NETINT

Quadra Video Server

Video Processing Unit | Quadra ASIC G5

Executive Summary

Early adoption of ASICs slingshots streaming platforms ahead of their competition due to their low operational cost basis.

Designed as a quickstart solution for high density live video encoding and transcoding, the Quadra 100 Video Server comprises ten Quadra video processing units (VPU) in a 1RU chassis that performs the equivalent work of up to 25 dedicated servers running a typical open-source FFmpeg and x264, x265, or SVT-AV1 configuration. The server delivers the lowest TCO of any solution in the market, and is a drop-in replacement for existing CPU and GPU encoding stacks.

NETINT application specific integrated circuits (ASICs) are the secret to replacing software-driven video encoding for video platforms and delivery services wanting to decrease carbon emissions, operating costs, and the number of servers needed by 90 - 95% for H.264 output, and as much as 99% for HEVC.

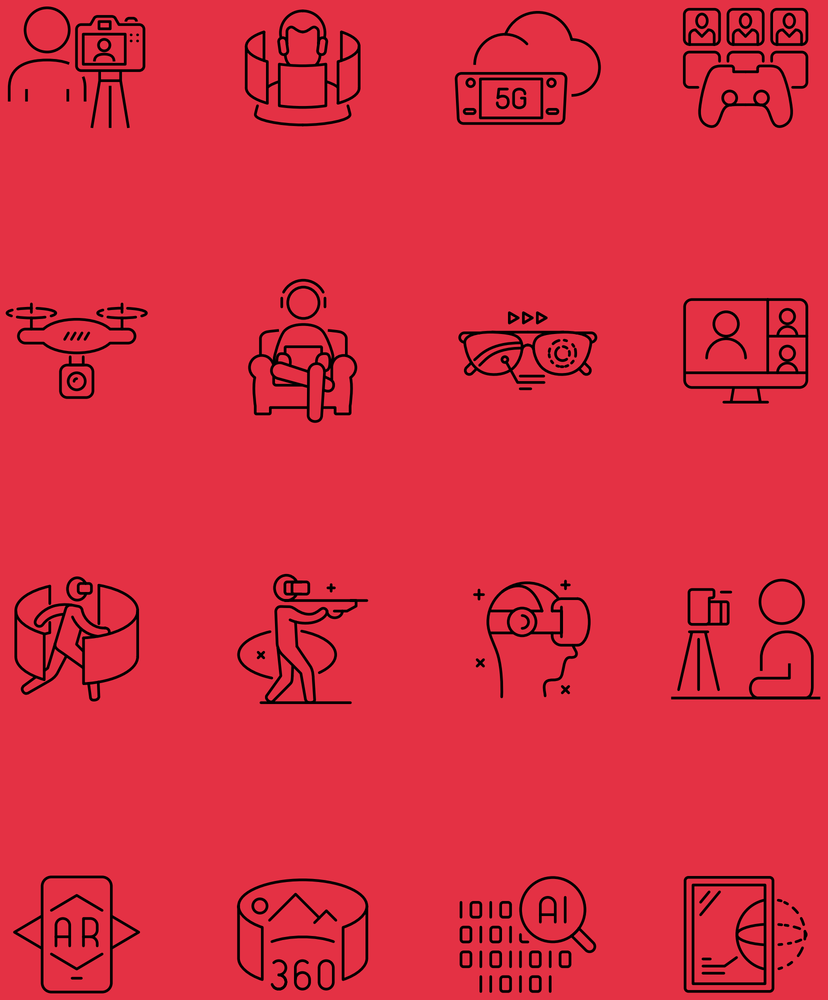
The results are profound and transformational.

ASICs will be the engine powering all future video streaming experiences

Live streaming experiences are seeing rapid adoption

Applications:

- Live events
- Interactive video
- Cloud gaming
- Real-Time video
- Virtual worlds
- 360 / VR / AR



The insatiable appetite of video consumers

They want nonstop, never-ending, high-resolution, non-buffering content accessible on any device. Now.

Viewers have developed an addiction to continuous content streaming. Video delivery and entertainment experiences are shifting from file-based to live where low latency and controlling operational costs are paramount.

