# NVIDIA Jetson Thor Series

## The Ultimate Platform for Physical AI and Robotics

Technical Brief

Aayush Pathak

**TB-12572-001** | August 2025

# Document History

**TB-12572-001**

| Version | Date | Authors | Description of Change |
|---|---|---|---|
| 1.0 | August 29, 2025 | Aayush Pathak | Initial release |

# Table of Contents


NVIDIA Jetson Thor Series - The Ultimate Platform for Physical AI and Robotics ..... 5
- Generative reasoning and multimodal sensor processing ..... 7
- Next Level of Gen AI Inferencing with FP4 and Speculative Decoding ..... 9

Jetson Thor Series Hardware Architecture ..... 11
- GPU ..... 14
- Multi-Instance GPU (MIG) ..... 15
- NVIDIA 5th Generation Tensor Cores ..... 16
- NVIDIA Transformer Engine ..... 16
- Memory ..... 17
- CPU ..... 18
- Video Codecs ..... 19
- Programable Vision Accelerator ..... 20
- Video Imaging Compositor ..... 21
- Optical Flow Accelerator ..... 21
- Vision Programming Interface ..... 22
- I/O ..... 23
- Power Profiles ..... 24

Safety ..... 25

Jetson Software ..... 27
- JetPack 7 ..... 28
- Get the most out of Blackwell GPU using NVIDIA Software Libraries ..... 29
- Supercharge humanoid robot development with NVIDIA GR00T workflows. ..... 30
- Visual Search and Summarization (VSS) ..... 32
- Video Search and Summarization Agent ..... 34
- Holoscan Sensor Bridge ..... 36

Jetson AGX Thor Developer Kit ..... 39


# List of Figures


Figure 1 NVIDIA Jetson Thor Module ..... 6
Figure 2 Jetson Thor Pillars ..... 7
Figure 3 Jetson Thor is up to 5x faster in generative reasoning compared to Jetson Orin ..... 8
Figure 4 Performance when handling multiple models and multimodal sensor inputs ..... 9

Figure 5 Jetson Thor Performance with Speculative Decoding ..... 10
Figure 6 Thor System-on-Chip Block Diagram ..... 12
Figure 7 Jetson Thor System-on-Module ..... 13
Figure 8 Jetson Thor functional block diagram ..... 18
Figure 9 CPU Complex ..... 19
Figure 10 Jetson Thor PVA Block Diagram ..... 21
Figure 11 Stereo Disparity Estimation Pipeline ..... 23
Figure 12 Jetson Software Stack ..... 27
Figure 13 JetPack7 Pillars ..... 28
Figure 14 Humanoid Sensors and AI Processing ..... 31
Figure 15 VSS Blueprint Architecture ..... 35
Figure 16 Holoscan Sensor Bridge IP Architecture ..... 37
Figure 17 Holoscan Sensor Bridge Reference Design Architecture ..... 38
Figure 18 NVIDIA Jetson AGX Thor Developer Kit ..... 39

# List of Tables

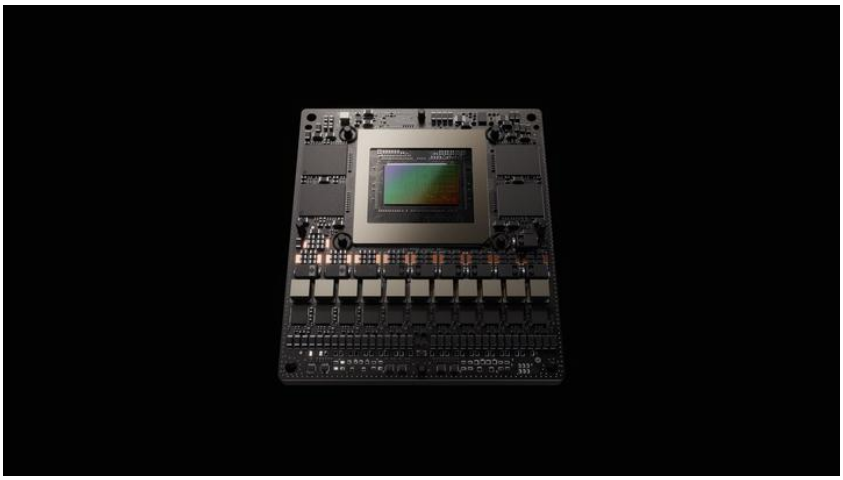
Table 1 Jetson Thor Series Technical Specifications ..... 13
Table 2 Jetson Thor Series Performance ..... 15
Table 3 Jetson T5000 I/O ..... 24

# NVIDIA Jetson Thor Series - The Ultimate Platform for Physical AI and Robotics

AI is stepping out of the digital realm and into the real world. Physical AI marks this breakthrough—the point where intelligence doesn't just analyze data but senses, moves, and reacts. No longer tied to static routines, systems can now see, understand, and handle the unexpected, adapting right beside us in workplaces, hospitals, and homes. This shift means machines are engaging directly with life's changing rhythms, learning from surroundings, and working with people in practical, approachable ways.

The NVIDIA® Jetson Thor™ is the ultimate platform for physical AI and robotics, offering unmatched performance and scalability. Designed to run large foundational models such as GR00T-N1 directly at the edge, Thor platform enables real-time perception, reasoning, and decision-making in complex environments. It represents a critical shift from task-specific and single-purpose automation to multi-function and general-purpose intelligence—laying the groundwork for more capable, autonomous, and versatile robotic systems. With Jetson Thor, robots no longer need to be reprogrammed for each new job. Jetson Thor is a super-computer for generative reasoning and multi modal multi sensor processing.

Figure 1 NVIDIA Jetson Thor Module

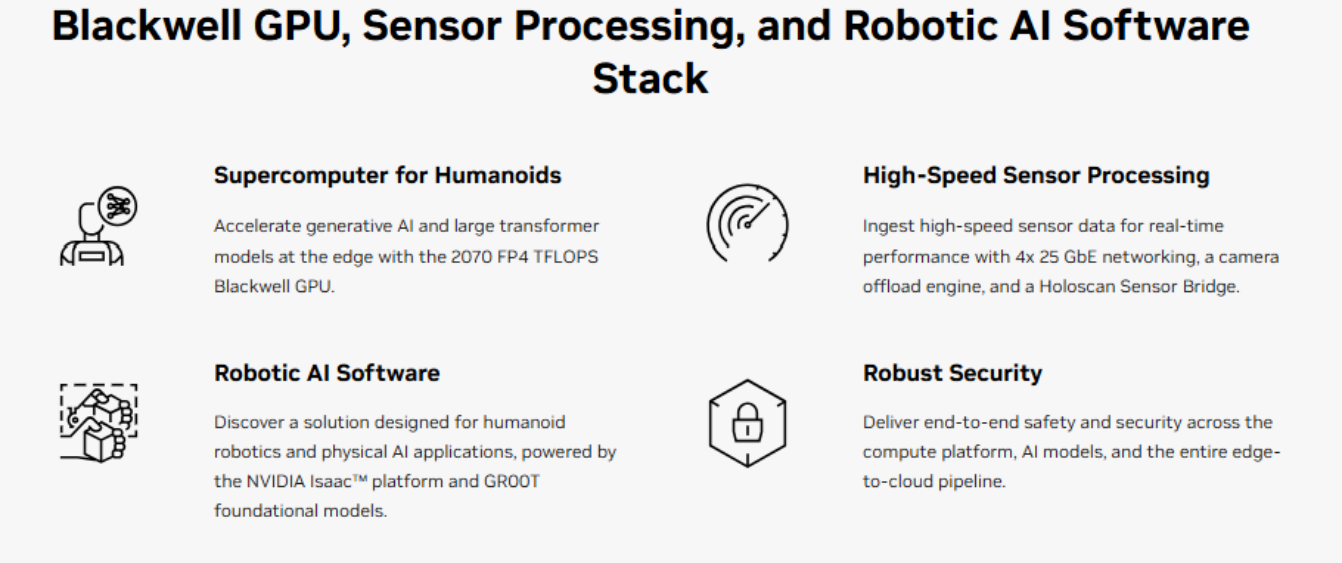


NVIDIA Thor series is a new class of high-performance platform purpose-built to support the next generation of Physical AI with a multi-faceted approach to safety. Powered by the NVIDIA Blackwell GPU and up to 128 GB of memory, it delivers up to 2 PetaFLOPS of FP4 AI compute to effortlessly run the latest generative AI models— all within a 130 W power envelope. Compared to NVIDIA Jetson AGX Orin™, it provides up to 7.5x higher AI compute and 3.5x better energy efficiency.

Jetson Thor accelerates low-latency, real-time applications with the new Blackwell Multi-Instance GPU (MIG) technology and a robust 14-core Arm® Neoverse®-V3AE CPU. It also includes a suite of accelerators, such as a third-generation Programmable Vision Accelerator (PVA), dual encoders and decoders, an optical flow accelerator, and more. For high-speed sensor fusion, the developer kit offers extensive I/O options, including a QSFP28 slot with 4x25GbE, a wired Multi-GbE RJ45 connector, multiple USB ports, and additional connectivity interfaces.

To deliver a seamless cloud-to-edge experience, Jetson Thor runs the NVIDIA AI software stack for physical AI applications, including NVIDIA Isaac™ for robotics, NVIDIA Metropolis™ for visual agentic AI, and NVIDIA Holoscan™ for sensor processing. Users can also build AI agents at the edge using NVIDIA's AI Blueprint for Video Search and Summarization (VSS) with Cosmos Reason as the VLM.

**Figure 2** Jetson Thor Pillars




NVIDIA's tremendous ecosystem of partners offers all the carrier boards, design services, cameras, and other sensors needed, as well as additional AI and system software to accelerate solution development in industries such as robotics, smart spaces, retail, industrial, medical, and more.

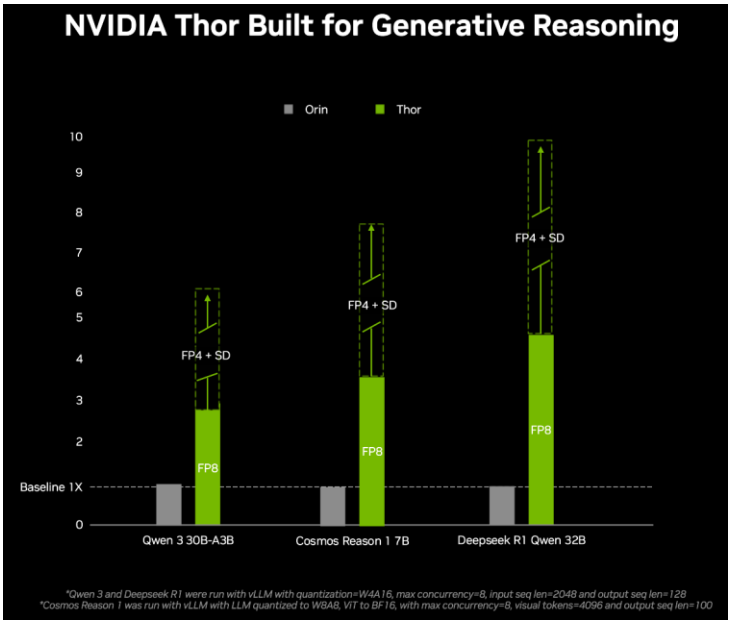
The Jetson AGX Thor Developer Kit provides massive AI performance and sensor capabilities in a compact form factor. This makes it the ideal platform for developers looking to unlock new possibilities for humanoid robotics and other physical AI applications. It is designed for seamless integration with existing humanoid robot platforms by allowing an easy tethering to jumpstart prototyping.

