VPX6U-BW5000E-DUAL-CX7

Dual NVIDIA Blackwell, 10496 CUDA Cores per GPU, ConnectX-7 and 100 GbE

Key Features

- 6U VPX Blackwell platform with two NVIDIA RTX™ 5000 (GB203) GPUs
  - 10,496 CUDA Cores and 320 Tensor Cores per GPU
- 24 GB GDDR7 256-bit ECC VRAM per GPU
- ConnectX-7 and a configurable PCIe Gen4 switch
- Up to 8 DisplayPort outputs, options for HDMI
- Module power: 210W to 350W, configurable

GPU Features

- Blackwell GPU parallel processing support:
  - CUDA® Toolkit 13, Compute Capability 12.0
  - CUDA-X AI and CUDA-X HPEC libraries
  - OpenCL™ 3.0, DirectX® 12 Ultimate, OpenGL 4.6, OpenGL ES 3.2, Vulkan™ 1.4
- 320 Tensor Cores (5th Gen)
- 80 Ray Tracing Cores (4th Gen)
- NVENC (9th Gen) and NVDEC (6th Gen) with H.264/H.265/AV1 support

Connectivity / System Management

- IPMI system management – Tier 3
- NVIDIA ConnectX-7 with GPUDirect RDMA & RoCE
- Configurable PCIe Gen4 switch

Mechanical / Open Systems Architecture

- High level of ruggedization:
  - Rugged conduction cooled (CC), air cooled (AC), or liquid flow through (LFT)
  - Operating temperature:
    CC: -40°C to +70°C standard
    AC: -40° to +60°C standard
  - Vibration (sine wave): 10G peak, 5 - 2000Hz
  - Shock: 40G peak for conduction cooled, 30G peak for air cooled
- Weight: ~AC: 1.5 kg, ~CC: 2.5 kg, ~LFT: 1.7 kg
- ANSI/VITA 48, and VITA 65 (VPX-REDI, OpenVPX) 10.2.7 Payload Slot Profile (User Defined for video)
- OpenVPX and SOSA Aligned profile 10.6.4, 10.6.3 Payload Slot Profile (P3/P6 depopulated no rear video)

Overview

The VPX6U-BW5000E-DUAL-VO-CX7 combines two NVIDIA RTX™ 5000 Blackwell embedded GPUs, an NVIDIA ConnectX-7 EN, and a configurable PCIe Gen4 switch in a rugged 6U VPX module. The module provides high-performance embedded computing (HPEC), artificial intelligence (AI), high-bandwidth networking, and sensor-processing capability in a single VPX payload.
Each NVIDIA RTX™ 5000 embedded GPU connects to the on-board PCIe Gen4 switch through a PCIe x16 interface and includes 24 GB of GDDR7 ECC memory. The ConnectX-7 EN is connected to the PCIe switch through two PCIe x16 interfaces and provides two 100G KR4 Ethernet interfaces, each using four lanes, routed through VPX connector P1. This architecture provides high-bandwidth data paths between the GPUs, network interface, VPX PCIe fabric and supports NVIDIA GPUDirect RDMA and RoCE.

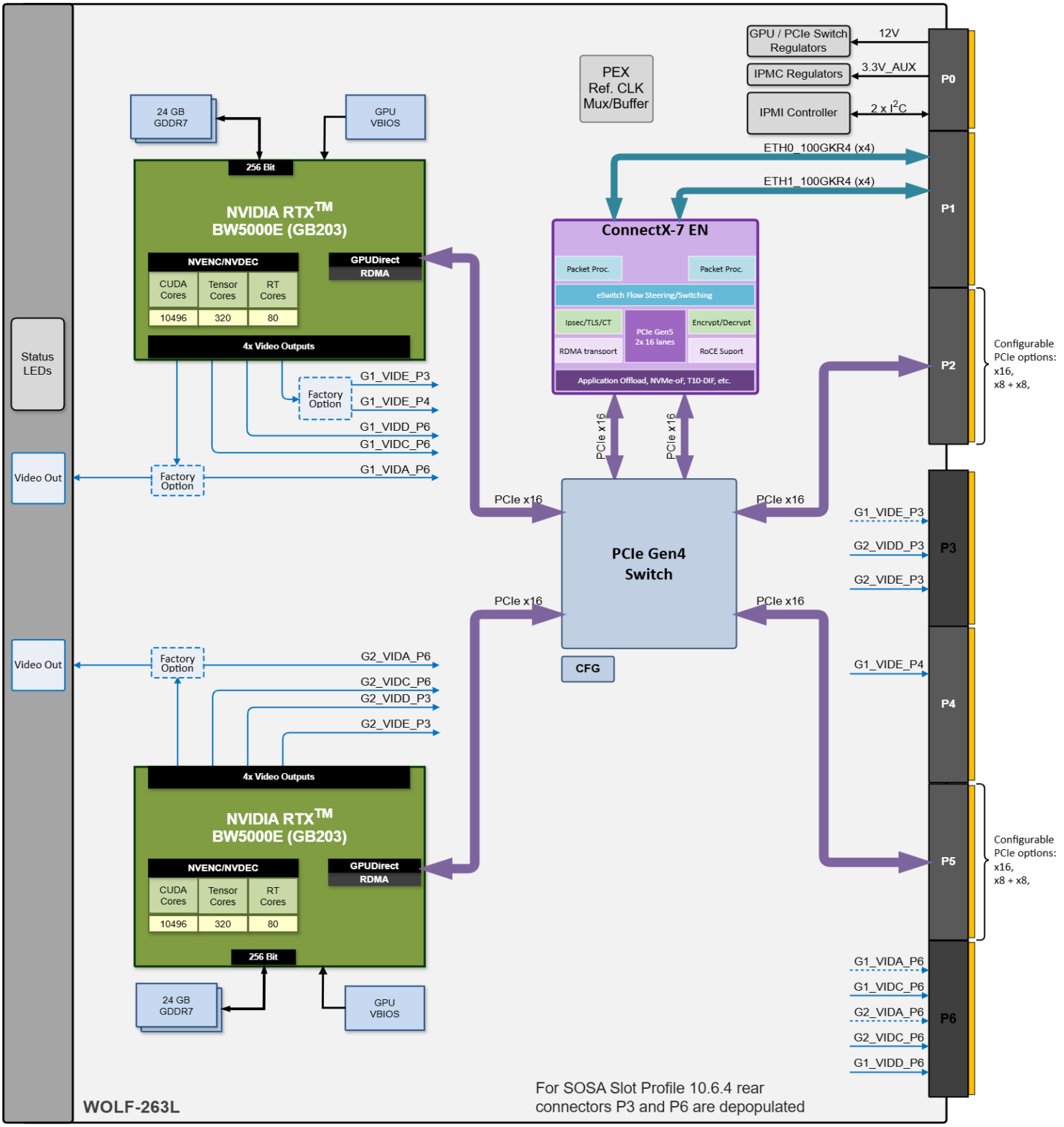
The NVIDIA Blackwell architecture includes CUDA® cores and 5th-generation Tensor Cores for AI and HPC. The improved architecture increases efficiency and supports GDDR7 memory, providing over 50% higher bandwidth than the previous generation. The GPUs connect to the module's PCIe Gen4 fabric for high-bandwidth data paths to the host, SmartNIC, and on-board PCIe devices.

