

Heterogeneous Computing AI Accelerator Module CXL 3.0 Compliance Guide - Official Technical Overview & Hardware Datasheet

EXECUTIVE SUMMARY

The Heterogeneous Computing AI Accelerator Module (HCAAM) with CXL 3.0 Compliance represents a paradigm shift in accelerated computing for next-generation telecom and enterprise infrastructure. Designed to address the insatiable demand for AI inference and training at the edge and in the core, this module integrates a multi-core AI accelerator fabric with a fully compliant CXL 3.0 host interface. The HCAAM enables cache-coherent, low-latency attachment of high-performance AI compute to standard x86 or Arm servers, dramatically reducing total cost of ownership while unlocking unprecedented computational density. This datasheet provides a comprehensive technical overview, architectural description, and compliance framework for system designers, network architects, and procurement engineers evaluating the HCAAM for carrier-grade deployments.



ARCHITECTURE & CHASSIS DESIGN

The HCAAM is built around a proprietary heterogeneous compute engine comprising an array of AI-optimized tensor cores, a high-throughput shared memory subsystem, and a dedicated CXL 3.0 controller. The module adheres to the CXL 3.0 specification, supporting PCIe 6.0 physical layer signaling, and implements both CXL.io and CXL.cache/CXL.mem protocols. The architecture enables direct, coherent memory access between the host CPU and accelerator memory, eliminating the need for explicit data copies and significantly lowering inference latency. The module is housed in a compact, single-slot, half-length PCIe add-in card form factor (HHHL) with a passive heatsink, suitable for dense server and edge chassis. For ruggedized deployments, a conduction-cooled variant with a sealed enclosure is available. The module includes an integrated management controller supporting NC-SI and MCTP over SMBus, enabling

out-of-band monitoring and control.

HARDWARE FEATURES

- Heterogeneous AI compute fabric with mixed-precision support (INT8, FP8, FP16, BF16, TF32).
- CXL 3.0 compliant host interface, backward compatible with CXL 2.0 and CXL 1.1.
- Cache-coherent memory access via CXL.cache and CXL.mem.
- Up to 256 GB of on-module HBM3e memory with 1.2 TB/s bandwidth.
- Multi-level hardware security including secure boot, memory encryption, and attestation.
- Integrated 100Gbps Ethernet network interface for scale-out AI clusters.
- Advanced RAS features: ECC, memory mirroring, and error containment.
- Comprehensive software stack support: PyTorch, TensorFlow, ONNX Runtime, and custom SDK.
- Hot-plug capable with CXL 3.0 hot-plug support.
- Redundant power inputs with on-board power management.

COMPLIANCE & STANDARDS

The HCAAM is designed to meet the most stringent industry standards and

regulatory requirements for telecom and enterprise environments. Key compliances include:

- CXL 3.0 specification (PCI-SIG).
- PCIe 6.0 specification (PCI-SIG).
- OCP Accelerator Module (OAM) form factor guidelines (where applicable).
- NEBS Level 3 (Telcordia GR-63-CORE, GR-1089-CORE) for the conduction-cooled variant.
- IEC/EN 62368-1 safety.
- FCC Part 15 Class A, EN 55032 Class A emissions.
- RoHS 3, REACH, and WEEE directives.
- CE and UKCA markings.
- FIPS 140-3 Level 2 cryptographic module validation (pending).

TECHNICAL SPECIFICATIONS

Parameter	Specification
Form Factor	HHHL (Half-Height Half-Length) PCIe card, single-slot
Host Interface	PCIe 6.0 x16, CXL 3.0 (backward compatible CXL 2.0/1.1)
AI Compute	Heterogeneous tensor core array, 512

	TFLOPS (INT8), 256 TFLOPS (BF16)
Memory	256 GB HBM3e, 1.2 TB/s bandwidth
CXL Protocols	CXL.io, CXL.cache, CXL.mem
Network Interface	2x 100Gbps Ethernet (QSFP28)
Power Consumption	250 W typical, 300 W max
Power Input	12V and 3.3V from PCIe slot, auxiliary 8-pin power connector
Cooling	Passive heatsink (air-cooled) or conduction plate (conduction-cooled)
Operating Temperature	0°C to 55°C (air-cooled), -40°C to 70°C (conduction-cooled)
Dimensions	167.65 mm x 111.15 mm (HHHL)
Management	NC-SI, MCTP over SMBus, IPMI 2.0
Security	Secure boot, memory encryption, attestation, TPM 2.0
Software Support	Linux, Windows Server, PyTorch, TensorFlow, ONNX Runtime
Regulatory	CE, UKCA, FCC Class A, NEBS Level 3 (conduction-cooled), RoHS 3

ORDERING OPTIONS

| SKU | Description |

|-----|-----|

| HCAAM-C3-256-A | Air-cooled HHHL module, 256GB HBM3e, CXL 3.0 |

| HCAAM-C3-256-C | Conduction-cooled HHHL module, 256GB HBM3e, CXL 3.0,

NEBS Level 3 |

| HCAAM-C3-128-A | Air-cooled HHHL module, 128GB HBM3e, CXL 3.0 |

| HCAAM-C3-128-C | Conduction-cooled HHHL module, 128GB HBM3e, CXL 3.0,

NEBS Level 3 |

