NVIDIA ADA GPU ARTISTRY

Showcasing the advances that are pushing the boundaries of how games and simulations are created and modded.

V1.0

Table of Contents

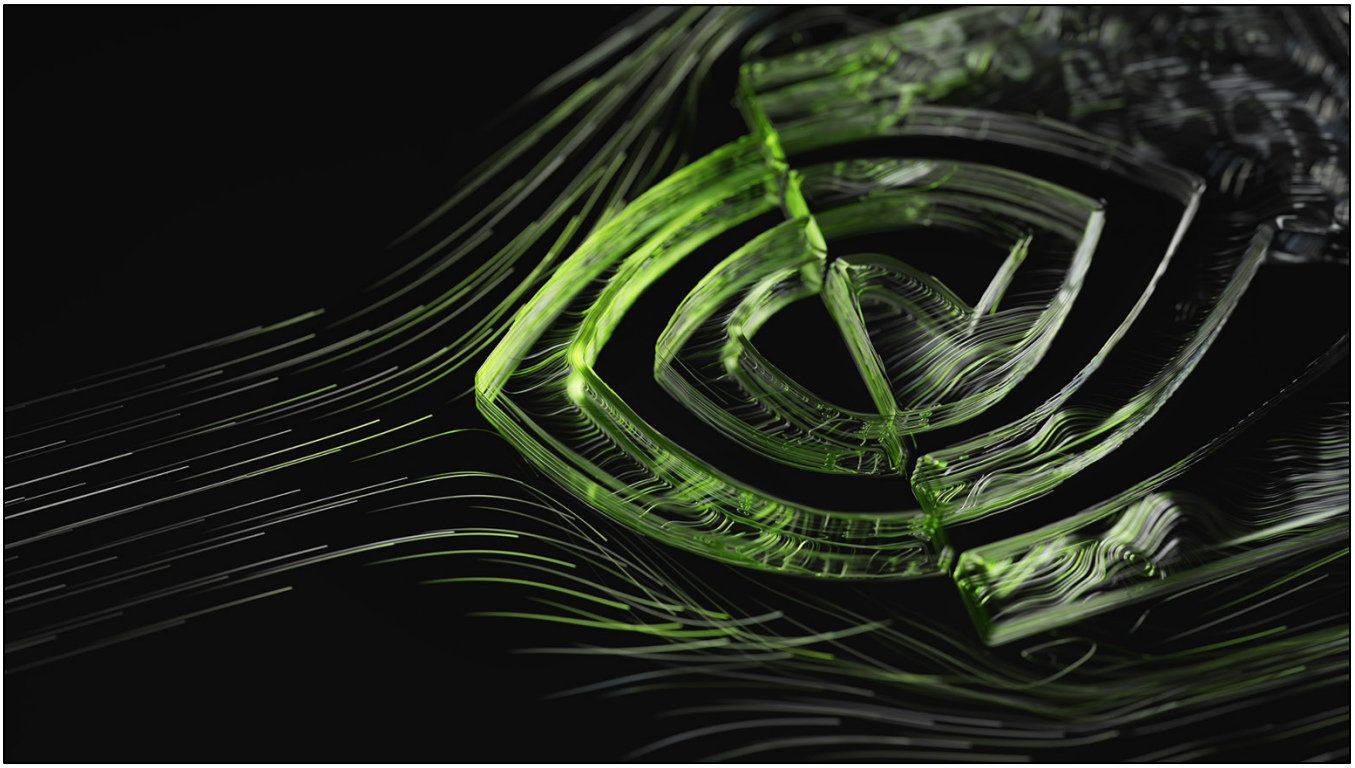
List of Figures

List of Tables

Introduction

Artistry is often used to describe the artistic quality of workmanship of an object, or the unique ability of a painter or a musician to create something profound and visionary. It's often used to denote a higher standard of quality or of expression that sets it apart from the ordinary. Computing is advancing at incredible speed, and it's being fueled by AI. These advances are not only pushing the boundaries of how games are accelerated and rendered on next-generation hardware, but also how they are being created. It's allowing games to be transformed in ways that allow designers and modders to enhance them with a higher standard level of artistry than ever before.

New breakthroughs in neural technologies like DLSS 3 use AI to massively boost performance by generating entirely new frames while maintaining great image quality and responsiveness. And now NVIDIA® Omniverse™ and NVIDIA Omniverse-connected applications like RTX Remix are putting powerful artistic tools into the hands of creators and game modders. These tools allow modders to easily replace classic game assets with greatly upscaled assets which include textures and roughness and normal maps that interact with ray-traced light realistically, all powered by AI.



In this whitepaper, we introduce RTX Remix, a modding platform built on NVIDIA Omniverse that allows modders to easily capture game assets, automatically enhance materials with powerful AI texture tools, modify assets in NVIDIA Omniverse-connected creator applications, while adding next-generation technologies including ray tracing and DLSS. Examples of demos and games built with RTX Remix in this paper include Bethesda Softworks' *The Elder Scrolls III: Morrowind* and Valve's classic game *Portal*. The fully dynamic simulation *Racer RTX*, built with NVIDIA Omniverse, is also discussed in detail along with the work that has gone into updating CD PROJECT RED's *Cyberpunk 2077* with advanced ray tracing and DLSS 3 technologies.

RTX Remix

NVIDIA typically puts a lot of time and effort into producing simulations and helping partners to upgrade games to showcase the next generation of immersive gaming technologies that new graphics architectures make possible. The effort required to upgrade classic games with updated graphics assets and new lighting and shadow technologies like ray tracing requires developers, artists, and QA experts months to years to develop, and that's with the benefit of source code and others' mods and tools. Millions of modders are downloading billions of mods each year, and nine of the ten most popular competitive games owe their existence to mods. To help this community, NVIDIA built a runtime and application environment to better help these modders.

NVIDIA engineers are passionate about gaming, and to that end, have built RTX Remix, a free modding platform built on NVIDIA Omniverse that allows modders to easily create stunning RTX remasters of classic DirectX 8 and DirectX 9 games with fixed-function graphics pipelines. In fact, RTX Remix was used to update *Portal* and *Morrowind* to showcase just how powerful RTX Remix can be to the modding community. RTX Remix consists of three key components: a USD (Universal Scene Description) extractor that converts game assets into an industry standard open 3D framework, and the RTX Remix Application to mod classic games and enhance the visual fidelity of assets. Essentially, RTX Remix allows modders to transform beloved classic games by easily capturing game assets, automatically enhancing materials with powerful AI tools, and quickly enabling RTX with ray tracing and DLSS.

