GPU TECHNOLOGY CONFERENCE

HIGH-PERFORMANCE GPU VIDEO ENCODING

ABHIJIT PATAIT

SR. MANAGER, NVIDIA

GPU TECHNOLOGY CONFERENCE

AGENDA

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

GPU TECHNOLOGY CONFERENCE

WHY GPU VIDEO ENCODING?

GPU TECHNOLOGY CONFERENCE

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

GPU TECHNOLOGY CONFERENCE

NVIDIA VIDEO ENCODER CAPABILITIES

GPU TECHNOLOGY CONFERENCE

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

GPU TECHNOLOGY CONFERENCE

FEATURE COMPARISON

Kepler	Maxwell Gen 1 (GM10x)	Maxwell Gen 2 (GM20x)
H.264 only	H.264 only	H.264 and HEVC/H.265
Planar 4:4:4 & proprietary 4:4:4; no lossless encoding	Standard 4:4:4 and H.264 lossless encoding	Standard 4:4:4 and H.264 lossless encoding
~240 fps 2-pass encoding @ 720p	~500 fps 2-pass encoding @ 720p	~900 fps 2-pass encoding @ 720p
GRID K340/K520, K1/K2, Quadro, Tesla K10/K20	Maxwell-based 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 (Now)	NV Encode SDK 4.0+ (Now)	NV Encode SDK 5.0+ (Now)
GRID SDK 1.x, 2.2, 2.3 (Now)	GRID SDK 3.0+ (Now)	In development