- Increased public cloud provider costs are stressing businesses
- Live experiences are growing in resolution, color depth, and quality expectations
- Playback is expected on every device using its full capability
- More data centers are needed to handle capacity increases

2021

Social video viewing surpassed Google search traffic

> 1 billion active montly users on short-form video apps.

65%

Percentage of ALL internet traffic is video streaming, increasing 24% year over year.

40%

Percentage of people 18 to 24 turning to visual-based social media platforms for internet searches.

Why ASICs are needed.

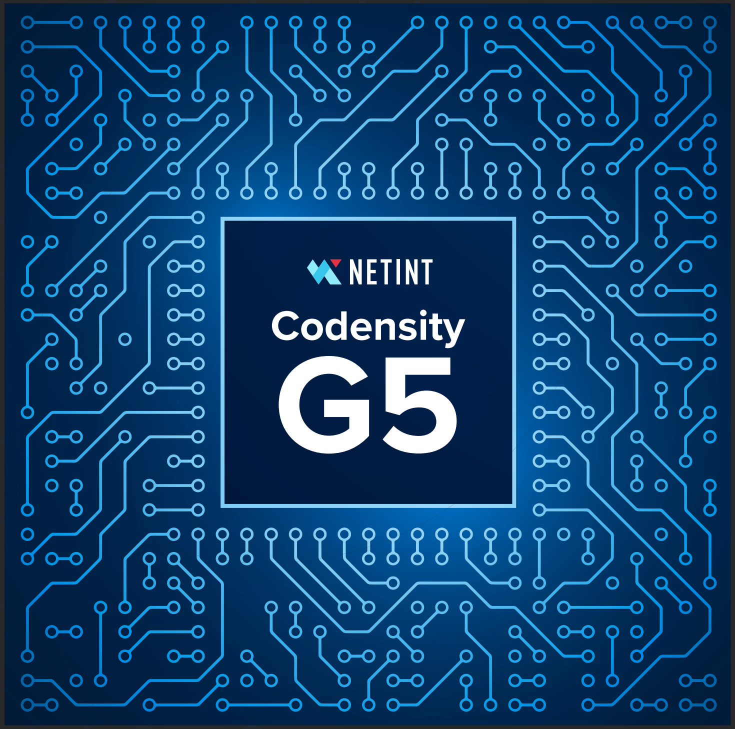
Density is a dirty expensive problem

Global corporations spend 20% of their annual OPEX powering data centers.

Data centers operate 24/7, massively consume energy, and are depleting our planet's resources at an accelerated and unsustainable rate. Today, there are 8,000 data centers globally and their collective consumption is expected to double by 2025.

Our solution.

We designed an ASIC to slash the encoding footprint up to 95%



By replacing video encoding software with ASIC-powered VPUs, you solve two problems:

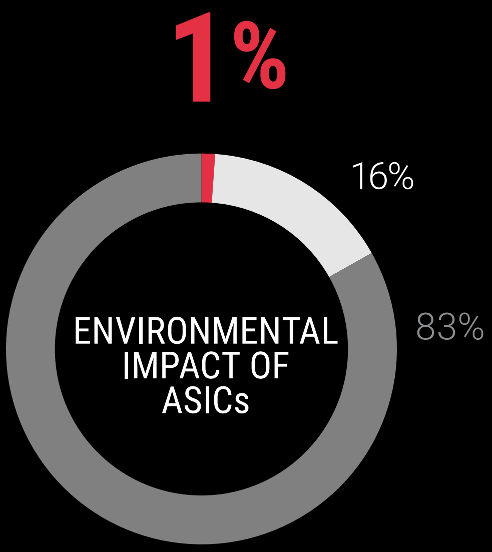
1. Massively increase capacity
2. Exponentially reduce power consumption

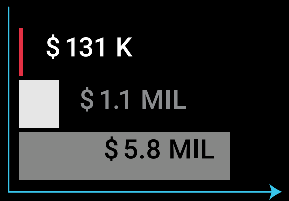
This saves your bottom line and the planet. That’s a win-win.

This is why Google built a custom chip for YouTube.

For everyone else who isn't Google, we did the heavy lifting for you.