## Generative reasoning and multimodal sensor processing

Generative reasoning models are crucial for robotics platforms that can simulate possible sequences of actions, anticipate consequences, reason over language or visual cues, and flexibly generate high-level plans or low-level motion policies. This leads to robotic systems that are significantly more flexible, adaptable, and capable of robust, human-level reasoning in real-world settings.

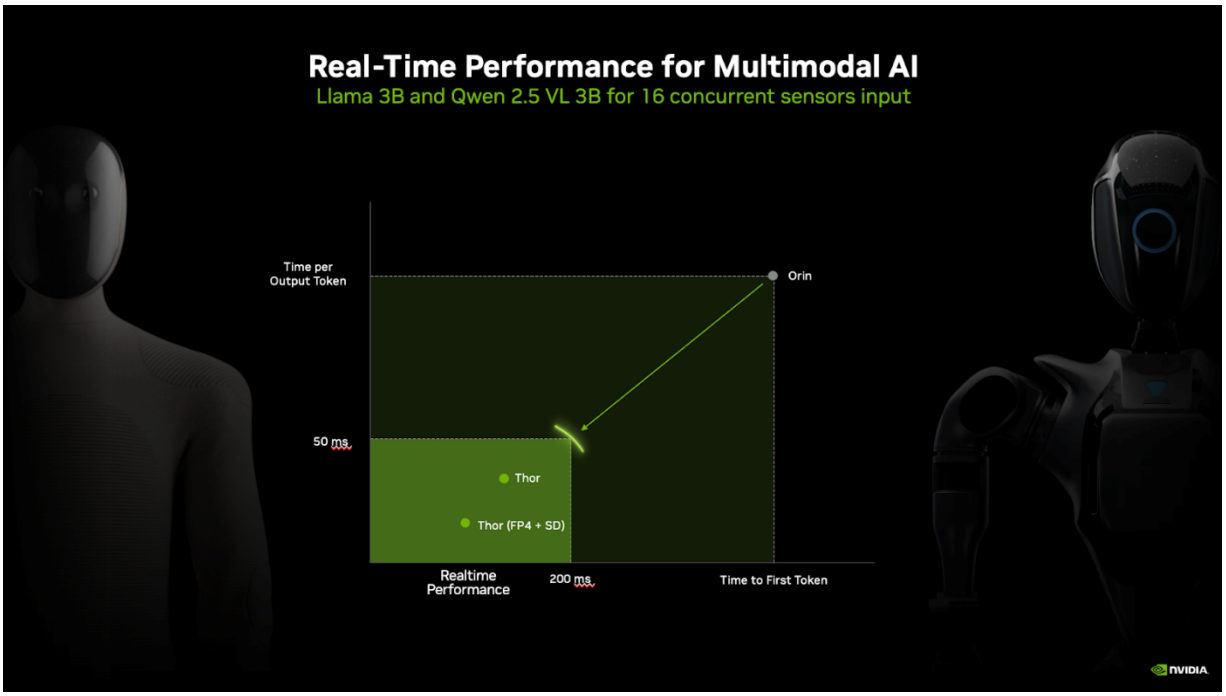
NVIDIA Jetson Thor delivers a giant leap in generative reasoning by providing speedups of up to 5x compared to Jetson Orin. With FP4 and speculative decoding, developers will be able to achieve an additional 2x performance speedup on Jetson Thor.

**Figure 3** Jetson Thor is up to 5x faster in generative reasoning compared to Jetson Orin




Jetson Thor also seamlessly handles multiple generative AI models and a large number of multimodal sensor inputs, delivering real-time responsiveness. Figure 5 demonstrates this capability using the Qwen2.5-VL-3B VLM and Llama 3.2 3B LLM, handling 16 simultaneous requests. Both models achieve a Time to First Token (TTFT) well below 200 milliseconds and a Time per Output Token (TPOT) well below 50 milliseconds—two key indicators of system responsiveness.

**Figure 4** Performance when handling multiple models and multimodal sensor inputs




## Next Level of Gen AI Inferencing with FP4 and Speculative Decoding

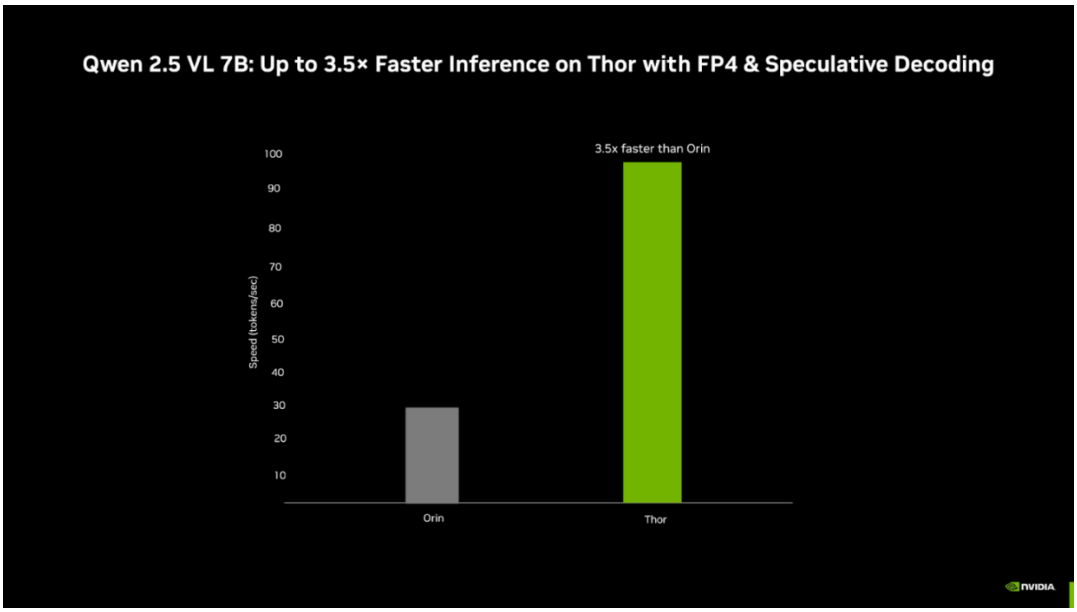
FP4 is natively supported on Jetson Thor in the 5th-gen Tensor Cores, letting core transformer layers run at up to multiple-x the speed of FP8 while cutting DRAM and L2 traffic, which directly translates to higher token, frame, and batch throughput at the edge. This low-precision path compounds with Blackwell's structured sparsity and L2/DRAM compression to further accelerate GEMMs and attention while keeping latency low, enabling real-time multimodal and robotics inference making FP4 the key to fitting larger models, increasing stream concurrency, and achieving rapid, accurate inference in embedded deployments.

Speculative decoding accelerates edge LLM inference by pairing a small, fast "draft" model with a larger "target" model to propose several tokens and verify them in one pass, which reduces the number of expensive target forward steps and can deliver 3× or more throughput gains at low latency on Jetson Thor running TensorRT-LLM. On edge, this draft-then-verify

workflow is especially effective because it amortizes verification over multiple tokens, better utilizes limited power budgets, and maintains output quality as long as the draft model is both faster and sufficiently aligned to the target to keep acceptance rates high.

The graph below demonstrates that the Qwen 2.5 VL 7B model achieves up to 3.5× faster inference on Thor, with FP4 quantization and Eagle-based speculative decoding, compared to Jetson Orin operating with W4A16 (4-bit weights, 16-bit activations).

**Figure 5** Jetson Thor Performance with Speculative Decoding



# Jetson Thor Series Hardware Architecture

The NVIDIA® Jetson Thor™ provides unmatched performance and scalability, delivering up to 2070 FP4 TFLOPS of AI performance for powering humanoids and autonomous systems. The Jetson Thor series includes the Jetson T5000 and the Jetson T4000 modules. Jetson Thor modules feature the NVIDIA Thor SoC with the NVIDIA Blackwell architecture GPUs, Arm® Cortex® Neoverse CPUs, next-generation vision accelerators, video encoders and decoders, and a new camera offload engine to boost AI performance.

High speed IOs and up to 128GB of DRAM with 273 GB/s of memory bandwidth, enables these modules to feed multiple concurrent AI application pipelines. With the SOM design, NVIDIA has done the heavy lifting of designing around the SoC to provide not only the compute and I/O but also the power and memory design. For more details, reference the Jetson Thor Datasheet.