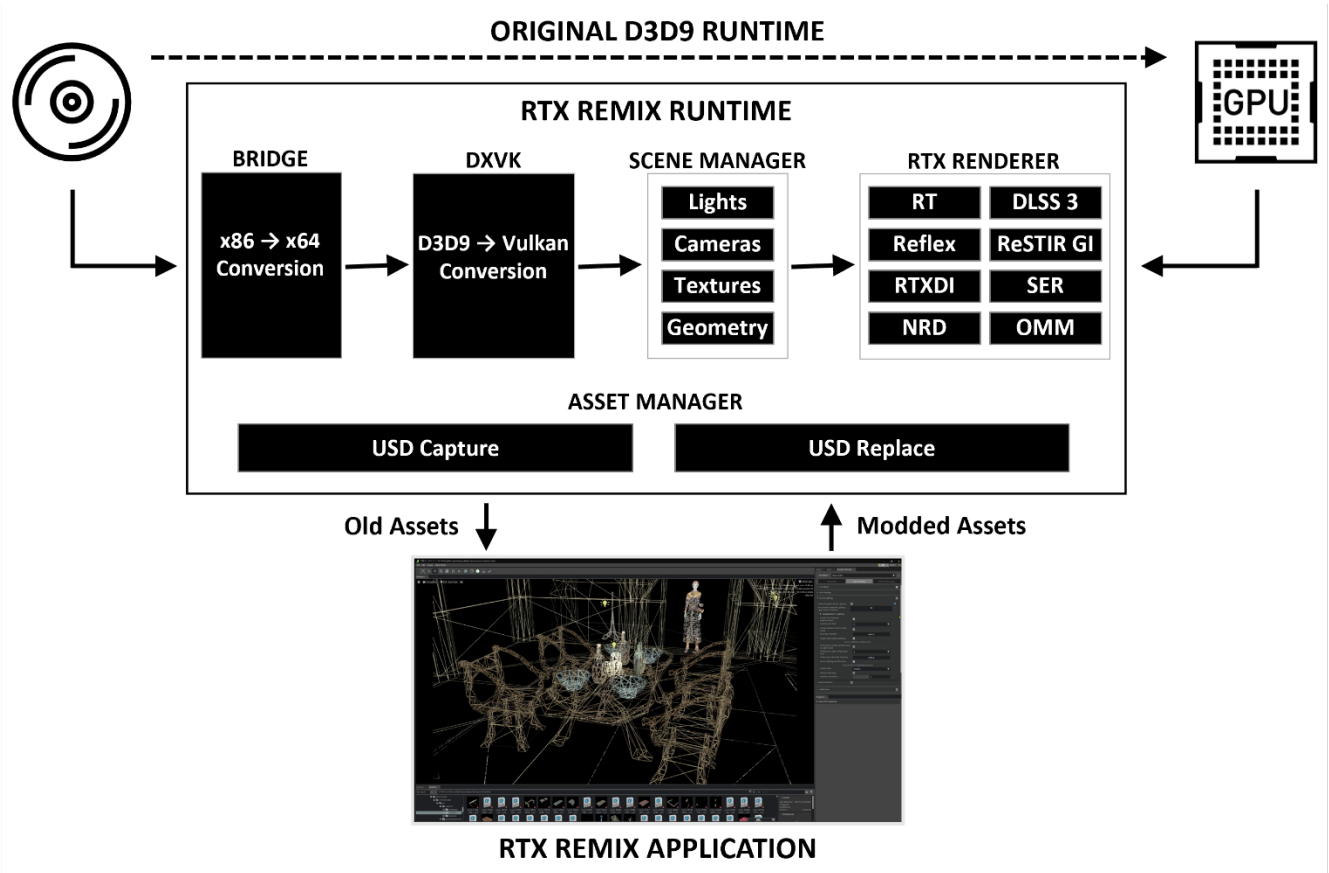


Figure 1. RTX Remix Runtime: The machinery behind the magic

The **RTX Remix Runtime** works by dropping the runtime application into a game's directory alongside the game's .exe and then launching the game. Typically, converting 32-bit/x86 code to

modern x64-bit code, required to add true ray tracing instead of screen-space mods, requires an exhaustive effort even with source code. Classic games are 32-bit and can't support ray tracing effectively because they are limited to a maximum of 3GB of system memory, which isn't enough for the high-res, enhanced assets needed to take advantage of ray tracing. Also, ray tracing APIs don't exist for older versions of DirectX games. RTX Remix state-of-the-art Bridge converts 32-bit game code to 64-bit while replacing the old rendering and system DirectX APIs with RTX Remix Runtime's 64-bit Vulkan renderer, while also upgrading the visuals on the fly.

The **Scene Manager** captures assets from the game including textures, geometry, lighting, and cameras thanks to the innovative RTX Remix Runtime. Classic games like *Morrowind* use the D3D9 runtime to send draw calls (rendering instructions) to the GPU. RTX Remix Runtime intercepts those draw calls and recognizes groups of triangles that belong to the same asset, as well as the motion vectors associated with that asset from inspecting its vertex data across frames. The Runtime actually understands assets with the same level of detail that they were programmed with in the original game, and then reassembles those assets into a cohesive, temporally coherent identical scene, which is passed to the GPU to be raytraced. NVIDIA RTX Remix Mods work alongside existing gameplay mods downloaded from Nexus Mods or other sites, giving games with rich modded content an instant ray-traced upgrade. Best of all, games can benefit from the addition of NVIDIA DLSS 3, NVIDIA Reflex, and a number of other advanced gaming technologies through the **RTX Renderer**.

Table 1. Technologies provided by the RTX Remix Renderer

RT *(Ray Tracing)* For decades, rendering ray-traced scenes with physically correct lighting in real time has been considered the holy grail of graphics. At the same time, geometric complexity of environments and objects continues to increase as 3D games and graphics continually strive to provide the most accurate representations of the real world. After spending years learning how to efficiently map ray tracing to RTX and investing considerable research into optimizing scene representations to maximize performance, including RT core traversal performance and memory access, we're seeing over 2x speedup that allows us to support real-time ray tracing. The ray tracing in RTX Remix is also known as "path tracing".
DLSS 3 *(Deep Learning Super Sampling)* A revolutionary breakthrough in AI-powered graphics that massively boosts performance, and can increase frame rates of the simulations, games, and demos outlined in this paper by up to 3x with max quality settings.
NVIDIA Reflex A new set of APIs for game developers to reduce and measure rendering latency. By integrating directly with the game, Reflex Low Latency Mode aligns game engine work to complete just-in-time for rendering, eliminating the GPU render queue and reducing CPU back pressure in GPU intensive scenes. This delivers latency reductions and increases responsiveness above and beyond existing driver-only techniques.
ReSTIR GI *(NVIDIA Reservoir Spatio Temporal Importance Resampling Global Illumination)* A new global illumination technique that calculates per-pixel correct indirect

lighting for multiple bounces in a highly performant way. ReSTIR GI allows indirect light to bathe a scene and can illuminate dark corners that are not lit directly.

RTXDI *(NVIDIA RTX Direct Illumination)*
Enables you to add millions of dynamic lights to your game environments without worrying about performance or resource constraints. Imagine thousands of tiny LEDs, sci-fi control panels, Times Square billboards, or even exploding fireballs. RTXDI makes this possible while rendering in real time, enabling geometry of any shape to emit light, cast appropriate shadows, and move freely and dynamically.

SER *(Shader Execution Reordering)*
Adds a new stage in the raytracing pipeline, enabling rays to be traced faster without any compromise on visual quality. It works by reordering and grouping the secondary hit shading to have better execution locality, thus much higher overall path-traced shading efficiency.

NRD *(NVIDIA Real-Time Denoiser)*
A spatio-temporal ray tracing denoising library that assists in denoising low ray-per-pixel signals with real-time performance. Compared to previous-gen denoisers, NRD improves quality and ensures the computationally intensive ray-traced output is noise-free, without performance tradeoffs.

OMM *(Opacity MicroMaps)*
With OMMs, developers can very compactly describe irregularly shaped or translucent objects, like ferns or fences, and directly ray trace them with the NVIDIA Ada Lovelace RT Core more efficiently. Although they don't make any difference visually, gamers will get a performance boost. OMM speeds up ray tracing of alpha-tested geometry by a factor of 2x.

NOTE: Refer to the ADA Launch Whitepaper | Technology for more information on DLSS 3, Shader Execution Reordering (SER), and Opacity MicroMaps (OMM).

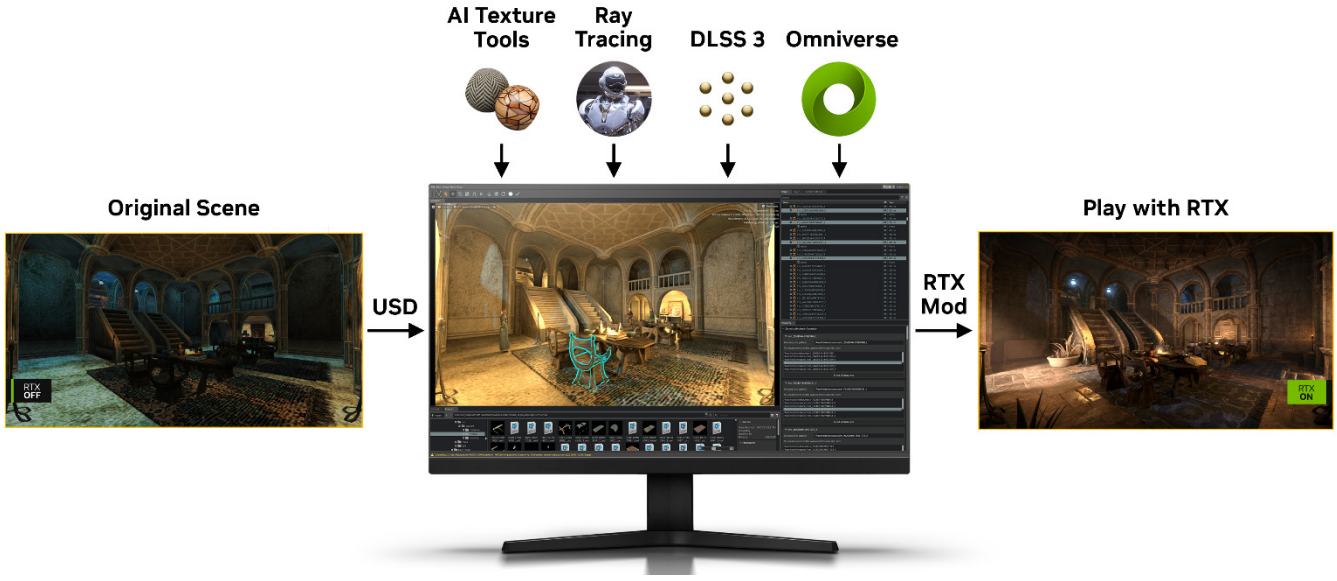


Figure 2. RTX Remix: Remaster the classics with RTX

The **Asset Manager** contains an extractor called USD Capture that captures the game's materials, meshes, and levels while the game is played, converting them to the industry standard Universal Scene Description (USD) open 3D framework, which is the foundation of the NVIDIA Omniverse platform for building and operating custom 3D pipelines. USD is widely adopted and can be opened by most art and modeling applications and shared by NVIDIA Omniverse applications. With permission from Bethesda, we used ***Elder Scrolls III: Morrowind***, one of the top modded games of all time, as an example of what can be accomplished with RTX Remix.