GPU TECHNOLOGY CONFERENCE

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

GPU TECHNOLOGY CONFERENCE

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

GPU TECHNOLOGY CONFERENCE

SOFTWARE API

GPU TECHNOLOGY CONFERENCE

USING NVENC

NVENC SDK

Direct Encode

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

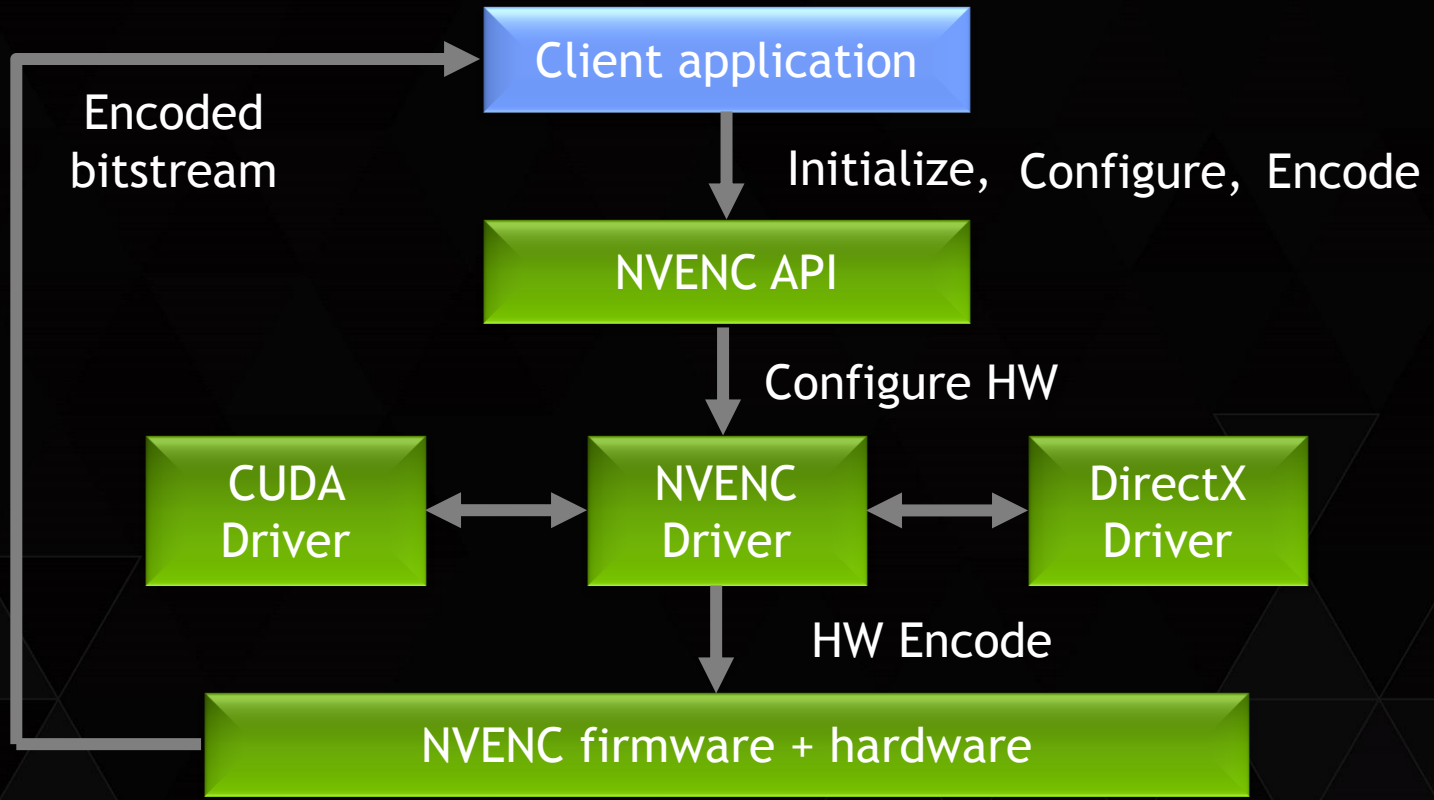
GRID SDK

Capture + Encode

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

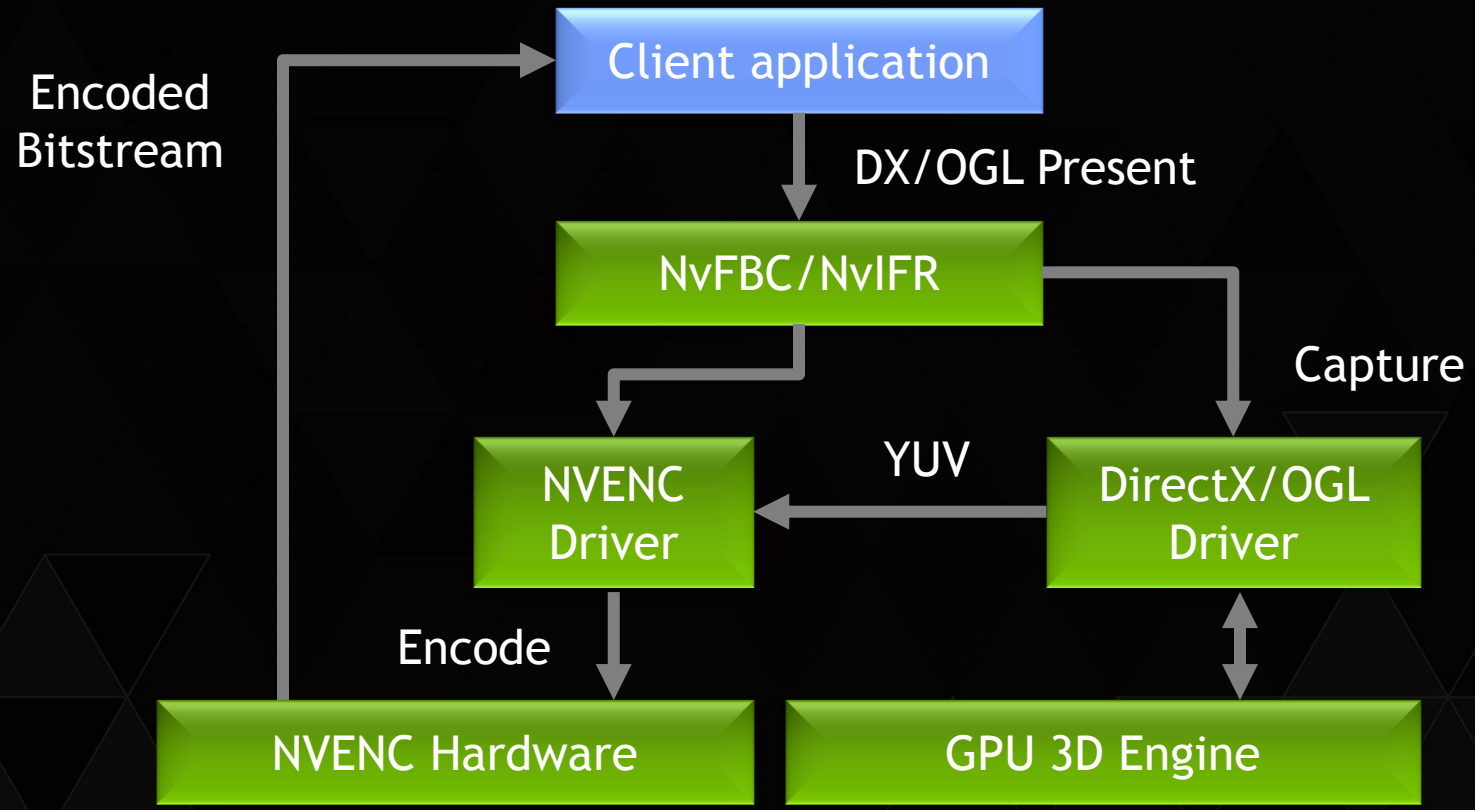
GPU TECHNOLOGY CONFERENCE

DIRECT ENCODE (NVENC SDK)



