

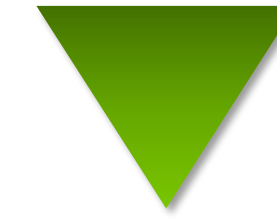
An abstract background featuring a complex, repeating pattern of green, three-dimensional rectangular blocks. These blocks are arranged in a way that creates a sense of depth and movement, resembling a stylized, digital landscape or a complex architectural structure. The pattern is dense and covers the entire frame, with varying shades of green and black, giving it a high-tech, digital feel.

현실 세계에서 옴니버스 세계까지 가속화하는 NVIDIA RTX

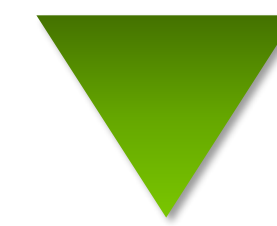
2022. 01. 21
BNI&C 김영호상무



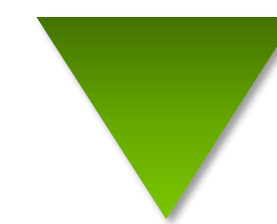
World is Changing



People are Changing



Values are Changing



Business is Changing

Platform & Tools are Changing

Data

TEXT ► Image ► Audio&Video ► ?

AI/ML

Metaverse

Digital
Transformation

XR

NVIDIA RTX IN REAL WORLD



NVIDIA RTX PLATFORM

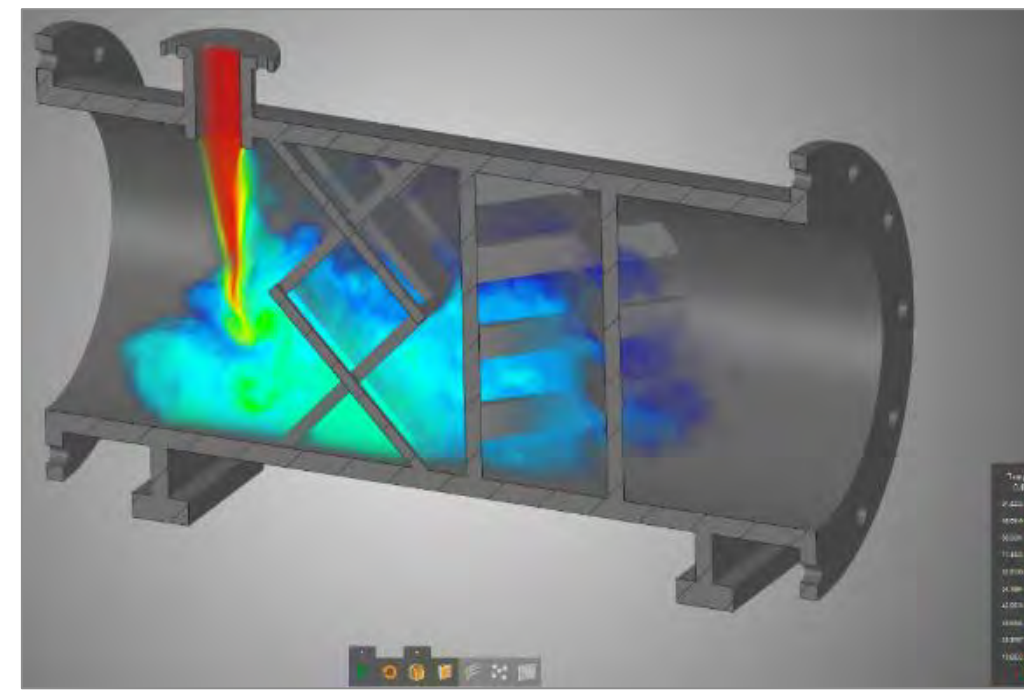
RENDERING



VIRTUAL
WORKSTATION



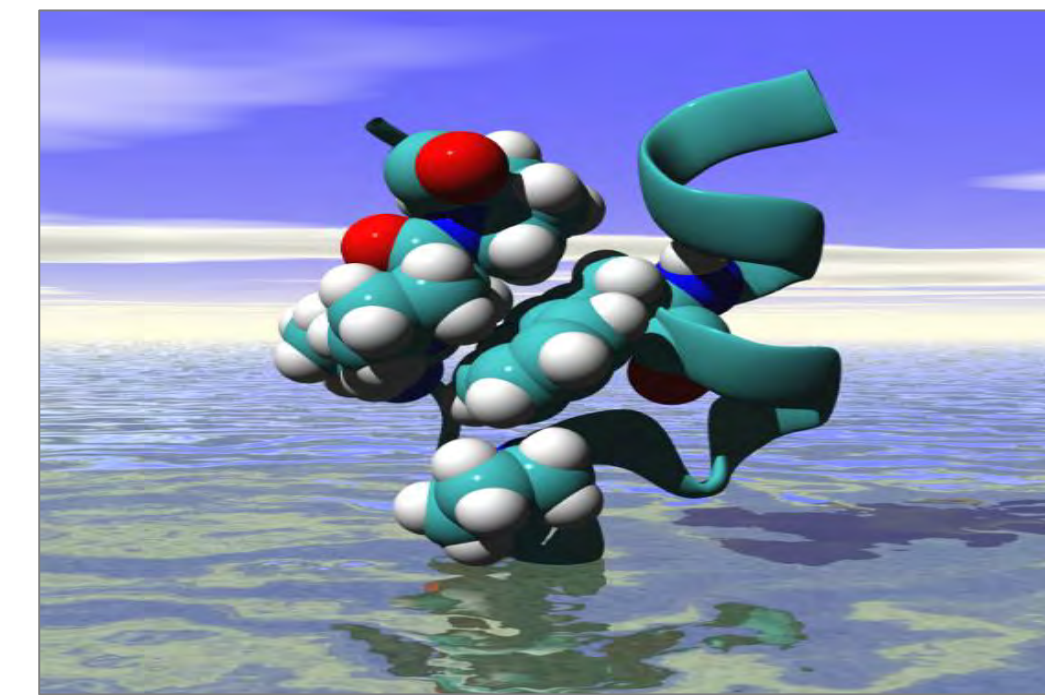
SIMULATION



DATA SCIENCE



COMPUTE



XR



NVIDIA OMNIVERSE ENTERPRISE



LAPTOPS



DESKTOPS



DATA CENTER

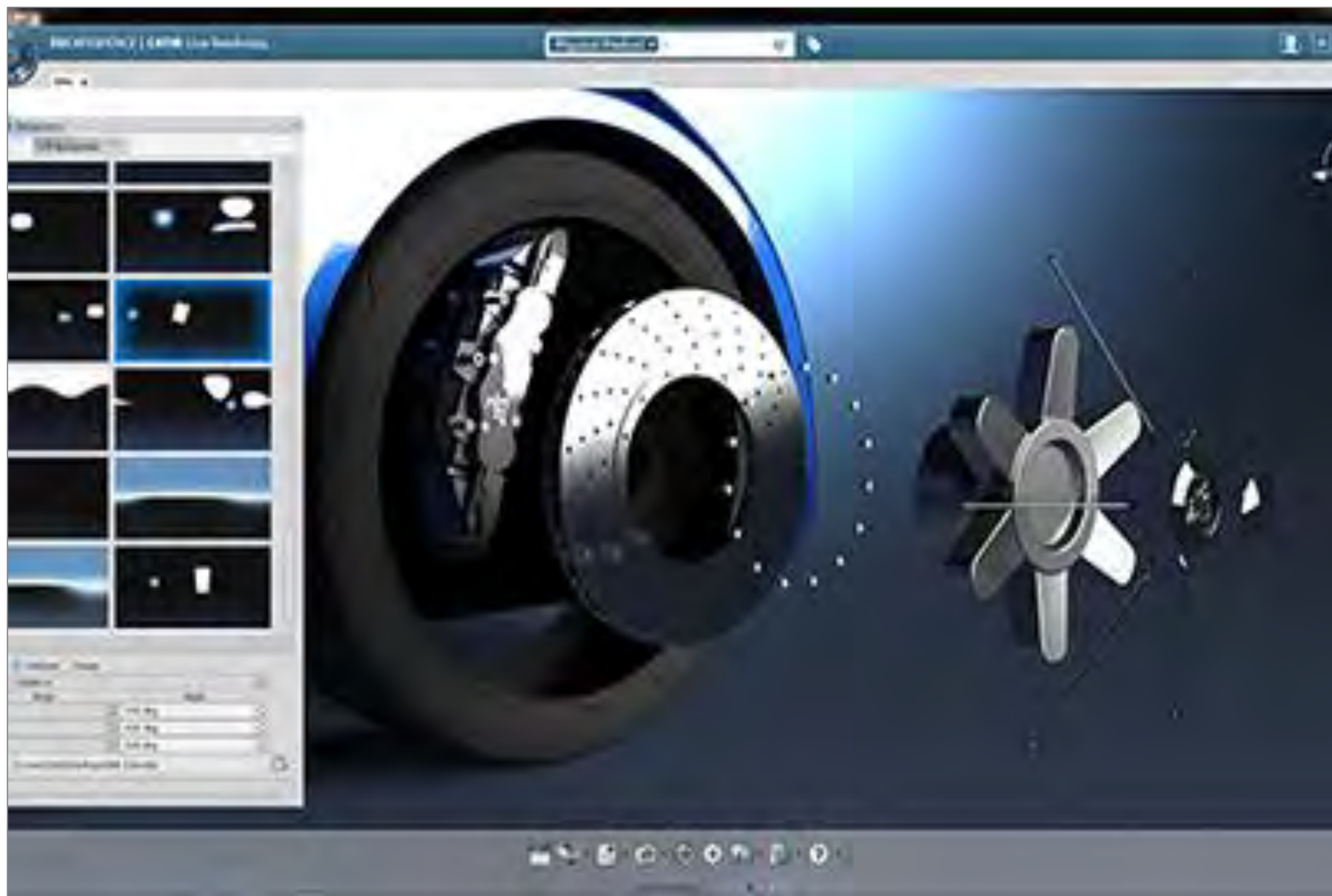


EMBEDDED

THE NEW ERA OF WORK

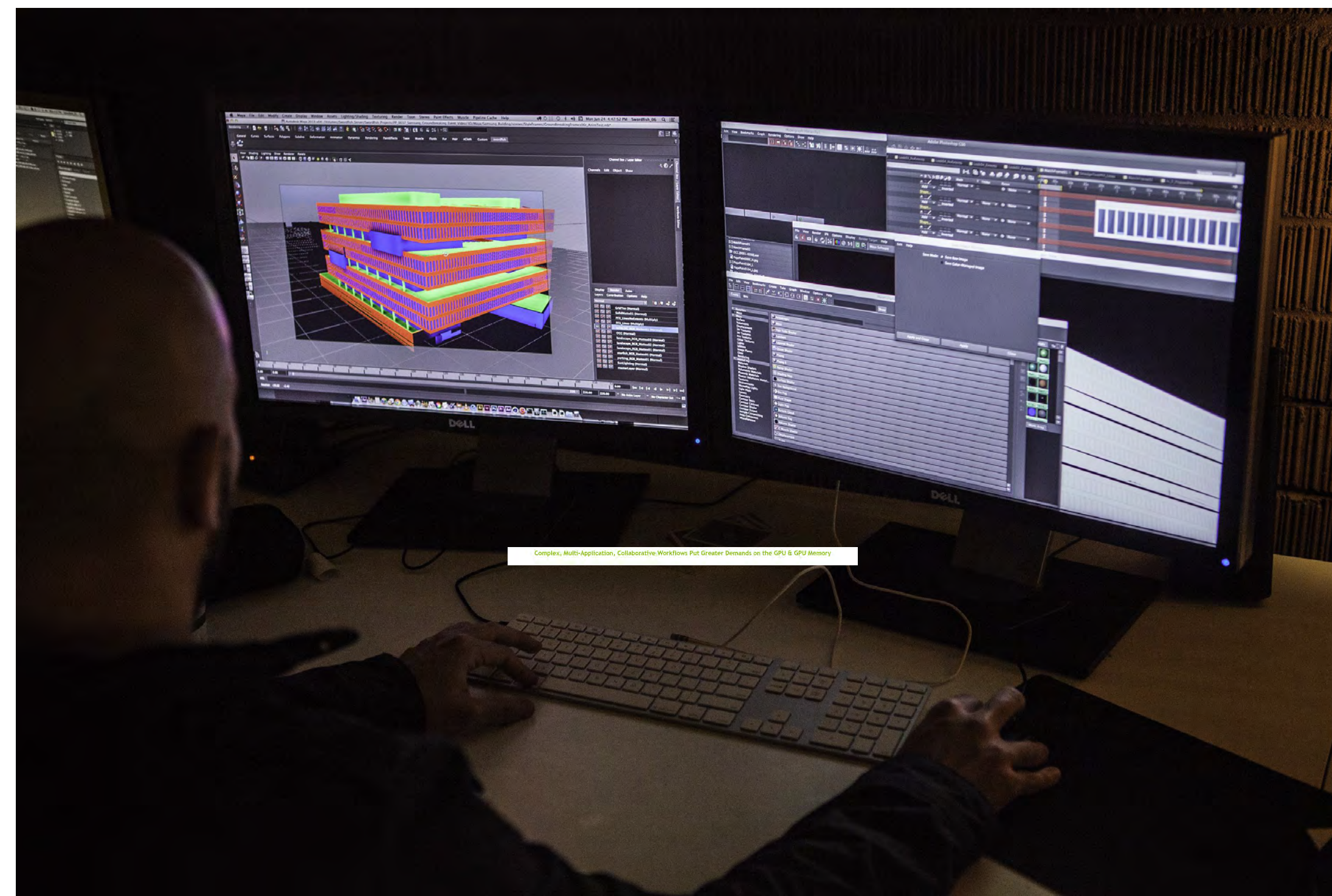
Complex, Multi-Application, Collaborative Workflows Put Greater Demands on GPU Memory

VISUALLY COMPLEX CONTENT



High Fidelity 2D, 3D, Video Content

MULTI-APPLICATION WORKFLOWS



Modern Workflows Require Multiple Applications

REMOTE COLLABORATION



Do Any Work from Anywhere

Complex, Multi-Application, Collaborative Workflows Put Greater Demands on the GPU & GPU Memory

