



# **NVENC Application Note**

***Release 13.0***

**NVIDIA Corporation**

**Dec 16, 2025**



# Contents

|          |                               |           |
|----------|-------------------------------|-----------|
| <b>1</b> | <b>Introduction</b>           | <b>1</b>  |
| <b>2</b> | <b>NVENC Capabilities</b>     | <b>3</b>  |
| <b>3</b> | <b>NVENC Licensing Policy</b> | <b>7</b>  |
| <b>4</b> | <b>NVENC Performance</b>      | <b>9</b>  |
| <b>5</b> | <b>Programming NVENC</b>      | <b>13</b> |
| <b>6</b> | <b>FFmpeg Support</b>         | <b>15</b> |



---

# Chapter 1. Introduction

NVIDIA GPUs - beginning with the Kepler generation - contain a hardware-based encoder (referred to as NVENC in this document) which provides fully accelerated hardware-based video encoding and is independent of graphics/CUDA cores. With end-to-end encoding offloaded to NVENC, the graphics/CUDA cores and the CPU cores are free for other operations. For example, in a game recording scenario, offloading the encoding to NVENC makes the graphics engine fully available for game rendering. In the video transcoding use-case, video encoding/decoding can happen on NVENC/NVDEC in parallel with other video post-/pre-processing on CUDA cores.

The hardware capabilities available in NVENC are exposed through APIs referred to as NVENCODE APIs in the document. This document provides information about the capabilities of the hardware encoder and features exposed through NVENCODE APIs.



---

## Chapter 2. NVENC Capabilities

NVENC can perform end-to-end encoding for H.264, HEVC 8-bit, HEVC 10-bit, AV1 8-bit and AV1 10-bit. This includes motion estimation and mode decision, motion compensation and residual coding, and entropy coding. It can also be used to generate motion vectors between two frames, which are useful for applications such as depth estimation, frame interpolation, encoding using other codecs not supported by NVENC, or hybrid encoding wherein motion estimation is performed by NVENC and the rest of the encoding is handled elsewhere in the system. These operations are hardware accelerated by a dedicated block on GPU silicon die. NVENCODE APIs provide the necessary knobs to utilize the hardware encoding capabilities.

Below Tables summarizes the capabilities of the NVENC hardware exposed through NVENCODE APIs.

### NVENC H264 Hardware Capabilities

Table 1: NVENC H264 Hardware Capabilities

| Hardware Features   | 1 <sup>st</sup> Gen Maxwell GPUs | 2 <sup>nd</sup> Gen Maxwell GPUs | Pascal GPUs | Volta and TU117 GPUs | Ampere and TU GPUs | Ada GPUs | Jetson Thor | Blackwell GPUs |
|---|----------------------------------|----------------------------------|-------------|----------------------|--------------------|----------|-------------|----------------|
| Baseline main and high profile: Capability to encode YUV 420 sequence | Y                                | Y                                | Y           | Y                    | Y                  | Y        | Y           | Y              |
| High10 profile: Support for encoding 10-bit content.                  | N                                | N                                | N           | N                    | N                  | N        | N           | Y              |
| 4:2:2 encoding  | N                                | N                                | N           | N                    | N                  | N        | N           | Y              |
| 4:4:4 encoding (only CAVLC)   | Y                                | Y                                | Y           | Y                    | Y                  | Y        | Y           | Y              |

