GPU TECHNOLOGY CONFERENCE

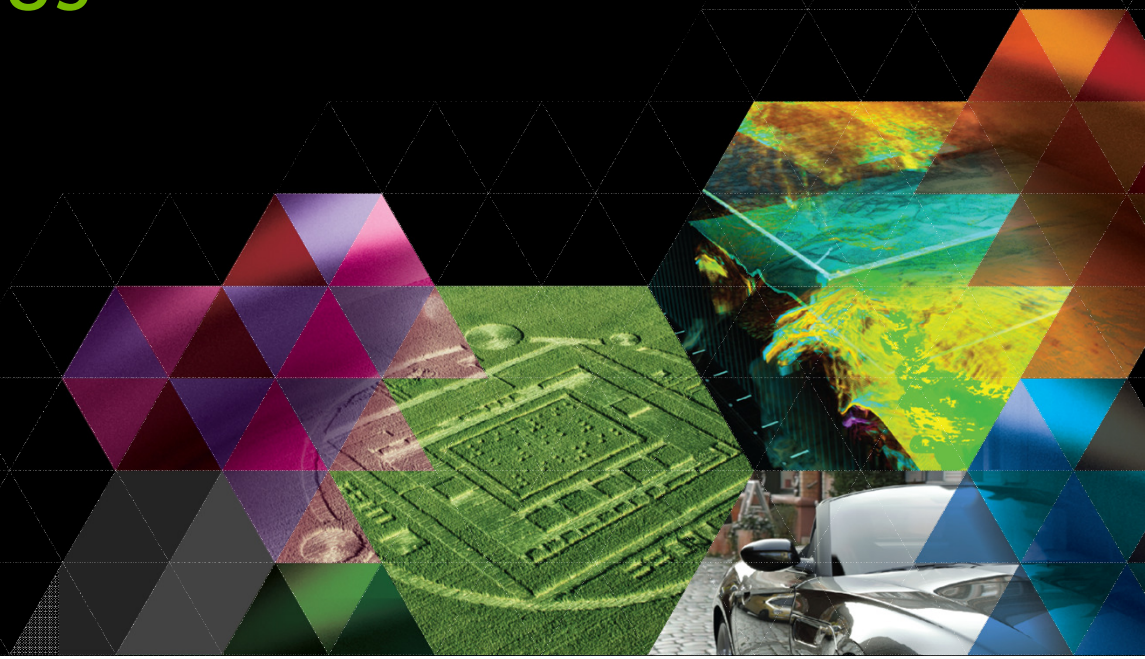
NVIDIA

HIGH PERFORMANCE VIDEO ENCODING USING NVIDIA GPUS

Abhijit Patait

Sr. Manager,

GPU Multimedia SW



AGENDA

- Overview GPU Video Encoding
- NVIDIA Video Encoding Capabilities
  - Kepler vs Maxwell GPU capabilities
  - Roadmap
- Software API
- Performance & Quality

WHY GPU VIDEO ENCODING?

BENEFITS OF ENCODING ON GPU

- Low power
  - Fixed function hardware
  - Reduced memory transfers
- Low latency
- High performance
- Higher density
- Scalability
- Ease of Programming
  - Linux, Windows, C/C++, Application portability

NVIDIA GPU VIDEO ENCODING CAPABILITIES

NVIDIA GPU ENCODING CAPABILITIES

Feature	Benefits
H.264 base, main, high profiles	Wide range of use-cases
High performance (Up to 16x HD)	“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
MVC	Full resolution stereo encode
Up to 4096 × 4096 in HW	High resolution encode
API - NV Encode SDK & GRID SDK	Flexible, Win/Linux, DirectX/CUDA
Independent of CUDA	Use CUDA and encode simultaneously

VIDEO ENCODING — KEPLER VS. MAXWELL

Kepler (GK104, GK107, GK106, GK110, GK208)	Maxwell (GM107)
Planar 4:4:4	Standard 4:4:4 and H.264 lossless encoding
~240 fps 2-pass encoding @ 720p	~500 fps 2-pass encoding @ 720p
GRID K340/K520, K1/K2, Quadro, Tesla K10/K20	Current and future Maxwell GPU-boards
GeForce – 2 full-speed encode sessions/GPU	GeForce – 2 full-speed encode sessions/GPU
NV Encode SDK 1.0, 2.0, 3.0 (Now)	NV Encode SDK 4.0+ (May 2014)
GRID SDK 1.x, 2.2, 2.3 (Now)	GRID SDK 3.0+ (June 2014)