AI-ASSISTED REAL-TIME RAY TRACING

Concurrent
Float & Int



Faster parallel processing
and better multi-tasking

Variable Rate
Shading



Accelerates performance of visually rich and
geometrically complex scenes

Mesh Shaders



RT Cores



Bring real-time ray
tracing to professional
graphics workflows

Tensor Cores



Enable AI-augmented
tools & applications like
denoising, generative
design

PHYSX



BLAST



BLAST



FLOW



FLOW



FLOW

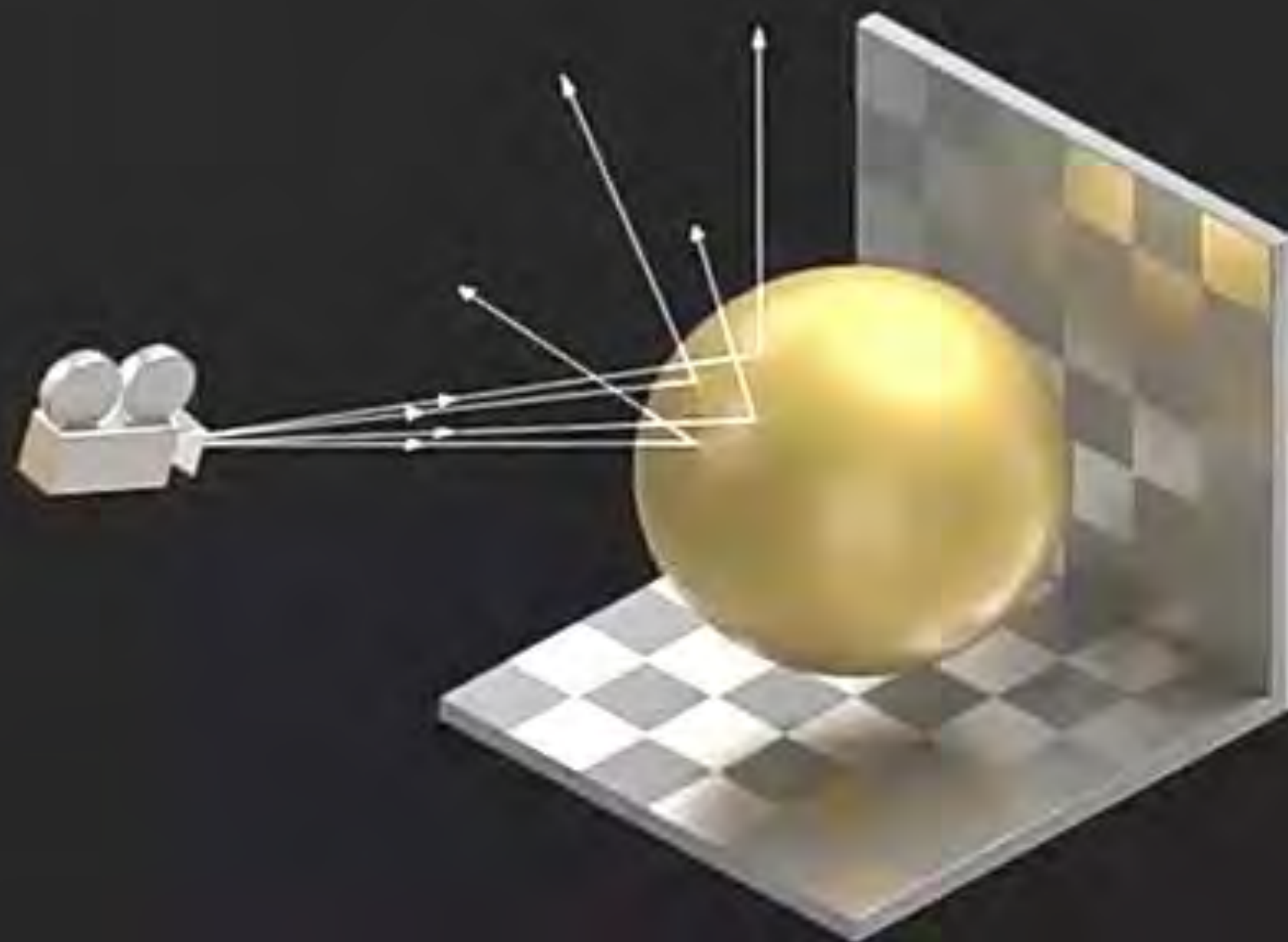


AMPERE ARCHITECTURE — 2nd Gen RTX



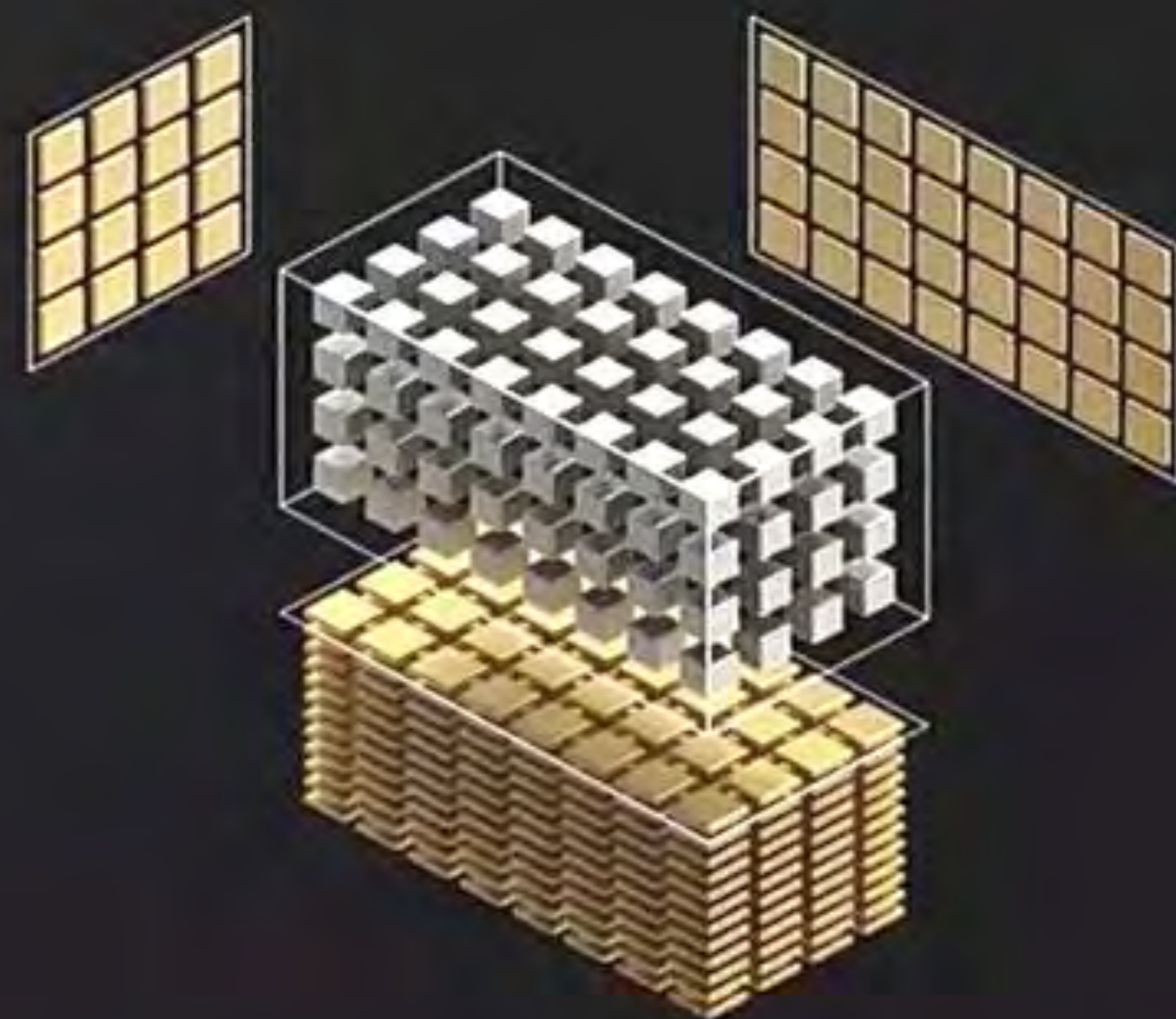
NEW SM
30 FP32 TFLOPS

2.7X



NEW RT CORE
58 RT TFLOPS

1.7X

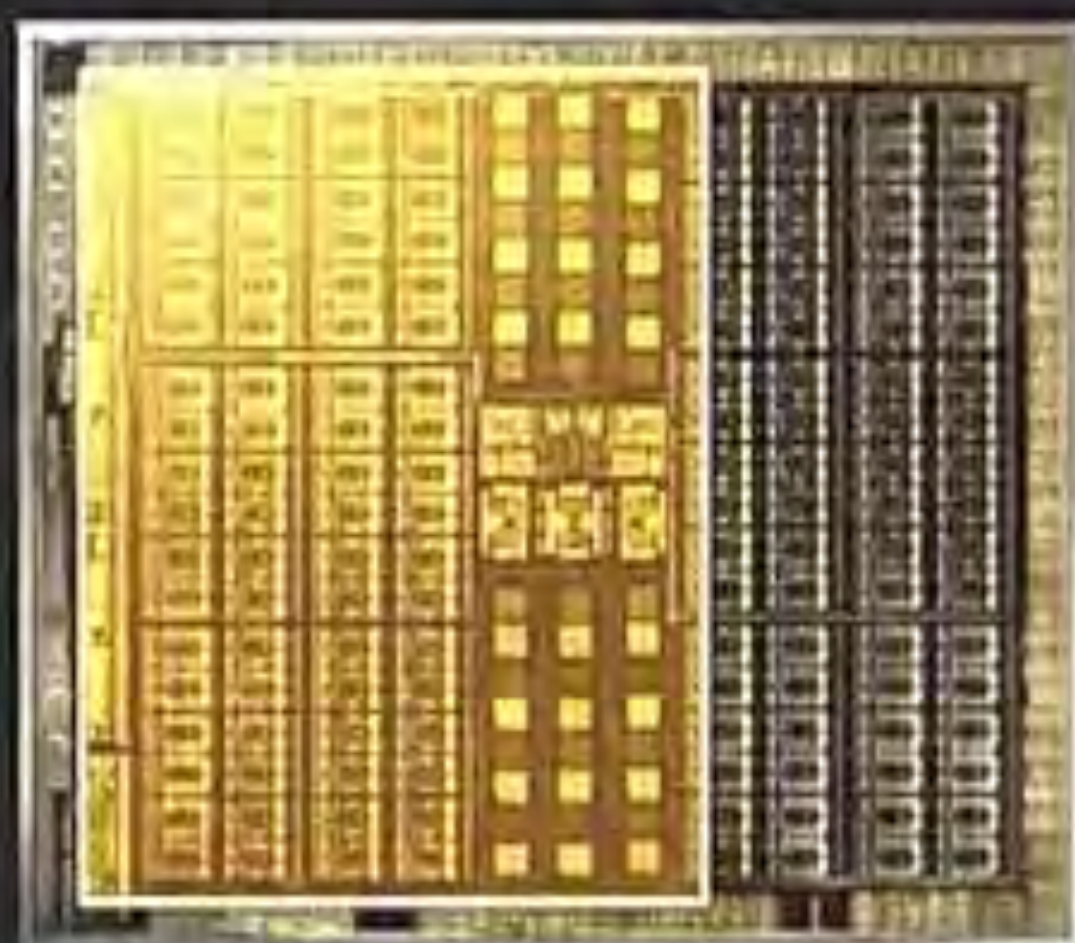


NEW TENSOR CORE
238 Tensor TFLOPS

2.7X

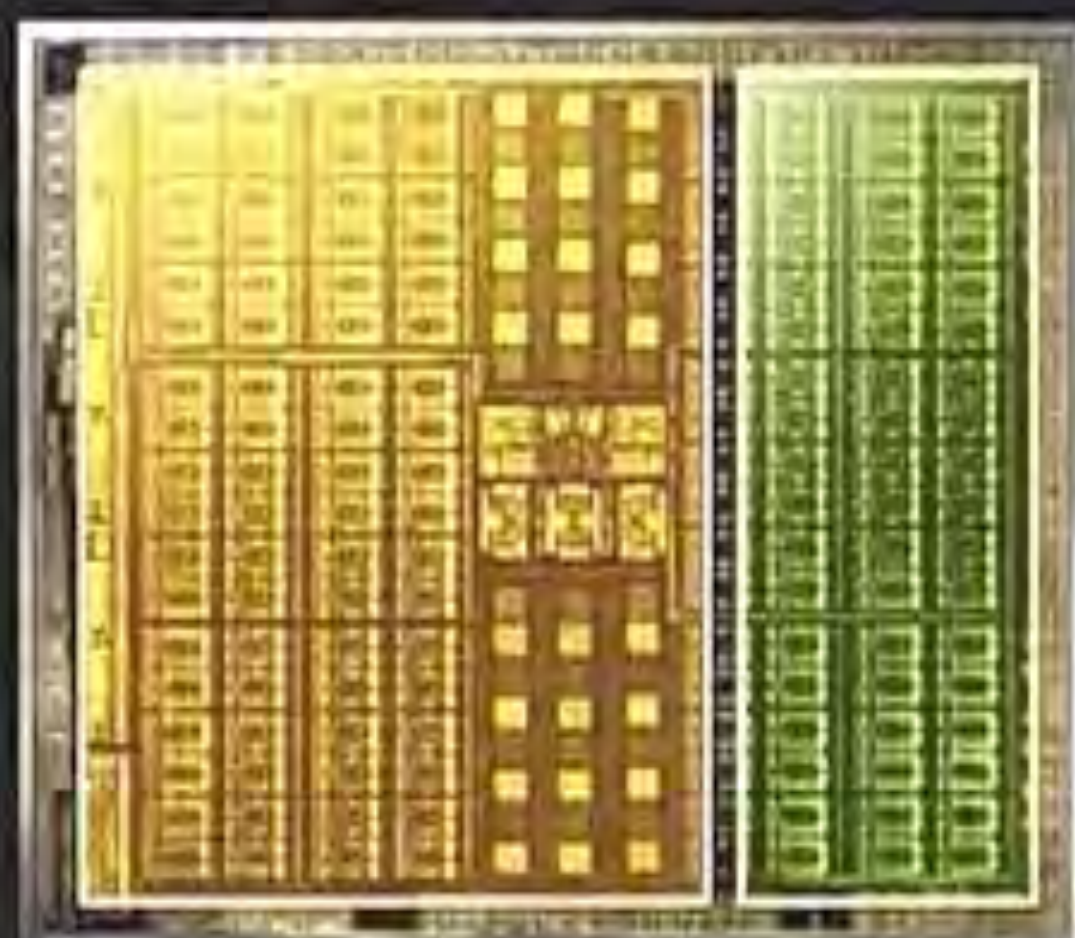
RTX - Dedicated cores to maximize ray tracing performance

Shader



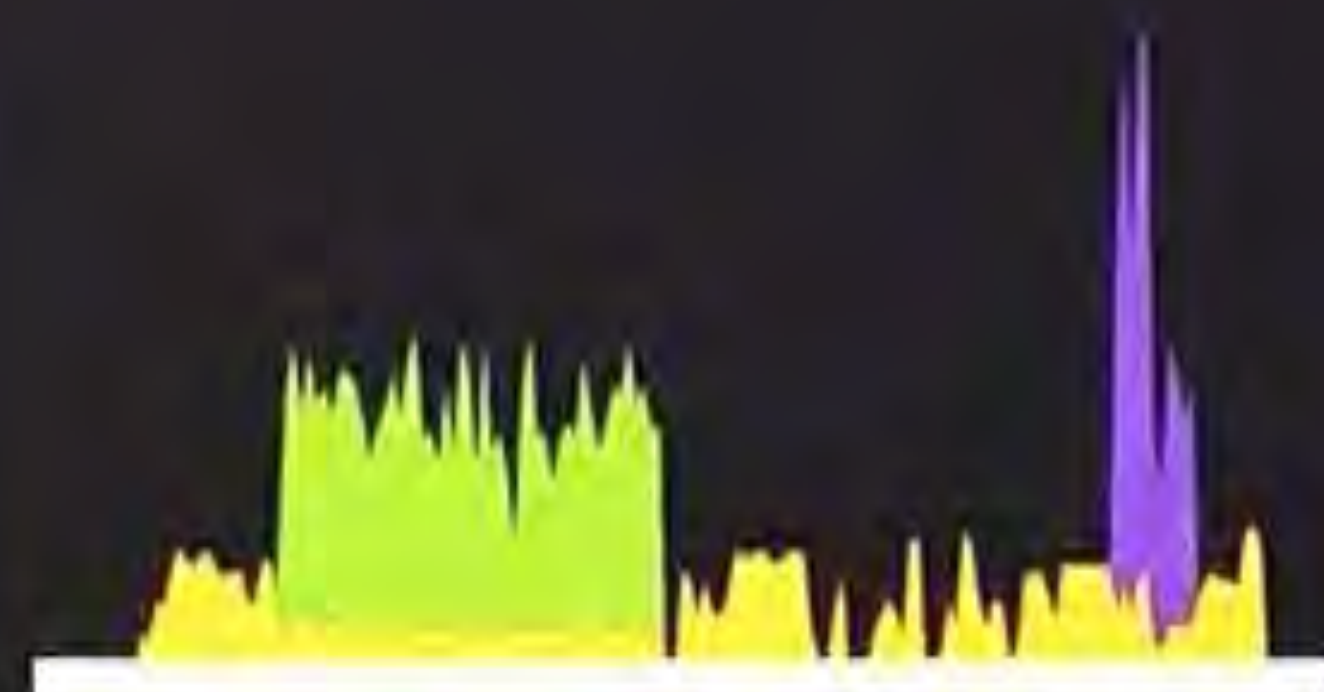
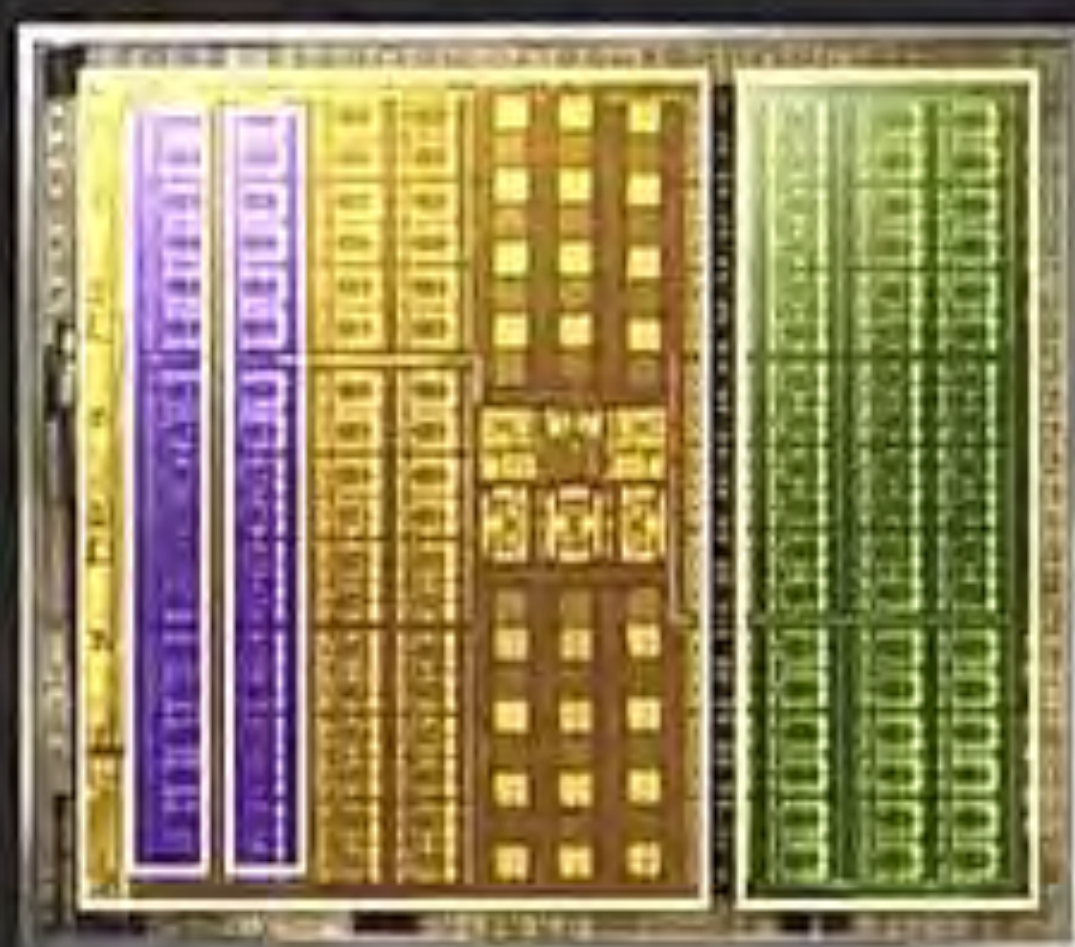
From 92ms to 51ms

