VPX3U-RTX5000E-SWITCH

WOLF ADVANCED TECHNOLOGY

NVIDIA Turing HPEC, PCIe Gen4 Switch, 10GigE Switch

KEY FEATURES

- NVIDIA Quadro Turing TU104 GPU provides 10.9 TFLOPS peak performance* of HPEC processing
- PCIe Gen4 Switch
- 10 Gig Ethernet Switch
- Module power: 60-160W
  *Peak performance requires the highest power configuration mode

ADDITIONAL FEATURES

- Turing GPGPU parallel processing:
  - □ 3072 CUDA® cores
  - □ CUDA Toolkit 10, CUDA Compute capability 7.5
  - □ OpenCL™ 1.2, DirectX® 12, OpenGL 4.6, OpenGL ES 3.2, Vulkan™ 1.0
- 384 Turing Cores for advanced AI inference processing
- 16 GB GDDR6 256-bit memory with up to 448 GB/s
- GDDR6 memory provides twice the bandwidth of the previous generation of GDDR5 memory
- HEVC (H.265) and AVC (H.264) Turing NVENC and NVDEC hardware acceleration with up to 8K encode resolution and B-frame support
- Optional Fiber to PCIe interface
- IPMI Controller for module management
- Windows and Linux drivers

SPECIFICATIONS

- High level of ruggedization:
  - □ Rugged Conduction-cooled
  - □ Operating temperature: -40° to +85°C
  - □ Vibration (sine wave): 10G peak, 5 - 2000Hz
  - □ Shock: 40G peak for conduction-cooled
- Dimensions: 160mm x 100mm x 25.4mm
- Weight: To be Announced
- +12V only power source (as per SOSA requirements)
- SOSA™ Aligned with SOSA switch slot profile supported
- ANSI/VITA 48, 65 (VPX-REDI, OpenVPX)

OVERVIEW

The VPX3U-RTX5000E-SWITCH module includes an NVIDIA® Quadro® Turing™ RTX5000 GPU for high performance embedded computing (HPEC), a PCIe Gen4 switch, and a 10Gig Ethernet switch, all on one slot saving module.

Adding an RTX5000 GPU to a switch module provides system integrators with a novel way to increase the processing capability of their system without using an additional slot. Data can be sent to/from the GPU via PCIe, providing access to the Turing GPU's advanced parallel processing and AI inference processing capabilities.

The module includes a 10Gig Ethernet switch, providing fast data movement from the module over the system backplane. An optional fiber to PCIe interface is also available.

The default profile for this module is the SOSA switch profile (SLT3-SWH-6F8U-14.4.15). OpenVPX profiles are also supported to provide flexible integration options.

Unlocking the best performance requires the best cooling capability. WOLF's advanced cooling technology is designed to move heat using a low weight, high efficiency pipeline from the GPU die to the wedgelocks.

NVIDIA® QUADRO
CERTIFIED
▪ EMBEDDED

DESIGNED FOR SYSTEM INTEGRATION

The VPX architecture is diverse, spanning custom backplanes, an ambiguous system specification and differing input and output methodologies. That is precisely why WOLF modules come with factory configuration options to solve virtually all system integration challenges.

This 3U VPX module includes a new Interface Control Document (ICD) that includes factory configurable options which can be used to maintain compliance with many Open COTS initiatives. By default the module has been designed to support SOSA switch slot profile (SLT3-SWH-6F8U-14.4.15).

