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
<table>
<thead>
<tr>
<th>Feature</th>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>H.264 base, main, high profiles</td>
<td>Wide range of use-cases</td>
</tr>
<tr>
<td>High performance (Up to 16x HD)</td>
<td>“Blazing-speed” encoding</td>
</tr>
<tr>
<td>YUV 4:2:0 and 4:4:4 support</td>
<td>High quality encoding without chroma subsampling</td>
</tr>
<tr>
<td>QP maps</td>
<td>Customizable quality, region of interest encoding</td>
</tr>
<tr>
<td>MVC</td>
<td>Full resolution stereo encode</td>
</tr>
<tr>
<td>Up to 4096 × 4096 in HW</td>
<td>High resolution encode</td>
</tr>
<tr>
<td>API - NV Encode SDK &amp; GRID SDK</td>
<td>Flexible, Win/Linux, DirectX/CUDA</td>
</tr>
<tr>
<td>Independent of CUDA</td>
<td>Use CUDA and encode simultaneously</td>
</tr>
</tbody>
</table>
## Video Encoding — Kepler vs. Maxwell

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

**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
DIRECT ENCODE (NVENC SDK)

Client application

- Initialize, Configure, Encode

NVENC API

- Configure HW

CUDA Driver

NVENC Driver

DirectX Driver

NVENC firmware + hardware

Encoded bitstream
NVENC SDK

- Available on NVIDIA developer zone
  - 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
  - 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 (CONT'D.)

- 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

NVENC Performance at 720p, Low-Latency HP preset

Rate control modes

- CBR_IFRAME_2PASS
  - Kepler (GRID): 231 fps
  - Maxwell: 504 fps

- 2_PASS_FRAMESIZE_CAP
  - Kepler (GRID): 232 fps
  - Maxwell: 503 fps

- 2_PASS_QUALITY
  - Kepler (GRID): 232 fps
  - Maxwell: 505 fps

Performance measured on GRID K520 with GRID SDK NVENC performance benchmarking application
PERFORMANCE - 1080P

NVENC Performance at 1080p, Low-Latency HP preset

- CBR_IFRAME_2PASS: 118 fps, 239 fps
- 2_PASS_FRAMESIZE_CAP: 118 fps, 240 fps
- 2_PASS_QUALITY: 119 fps, 238 fps

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

Titan Fall 720p, 5 Mbps, Low-latency HQ

PSNR Y (dB)

SSIM Y
NVENC/X264 QUALITY COMPARISON

Bunny 1080p, 12 Mbps, Low-latency HQ
QUALITY COMPARISON - PSNR

PSNR Comparison - x264 vs NVENC

<table>
<thead>
<tr>
<th>Video Game</th>
<th>1080p PSNR (dB)</th>
<th>720p PSNR (dB)</th>
<th>Noiseless PSNR (dB)</th>
<th>PSNR Difference (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bunny</td>
<td>47.24</td>
<td>34.05</td>
<td>35.51</td>
<td>3.52</td>
</tr>
<tr>
<td>NFS Rivals</td>
<td>43.71</td>
<td>33.18</td>
<td>34.39</td>
<td>1.12</td>
</tr>
<tr>
<td>Titan Fall</td>
<td>43.71</td>
<td>33.18</td>
<td>34.39</td>
<td>1.12</td>
</tr>
<tr>
<td>WoT - 3</td>
<td>33.87</td>
<td>33.87</td>
<td>33.87</td>
<td>0.00</td>
</tr>
<tr>
<td>WoT - 12</td>
<td>33.87</td>
<td>33.87</td>
<td>33.87</td>
<td>0.00</td>
</tr>
</tbody>
</table>

PSNR Y (dB)
QUALITY COMPARISON - SSIM

SSIM Comparison - x264 vs NVENC

<table>
<thead>
<tr>
<th></th>
<th>SSIM NVENC</th>
<th>SSIM x264</th>
<th>SSIM Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bunny 1080p</td>
<td>0.9874</td>
<td>0.9808</td>
<td>-0.0069</td>
</tr>
<tr>
<td>NFS Rivals 720p</td>
<td>0.9217</td>
<td>0.9103</td>
<td>-0.0114</td>
</tr>
<tr>
<td>NFS Rivals 1080p</td>
<td>0.9388</td>
<td>0.9269</td>
<td>-0.0119</td>
</tr>
<tr>
<td>Titan Fall 720p</td>
<td>0.8350</td>
<td>0.8073</td>
<td>-0.0277</td>
</tr>
<tr>
<td>Titan Fall 1080p</td>
<td>0.8309</td>
<td>0.8567</td>
<td>0.0258</td>
</tr>
<tr>
<td>WoT - 3 1280 × 768</td>
<td>0.9101</td>
<td>0.8930</td>
<td>-0.0171</td>
</tr>
<tr>
<td>WoT - 12 1280 × 768</td>
<td>0.9169</td>
<td>0.9027</td>
<td>-0.0142</td>
</tr>
</tbody>
</table>
QUESTIONS?