**Figure 6** Thor System-on-Chip Block Diagram

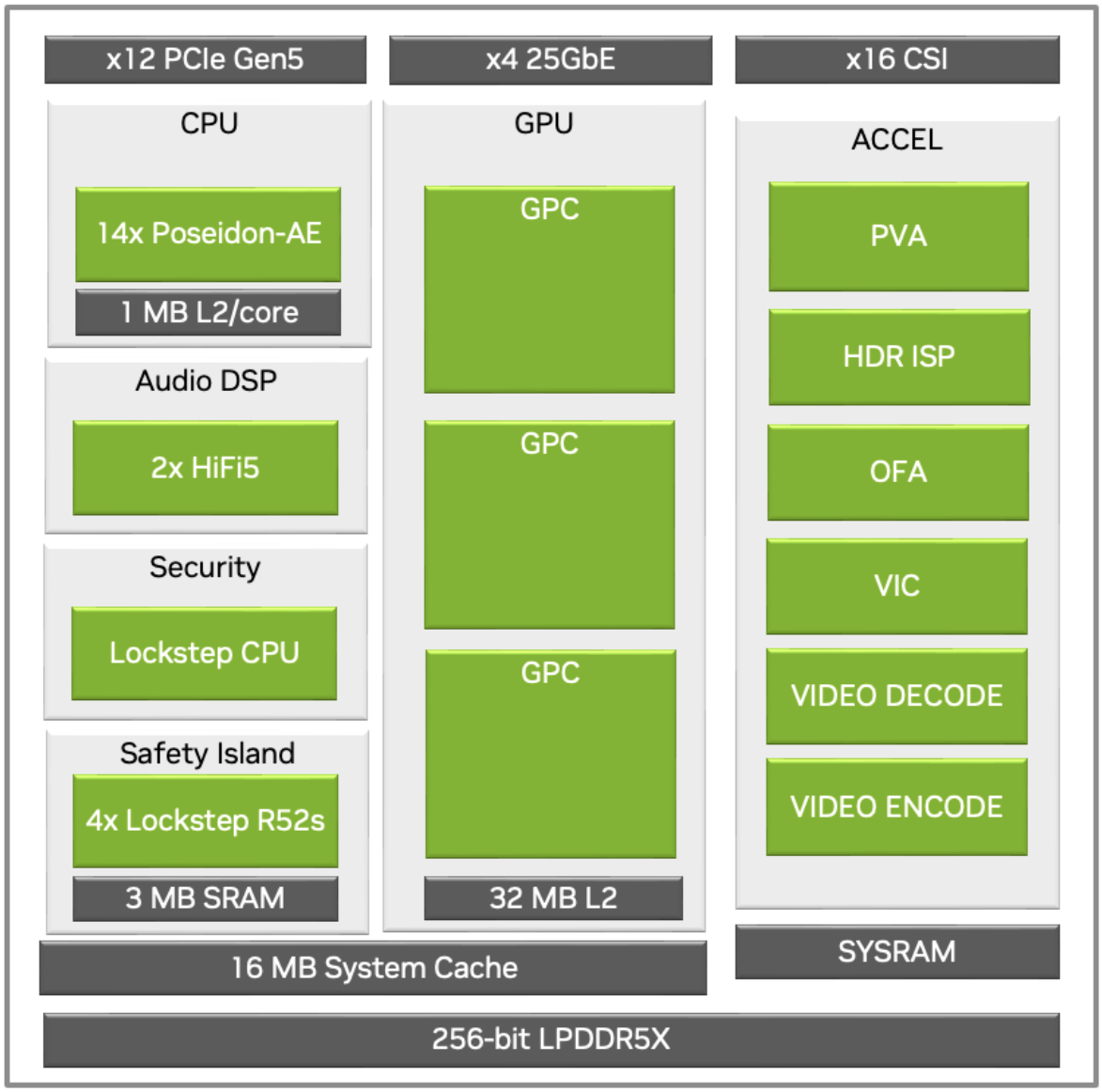


**Figure 7** Jetson Thor System-on-Module

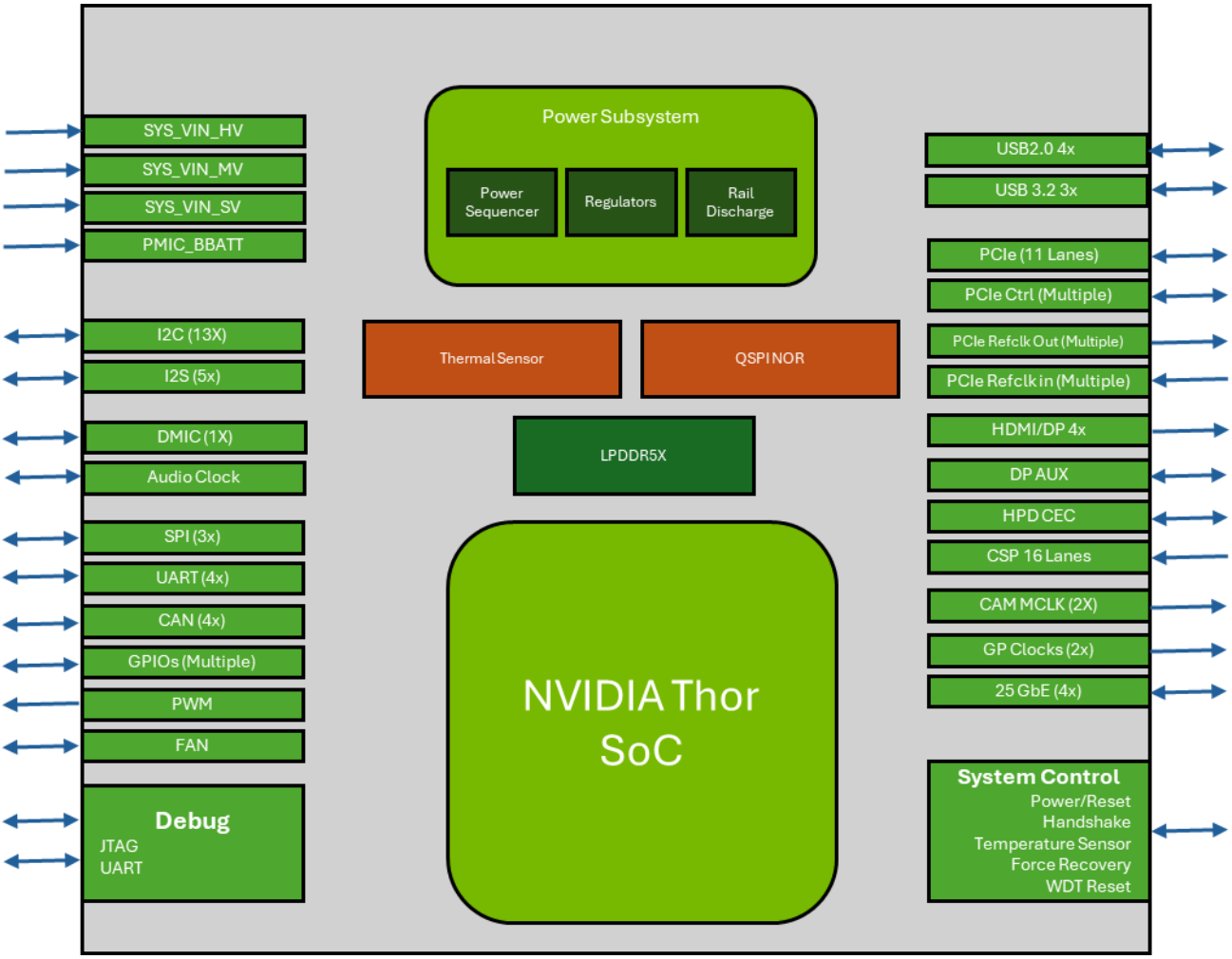



**Table 1** Jetson Thor Series Technical Specifications

| | **Jetson T5000** | **Jetson T4000** |
|---|---|---|
| AI Performance | 2070 TFLOPS (FP4) | 1200 TFLOPS (FP4) |
| GPU | 2560-core NVIDIA Blackwell architecture GPU with 96 5th GEN Tensor cores<br>Multi Instance GPU with 10 TPCs | 1536-core NVIDIA Blackwell architecture GPU with 64 5th GEN Tensor cores<br>Multi Instance GPU with 6 TPCs |
| GPU Max Freq. | 1.57 GHz | 1.57 GHz |

| | | |
|---|---|---|
| CPU | 14-core Arm® Neoverse®-V3AE 64-bit CPU<br>1 MB L2 Cache per core<br>16 MB Shared System L3 Cache | 12-core Arm® Neoverse®-V3AE 64-bit CPU<br>1 MB L2 Cache per core<br>16 MB Shared System L3 Cache |
| CPU Max Freq. | 2.6 GHz | 2.6 GHz |
| Vision Accelerator | 1x PVA v3 | 1x PVA v3 |
| Memory | 128 GB LPDDR5x<br>273 Gb/s | 64 GB LPDDR5x<br>273 Gb/s |
| Storage | Supports NVMe through PCIe<br>Supports SSD through USB3.2 | Supports NVMe through PCIe<br>Supports SSD through USB3.2 |
| CSI Camera | Up to 20 cameras via HSB<br>Up to 6 cameras through 16x lanes MIPI CSI-2<br>Up to 32 cameras using Virtual Channels<br>C-PHY 2.1 (10.25 Gbps), D-PHY 2.1 (40 Gbps) | Up to 20 cameras via HSB<br>Up to 6 cameras through 16x lanes MIPI CSI-2<br>Up to 32 cameras using Virtual Channels<br>C-PHY 2.1 (10.25 Gbps), D-PHY 2.1 (40 Gbps) |
| Video Encode | 2x NVENC (supports H.265 and H.264) | 1x NVENC (supports H.265 and H.264) |
| Video Decode | 2x NVDEC (supports H.265, H.264, VP9, VP8, AV1, MPEG-4, MPEG-2, VC-1) | 1x NVDEC (supports H.265, H.264, VP9, VP8, AV1, MPEG-4, MPEG-2, VC-1) |
| UPHY | Up to 1x8, 1 x4, 1 x2, and 2 x1 Gen 5 PCIe Lanes<br>Up to 4 XFI<br>3x USB3.1 | Up to 1x8, 1 x4, 1 x2, and 2 x1 Gen 5 PCIe Lanes<br>Up to 4 XFI<br>3x USB3.1 |
| Networking | 4x 25GbE | 3x 25GbE |
| Display | 4x shared HDMI2.1<br>VESA DisplayPort 1.4a - HBR2, MST | 4x shared HDMI2.1<br>VESA DisplayPort 1.4a - HBR2, MST |
| Other I/O | 5x I2S/2x Audio Hub (AHUB), 2X DMIS, 4x UART, 4x CAN, 3x SPI, 13x I2C, 6x PWM outputs | 5x I2S/2x Audio Hub (AHUB), 2X DMIS, 4x UART, 3x SPI, 13x I2C, 6x PWM outputs |
| Power | 40W-130W | 40W-90W |
| Mechanical | 100mm x 87mm<br>699 pin B2B Connector<br>Integrated Thermal Transfer Plate (TTP) with Heat-pipe | 100mm x 87mm<br>699 pin B2B Connector<br>Integrated Thermal Transfer Plate (TTP) with Heat-pipe |

# GPU

Jetson Thor modules contain an integrated Blackwell GPU composed of up to 3 Graphic Processing Clusters (GPCs), and 10 Texture Processing Clusters (TPCs) in T5000. To unlock superior AI performance, the SoC has up to 20 Streaming Multiprocessors (SMs). There are 128 CUDA cores per SM, compared to 64 CUDA cores per SM for Volta. Thor GPU contains four 5th Generation Tensor Cores per SM. In total, Jetson Thor modules have up to 2560 CUDA cores with 96 Tensor Cores, providing up to 2070 TFLOPS of FP4 compute. The new transformer engine in NVIDIA Thor allows a dynamic

switch between FP4 and FP8 operation based on the model requirements and thus enhancing the dynamic performance.

**Table 2** Jetson Thor Series Performance

| Precision | T5000 Performance |
|---|---|
| **FP4 Sparse** | 2070 TFLOPS |
| **FP4 Dense** | 1035 TFLOPS |
| **FP8 Sparse** | 1035 TFLOPS |
| **FP8 Dense** | 517 TFLOPS |
| **INT8 Dense** | 517 TOPs |
| **FP16 Dense** | 258 TFLOPS |

## Multi-Instance GPU (MIG)

Multi-Instance GPU has been a part of several NVIDIA products and now is being introduced in Jetson Thor. MIG partitions a GPU into multiple instances, and they each run simultaneously each with its own memory, cache and streaming multiprocessors, thus boosting the GPU utilization. Because MIG walls off GPU instances, it provides fault isolation – a key for mixed criticality workloads.

MIG allows the GPU to be split into two physically separated GPUs to provide isolation, for example allowing allocation of one MIG instance for critical workloads and another MIG instance for other applications. MIG separation is at the GPC level in the GPU. Graphics can only be supported on one MIG partition, or across the whole GPU, when MIG is not being used. Compute functionality is available on all MIG partitions.