Since RTX Remix is built on NVIDIA Omniverse, USD game assets can easily be imported into the RTX Remix application or any other NVIDIA Omniverse app or connector, including game industry-standard apps, such as *Adobe Substance 3D Painter, Adobe Photoshop, Autodesk Maya, 3ds Max, Blender, SideFX Houdini*, and Epic Games' *Unreal Engine*.

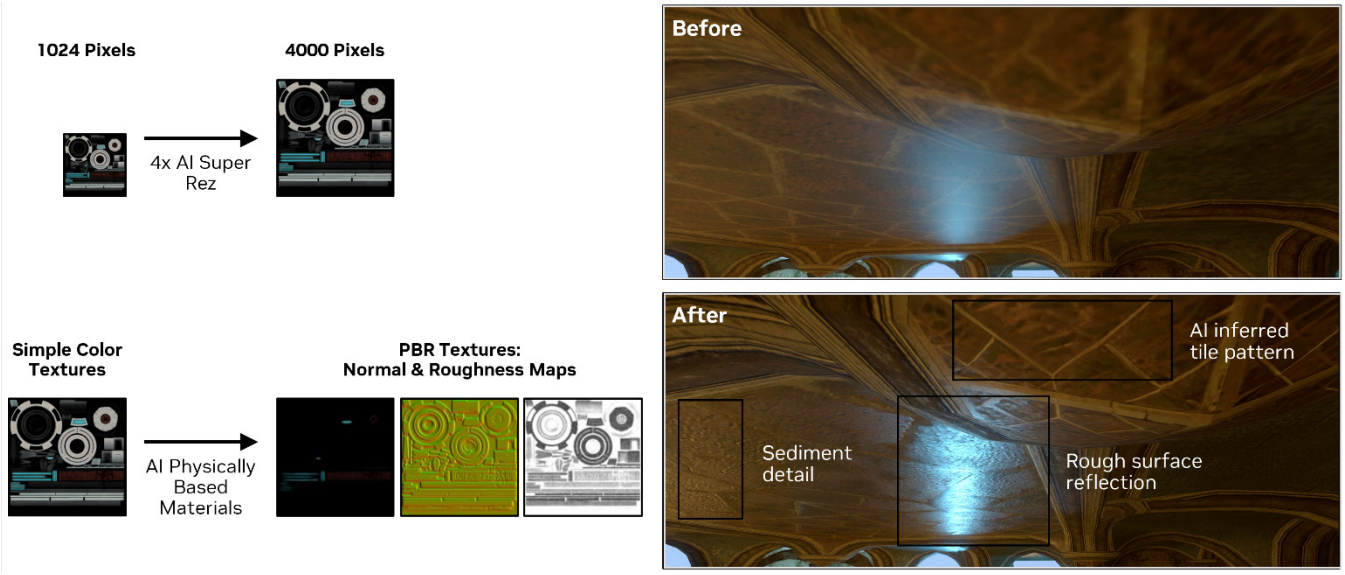


Figure 3. RTX Remix: AI Texture Tools

RTX Remix simplifies and accelerates the art remastering process using artificial intelligence. An AI-assisted toolset features deep learning tools to automatically upscale assets by 4x, turning 1080p-ready textures and assets into higher-quality 4K resolution ready assets. Also, an AI PBR (Physics-Based Rendering) replacer scans the environment, identifies the materials the scene is composed of, and renders those materials with modern, physically-accurate properties. For example, a cobblestone floor that would otherwise be a basic flat-looking texture gets turned into a detailed stone surface with a roughness map that interacts with light realistically.

The RTX Remix application isn't just for tweaking material properties of game assets. It allows modders to view lighting and shadows in isolation through a white mode, and with little effort, create new lights that help bring realistic global illumination, shadows, ambient occlusion, and reflections to a game. Modders and gamers can customize their experience during gameplay from the overlay, adjusting lights dynamically and even tweaking volumetrics to make for a more immersive scene. Other effects include customized lighting intensity, fog, and post-processing effects like bloom and motion blur. Advanced users can even use diagnostic and heuristic tools to highlight every changed asset in an RTX Mod, enabling a collaborative mod team to easily review an in-progress mod as they play, leading to a better public release.

In fact, NVIDIA artists used RTX Remix to remaster the scene to the right. The team loaded the assets into a shared RTX Remix project via NVIDIA Omniverse, and as one member updated an object's material properties, or an asset's mesh in NVIDIA Omniverse-connected applications like *Substance* or *Maya*, the game scene would sync in the Remix viewport, making collaboration a breeze.

A second artist could then spend their time in the Remix application, meticulously moving each completed asset to its perfect position, and relighting the scene with the rebuilt objects in mind. It's that simple, enabling round-the-clock collaboration and development with a mod team spread across the planet. Previously, artists would need to wait for one person's work to be finished and sent. Now, there are no delays, and another can pick up where the first left off.