NVIDIA VIDEO ENCODING ROADMAP

- Performance improvements
- Quality improvements
  - 4:4:4 & lossless encoding
  - Rate control enhancements
  - Adaptive quantization
  - ROI, ME-only mode
- New video standards

NVENC SOFTWARE APIS

USING NVENC

Direct Encode

NVENC SDK

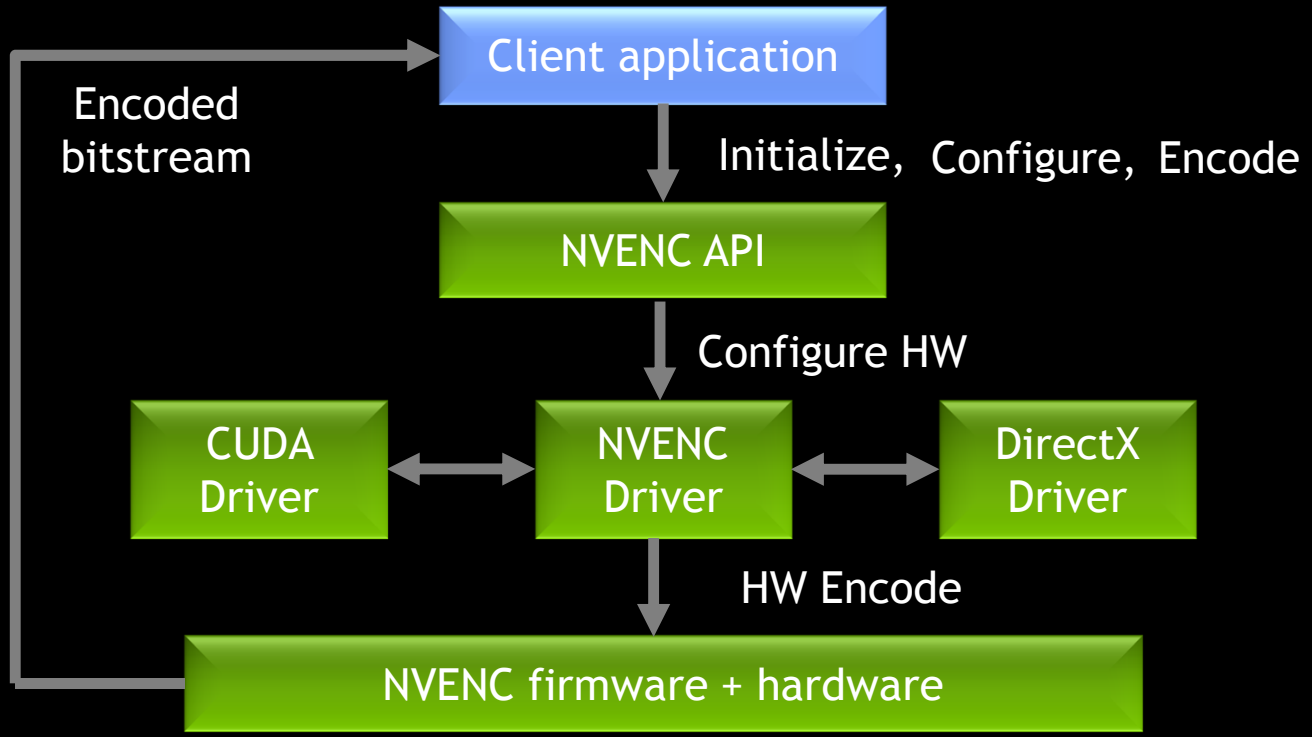
- No capture
- Transcoding
- Archiving
- Video editing
- CUDA pre-process + encoding
- Granular encoder settings
- D3D, CUDA interop