As an example, MIG_0: can have 1 GPC (GPC_0 with 4 TPCs), and MIG_1 can have 2 GPCs (GPC_1 and GPC_2 with 6TPCs)

## NVIDIA 5th Generation Tensor Cores

NVIDIA 5th Generation Tensor Cores are programmable fused matrix-multiply-and-accumulate units that execute concurrently alongside the CUDA cores. Tensor cores implement floating point HMMA (Half-Precision Matrix Multiply and Accumulate) and IMMA (Integer Matrix Multiple and Accumulate) instructions for accelerating dense linear algebra computations, signal processing, and deep learning inference.

NVIDIA Blackwell brings fifth generation Tensor cores to provide the performance necessary to accelerate next generation AI applications. It significantly expands support for a range of numerical precisions: TF32, bfloat16, FP16, FP8, FP4, and INT8. This broad compatibility ensures optimal performance and adaptability for various AI and HPC workloads. Notably, TensorFloat-32 (TF32) leverages the same 10-bit mantissa as FP16 but incorporates the 8-bit exponent from FP32 Tensor cores, achieving high precision required for most AI applications.

In addition, Blackwell features native support for structured sparsity. By zeroing out select model parameters, structured sparsity reduces computation without impacting accuracy, enabling up to 2x performance improvement for sparse model inference on Tensor Cores.

Compute Data Compression is also supported in Blackwell, targeting efficient acceleration for unstructured sparsity and other forms of compressible data. Compression at the L2 cache level allows for up to 4x higher DRAM read/write bandwidth, up to 4x greater L2 read bandwidth, and up to 2x increased L2 cache capacity. These enhancements improve memory efficiency and overall accelerator throughput for demanding AI and scientific computing tasks.

## NVIDIA Transformer Engine

The second-generation Transformer Engine uses custom Blackwell Tensor Core technology combined with NVIDIA®'s CUDA-X libraries and AI serving frameworks such as vLLM, SGLang and EdgeLLM SDK, to accelerate inference and training for large language models (LLMs) and mixture-of-experts (MoE) models. The NVIDIA Blackwell Transformer Engine utilizes fine-grain scaling techniques called micro-tensor scaling, to optimize performance and accuracy enabling 4-bit floating point (FP4) AI. This

doubles the performance and size of next-generation models that memory can support while maintaining high accuracy for current and next-generation MoE models. For more details on the latest transformer engine refer the Transformer Engine Documentation.

## Memory

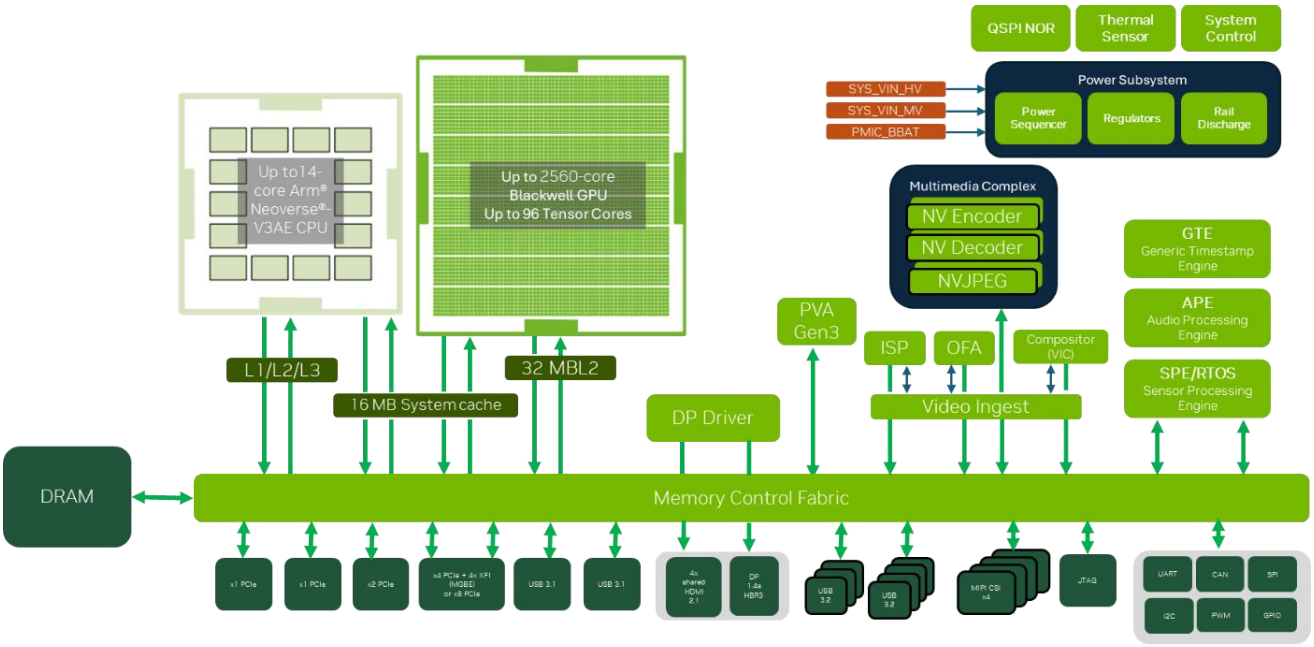
Jetson T5000 module brings support for 1.4X the memory bandwidth and 2X the storage of Jetson AGX Orin, enabling 128 of 256-bit LPDDR5. The DRAM supports a max clock speed of 4266 MHz, with 6400 Gbps per pin, enabling 273 GB/s of memory bandwidth. The superior 273GB/s memory bandwidth unlocks the power of large models that can now be run on edge devices.

This significant increase is crucial because state-of-the-art AI models, particularly transformer-based and generative models, demand massive and continuous data movement between memory and compute units. High bandwidth ensures these large models can be accessed and executed smoothly, minimizing latency and maximizing throughput—two key requirements for real-time edge deployments. With bandwidth on par with modern data center accelerators, Jetson Thor supports running complex neural networks for vision, speech, robotics, and multimodal applications directly at the edge, rather than relying on remote servers.

The high memory bandwidth is especially important for edge deployments where applications must process large, high-resolution sensor data streams or execute multiple advanced AI tasks in parallel. By providing 273GB/s memory bandwidth, Jetson Thor paves the way for deploying next-generation AI systems outside the data center, transforming what's possible for edge AI.

Figure 7 highlights how each of the various components interact with the Memory Controller Fabric and the DRAM.

**Figure 8** Jetson Thor functional block diagram

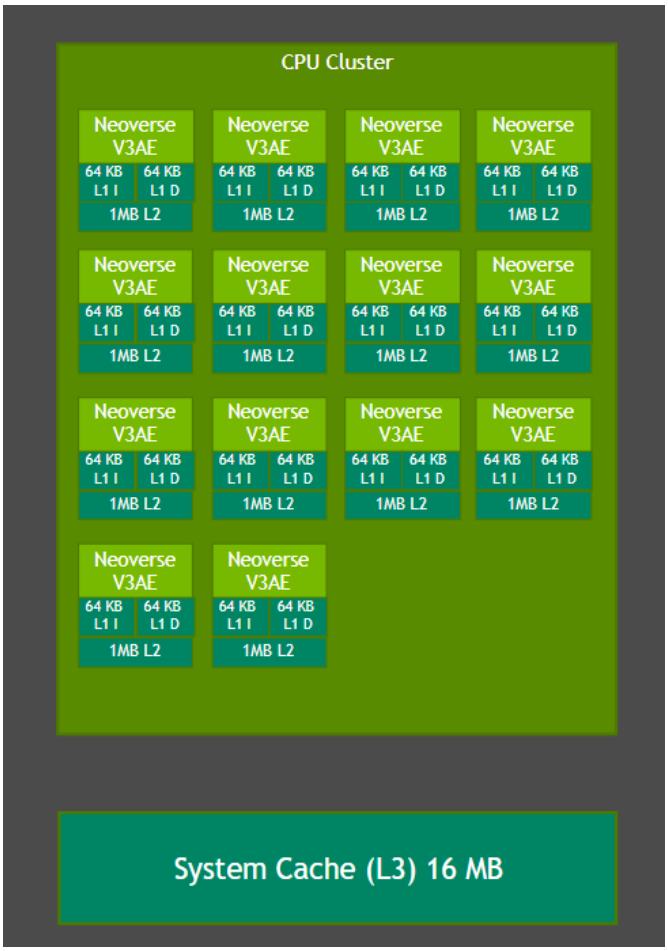



# CPU

Jetson Thor series modules feature Arm® Neoverse®-V3AE 64-bit CPU cores. The modules' CPU complex has up to 14 CPU cores. Each core includes 64KB Instruction L1 Cache and 64KB Data Cache, and 4x256 KB of L2 Cache. The CPU complex also features a 16MB Shares System LLC. The maximum supported CPU frequency is 2.6 GHz.

The 14-core CPU on Jetson T5000 enables almost 2.6 times the CPU performance compared to Jetson Orin. Users can leverage the enhanced capabilities of the Cortex-Neoverse V3AE including higher performance and enhanced cache to optimize their CPU implementations.

**Figure 9** CPU Complex




## Video Codecs

Jetson Thor modules contain up to two Multi-Standard Video Encoders (NVENC), a Multi-Standard Video Decoders (NVDEC), and a JPEG processing block (NVJPEG). NVENC enables full hardware acceleration for various encoding standards including H.265 and H.264. NVDEC enables full hardware acceleration for various decoding standards including H.265, H.264, AV1, VP9.

The dual NVENC/NVDEC video encoder and decoder engines in NVIDIA Jetson Thor enable simultaneous support for a high number of video streams—up to 6x 4Kp60 in H.265 to 48x 1080p30 in H.264. This capability is critical for NVIDIA Metropolis use cases, which often involve real-time analysis of multiple high-resolution video feeds from smart city cameras, retail stores, industrial sites, or traffic intersections.

The advanced encoding and decoding hardware ensure low-latency, high-throughput video processing, making it possible to efficiently ingest, analyze, and store video streams at the edge without overloading the network or requiring cloud roundtrips. This unlocks powerful video analytics—such as object detection, behavior analysis, and multi-camera tracking—directly within physical environments where timely insights are vital. The high parallelism and codec flexibility support scalable deployments and futureproof solutions, empowering the Metropolis platform to deliver robust AI vision agents for next-generation smart infrastructure, safety, and operational efficiency.