GPU TECHNOLOGY CONFERENCE

CAPTURE & ENCODE (GRID SDK)



GPU TECHNOLOGY CONFERENCE

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

GPU TECHNOLOGY CONFERENCE

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

GPU TECHNOLOGY CONFERENCE

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

GPU TECHNOLOGY CONFERENCE

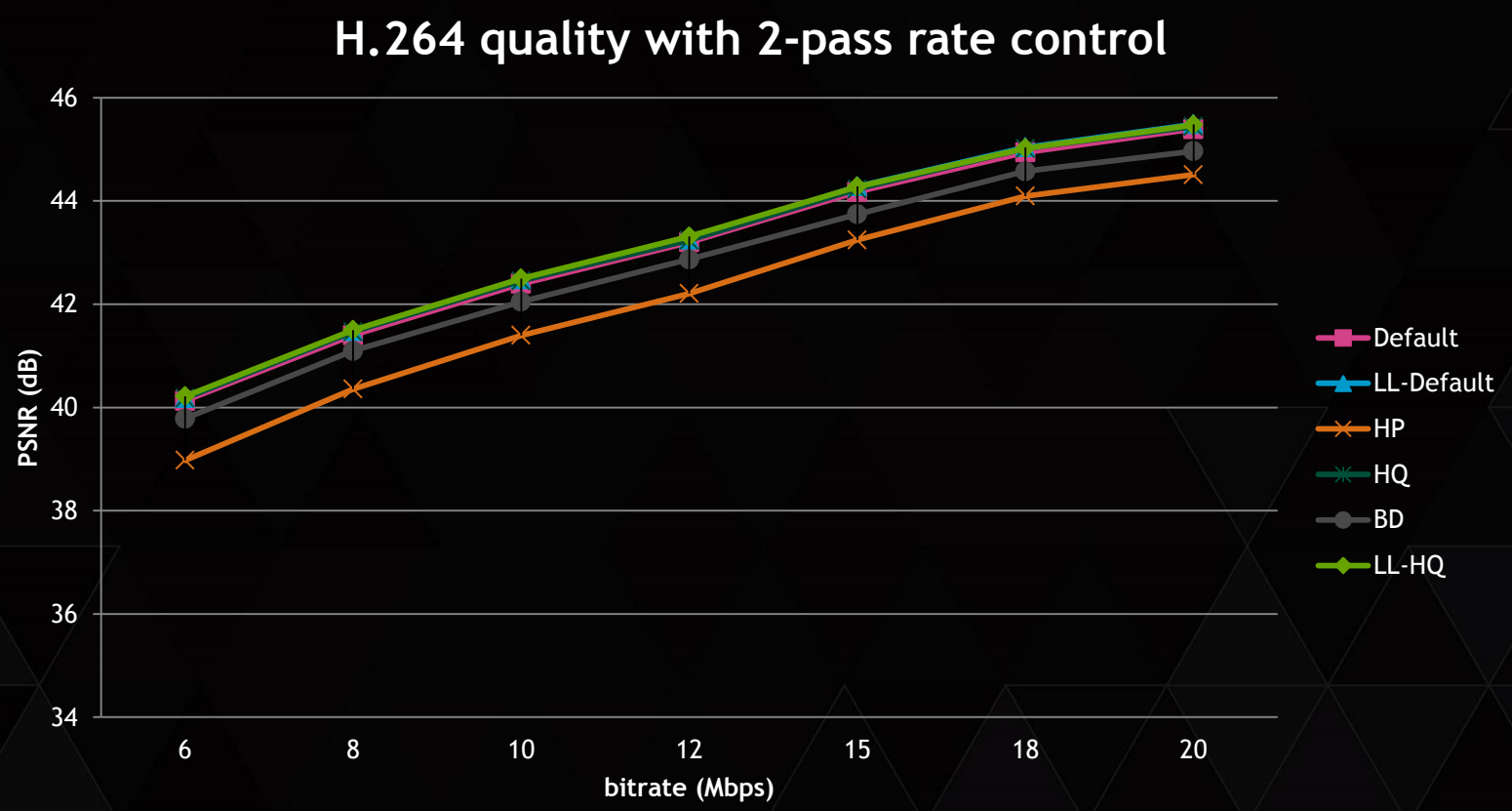
QUALITY

GPU TECHNOLOGY CONFERENCE

H.264 QUALITY - 1-PASS ENCODING

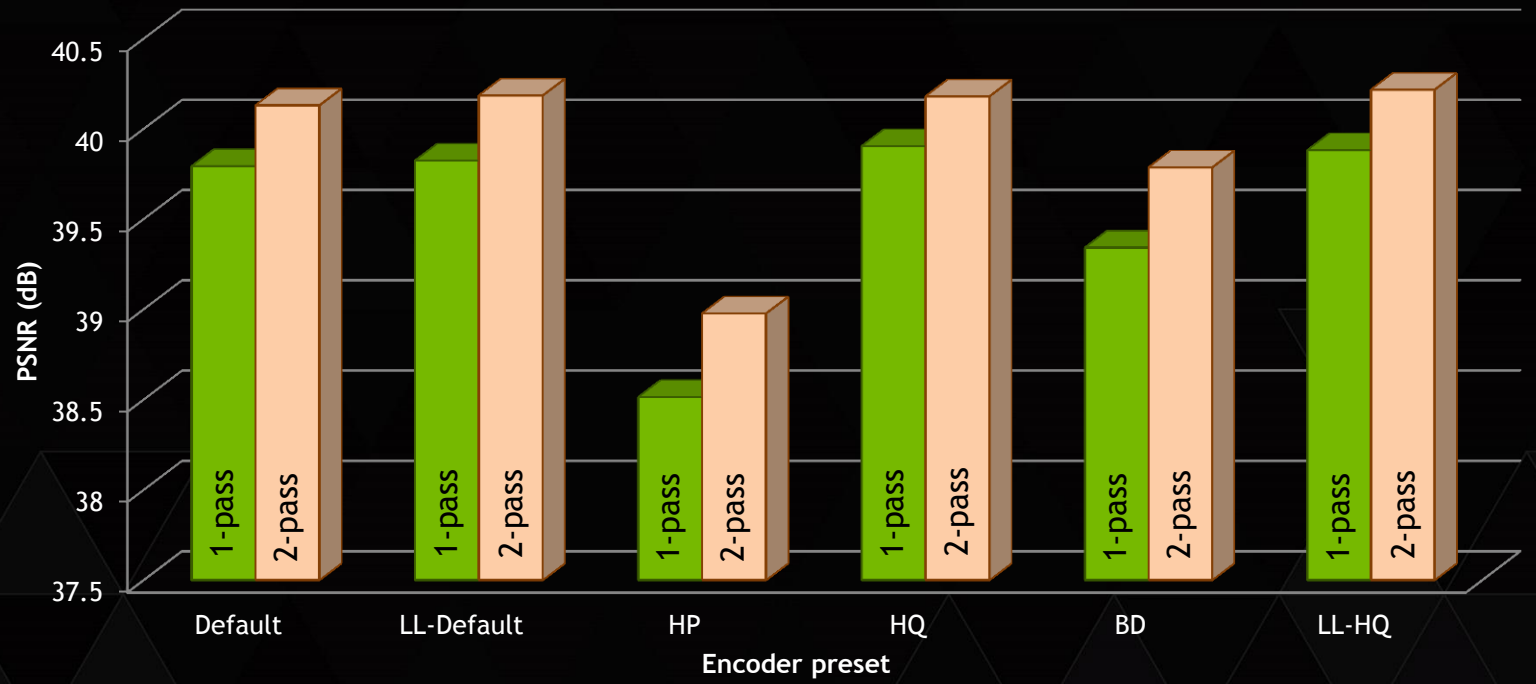
GPU TECHNOLOGY CONFERENCE