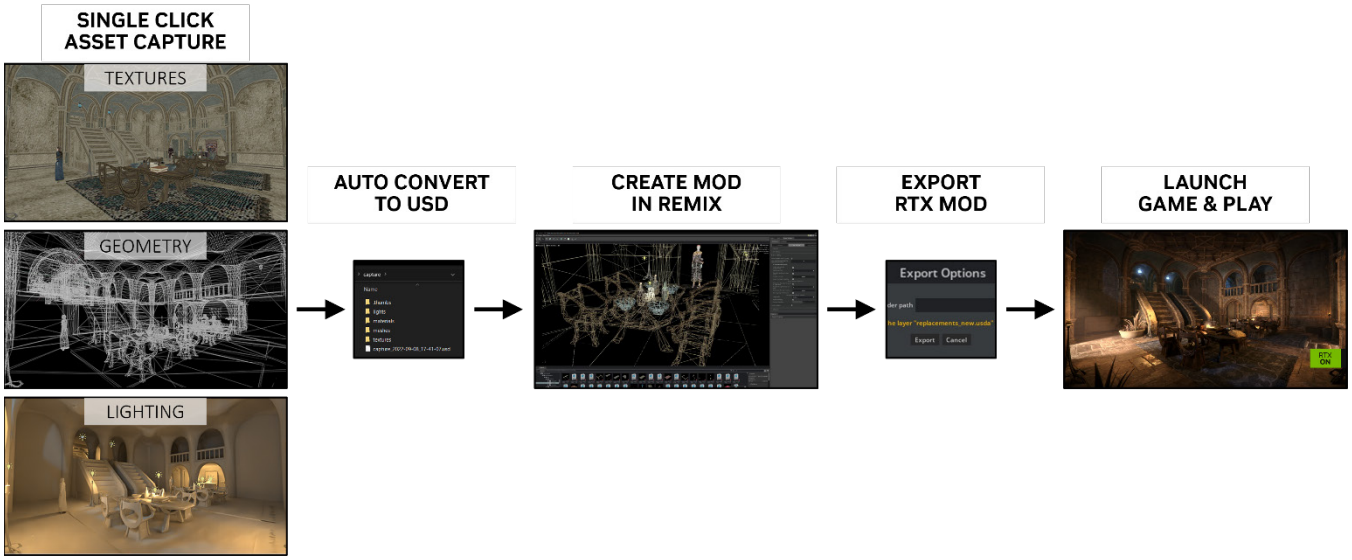


Figure 4. RTX Remix: Simplifies Asset Capture, Import, & Export

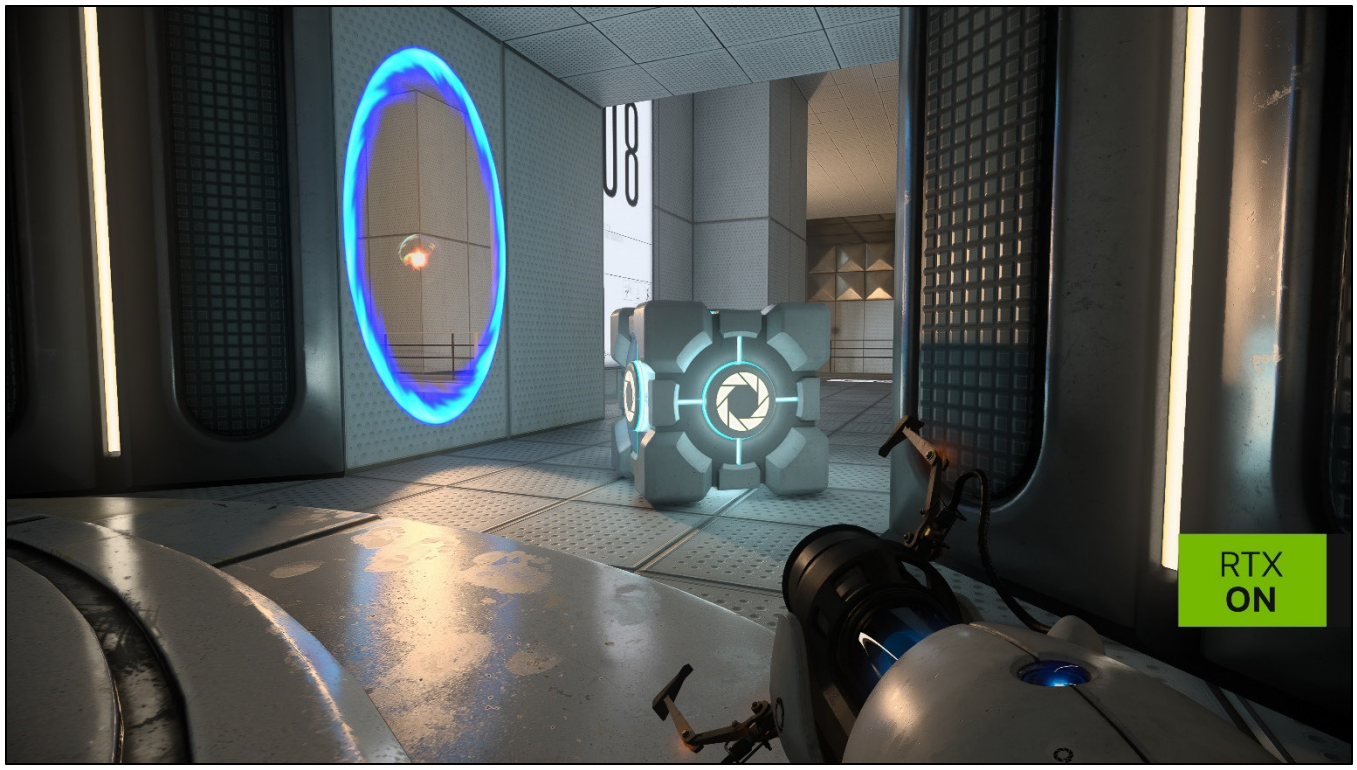
When the RTX Remix Mod is ready, it's easy to export and share with fellow gamers online - they simply download the mod and drop it into the game's directory alongside the .exe and launch the game. The NVIDIA RTX Remix Runtime does the rest, replacing the old rendering APIs and systems with RTX Remix Runtime's 64-bit Vulkan renderer, and upgrading the visuals on the fly. The player also benefits from the instant addition of NVIDIA DLSS 3, for incredibly smooth frame rates, and NVIDIA Reflex, which improves responsiveness by reducing end-to-end system latency. And best of all, NVIDIA RTX Remix Mods work alongside existing gameplay mods downloaded from Nexus Mods, Mod DB and other sites, giving games with rich modded content an instant ray-traced upgrade.

Portal with RTX

Portal with RTX is the work of NVIDIA Lightspeed Studios, in close association with Valve, that reimagines the classic game with full path tracing and updated meshes and materials remastered in NVIDIA Omniverse using NVIDIA RTX Remix. Observant *Portal* enthusiasts will notice that although the game is based on *Portal*, many of the higher-detailed models and textures were taken and expanded from *Portal 2*.

Portal, a first-person platformer with uniquely challenging puzzles that required a portal gun that creates portals in space, was launched in October 2007 and was part of *Valve's Orange Box*. It is still regarded as one of the greatest games of all time.

The rendering technology at the time of *Portal's* release was solely rasterization, particularly from a lighting model perspective, which used traditional forward lighting and pre-baked static light maps with no support for dynamic object shadows.



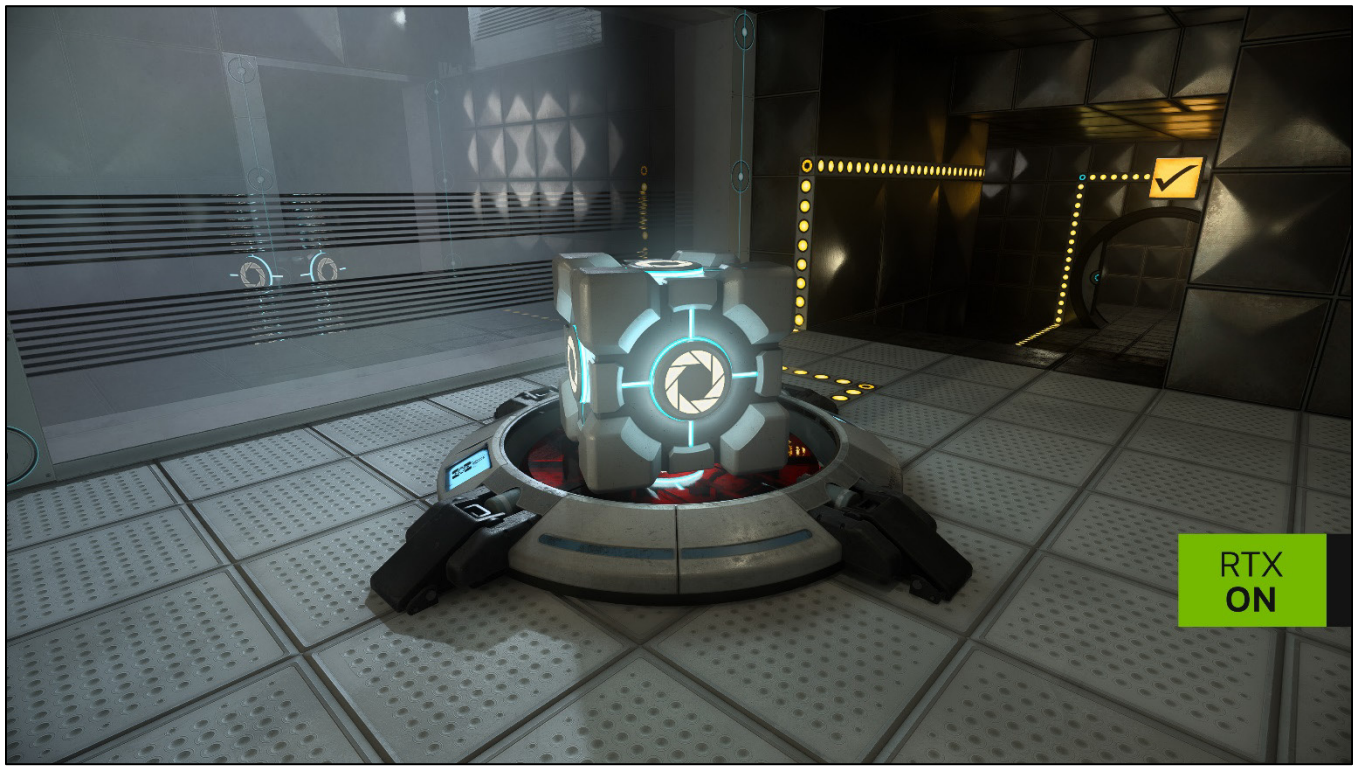
The most important new aspect of light in *Portal* is that now light itself comes through the portals, which is a mechanic that is completely missing in the original game. Due to the number of light bounces happening in this scene, up to four at times, light from the portal can be seen reflecting off the floor and then again off the wall to the right.

Figure 5. *Portal with RTX:* Light Reflecting Off Multiple Surfaces

Portal with RTX uses meshes and materials remastered in NVIDIA Omniverse, and is rendered in real-time with path-tracing and denoising provided by NVIDIA RTX Remix's 64-bit Vulkan renderer. By using the RTX Remix platform to implement path tracing, NVIDIA Lightspeed Studios has brought an entirely new lighting model to *Portal*. Path Tracing is an advanced form of Ray Tracing that is capable of tracing many more rays through each pixel, and then following those rays through numerous bounces off or through objects, before reaching the light source in order to collect color and lighting information to bring back to a pixel. It provides more accurate reflections and global illumination to light rooms and environments with indirect lighting by bouncing rays off primary rasterized surfaces.

The first thing gamers will notice is that suddenly, there is complete lighting on every object including companion cubes. The portals themselves are now illuminated, which helps the player understand where to go. These objects are not pre-baked anymore. When you pick up a cube, or when you rotate around with the camera, you see reflections in the shadows on the portal gun and on the objects you're holding, which really cements those objects in the scene in a way that wasn't possible before. Reflections now happen in world space, so that light is traced accurately in the room that you're in and in adjacent rooms, which makes lighting through the portals possible, because now that gamers can see through portals, the light can travel through portals as well.