NVJPG is responsible for JPEG (de)compression calculations (based on the JPEG still image standard), image scaling, decoding (YUV420, YUV422H/V, YUV444, YUV400) and color space conversion (RGB to YUV). Please reference the Jetson Thor Series Data Sheet for a full list of standards. Customers can leverage NVIDIA Jetson's Multimedia API to power these engines. The Multimedia API is a collection of low-level APIs that supports flexible application development across these engines

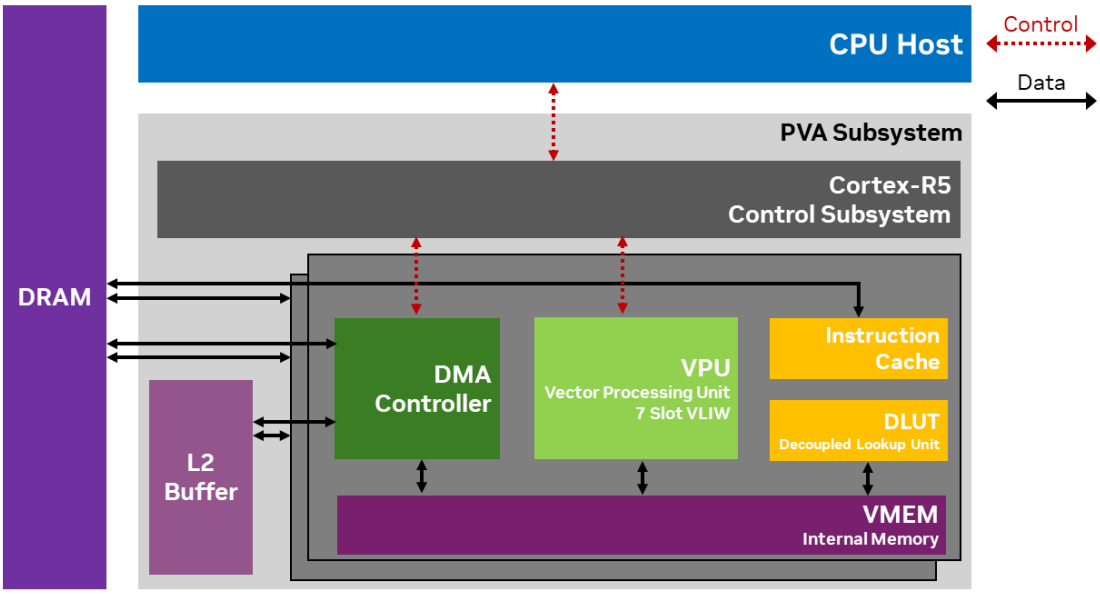
## Programable Vision Accelerator

Jetson Thor PVA engine is the third generation of NVIDIA's vision DSP architecture, which is an application-specific instruction vector processor that targets computer-vision along with virtual and mixed reality applications.

Jetson Thor PVA engine includes control processor, Cortex processor, 2x Vector Processing Subsystems (VPS) as data processing engines, 2x Pixel Processing engine subsystems (PPS), and 2x DMAs as data movement engines. The PVA enables support for various computer vision kernels such as filtering, warping, image pyramid, feature detection, and FFT. Some common computer vision applications using the PVA include feature detector, feature tracker, object tracker, stereo disparity, and visual perception.

It boosts system performance by offloading non-AI tasks—like feature extraction and object localization in multi-object tracking—to free up GPUs for AI-driven detection, tracking, and safety-critical functions.

**Figure 10** Jetson Thor PVA Block Diagram




The PVA SDK provides runtime APIs, tools, and tutorials for the development and deployment of CV and DL/ML algorithms. PVA Solutions provide operators, primitives, and samples powered by the PVA SDK. Source-code availability lets you use them as-is or customize them as needed. More details can be found in the PVA documentation.

## Video Imaging Compositor

The Thor SoC contains a Gen 4.2 Video Imaging Compositor (VIC) 2D Engine. The VIC enables support for various image processing features like lens distortion correction and enhanced temporal noise reductions, video features like sharpness enhancement, and general pixel processing features like color space conversion, scaling, blend, and composition.

## Optical Flow Accelerator

NVIDIA Optical Flow Accelerator (OFA) is hardware accelerator for computing optical flow and stereo disparity between the frames. Optical flow is useful in various use-case such as object detection and tracking,

while Stereo disparity is used in depth estimation. The hardware capabilities of OFA are exposed through NvMedia IOFA APIs.

OFA can operate in two modes:

- Stereo Disparity Mode: In this mode, OFA generates disparity between rectified left and right images of stereo capture. The output stereo disparity format is fixed signed 10.5 (2 bytes per disparity output). We need to divide the output values by 32 to get a disparity value in terms of pixel units.

- Optical Flow Mode: In this mode, OFA generates optical flow between two given frames, returning X and Y component of flow vectors.

The input to OFA in this mode is image pyramid of input and reference frames with fixed scale factor of 2. As search range of single layer is small, each pyramid level will search around output of previous pyramid level. • OFA generates a flow vector has X and Y component that represent motion in X and Y direction. The output flow format is fixed signed 10.5 (4 bytes per flow vector). We need to divide the output values by 32 to get a disparity value in terms of pixel units.

OFA generates disparity and flow vector block-wise, one output for each input block of 4x4, 2x2, and 1x1 pixels (referred as output grid size). The generated output can be further post-processed to improve accuracy, up sampled to produce dense map.

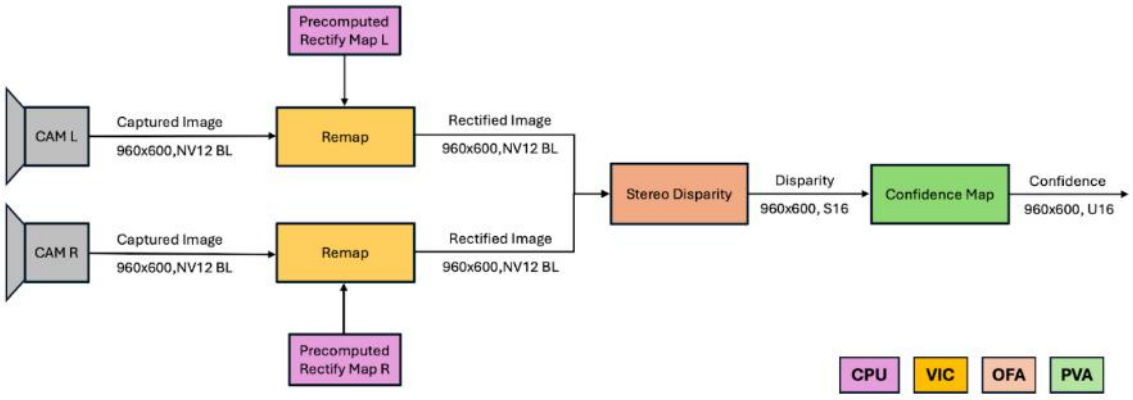
## Vision Programming Interface

Vision Programming Interface (VPI) is a software library which implements computer vision and image processing algorithms on several NVIDIA Jetson hardware components including PVA, VIC, CPU, and GPU. With VPI algorithm support on the PVA and the VIC, one can offload computer vision and image processing tasks to them and prioritize the CPU and the GPU for other tasks.

As an example, a complete Stereo Disparity Estimation pipeline using VPI can efficiently use several backends including the VIC, PVA, and OFA. The pipeline receives the input from a stereo camera, which are left and right images of a stereo pair. The VIC works on this input to correct lens distortion and optionally scales the images, resulting in a rectified stereo pair. This pair

is fed into the OFA backend, and PVA is used to compute the confidence map. The output is an estimate of the disparity between the input images, which is related to the scene depth.

**Figure 11** Stereo Disparity Estimation Pipeline




VPI comes with several algorithms ranging from image processing building blocks like box Filtering, Convolution, Image Rescaling and Remap, to more complex computer vision algorithms like Harris Corners Detection, KLT Feature Tracker, Optical Flow, Background Subtraction, and more. Please check out the VPI Webinars here to learn more about using VPI to accelerate computer vision applications.

VPI is a fixed functionality library, and for users who want to implement their own algorithms, we enable PVA SDK to program the PVA. Users can refer the SDK documentation here.

# I/O

Jetson is used across a variety of applications with various I/O requirements. For example, Autonomous Ground Vehicles could leverage the CSI cameras for a surround view around the robot, I2S for voice commands, HDMI for display, PCIe for Wi-Fi, GPIO & I2C and more. A video analytics application like traffic management at an intersection might require many GigE Cameras and Ethernet for networking purposes. As autonomous machines continue to perform more advanced tasks, more I/O is needed to interface more sensors.

Jetson Thor series modules contain plenty of high speed I/O including 12 lanes of PCIe Gen5, 4x 25Gigabit Ethernet, 1.4a Display Port, HDMI 2.2, 16 lanes of MIPI CSI-2, USB3.2 interfaces as well as various other I/O like I2C, I2S, SPI, CAN, GPIO, USB 2.0. Users can leverage the UPHY lanes for USB 3.2, PCIe, and MGBE, and some of the UPHY lanes are shared between these interfaces. The Display Port can support 2 displays using the Multi-stream mode on DP1.4. The HDMI can support up to 4 displays.

**Table 3** Jetson T5000 I/O

| Interface | |
|---|---|
| **Camera** | Up to 20 cameras via HSB<br>Up to 6 cameras through 16x lanes MIPI CSI-2<br>Up to 32 cameras using Virtual Channels<br>C-PHY 2.1 (10.25 Gbps)<br>D-PHY 2.1 (40 Gbps) |
| **UPHY** | Up to 1 x8, 1 x4, 1 x2, 2 x1 Gen 5 PCIe lanes,<br>3x USB3.1<br>Up to 4 XFI |
| **Networking** | 4x 25GbE<br>1x 5GbE |
| **Display** | 4x shared HDMI2.1<br>VESA DisplayPort 1.4a - HBR2, MST |
| **Other I/O** | 5x I2S/2x Audio Hub (AHUB), 1x DMIC, 4x UART, 4x CAN, 3x SPI, 13x I2C, 6x PWM outputs |

# Power Profiles

Jetson Thor series modules are designed with a high efficiency Power Management Integrated Circuit (PMIC), voltage regulators, and power tree to optimize power efficiency. Jetson T5000 supports 75W, 90W, 120W, and MAXN power modes. Each power mode caps various component frequencies, and the number of online CPU, GPU TPC, and PVA cores, and custom power modes can be configured down to 40W. Customers can leverage the *nvpmodel* tool in Jetson Linux to use one of these pre-optimized power modes or to customize their own power mode within the design constraints provided in our documentation.

# Safety

The NVIDIA IGX Thor platform is architected for deterministic, safety-critical AI workloads in industrial automation, robotics, and medical technology. It embraces NVIDIA Halos principles and leverages the NVIDIA research in architecting safety for autonomous driving.

The IGX Thor hardware has been developed to meet ISO 26262 and IEC 61508 systematic failures requirements. The IGX Thor safety application note provides guidance on other functional safety standards such as ISO 13849, allowing developers to reuse safety-certified building blocks and shorten the path to certification. It also provides guidance on how AI can be safely developed based on ISO/IEC TS 22440 requirements.

Thor's design integrates a multi-core lockstep-based Functional Safety MCU into a dedicated Functional Safety Island (FSI) with its own clock, power, and memory domains. The safety island includes watchdog timers, ECC-protected memory, CRC checks, and direct actuator control interfaces — critical for executing hard real-time fail-safe or fail-operational routines.

Thor uses hardware fault containment regions (FCRs) to keep safety-critical tasks isolated from general-purpose AI workloads, preventing fault propagation. This enables mixed-criticality execution, where certified motion control logic can run alongside high-throughput perception pipelines on the GPU/DLA accelerators without violating safety constraints.

