

# NVIDIA® OptiX™ Ray Tracing SDK

## Release Notes

---

**Version 7.3.0**

**April 2021**

Welcome to the 7.3.0 release of the NVIDIA OptiX SDK. This release includes support for temporal denoising and faster ray tracing of curve primitives. The NVIDIA OptiX Demand Loading library is now fully asynchronous, with sparse texture tiles loaded in the background by multiple CPU threads in parallel with OptiX kernels executing on the GPU.

Upgrading to 7.3.0 may require source code changes, though they should be minimal. Applications compiled with the 7.3.0 SDK headers will require driver version 465 or later. Applications compiled with earlier SDK headers will continue to work, but recompiling with 7.3.0 will expose new features and may also improve performance.

### System Requirements

(for running binaries referencing NVIDIA OptiX)

#### *Graphics Hardware:*

- All NVIDIA GPUs of Compute Capability 5.0 (Maxwell) or higher are supported.

#### *Graphics Driver:*

- NVIDIA OptiX 7.3.0 requires that you install a r465+ driver.
- Windows 8.1/10 64-bit; Linux RHEL 4.8+ or Ubuntu 10.10+ 64-bit

#### *CUDA Toolkit*

- It is **not** required to have any CUDA toolkit installed to be able to run NVIDIA OptiX-based applications.

### Development Environment Requirements

(for compiling with NVIDIA OptiX)

- CUDA Toolkit 11.1

This release has been tested with PTX generated from CUDA Toolkit 11.1. Other toolkit versions should also work, but 11.1 is recommended.

Version 11.1 of the CUDA toolkit introduces CUDA sparse textures, which are used in the NVIDIA OptiX Demand Loading library. This library is built only when the NVIDIA OptiX SDK is compiled with CUDA 11.1 or later.

- C/C++ Compiler

A compiler compatible with the used CUDA Toolkit is required. Please see the CUDA Toolkit documentation for more information on supported compilers.

## What's new in this version

- Denoiser for temporal images and API simplifications
  - Temporal denoising is now supported with a new built-in model kind. In this mode a sequence of images can be denoised. The AI network was trained to reduce flickering for camera or geometry animations. It requires the denoised beauty image from the previous frame as input as well as flow (motion) vectors. Currently temporal denoising is not supported for AOVs.
    - Enable by using the new `OPTIX_DENOISER_MODEL_KIND_TEMPORAL` enum value from `OptixDenoiserModelKind`.
  - `optixDenoiserCreate` and `optixDenoiserSetModel` have been merged. `optixDenoiserCreate` now uses an enum to select the built-in model kind and `optixDenoiserCreateWithUserModel` that takes a user model.
  - `optixDenoiserInvoke`'s input has changed to specify guide layers specified explicitly in `OptixDenoiserGuideLayer` (e.g. albedo, normal or pixel flow layers) and separately from beauty layers which are now specified in `OptixDenoiserLayer`. `optixUtilDenoiserInvokeTiled` has similarly been updated to the new interface.
  - `OptixDenoiserOptions` changed to specify whether albedo and normal guide layers will be provided during `optixDenoiserInvoke`
  - New `OptixPixelFormat`
    - `OPTIX_PIXEL_FORMAT_HALF2`
    - `OPTIX_PIXEL_FORMAT_FLOAT2`
- Demand Loading
  - The `launchPrepare` and `processRequests` methods are now asynchronous, taking a CUDA stream argument on which operations are enqueued.
  - Requests for sparse texture tiles are now processed in the background by multiple CPU threads. This greatly improves concurrency, keeping the CPU busy with I/O and texture decompression while OptiX kernels perform rendering work on GPU.
  - Multiple streams are now supported. A degree of latency hiding can be accomplished by rendering a framebuffer as multiple tiles in round-robin fashion: while the texture data for one framebuffer tile is loading, work can proceed on subsequent tiles, returning to the first when its data is ready.
  - NVIDIA OptiX now provides the following texture footprint functions, which are used by the NVIDIA OptiX Demand Loading library to quickly determine which sparse texture tiles are required. These functions are hardware-accelerated on Turing and Ampere GPUs, with software emulation on older architectures.
    - `optixTexFootprint2D`
    - `optixTexFootprint2DLod`
    - `optixTexFootprint2DGrad`
- Improved curve intersectors
  - There is a new faster intersector for cubic and quadratic curves (based on Reshetov's [Phantom intersector](#)).