H.264 QUALITY – 2-PASS ENCODING



GPU TECHNOLOGY CONFERENCE

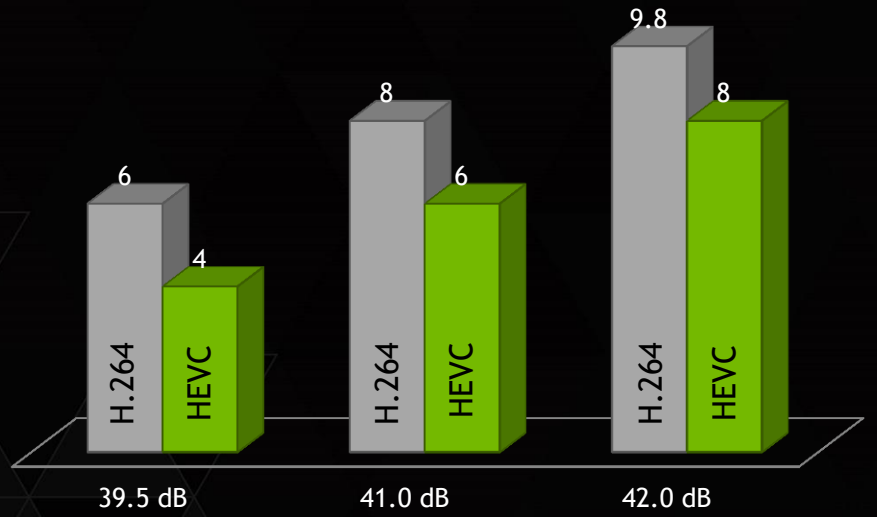
COMPARISON: 1-PASS VS 2-PASS



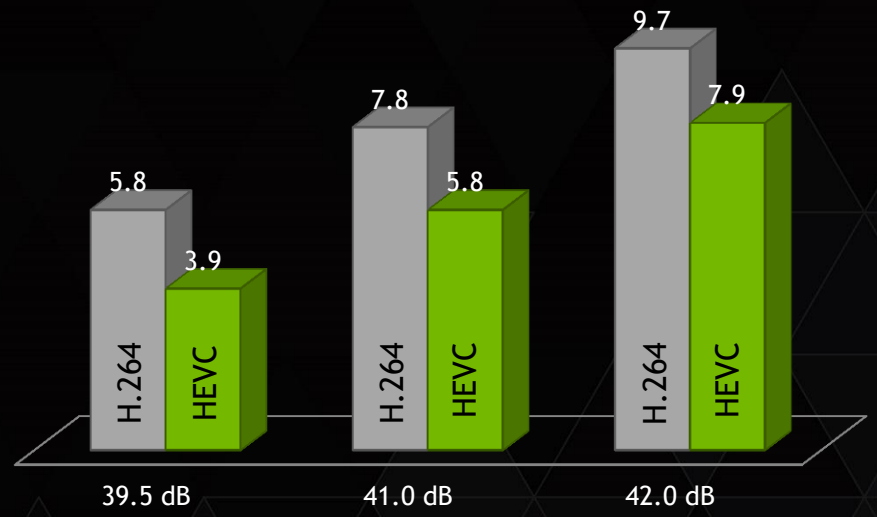
GPU TECHNOLOGY CONFERENCE

BITRATE SAVINGS

Bitrate savings - Default preset



