

**GPU** TECHNOLOGY  
CONFERENCE

# HIGH-PERFORMANCE GPU VIDEO ENCODING

ABHIJIT PATAIT  
SR. MANAGER, NVIDIA

# AGENDA

- ▶ GPU Video Encoding Overview
- ▶ NVIDIA Video Encoding Capabilities
  - ▶ Kepler, Maxwell Gen 1, Maxwell Gen 2
- ▶ Software API
- ▶ Performance & Quality
- ▶ Roadmap

# WHY GPU VIDEO ENCODING?

# BENEFITS

- ▶ Low power
  - ▶ Fixed function hardware, free CPU
  - ▶ Reduced memory transfers
- ▶ Low latency
- ▶ High performance
- ▶ Higher density
- ▶ Scalability
  - ▶ Automatic benefit from improvements in hardware
- ▶ Ease of programming
  - ▶ Linux, Windows, C/C++, Application portability

# NVIDIA VIDEO ENCODER CAPABILITIES

# MAIN FEATURES

Feature	Benefits
H.264 base, main, high profiles	Wide range of use-cases
H.265/HEVC main profile	Lower bitrates at same quality
High performance (4K @ 60 fps)	“Blazing-speed” encoding
YUV 4:2:0 and 4:4:4 support	High quality encoding without chroma subsampling
QP maps	Customizable quality, region of interest encoding
4K encoding in hardware	High resolution encode
API - NV Encode SDK & GRID SDK	Flexible, Win/Linux, DirectX/CUDA
Independent of CUDA	Use CUDA and encode simultaneously

# FEATURE COMPARISON

Kepler	Maxwell Gen 1 (GM10x)	Maxwell Gen 2 (GM20x)
H.264 only	H.264 only	H.264 and <a href="#">HEVC/H.265</a>
Planar 4:4:4 & proprietary 4:4:4; <b>no lossless</b> encoding	Standard 4:4:4 and <a href="#">H.264 lossless</a> encoding	Standard 4:4:4 and <a href="#">H.264 lossless</a> encoding
~ <b>240 fps</b> 2-pass encoding @ 720p	~ <b>500 fps</b> 2-pass encoding @ 720p	~ <b>900 fps</b> 2-pass encoding @ 720p
GRID <a href="#">K340/K520</a> , <a href="#">K1/K2</a> , Quadro, Tesla <a href="#">K10/K20</a>	<b>Maxwell-based</b> GRID & Quadro products	TBA
GeForce – 2 full-speed encode sessions/system	GeForce – 2 full-speed encode sessions/system	GeForce – 2 full-speed encode sessions/system
NV Encode SDK 1.0-5.0 ( <b>Now</b> )	NV Encode SDK 4.0+ ( <b>Now</b> )	NV Encode SDK 5.0+ ( <b>Now</b> )
GRID SDK 1.x, 2.2, 2.3 ( <b>Now</b> )	GRID SDK 3.0+ ( <b>Now</b> )	<a href="#">In development</a>

# WHAT'S NEW - HARDWARE

## ▶ HEVC

- ▶ 8-bit encoding
- ▶ Main8 profile
- ▶ Optimized for low-latency applications (I and P frames)
- ▶ > 300 fps at very high quality 720p

## ▶ H.264

- ▶ Improved performance (~80% higher compared to 1st Gen Maxwell)
- ▶ 4:4:4 and lossless



# WHAT'S NEW - SOFTWARE

- ▶ NVENC SDK 5.0
- ▶ NVIDIA GPU driver 347.18 and above
- ▶ HEVC
  - ▶ Unified API for H.264 and HEVC
  - ▶ Linux & Windows
  - ▶ Intra refresh, ref-pic invalidation, etc. for H.264 and HEVC
- ▶ Support for all NVENC hardware up to GM20x
- ▶ Adaptive quantization
- ▶ Quality improvements
- ▶ All-new sample applications, including a performance application

