

NVIDIA Turing 11 TFLOPS, WOLF FGX, Coaxial Connections

PRELIMINARY INFORMATION

KEY FEATURES

- NVIDIA Quadro Turing TU104 GPU, 10.9 TFLOPS peak performance*, 3072 CUDA Cores, 384 Tensor Cores
- SOSA™ Aligned with support for SOSA slot profile
- WOLF FGX provides support for additional video formats such as SDI and Analog
- Module power: configurable from 50 to 150W
*Peak performance requires the highest power configuration mode

ADDITIONAL FEATURES

- Four DisplayPort 1.4 digital video outputs:
 - Support for High Dynamic Range (HDR) video
 - 4K at 120Hz or 8K at 60Hz with 10-bit color depth
 - HDMI and DVI options
- Two 3G/HD-SDI and two CVBS/NTSC/PAL outputs
- 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
- 48 Ray Tracing Cores for advanced rendering
- 16 GB GDDR6 256-bit memory with up to 448 GB/s
- HEVC (H.265) and AVC (H.264) Turing NVENC and NVDEC hardware acceleration with up to 8K encode resolution and B-frame support
- PCI Express x16
- Daisy Chain and NTB Options Available
- 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
- Dimensions: 160mm x 100mm x 25.4mm
- Weight: To be Announced
- +12V Only Power Source (as per SOSA requirements)
- SOSA Aligned with SOSA payload slot profile and VITA 67.3 connector with coaxial connections on P2
- ANSI/VITA 48 (VPX-REDI), OpenVPX compatible

OVERVIEW

The VPX3U-RTX5000E-COAX-CV module includes both an NVIDIA® Quadro® Turing™ RTX5000 embedded GPU and a WOLF FGX in a rugged 3U VPX module. The RTX5000 includes CUDA cores for parallel processing, Tensor cores for dedicated AI inference and ray tracing cores for superior rendering speeds. The WOLF FGX provides video conversion to formats which are not native to the GPU, such as SDI and analog formats. The module is SOSA™ Aligned with support for the SOSA payload slot profile.

Getting data into and out of the module is an important consideration for this high-speed GPU. Support for GDDR6 memory provides twice the bandwidth of the previous generation's GDDR5 memory.

The Turing GPU with its Tensor cores provides this module with the underlying architecture required for an efficient AI inference engine. Intended to work in conjunction with TensorRT, CUDA and CuDNN, the Turing Tensor Core design adds INT8 and INT4 matrix operations, while continuing to support FP16 for higher precision workloads.

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.



This datasheet is preliminary and is subject to change

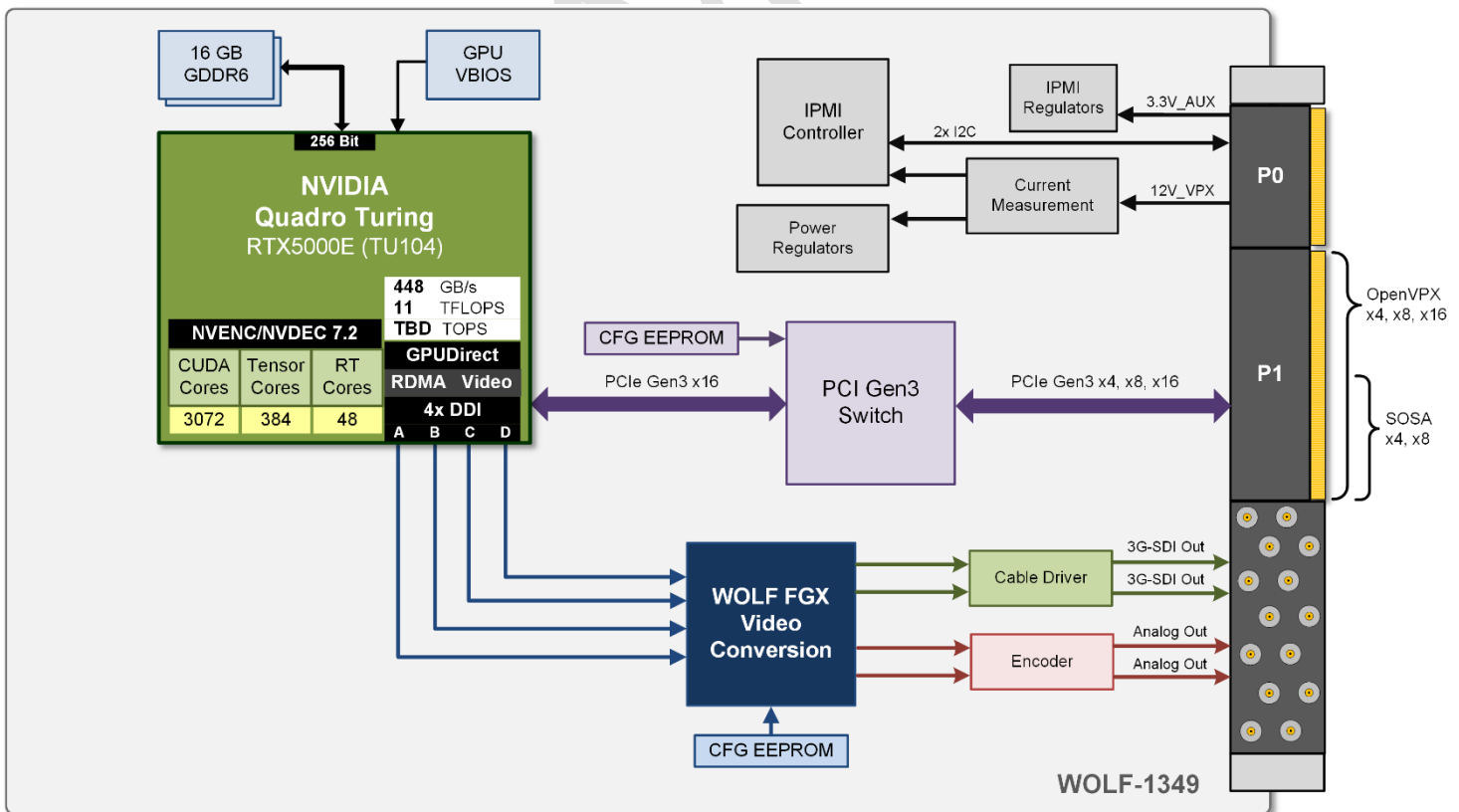
NVIDIA TURING STREAMING MULTIPROCESSOR (SM)

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.

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.



This datasheet is preliminary and is subject to change

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 development of applications for AI and HPC.

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.

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 WOLF module has been designed to include a WOLF FGX to support additional video output protocols that are not native to the Turing GPU. Through MCOTS services, the number of outputs and the protocols used can be changed to meet the requirements of many different system architectures.

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 Payload slot profile SLT3-PAY-1F1U1S1S1U1U2F1H-14.6.11-n.

This module or others from the same family can also be modified to support OpenVPX profiles. Please contact WOLF to discuss your system level requirements.

PRELIMINARY INFORMATION

ORDERING CODES

The following table defines series of common order codes for the VPX3U-RTX5000E-CV 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:

- Display Interfaces
- Default Power Threshold
- COTS, MCOTS or Locked
- Conformal Coating Type
- Converted video formats

Ordering Number	Description
3U VPX Turing RTX5000 Single Slot Configurations	
134933-F***-***VA0	3U VPX, Conduction Cooled, 1", NVIDIA RTX5000, WOLF FGX, VITA 67.3 connector with coaxial connections on P2, 2x SDI and 2x CVBS outputs

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)



This datasheet is preliminary and is subject to change