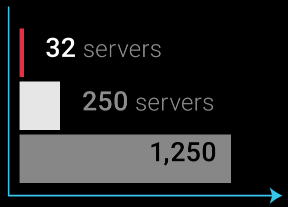
Google's ASIC called 'Argos,' helps YouTube encode and process videos much more efficiently. Argos chips provide "up to 20-33x improvements in compute efficiency compared to its previous traditional server set up," according to a Google executive. Another report suggests that Argos replaced over 10 million Intel CPUs in YouTube.





ANNUAL OPEX

Required to deliver 10K concurrent live HD streams



SERVER DENSITY

Servers required to deliver 10K concurrent live HD streams

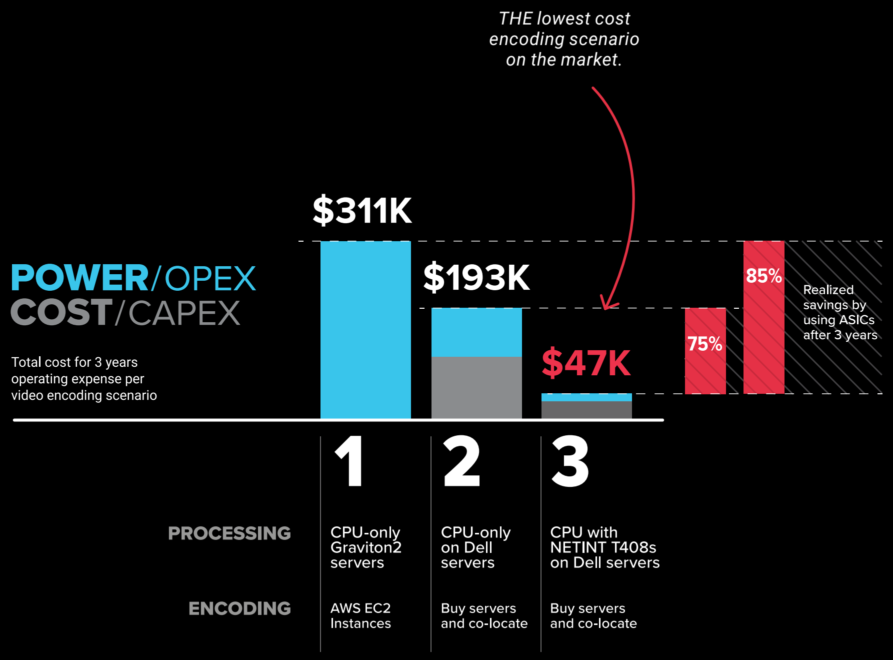
ASIC	*NETINT ASIC video processing units (VPU)*
GPU	*NVIDIA T4 graphic processing units*
CPU	*Intel SVT with CPU-based encoding*

The real cost of live streaming

CAPEX and OPEX comparison using 3 video process/encoding scenarios.

Test assumptions:

- Servers run 100 concurrent five-rung encoding ladders
- x264 very fast preset used for CPU-only processing



SOURCE: https://info.netint.com/hubfs/Whitepaper/Slash_CAPEX_OPEX_Carbon-Emissions-extended.pdf

Quadra
Video Server

VPU | Codensity Quadra G5 ASIC

Built on the Supermicro 1114S-WN10RT server platform, NETINT's Quadra Video Server boasts ultra-high density encoding capacity enabled by ten Quadra video processing units (VPUs).

Supports:

- **HEVC, H.264 and AV1 video encoding**
- **HEVC, H.264, and VP9 video decoding**
- **Up to 8K resolution**
- **10-bit HDR**

Ultra-low latency encoding of up to 320 broadcast quality 1080p30 streams in a compact 1RU form factor. Massive transcoding capacity enables breakthrough reductions of up to 90-95% in OPEX and CAPEX costs compared to software-based encoding systems.

Performance results in this brochure are for the NETINT Quadra Video Server powered by an AMD EPYC™ 7543P (32-core) CPU. For encoding workloads with different encoding demands, the server is also available with the AMD EPYC 7232P (8-core) and 7713P (64-core) CPUs.