- o There is a new intersector for piecewise linear curves that is slightly faster and higher precision.
- o The new curve intersector cull backfaces, treating the swept curve primitives as hollow. Rays originating inside the primitive will now exit the primitive.
- o Caveat: Internal rays can still hit internal end caps (invisible end caps between segments of a strand).
- o A bug in compaction of acceleration structures with curves has been fixed.
- New SDK samples:
  - o `optixDemandLoadSimple` - Demonstrates a simple use of the Demand Loading library.
  - o `optixModuleCreateAbort` - Compiles modules in separate processes, which can be interrupted.
  - o `optixMotionGeometry` - Demonstrates motion blur for vertex positions, SRT transforms, matrix transforms, and combined motion.
  - o `optixVolumeViewer` - Demonstrates incorporating OpenVDB volumes in an OptiX render, by using the open source [NanoVDB](#) library.
- Improved SDK samples:
  - o `optixDenoiser`
    - New mode for temporal denoising
    - Tiling operation now supported.
  - o `optixDemandTexture` - Now demonstrates the use of multiple streams for demand loading.
- `OptixPipelineCompileOptions` struct has changed. Please make sure to zero initialize this struct to avoid compilation errors.
  - o `OptixPipelineCompileOptions pipelineCompileOptions = {};`
- Validation mode now checks the stream state before executing API functions that take a stream.
- New device function `optixGetInstanceTraversableFromIAS`
  - o Return the traversable handle of a given instance in an Instance Acceleration Structure (IAS). This handle is not directly traversable, but can be used to query additional information.
- New OptixBuildFlag `OPTIX_BUILD_FLAG_ALLOW_RANDOM_INSTANCE_ACCESS`. This flag is required to call `optixGetInstanceTraversableFromIAS` and `optixGetTriangleVertexData`.
- New device function `optixGetInstanceChildFromHandle`
  - o Returns child traversable handle from an `OptixInstance` traversable
- New exception codes that can be triggered by `optixGetTriangleVertexData` and `optixGetInstanceTraversableFromIAS` when exceptions are enabled.
  - o `OPTIX_EXCEPTION_CODE_INVALID_VALUE_ARGUMENT_0`
    - The value passed in `ias` to `optixGetInstanceTraversableFromIAS` is not a valid Instance AS.
    - The value passed in `gas` to `optixGetTriangleVertexData` is not a valid Geometry AS.
  - o `OPTIX_EXCEPTION_CODE_INVALID_VALUE_ARGUMENT_1`
    - The `index` passed is out of range.
  - o `OPTIX_EXCEPTION_CODE_INVALID_VALUE_ARGUMENT_2`
    - The `sbtGASIndex` passed to `optixGetTriangleVertexData` is out of range.
  - o `OPTIX_EXCEPTION_CODE_UNSUPPORTED_DATA_ACCESS`

- `optixGetTriangleVertexData` or `optixGetInstanceTraversableFromIAS` was called on an acceleration structure built without `OPTIX_BUILD_FLAG_ALLOW_RANDOM_VERTEX_ACCESS` set or an acceleration structure built with motion used in a pipeline without motion enabled.
- If the OptiX compile cache is corrupted, OptiX will now attempt to delete and reinitialize the cache.
- Added `OPTIX_CACHE_MAXSIZE` environment variable to control the size of the disk cache.
- AS builds have been optimized for faster trace times.
- PTX must be recompiled with the new SDK headers if the host uses the new SDK.
- NVRTC is no longer enabled by default when compiling the SDK samples. It can be enabled by setting the CMake variable `CUDA_NVRTC_ENABLED`.
- Bug fixes:
  - Fixed bug to allow `optix_stack_size.h` to be able to be included in more than one compilation unit.
  - Fixed bug with `surf2Dwrite` calls having no effect.
  - Fixed bug with some denormal floats being treated as zero.
  - Fixed some bugs that would prevent correct line information from being used in profiling and debugging.

## Known Issues

1. Pixel formats `OPTIX_PIXEL_FORMAT_UCHAR3` and `OPTIX_PIXEL_FORMAT_UCHAR4` are not supported by the Denoiser.
2. Concurrent launches from the same pipeline will serialize automatically on the device.
3. `OPTIX_COMPILE_DEBUG_LEVEL_FULL` does not currently generate debug information necessary for `cuda-gdb` or `Nsight Compute VSE`.
4. The demand loading library may not work correctly with NVRTC enabled.