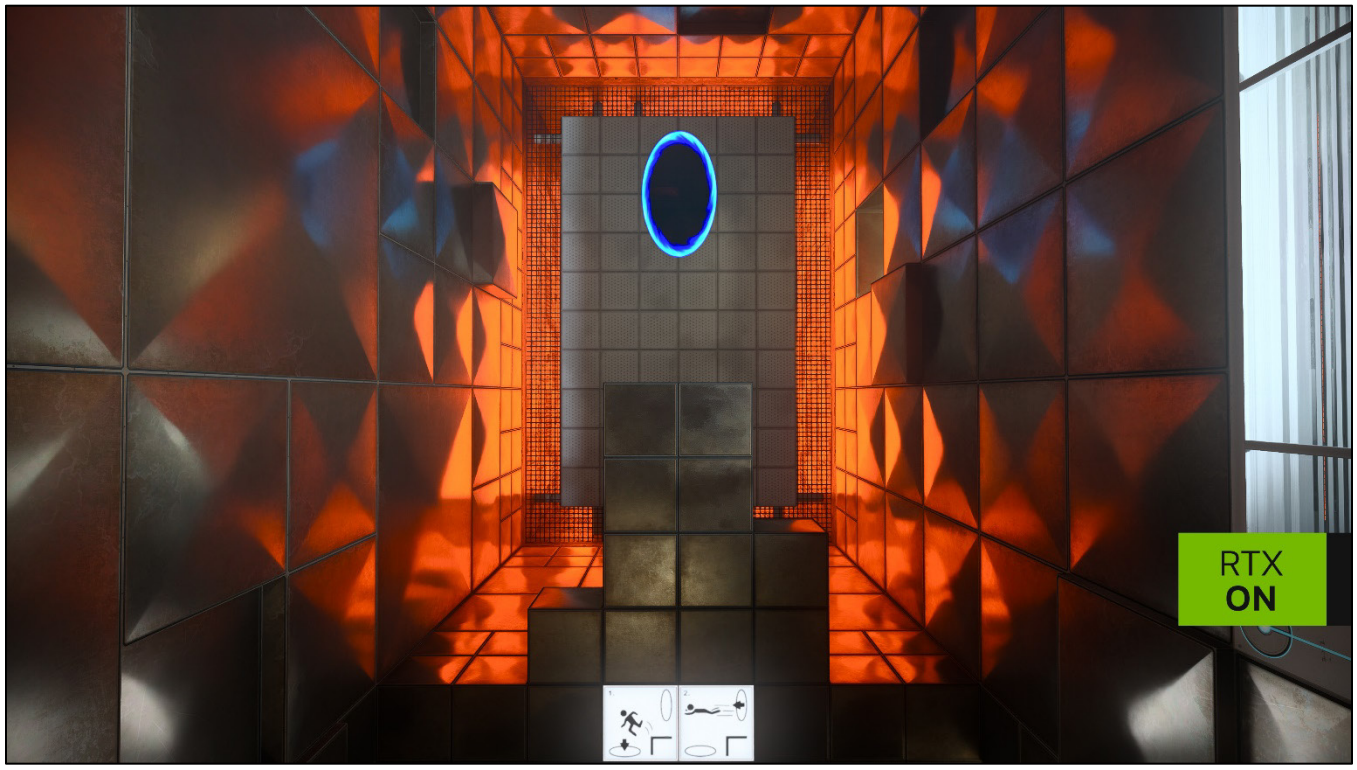
With RTX Remix, new light sources can be easily created and placed and there are tools for determining how emissive they are when attached to walls and objects. Also, there are tools that allow modders to tweak materials and surface colors and how they light a room. Before, only point lights were available, which were shaped as spotlights, but now, light sources can be converted into spheres, cylinder lights, disc lights, and area lights. Particles are difficult to render with ray tracing, but in Portal With RTX, particles are modeled accurately in three dimensions, allowing them to be reflected properly–a feat that is uncommon even in AAA development. Particles even emit light, alongside the Cube, Portal rings, floor lighting strips, and energy balls.



All of the objects in *Portal* that can be interacted with have a specific queue and they all emit light in some way. This includes the buttons, which are crucial to gameplay, and the cubes that are placed on the buttons. It also includes the turrets that shoot at you. And with path tracing and dynamic lighting, those objects stand out just a little bit more.

Figure 6. *Portal with RTX:* Portals and Objects Now Emit Lights

Other changes were made to *Portal with RTX* such as adding higher-resolution materials, because the original 512x512 textures were not good enough by modern standards. Using RTX Remix, 100% of textures and assets were replaced. Many of the new textures were recreated and swapped by texture artists that were faithful to the original textures. A total of 195 models were exported and rebuilt from scratch, preserving the silhouettes while increasing fidelity, meshes, and resolutions.



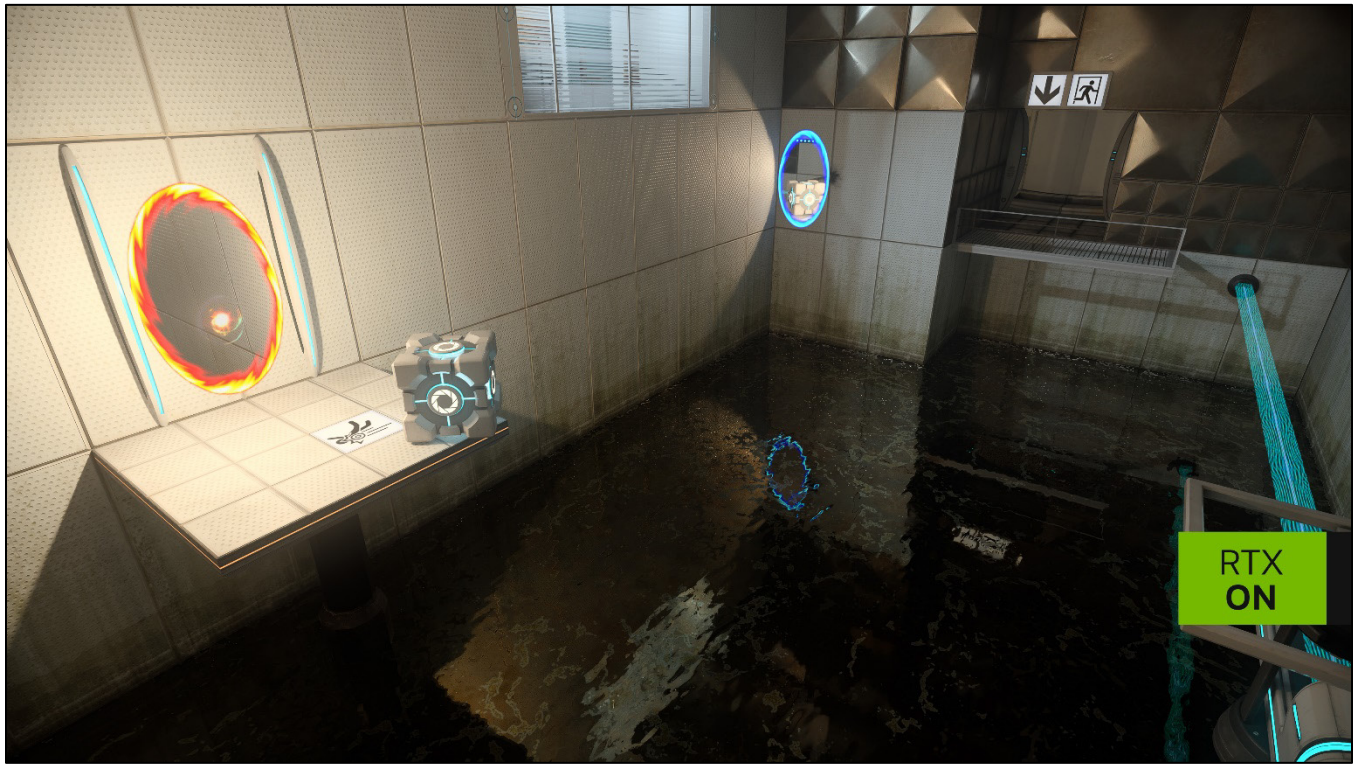
This scene showcases indirect illumination from multi-bounce reflections being cast from the red lights. The bounces from this volumetric lighting combine to create specular reflections on the walls, helping to set the atmosphere for the scene. Lights can bounce off of four surfaces, although not every pixel will require four bounces to be shaded effectively, some may gather sufficient shading information from one or two bounces.

Figure 7. *Portal with RTX:* Indirect Illumination with Multi-Bounce Reflections

Updating the assets started with the original game. With the RTX Remix Runtime loaded, a “print-screen” captured the game to USD (Universal Screen Description). This both works for a single frame capture, as well as a time-sampled USD animation of gameplay. A USD file containing the current representation of the world, and all associated assets were dumped to disk so that it could be loaded into any NVIDIA Omniverse-supported software, essentially creating a facsimile of the game world. The Lightspeed Studios team designed new materials through a PBR (Physics-Based Rendering) virtual material pipeline to represent what the game had originally, rather than having to redesign from scratch. This pipeline allowed the modders to simulate any kind of physical material to particularly improve a 3D model, including metallic materials now that screen space ambient occlusion (SSAO) is being used. Utilizing USD Layering, remastered assets are

applied on top of the original game's assets, both in NVIDIA Omniverse and in the RTX Remix runtime.