Capture + Encode

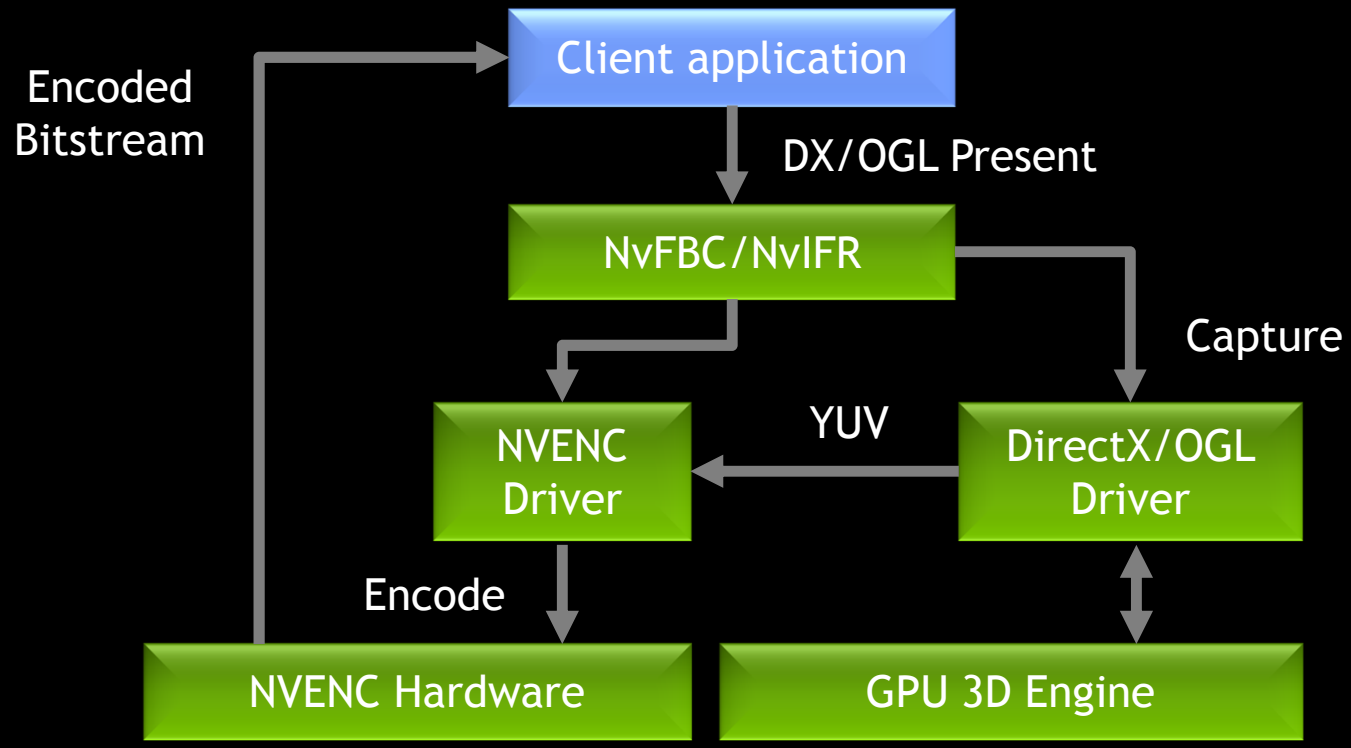
GRID SDK

- 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 AND ENCODE (GRID SDK)



NVENC SDK

- Available on NVIDIA developer zone
  - https://developer.nvidia.com/nvidia-video-codec-sdk
  - Current release 3.0
  - Release 4.0 in May 2014 with Maxwell support
- Interface header, documentation, sample application
  - .dll/.so included in the driver
- Unified API for Windows and Linux
- Works on x86/x64
- Various API's, presets, rate control modes for
  - Transcoding
  - Video conferencing
  - GTC Session S4654

NVENC SDK (CONTD.)

- Advantages
  - Flexibility
    - Dynamic resolution/bitrate change
    - CABAC vs CAVLC; low-level encoder settings, B-frames, sync vs async, custom QP
    - Linux, Windows, DirectX, CUDA, OGL (via CUDA)
    - Also works on GeForce hardware (2 sessions/GPU)
  - Error concealment
    - Reference picture invalidation
    - Intra-refresh
  - Quality
    - Two-pass modes for higher quality
    - Various presets with quality/performance trade-off
    - 4:4:4 & lossless encoding (Maxwell only)