## What's New in 7.2.0

- Specialization is a powerful new feature that allows renderers to maintain generality while increasing performance on specific use cases. A single version of the PTX can be supplied to OptiX and specialized to toggle specific features on and off. The OptiX compiler is leveraged to fold constant values and elide complex code that is not required by a particular scene setup. Specialized values are supplied during module creation with `OptixModuleCompileOptions::boundValues`. See the Programming Guide section 6.3.1, "Parameter specialization", and the `optixBoundValues` sample.
- Demand loading source library
  - Enables textures to be loaded on demand, which greatly reduces memory requirements, start-up time, and disk I/O compared to preloading textures.
  - Requires the CUDA 11.1 toolkit.
  - See section 14 "Demand-loaded sparse textures" of the NVIDIA OptiX Programming Guide for a detailed technical introduction.
  - The `optixDemandTexture` sample demonstrates how to use this library.
  - This library supersedes the lower-level `optixPaging` library. The `optixDemandPaging` sample shows how to use `optixPaging` directly.
  - Known issue: wrap mode and mirror mode are not yet supported for CUDA sparse textures.

- There is a new mode in the denoiser that uses a neural network to predict a filter kernel instead of the final image. The filter weights can be applied to multiple layers or AOVs with just an incremental cost.
  - To select this mode, use `OptixDenoiserModelKind::OPTIX_DENOISER_MODEL_KIND_AOV`
  - Additional quality may be achieved by computing the average color using a new API function `optixDenoiserComputeAverageColor` and supplying the value to the denoiser using `OptixDenoiserParams::hdrAverageColor`
- To aid in debugging there is a new validation mode that runs additional checks during runtime and enables all debug exceptions.
  - Enable validation mode with: `OptixDeviceContextOptions::validationMode`
  - New error code when validation catches an error: `OPTIX_ERROR_VALIDATION_FAILURE`
  - APIs that take `CUstream` arguments are synchronized on the stream to check for errors before proceeding.
  - `optixLaunch` will synchronize after the launch and report errors
  - If any OptiX debug exceptions were thrown during launch, a CUDA launch error is triggered to prevent proceeding.
  - Validation mode reduces performance. Remember to turn it off.
- New debug exceptions
  - `OPTIX_EXCEPTION_CODE_CALLABLE_INVALID_SBT`
    - The callable program SBT record index was out of bounds
  - `OPTIX_EXCEPTION_CODE_CALLABLE_NO_DC_SBT_RECORD`
    - The callable program SBT record does not contain a direct callable program
  - `OPTIX_EXCEPTION_CODE_CALLABLE_NO_CC_SBT_RECORD`
    - The callable program SBT record does not contain a continuation callable program
- When linking, unresolved and multiply defined symbols will produce detailed error messages in the logs.
- Instance bounds are now computed automatically, even with motion and motion transforms. Instance acceleration structure build inputs no longer have the optional `OptixBuildInputInstanceArray::aabbs`.
  - Note OptiX will compute AABBs for all applications running against driver 455+ regardless of the version of the SDK it was built against. OptiX will ignore the AABBs supplied by applications compiled with earlier SDKs.
- You can now unload the NVIDIA OptiX driver DLL or DSO. See `optixUninitWithHandle()` in the `optix_stubs.h` header file.
  - If other threads in the process have a handle to the DLL it will continue to be loaded until the last handle is released.
  - There is a new error code when `optixUninitWithHandle` fails: `OPTIX_ERROR_LIBRARY_UNLOAD_FAILURE`
- Fixed a bug where PTX compiled with `-lineinfo` in CUDA 11 could cause errors when loading into OptiX.
- Fixed bugs related to large numbers of curves.

- Optimized traversal when only triangle geometry is enabled with single level instancing, that is, `OptixPipelineCompileOptions::usesPrimitiveTypeFlags` equals exactly `OPTIX_PRIMITIVE_TYPE_FLAGS_TRIANGLE` and `OptixPipelineCompileOptions::traversableGraphFlags` equals exactly `OPTIX_TRAVERSABLE_GRAPH_FLAG_ALLOW_SINGLE_LEVEL_INSTANCING`
- SDK
  - By default the SDK will target PTX compilation for SM 60 (Volta+). Set CMake variables `CUDA_NVCC_FLAGS` and `CUDA_NVRTC_FLAGS` to use an older SM version if desired.
  - New samples
    - `optixBoundValues`
    - `optixDemandTexture`
    - `optixDenoiser`
    - `optixDynamicGeometry`