Shader
+
RT Core



20ms

Shader
+
RT Core
+
Tensor Core



12ms

THE MAGIC OF DEEP LEARNING



CHARACTER
CONCEPTING



AUDIO TO FACIAL
ANIMATION



CHARACTER
LOCOMOTION



VIDEO TO 3D



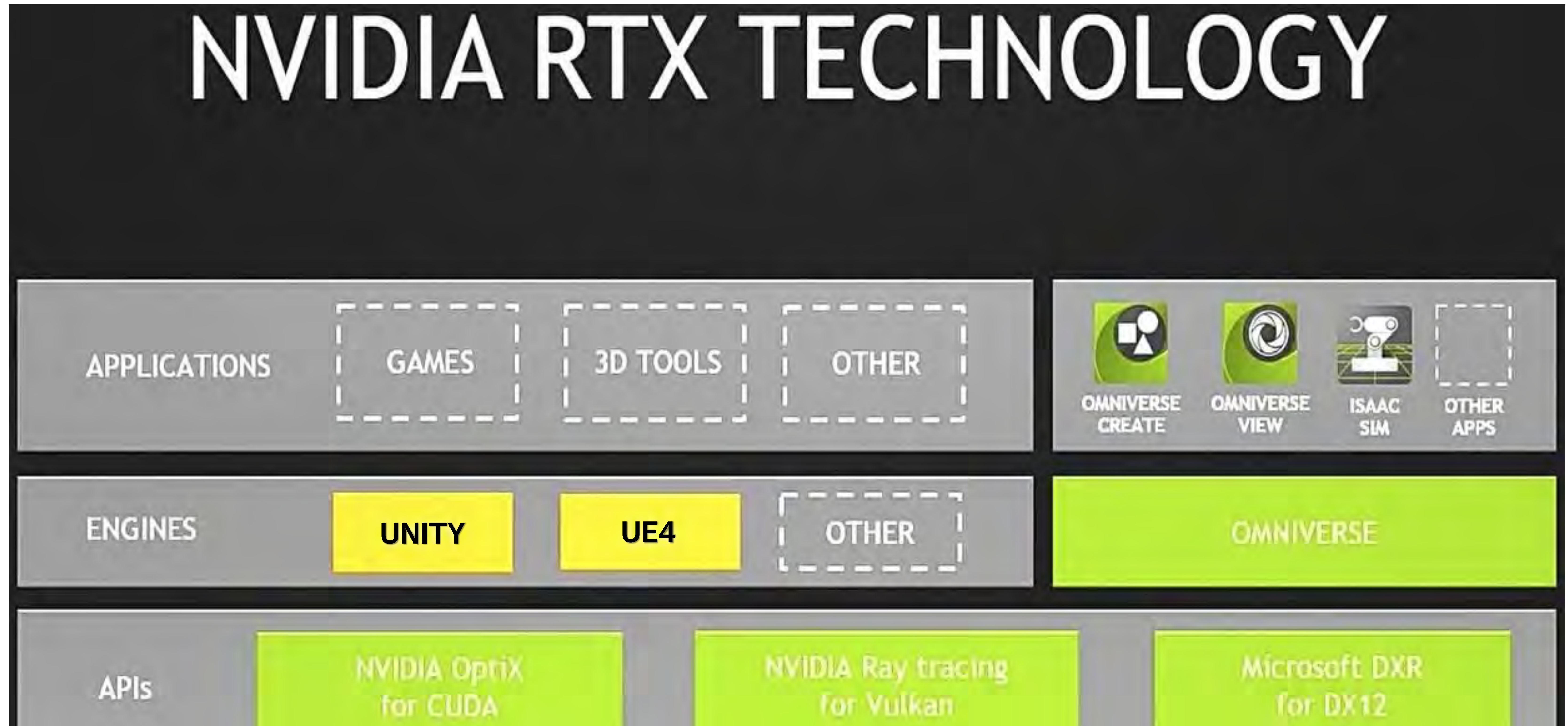
PHYSICS SIMULATION

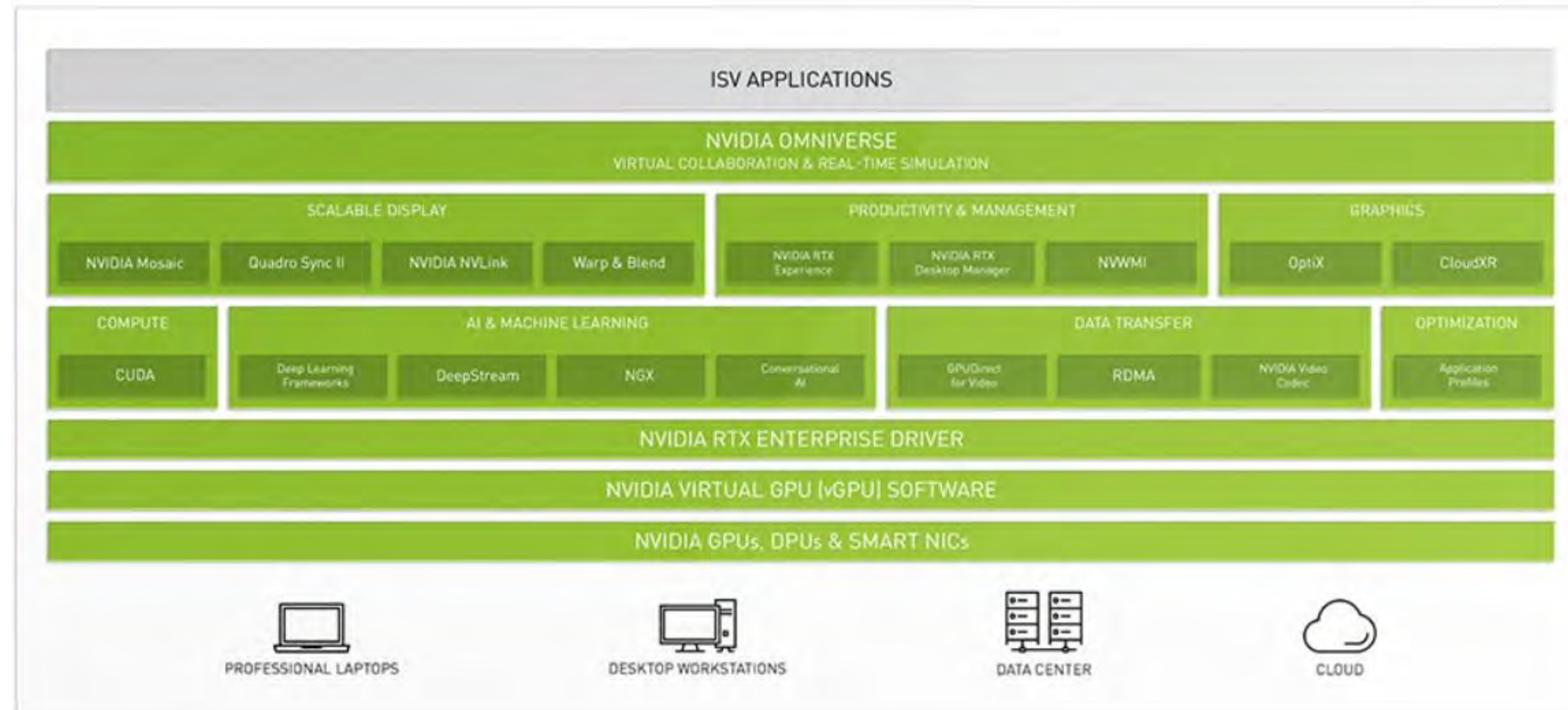


RT DENOISING

Clothing models from UC Berkeley Garment Library

NVIDIA RTX TECHNOLOGIES



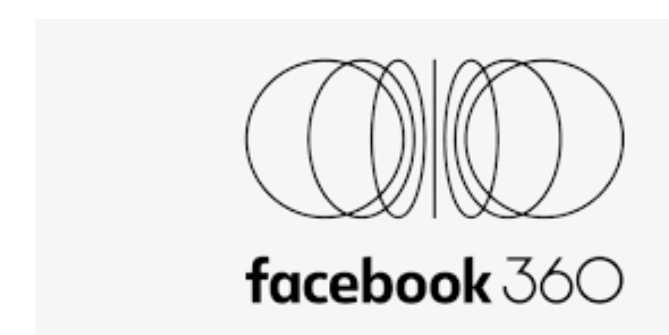


METaverse IN OMNIVERSE



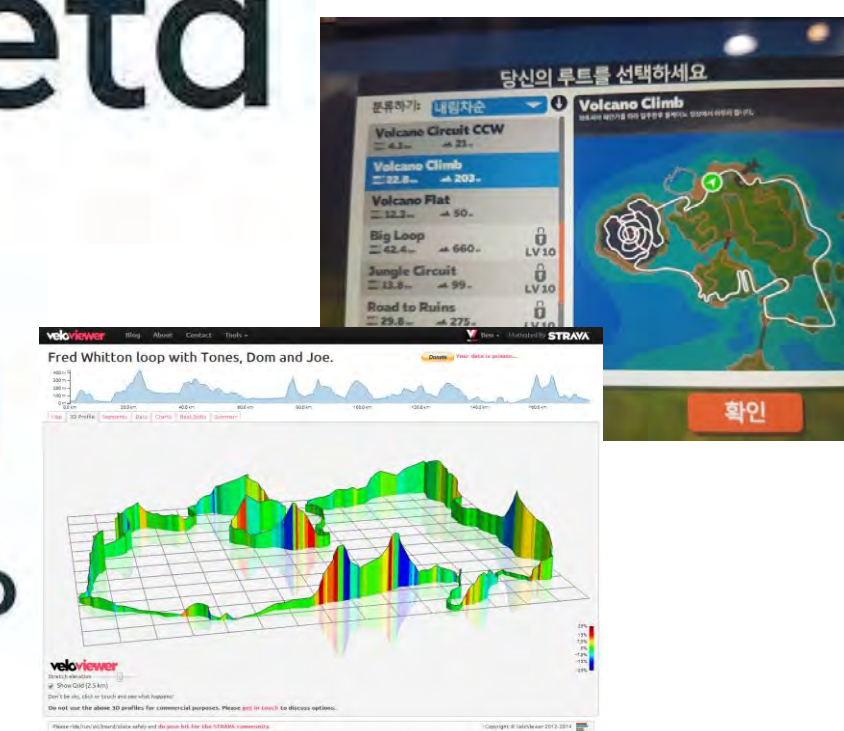


Augmentation



Augmented Reality

Lifelogging



Interface & Networks

External

Sensors & Everyware

Identity & Interaction

Intimate

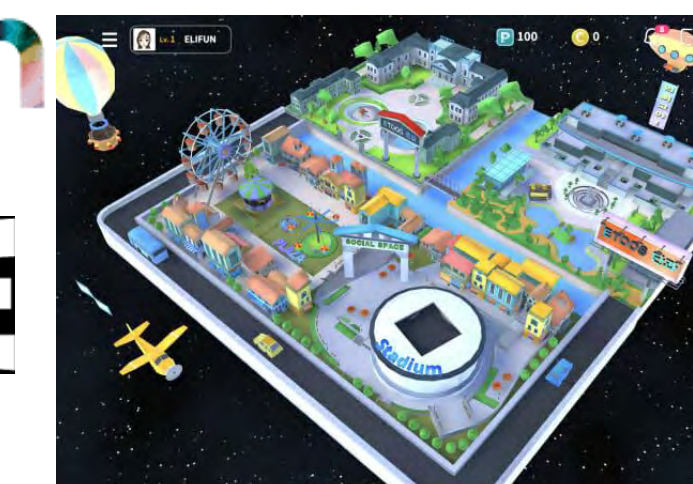
Mirror World

Virtual World

Models & Immersion



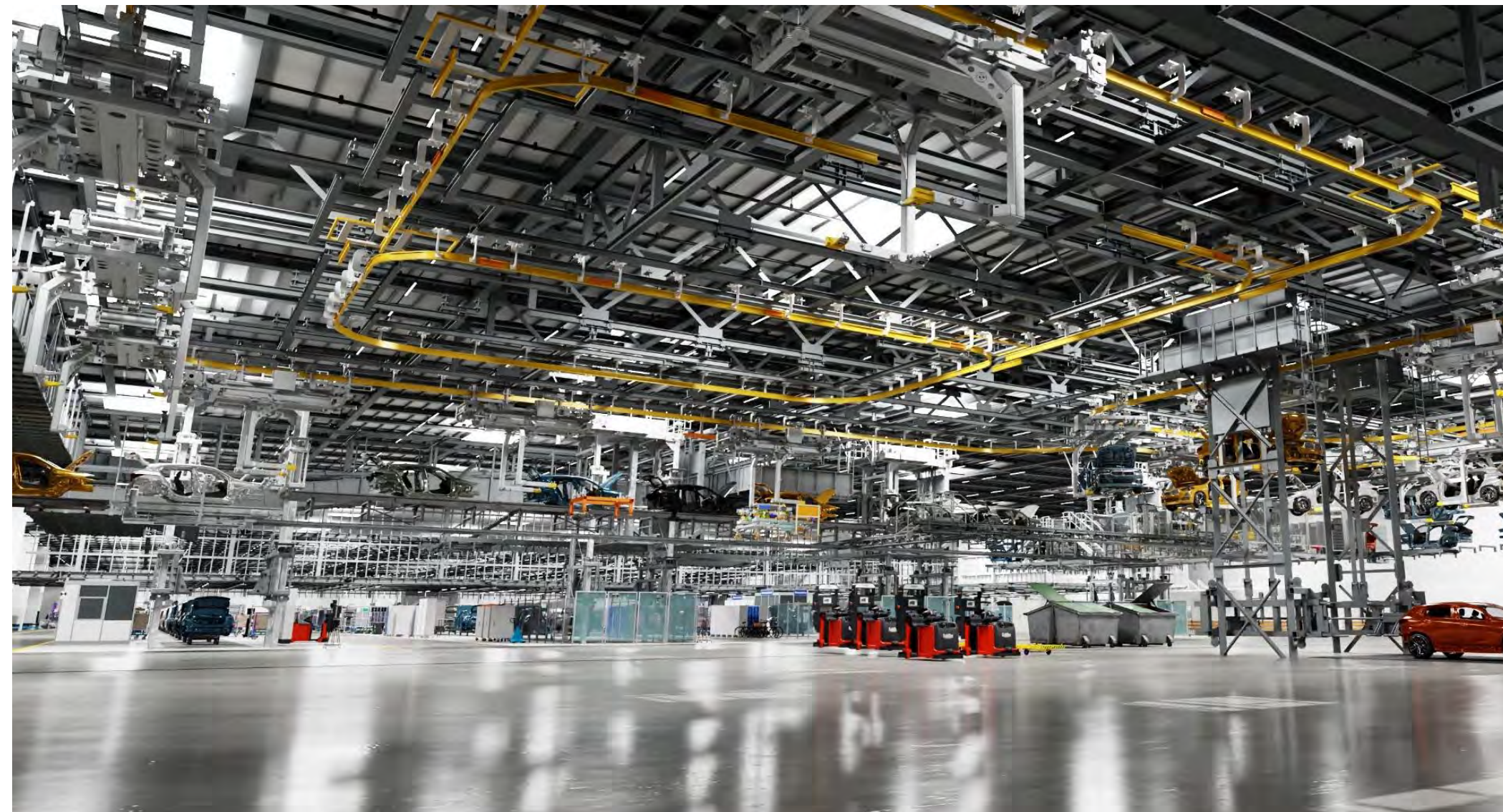
Simulation



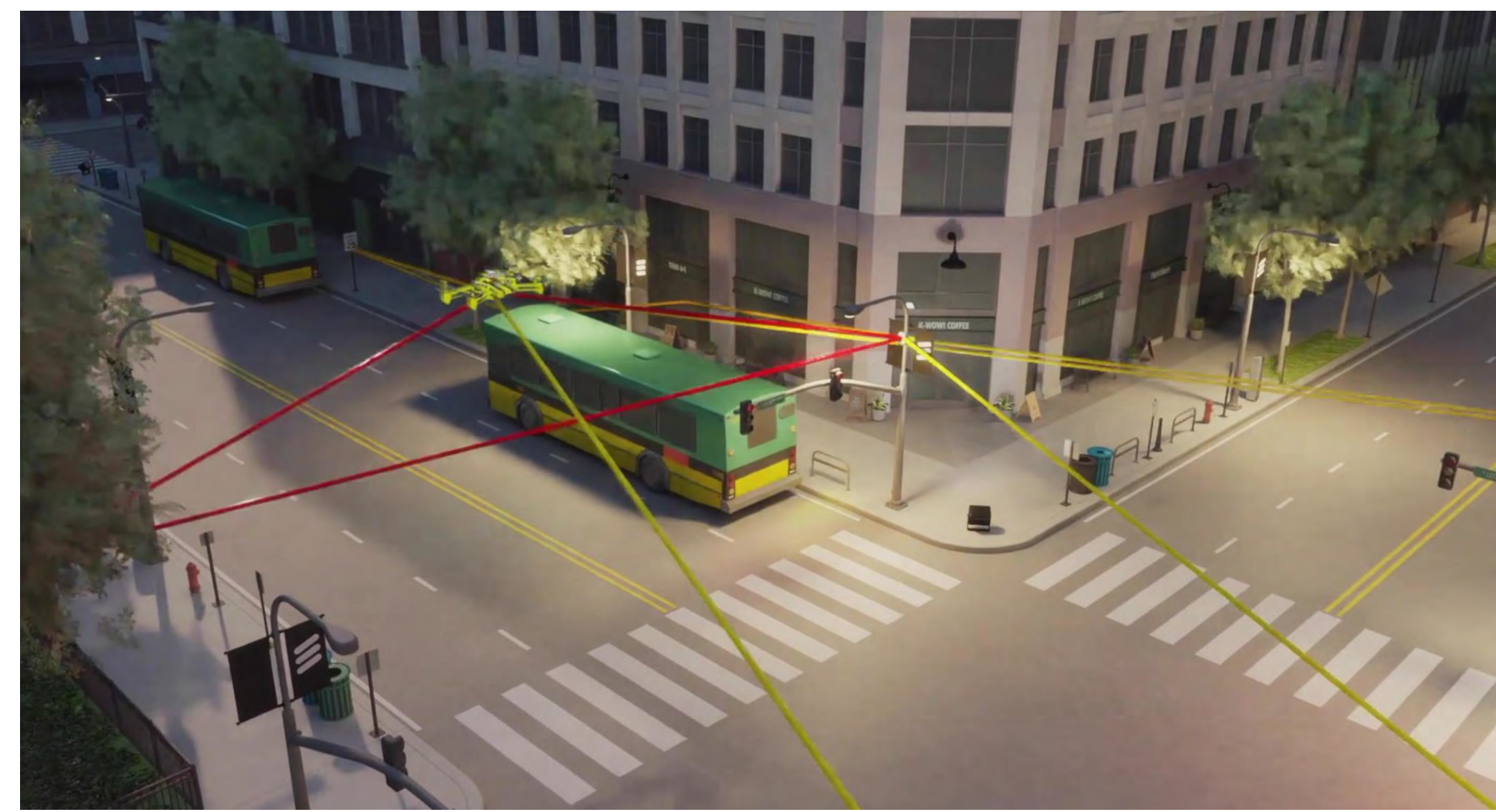
♥ NVIDIA OMNIVERSE PLATFORM & APPS ♥

NVIDIA OMNIVERSE

Active Evaluation Across Industry Leaders



BMW Group
Inventing the Factory of the Future



Ericsson
Simulating 5G Signal Propagation



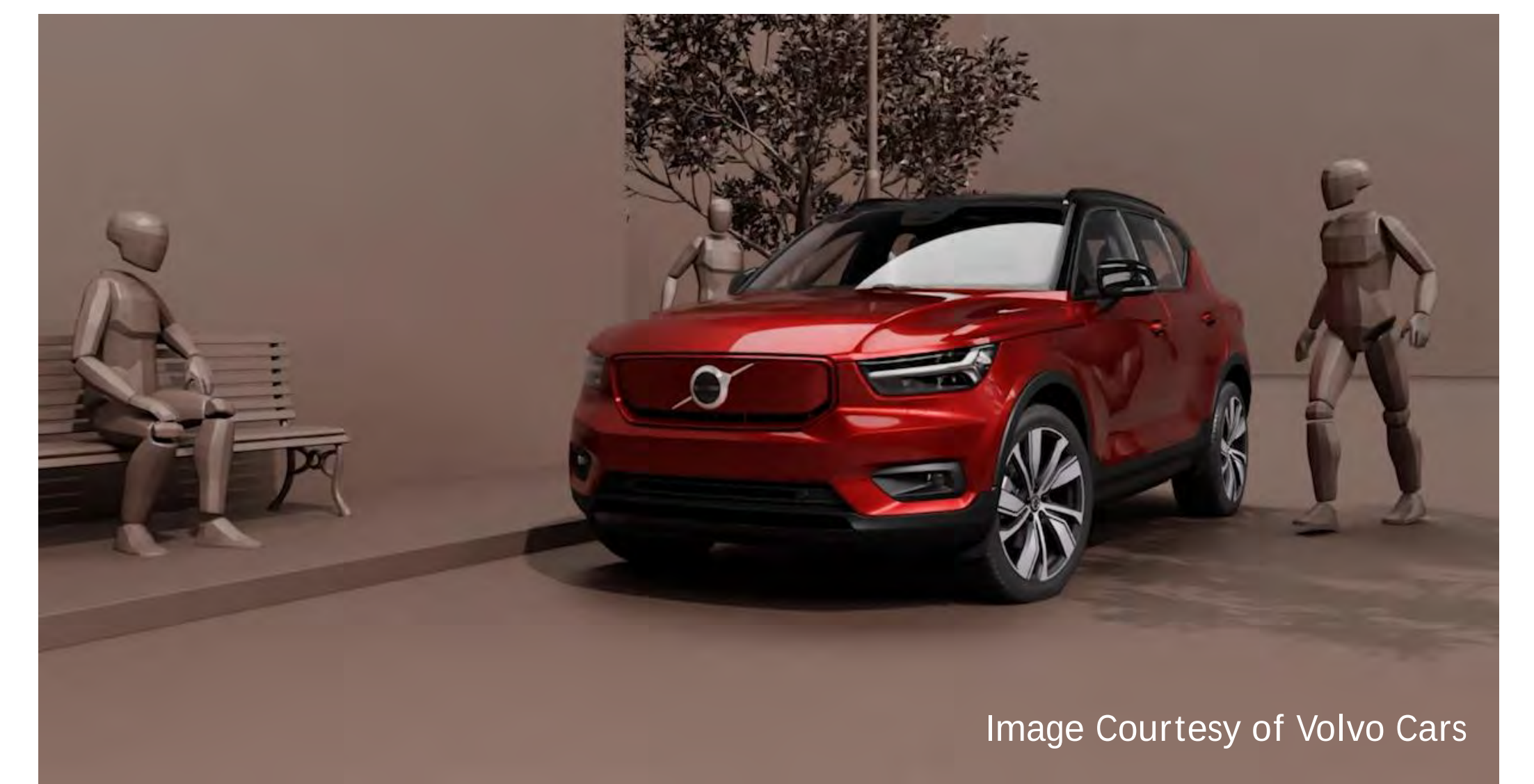
Industrial Light and Magic
Transforming VFX Workflows



Kohn Pedersen Fox
Paving the Way to Next-Gen Design



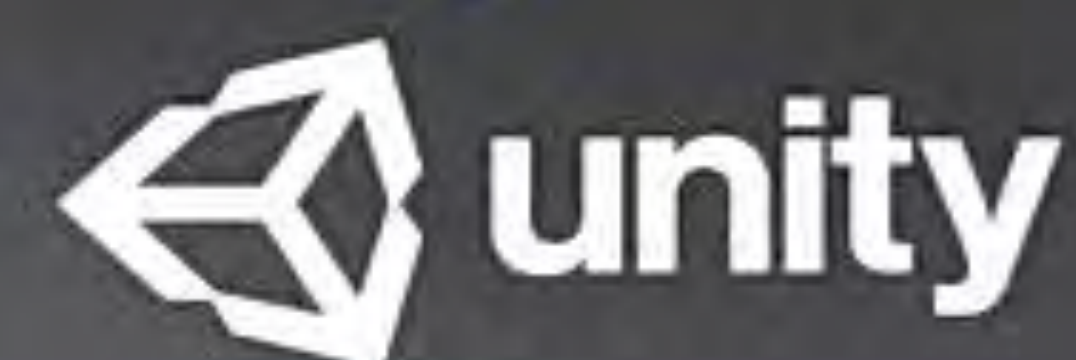
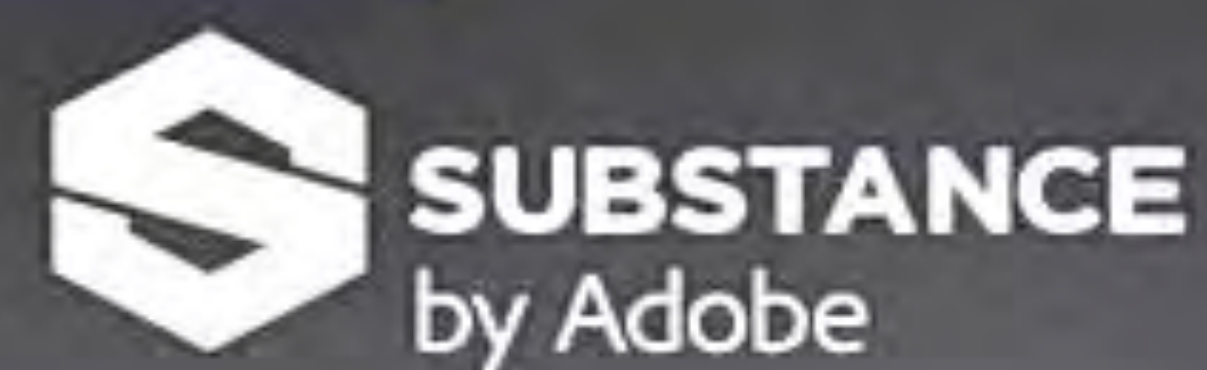
WPP
Building the Future of Advertising



