HighPoint Launches Gen5 AI Upgrade Solutions: Bringing Blackwell GPU Compute and Gen5 Storage to Edge Infrastructure

Technical dedicated Solutions Page Enables x86 Edge Infrastructure to Run 600W Blackwell Accelerators and Gen5 Storage at Full 64 GB/s Line Rate Without Full Platform Refreshes.

FREMONT, CA — October 2026 — HighPoint Technologies, Inc., a pioneer in high-performance PCIe switching architectures and NVMe storage connectivity, has unveiled its new portfolio of **Gen5 AI Upgrade Solutions**. Engineered for systems integrators and AI infrastructure professionals, the new dedicated landing page details deployment architectures for adding NVIDIA Blackwell RTX 6000 Pro GPUs and high-density Gen5 NVMe arrays to existing x86 Intel and AMD systems via autonomous hardware-level P2P switching fabric.

Overcoming the Physical and Bus Bottlenecks of Edge AI

As large language models, multimodal vision pipelines, and real-time inference expand into decentralized edge environments, engineering teams face critical infrastructure roadblocks:

The Gen4 Replacement Dilemma: Swapping out entire server backbones for native Gen5 hardware demands significant capital expenditure and introduces extensive project delays driven by persistent market supply constraints for enterprise Gen5 CPUs and DDR5 memory.

The Physical Edge Barrier: Compact 1U/2U server chassis and short-depth edge enclosures often face severe slot limitations, insufficient onboard power, and thermal constraints that physically prevent housing 600W active-cooled GPUs internally.

HighPoint's Gen5 AI Upgrade Solutions resolve both challenges at the silicon layer, unlocking native NVIDIA GPUDirect P2P and GPUDirect Storage (GDS) compatibility across x86 Intel and AMD platforms. Powered by an integrated 48-lane Broadcom PCIe Gen5 switch and PCI-SIG CopprLink connectivity, these solutions establish a localized, autonomous downstream sub-fabric. High-throughput AI workloads execute via hardware-level Peer-to-Peer (P2P) DMA at full line-rate (up to 64 GB/s bi-directional)—transferring data directly between local NVMe storage, edge accelerators, and GPUs while completely bypassing the host CPU, system RAM, and host motherboard bus

"Edge AI architects are caught between two bad options: replace entire server fleets prematurely, or sacrifice GPU performance due to host bus bottlenecks," said May Hwang, Marketing Director at HighPoint Technologies. "Our Gen5 AI Upgrade Solutions offers an immediate third path. By combining dedicated 64 GB/s CopprLink cabling with our autonomous 48-lane switching fabric, system builders can integrate Blackwell-class compute and Gen5 storage into their existing x86 Intel and AMD edge footprints at full line rate."

Three Targeted Architectural Deployment Models

The newly launched solution details three modular hardware deployment blueprints tailored to real-world edge environments:

1. **Disaggregated External Compute:** Couples the Rocket 7634D/7638D/1638E host adapter with the RocketStor 8631D-1300W enclosure over dedicated CopprLink™ CDFP x16 cabling, offloading 1300W of power and active cooling outside 1U/2U servers to run 600W Blackwell GPUs at continuous peak boost clocks.

2. **Vertical / Cascading PCIe Expansion:** Bridges the **Rocket 1600A Series** to secondary adapters like the **Rocket 1528D** via high-density MCIO connections, multiplying a single host x16 slot into an expansive hierarchy supporting 8+ NVMe drives alongside edge GPUs and NPU accelerators without chassis overhauls.

3. **Horizontal Direct PCIe Expansion:** Connects GPUs and Gen5 NVMe scratch pools directly to a single-tier crossbar matrix (such as the **Rocket 7638D**), delivering zero-hop, ultra-low latency (~100–150 ns) for live inference, token caching, and prompt fine-tuning.

Learn More

To explore architectural diagrams, request evaluation units, and access technical blogs comparing CopprLink™ vs. OCuLink, please visit HighPoint's Gen5 AI Upgrade Solutions.