Heterogeneous Computing AI Accelerator Module Interconnect Protocol Standards - Official Technical Overview & Hardware Datasheet

EXECUTIVE SUMMARY

The Heterogeneous Computing AI Accelerator Module Interconnect Protocol Standards define a comprehensive framework for high-bandwidth, low-latency communication between disparate AI accelerator modules within a unified compute complex. As artificial intelligence workloads grow in complexity and scale, the need for seamless interoperability between Graphics Processing Units (GPUs), Tensor Processing Units (TPUs), Field Programmable Gate Arrays (FPGAs), and custom Application-Specific Integrated Circuits (ASICs) becomes paramount. This specification establishes the physical, electrical, and logical layers required to achieve deterministic, cache-coherent data exchange across heterogeneous compute elements. Designed for carrier-grade infrastructure, these standards ensure that AI training and inference platforms can scale linearly while maintaining strict power efficiency and thermal envelopes. This document serves as the authoritative reference for system architects, hardware engineers, and pre-sales professionals evaluating next-generation AI compute fabrics.

ARCHITECTURE & CHASSIS DESIGN

The interconnect architecture is predicated on a switched-fabric topology that supports point-to-point, multi-drop, and mesh configurations. The physical layer utilizes a high-density, differential signaling scheme capable of operating over both copper backplanes and optical interconnects. The chassis design accommodates modular accelerator trays, each housing up to eight heterogeneous compute modules. A centralized fabric switch ASIC manages traffic routing, quality of service (QoS), and congestion control. The backplane supports orthogonal mid-plane connections, enabling hot-swappable module replacement without disrupting adjacent compute lanes. Thermal design incorporates liquid-cooling-ready cold plates and high-static-pressure air shrouds to maintain operational thresholds in dense rack environments.

HARDWARE FEATURES

The protocol standard supports a range of advanced hardware features essential for modern AI infrastructure. These include:

- Cache Coherency Engine: Implements a directory-based protocol (MESI-F) to maintain data consistency across heterogeneous memory spaces.
- Dynamic Bandwidth Allocation: Real-time arbitration logic that prioritizes latency-sensitive inference traffic over bulk training data transfers.
- Error Correction and Containment: End-to-end CRC, retry buffers, and link-level replay mechanisms ensure data integrity at speeds exceeding 400 Gbps per lane group.
- Multi-Protocol Encapsulation: Native support for PCIe Gen5/6, CXL 2.0/3.0, and proprietary accelerator interconnects.
- Security Engine: Line-rate encryption (AES-256-GCM) and replay protection for inter-module communication.

COMPLIANCE & STANDARDS

The Heterogeneous Computing AI Accelerator Module Interconnect Protocol Standards are aligned with and contribute to the following industry specifications:

- PCI Express Base Specification 5.0 and 6.0

- Compute Express Link (CXL) 2.0 and 3.0

- OCP Accelerator Module (OAM) 2.0

- IEEE 802.3cd (for optical management interfaces)

- ISO 9001:2015 Quality Management Systems

- RoHS 3 (EU 2015/863) and REACH compliance

- NEBS Level 3 (where applicable for telecom edge deployments)

TECHNICAL SPECIFICATIONS

[IMAGE_1]

Parameter	**Specification**
Form Factor	Modular 2RU chassis per compute tray, up to 8 trays per rack
Interconnect Topology	Switched fabric, mesh, and point-to-point
Per-Lane Data Rate	Up to 112 Gbps PAM4 (optical) / 56 Gbps NRZ (copper)
Aggregate Bandwidth	14.4 Tbps per tray (bidirectional)
Protocol Support	PCIe Gen5/6, CXL 2.0/3.0, OAM 2.0, proprietary
Cache Coherency	Directory-based MESI-F

Latency	< 100 ns module-to-module (cut-through)
Power Supply	1+1 Redundant AC/DC (48V DC nominal)
Cooling	Liquid-cooling ready, air-cooled option
Operating Temperature	0°C to 50°C (standard); -40°C to 70°C (ruggedized)
Management Interface	Redfish, SNMP v3, CLI over 1000BASE-T
Security	AES-256-GCM, replay protection, secure boot

ORDERING OPTIONS

Heterogeneous Computing AI Accelerator Module Interconnect Protocol Standards are available in the following configurations:

- HC-AI-IPS-100: Base specification document and reference architecture (digital delivery).

- HC-AI-IPS-200: Full hardware development kit including fabric switch ASIC, accelerator trays, and backplane.

- HC-AI-IPS-300: Carrier-grade rack integration with redundant power and cooling.

- HC-AI-IPS-400: Ruggedized edge variant with extended temperature range.

- HC-AI-IPS-SVC: Annual support and protocol update subscription.