Future-Proofing Approach

Commercial and defense compute-intensive applications including radar, EO/IR imagery, AI, urban air mobility, autonomous driving, cognitive EW and sensor fusion generate huge volumes of sensor and network data which require big processing resources to extract its actionable information in real-time. By off-loading compute-intensive functions such as fast Fourier transforms (FFTs), matrix multiplications, constant false alarm rate (CFAR), QR decomposition (QRD), video codecs (H.264, JPEG2000), pattern recognition and deep packet inspection to powerful GPU co-processors, system architects can engineer solutions that exceed today’s processing requirements — with headroom for future application capability. Powered by NVIDIA® Quadro® TU104 GPUs, rugged GSC6204 co-processing modules deliver the efficient processing performance required for efficient compute off-loading.

NVIDIA Turing Architecture

These professional-grade NVIDIA Turing architecture processors feature 3072 CUDA cores to perform multiple accumulate floating-point operations (FLOPS). This delivers 10.9 TFLOPS of single-precision processing power per GPU. Both GPU processors are supported with 16 GB of GDDR6 (448 GB/s bandwidth). Collectively, each GSC6204 module simultaneously delivers 21.7 TFLOPS of single-precision processing power with almost 1 TB/s of memory bandwidth. Fabricated using a 12nm resolution process, each GPU processor is SWaP-efficient, even with their increased compute performance and capabilities.

For enhanced performance, each GPU features 384 Tensor Cores for 30 TFLOPS of mixed-precision matrix-multiply and accumulate calculations in single operations, which are a key processing attribute for AI, deep learning, and other signal processing and fusion workloads. Tensor cores and half-precision floating point (FP16) support a single GPU core that produces unmatched floating-point performance. For graphic acceleration, ray tracing (RT) hardware cores enable accelerated rendering performance and new functionality.

Similar to earlier Pascal GPUs, each Turing processor connects via a PCIe 3.0x16 interface to an onboard PCIe switch. However, unlike previous generation GPUs, data is now also sharable between GPUs using NVIDIA’s new NVlink™ high-speed bus, which adds an additional 50 GB/s of bandwidth between processors. This enables GSC6204 with their dual GPUs to function as a single VGPU with 6144 cores, 32GB of memory and a bandwidth of 896GB/s. The GSC6204 is the only OpenVPX module available today with this processing enhancement.
Composable PCIe HPEEC

Rugged Packaging
GSC6204 modules protect electronics, keeping them cool for long, reliable service lives and delivering consistent switch fabric performance across a broad temperature range. Our proven fifth generation of advanced packaging, cooling and interconnects technologies enable GSC6204 modules to be the first GPU powered OpenVPX co-processors that are truly rugged for applications requiring protection from harsh environments. For extreme environmental protection GSC6204 modules are available with optional MOTS+ technologies.

Optional MOTS+ for Extreme Environmental Protection
GSC6204 GPU co-processor modules are OpenVPX compliant and are optionally available with modified off the shelf plus (MOTS+) technology for extreme durability and environmental protection. MOTS+ configurations leverage enhanced commercial components, board fabrication rules, and subsystem design techniques for extra durability and the ability to withstand extreme temperature cycles better than other rugged designs.

Truly Composable Embedded Architecture
GSC6204 co-processors are the first trusted and truly rugged OpenVPX GPU co-processing engines and critical high-performance embedded edge computing (HPEEC) component. These modules, when integrated with other Mercury OpenVPX processing elements including Intel Xeon Scalable server blades, wideband PCIe switches and fast memory, form a truly composable datacenter architecture (Fig. 1) that is ready for deployment in the harshest and most SWaP-constrained industrial, aerospace and defense applications.

System Management
Each GSC6204 GPU co-processor module implements the advanced system management functionality architected in the OpenVPX standard to enable remote monitoring, alarm management, and hardware revision and health status.

Using the standard I2C bus, intelligent platform management controller (IPMC), and IPMI protocol, the onboard system-management block implementation is designed to comply with VITA 46.11. This allows each GSC6204 co-processor module to:

- Read sensor values
- Read and write sensor thresholds, enabling an application to react to thermal, voltage, or current variations that exceed those thresholds
- Reset the entire blade
- Power up/down the entire blade
- Retrieve blade field replaceable unit (FRU) information
- Be managed remotely by a chassis management controller at the system level

Open Software Environment
Mercury leverages over 35 years of multi-computing software expertise across our many platforms, including GSC6204 co-processor modules. The same Linux® development and run-time environment, and other off-the-shelf open software including OpenCL, NVIDIA CUDA, OFED and OpenMPI, are supported.

Extended Support
GSC6104 co-processing modules have extended 10+ year product support, are OpenVPX compliant and Sensor Open System Architecture (SOSA) aligned, making them ideal for technology refreshes.

Figure 1. Truly composable high performance embedded edge computing architecture
Trusted

GSC6104 modules are designed, manufactured, coded and supported in the USA from DMEA-certified facilities using devices from a managed supply chain.

Technical Specifications

**GPU**
- Two embedded NVIDIA Quadro TU104 GPU with the Turing architecture
- 6144 total processing cores (3072 lanes per GPU), 384 Tensor cores, 48 Ray Tracing cores
- 21.7 peak theoretical single-precision (FP32) TFLOPS
- 32 total PCIe 3.0 lanes (16 lanes per processor)
- 32 GB total GDDR6 Memory (16 GB per processor)
- 512-bit memory interface (256-bit per processor)
- 896 GB/s memory bandwidth (448 GB/s per processor)
- 4x display port display outputs (2 per GPU) to front panel (air-cooled only) and OpenVPX backplane
- 2x analog VGA display outputs (1 per processor) to front panel (air-cooled only) and OpenVPX backplane

**64-Lane Configurable PCIe Switch**
- Configurable switch enables multiple system-level configurations:
  - Non-transparent bridging and enumeration
  - x16 PCIe 3.0 connections to each MXM site (32 lanes total)
  - x32 PCIe 3.0 total connections to backplane
  - x16 PCIe 3.0 OpenVPX P2 expansion plane
  - x16 PCIe 3.0 OpenVPX P5 expansion plane

**IPMI (System Management)**
- On-board IPMI controller
- Voltage and temperature monitor
- Geographical address monitor
- Power/reset control
- Onboard CPLD, FRU EEPROM interfaces

**OpenVPX Multi-Plane Architecture**
- System management via IPMB-A and IPMB-B link on P0 management plane
- Dual full x16 or dual x8 PCIe on P2 and P5 expansion plane
- 4x display port display outputs on P6 mezzanine I/O plane
- 2x analog VGA outputs on P3 mezzanine I/O plane

![Figure 2. GSC6204 functional block diagram](image)
Mechanical and Open Systems Architecture

- OpenVPX (VITA 65) and VPX-REDI
- VITA 46.0, 46.3, 46.4, 46.6, 46.11, and VITA 48.1, 48.2, 48.4, 48.7 (REDI)
- VITA 65 module profile MOD6-PAY-4F1Q2U2T-12.2.1-n (where n can vary based on ConnectX-5 configuration)
- 1.0” pitch, single-slot
- Cooling options: air-cooled, air flow-by, conduction-cooled, or liquid flow-through
- Optimal MOTS+ extreme environmental protection packaging
- Optional SOSA-aligned configurations

Refer to separate publication “Rugged Embedded Packaging & Next Generation Cooling” for environmental performance.

Power Consumption

- Typically 170W per accelerator, power is configurable