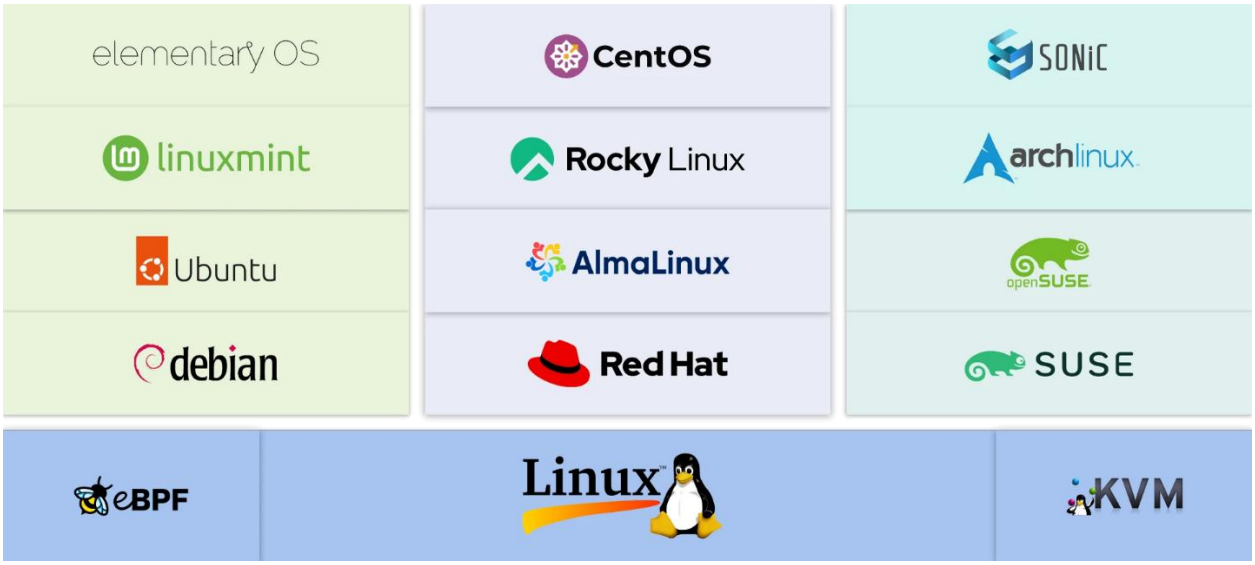
Additionally, users have the flexibility to install new software or develop their own software using the built-in toolchain as needed to address additional use cases.

Operating System

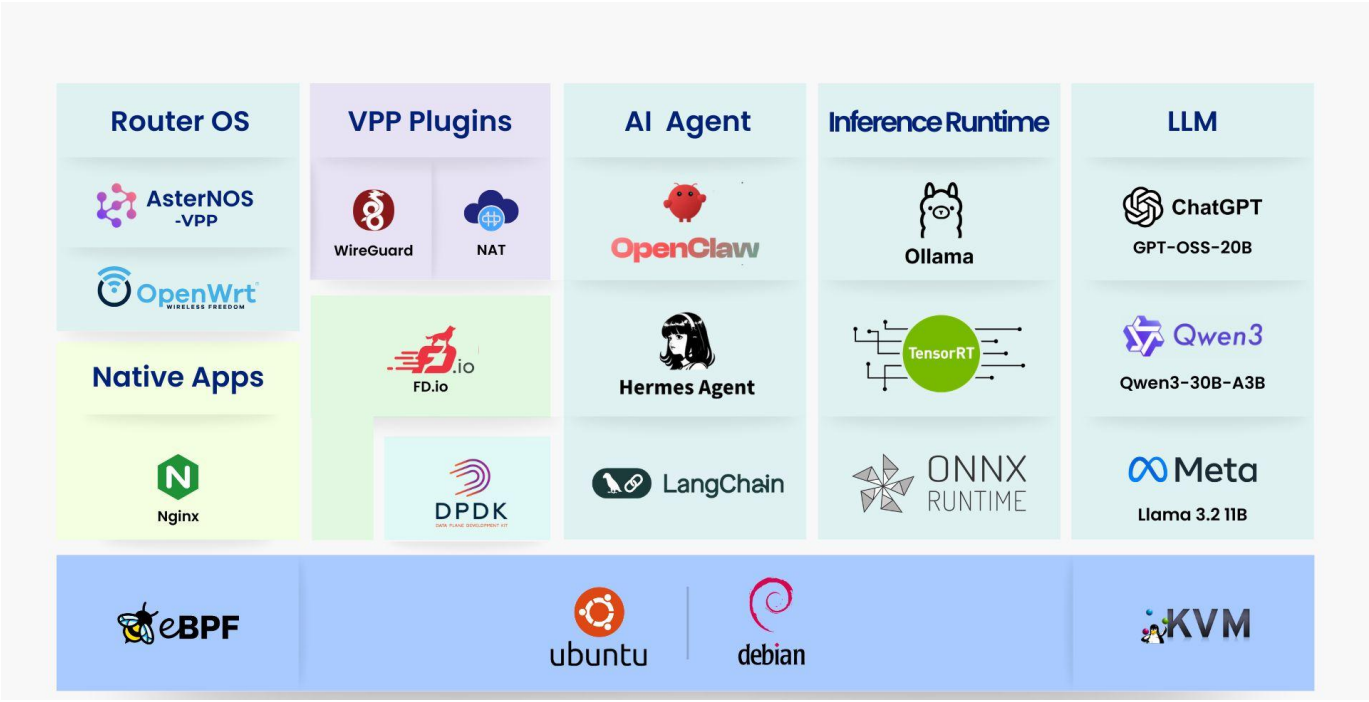
- Supports Ubuntu, Debian, SONiC and other Linux distribution, such as CentOS, OpenSUSE, Arch Linux, AlmaLinux, Rocky Linux, Linux Mint and Elementary OS.
- Install and upgrade the OS using a USB disk with Arm Trusted Firmware and UEFI.
- Embedded eBPF (extended Berkeley Packet Filter) in Linux kernel via XDP.