GRID SDK ENCODE

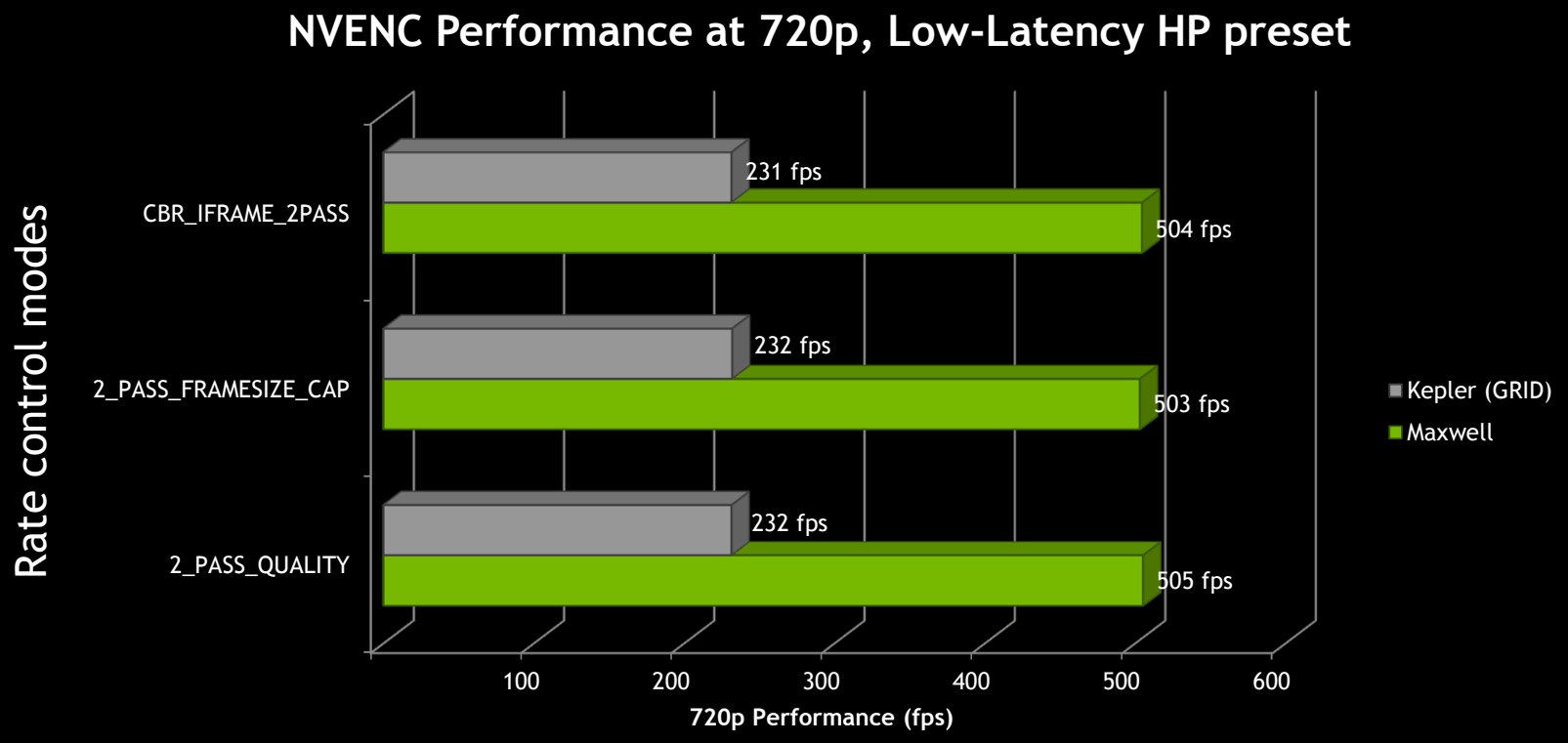
- Available on NVIDIA developer zone
  - https://developer.nvidia.com/grid-app-game-streaming
  - Current release: 2.2
- Interface header, documentation, sample apps
  - .dll/.so included in the driver
- Windows and Linux
- Works on x86/x64
- Various presets and API's for
  - Remote graphics (Cloud gaming, remote desktop, capture & stream)
- Optimized for low latency

GRID SDK (CONTD.)

- Advantages
  - Simplicity
    - Very simple API; single function call for capture + H.264 encode
  - Low-latency, high performance
    - Optimized API
  - Error concealment
    - Reference picture invalidation
    - Intra-refresh
  - Quality
    - Two-pass modes for higher quality
    - 4:4:4 & lossless encoding (Maxwell only)