### What's New in 7.1.0

- Added curves as a new type of geometric primitive. Curves are swept surfaces used to represent long thin strands, such as for hair, fur, or cloth fibers. Linear, quadratic, and cubic B-spline bases are supported. Motion blur is supported.
  - Added new GAS build input type for curves. See the NVIDIA OptiX Programming Guide section 5.2, “Curve build inputs”.
  - Hit programs can access the curve parameter value at the hit point, and the curve’s geometric data which is stored in the acceleration structure (GAS). Utility code is provided in the SDK (`cuda/curve.h`) to compute the curve surface position, tangent, and normal.
  - Each SBT program group for curves requires a built-in curves intersection program, returned by the new host function `optixBuiltinISModuleGet`. Motion blur for curves is enabled here.
  - Pipelines can now indicate which primitive types they support via `OptixPipelineCompileOptions::usesPrimitiveTypeFlags`. If your scene contains curves, they must be enabled here. If your scene geometry is all triangles (no curves and no custom primitives), set these flags to enable only triangles, for optimal performance.
  - In hit programs, the recommended method to discriminate among hit types is to use the new methods `optixGetPrimitiveType`, `optixIsFrontFaceHit`, and `optixIsBackFaceHit`. The old methods, e.g. `optixIsTriangleFrontFaceHit`, still work.
  - See the NVIDIA OptiX Programming Guide section 8, “Curves”, for additional information.
- The denoiser has several improvements for quality and performance in addition to some API changes.
  - Added support for `OptixDenoiserInputKind::OPTIX_DENOISER_INPUT_RGB_ALBEDO_NORMAL`.
  - Added support for tiling. See `optix_denoiser_tiling.h` in the SDK for helpers to use tiling.
  - `OptixDenoiserOptions::pixelFormat` has been removed, because it is no longer needed.

- `OptixDenoiserSizes::minimumScratchSizeInBytes` and `recommendedScratchSizeInBytes` have been removed and replaced with `withOverlapScratchSizeInBytes` and `withoutOverlapScratchSizeInBytes`.
- Increase instancing limits. Query limit with `OPTIX_DEVICE_PROPERTY_LIMIT_MAX_INSTANCES_PER_IAS`. The limit has been increased to  $2^{28}$  instead of  $2^{24}$ . `OptixInstance::sbtOffset` now also supports values up to  $2^{28}$ .
- Removed `OptixPipelineLinkOptions::overrideUsesMotionBlur`. This option is no longer needed.
- Added `OptixTransformFormat` used in `OptixBuildInputTriangleArray::transformFormat`. Allows to specify the format of `OptixBuildInputTriangleArray::preTransform`. Value should be `OPTIX_TRANSFORM_FORMAT_MATRIX_FLOAT12` when `preTransform` will be supplied to the build and `OPTIX_TRANSFORM_FORMAT_NONE` when `preTransform` is unused. A `nullptr` can now be used for `optixAccelComputeMemoryUsage` instead of supplying a valid or dummy pointer.
- Several new device exceptions were added to catch common errors. They are active when debug exceptions are enabled in `OptixPipelineCompileOptions::exceptionFlags`.
  - Invalid ray exception
    - Checks for NaN and Inf in the ray parameters passed `optixTrace`
    - New exception code: `OPTIX_EXCEPTION_CODE_INVALID_RAY`
    - `optixGetExceptionInvalidRay` returns details of the exception in a struct called `OptixInvalidRayExceptionDetails`
  - Callable program parameter mismatch
    - Currently only checks the number of parameters and not the types, since PTX can lose type information.
    - New exception code: `OPTIX_EXCEPTION_CODE_CALLABLE_PARAMETER_MISMATCH`
    - `optixGetExceptionParameterMismatch` returns details of the exception in a struct called `OptixParameterMismatchExceptionDetails`
  - Ensure that the built-in intersection program assigned to the SBT matches the GAS
    - New exception code: `OPTIX_EXCEPTION_CODE_BUILTIN_IS_MISMATCH`
  - Ensure that when `optixTrace` is called with a single level GAS as the trace target that `OptixPipelineCompileOptions::traversableGraphFlags` is set to either `OPTIX_TRAVERSABLE_GRAPH_FLAG_ALLOW_SINGLE_GAS` or `OPTIX_TRAVERSABLE_GRAPH_FLAG_ALLOW_ANY`.
    - new exception code: `OPTIX_EXCEPTION_CODE_UNSUPPORTED_SINGLE_LEVEL_GAS`
  - The shader encountered an unsupported primitive type when calling `optixGetLinearCurveVertexData`, `optixGetQuadraticBSplineVertexData`, `optixGetCubicBSplineVertexData`, or `optixGetPrimitiveType`. Supported primitive types are set with `OptixPipelineCompileOptions::usesPrimitiveTypeFlags`.
    - New exception code: `OPTIX_EXCEPTION_CODE_UNSUPPORTED_PRIMITIVE_TYPE`