# SOFTWARE API

# USING NVENC

## Direct Encode

### NVENC SDK

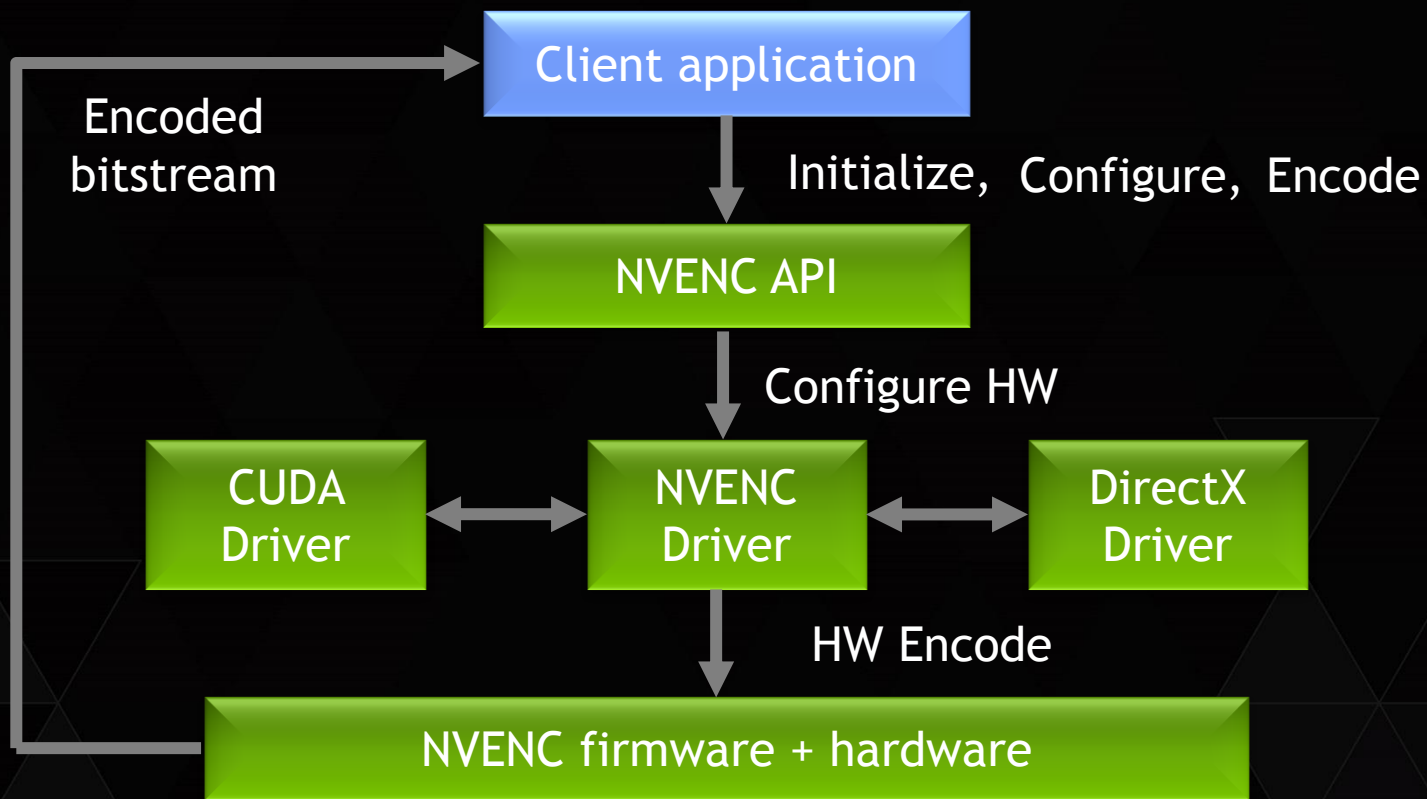
- No capture
- Transcoding
- Archiving
- Video editing
- CUDA pre-process + encoding
- Granular encoder settings
- D3D, CUDA interop