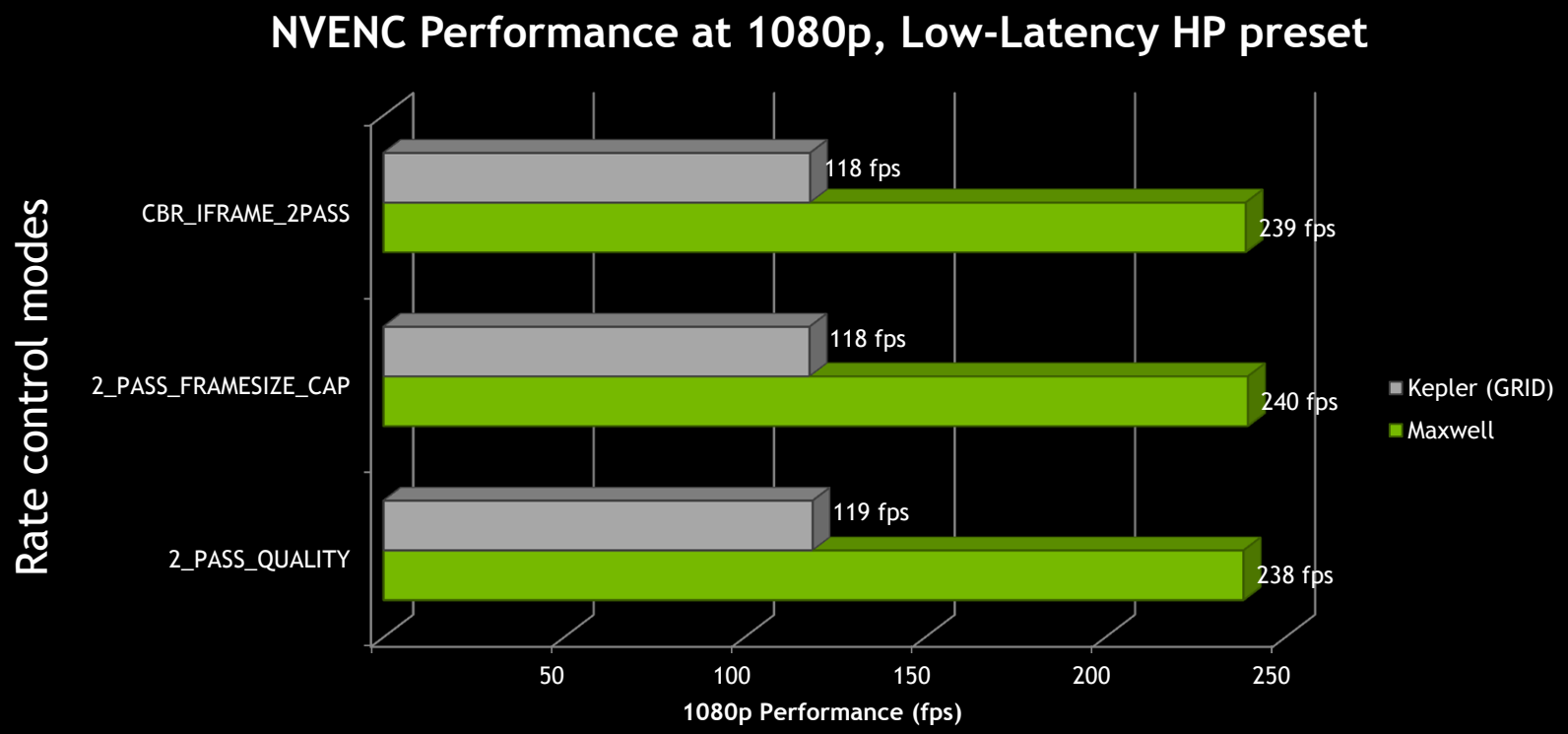
PERFORMANCE AND QUALITY

PERFORMANCE – 720P



Performance measured on GRID K520 with GRID SDK NVENC performance benchmarking application