This module or others from the same family can also be modified to support OpenVPX (VITA 65) profiles. Please contact WOLF to discuss your system level requirements.
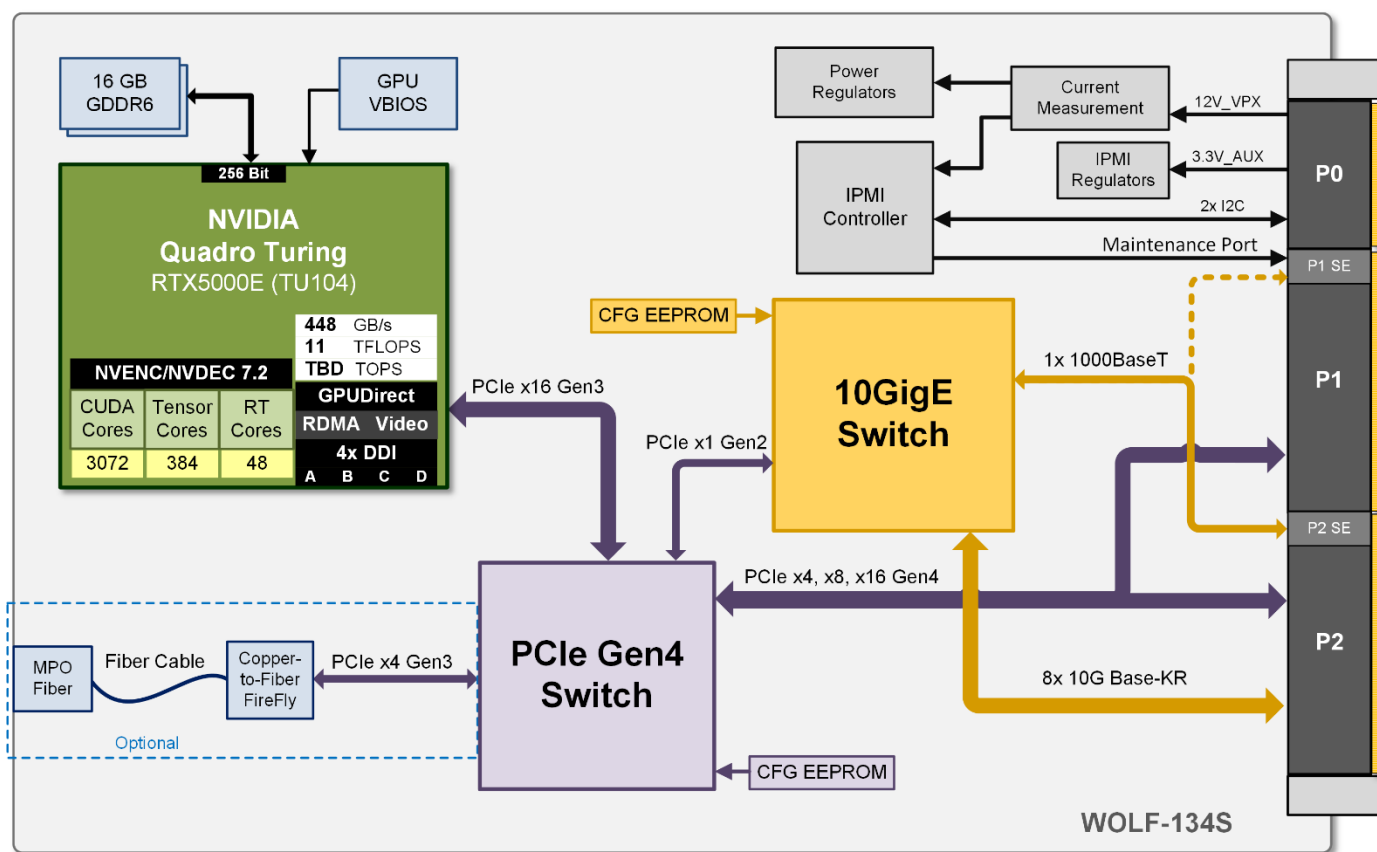
— — — Optional Fiber Interface

— — — Optional routing for OpenVPX slot profiles

NVIDIA TURING STREAMING MULTIPROCESSOR (SM)

The NVIDIA Turing architecture provides a 50% improvement in delivered performance per CUDA core compared to the Pascal generation. This is due to the new Turing Streaming Multiprocessor's independent integer datapath, allowing execution of concurrent integer and floating-point instructions, and the redesigned memory path which provides two times the bandwidth and more than two times the capacity for common workloads.