## Capture + Encode

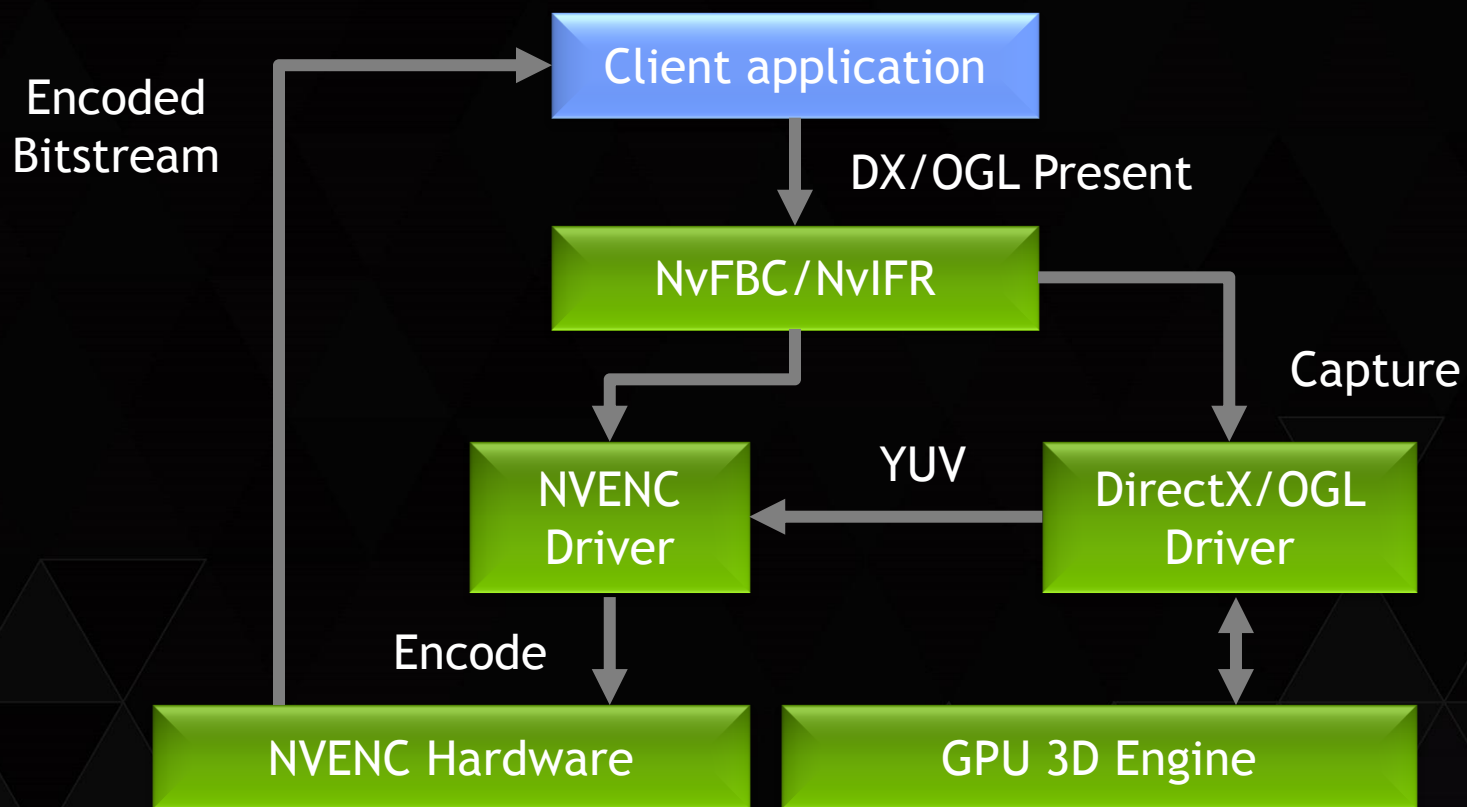
### GRID SDK

- Capture + encode
- Optimized for low-latency apps
- Capture + CUDA pre-process + encoding
- Encoder settings optimized for streaming
- D3D, CUDA interop

# DIRECT ENCODE (NVENC SDK)



# CAPTURE & ENCODE (GRID SDK)



# NVENC SDK (1/2)

- ▶ Available on NVIDIA developer zone
  - ▶ <https://developer.nvidia.com/nvidia-video-codec-sdk>
  - ▶ Current release: 5.0
- ▶ Interface header, documentation, sample application
- ▶ .dll/ .so included in the driver
- ▶ Unified API for Windows and Linux
- ▶ Works on x86/x64
- ▶ API's, presets, rate control modes for
  - ▶ Low-latency streaming
  - ▶ Transcoding
  - ▶ Video conferencing

# NVENC SDK (2/2)

- ▶ **Unified API for H.264 and HEVC**
- ▶ **Flexibility**
  - ▶ Dynamic resolution/bitrate change
  - ▶ Low-level encoder settings
  - ▶ Windows, Linux, DirectX, CUDA, OGL (via CUDA)
  - ▶ Works on GeForce (2 sessions/system)
- ▶ **Error concealment**
  - ▶ Reference picture invalidation
  - ▶ Intra-refresh
- ▶ **Greater flexibility for quality/performance trade-off**
- ▶ **Lossless encoding only in NVENC SDK**