PERFORMANCE – 1080P

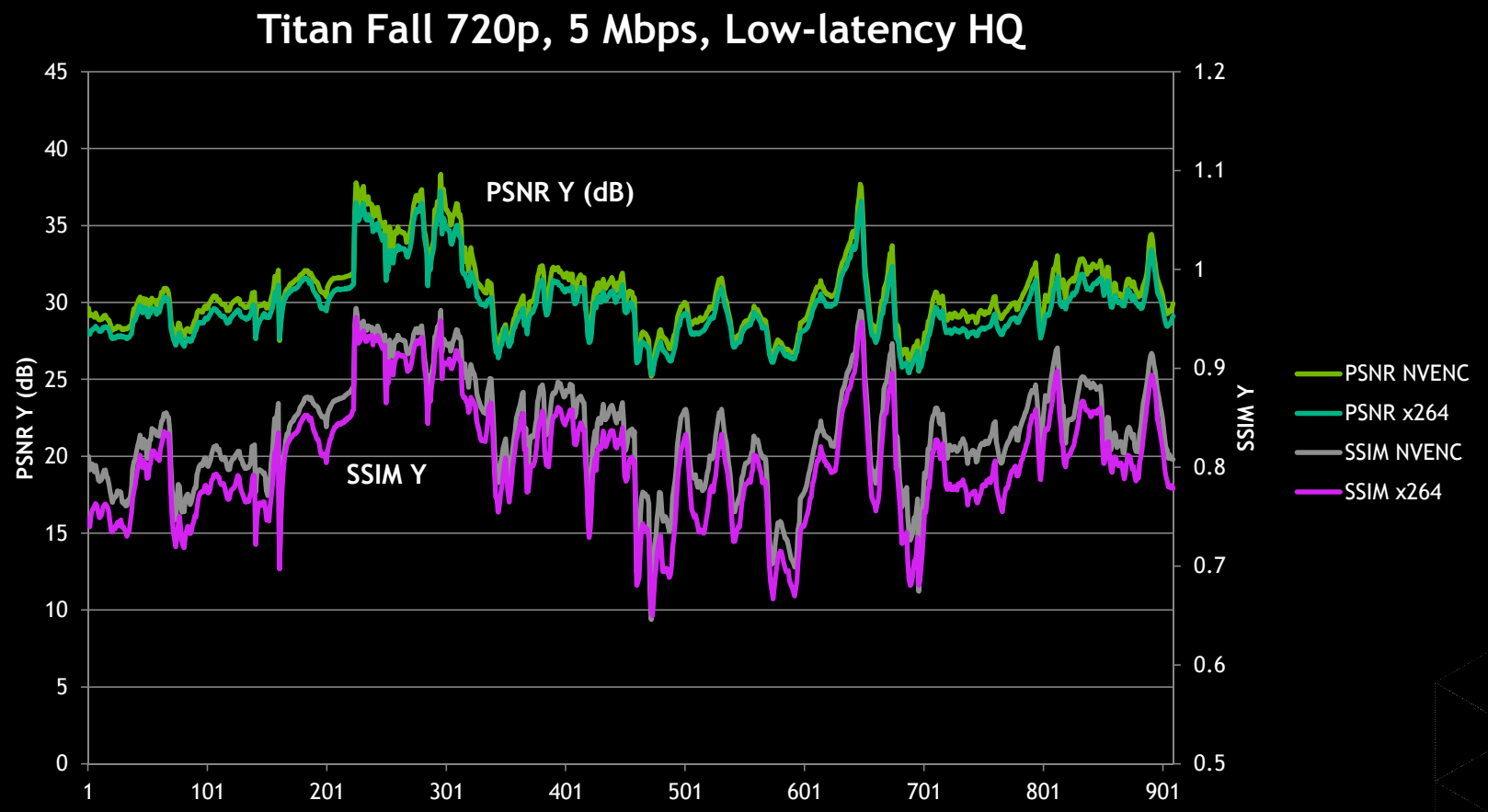


Performance measured on GRID K520 with GRID SDK NVENC performance benchmarking application