Adding different levels of gloss and metallics to objects allows objects, including the portal gun as mentioned earlier, to be reflected in the environment as light changes and moves through the scene, allowing simple assets to better come to life. However, some of the materials needed to be hand-painted to match the look of the original textures. The Lightspeed team also attempted to remove some of the excess repetition and tiling, now that the game had access to higher resolution textures with larger texel density.



Portals themselves emit light, and not only in immediate view, but also from bounces of secondary rays in the steam and dirty water below. This allows the players to have an intuition of where the portals are located within a scene, even if they're not looking directly at them because they may be visible in a reflection on the water.

Figure 8. *Portal with RTX:* Dynamic Lighting Enhances Game Mechanics

Despite all of these changes, there was a mandate to ensure changes did not affect gameplay. In fact, specific tools were provided to help ensure the accuracy of lights and shadows in a scene so that new features like reflections did not obscure gameplay cues. There are areas in the original game where reflections were included that allowed a robot to be seen around the corner on a pipe. We had to be mindful when using these new tools to ensure that the intensities of lights and shadows were matched with ray tracing and global illumination, because they were originally used for sight lines to help guide the player through the environment and to highlight specific spots and places where they needed to look, such as a button. Therefore, clever light sources had to be created rather than resorting to fake-looking floating light sources.

Racer RTX

Racer RTX is a fully dynamic simulation built with NVIDIA Omniverse™, an easily extensible platform for real-time 3D design collaboration. Take control of realistically rendered and physics-based RC cars and navigate them around four unique sand-box style environments while interacting with all the physically modeled objects scattered about. Each environment is filled with photo-real objects, simulated in real time in NVIDIA Omniverse.

Racer RTX was built to showcase a number of cutting-edge technologies including advanced real-time ray tracing, vehicle AI control, PhysX car control, real-time rigid bodies physics, volumetrics, and real-time Flow effects. All of these technologies and effects can be controlled through exposed render settings in the simulation. Light is not pre-rendered or artificially baked in, it's physically accurate with every reflection and refraction of light dynamically adjusted in real-time within the world space. Using these technologies, modders and gamers are able to assemble obstacle courses to run cars that they design and build, and then ultimately record videos of their cars racing around the simulation using an AI path tool.

The RC cars are modeled in unprecedented detail, with 50-70 parts in the suspension/joint area, simulated in real time with real world parameters using PhysX.

Figure 9. *Racer RTX:* RC Cars Modeled in Unprecedented Detail

The cars in *Racer RTX* are not models moving around with simple code, they're actually a physical simulation of accurately modeled rigid-body cars. The physical properties of these vehicles like body length, wheel size, and center of mass are important to how the different vehicles in the simulation are differentiated, because physics dictates how they perform and drive. When accelerated, PhysX is used to apply force and torque, and the spring coefficient and the geometry of the suspension actually makes a difference in how the vehicle drives and behaves over and through different terrain. These vehicles are dampened such that energy is drained as they need to overcome frictional or other resistive forces such as ground clutter or soft terrain.

Once the world is built, it needs to move.

Figure 10. *Racer RTX:* Physically Based Rendered Objects

The Universal Scene Description (USD)-based levels composed of 1,811 hand-modeled, textured and simulated assets objects, were built in 3 months by NVIDIA artists across 12 time zones using NVIDIA Omniverse. Artists collaboratively contributed to the shared NVIDIA Omniverse world using their preferred design and content creation tools, such as *Autodesk 3ds Max, Maya, Blender, Modo, Maxon ZBrush, Adobe Substance 3D Painter, Substance 3D Designer, Photoshop, Illustrator, Rizom UV,* and *SideFX Houdini*, achieving interoperability via the USD file format.

The NVIDIA Material Definition Language (MDL) was used to build a library of materials that were shared in NVIDIA Omniverse for all of the models and rigid bodies used. These materials use physically-based rendering (PBR) to describe the visual properties of the surfaces and materials to ensure realistic results under all lighting conditions. The clutter littered through the simulation are not static props. All of the jigsaw puzzle pieces, Lego blocks, toys, robots, dice, and other potential obstructions are physically simulated rigid bodies that interact with physics and can be moved by the car, or not, if they're too big or heavy to be moved. All are tied to the Physics simulation extension that's built into NVIDIA Omniverse. The throw blanket on the couch is controlled by NvCloth, a library that provides low level access to a cloth solver that provides real-time cloth simulation and collision detection. The dust kicked up by the car in the dirt outside is handled by NVIDIA Flow, a GameWorks feature that offers support for real-time liquid, fire, dust, and smoke simulation, and the car leaves tracks in the sand with support from real-time terrain deformation.

The next important step is how that light interacts with your environment through materials. Materials are incredibly important in 3D worlds and they need to be physically accurate.

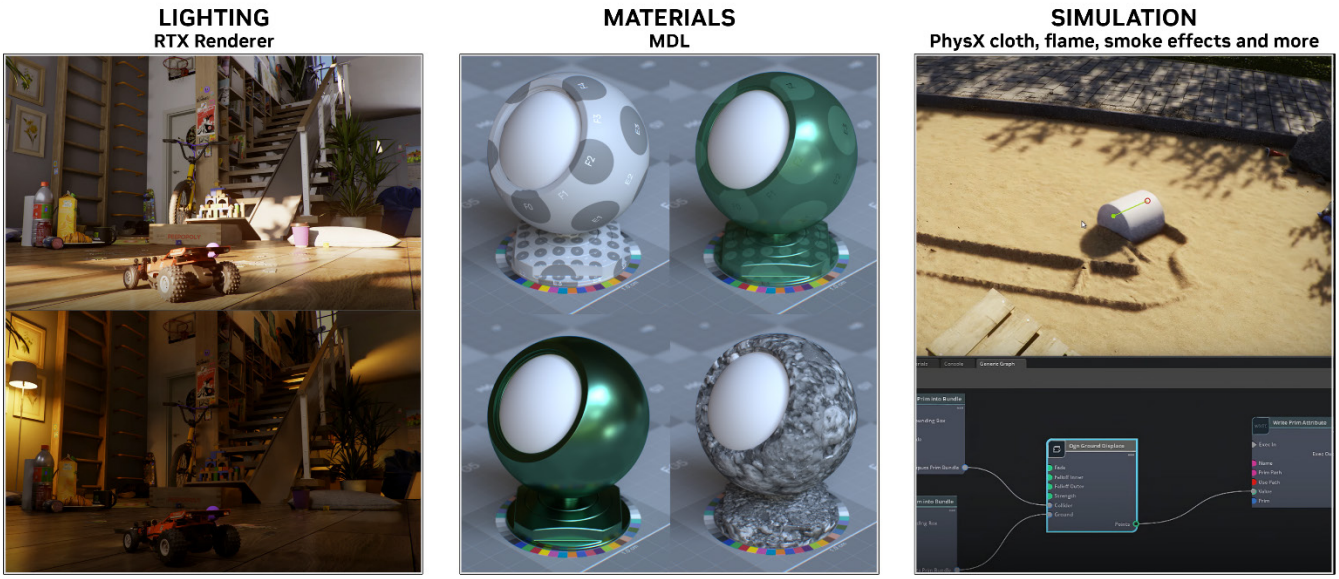
Figure 11. *Racer RTX:* Light Interacts with the Environment Through Metals

None of the lights in *Racer RTX* are pre-baked, so moving a light will affect the way the scene is lit and shadowed in real-time. The dynamic nature of lighting is due to real-time ray tracing that helps to make content look more photoreal. However, a secondary benefit is to introduce offline workflows in the real-time world, which helps the overall process of creating games and simulations like *Racer RTX*. In the end, it's faster, cheaper, and easier when designers don't need to worry about how to bake in lights or occluders when all they need to do is add content, turn on the lights, and it all works. *Racer RTX* contains an exposed menu system that allows for the control of every lighting aspect in the game in real time, including the time of day.

With real-time ray tracing, it's important to understand that the number of rays used in a scene is adaptive and dependent upon what happens in the scene, but on average, ~20 rays are used per pixel. The nature of each traced ray is different - some are cheaper and shorter, while others require a lot more calculations, or are needed to be traced to longer distances. For some rays, like primary rays, there is a lot of computations needed to gather the texture data and the material scattering properties, while for others, like ambient occlusion rays, this is not needed.

When a primary ray hits somewhere, we compute all the direct lighting there, and then we collect and compute indirect light contribution from the bounce. We use multiple different passes to compute the different parts of the rendering equation. This allows for the separation of different light components and is useful for the denoising process later. The indirect diffuse global illumination is computed in real time using a combination of ray tracing and world space caching. The result is a completely dynamic, physically based light transport that interacts accurately with every object and light in the scene. The world space cache allows for accurate simulation of many indirect light bounces, without paying the price of recomputing the same data multiple times.