Unlocking this performance requires robust cooling that can be tailored per deployment; conduction-cooled (CC), air cooled (AC), and liquid flow through (LFT) variants, using a low-weight, high-efficiency path from the GPU die to the wedgelocks

NVIDIA RTX

Information Subject to Change

VPX6U-BW5000E-DUAL-CX7



Power and Performance

An NVIDIA GPU clock speed is dependent on the TGP (total GPU subsystem power) and the GPU temperature. The highest clock speeds are available at the highest TGP power allowed by the GPU. When the TGP setting is decreased the clock speed will also decrease resulting in a decrease in processing speed. If the GPU temperature exceeds 87°C the GPU clock speed will also decrease to protect the GPU from heat damage. If the GPU temperature is below 86.5°C the GPU can operate at maximum boost clock speeds at the currently available power when the GPU detects that higher processing is required.

Each Blackwell GB203 GPU in this 6U VPX module defaults to a TGP of 100 W per GPU. At 100 W, the base clock of 1125 MHz provides up to 23.6 TFLOPS per GPU, while at higher GPU loads the boost clock can reach up to 1792 MHz, providing up to 37.6 TFLOPS per GPU. A higher TGP of up to 150 W per GPU can be configured when sufficient cooling is available, enabling a maximum boost clock of 2370 MHz and performance of up to 50 TFLOPS per GPU.

NVIDIA Blackwell GPU

NVIDIA Blackwell GPUs have improved architecture, which provides increased efficiency. Blackwell GPUs have CUDA cores that can all handle either FP or INT operations, whereas previous generations restricted half of the CUDA cores for FP operations only. This is important for tasks that require lots of small, frequent lookups resulting in frequent address computations, which are commonly required by the matrix operations used by AI and HPC calculations. Optimizing these lookups ensures the data can reach the processing cores efficiently, permitting full use of the processing power. The new Blackwell architecture also supports GDDR7 memory, which provides 55% more memory bandwidth. With the increased performance and memory handling abilities, and improved next-generation Tensor cores, the Blackwell GPUs provide significant performance increases compared to the previous generation.

Tensor Cores for Artificial Intelligence and HPC

Tensor Cores are designed to speed up the tensor / matrix computations used for deep learning neural network training and inferencing operations. NVIDIA Blackwell architecture GPUs include the fifth-generation Tensor Core design which supports many data types for improved performance, efficiency, and programming flexibility, including support for new INT4 and INT6 precision modes and micro-scaling formats. NVIDIA provides CUDA-X AI and CUDA-X HPEC libraries designed to work with NVIDIA GPUs to provide the tools needed to accelerate development of applications for AI and HPEC.

Hardware Accelerated Video Encode / Decode

The Blackwell GPU includes the NVENC video encode and NVDEC decode hardware acceleration engine. Using the GPU for video encoding provides an efficient, high quality method to achieve real time 8K and 4K encoding without burdening the system CPU. The GPU also supports hardware-accelerated AV1 encoding and decoding. The NVIDIA Video Codec SDK provides a complete set of APIs, samples and documentation for hardware accelerated video encode and decode.

SOSA and OpenVPX Slot Profile Support

This module's configurable switch provides support for SOSA aligned and OpenVPX slot profiles. The module can be configured with pin mappings that are compatible with previous generation WOLF-2538/2348 modules, which allows the WOLF-263L to be a plug-in upgrade for previous WOLF products. The following SOSA aligned profile is supported:

- 10.6.3 and 10.6.4 Payload Slot Profile (P3 and P6 depopulated)