# GRID SDK ENCODE

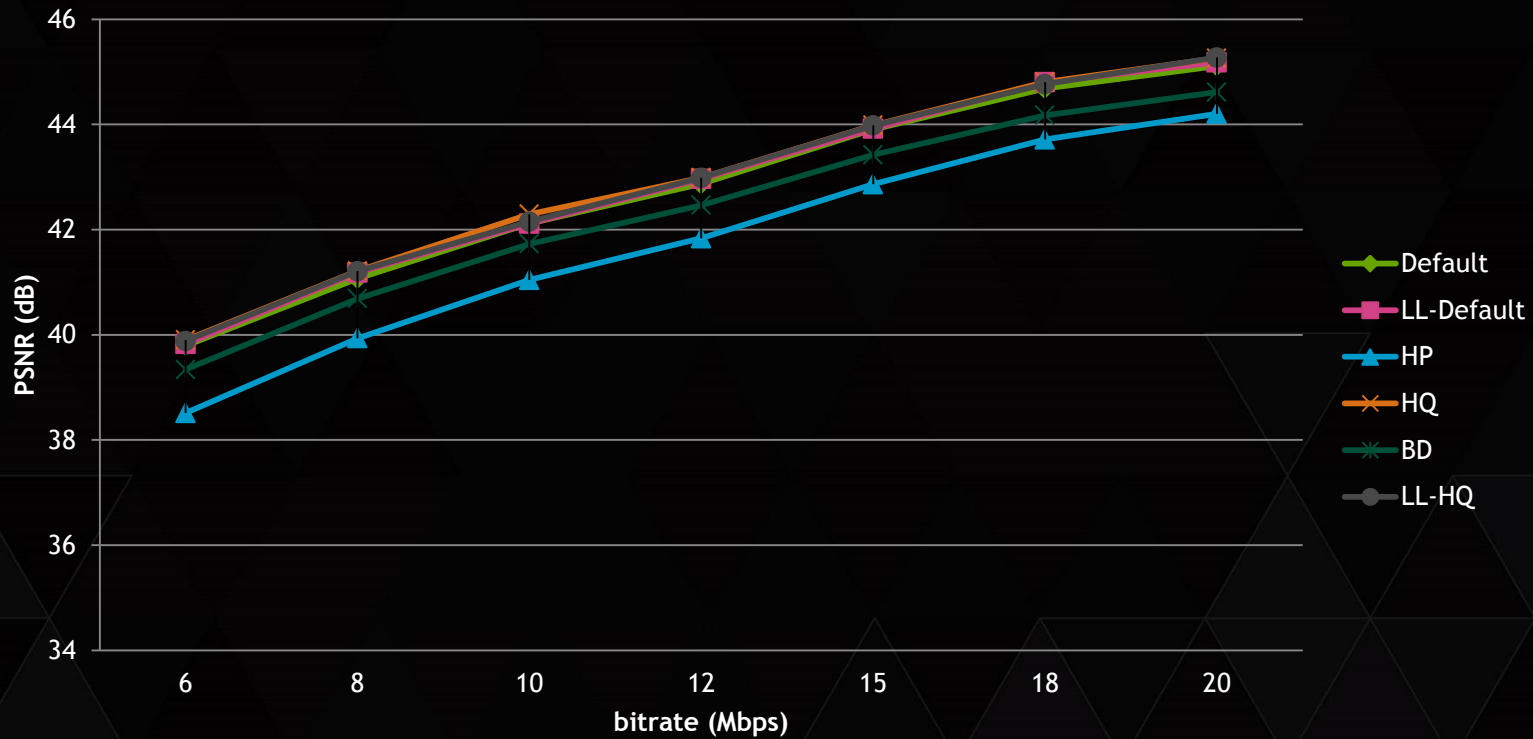
- ▶ NDA only - older release available on NV developer zone
  - ▶ <https://developer.nvidia.com/grid-app-game-streaming>
- ▶ Current release: **3.1 (Now - NDA)**, **2.3 (Public)**
- ▶ Interface header, documentation, sample apps
- ▶ .dll/.so included in the driver
- ▶ **Windows** and **Linux**
- ▶ Works on x86/x64
- ▶ Presets and API's for
  - ▶ Remote graphics (Cloud gaming, remote desktop, capture & stream)
  - ▶ Optimized for low latency