Software Ecosystem

- Optimized DPDK (Data Plane Development Kit) tied to HW Acceleration.
- Open-source routers, including VPP (Vector Packet Processing), AsterNOS-VPP, OpenWRT, etc.
- Open-source firewalls, including iptables, UFW, nftables, Firewalld, etc.
- Cloud-native deployment with Docker, Kubernetes, and containerized edge applications.
- Open-source load balances, including Nginx.
- Rich VPP plugins, including Marvel device plugin, QUIC, SRv6, LLDP, NAT64, SRTP etc.
- GCC, GDB, BinUtils, Buildroot and other tool chains.
- C/C++/Python/Go/Rust/Java/Lua and other programming languages.
- AI software ecosystem supporting PyTorch, TensorFlow Lite, Keras, ONNX Runtime, llama.cpp and Ollama.
- Applications from other Linux distributions on Ubuntu/Debian using Docker with direct access to the host network.
- Any software for ARM64 + Linux.

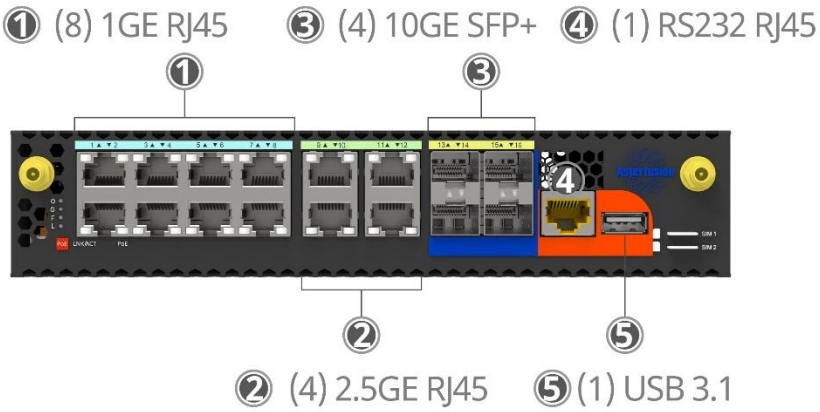


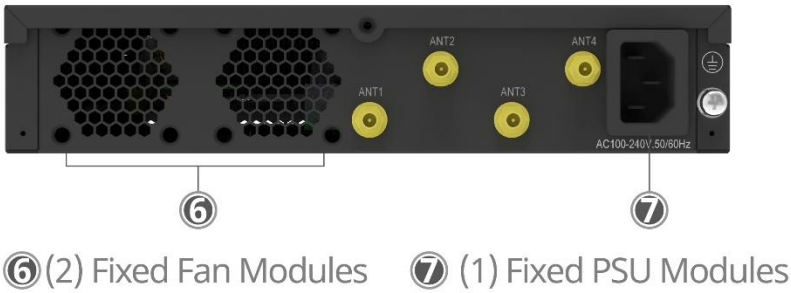
Operating System



Software Ecosystem