This way, the price of getting the indirect global illumination is just a single ray trace and query to the global illumination cache. Doing this allows multi-bounce global illumination contributions to be added not only to the primary rays, but also to each reflection bounce, which is incredibly important for achieving photorealism.



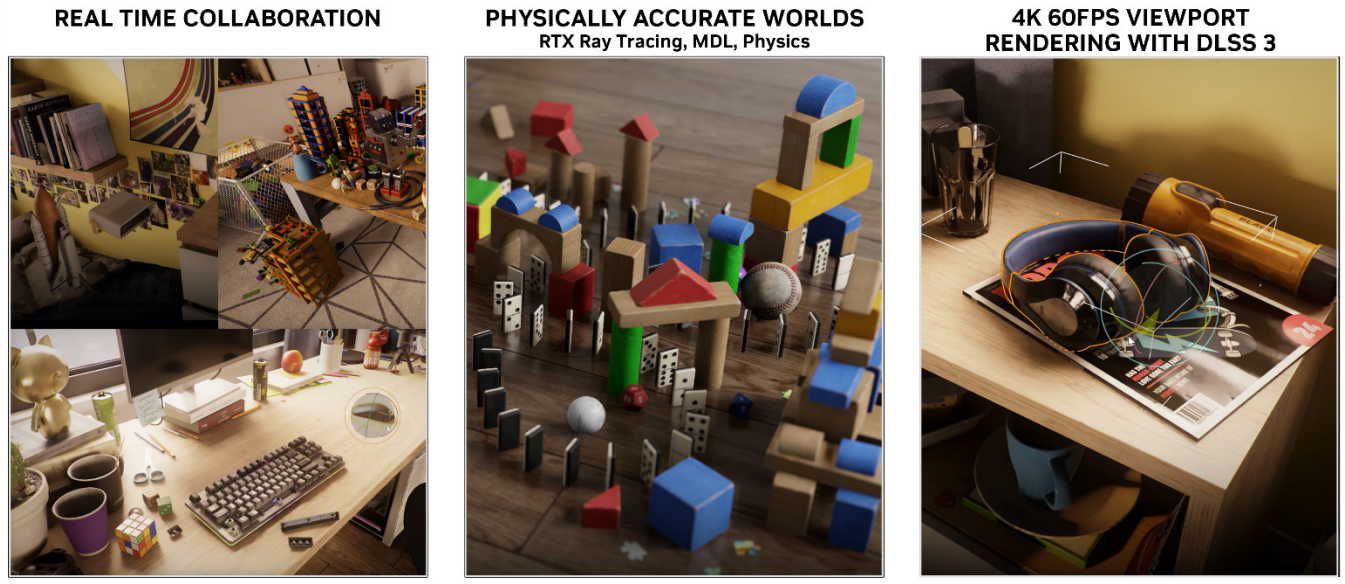
No Precompute or Baking — Everything Behaves as in Reality.

Figure 12. NVIDIA Omniverse Enables Physically Accurate Worlds

Reflections and refractions in *Racer RTX* are simulated accurately, both for materials that are purely specular - like perfect mirrors and glass - and for materials that have more complicated properties. In the real world, most materials are neither perfectly diffuse, nor perfectly specular and their light scattering properties are modeled with a variety of scattering functions. This requires the computation for how the lighting interacts with each of the layers of materials on each of the ray bounces. Ray tracing in environments that are made of physically-based and accurately modeled materials is often an order of magnitude more computationally challenging than perfectly specular materials.

The number of rays for ambient occlusion used is adaptive and can be anywhere from 3-9 per pixel, depending on the scene. These rays are short and do nothing after they hit a destination; only the length of the ray influences the AO calculation.

Direct illumination, due to an usually high frequency of details in a scene, is the most challenging part of the lighting equation, both for computing and denoising. Direct illumination in *Racer RTX* is modeled by a variety of area, distant, and dome lights. For every frame, a number of shadow rays are traced towards the light sources with the goal of achieving a variance-free result as much as possible. The number of shadow rays needed to achieve high quality results depends on the scene and the lighting setup (time of day). The materials used in *Racer RTX* are production-quality and have non-trivial scattering properties. For most scenes, 4 to 8 shadow rays are enough to achieve low noise results that are suitable to be denoised, while preserving all the details of the production materials used in the scene.



With all these technologies together, we finally have the full 3D real time design experience.

Figure 13. NVIDIA Omniverse Enables a New Paradigm for 3D Design

NVIDIA DLSS uses AI to deliver a significant breakthrough in graphics performance and image quality. With DLSS 3, *Racer RTX* benefits from smoother frame rates with little-to-no impact on latency while delivering 81 FPS at 4K on GeForce RTX® 40 Series graphics cards. Performance also gets a boost from all-new Optical Multi-frame Generation, allowing even the most computationally intense ray traced 4K scenes to be delivered with buttery smooth frame rates. Racer RTX did not require Ada's Shader Execution Reordering (SER) technology in every part of the ray tracing pipeline, such as for direct lighting, due to the coherency of primary rays hitting mostly the same things. However, SER *is* used for the reflections pass, which is more computationally intensive.

Cyberpunk 2077

NVIDIA and CD PROJEKT RED teamed up in 2020 to bring DirectX Ray Tracing (DXR) and NVIDIA DLSS to *Cyberpunk 2077*, enabling new heights of immersion with ray-traced shadows, reflections, global illumination, diffuse illumination, and ambient occlusion— with DLSS providing state-of-the-art AI super resolution. The game launched in December of that year, and both companies continued working together post-release to improve the image quality of ray tracing, including the addition of ray-traced local shadows and sun-based shadows. And now, even more improvements are being made, including advanced ray tracing with global illumination, multi-bounce reflections, direct lighting through NVIDIA RTX Direct Illumination (RTXDI) and support for cast shadows, and better occlusion. Further improvements include opacity micromaps, Shader Execution Reordering (SER), and support for DLSS Super Resolution, NVIDIA Reflex, and DLSS 3 with Frame Generation, all of which provide large improvements to performance.

When *Cyberpunk 2077* was originally released, ray-traced global Illumination (indirect lighting) was a hybrid ray tracing technique that combined traditional rasterization and ray tracing, and it only supported a single Sun bounce that was subtle, but very realistic. Also, the original ray-traced shadows were for the Sun and Moon only. However, it only worked when the game was set to the Psycho level setting. Global illumination can be thought of as the complement to ambient occlusion: a lighting technique that models how light scatters off of various surfaces to indirectly illuminate other objects in the scene. Some rays directly reach your eyes from light sources, others may be blocked by objects in the scene causing shadows, and still others reflect or refract off other objects before reaching your eyes. When a ray of light strikes a surface, surrounding game elements will be naturally illuminated. In addition, if the contact surface or light source happens to be colored, that color will naturally spread to surrounding detail.

Ray-traced indirect lighting and reflections with a single bounce, combined with denoising, can result in artifact-free and physically correct reflections for shiny surfaces, as well as semi-glossy

and rougher ones, like glass panes, painted handrails, and brushed metals. But now, *Cyberpunk 2077* supports path-traced multi-bounce reflections, resulting in more accurate, realistic, and immersive global illumination, reflections, and self-reflections. This makes specular reflections possible when lights shining on the floor are highlighted on the wall next to it. More importantly, path-traced refraction enables glass and water to be rendered with a realistic index of refraction, so that rays change direction when reflecting on those surfaces like it does in the real world.

This is the scene without path-traced global illumination.

Figure 14. *Cyberpunk 2077:* Without Path-Traced Global Illumination

Adding Path-traced global illumination simulates the true physics of light and provides better occlusion and gradients through multi-bounce rays.

Figure 15. *Cyberpunk 2077:* With Path-Traced Global Illumination

In the original release of *Cyberpunk 2077* with ray tracing, diffuse Illumination (direct lighting) handled emissive surfaces like the neon signs and glowing advertisements, and although it was not screen space, it did not support reflections with game characters. Any surface that emits light can be considered an emissive surface, while the entire skydome acts as one massive area light. With regards to local light sources, if they had shadows, ambient occlusion, reflections, and global illumination, they were all traditional raster techniques that have been utilized for decades.

Cyberpunk 2077 now features a new ray tracing mode called Overdrive that greatly enhances the game's already-incredible visuals. This updated version of the game is the first modern AAA game to take full advantage of NVIDIA RTX SDKs including NVIDIA RTX Direct Illumination (RTXDI), the direct illumination SDK that features NRD and video real-time denoisers to deliver a unified renderer for all effects so they don't need to be done individually. RTXDI is specifically used for direct illumination to light objects, characters, and pedestrians. RTXDI enables shadows for all lights, and it allows for more accurate lights than before, including emissive surfaces such as street lights, LED billboards, TV screens, neon signs, and other lights, bathing all objects near them with accurate, colored lighting.