Volvo Cars
Reinventing Manufacturing, Marketing, and the Customer Experience



NVIDIA OMNIVERSE



THE OMNIVERSE PLATFORM



Top Industry Tools

+



+



Open Standards

Realtime Collaboration

OMNIVERSE APPLICATIONS



FOR GAME DEVELOPERS, ANIMATORS



FOR DESIGNERS, CREATORS, ENGINEERS



FOR 3D DEEP LEARNING RESEARCHERS



FOR GEFORCE RTX GAMERS

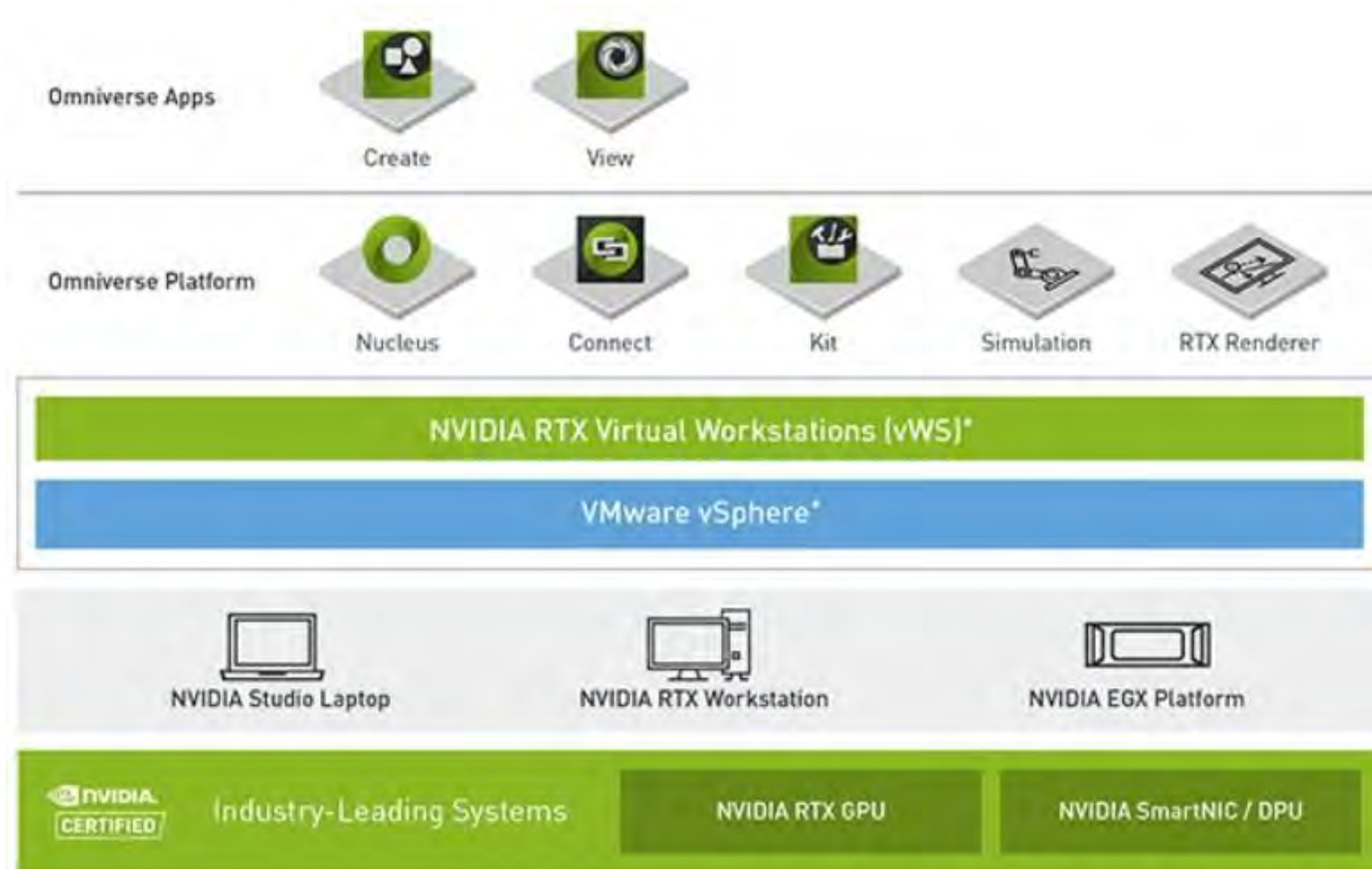


FOR ARCHITECTS, DESIGNERS, ENGINEERS



FOR ROBOTICISTS, SIMULATION SPECIALISTS

NVIDIA OMNIVERSE STACK



**VMware vSphere and NVIDIA RTX vWS only required for virtualized deployment*

NVIDIA OMNIVERSE STACK

OMNIVERSE

NUCLEUS



Core Services / On Prem / Cloud

CONNECT



Connection SDK / Plugins

KIT



Viewer / Editor / Framework

SIMULATION



Physics / AI / Animation / Behavior

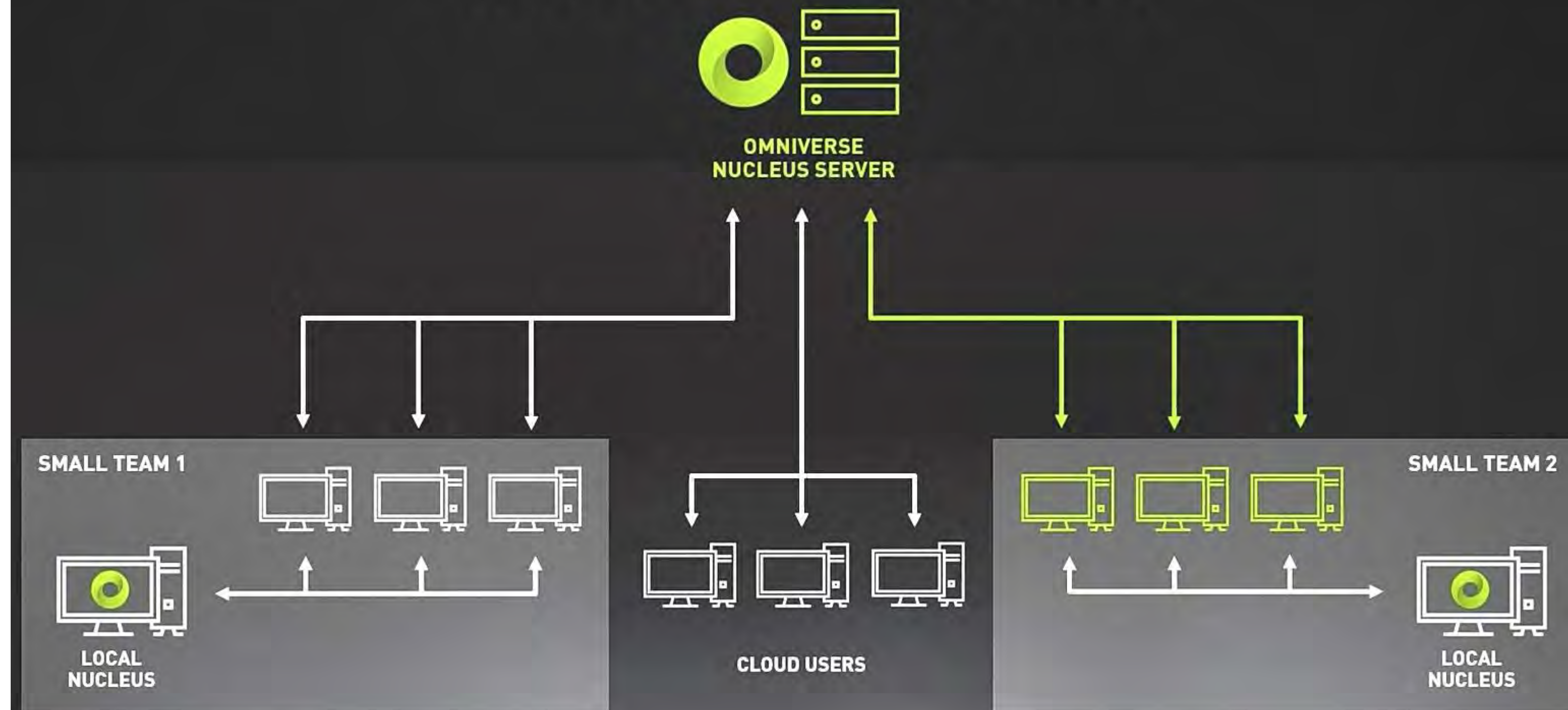
RTX RENDERER



Realtime / Scalable / Accurate / MDL

NVIDIA OMNIVERSE STACK

CONNECTED ASSETS



MDL & USD SCHEMA

```
def Material "flex_material" {  
  ...  
  def Shader "flex_material" {  
    uniform token info:implementationSource = "sourceAsset"  
    uniform asset info:mdl:sourceAsset = @nvidia/core_definitions.mdl@  
    uniform token info:mdl:sourceAsset:subIdentifier =  
      "::nvidia::core_definitions::flex_material"  
  }  
}
```



UNIVERSAL SCENE DESCRIPTION

THE "HTML" OF 3D VIRTUAL WORLDS

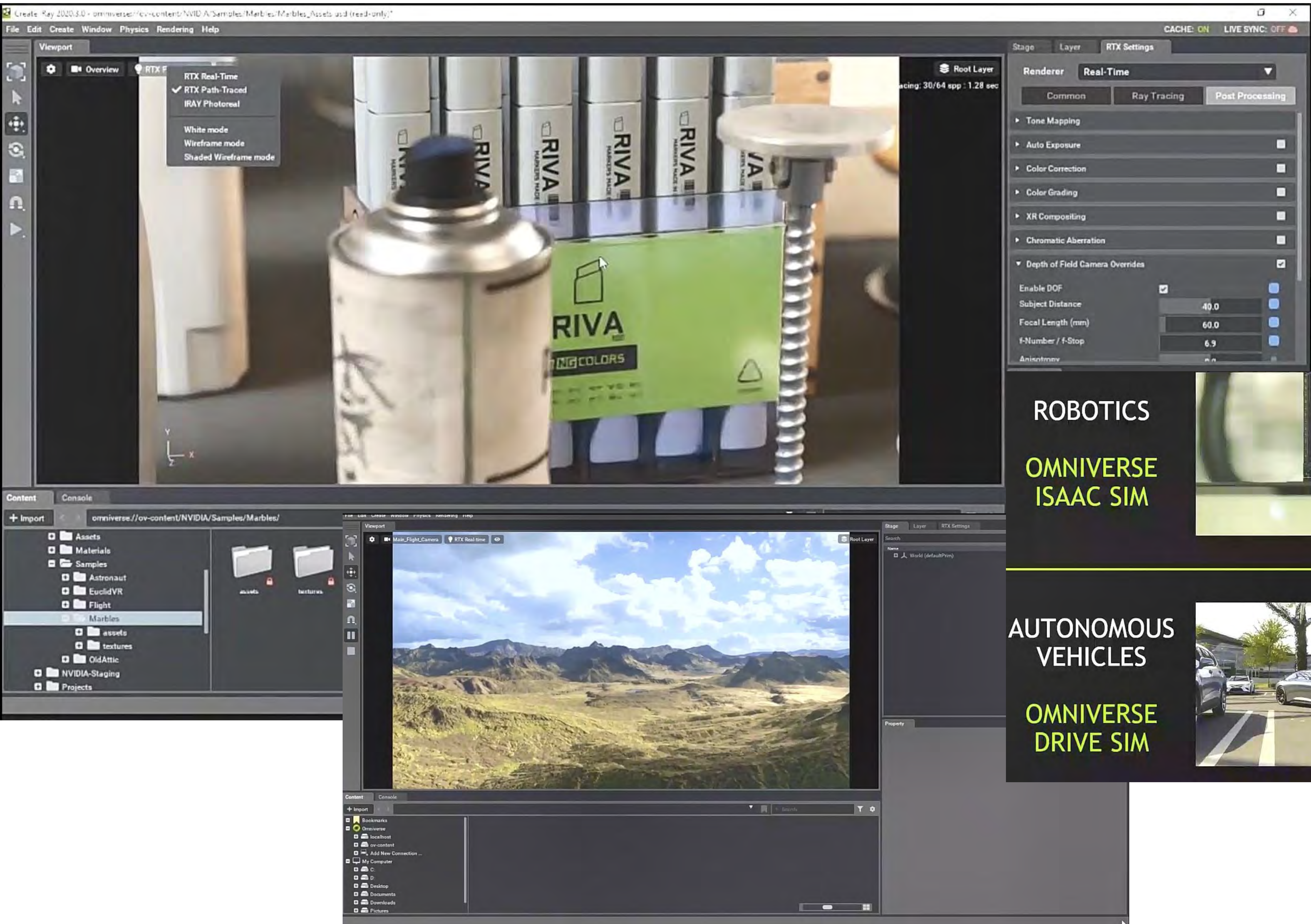
- ▶ Developed by Pixar
- ▶ Foundation for NVIDIA Omniverse
- ▶ Open-sourced API and file format for complex scene graphs
- ▶ Easily extensible, simplifies interchange of assets between industry software
- ▶ Introduces novel concept of layering
- ▶ Enables simultaneous collaboration for large teams in different department working on the same scene
- ▶ Originated in M&E, now becoming a standard across industries including AEC, Manufacturing, Product Design, Robotics



LIVE-LINK DIFFERENT APPS



NVIDIA OMNIVERSE STACK



ROBOTICS

OMNIVERSE
ISAAC SIM



AUTONOMOUS
VEHICLES

OMNIVERSE
DRIVE SIM



NVIDIA OMNIVERSE STACK

KIT STACK

Developer View

ASSEMBLIES

APPS

CREATE

VIEW

MACHINIMA

ISAAC SIM

AUDIO2FACE

...

MICROSERVICES

FORMAT CONVERTER

DEEP TAG

THUMBNAILS

BATCH RENDERING

...

EXTENSIONS

ASSET IMPORT/EXPORT

ANIMATION TOOLS

JARVIS

...

CORE EXTENSIONS

CONTENT BROWSER

MESH TOOLS

TEXTURE TOOLS

...

OMNIVERSE KIT

USD

PHYSX

OMNIUI

PYTHON

RTX

OMNIGRAPH

NUCLEUS CLIENT

...

NVIDIA OMNIVERSE STACK

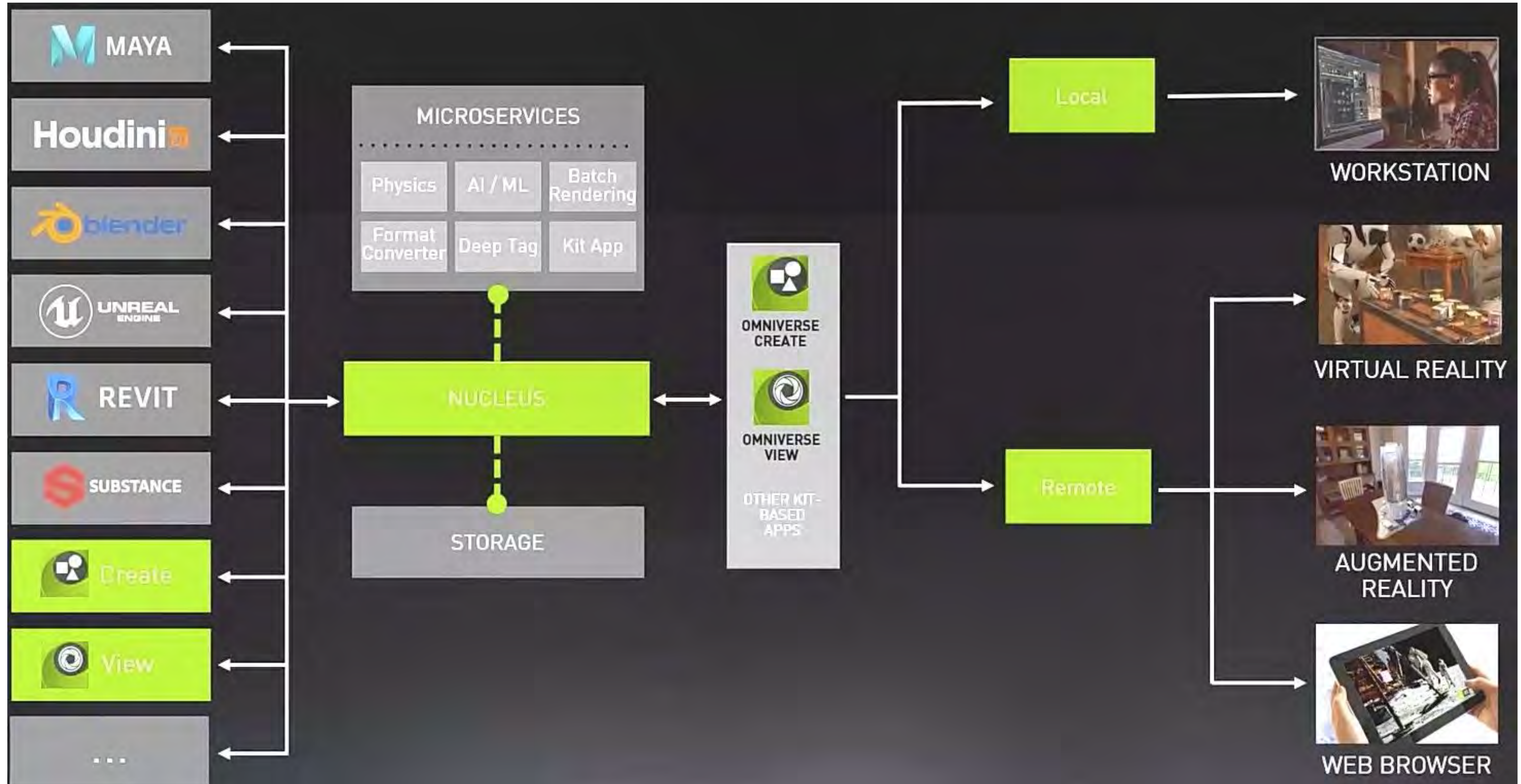
NUCLEUS-BASED MICROSERVICES

USE NUCLEUS FOR SERVICES TO COMMUNICATE

- Many Applications can communicate using Nucleus USD content or Nucleus Channels
- Kit Apps can contribute and represent Micro-services Execution Engine
- Other Application can be Microservices too, using Unreal Editor or Houdini to author or control Data