Quadra T1U
VPU

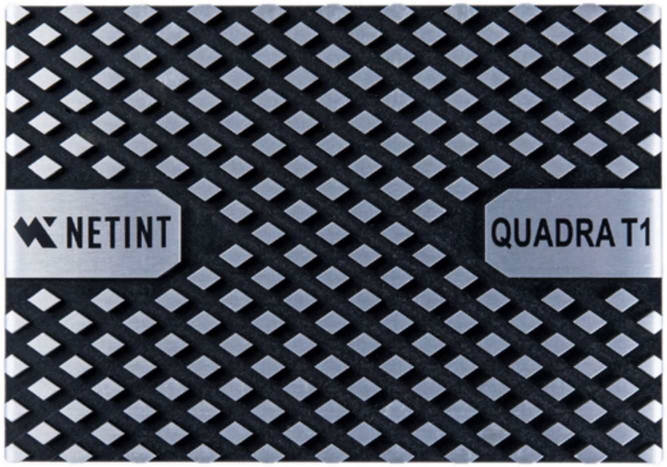
Codensity ASIC G5

The video engine inside.

The Quadra T1U is a U.2 form-factor video processing unit with one Codensity G5 ASIC. Operating in x86 and Arm-based servers, the T1U enables video operators to move from software to hardware-based encoding to power real-time video applications and the metaverse at a TCO that is up to 85-99% lower than CPU-based solutions.

The Quadra T1U uses the Codensity G5 ASIC to support AV1, HEVC, and H.264 real-time video encoding at up to 8K resolution with 10-bit HDR. The Quadra T1U's exceptional throughput enables ultra-low latency encoding of up to 16 broadcast quality 1080p60 streams using AV1, HEVC, or H.264.

The addition of two Deep Neural Network engines capable of 18 trillion operations per second (TOPS), enables functions such as object detection, classification and segmentation, and ROI to provide additional data to the encoding engine for image quality improvement and content-adaptive rate control.



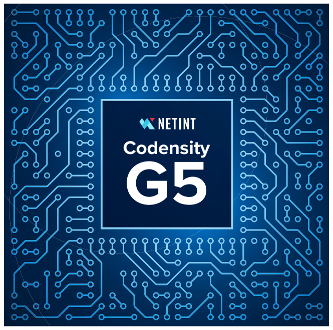
Codensity G5 ASIC

Application Specific Integrated Circuit

ASIC Video Transcoder

The Codensity G5 architecture uniquely combines on-chip AV1, HEVC, and H.264 video encoding and AI processing engines to deliver scalability for metaverse, live streaming, and interactive applications.

The core of NETINT's Codensity technology is an in-house built ASIC that increases encoding density compared to CPU-based software encoding solutions. This increase in encoding density expands the number of channels that can be encoded without increasing the rack footprint. This reduces power and HVAC costs to deliver a lower TCO without sacrificing video quality or latency.



8K UHD Video Encoding

The Codensity G5 ASIC enables up to 8K video transcoding using the HEVC and H.264 codecs (AV1 is limited to 4K). Advanced codecs like AV1 and HEVC deliver superior quality to H.264 with up to a 60% reduction in bitrate, but when produced by CPU-only encoders, can require up to 10x the processing power, limiting throughput severely. HEVC and AV1 output with the Codensity G5 ASIC should be similar to H.264, making 4K and 8K live resolutions affordable and scalable for the first time.

Flexible Architecture

The Codensity G5 is built on a programmable microprocessor architecture to optimize the firmware and pipeline processing for improved performance and increased video quality. This counters a criticism that silicon-based encoders lack upgrade flexibility.

AI Engine

Two Deep Neural Network engines capable of 18 trillion operations per second (TOPS) enable object detection, classification, and segmentation to provide additional data to the encoding engine for image quality improvement and content-adaptive rate control for advanced performance and functionality. Seamlessly integrated for region-of-interest (ROI) encoding and background replacement. Additional features to be released.