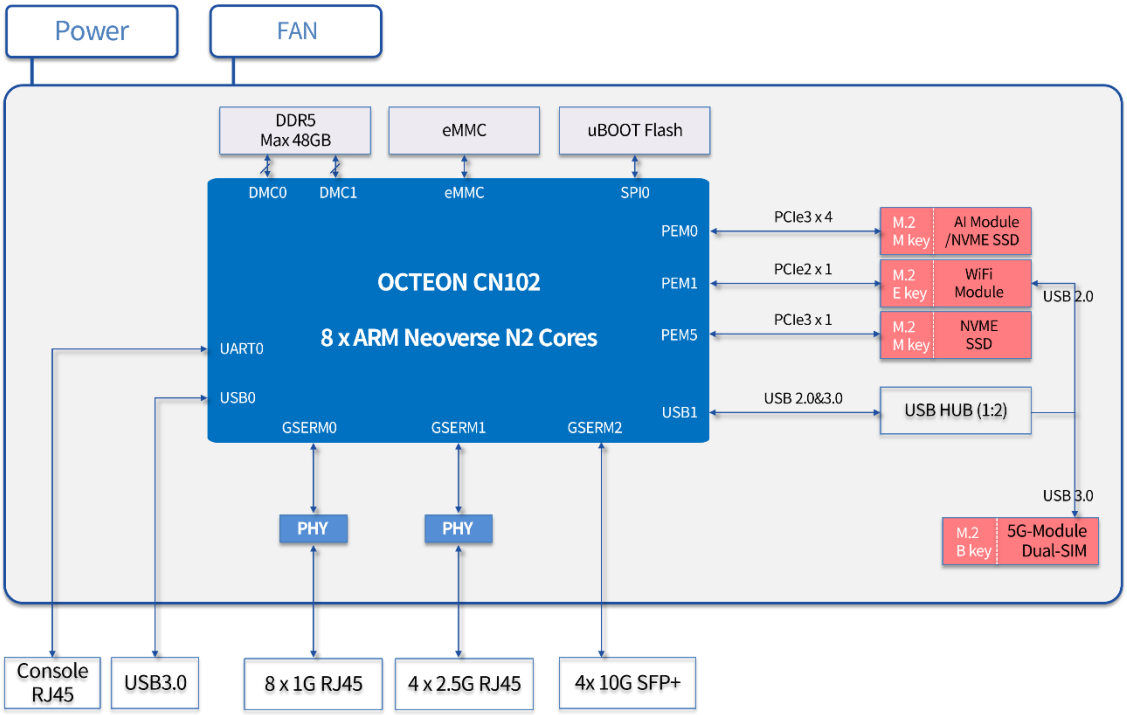
Interfaces





ET2500

System Architecture



ET2500 Edge Network Appliance system architecture

ET2500 Platform Specification

Product Model		ET2500
Processor	CPU	8 Core ARM64 N2 @2.5 GHz
	Cache	L2 8MB, L3 16MB
Memory and Storage	RAM	16GB DDR5 SODIMM, up to 48GB
	Flash	64GB eMMC 5.1
	NVME SSD (Option)	up to 4TB, M.2 M key, one share slot with AI accelerator module
Network Interface	10GE (SFP+)	4
	2.5GE (RJ45)	4
	1GE (RJ45)	8
	PoE	2.5GE ports POE++, 1GE ports POE+
	WiFi (Option)	WiFi6E/7, M.2 E key
	5G/LTE (Option)	2 SIM cards, M.2 B key
	Antenna	6
Network Performance	Routing capacity	58Gbps
	Encryption and Decryption capacity	50Gbps
	PTP/SyncE accuracy	20ns
	PTP/SyncE holdover	> 8hours
Diagnostic Interface	USB	1 x USB3.0
	Console	1 x Console RJ45
Electrical Characteristics	Fan	2
	Power Module	1 x 150W (w/o PoE) 1 x 270W (PoE)
	Input voltage	100~240VAC
	Maximum power consumption	60W (FULL configuration and workload)
	PoE budget	150W