NVIDIA OMNIVERSE STACK



NEW RTX PRODUCT PORTFOLIO

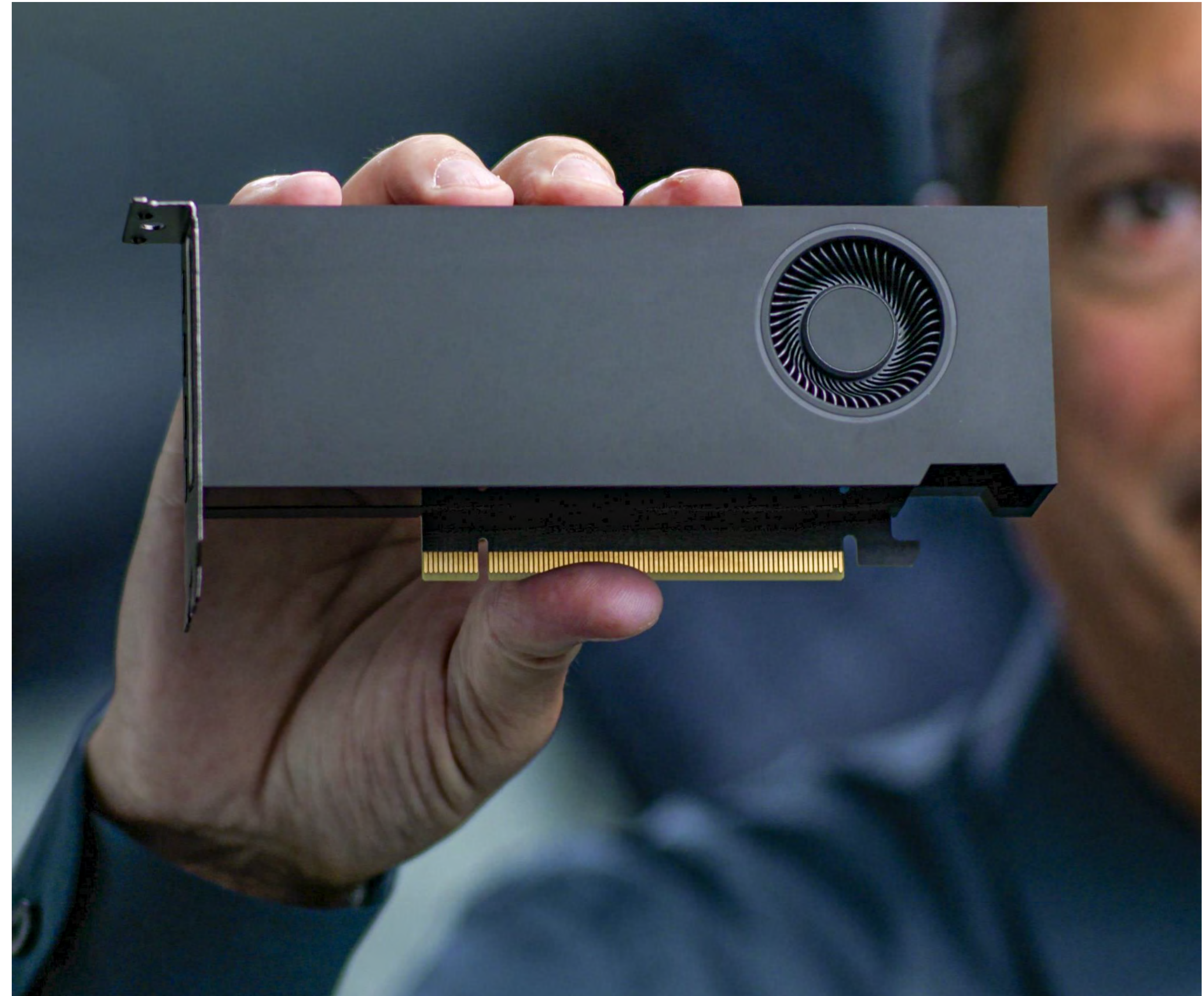


NVIDIA RTX A2000 12GB

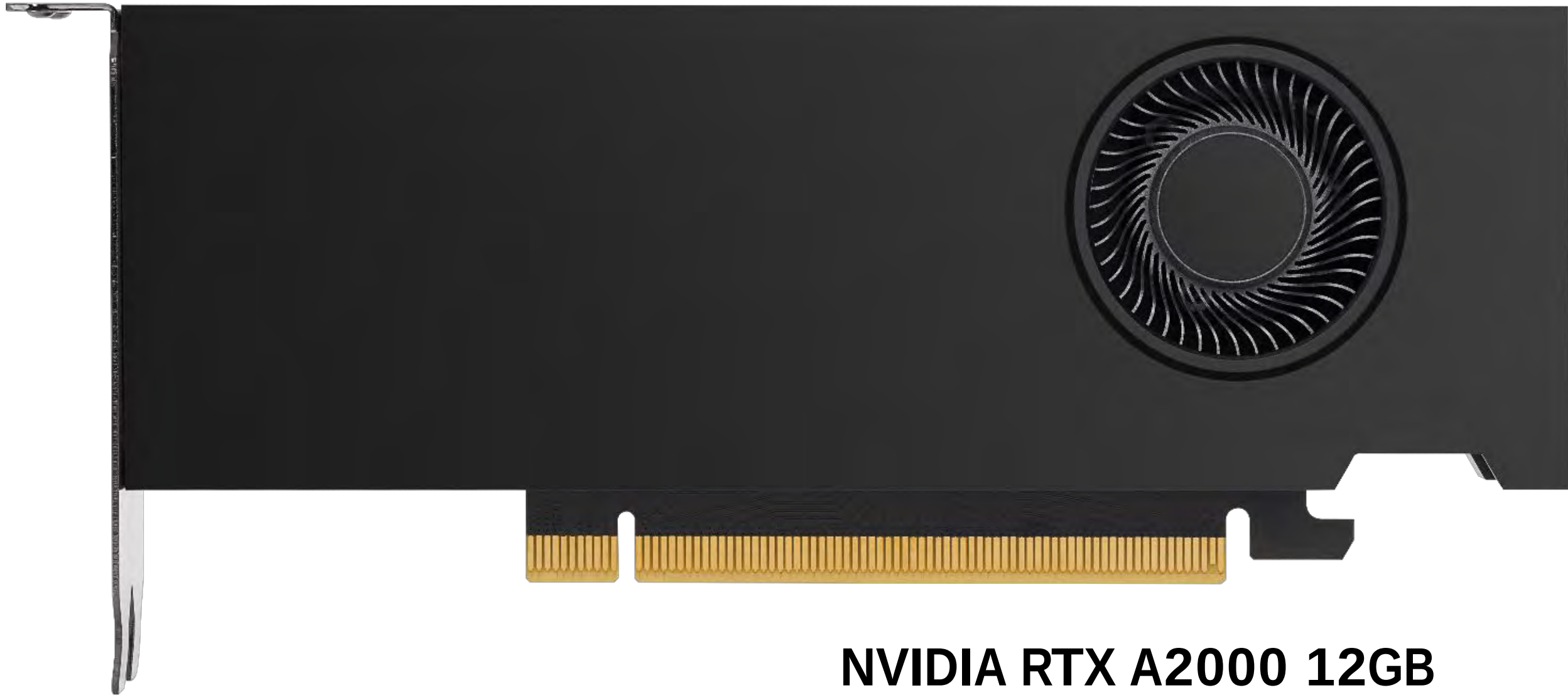
Expanding Access to the Power of NVIDIA RTX

Innovative Form Factor Incredible Performance and Features

- ▶ Low Profile, Dual Slot Design
- ▶ NVIDIA RTX Tensor, RT Cores
- ▶ 12 GB GDDR6 w/ECC Memory
- ▶ VR Ready



NVIDIA RTX A2000 12GB SPECIFICATIONS



- Brings 12GB of GPU memory to 2000 series GPUs
- Provides performance benefits to memory intensive applications & workflows
- Ready for today’s multi-application, higher-resolution workflows
- Future proofs your investment for the larger data sets and higher resolution work of tomorrow

	NVIDIA RTX A2000 12GB	NVIDIA RTX A2000	Quadro P2200
GPU Architecture	Ampere	Ampere	Pascal
CUDA Cores	3328	3328	1280
Tensor Cores	104 Ampere Arch Cores	104 Ampere Arch Cores	-
RT Cores	26 Ampere Arch Cores	26	-
Peak Single-Precision Performance	8 TFLOPS	7.99 TFLOPS	3.8 TFLOPS
Memory Size	12 GB GDDR6 w/ECC	6 GB GDDR6 w/ECC	5 GB GDDR5X
Memory Bandwidth	288 GB/s	288 GB/s	200 GB/s
VR Ready	Yes	Yes	No
Display Connectors	4x mDP 1.4	4x mDP 1.4	4x DP 1.4
Form Factor	2.7”H x 6.6”L Dual Slot		4.4” H x 7.9”L Single Slot
Max Power Consumption	70W	70W	75W
Graphics Bus	PCI Express Gen 4 x 16	PCI Express Gen 4 x 16	PCI Express Gen 3 x 16

NVIDIA RTX A4500 SPECIFICATIONS



NVIDIA RTX A4500

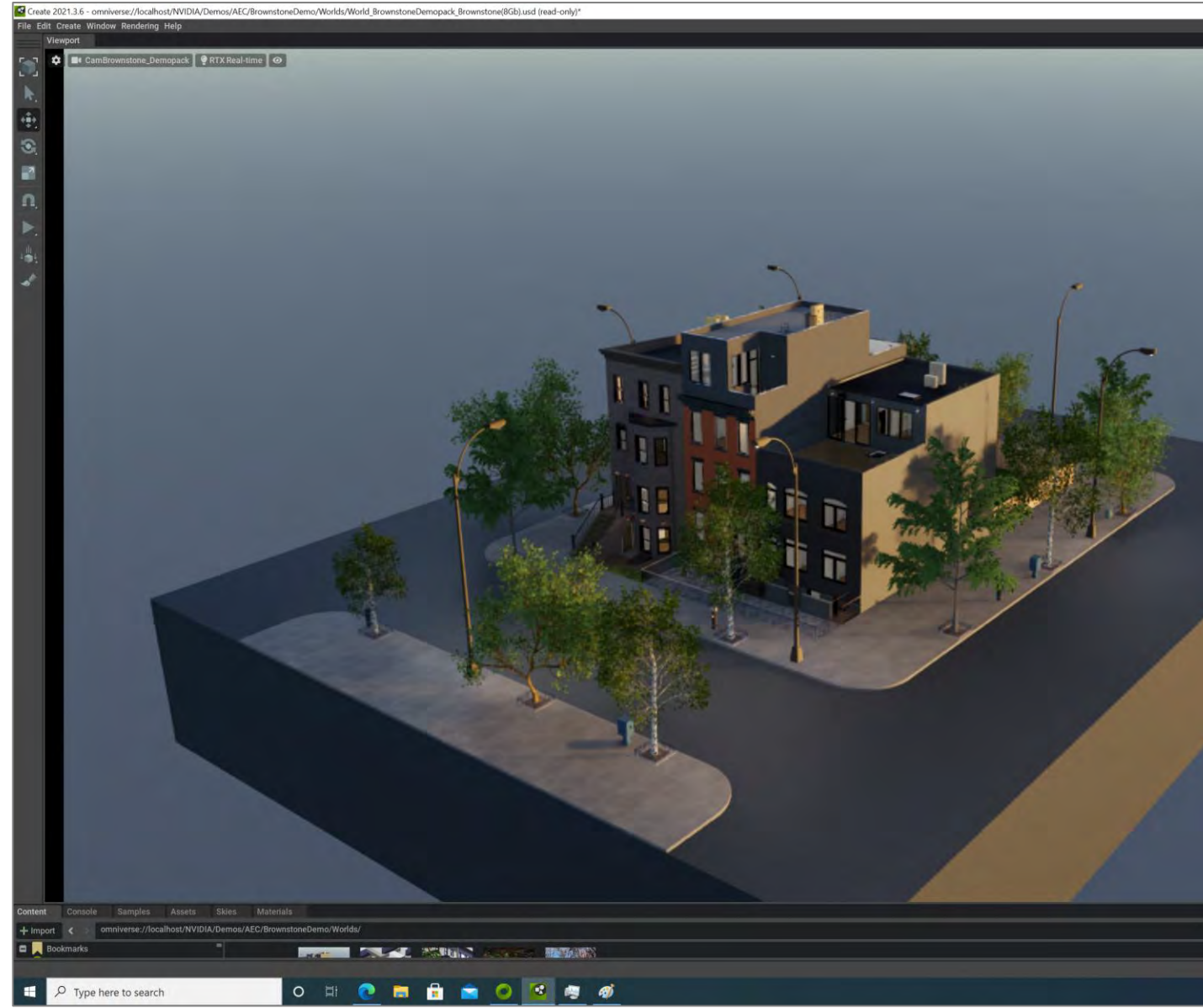
- Brings 20GB of GPU memory to 4000 series GPUs
- Additional CUDA, RT, and Tensor cores provide performance benefits to professional applications & workflows
- NVLink support, another first for 4000 series GPUs, future proofs your investment, letting you expand GPU memory for even larger models and data sets of tomorrow

	RTX A4500	RTX A4000	RTX 4000
GPU Architecture	Ampere	Ampere	Turing
CUDA Cores	7168	6144	2304
Tensor Cores	224 Ampere Arch Cores	192 Ampere Arch Cores	288 Turing Arch Cores
RT Cores	56	48	36
Peak Single-Precision Performance	23.7 TFLOPS	19.2 TFLOPS	7.1 TFLOPS
Memory Size	20 GB GDDR6 w/ECC	16 GB GDDR6 w/ECC	8 GB GDDR6
Memory Bandwidth	640 GB/s	512 GB/s	416 GB/s
Display Connectors	4x DP 1.4	4x DP 1.4	3x DP 1.4 + 1x USB-C
Max Power Consumption	200W	140W	125W*
Power Connector	8-Pin PCIe	6-Pin PCIe	8-Pin PCIe
Graphics Bus	PCI Express Gen 4 x 16	PCI Express Gen 4 x 16	PCI Express Gen 3 x 16
NVLink Support	Yes, 112.5 GB/s (bidirectional)	No	No
Form Factor	4.4” H x 10.5” L Dual Slot	4.4”H x 9” L Single Slot	4.4”H x 9.5”L Single Slot