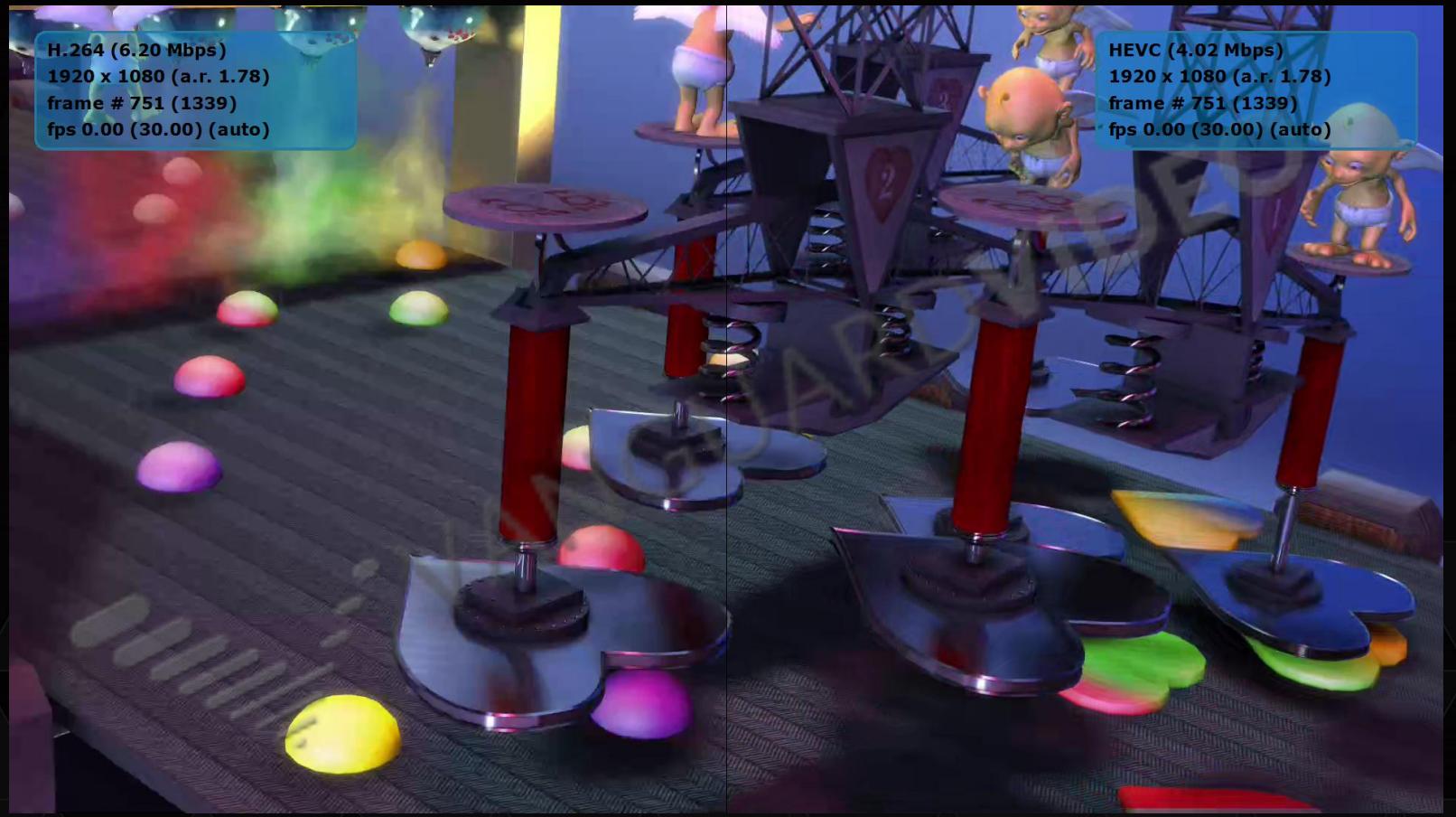
Bitrate savings - HQ preset



33%	25%	18%	Bitrate savings	33%	26%	19%

GPU TECHNOLOGY CONFERENCE

H.264 VS HEVC



Courtesy: Vanguard video

GPU TECHNOLOGY CONFERENCE

H.264 VS HEVC

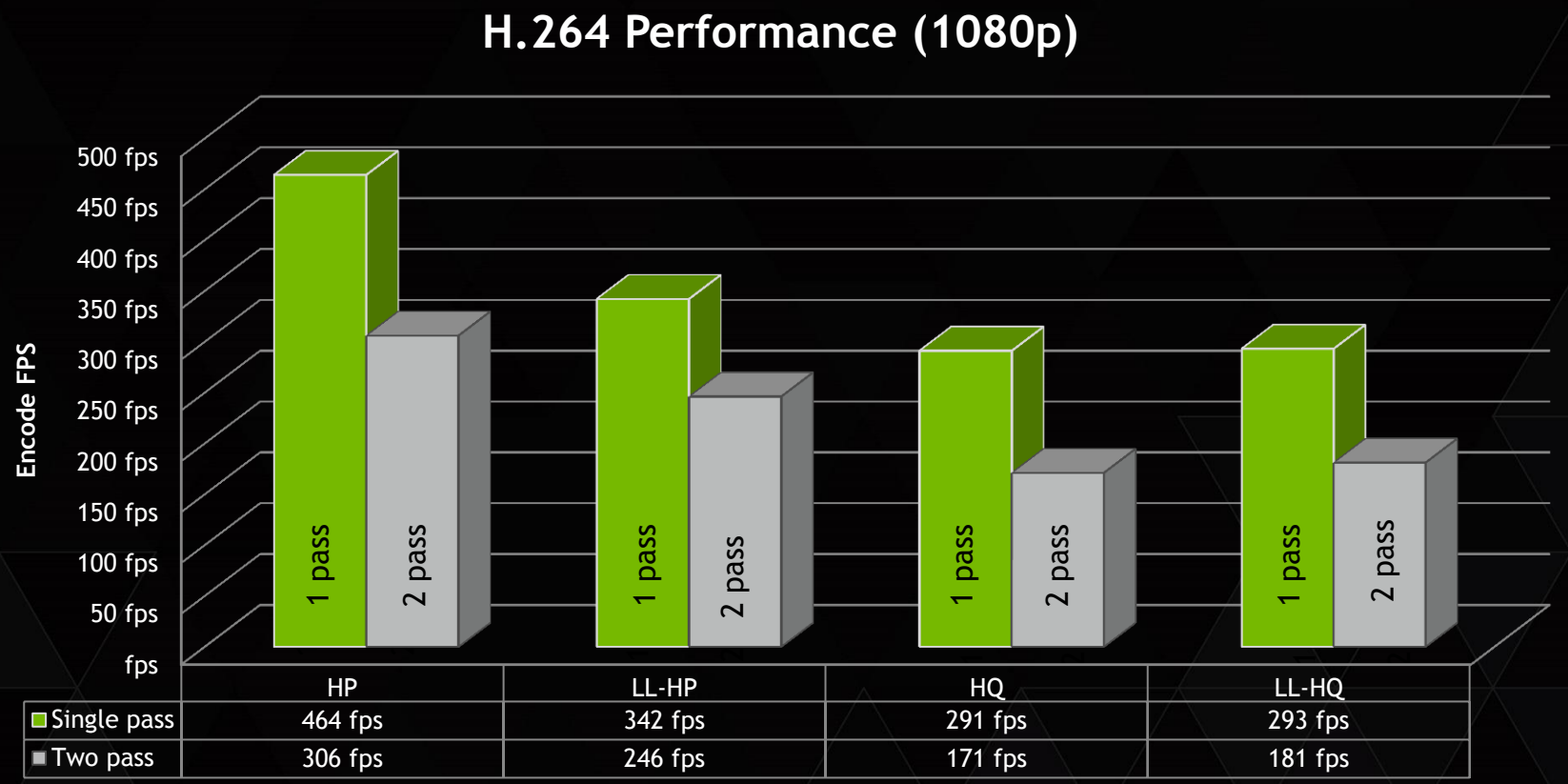