continues on next page

Table 1 – continued from previous page

| Hardware Features  | 1 <sup>st</sup> Gen Maxwell GPUs | 2 <sup>nd</sup> Gen Maxwell GPUs | Pascal GPUs | Volta and TU117 GPUs | Ampere and TU GPUs | Ada GPUs | Jetson Thor | Blackwell GPUs |
|--|----------------------------------|----------------------------------|-------------|----------------------|--------------------|----------|-------------|----------------|
| Lossless encoding  | Y                                | Y                                | Y           | Y                    | Y                  | Y        | Y           | Y              |
| Motion estimation (ME) only mode: Capability to provide macro-block level motion vectors and intra or inter modes. | Y                                | Y                                | Y           | Y                    | Y                  | Y        | Y           | Y              |
| Field encoding   | Y                                | Y                                | Y           | Y                    | N                  | N        | N           | Y              |
| Weighted prediction  | N                                | N                                | Y           | Y                    | Y                  | Y        | Y           | Y              |
| Encoding support ARGB content  | Y                                | Y                                | Y           | Y                    | Y                  | Y        | Y           | Y              |
| Multiple reference frames : Capability to use different reference frames   | N                                | N                                | N           | N                    | Y                  | Y        | Y           | Y              |

► **Y:** Supported, **N:** Not supported

► \*Supported in select Pascal generation GPUs

### NVENC HEVC Hardware Capabilities

Table 2: NVENC HEVC Hardware Capabilities

| Hardware Features                                      | 1 <sup>st</sup> Gen Maxwell GPUs | 2 <sup>nd</sup> Gen Maxwell GPUs | Pascal GPUs | Volta and TU117 GPUs | Ampere and TU GPUs | Ada GPUs | Jetson Thor | Blackwell GPUs |
|--|----------------------------------|----------------------------------|-------------|----------------------|--------------------|----------|-------------|----------------|
| Main profile: Capability to encode YUV 4:2:0 sequence. | N                                | Y                                | Y           | Y                    | Y                  | Y        | Y           | Y              |

continues on next page



Table 2 – continued from previous page

| Hardware Features   | 1 <sup>st</sup> Gen<br>Maxwell<br>GPUs | 2 <sup>nd</sup> Gen<br>Maxwell<br>GPUs | Pascal<br>GPUs | Volta<br>and<br>TU117<br>GPUs | Am-<br>pere<br>and TU<br>GPUs | Ada<br>GPUs | Jetson<br>Thor | Black-<br>well<br>GPUs |
|---|--|--|----------------|-------------------------------|-------------------------------|-------------|----------------|------------------------|
| Lossless encoding   | N                                      | N                                      | Y              | Y                             | Y                             | Y           | Y              | Y                      |
| Main10 profile: Support for encoding 10-bit content.  | N                                      | N                                      | Y              | Y                             | Y                             | Y           | Y              | Y                      |
| 4:2:2 encoding  | N                                      | N                                      | N              | N                             | N                             | N           | N              | Y                      |
| 4:4:4 encoding  | N                                      | N                                      | Y              | Y                             | Y                             | Y           | Y              | Y                      |
| Motion estimation (ME) only mode: Capability to provide CTB level motion vectors and intra/inter modes. | N                                      | N                                      | Y              | Y                             | Y                             | Y           | Y              | Y                      |
| 8K encoding: Support for encoding 8192 x 8192 Content.  | N                                      | N                                      | Y*             | Y                             | Y                             | Y           | Y              | Y                      |
| Weighted prediction   | N                                      | N                                      | Y              | Y                             | Y                             | Y           | Y              | Y                      |
| Sample adaptive offset (SAO): Improves encoded video quality.   | N                                      | N                                      | Y              | Y                             | Y                             | Y           | Y              | Y                      |
| B frame: Improves encoded quality.  | N                                      | N                                      | N              | N                             | Y                             | Y           | Y              | Y                      |
| Multiple reference frames : Capability to use different reference frames                                | N                                      | N                                      | N              | N                             | Y                             | Y           | Y              | Y                      |

continues on next page

Table 2 – continued from previous page

| Hardware Features | 1 <sup>st</sup> Gen<br>Maxwell<br>GPUs | 2 <sup>nd</sup> Gen<br>Maxwell<br>GPUs | Pascal<br>GPUs | Volta<br>and<br>TU117<br>GPUs | Am-<br>pere<br>and TU<br>GPUs | Ada<br>GPUs | Jetson<br>Thor | Black-<br>well<br>GPUs |
|-------------------|--|--|----------------|-------------------------------|-------------------------------|-------------|----------------|------------------------|
|                   |  |  |                |                               |                               |             |                |                        |