IGX Thor can be integrated directly within physical robots, utilizing onboard sensors to implement safety functions in an inside-out manner—that is, perceiving the world from the robot's perspective. Additionally, IGX Thor can be deployed throughout facility infrastructure where robots operate, leveraging external sensors for an outside-in viewpoint to provide an added layer of safety through AI. AI models developed according to ISO/IEC TS 22440 run on deterministic execution tracks, while redundant secondary models cross-check primary outputs. Any discrepancy triggers automated mitigations such as speed reduction, controlled stops, or a switch to

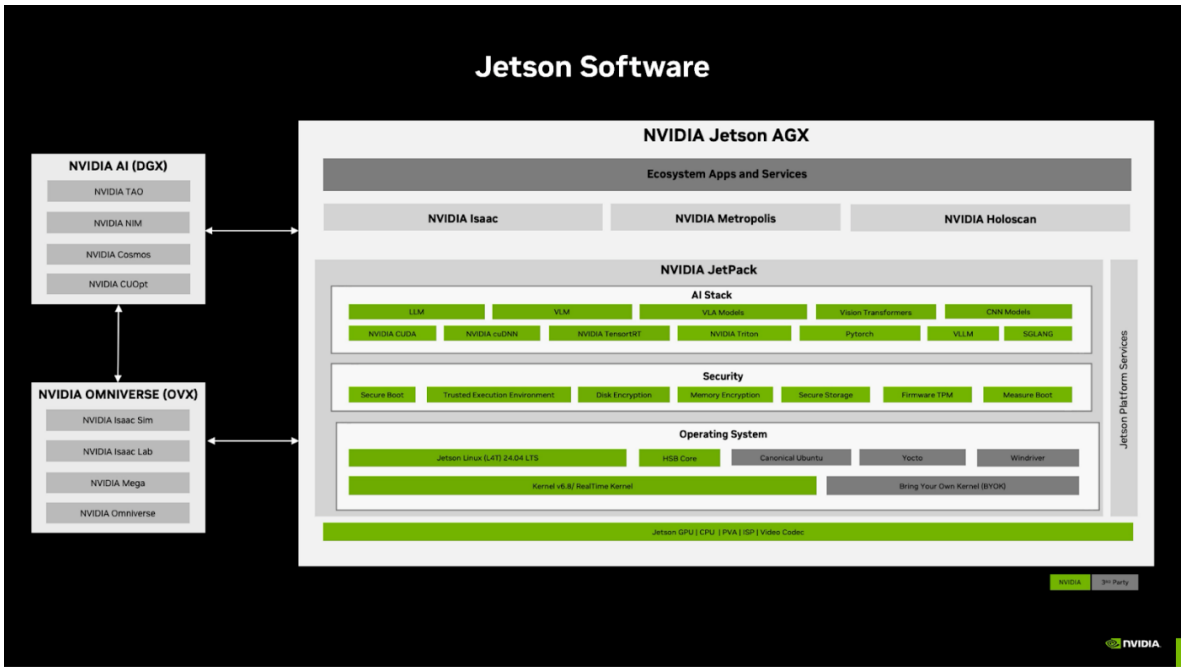
teleoperation—coordinated by the FSI, the external safety MCU or external PLC.

Security underpins the safety model, starting with a hardware root of trust in Thor's security island. The system enforces secure boot, cryptographic firmware signing, and runtime integrity verification, ensuring only authenticated safety logic and AI models are in operation. Partitioned OS environments keep safety-critical and non-critical software separated, while the Baseboard Management Controller (BMC) enables secure over-the-air updates and remote diagnostics. Together, these capabilities make IGX Thor a certifiable, millisecond-response platform designed for mission-critical AI at the edge.

# Jetson Software

Jetson Software is NVIDIA's full-stack edge AI and robotics platform, built for high performance across industries like robotics, healthcare, logistics, and autonomous systems. Designed for real-time, runtime generative AI, it enables fast development and deployment of advanced AI applications. With optimized system software, tight hardware integration, and cloud-native support, Jetson software accelerates innovation at the edge, where latency and efficiency matter most. It powers complex use cases such as sensor processing, multicamera tracking, and robotic perception and control, enabling intelligent systems like humanoid robots, autonomous platforms, and precision manipulators.

**Figure 12** Jetson Software Stack

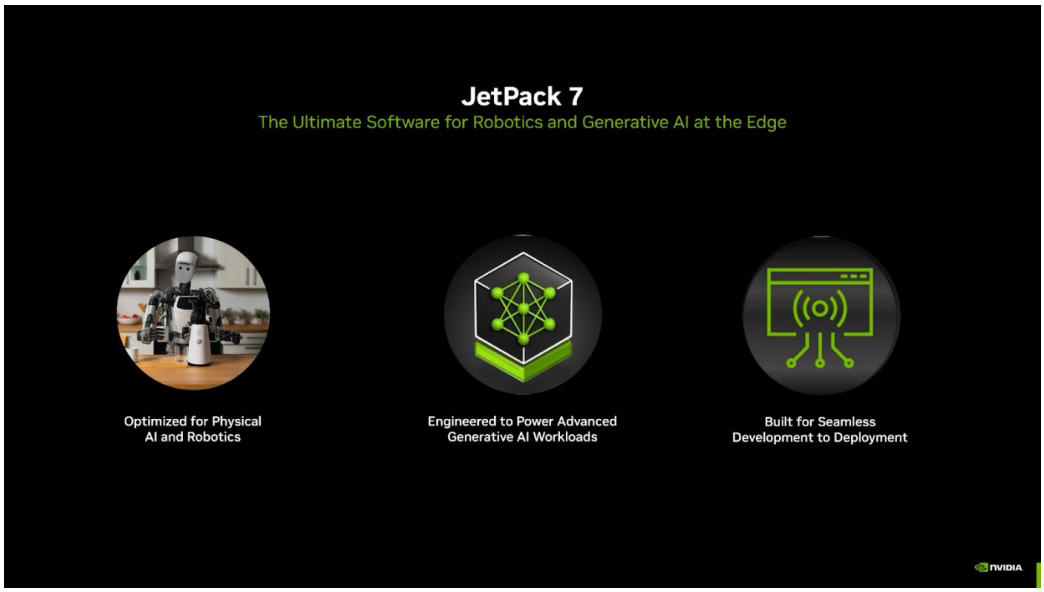


Jetson software is powered by JetPack7, a modern software stack built on Linux kernel 6.8 and Ubuntu 24.04 LTS. JetPack7 provides full support for generative AI models and is optimized for the latency and determinism required in real-time applications such as humanoid robotics, medical imaging, and industrial automation. It leverages the latest NVIDIA AI stack, from cloud to edge, with microservices and agentic workflow support that simplify the integration of perception, planning, and control in intelligent systems.

## JetPack 7

JetPack 7 introduces key advancements for physical AI, including Holoscan Sensor Bridge, Multi-Instance GPU (MIG) support, a Preemptable Realtime Kernel, and NVIDIA's latest AI compute libraries. It supports leading generative AI serving frameworks like vLLM, SGLang, TensorRT, TensorRT-LLM, PyTorch, Llama.cpp, and MLC for efficient real-time model deployment.

**Figure 13** JetPack7 Pillars




Security is critical for physical AI applications, where intelligent machines interact with the real world, often in safety-critical or sensitive environments. Protecting data integrity, safeguarding AI models, and ensuring trusted execution are essential to prevent tampering,

unauthorized access, or malicious behavior. JetPack 7 delivers a comprehensive suite of security features spanning edge to cloud, including secure boot, disk encryption, runtime integrity, fTPM, and secure OTA updates, enabling developers to build and deploy trusted AI systems at scale.

Jetson software is backed by a rich collection of assets and ecosystem. NVIDIA Blueprints such as the Video Search and Summarization (VSS) makes it easy to build and customize visually perceptive and interactive AI Agents, integrating vision language models (VLM), vision foundation models, and LLMs, and. The Jetson AI Lab offers ready-to-deploy generative AI models and tutorials for Jetson edge devices. NVIDIA also maintains an extensive ecosystem of partners and ISVs that provide domain expertise to accelerate development across robotics and other edge AI applications.

## Get the most out of Blackwell GPU using NVIDIA Software Libraries

Users can accelerate their inferencing on the GPU using NVIDIA TensorRT and cuDNN. NVIDIA TensorRT is a runtime library and optimizer for deep learning inference that delivers lower latency and higher throughput across NVIDIA GPU products. TensorRT enables users to parse a trained model and maximize the throughput by quantizing models to INT8, optimizing use of the GPU memory and bandwidth by fusing nodes in a kernel, and selecting the best data layers and algorithms based on the target GPU.

cuDNN (CUDA Deep Neural Network Library) is a GPU-accelerated library of primitives for deep neural networks. It provides highly tuned implementations of routines commonly found in DNN applications like convolution forward and backward, cross-correlation, pooling forward and backward, softmax forward and backward, tensor transformation functions, and more. With the Ampere GPU and NVIDIA software stack, users can handle new, complex neural networks that are being invented every day.

NVIDIA also provides CUDA libraries tailored for robotics application development. These include cuVSLAM for visual-inertial simultaneous localization and mapping, nvblox for real-time 3D scene reconstruction,

cuMotion for GPU-accelerated optimal-time, minimal-jerk trajectories for serial robot arms, and cuVGL for high-performance visualization and pose estimation in robotics applications. Together, these libraries accelerate perception, planning, and control, enabling more advanced and efficient robotics systems.

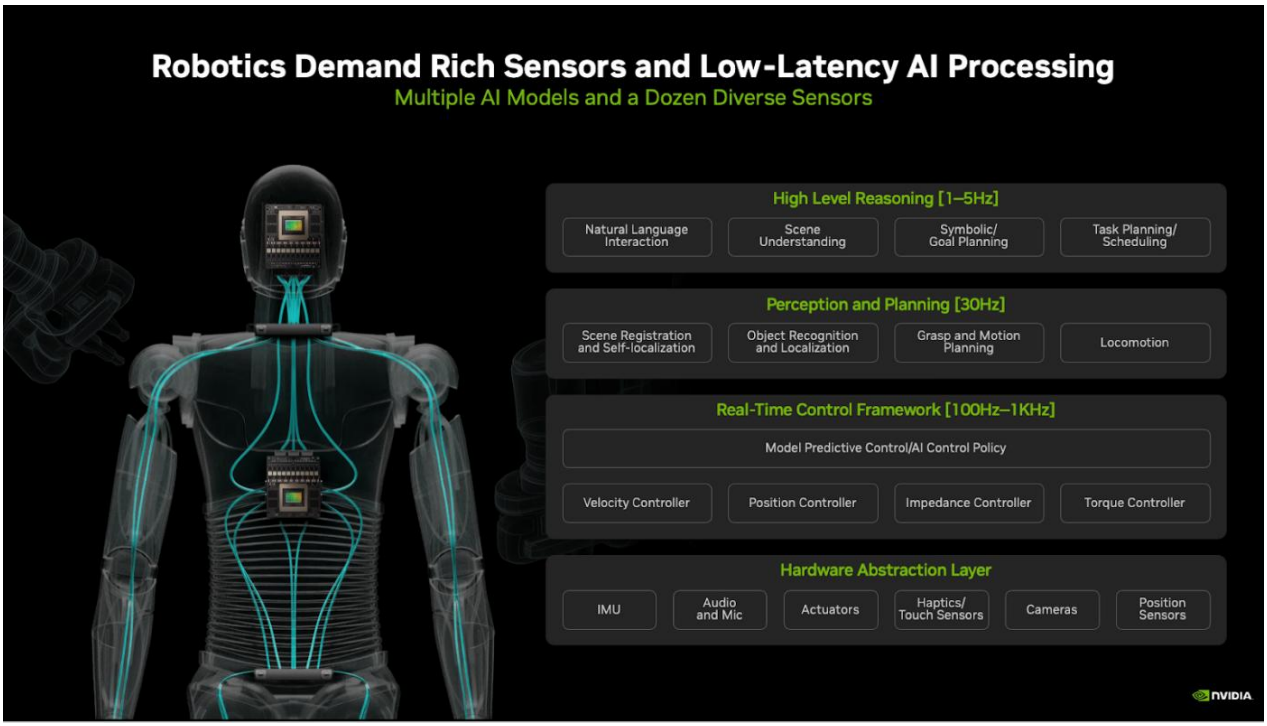
## Supercharge humanoid robot development with NVIDIA GR00T workflows.