# QUALITY

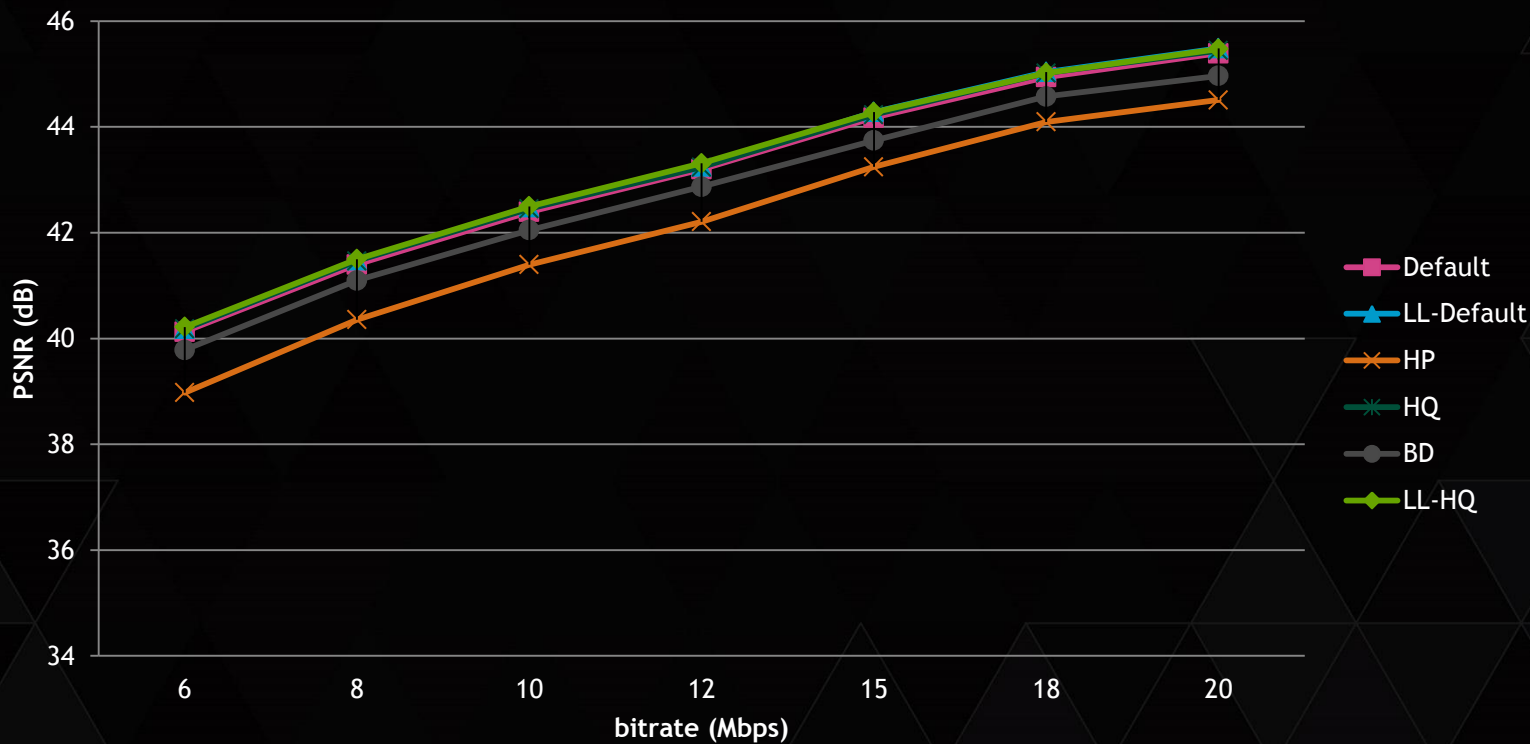
# H.264 QUALITY - 1-PASS ENCODING

## H.264 quality with 1-pass rate control



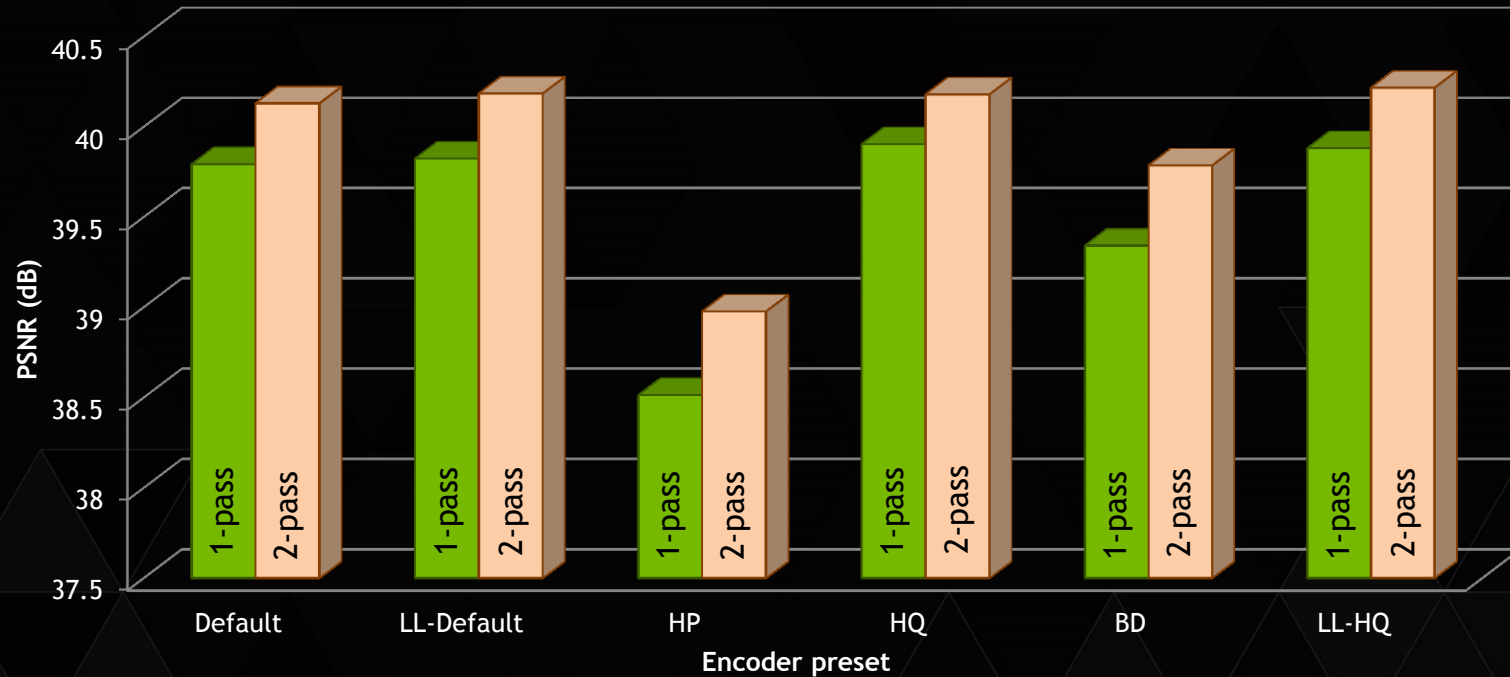
# H.264 QUALITY - 2-PASS ENCODING

## H.264 quality with 2-pass rate control



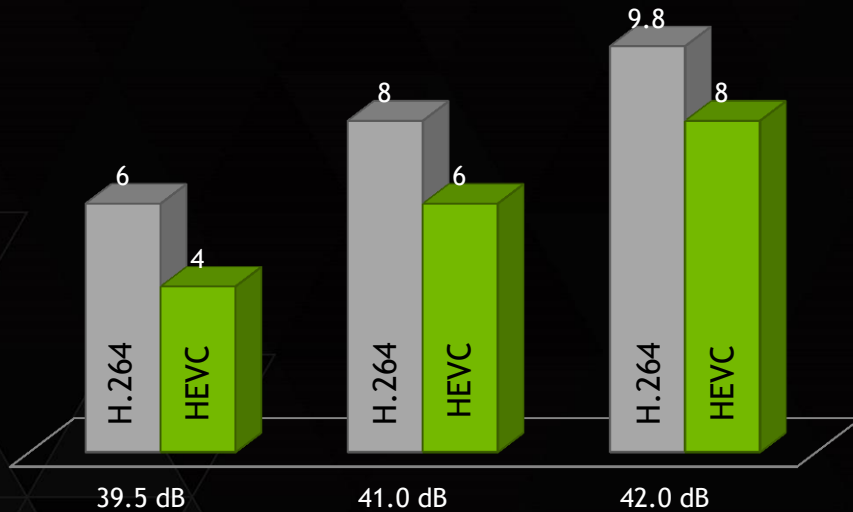