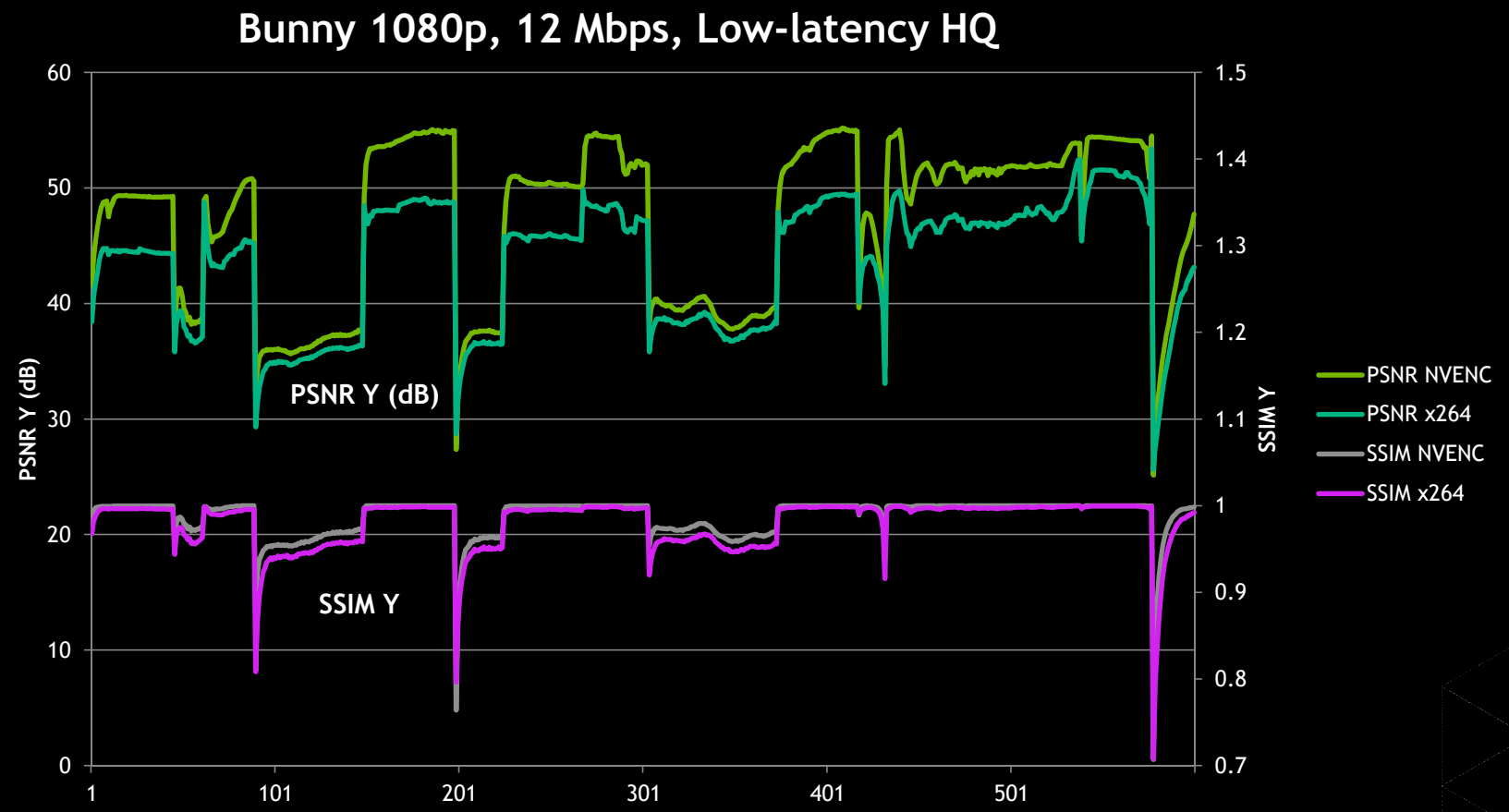
ENCODING QUALITY VS X264 – ASSUMPTIONS

- Infinite GOP IPPP…
- VBV buffer = bitrate/framerate
- x264
  - Zero latency
  - CRF = 24
  - Preset = faster
- NVENC
  - Preset = LOW_LATENCY_HQ
  - RC = 2-pass-quality

