

GeForce RTX 30 Series vs. RTX 40 Series GPUs: Key Differences for Gamers

Discover how NVIDIA's GeForce RTX 40 Series GPUs build on the RTX 30 Series' success, elevating gaming with enhanced ray tracing, DLSS 3 and a new ultra-efficient architecture.

Author: Brian Caulfield

What's the difference between NVIDIA GeForce RTX 30 and 40 Series GPUs for gamers?

To briefly set aside the technical specifications, the difference lies in the level of performance and capability each series offers.

Both deliver great graphics. Both offer advanced new features driven by NVIDIA's global AI revolution a decade ago. Either can power glorious high-def gaming experiences.

But the RTX 40 Series takes everything RTX GPUs deliver and turns it up to 11.

"Think of any current PC gaming workload that includes 'future-proofed' overkill settings, then imagine the RTX 4090 making like Grave Digger and crushing those tests like abandoned cars at a monster truck rally," writes Ars Technica.

That said, the RTX 30 Series and 40 Series GPUs have a lot in common.

Both offer hardware-accelerated ray tracing thanks to specialized RT Cores. They also have AI-enabling Tensor Cores that supercharge graphics. And both come loaded with support for next-generation AI and rendering technologies.

But NVIDIA's GeForce RTX 40 Series delivers all this in a simply unmatched way.

Unveiled in September 2022, the RTX 40 Series GPUs consist of four variations: the RTX 4090 , RTX 4080, RTX 4070 Ti and RTX 4070.

All four are built on NVIDIA's Ada Lovelace architecture , a significant upgrade over the NVIDIA Ampere architecture used in the RTX 30 Series GPUs.

While both 30 Series and 40 Series GPUs utilize Tensor Cores, Ada's new fourth-generation Tensor Cores are unbelievably fast, increasing throughput by up to 5x, to 1.4 Tensor-petaflops using the new FP8 Transformer Engine, first introduced in NVIDIA's Hopper architecture H100 data center GPU .

NVIDIA made real-time ray tracing a reality with the invention of RT Cores, dedicated processing cores on the GPU designed to tackle performance-intensive ray-tracing workloads.

Stay updated on the latest news, features, and tips for gaming, creating, and streaming with NVIDIA GeForce; check out GeForce News – the ultimate destination for GeForce enthusiasts .

Advanced ray tracing requires computing the impact of many rays striking numerous different material types throughout a scene, creating a sequence of divergent, inefficient workloads for the shaders to calculate the appropriate levels of light, darkness and color while rendering a 3D scene.

Ada's third-generation RT Cores have up to twice the ray-triangle intersection throughput, increasing RT-TFLOP performance by over 2x vs. Ampere's best.

And Ada's new Shader Execution Reordering technology dynamically reorganizes these previously inefficient workloads into considerably more efficient ones. SER can improve shader performance for ray-tracing operations by up to 3x and in-game frame rates by up to 25%.

As a result, 40 Series GPUs excel at real-time ray tracing, delivering unmatched gameplay on the most demanding titles, such as Cyberpunk 2077 that support the technology.

Ada also advances NVIDIA DLSS, which brings advanced deep learning techniques to graphics, massively boosting performance.

Powered by the new fourth-gen Tensor Cores and Optical Flow Accelerator on GeForce RTX 40 Series GPUs, DLSS 3 uses AI to create additional high-quality frames.

As a result, RTX 40 Series GPUs deliver buttery-smooth gameplay in the latest and greatest PC games.

NVIDIA GeForce RTX 40 Series graphics cards also feature new eighth-generation NVENC (NVIDIA Encoders) with AV1 encoding, enabling new possibilities for streamers, broadcasters, video callers and creators.

AV1 is 40% more efficient than H.264. This allows users streaming at 1080p to increase their stream resolution to 1440p while running at the same bitrate and quality.

Remote workers will be able to communicate more smoothly with colleagues and clients. For creators, the ability to stream high-quality video with reduced bandwidth requirements can enable smoother collaboration and content delivery, allowing for a more efficient creative process.

RTX 40 Series GPUs are also built at the absolute cutting edge, with a custom TSMC 4N process. The process and Ada architecture are ultra-efficient.

And RTX 40 Series GPUs come loaded with the memory needed to keep its Ada GPUs running at full tilt.

All that said, RTX 30 Series GPUs remain powerful and popular.

Launched in September 2020, the RTX 30 Series GPUs include a range of different models, from the RTX 3050 to the RTX 3090 Ti.

All deliver the grunt to run the latest games in high definition and at smooth frame rates.

But while the RTX 30 Series GPUs have remained a popular choice for gamers and professionals since their release, the RTX 40 Series GPUs offer significant improvements for gamers and creators alike, particularly those who want to crank up settings with high frames rates, drive big 4K displays, or deliver buttery-smooth streaming to global audiences.

With higher performance, enhanced ray-tracing capabilities, support for DLSS 3 and better power efficiency, the RTX 40 Series GPUs are an attractive option for those who want the latest and greatest technology.

NVIDIA Ada Lovelace Architecture: Ahead of its Time, Ahead of the Game

NVIDIA DLSS 3: The Performance Multiplier, Powered by AI

NVIDIA Reflex: Victory Measured in Milliseconds

NVIDIA Studio: Your Creativity AI Take

How to Build a Gaming PC with an RTX 40 Series GPU

The Best Games to Play on RTX 40 Series GPUs

How to Stream Like a Pro with an RTX 40 Series GPU

Original URL: <https://blogs.nvidia.com/blog/2023/04/21/geforce-rtx-30-40-series/>