# COMPARISON: 1-PASS VS 2-PASS

H.264 quality comparison: 1-pass vs 2-pass



# BITRATE SAVINGS

### Bitrate savings - Default preset



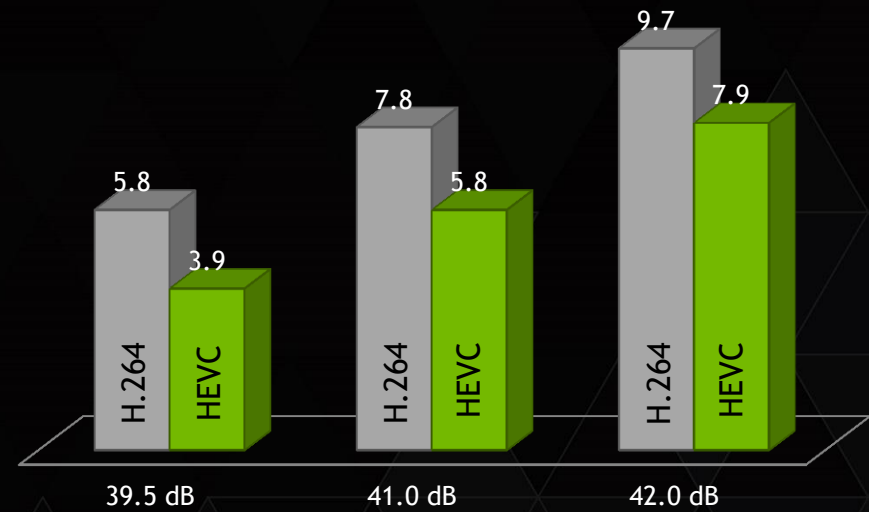
33%

25%

18%

Bitrate savings

### Bitrate savings - HQ preset



33%

26%

19%

# H.264 VS HEVC



# H.264 VS HEVC



# PERFORMANCE



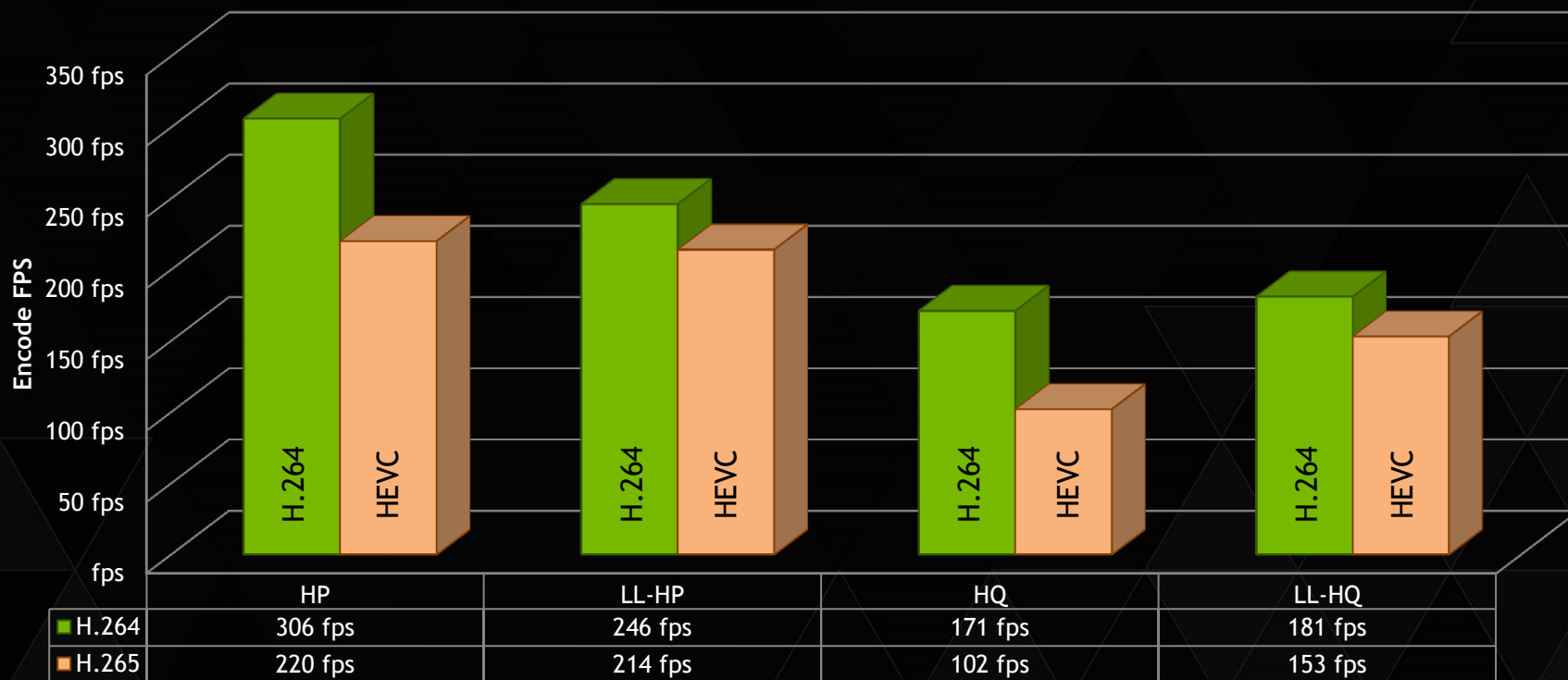