- **Y:** Supported, **N:** Not supported
- \*Supported in select Pascal generation GPUs

### NVENC AV1 Hardware Capabilities

Table 3: NVENC AV1 Hardware Capabilities

| Hardware Features  | 1 <sup>st</sup> Gen<br>Maxwell<br>GPUs | 2 <sup>nd</sup> Gen<br>Maxwell<br>GPUs | Pascal<br>GPUs | Volta<br>and<br>TU117<br>GPUs | Am-<br>pere<br>and TU<br>GPUs | Ada<br>GPUs | Jetson<br>Thor | Black-<br>well<br>GPUs |
|--|--|--|----------------|-------------------------------|-------------------------------|-------------|----------------|------------------------|
| Main profile: Capability to encode YUV 420 8-bit and 10-bit content up to 8192 x 8192 content. | N                                      | N                                      | N              | N                             | N                             | Y           | N              | Y                      |

- **Y:** Supported, **N:** Not supported
- \*Supported in select Pascal generation GPUs

---

## Chapter 3. NVENC Licensing Policy

As far as NVENC hardware encoding is concerned, NVIDIA GPUs are classified into two categories: “qualified” and “non-qualified”. On qualified GPUs, the number of concurrent encode sessions is limited by available system resources (encoder capacity, system memory, video memory etc.). On non-qualified GPUs, the number of concurrent encode sessions is limited to 8 per system. This limit of 8 concurrent sessions per system applies to the combined number of encoding sessions executed on all non-qualified cards present in the system.

For a complete list of qualified and non-qualified GPUs, refer to <https://developer.nvidia.com/nvidia-video-codec-sdk..>

For example, on a system with one Quadro RTX4000 card (which is a qualified GPU) and three GeForce cards (which are non-qualified GPUs), the application can run  $N$  simultaneous encode sessions on Quadro RTX4000 card (where  $N$  is defined by the encoder/memory/hardware limitations) and 8 sessions on all the GeForce cards combined. Thus, the limit on the number of simultaneous encode sessions for such a system is  $N + 8$ .



---

## Chapter 4. NVENC Performance

With every generation of NVIDIA GPUs (Maxwell 1st/2nd gen, Pascal, Volta, Turing, Ampere and Ada), NVENC performance has increased steadily. [NVENC H264 Performance](#) provides *indicative*<sup>1</sup> NVENC performance on Pascal, Turing, and Ada GPUs for different presets and rate control modes (these two factors play a major role in determining the performance and quality). Note that performance numbers in [NVENC H264 Performance](#) for H264, [NVENC HEVC Performance](#) for HEVC, [NVENC AV1 Performance](#) for AV1 are measured on GeForce hardware with assumptions listed under the table. The performance varies across GPU classes (e.g. Quadro, Tesla), and scales (almost) linearly with the clock speeds for each hardware.

While first-generation Maxwell GPUs had one NVENC engine per chip, certain variants of the second-generation Maxwell, Pascal, Volta and Ada GPUs have two/three NVENC engines per chip. This increases the aggregate encoder performance of the GPU. NVIDIA driver takes care of load balancing among multiple NVENC engines on the chip, so that applications don't require any special code to take advantage of multiple encoders and automatically benefit from higher encoder capacity on higher-end GPU hardware. The encode performance listed in [NVENC H264 Performance](#) for H264, [NVENC HEVC Performance](#) for HEVC, [NVENC AV1 Performance](#) for AV1 is given *per NVENC engine*. Thus, if the GPU has 2 NVENCs (e.g. GP104, AD104), multiply the corresponding number in [NVENC H264 Performance](#) for H264, [NVENC HEVC Performance](#) for HEVC, [NVENC AV1 Performance](#) for AV1 by the number of NVENCs per chip to get aggregate maximum performance (applicable only when running multiple simultaneous encode sessions). Note that unless Split Frame Encoding is enabled, performance with single encoding session cannot exceed performance per NVENC, regardless of the number of NVENCs present on the GPU. Multi NVENC Split Frame Encoding is a feature introduced in SDK12.0 on Ada GPUs for HEVC and AV1. Refer to the NVENC Video Encoder API Programming Guide for more details on this feature.