Designed for the Cloud

High-density live UHD transcoding

The NETINT Quadra VPU takes full advantage of the video processing capability inside the Codensity G5 ASIC to support H.264, HEVC, and AV1 HEVC live encode functionality of up to 8K UHD video in SDR and HDR with HDR10 and other popular high dynamic range standards. By offloading complex encode and decode processing to the Codensity G5 ASIC, the Quadra VPU minimizes host CPU utilization. The result is a significant improvement in real-time transcoding density compared to any software or GPU-based transcoding solution.

Every NETINT Quadra Video Server installed in a data center would replace as many as 25 software-based video encoding servers. *(See appendices for details).*

High power efficiency

Each NETINT Quadra U.2 module consumes only 20W of power at full load. This makes the Quadra Video Server, the most energy efficient video transcoder available.

Enterprise NVMe integration

Deployed in a U.2 form factor (and also available in HHHL AIC form factor), Quadra offers a simple upgrade path from CPU-based software to ASIC video encoding on any enterprise-class server.

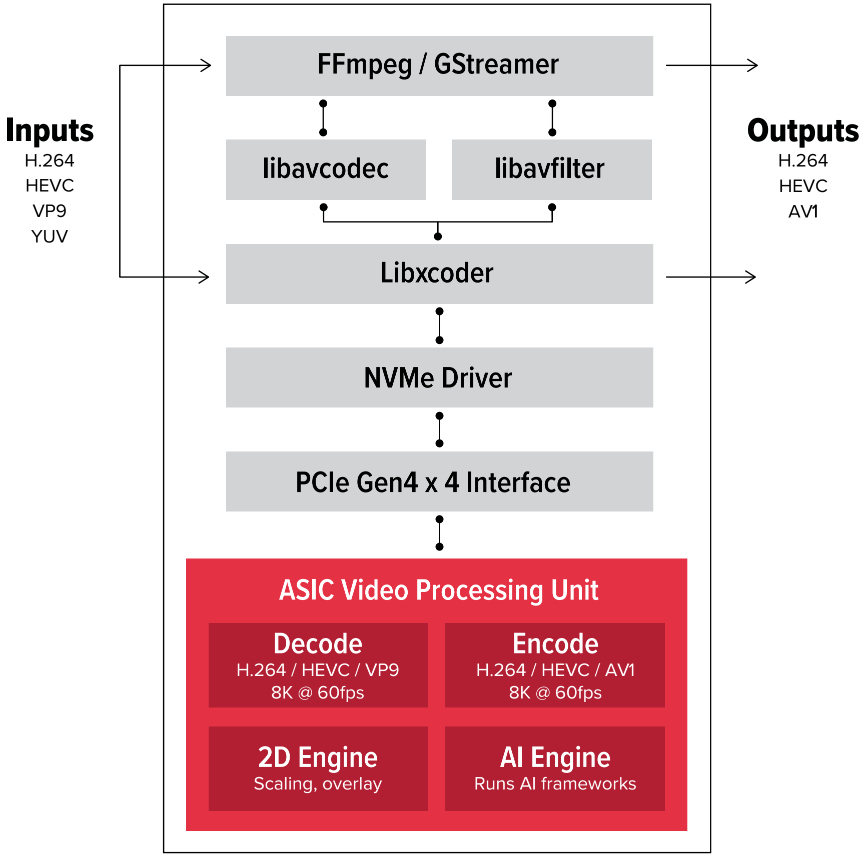
NETINT's Quadra Video Server hosts ten Quadra VPUs supporting up to 320 simultaneous live 1080p30 encoding sessions.

Simple Integration

Open-source suite of processing tools.

Many video processing and transcoding applications developers use FFmpeg and GStreamer, two open-source software libraries offering a vast suite of video processing functions. The Quadra video server includes highly efficient FFmpeg and GStreamer compatible SDKs, allowing operators to apply an FFmpeg or GStreamer patch to complete the integration.

The libavcodec patch on the host server functions between the Quadra NVMe interface and the FFmpeg and GStreamer software layers simplifying integration and enabling fast and efficient performance and capacity upgrades.