# H.264 PERFORMANCE - GM20X

## H.264 Performance (1080p)



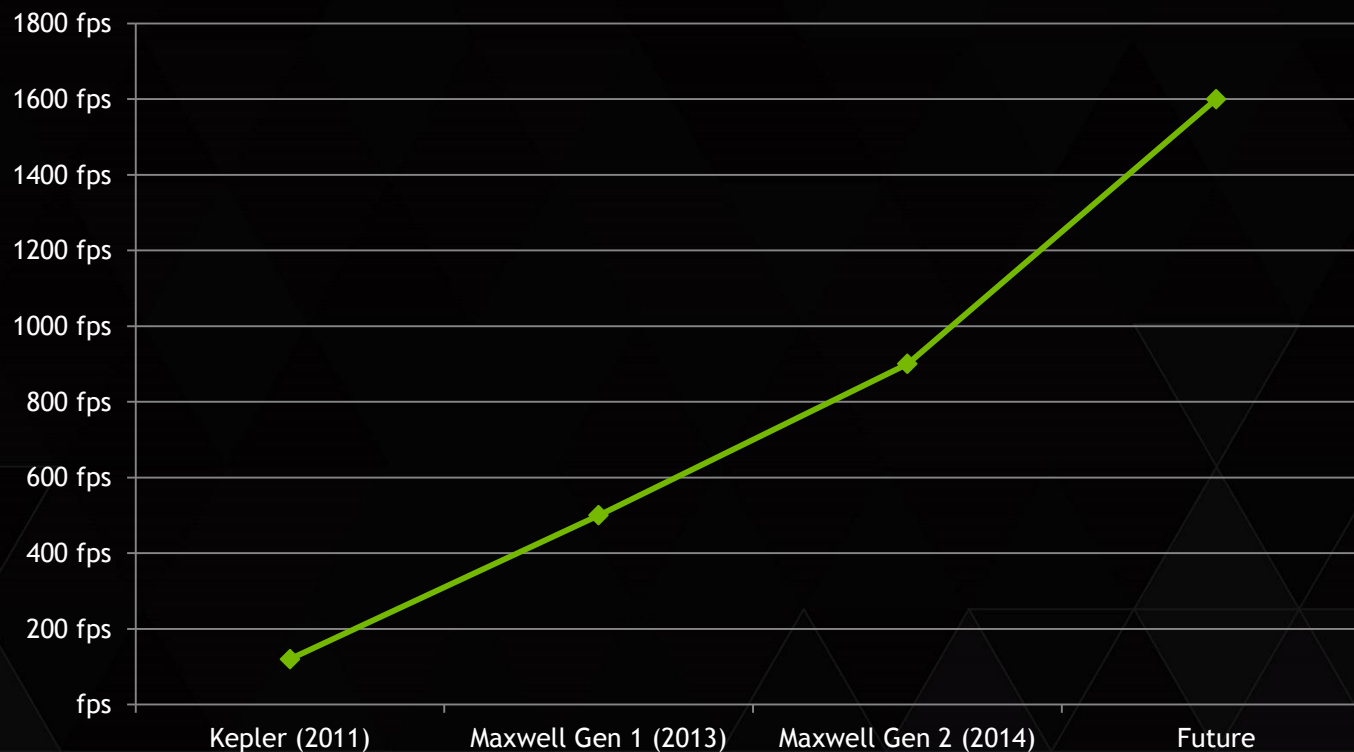
# H.264/HEVC PERF COMPARISON

## H.264/HEVC Performance: 2-pass



# PERFORMANCE - TREND

Performance



# ROADMAP

# ROADMAP

- ▶ Core GPU chip IP
- ▶ Motion estimation only mode - **2H2015**
- ▶ SAO, 10/12-bit, HEVC B-frames
- ▶ Lossless/4:4:4
- ▶ Improved quality for screen content encoding
- ▶ ME performance and quality enhancements
- ▶ **Today:** 4K@60fps
- ▶ **Next:** 8K@??

**GPU** TECHNOLOGY  
CONFERENCE

**THANK YOU**

**APATAIT@NVIDIA.COM**

JOIN THE CONVERSATION

**#GTC15**   