NVENC hardware natively supports multiple hardware encoding contexts with negligible context-switching penalty. As a result, subject to the hardware performance limit and available memory, an application can encode multiple videos simultaneously. NVENC API exposes several presets, rate control modes and other parameters for programming the hardware. A combination of these parameters enables video encoding at varying quality and performance levels. In general, one can trade performance for quality and vice versa.

---

<sup>1</sup>

## NVENC H264 encoding performance in frames/second (fps)

Table 1: NVENC H264 Performance

| Preset | RC Mode | Tuning Info | Pascal | Turing | Ampere | Ada | Jetson Thor | Blackwell |
|--------|---------|-------------|--------|--------|--------|-----|-------------|-----------|
| P1     | CBR     | LL          | 667    | 855    | 868    | 910 | 724         | 977       |
| P1     | VBR     | HQ          | 692    | 833    | 846    | 885 | 713         | 948       |
| P3     | CBR     | LL          | 649    | 600    | 613    | 652 | 529         | 718       |
| P3     | VBR     | HQ          | 398    | 602    | 617    | 647 | 527         | 708       |
| P5     | CBR     | LL          | 363    | 271    | 273    | 291 | 236         | 323       |
| P5     | VBR     | HQ          | 327    | 264    | 266    | 283 | 230         | 317       |
| P7     | CBR     | LL          | 321    | 229    | 231    | 247 | 219         | 264       |
| P7     | VBR     | HQ          | 250    | 207    | 213    | 211 | 202         | 227       |

## NVENC HEVC encoding performance in frames/second (fps)

Table 2: NVENC HEVC Performance

| Preset | RC Mode | Tuning Info | Pascal | Turing | Ampere | Ada  | Jetson Thor | Blackwell |
|--------|---------|-------------|--------|--------|--------|------|-------------|-----------|
| P1     | CBR     | LL          | 539    | 932    | 943    | 1055 | 860         | 1134      |
| P1     | VBR     | HQ          | 506    | 920    | 939    | 1037 | 850         | 1119      |
| P3     | CBR     | LL          | 442    | 463    | 467    | 494  | 402         | 529       |
| P3     | VBR     | HQ          | 443    | 552    | 557    | 706  | 579         | 947       |
| P5     | CBR     | LL          | 370    | 305    | 307    | 343  | 279         | 506       |
| P5     | VBR     | HQ          | 371    | 334    | 335    | 411  | 336         | 521       |
| P7     | CBR     | LL          | 345    | 306    | 308    | 343  | 279         | 464       |
| P7     | VBR     | HQ          | 260    | 171    | 171    | 181  | 148         | 181       |

## NVENC AV1 encoding performance in frames/second (fps)

Table 3: NVENC AV1 Performance

| Preset | RC Mode | Tuning Info | Ada  | Blackwell |
|--------|---------|-------------|------|-----------|
| P1     | CBR     | LL          | 1090 | 1076      |
| P1     | VBR     | HQ          | 741  | 957       |
| P3     | CBR     | LL          | 774  | 798       |
| P3     | VBR     | HQ          | 549  | 678       |
| P5     | CBR     | LL          | 512  | 624       |
| P5     | VBR     | HQ          | 440  | 552       |
| P7     | CBR     | LL          | 356  | 395       |
| P7     | VBR     | HQ          | 323  | 401       |