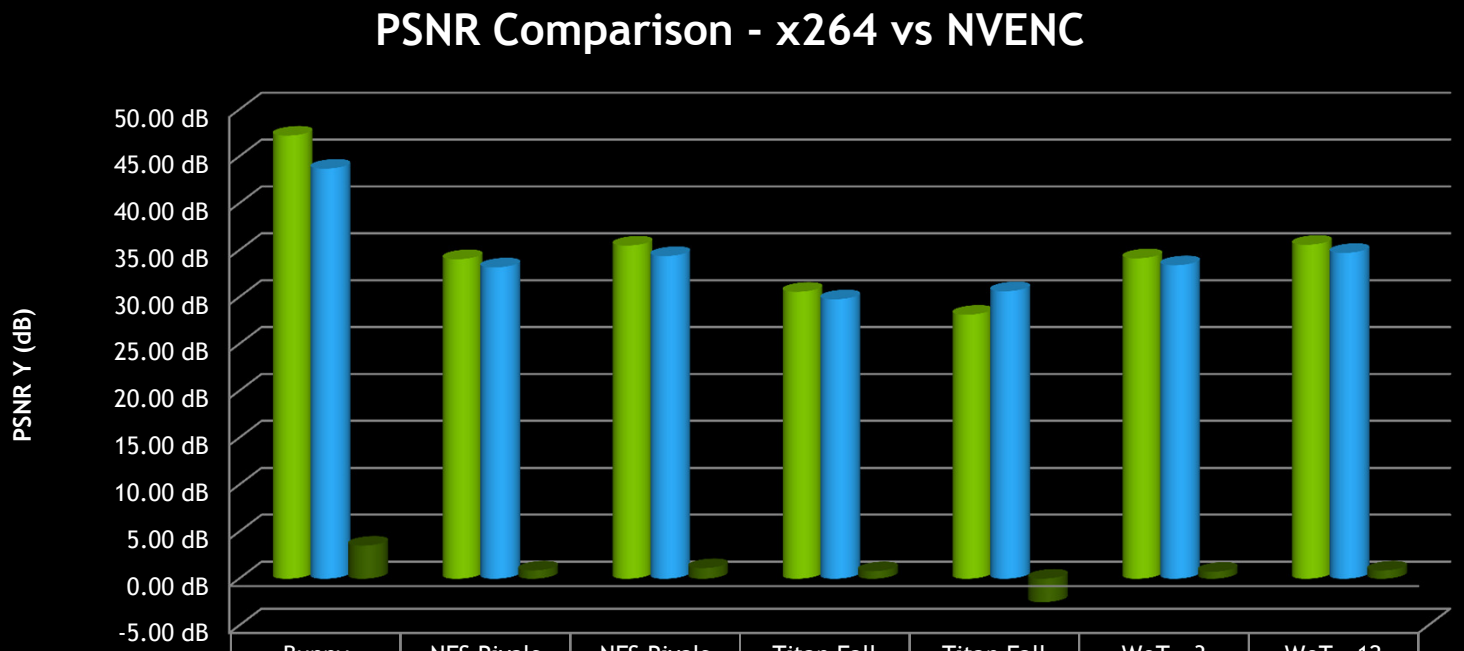
NVENC/X264 QUALITY COMPARISON



NVENC/X264 QUALITY COMPARISON

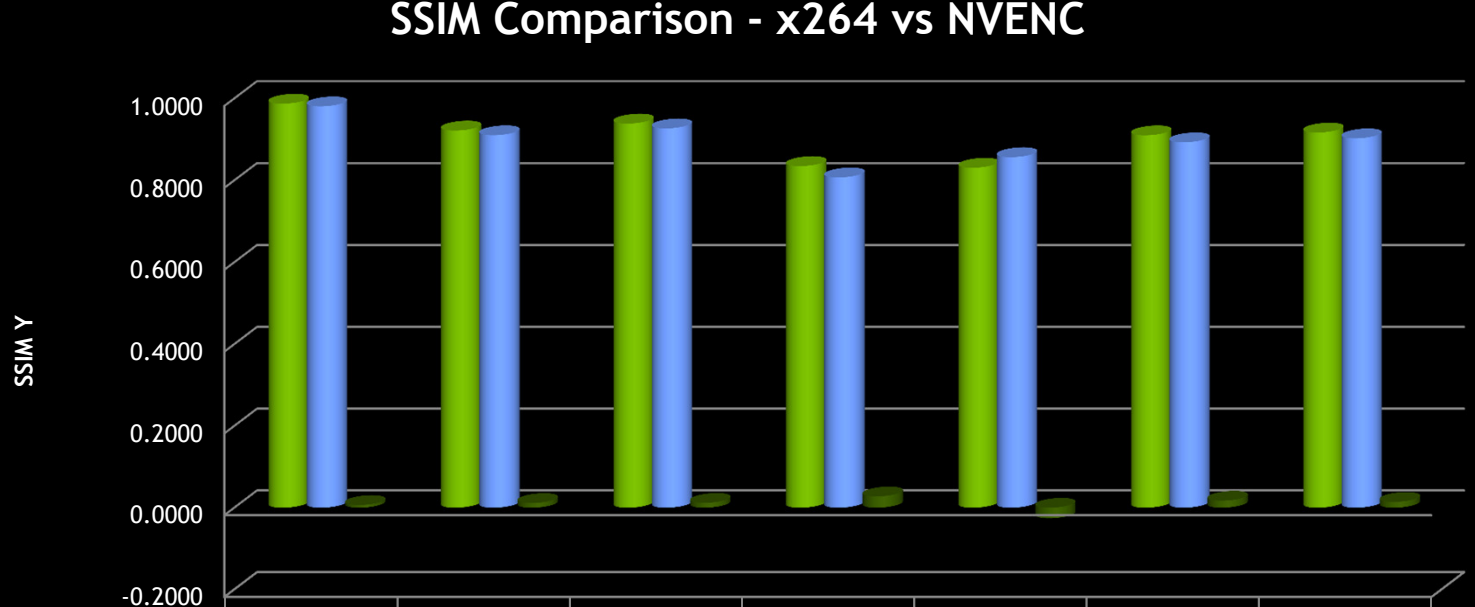


QUALITY COMPARISON – PSNR



	Bunny 1080p	NFS Rivals 720p	NFS Rivals 1080p	Titan Fall 720p	Titan Fall 1080p	WoT - 3 1280 × 768	WoT - 12 1280 × 768
PSNR NVENC	47.24 dB	34.05 dB	35.51 dB	30.58 dB	28.13 dB	34.15 dB	35.60 dB
PSNR x264	43.71 dB	33.18 dB	34.39 dB	29.78 dB	30.63 dB	33.41 dB	34.72 dB
PSNR Difference	3.52 dB	0.87 dB	1.12 dB	0.80 dB	-2.50 dB	0.74 dB	0.87 dB

QUALITY COMPARISON – SSIM



	Bunny 1080p	NFS Rivals 720p	NFS Rivals 1080p	Titan Fall 720p	Titan Fall 1080p	WoT - 3 1280 × 768	WoT - 12 1280 × 768
SSIM NVENC	0.9874	0.9217	0.9388	0.8350	0.8309	0.9101	0.9169
SSIM x264	0.9808	0.9103	0.9269	0.8073	0.8567	0.8930	0.9027
SSIM Difference	0.01	0.01	0.01	0.03	-0.03	0.02	0.01

QUESTIONS?