Dimensions	Height	1U
	Dimensions (W x H x D, mm)	220 x 44 x 310
Operating Conditions	Operating temperature	0 – 45°C
	Relative humidity	5% - 95% (non-condensing)

AI Acceleration Module Specification

Model	M2-AI-INFERENCE-24GB
AI Compute Perf.	• Up to 160 TOPS elastic AI computing performance • Up to 100 TFLOPS @ bFP16
Memory	• 24 GB LPDDR5 / LPDDR5X • Memory bandwidth: 153.6 GB/s
Storage	1 × 32 MB NOR Flash
Interface Type	M.2 Key M
PCIe Interface	1 × PCIe Gen4 interface Supports up to x4 lanes, compatible with x3 / x2 / x1 Maximum bandwidth up to 8 GB/s per interface
UART	Supports 2 × UART interfaces Standard baud rates supported, up to 921600 bps
I2C	1 × I²C controller Supports 100 kHz and 400 kHz communication rates
Dimensions (mm)	22 x 80 x 2.58
Weight	9 g (±10 g)
Operating Temperature	-10°C to +60°C
Operating Humidity	10% RH – 90% RH (non-condensing)
Supply Voltage	3.3 V ± 5%
Maximum Continuous Operating Current	9 A

Supported LLMs:

The AI module supports stable inference of large language models in the 7B–30B parameter range, using modern quantization (INT4/INT8).

- LLM: Qwen3-14B, Qwen2.5-7B, deepseek-8B, Qwen3-30B-A3B, GPT-OSS-30B-A3B, etc.
- MLLM: minicpm-o-2_6
- VLM: Qwen2.5-VL-7B, Qwen3-VL-2B, Qwen3-VL-4B, Qwen3-VL-8B, Qwen3-VL-30B-A3B

Ordering Information

Part Number	Description
ET2508-4S4M8	Edge Network Appliance, 8 Core ARM64 N2 CPU, 16GB DDR5, 64GB eMMC, 4 x 10GE + 4 x 2.5GE + 8 x 1GE, 150W Power
ET2508-4S4M8-SWP	Edge Network Appliance, 8 Core ARM64 N2 CPU, 16GB DDR5, 64GB eMMC, 4 x 10GE + 4 x 2.5GE + 8 x 1GE, 4 x PoE@150W on 2.5GE/1GE ports, 270W Power
ET2508-4S4M8-2	Edge Network Appliance, 8 Core ARM64 N2 CPU, 32GB DDR5, 64GB eMMC, 4 x 10GE + 4 x 2.5GE + 8 x 1GE, 150W Power
ET2508-4S4M8-2-SWP	Edge Network Appliance, 8 Core ARM64 N2 CPU, 32GB DDR5, 64GB eMMC, 4 x 10GE + 4 x 2.5GE + 8 x 1GE, 4 x PoE@150W on 2.5GE/1GE ports, 270W Power
Optional components and spares	
M2-AI-INFERENCE-24GB	AI Accelerator Module, 160 Tera-Operations Per Second, 32MB Flash, 24GB LPDDR5 with memory bandwidth 153.6GB, M.2 M Key 2280
SSD-M2-NVME-1TB	SSD,1TB, NVME, M.2 M Key 2280, PCIe3 x4
SSD-M2-NVME-2TB	SSD,2TB, NVME, M.2 M Key 2280, PCIe3 x4
SSD-M2-NVME-4TB	SSD,4TB, NVME, M.2 M Key 2280, PCIe3 x4
5G-M2-5Gbps	5G Module, PCIe 3x1, M.2 B key, 5G NR, LTE, GNSS, 5.0Gbps (DL) / 650Mbps (UL)
WiFi7-M2-9Gbps	WiFi6E/7 Module, PCIe 3x1, M.2 E key, Tri-Band 802.11be/ax, Bluetooth 5.3, 6-Stream 9.3Gbps
PTP-20ns	PTP module, 20ns accuracy, BC support with holdover > 8hours
SVC-Basic-1Y-ET	Basic H/W Service and Warranty

Warranty and Service Support

Asterfusion ET2500 series come with 2-year Basic H/W service and warranty.

To acquire more info about company, products, and solutions: *www.cloudswit.ch*
Sales: *bd@cloudswit.ch*