- ▶ *Resolution/Input Format/Bit depth:* 1920 × 1080/YUV 4:2:0/8-bit
- ▶ Above measurements are made using the following GPUs: GTX 1060 for Pascal, RTX 8000 for Turing, RTX 3090 for Ampere, RTX 4090 for Ada, and NVIDIA® Jetson Thor™ for Thor GPUs. All measurements are done at the highest video clocks as reported by nvidia-smi (i.e. 1708 MHz, 1950 MHz, 1950 MHz, 2415 MHz for GTX 1060, RTX 8000, RTX 3090, and RTX 4090 respectively). The performance should scale according to the video clocks as reported by nvidia-smi for other GPUs of every individual family. Information on nvidia-smi can be found at <https://developer.nvidia.com/nvidia-system-management-interface>.
- ▶ H.264 and HEVC encoding fps for Volta GPU can be obtained by multiplying the Pascal fps in the above table by ratio of the clocks, as reported by nvidia-smi.
- ▶ *Software:* Windows 11, Video Codec SDK v13.0; Thor GPU performance is measured on NVIDIA® Jetson™ Linux
- ▶ CBR: Constant bitrate rate control mode, VBR: Variable bitrate rate control mode, LL : Low latency tuning info, HQ: High quality tuning info





---

## Chapter 5. Programming NVENC

The NVENCODE API provides access to the video encoding features of NVENC described in the previous chapters and provides control over encoding parameters. The NVENC hardware takes YUV/RGB as input and generates H.264/HEVC/AV1 compliant video bit streams.

Refer to the SDK release notes for information regarding the required driver version.

Refer to the documents and the sample applications included in the SDK package for details on how to program NVENC.



---

## Chapter 6. FFmpeg Support

FFmpeg is the most popular multimedia transcoding tool used extensively for video and audio transcoding.

The video hardware accelerators in NVIDIA GPUs can be effectively used with FFmpeg to significantly speed up the video decoding, encoding and end-to-end transcoding at very high performance. For more information on how to use NVENC or NVDEC with FFmpeg, please refer to the FFmpeg guide in the Video Codec SDK.

Note that FFmpeg is open-source project and its usage is governed by specific licenses and terms and conditions for FFmpeg.

### Notices

#### Notice

This document is provided for information purposes only and shall not be regarded as a warranty of a certain functionality, condition, or quality of a product. NVIDIA Corporation (“NVIDIA”) makes no representations or warranties, expressed or implied, as to the accuracy or completeness of the information contained in this document and assumes no responsibility for any errors contained herein. 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 document is not a commitment to develop, release, or deliver any Material (defined below), code, or functionality.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice.

Customer should obtain the latest relevant information 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 acknowledgment, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer (“Terms of Sale”). NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this document. No contractual obligations are formed either directly or indirectly by this document.

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 this document will be suitable for any specified use. Testing of all parameters of each product is not necessarily performed by NVIDIA.

It is customer's sole responsibility to evaluate and determine the applicability of any information contained in this document, ensure the product is suitable and fit for the application planned by customer, and perform 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 document. NVIDIA accepts no 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 document or (ii) customer product designs.

## Trademarks

NVIDIA, the NVIDIA logo, and cuBLAS, CUDA, CUDA Toolkit, cuDNN, DALI, DIGITS, DGX, DGX-1, DGX-2, DGX Station, DLProf, GPU, Jetson, Kepler, Maxwell, NCCL, Nsight Compute, Nsight Systems, NVcaffe, NVIDIA Deep Learning SDK, NVIDIA Developer Program, NVIDIA GPU Cloud, NVLink, NVSHMEM, PerfWorks, Pascal, SDK Manager, Tegra, TensorRT, TensorRT Inference Server, Tesla, TF-TRT, Triton Inference Server, Turing, and Volta are trademarks and/or registered trademarks of NVIDIA Corporation in the United States and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

## Copyright

©2022-2025, NVIDIA Corporation & affiliates. All rights reserved