RTX A4500 - DO MORE

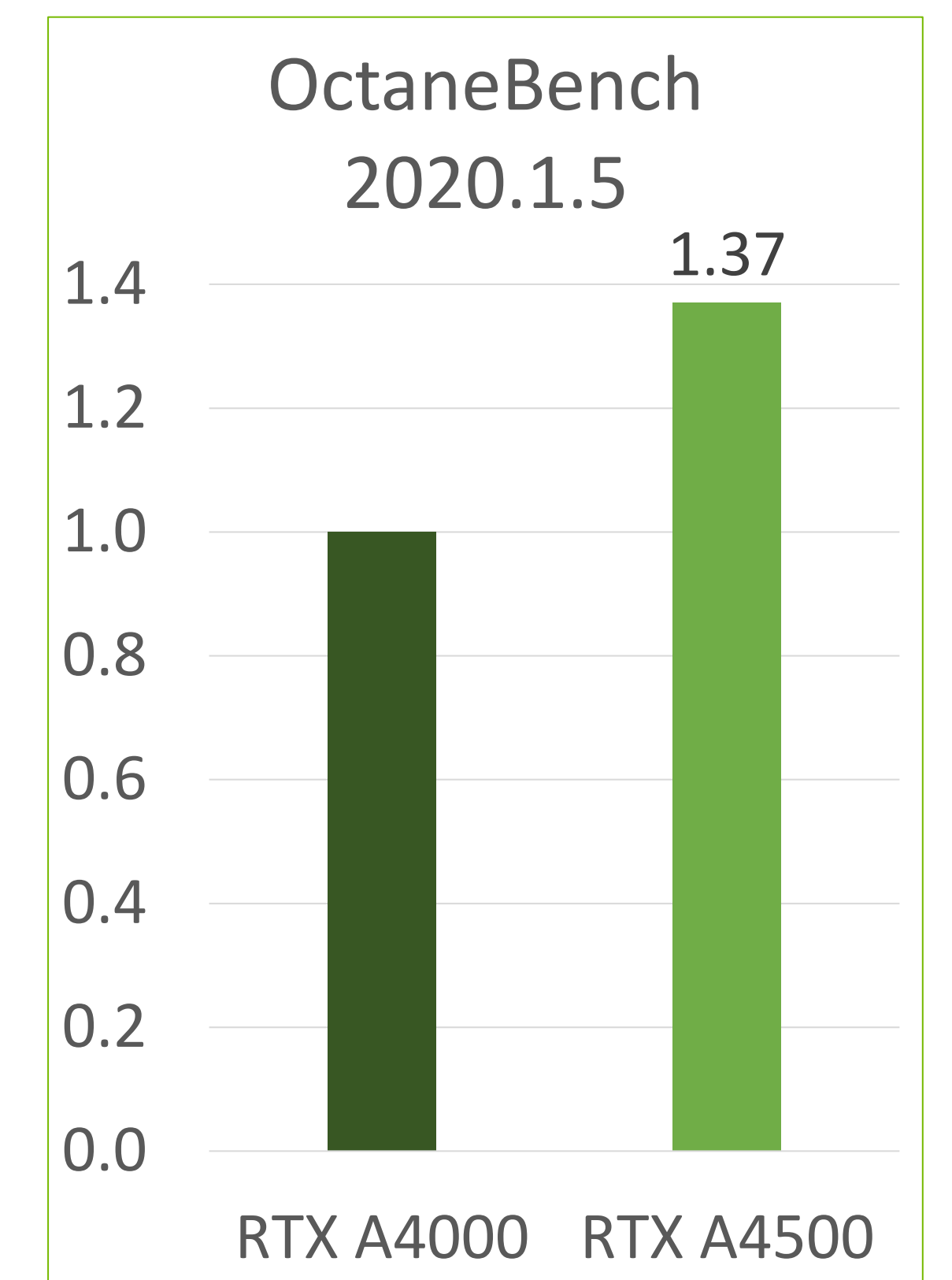
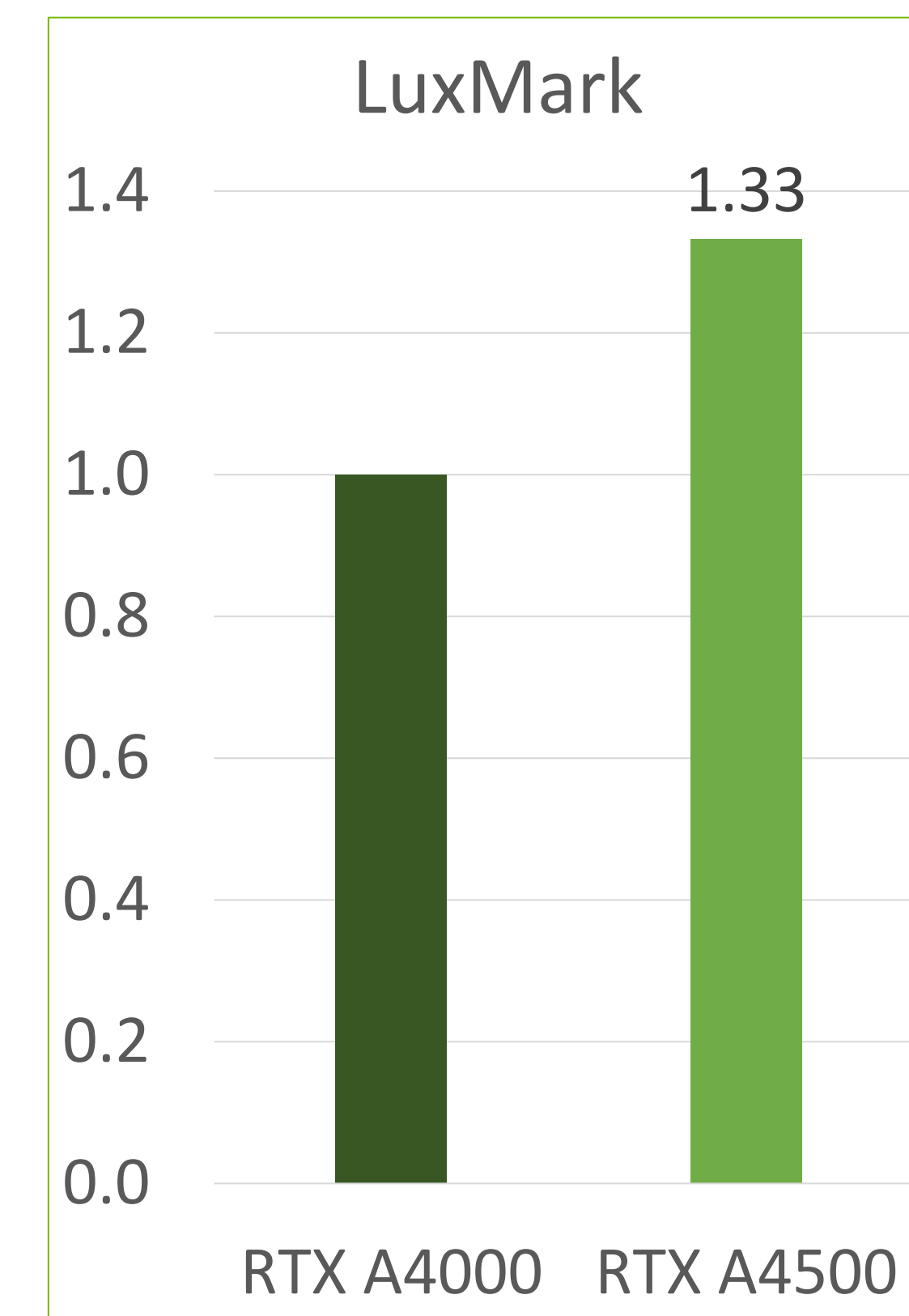
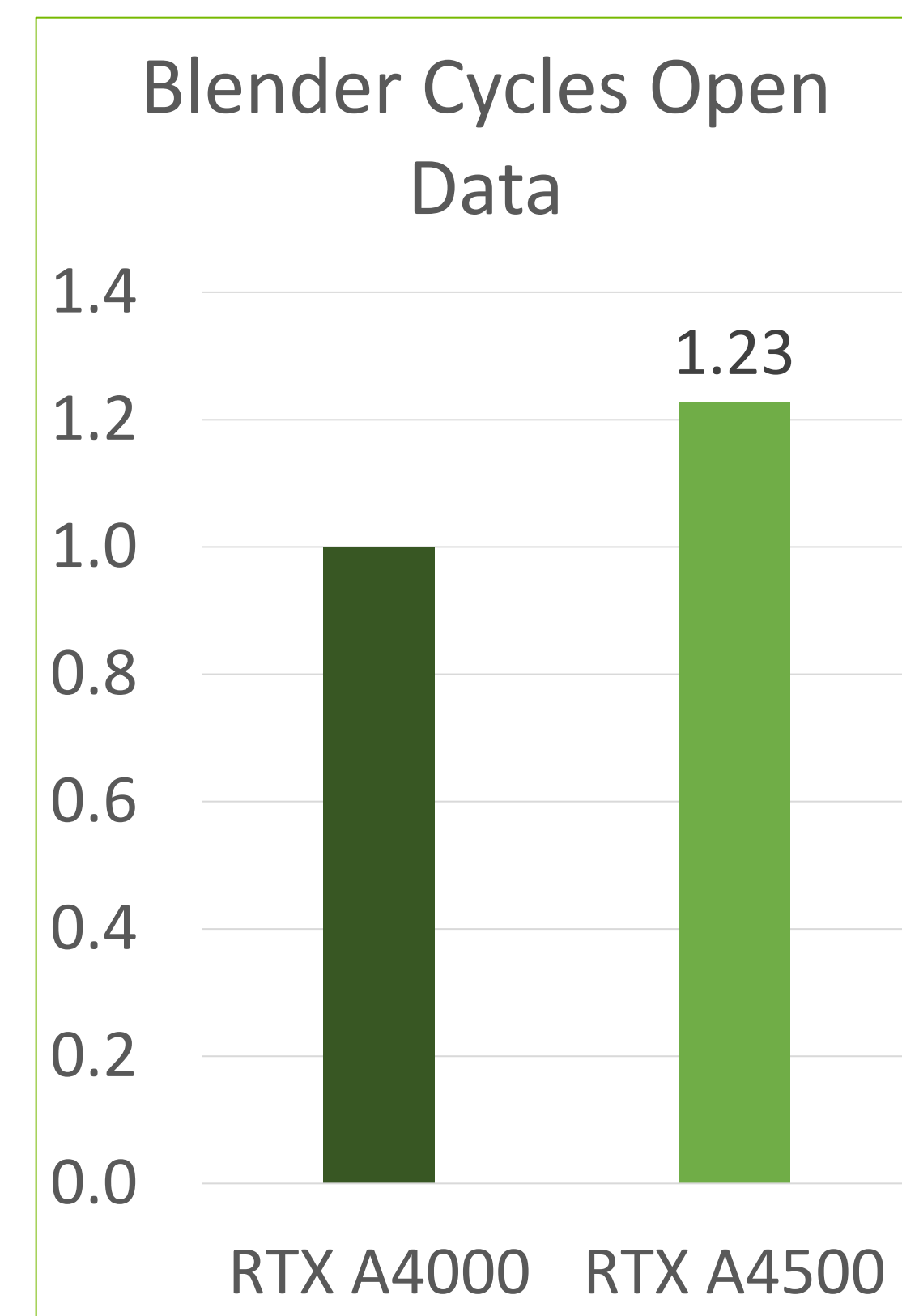
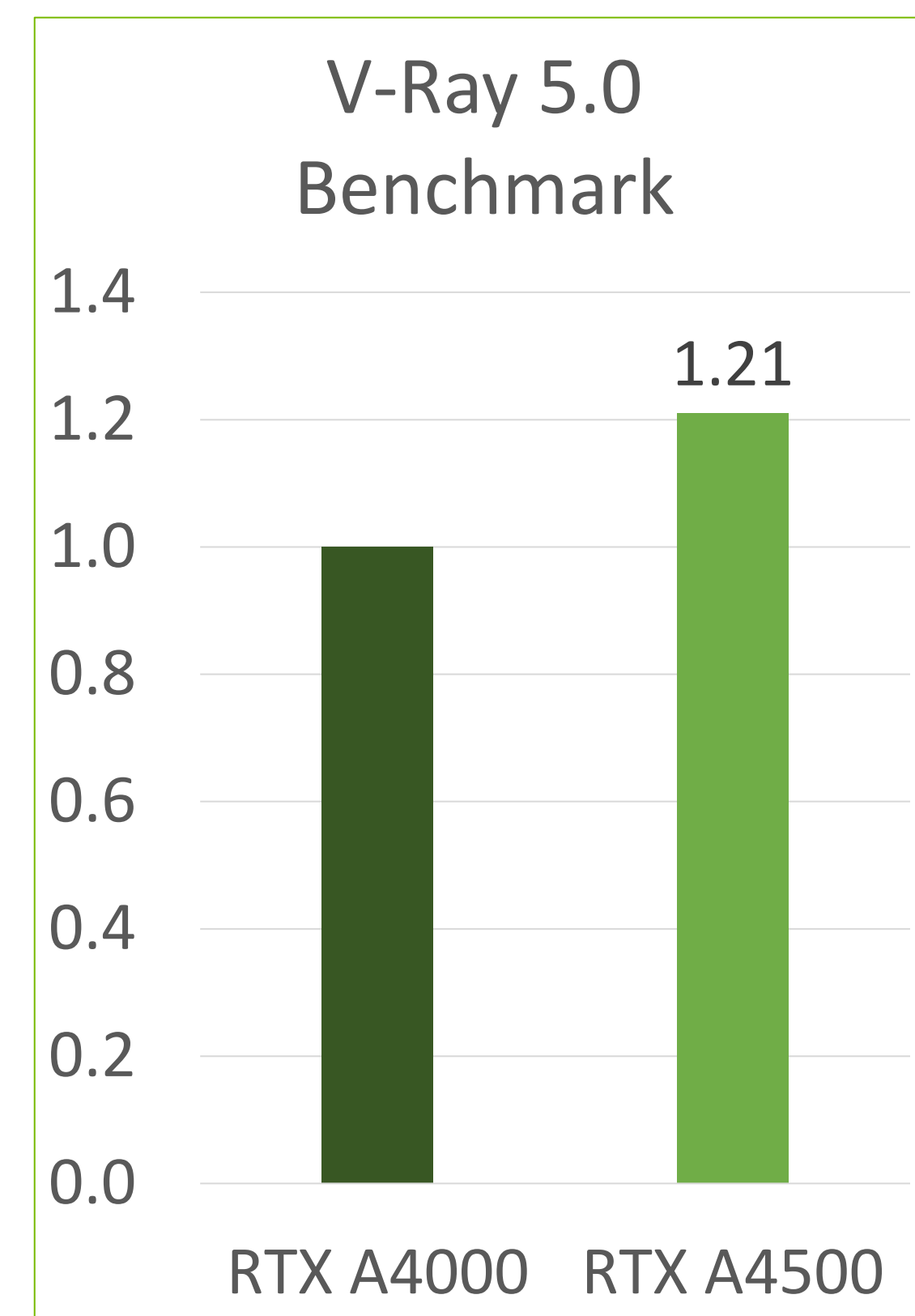
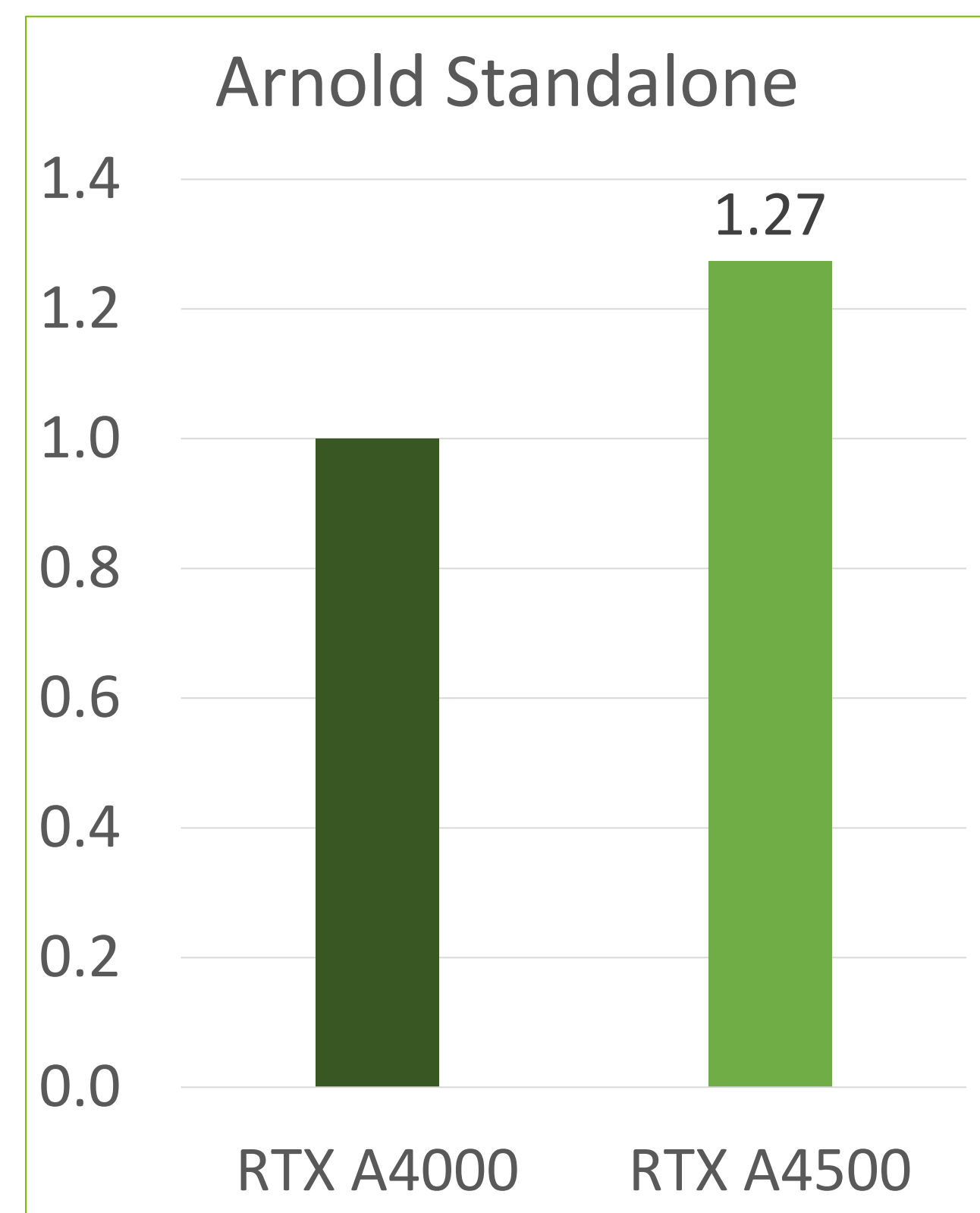
20GB of GPU Memory

8GB GPU Memory



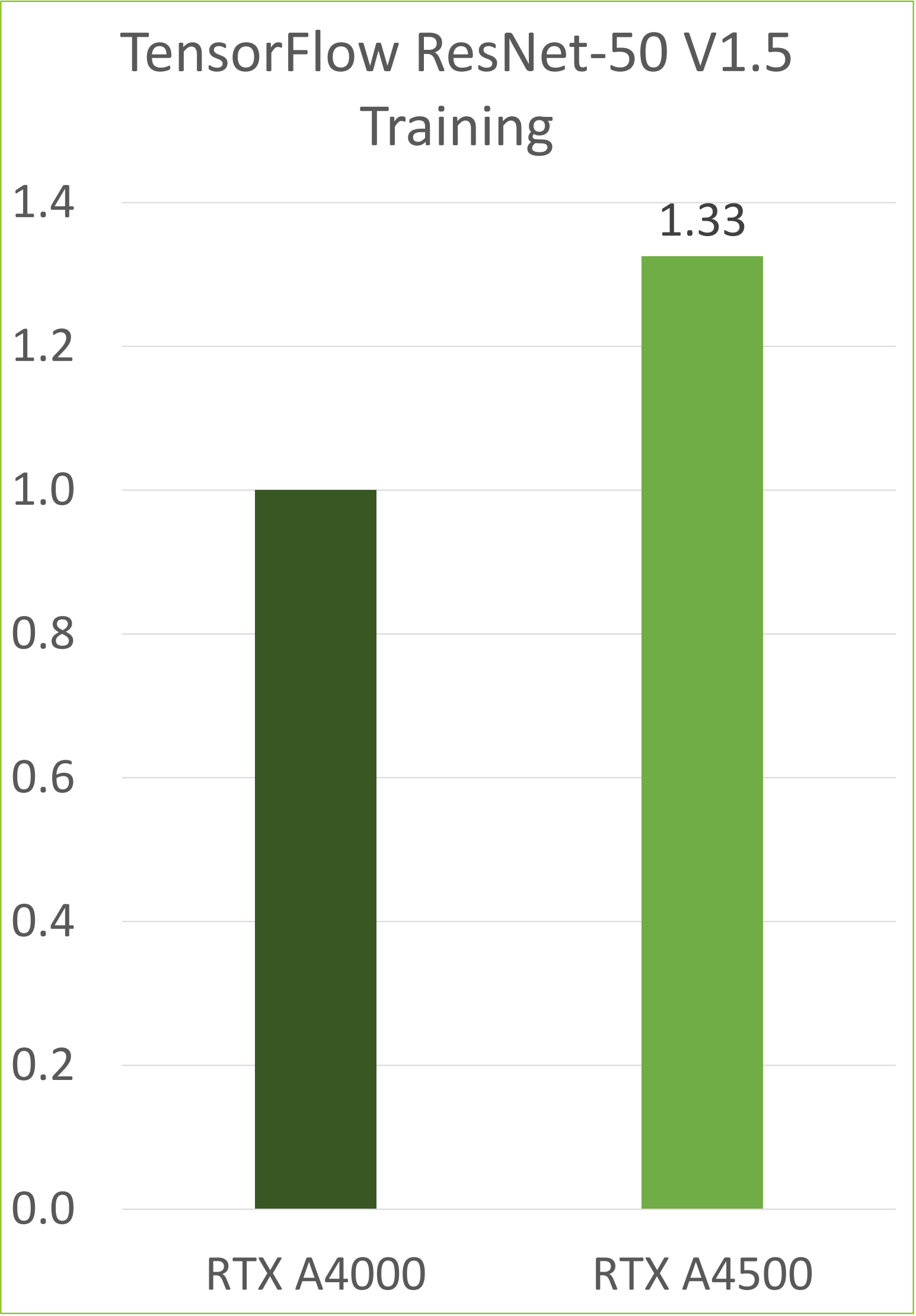
A4500 RENDERING PERFORMANCE

Faster Rendering Performance than Previous Generation 4000 Series GPUs

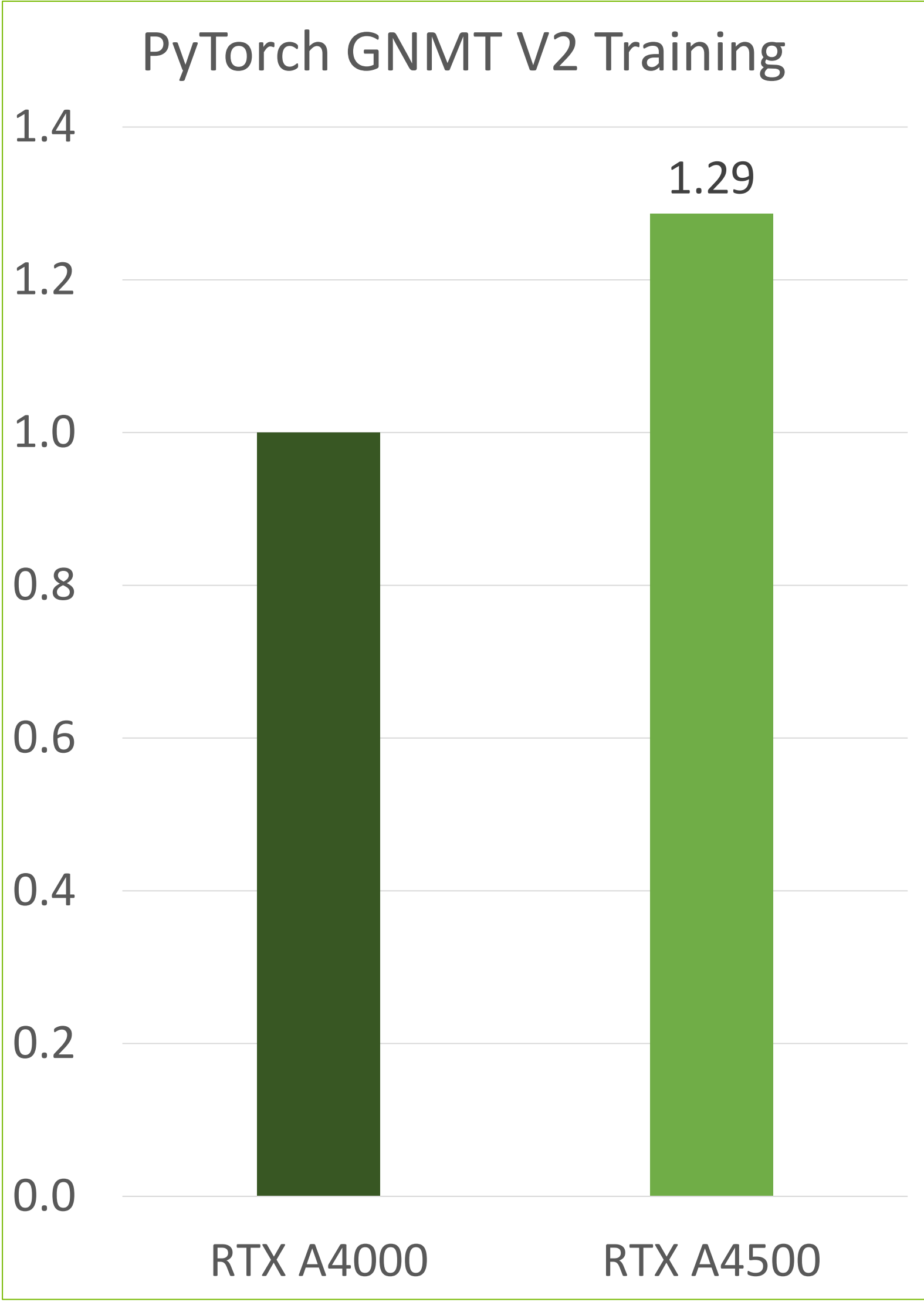


A4500 PERFORMANCE (PRELIMINARY)