- When `OPTIX_EXCEPTION_CODE_TRAVERSAL_INVALID_HIT_SBT` is thrown, `optixGetPrimitiveIndex` is no longer supported. `optixGetSbtGASIndex` should be used instead. See `optixDumpExceptionDetails` for details.
- `optixThrowException` is now elided when user exceptions are disabled instead of producing an error.
- Added `optixGetExceptionLineInfo` device function accessible in exception programs.
  - Returns a string that includes information about the source location that caused the current exception. Only supported for certain exceptions, and requires line information in the PTX (`--lineinfo`) and a debug level that supports line info (`OPTIX_COMPILE_DEBUG_LEVEL_LINEINFO` and `OPTIX_COMPILE_DEBUG_LEVEL_FULL`).
- Change `OptixCompileOptimizationLevel` and `OptixCompileDebugLevel` enum values.
  - Added a `DEFAULT` value of 0 (used when zero initializing the structs).
  - `OptixCompileOptimizationLevel::OPTIX_COMPILE_OPTIMIZATION_DEFAULT` continues to mean `OPTIMIZATION_LEVEL_3`.
  - `OptixCompileDebugLevel::OPTIX_COMPILE_DEBUG_LEVEL_DEFAULT` (new) adds `lineinfo` (same as `OPTIX_COMPILE_DEBUG_LEVEL_LINEINFO`).
- Fixed support in `optix` headers for cuda runtime compilation using `nVRTC`.
- Enable compaction support for acceleration structures on non-RTX GPUs.
- OptiX will attempt to reset a corrupted compile disk cache.
- Added `OptixIndicesFormat::OPTIX_INDICES_FORMAT_NONE` which must be used when vertex indices are not present (e.g. triangle soup).
- Added `OptixVertexFormat::OPTIX_VERTEX_FORMAT_NONE` for use when initializing `OptixBuildInputTriangleArray::vertexFormat`. The value, `vertexFormat`, must be set to something other than `OPTIX_VERTEX_FORMAT_NONE` before `OptixBuildInputTriangleArray` can be used.
- Additional checks are now in place when calling `optixPipelineSetStackSize`. If the `maxTraversableGraphDepth` is greater than maximum supported by the device an error is generated. `maxTraversableGraphDepth` also must be greater than zero. `maxTraversableGraphDepth` must also be compatible with `OptixPipelineCompileOptions::traversableGraphFlags`.
- When used with Nsight Compute, `optixLaunch` and `optixAccelBuild` are now marked in the timeline.
- Very large AABBs are now clamped to +/- 2^40 for non-motion acceleration structures.
- `optixGetTriangleVertexData` support has been extended to all supported GPUs.
- `optixGetTriangleVertexData` support has been fixed for large meshes.
- Fixed a crash on non-RTX GPUs when rendering with a refit IAS.
- New SDK samples.
  - `optixCallablePrograms` - Demonstration of *callable program* usage and SBT setup.
  - `optixCurves` - Minimal curve API usage example.



- optixDenoiser - Demonstration of NVIDIA OptiX AI Denoiser usage.
- optixDynamicGeometry - Shows typical setup for changing geometry between frames.
- optixHair - More complicated curve API example.
- optixNVLink - Demonstrates NVLink usage within an OptiX application.

### **What's New in 7.0**

- 7.0 introduced the NVIDIA OptiX 7 API, a new low-level CUDA-centric API giving application developers direct control of memory, compilation, and launches while maintaining the programming model and shader types from previous versions of OptiX.
- Minimal host state is maintained. Scene graphs, materials, etc., are managed by the application rather than by OptiX.
- GPU memory is managed by the application using CUDA. (No OptiX buffers or variables)
- GPU launches are explicit and asynchronous using CUDA streams.
- Shader compilation is explicit. (Similar to DXR or Vulkan)
- All host functions are thread-safe.
- Source code for demand loading library is included and designed for direct inclusion in production applications.
- Multi-GPU operation is managed by the application.