NVIDIA RTX Direct Illumination (RTXDI) greatly enhances the game by providing real lighting, shadows, and reflections.

Figure 16. *Cyberpunk 2077:* With NVIDIA RTX Direct Illumination (RTXDI)

Depending on the scene and the number of reflective surfaces, in terms of rays per pixel that are generated, the modes cast 8 rays per pixel for Ultra mode, 10 rays per pixel for Psycho mode, and 22 rays per pixel for Overdrive mode. Approximately 1-2 shadow rays are cast per pixel to the most important lights in a scene, allowing for scenes with unlimited shadow-casting and dynamic lights in real time, without worrying about performance or resource constraints. Now, there's no limit on the number of lights that can be used or with emissive surfaces, which are very prevalent in the game, and there's no need for ambient occlusion because it essentially comes for free with the unified renderer.

Traditionally with prior generation GPUs, the performance hit with advanced ray tracing techniques like path tracing was so high it was only feasible to integrate it into classic games, like *Quake 2* and *Minecraft*. At the time of the release of ray tracing with *Cyberpunk 2077*, it was not possible to play with ray tracing enabled with Ultra settings at the native 4k resolution. It was only possible with DLSS. With the introduction of Ada, games as graphically advanced as *Cyberpunk 2077* can be run with advanced ray tracing at frame rates as high as 120 FPS at 4K with DLSS 3. This is due to the support of several NVIDIA technologies that greatly accelerate and improve the quality of advanced ray tracing workloads on GeForce RTX 40 Series graphics cards and is offered with the new Ray Tracing: Overdrive Mode. These technologies include NVIDIA Real-Time Denoisers (NRD), Opacity Micromaps (OMMs), and Shader Execution Reordering (SER).

Denoising is a technique to reduce the number of rays that need to be cast per pixel, smoothing out the results of ray tracing and RTXDI to produce an accurate result. NRD is a spatio-temporal, API-agnostic denoising library that's designed to work with low ray-per-pixel signals, and is used to deliver real-time performance results with improved quality that are comparable to ground-

truth images. Compared to previous-gen denoisers, NRD improves quality and ensures the computationally intensive ray-traced output is noise-free, without performance tradeoffs.

OMMs accelerate ray tracing workloads by encoding the surface opacity directly onto the geometry, drastically reducing expensive opacity evaluation during ray traversal, and enabling higher quality acceleration structures to be constructed. With OMMs, developers can very compactly describe irregularly shaped or translucent objects, like ferns or fences, and directly ray trace them with the Ada RT Core more efficiently. Although they don't make any difference visually, gamers will get a performance boost. The new Opacity Micromap Engine speeds up ray tracing of alpha-tested geometry by a factor of 2x, by significantly reducing shader-based alpha computations. On GeForce RTX 40 Series graphics cards, the Opacity Micromap format is directly decodable by ray tracing hardware, improving performance even further.

Raw ray tracing horsepower is not enough to ensure high frame rates with advanced ray-traced content, as these workloads can be bottlenecked by a number of factors. In particular, divergent ray-traced shaders are increasingly becoming a limiter, for example when executing multi-bounce, stochastic path tracing algorithms, or ray tracing complex materials. Divergence takes two forms: execution divergence where different threads execute different shaders or code paths within a shader, and data divergence, where threads access memory resources that are hard to coalesce or cache. Ada includes a new technology designed to enhance the efficiency of RT shader execution by tackling the divergence problem. SER is a new scheduling system that reorders and parallelizes the execution of threads that trace rays, without compromising image quality and it does this work on-the-fly for better execution and data locality. Years of research and development have been invested in SER in order to minimize overheads and maximize its effectiveness. The Ada hardware architecture was designed with SER in mind and includes optimizations to the memory system specifically targeted at efficient thread reordering.

The largest performance increase for *Cyberpunk 2077* is provided by DLSS 3.0, which takes DLSS to the next level with a new revolutionary capability made possible with deep learning. DLSS 3.0 includes new Optical Multi Frame Generation technology which uses deep learning, the Ada Optical Flow Accelerator, and the same type of game motion vectors DLSS 2.0 uses to create an AI-generated frame for each game-rendered frame. When the game renders a new frame, DLSS Frame Generation evaluates that new frame and the prior frame to discover how the scene is changing. DLSS Frame Generation essentially generates entirely new, extra rendered frames, boosting game performance for extra smooth gameplay by up to 2x.

Path Tracing SDK

A new NVIDIA RTX™ Path Tracing Software Development Kit (SDK) is also being released that merges years of best practices for real-time ray tracing and neural graphics development to build a real-time path tracer. Whether rasterizing, tracing rays, or both, this reference path tracer will ensure the photorealistic lighting produced is true to life in real time to deliver ground truth lighting in real time. It includes a highly customizable toolkit that provides the flexibility to use individual components as needed or use the entire path tracer and integrate it directly into a game engine. This path tracer leverages new hardware in NVIDIA GeForce RTX® 40 Series GPUs, enabling the highest-quality rendering modes with uncompromised performance.

Conclusion

RTX Remix was built to enable creators and game modders with an easy way to make their own RTX remasters of classic games. A powerful capture tool and the RTX Remix runtime application, when placed next to the game files, works as the games primary renderer rather than the renderer built into the game, intercepting draw calls (rendering instructions) to the GPU and converting everything in the game, including textures, geometry, and lights, to the Universal Scene Description (USD) format, an industry standard 3D file format. Because RTX Remix is built with NVIDIA Omniverse, creators can leverage the entire Omniverse platform to reimagine any asset. That includes the Omniverse-connected game development apps like *Adobe Substance 3D Painter, Adobe Photoshop, Autodesk Maya, 3ds Max, Blender, SideFX Houdini*, and Epic Games' *Unreal Engine*. When manipulating USD files in NVIDIA Omniverse, they're synced, so assets that are modified can be seen in RTX Remix.

The AI texture tools in RTX Remix can upres textures by up to 4x with a single click. Original legacy materials can be transformed into modern physically-based material using a neural network that has been trained on thousands of images of game textures. The AI can look at single color textures that make up objects and understand what material the object is supposed to be made of. The AI can then construct roughness maps to model that material accurately so that ray-traced light can interact with surfaces realistically.

The RTX Remix application isn't just for tweaking material properties of game assets. It allows modders to view lighting and shadows in isolation through a white mode, and with little effort, create new lights that help bring realistic global illumination, shadows, ambient occlusion, and reflections to a game. Modders and gamers can customize their experience during gameplay from the overlay, adjusting lights dynamically and even tweaking volumetrics to make for a more immersive scene. Other effects include customized lighting intensity, fog, and post-processing effects like bloom and motion blur.

New breakthroughs in neural technologies like DLSS 3 use AI to massively boost performance by generating entirely new frames while maintaining great image quality and responsiveness. A new Opacity Micromap Engine in the Ada RT Core speeds up alpha traversal by 2x. With this new capability, developers can very quickly assign opacity values to irregularly shaped objects (like ferns and fences) or translucent items (like flames or smoke) allowing the Ada RT Core to directly alpha test this geometry instead of relying on the GPU's SM. In addition, Ada introduces the Shader Execution Reordering (SER) scheduling system. Shader Execution Reordering organizes and reorders workloads on the fly so they can be processed by the SM and RT Core more efficiently. Shader Execution Reordering is as big of an innovation for GPUs as out-of-order execution was for CPUs back in the 1990s, offering 2-3x speedups for some RT workloads.

The NVIDIA Ada Lovelace architecture delivers a quantum leap in GPU performance and capabilities, giving GeForce RTX 40 Series users the power to experience the next generation of fully ray-traced games. And with RTX Remix, the gaming community has a new powerful tool to help remaster classic games and possibly help build the next generation of games.

Notice - The information provided in this specification is believed to be accurate and reliable as of the date provided. However, NVIDIA Corporation (“NVIDIA”) does not give any representations or warranties, expressed or implied, as to the accuracy or completeness of such information. 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 publication supersedes and replaces all other specifications for the product that may have been previously supplied.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and other changes to this specification, at any time and/or to discontinue any product or service without notice. Customer should obtain the latest relevant specification 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. NVIDIA hereby expressly objects to applying any customer general terms and conditions with regard to the purchase of the NVIDIA product referenced in this specification.

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 these specifications will be suitable for any specified use without further testing or modification. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer’s sole responsibility to ensure the product is suitable and fit for the application planned by customer and to do 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 specification. NVIDIA does not accept any 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 specification, 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 specification. 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 specification is permissible only if reproduction is approved by NVIDIA in writing, is reproduced without alteration, and is accompanied by all associated conditions, limitations, and notices.

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. 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 NVIDIA terms and conditions of sale for the product.

Trademarks - NVIDIA, the NVIDIA logo, GeForce, and GeForce RTX 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.