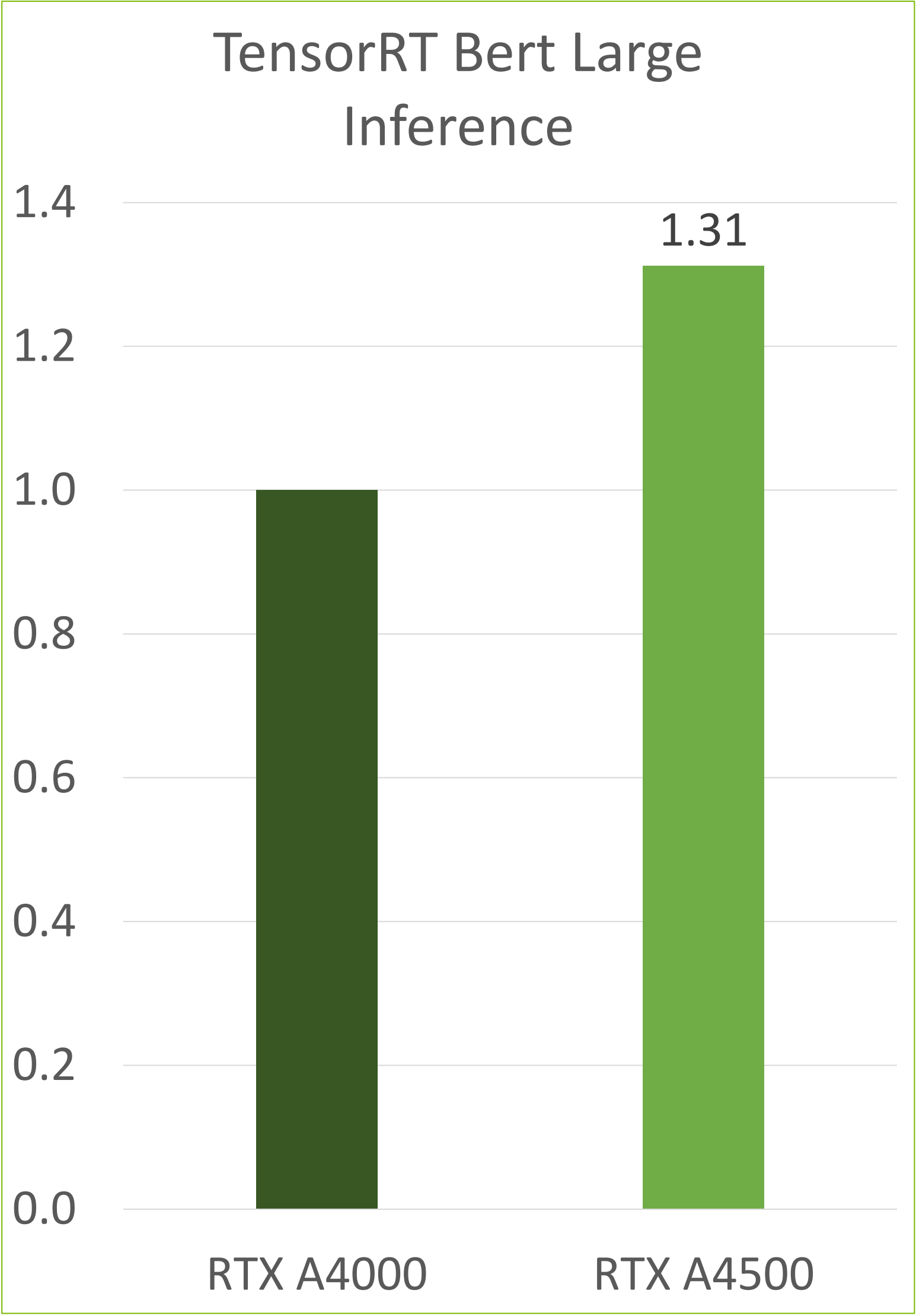
Faster AI 4000 Series GPUs



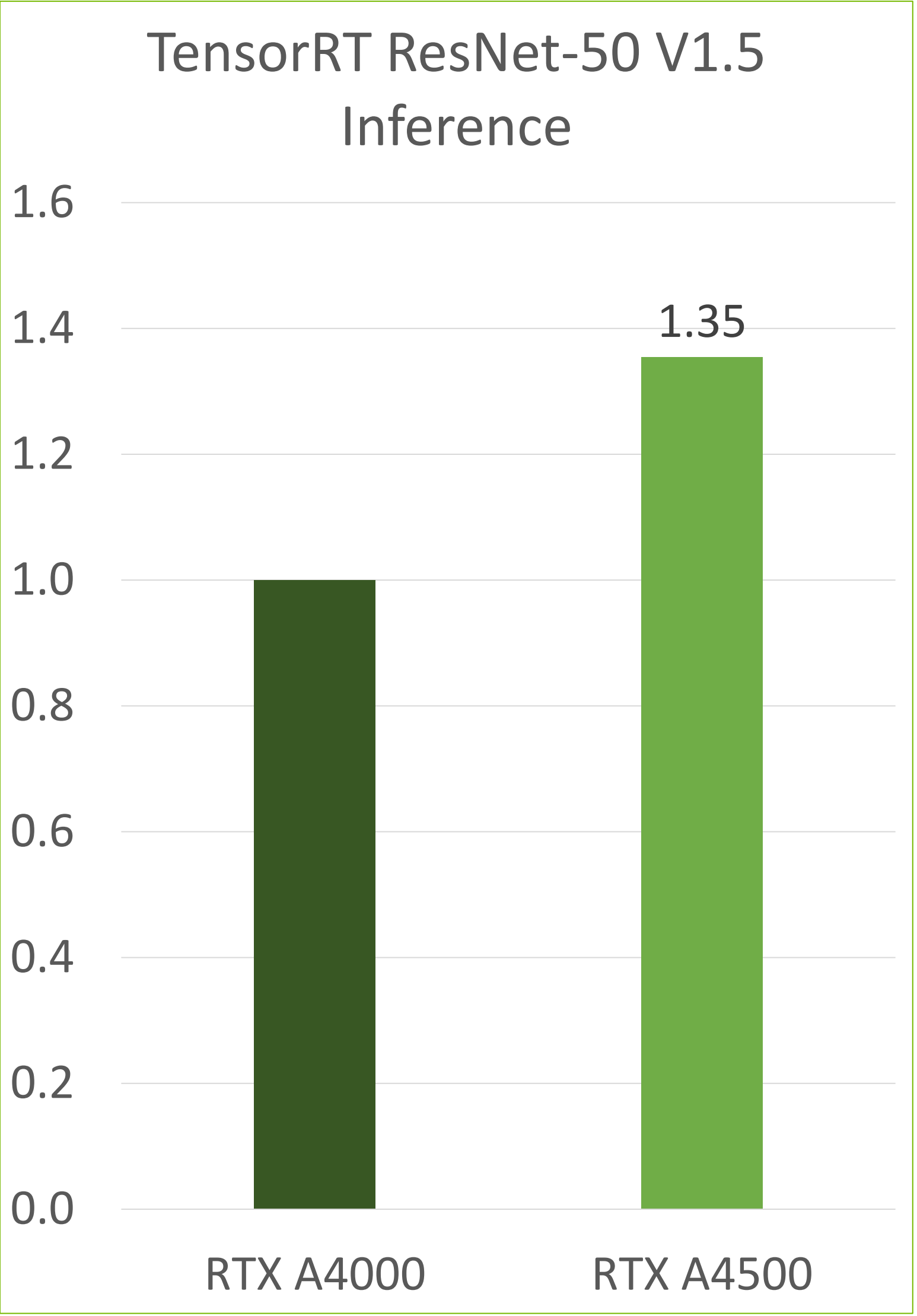
Batch Size=256*# of GPUs; Precision=Mixed; AMP=Yes; Data=Real; XLA=Yes; cuDNN Version=8.2.4.15; NCCL Version=2.11.4; Baseline=DL 21.10; Installation Source=NGC;



Batch Size=128*# of GPUs; Precision=Mixed; Data=Real; cuDNN Version=8.2.4.15; NCCL Version=2.11.4; Baseline=DL 21.10; Installation Source=NGC;



Batch Size=128*# of GPUs; Precision=Mixed; Data=Synthetic; Sequence Length=128; cuDNN Version=8.2.4.15; Baseline=DL 21.10; Installation Source=NGC;



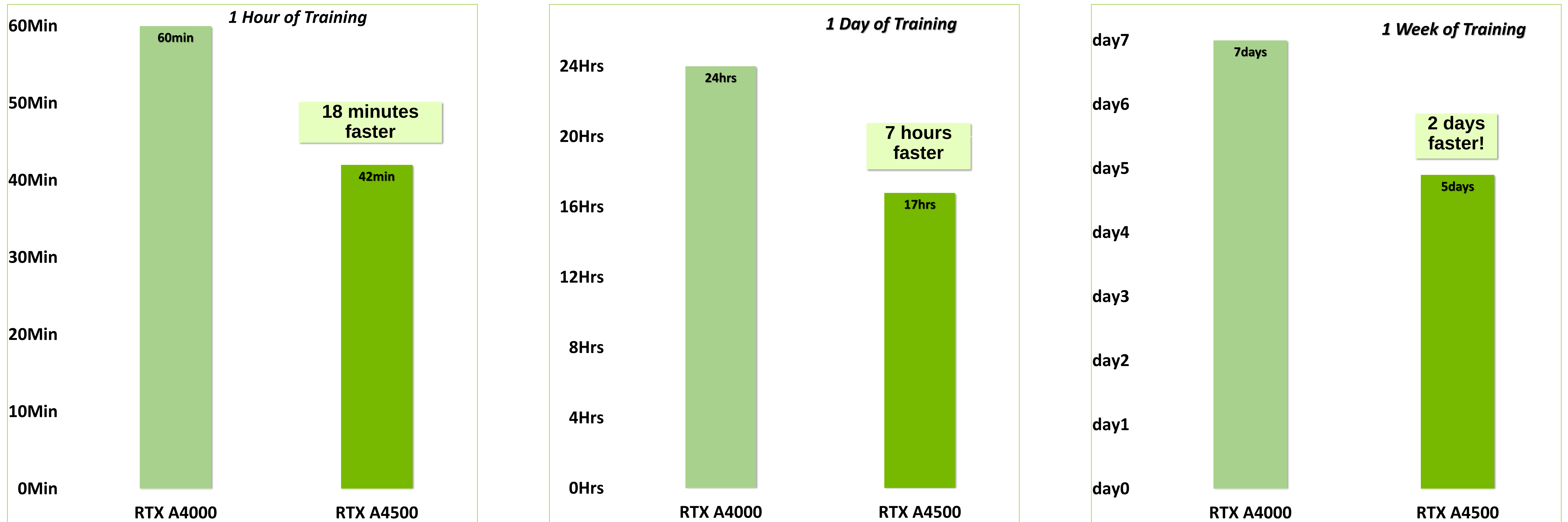
Batch Size=128*# of GPUs; Precision=Mixed; Data=Synthetic; cuDNN Version=8.2.4.15; Baseline=DL 21.10; Installation Source=NGC;

Tests un on an Intel Xeon Gold 6154 3GHz, 3.7GHz Turbo, 64GB RAM, Ubuntu 20.04 LTS x64, NVIDIA Driver 470.62

WORKFLOW EXAMPLE: UNDERSTANDING TRUE VALUE

Even a Small Performance Advantage can Result in Huge Productivity Gains

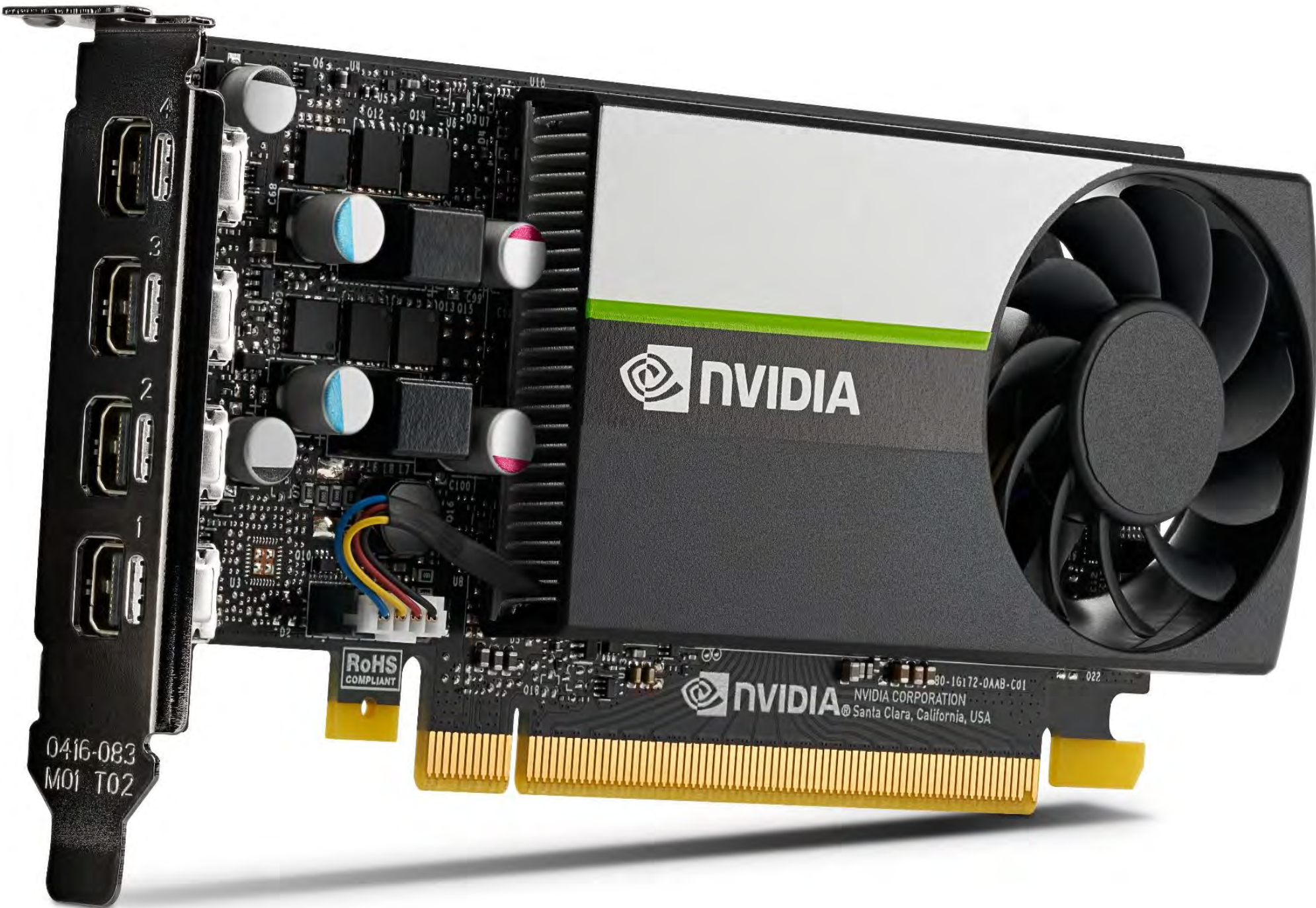
Example of a 30% performance advantage for AI training



(lower is better)

AI training workflows can take days, weeks to train & validate models - just comparing price/performance would result in wrong decision

NVIDIA T1000 SPECIFICATIONS

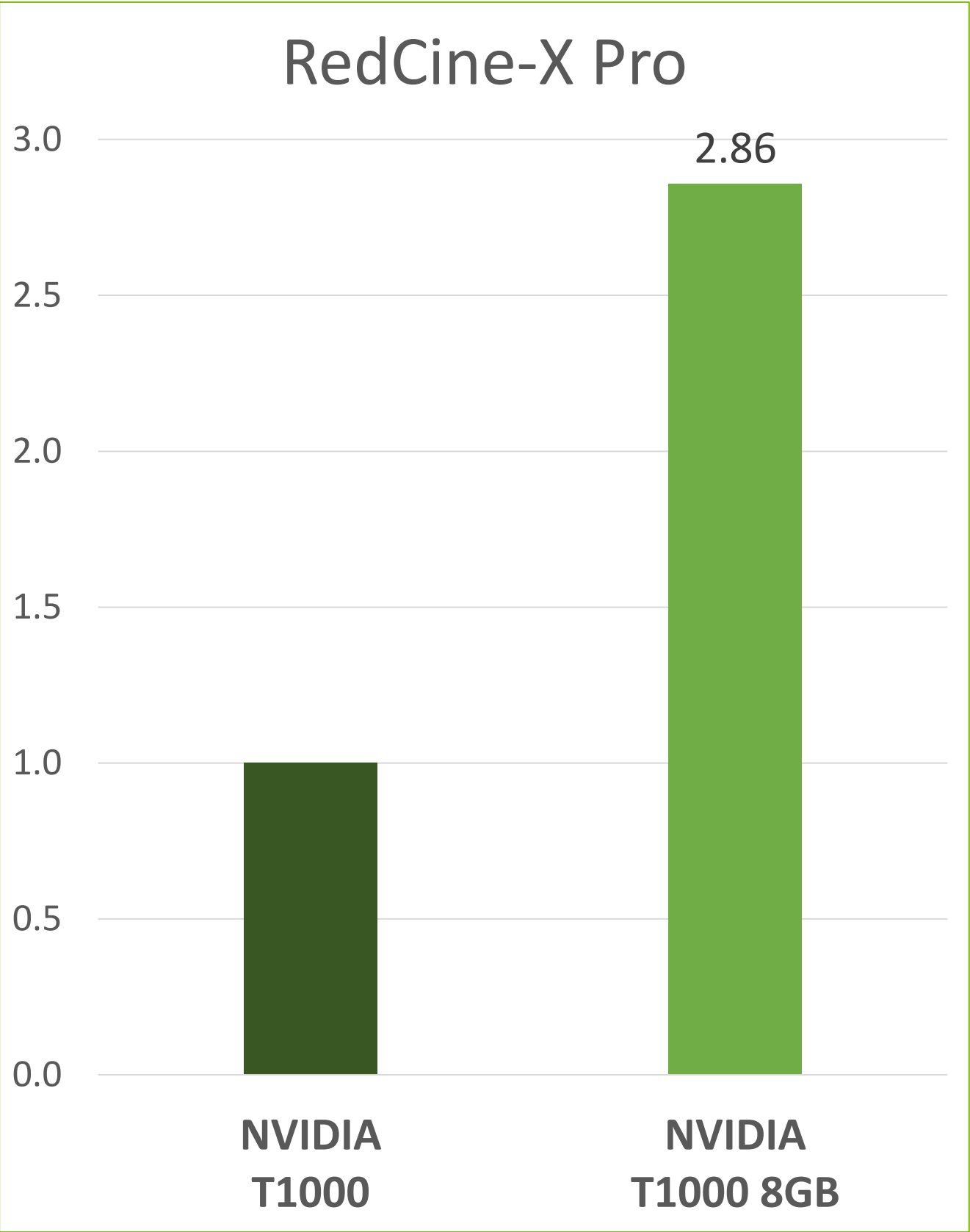
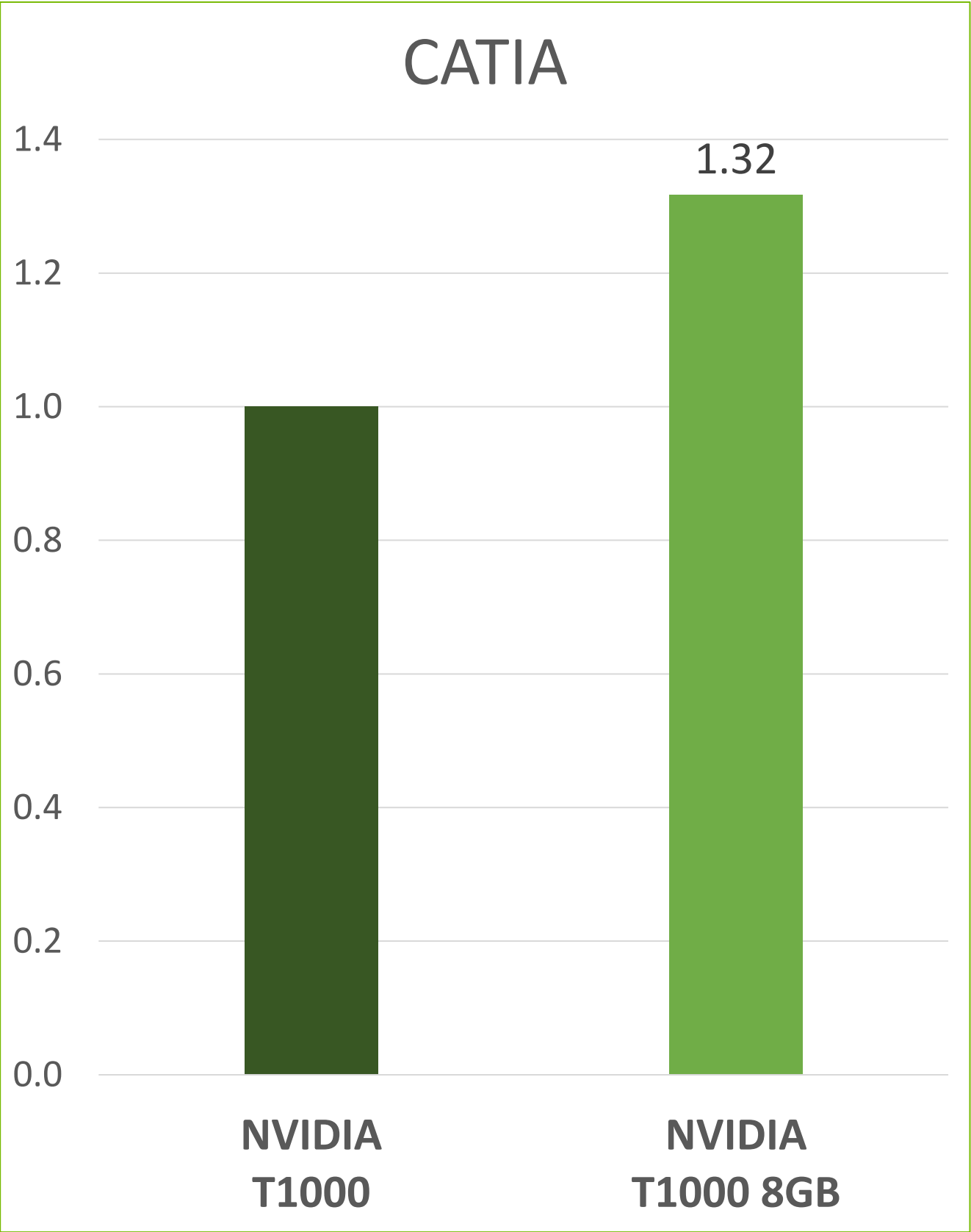
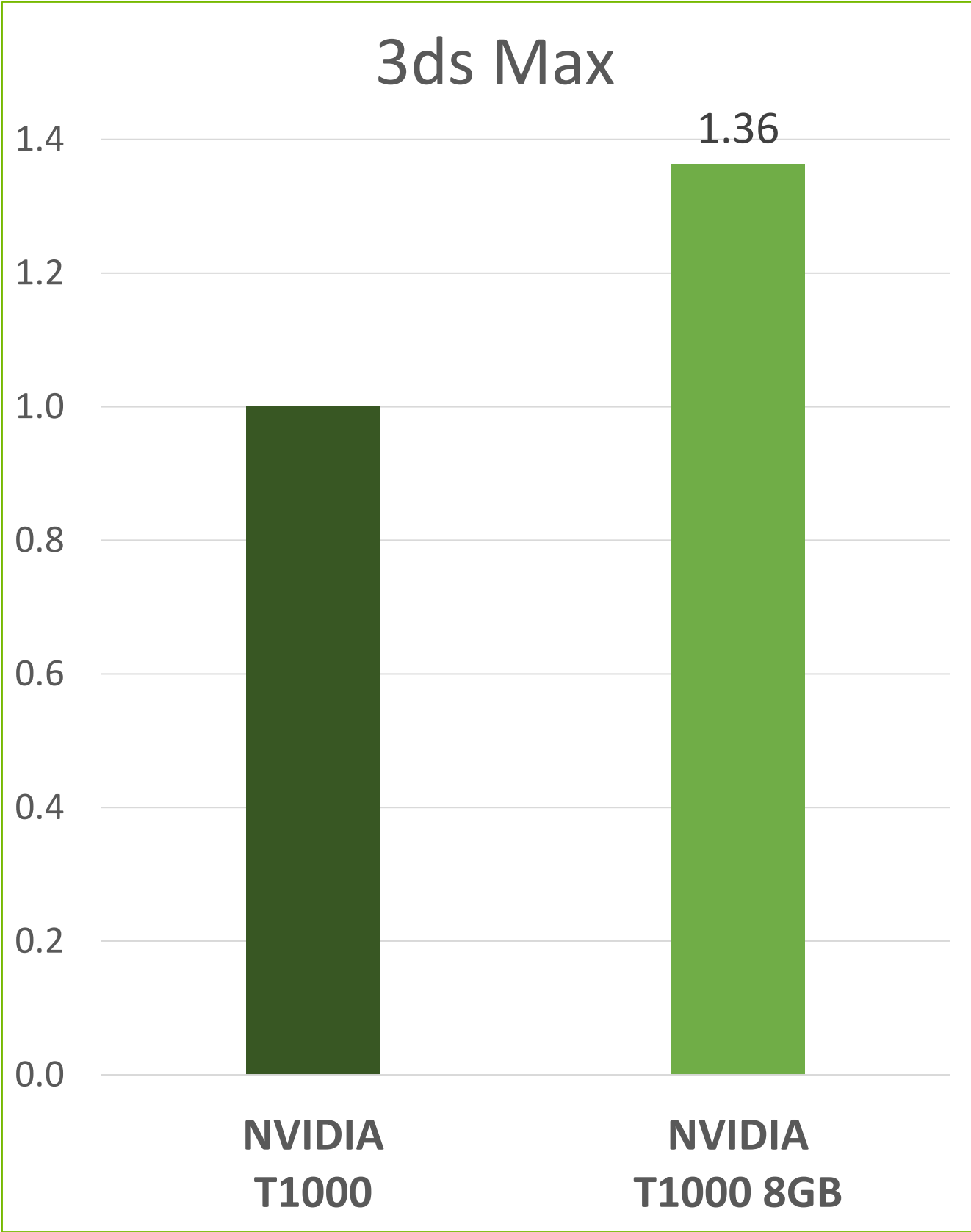