**What does it take to build a Generalist Robot like a Humanoid?**

A typical humanoid robot is organized into four essential layers:

- Hardware Abstraction Layer: Integrates all key sensing and actuation modalities, enabling the robot to perceive its environment and interact physically with the world.
- Real-Time Control Framework: Manages precise, low-latency control of the robot's movement, where minimizing latency is critical for safe and responsive operation.
- Perception and Planning: Equips the robot with environmental understanding, grasp and motion planning, locomotion, object recognition, and localization—allowing effective interaction with the surrounding world. Unlike real-time control, this layer allows for slightly more processing time to ensure thoughtful, accurate decisions.
- High-Level Reasoning: Powers advanced functions such as scene understanding, complex task planning, and natural language interaction increasingly enabled by reasoning Vision-Language-Action (VLA) models where longer processing times are acceptable to support deeper reasoning and adaptability.

**Figure 14** Humanoid Sensors and AI Processing




Robotic systems require software that can isolate critical workloads and deliver real-time, deterministic responses to ensure efficient operation. They must also support diverse generative AI models alongside fast-serving frameworks. Jetson software addresses these needs with a preemptable real-time kernel for low-latency, deterministic performance, essential for machines that interact with the physical world. Support for Multi-Instance GPU (MIG) further enables mixed-criticality workloads by isolating safety-critical tasks from less sensitive processes. Jetson software enables real-time sensor processing, multi-camera tracking, and advanced robotics capabilities such as manipulation and navigation, all supported within a robust AI ecosystem. Developers also benefit from integrated software tools such as integrated Holoscan Sensor Bridge for high-bandwidth, low-latency sensor streaming and Video Search and Summarization blueprint from NVIDIA Metropolis for building video analytics AI agents. Jetson software meets these demands with a unified, scalable foundation that spans from prototyping to production across the full Jetson platform.

Unlike generative AI domains such as language and vision that rely on abundant public datasets, humanoid robotics struggles with limited access to rich, task-specific, multi-modal sensor data.

Jetson Thor is purpose-built as the high-performance runtime platform for Isaac GR00T workflows - NVIDIA's canonical path for sim-to-real robotics. GR00T combines foundation models like GR00T N1.5, a reasoning vision-language-action model (VLA), with simulation tools built on NVIDIA Omniverse and Cosmos, and pipelines that fuse synthetic and real-world sensor data through Holoscan. While GR00T offers the most integrated experience, Thor also supports other reasoning Vision-Language-Action (VLA) foundation models and specialist end-to-end policies, giving developers flexibility to deploy their own innovations on the same deterministic runtime.

With Jetson Thor's powerful compute, low-latency architecture, and support for synchronized multi-sensor input, developers can efficiently train and deploy large robotics models for real-time inference and continuous learning, accelerating the entire development lifecycle from simulation to real-world deployment.

Jetson Thor gives you the performance you need for real-time control and multi-sensor processing, while GR00T's comprehensive AI software stack supports a wide range of generative AI models and seamless cloud-to-edge integration. Together, they offer a powerful, integrated solution that accelerates the development and deployment of sophisticated humanoid robots, making them more adaptable, responsive, and capable.

## Visual Search and Summarization (VSS)

NVIDIA AI Blueprint for Video Search and Summarization (VSS) is an advanced system designed to efficiently analyze, search, and summarize large-scale video and image data. Its architecture brings together vision-language models (VLMs), large language models (LLMs), and retrieval-augmented generation (RAG) to make video analytics interactive and scalable—all while leveraging the acceleration power of NVIDIA GPUs.

At a high level, the VSS pipeline works as follows:

- Ingestion and Encoding:

    - Video streams or files are divided into chunks and sampled for key frames.
    - Each frame is processed using visual encoders to generate rich feature embeddings.
    - Multimodal enrichment includes computer vision metadata (object detection, tracking, segmentation) and audio transcription.
- Semantic Understanding & Storage:
    - Encoded frames are passed through VLMs to produce dense captions, object contexts, and scene summaries.
    - Embeddings and structured metadata are stored in a vector database and knowledge graph for quick retrieval.
    - The RAG module links video segments, building a graph of objects, actions, and relationships for deeper contextual understanding.
- Search and Summarization:
    - Users can query for specific events, objects, or activities.
    - The system supports natural language search across time and across sources.
    - Summaries are dynamically generated, highlighting relevant scenes, tracking objects/people, or aggregating key events.
- Performance & Flexibility:
    - The architecture supports burst workloads and real-time summarization with GPU acceleration.
    - Chunk and batch size can be tuned to balance accuracy and throughput.
    - The modular pipeline allows plug-in of custom vision or language models and adapts to a range of industries—from surveillance to compliance auditing.

Security and compliance are maintained by supporting access controls, policy-based data handling, and scalable event-driven notifications for sensitive environments.

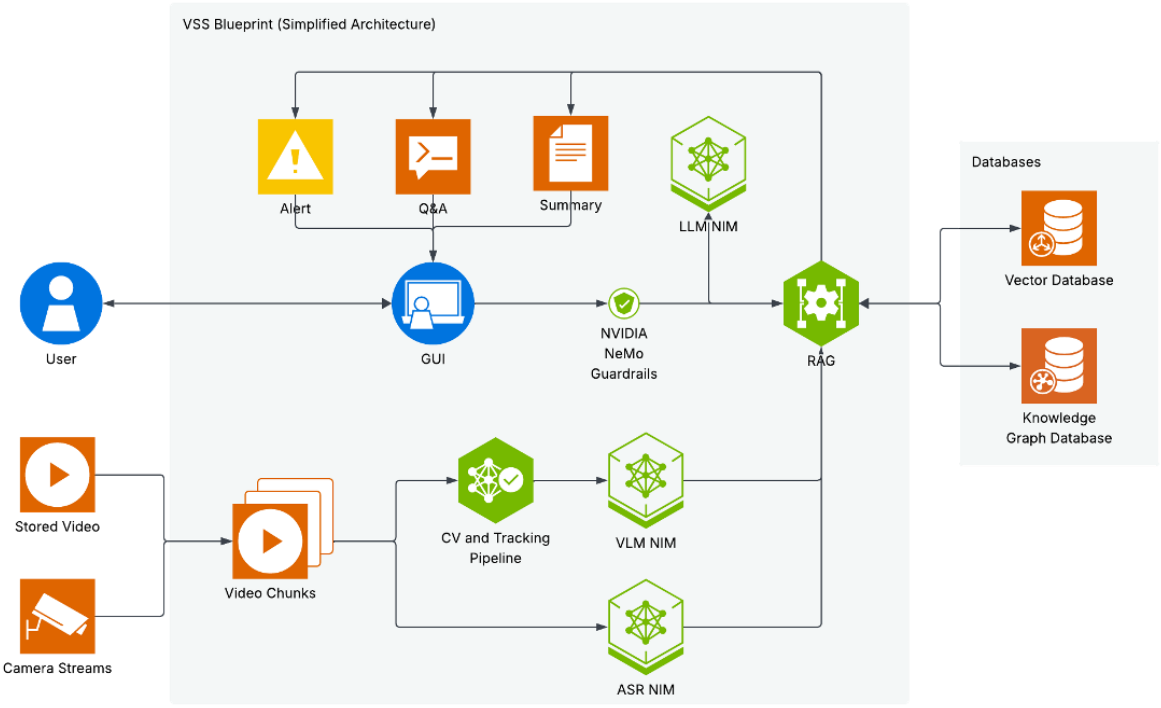
## Video Search and Summarization Agent

Building a video analytics AI agent capable of understanding long-form videos requires a combination of VLMs and LLMs ensembled together with datastores. The blueprint provides a recipe for combining all of these components to enable scalable and GPU-accelerated video understanding agents that can perform several tasks such as summarization, Q&A, and detecting events on live streaming video.

The blueprint consists of the following components:

- Stream handler: Manages the interaction and synchronization with the other components such as NeMo Guardrails, CA-RAG, the VLM pipeline, chunking, and the Milvus Vector DB.
- NeMo Guardrails: Filters out invalid user prompts. It makes use of the REST API of an LLM NIM microservice.
- VLM pipeline – Decodes video chunks generated by the stream handler, generates the embeddings for the video chunks using an NVIDIA Tensor RT-based visual encoder model, and then makes use of a VLM to generate per-chunk response for the user query. It is based on the NVIDIA DeepStream SDK.
- VectorDB: Stores the intermediate per-chunk VLM response.
- CA-RAG module: Extracts useful information from the per-chunk VLM response and aggregates it to generate a single unified summary. CA-RAG (Context Aware-Retrieval-Augmented Generation) uses the REST API of an LLM NIM microservice.
- Graph-RAG module: Captures the complex relationships present in the video and stores important information in a graph database as sets of nodes and edges. This is then queried by an LLM for interactive Q&A.

**Figure 15** VSS Blueprint Architecture




VSS on Jetson Thor platform provides a new *event verification* feature, enabling VSS to act as an intelligent add-on to any existing computer vision (CV) pipeline.

While standard CV pipelines can detect objects like people appearing in view or apply analytics to identify events such as possible vehicle collisions, they often generate false positives and lack deeper scene understanding.

VSS enhances these pipelines by analyzing short video clips flagged by the CV system, verifying the detected events, and uncovering additional insights using that traditional methods may miss. VSS accomplishes this by using a Vision Language Model (VLM) - Cosmos Reason.

# Holoscan Sensor Bridge

Connecting different cameras, radars, lidars, and RF sensors often require unique, proprietary data protocols that are costly and complex to develop, maintain, and scale. The NVIDIA Holoscan Sensor Bridge is a sensor-over-Ethernet technology designed to enable real-time data streaming and simplify high-speed sensor fusion and actuator integration on NVIDIA edge-AI platforms. It provides a standard API and open-source software that streams high-speed sensor data directly to GPU memory through FPGA interfaces. For mission-critical applications like healthcare, instrumentation, robotics, signal processing, and beyond, you can achieve ultimate low latency from sensor data to compute with more flexibility and scalability.