NVIDIA 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. Turing GPUs include a new version of the Tensor Core design that has been enhanced for inferencing. Turing Tensor Cores add new INT8 and INT4 precision modes for inferencing workloads that can tolerate quantization and don't require FP16 precision.

NVIDIA provides CUDA-X AI and CUDA-X HPC libraires which are specialized libraires built on top of CUDA. They have been designed to work with NVIDIA Tensor Core GPUs to provide the tools needed to accelerate

FAST GDDR6 MEMORY

Getting data into and out of a high performance GPU requires fast graphics memory to ensure that the memory does not become a system bottleneck. In moving from GDDR5 to GDDR6 the number of data transfers per clock cycle doubled from two to four, and memory chips can be read in dual-channel modes rather than just single channel modes. The newer GDDR6 memory does all of this while also slightly reducing the memory's average power consumption compared to using GDDR5 memory.

NVIDIA also uses memory compression technology, especially data color compression for reducing the amount of graphical information that needs to be transmitted. With Turing and the GDDR6 memory the latest generation compression technology provides a 20 to 30% memory compression efficiency increase.

HARDWARE ACCELERATED VIDEO ENCODE / DECODE

The RTX5000E chip includes the latest generation video encode/decode hardware acceleration engine (version 7.2). This adds support for HEVC (H.265) 8K encoding at 30 fps and B-Frame support. It also provides up to 25% bitrate savings for HEVC and up to 15% bitrate savings for AVC (H.264). Using the Turing encoding engine for video encoding provides an efficient, high quality method to achieve real time 8K and 4K encoding without burdening the system CPU.

As with previous versions of the encoding engine, NVENC supports CBR and VBR rate control, programmable intra-refresh for error resiliency, and a motion estimation (ME) only mode. The NVIDIA Video Codec SDK provides a complete set of APIs, samples and documentation for hardware accelerated video encode and decode on Windows and Linux.

ORDERING CODES

The following table defines series of common order codes for the VPX3U-RTX5000E-VO module. The asterisks denote characters of the part number that are defined based on common configuration options. Some common configuration options for this module are:

- Slot Profile
- Conformal Coating Type
- Default Power Threshold
- COTS, MCOTS or Locked

	Ordering Number	Description
	3U VPX Turing RTX5000, PCIe Gen4, 10GigE, Single Slot Configurations	
	134S33-F**0*VPX3v10	3U VPX, Conduction Cooled, 1", NVIDIA RTX5000, PCIe Gen4, 10GigE, no Fiber interface
	134S33-F**1*VPX3v10	3U VPX, Conduction Cooled, 1", NVIDIA RTX5000, PCIe Gen4, 10GigE, includes optional Fiber interface

Contact Sales for the latest Ordering Numbers and available options

MANUFACTURING AND QUALITY ASSURANCE

WOLF designs modules to pass the following environmental standards:

- MIL-STD-810 (United States Military Standard for Environmental Engineering Considerations and Laboratory Tests)
- MIL-HDBK-217 (Reliability Prediction of Electronic Equipment)
- RTCA DO-160 (Environmental Conditions and Test Procedures for Airborne Equipment) on request

WOLF complies with the following management systems:

- AS9100D: Quality Management System - Requirements for Aviation, Space and Defense Organizations (certified)
- ISO 9001:2015: Quality management systems (certified)
- AS5553: Counterfeit Electronic Parts; Avoidance, Detection, Mitigation, and Disposition (compliant)
- NIST SP 800-171: Protecting Controlled Unclassified Information in Nonfederal Systems (compliant)

Boards are manufactured to meet the following standards:

- IPC-A-610 CLASS 3 (Acceptability of Electronic Assemblies)
- IPC 6012 CLASS 3 (Qualification and Performance Specification for Rigid Printed Boards, Class 3 for High Reliability Electronic Products)
- IPC J-STD-001 (Requirements for Soldered Electrical and Electronic Assemblies)