NVIDIA T1000 8GB

- Brings 8GB of GPU memory to 1000 series GPUs
- Provides performance benefits to memory intensive applications & workflows
- Ready for today’s multi-application, higher-resolution workflows
- Future proofs your investment for the larger data sets and higher resolution work of tomorrow

	NVIDIA T1000 8GB	NVIDIA T1000	NVIDIA Quadro P1000
GPU Architecture	Turing	Turing	Pascal
CUDA Cores	896	896	640
Peak Single-Precision Performance	Up to 2.5 TFLOPS	Up to 2.5 TFLOPS	Up to 1.89 TFLOPS
Memory Size	8GB GDDR6	4GB GDDR6	4GB GDDR5
Memory Interface	128-bit	128-bit	128-bit
Memory Bandwidth	Up to 160 GB/s	Up to 160 GB/s	Up to 80 GB/s
Display Connectors	4x mDP 1.4	4x mDP 1.4	4x mDP 1.4
Max Display Resolution	8K @ 60Hz	8K @ 60Hz	5K @ 60Hz
Max Power Consumption	50 W	50 W	47 W
Graphics Bus	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16
Form Factor	2.713 inches H x 6.137 inches L Single Slot	2.713 inches H x 6.137 inches L Single Slot	2.713 inches H x 5.7 inches L Single Slot

APPLICATION PERFORMANCE DATA



Tests run on an Intel Xeon Gold 6154 @ 3GHz. 3.7GHz Turbo, 64GB RAM, Windows 10 x64, NVIDIA driver 472.06. Autodesk 3dsMax Dassault Systemes CATIA results from SPECviewperf 2020 3dsMax and CATIA subtest composite scores; RedCine-X Pro performance uses internal testing measuring performance with 6K and 8K video. DaVinci Resolve results from internal testing measuring average FPS of video rendering using various effects at resolutions from 1080p to 8K. NVIDIA T1000 failed to execute tests due to lack of GPU memory.

NVIDIA T400 4GB SPECIFICATIONS



NVIDIA T400 4GB

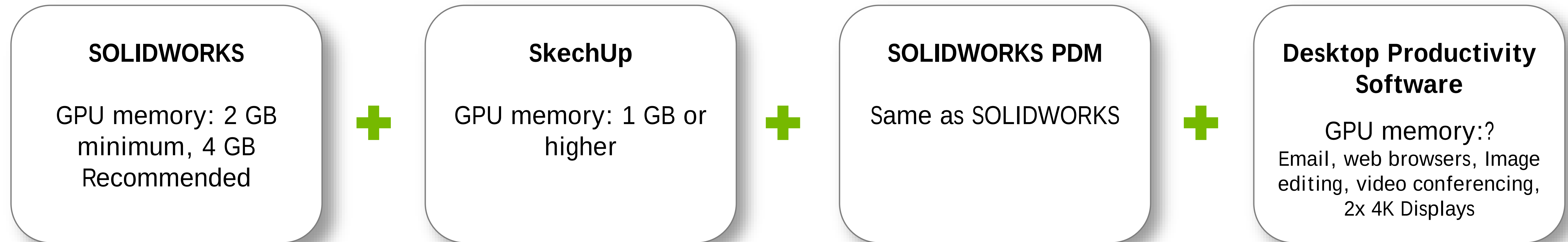
- Brings 4GB of GPU memory to 400 series GPUs
- Same compact, power efficient form factor
- Ready for multi-application, higher-resolution workflows
- Future-proofs your investment for the larger data sets and higher resolution workloads of tomorrow

	NVIDIA T400 4GB	NVIDIA T400	NVIDIA Quadro P400
GPU Architecture	Turing	Turing	Pascal
CUDA Cores	384	384	256
Peak Single-Precision Performance	Up to 1.09 TFLOPS	Up to 1.09 TFLOPS	Up to .64 TFLOPS
Memory Size	4GB GDDR6	2GB GDDR6	2GB GDDR5
Memory Interface	64-bit	64-bit	64-bit
Memory Bandwidth	Up to 80 GB/s	Up to 80 GB/s	Up to 32 GB/s
Display Connectors	3x mDP 1.4	3x mDP 1.4	3x mDP 1.4
Max Display Resolution	5K @ 60Hz	5K @ 60Hz	5K @ 60Hz
Max Power Consumption	30 W	30 W	30 W
Graphics Bus	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16
Form Factor	2.713 inches H x 6.137 inches L Single Slot	2.713 inches H x 6.137 inches L Single Slot	2.713 inches H x 5.7 inches L Single Slot

MULTI-APPLICATION WORKLOADS KEY TO MAXIMIZE PRODUCTIVITY

Need to Consider the GPU Memory Requirements of all the Apps in the Workflow

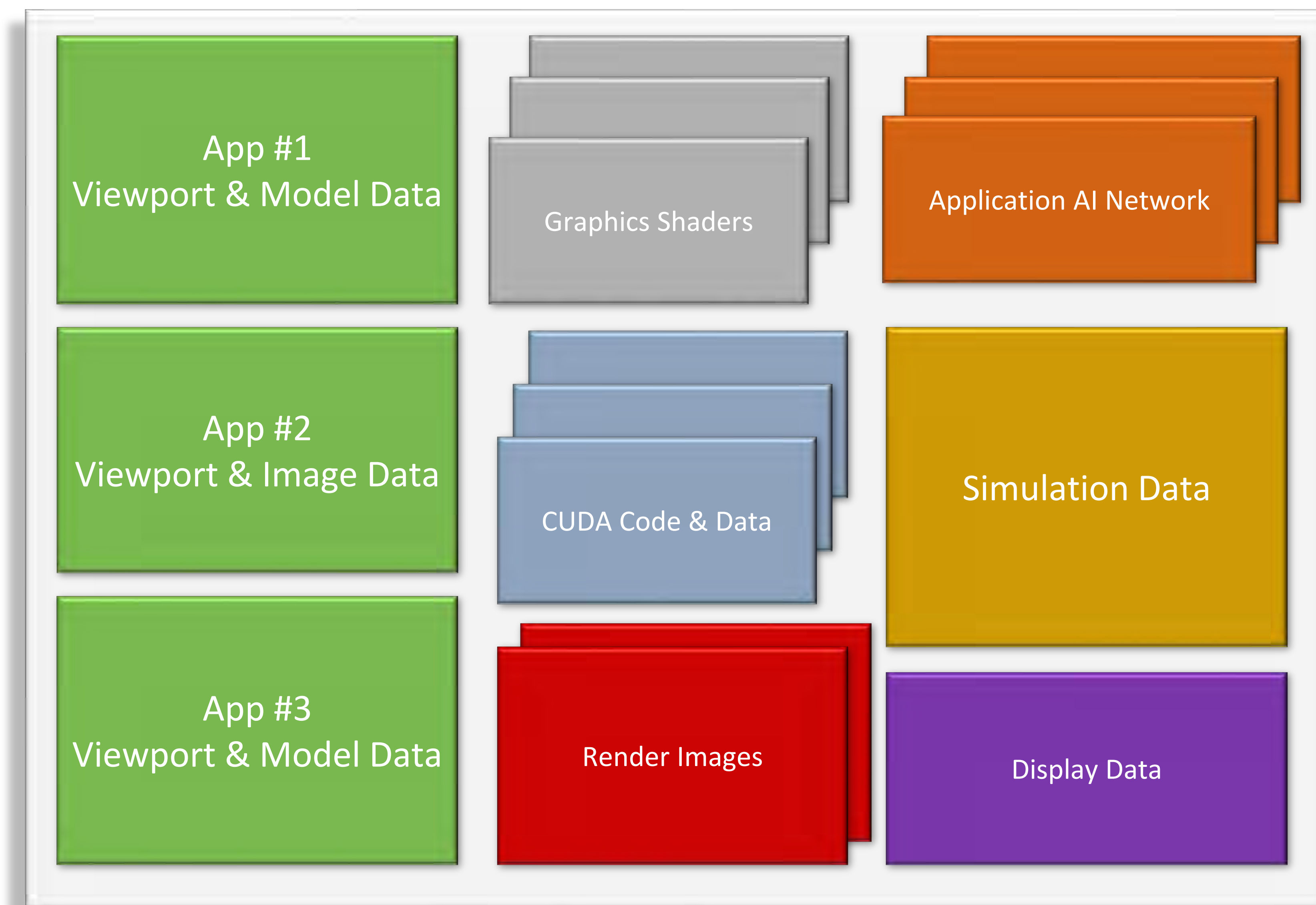
Sample CAD Design Workflow



Will the GPU purchased today handle the workloads of the next 6, 12, 24, 36 months?

GPU MEMORY - WORKFLOW CONSUMPTION

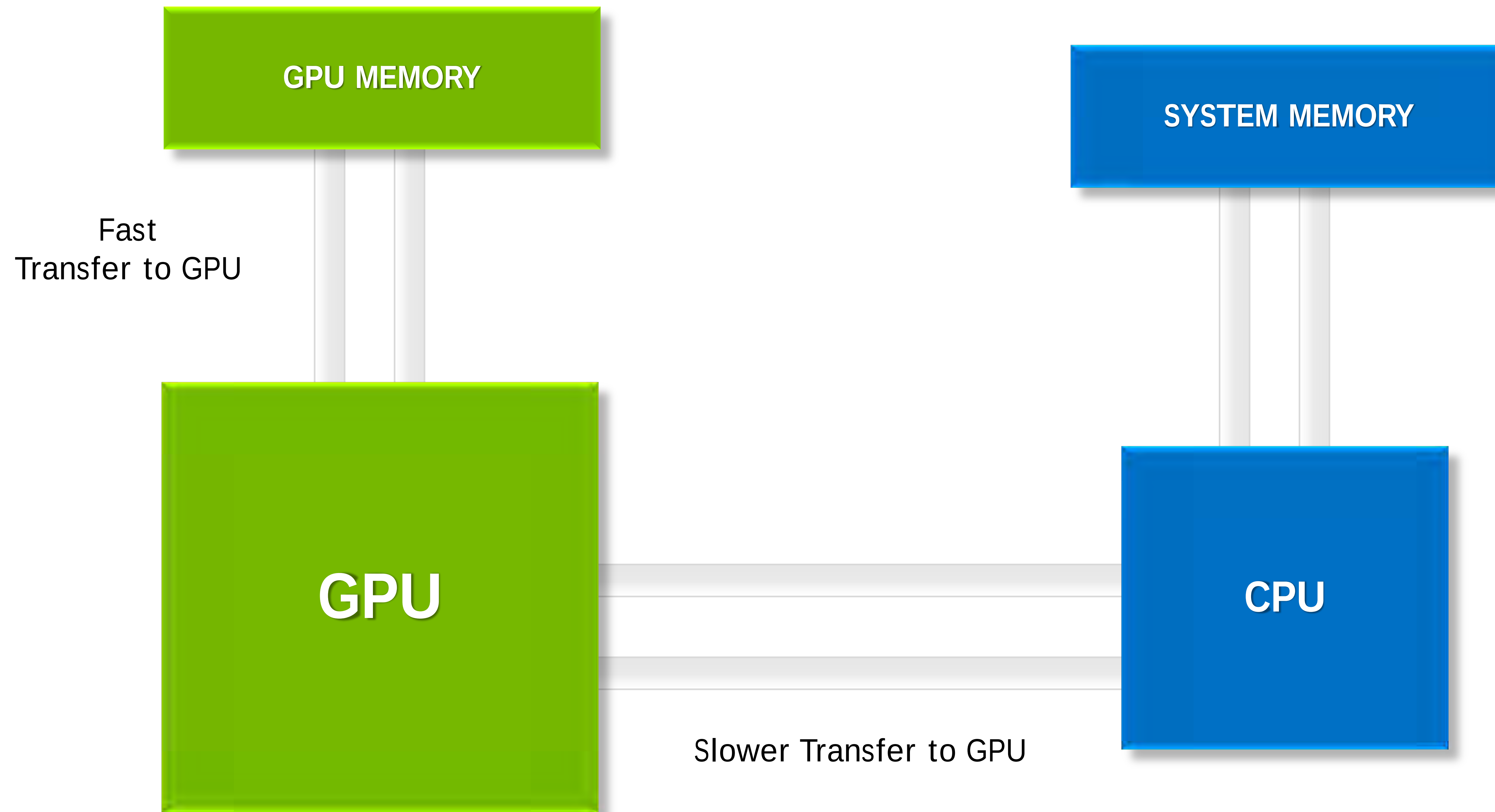
Each Application will Consume GPU Compute Cycles and GPU memory



GPU accelerated applications rely on data being in GPU memory for maximin performance

- Insufficient GPU can cause applications to revert to slower paths or fail
- Plug-ins, productivity applications, other tools (like NVIDIA Broadcast app) can add to demands of GPU memory - don't forget to include them
- Display data needs to be in GPU memory too, multiple high-resolution displays, high-resolution HMDs will add to memory consumption
- Model size continues to grow, use of materials, photo real models, overall complexity means more GPU memory

GPU PERFORMANCE



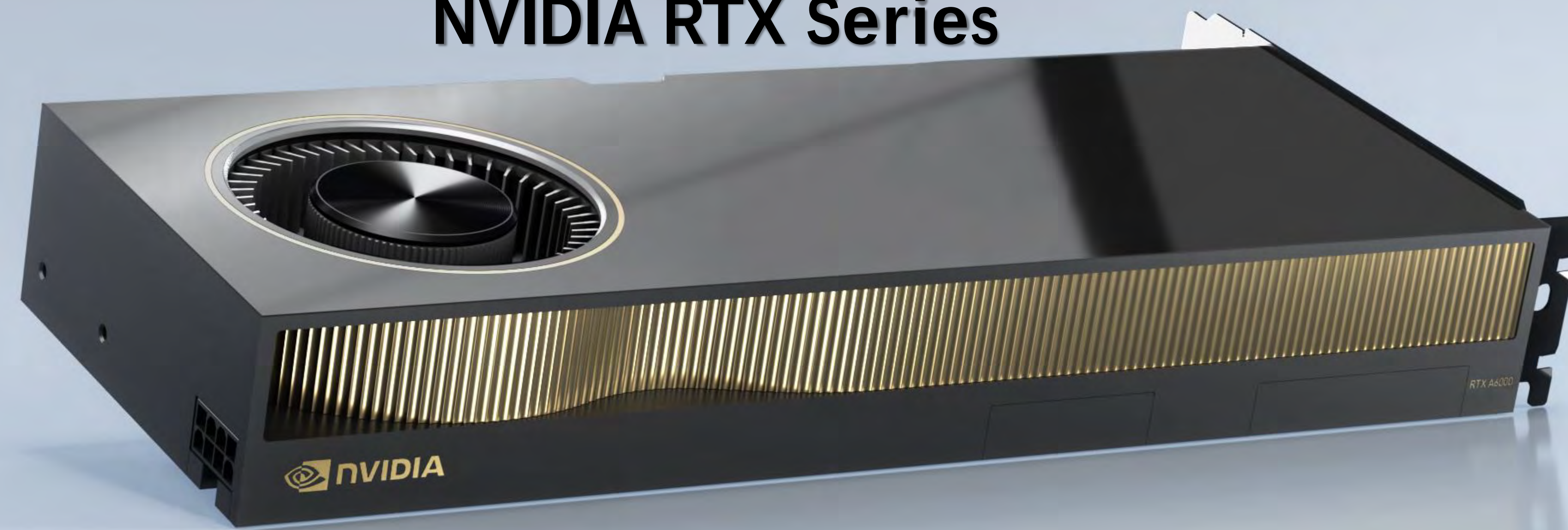
For Best Performance, Data Needs to be In Local GPU Memory

- Applications may run slower or fail to run if there is insufficient GPU memory

Boosting the World

AI Based Immersive Real-time 3D

NVIDIA RTX Series





Thank you