Courtesy: Vanguard video

GPU TECHNOLOGY CONFERENCE

PERFORMANCE

GPU TECHNOLOGY CONFERENCE

H.264 PERFORMANCE - GM20X



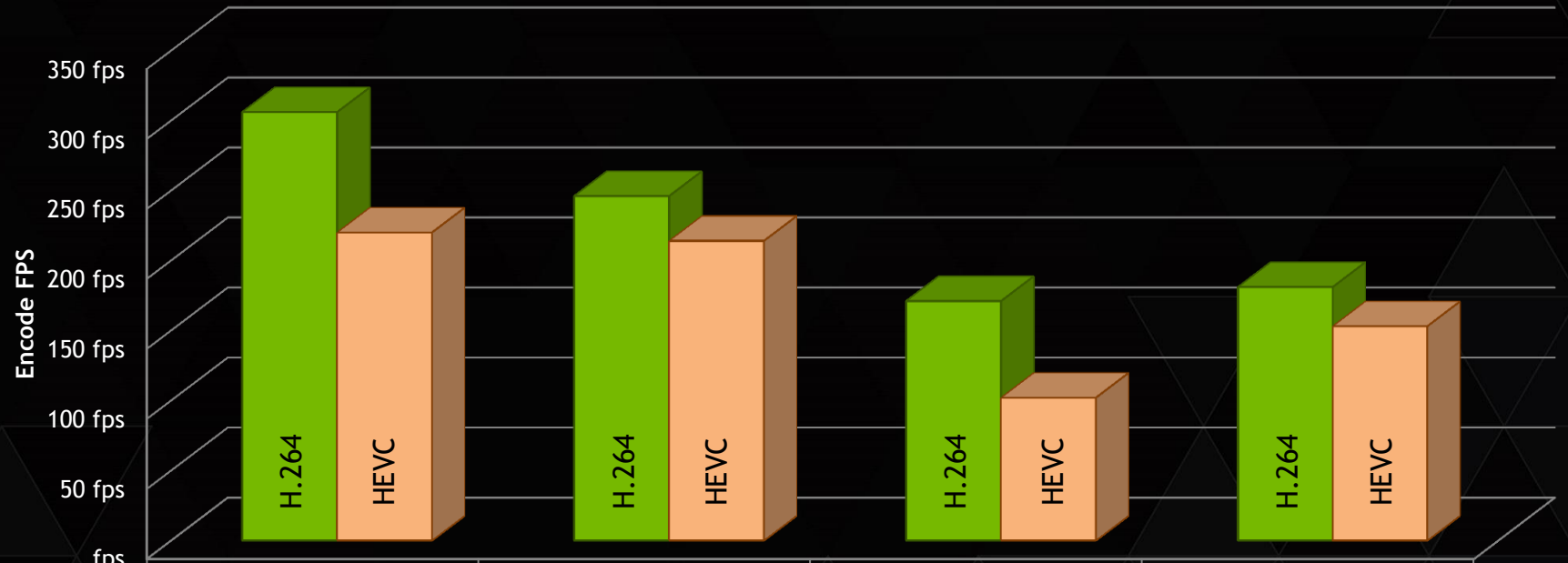
	HP	LL-HP	HQ	LL-HQ
Single pass	464 fps	342 fps	291 fps	293 fps
Two pass	306 fps	246 fps	171 fps	181 fps