Specifications

Quadra
Video Server

VPU | Codensity Quadra G5

CPU Options	AMD EPYC™ 7232P Server Processor (8-core)
	AMD EPYC 7543P Server Processor (32-core)
	AMD EPYC 7713P Server Processor (64-core)
Operating System	Ubuntu 20.04.05 LTS *(as of May 2023)*
Memory	8x 16GB DDR4-3200
Storage	400GB M.2 SSD
NVMe Support	10x
PCIe Expansion	Up to 3x PCIe slots
Network Options	Dual 10GBase-T LAN
Power Supply	700W: 100 - 140Vac
	750W: 200 - 240Vac
	750W: 200 - 240Vdc (CCC only)
Transcoders	10x NETINT Quadra T1U
Encoding Capacity	Up to 40 4Kp60 or 320 1080p30
Codec Support	H.264 - Encode/Decode
	HEVC - Encode/Decode
	VP9 - Decode
	AV1 - Encode
Transcoder Software	FFmpeg, GStreamer

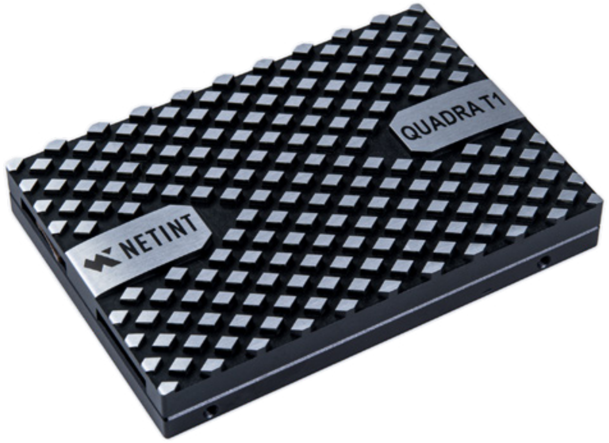
Physical Dimensions	W: 17.2” (437mm), H: 1.7” (43mm), D: 23.5” (597mm)
Rack Size	1U
Weight	39 lbs (17.69 kg) *(includes 10 processors)*
Environmental	50 degrees F to 95 degrees F Operating Temperature, 8% to 90% Operating Relative Humidity
Power Inputs	100 - 140Vac / 8 - 6V / 50-60Hz
	200 - 240Vac / 4.5 - 3.8A / 50-60Hz
	200 - 240Vdc / 4.5 - 3.8A (CCC Only)
Certifications	RoHS Compliant, UL Approved

NETINT

Specifications

Quadra T1U VPU

Codensity ASIC G5



Form Factor	U.2 (SFF-8639)
Interface	PCIe 4.0 x4
Protocol	NVMe
Power Consumption (Typ)	20W
Usage	24/7 Operation
Operation Temperature	0 degrees C to 50 degrees C
RoHS Compliance	Meets requirements of European Union (EU) ROHS Compliance Directives
Product Health Monitoring	Self-Monitoring, Analysis, and Reporting Technology (SMART) commands Temperature Monitoring and Logging
Hardware Interface	Available U.2 slot

	H.264 AVC Encode/Decode	**H.265 HEVC Encode/Decode**	**AV1 Encode**
Profile	CBP / BP / XP / MP / HiP / Hi10P	Main / Main 10	Main
Level	1 to 6.2	1 to 6.2 Main Tier	1 to 5.3
Min / Max Resolution	Min: 32 x 32 / Max: 8192 x 5120		
Scan Type	Progressive		
Bitrate	64kbit/s to 700Mbit/s		
VP9 decode	Profile 0, Profile 2, Level 6		
Software Integration	FFmpeg and GStreamer SDKs. Direct Integration with LibXcoder API		
Capacity	Up to 16x 1080p60, 4x 4Kp60, 1x 8Kp60		

Region of Interest (ROI)	ROI enables the quality of some regions to be improved at the expense of other regions
Closed Captioning	Quadra supports EIA CEA-708 closed captions for H.264 and HEVC encode and decode
High Dynamic Range (HDR)	Quadra supports HDR10 & HDR10+ for H.264 & HEVC encode and decode
Low Latency	Quadra supports sub-frame latency
IDR Insert	Forced IDR frame inserts at any location
Flexible GOP Structure	8 presets plus customizable GOP structure

NETINT

HYPERSCALE PROFITABLY

For more information on NETINT
encoding solutions, contact us.

go@netint.com
netint.com