Holoscan Sensor Bridge provides an IP that can run on an FPGA for low-latency sensor data processing using GPUs. Peripheral device data is acquired by the Holoscan Sensor Bridge device FPGA and sent via UDP over Ethernet to the host system. Interfaces can greatly lower latency by writing that UDP data directly into GPU memory.

## Holoscan Sensor Bridge FPGA IP Architecture

The NVIDIA Holoscan Sensor Bridge IP, designed to work seamlessly with Holoscan software, transforms your FPGA into a sensor-agnostic data gateway for Ethernet-connected hosts. Whether you're dealing with new sensor types or scaling up to more complex applications, this IP core streamlines and accelerates your FPGA design work, making it easy to adapt for a wide range of sensor-to-host use cases.

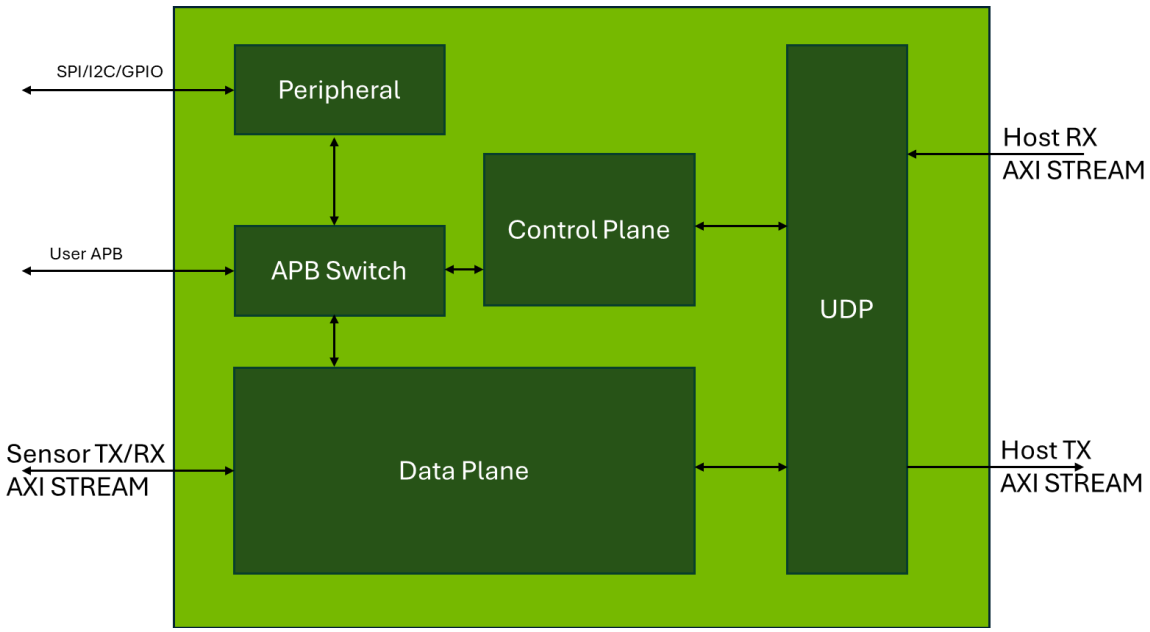
What does this mean in practice? The Sensor Bridge IP handles everything from encapsulating raw sensor data (over AXI-Stream) into Ethernet-bound UDP streams, ensuring your data is network-ready before it even reaches the host. It also efficiently manages key networking protocols—including BOOTP, ICMP, and NVIDIA's own Ethernet Control Bus (ECB)—right on the FPGA, freeing developers from complex low-level integration.

Beyond just data movement, the Holoscan Sensor Bridge can transmit automatic enumeration and event packets based on predefined triggers,

ensuring your system is always aware of device states and critical events. Plus, it offers built-in control over SPI, I2C, and GPIO peripheral links, making sensor configuration and management straightforward—all from a single, unified IP block. This combination of flexibility and turnkey functionality makes it an ideal foundation for both rapid prototyping and robust, production-ready sensor integration.

There are 2 main planes of data bridging in the Holoscan Sensor Bridge IP, the Data-plane and the Control-plane. The Data-plane is data transfer relevant to sensor data as it gets packetized to UDP packets. The Control-plane is data transfer relevant to host control. The Control-plane could be receiving and decoding UDP packets from the host to access a register or transmitting specific network protocols the Holoscan Sensor Bridge IP supports.

**Figure 16** Holoscan Sensor Bridge IP Architecture



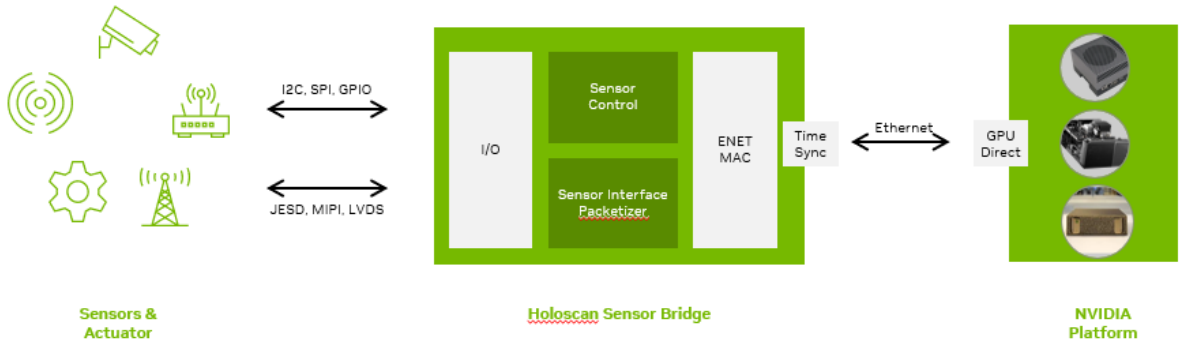

Holoscan Sensor Bridge offers multiple advantages over traditional architectures:

- Ultra-low Latency - The Holoscan Sensor Bridge ensures real-time data processing, reducing latency by up to 10x compared to

traditional systems. This means faster decision-making and more responsive applications, which are crucial in dynamic environments like autonomous vehicles and industrial automation.

- Scalability - The modular design of the Holoscan Sensor Bridge makes it easy to scale and adapt to different needs. Software-defined sensors allow for flexible and customizable configurations, ensuring that the system can grow and evolve alongside your project, whether it's a small-scale deployment or a large, complex setup.
- Safety and Security - The Holoscan Sensor Bridge is built with end-to-end safety protocols, meeting up to SIL 2 standards. This ensures that data streaming is not only processed efficiently, but also securely, providing a robust and reliable solution for critical applications where safety is paramount.
- Ease of use - The Holoscan Sensor Bridge not only lets you significantly cut down the time required to create and integrate sensor drivers but also facilitates efficient sensor fusion. The streamlined process reduces development time by up to 100x, allowing teams to focus more on innovation and less on tedious setup and configuration.

**Figure 17** Holoscan Sensor Bridge Reference Design Architecture



# Jetson AGX Thor Developer Kit

The Jetson AGX Thor Developer Kit will contain everything needed for developers to get up and running quickly. The Jetson AGX Thor Developer Kit includes the Jetson T5000 module with a reference carrier board, thermal solution, and a power supply. The Jetson AGX Thor Developer kit can be used to develop for all the Jetson Thor modules. With up to 2070 TFLOPS of AI performance and power configurable between 40W and 130 W, customers now have more than 7.5X the performance of Jetson AGX Orin for developing advanced humanoids and other autonomous machine products. The Jetson AGX Thor Developer Kit is available today.

**Figure 18** NVIDIA Jetson AGX Thor Developer Kit

### Notice

This document is provided for information purposes only and shall not be regarded as a warranty of a certain functionality, condition, or quality of a product. NVIDIA Corporation (“NVIDIA”) makes no representations or warranties, expressed or implied, as to the accuracy or completeness of the information contained in this document and assumes no responsibility for any errors contained herein. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This document is not a commitment to develop, release, or deliver any Material (defined below), code, or functionality.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice.

Customer should obtain the latest relevant information before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer (“Terms of Sale”). NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this document. No contractual obligations are formed either directly or indirectly by this document.

NVIDIA products are not designed, authorized, or warranted to be suitable for use in medical, military, aircraft, space, or life support equipment, nor in applications where failure or malfunction of the NVIDIA product can reasonably be expected to result in personal injury, death, or property or environmental damage. NVIDIA accepts no liability for inclusion and/or use of NVIDIA products in such equipment or applications and therefore such inclusion and/or use is at customer’s own risk.

NVIDIA makes no representation or warranty that products based on this document will be suitable for any specified use. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer’s sole responsibility to evaluate and determine the applicability of any information contained in this document, ensure the product is suitable and fit for the application planned by customer, and perform the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer’s product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this document. NVIDIA accepts no liability related to any default, damage, costs, or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this document or (ii) customer product designs.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this document. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA.

Reproduction of information in this document is permissible only if approved in advance by NVIDIA in writing, reproduced without alteration and in full compliance with all applicable export laws and regulations, and accompanied by all associated conditions, limitations, and notices.

THIS DOCUMENT AND ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, “MATERIALS”) ARE BEING PROVIDED “AS IS.” NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. TO THE EXTENT NOT PROHIBITED BY LAW, IN NO EVENT WILL NVIDIA BE LIABLE FOR ANY DAMAGES, INCLUDING WITHOUT LIMITATION ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, PUNITIVE, OR CONSEQUENTIAL DAMAGES, HOWEVER CAUSED AND REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF ANY USE OF THIS DOCUMENT, EVEN IF NVIDIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA’s aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the Terms of Sale for the product.

### Trademarks

NVIDIA, the NVIDIA logo, NVIDIA Jetson are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

### VESA DisplayPort

DisplayPort and DisplayPort Compliance Logo, DisplayPort Compliance Logo for Dual-mode Sources, and DisplayPort Compliance Logo for Active Cables are trademarks owned by the Video Electronics Standards Association in the United States and other countries.

### HDMI

HDMI, the HDMI logo, and High-Definition Multimedia Interface are trademarks or registered trademarks of HDMI Licensing LLC.

### Arm

Arm, AMBA, and Arm Powered are registered trademarks of Arm Limited. Cortex, MPCore, and Mali are trademarks of Arm Limited. All other brands or product names are the property of their respective holders. "Arm" is used to represent Arm Holdings plc; its operating company Arm Limited; and the regional subsidiaries Arm Inc.; Arm KK; Arm Korea Limited.; Arm Taiwan Limited; Arm France SAS; Arm Consulting (Shanghai) Co. Ltd.; Arm Germany GmbH; Arm Embedded Technologies Pvt. Ltd.; Arm Norway, AS, and Arm Sweden AB.

### OpenCL

OpenCL is a trademark of Apple Inc. used under license to the Khronos Group Inc.

### Copyright



NVIDIA Corporation | 2788 San Tomas Expressway, Santa Clara, CA 95051
http://www.nvidia.com