1080p, GM20x

GPU TECHNOLOGY CONFERENCE

H.264/HEVC PERF COMPARISON

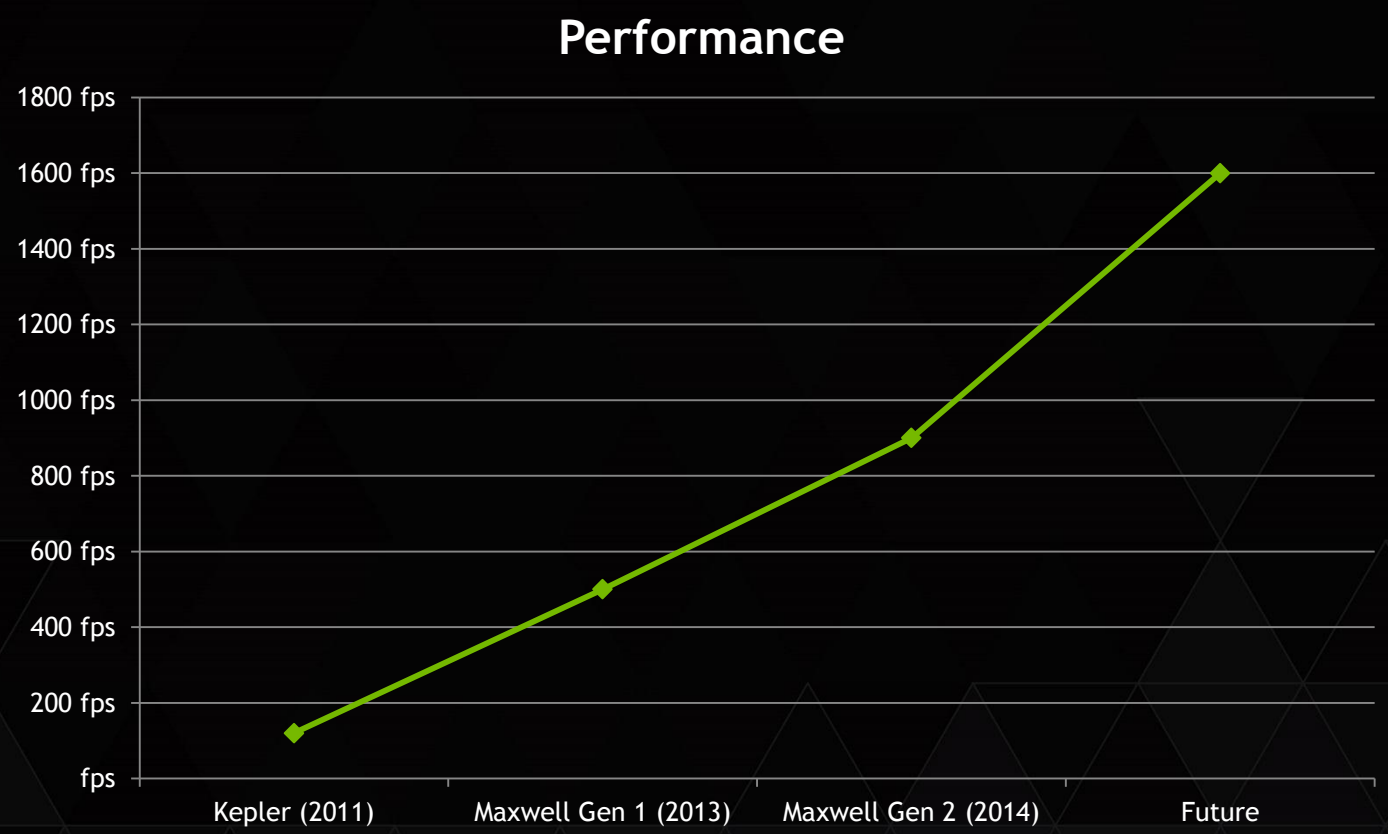
H.264/HEVC Performance: 2-pass



	HP	LL-HP	HQ	LL-HQ
H.264	306 fps	246 fps	171 fps	181 fps
H.265	220 fps	214 fps	102 fps	153 fps

GPU TECHNOLOGY CONFERENCE

PERFORMANCE - TREND



GPU TECHNOLOGY CONFERENCE

ROADMAP

GPU TECHNOLOGY CONFERENCE

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@??

THANK YOU

APATAIT@NVIDIA.COM

JOIN THE CONVERSATION

#GTC15