



SPECIÁLNÍ PŘENOSY OBRAZU A ZVUKU

Miloš Liška

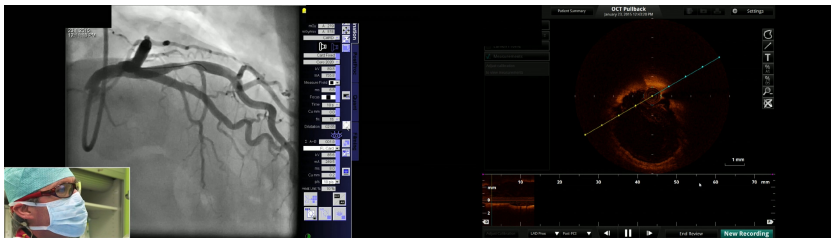
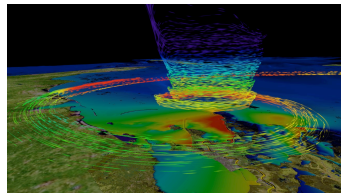
CESNET

prosinec 2017

Laboratoř pokročilých síťových technologií FI MU

- Přenosy obrazu s vysokým rozlišením (HD, 2K, 4K, 8K), vysoce kvalitním zvukem a velmi nízkým zpožděním
 - Detaily, imerze
 - Interaktivita, vzdáleně ovládané aplikace
 - Zachycení rychlých pohybů

- Vědecké vizualizace
- Medicína
 - Kardiologie, chirurgie, endoskopie



■ Výuka

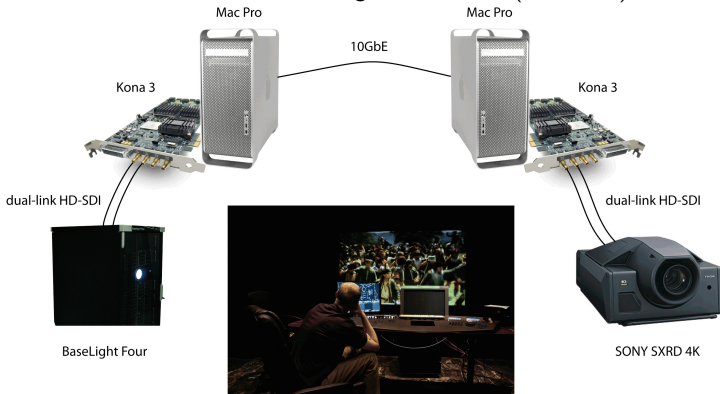
- výuka na dálku s vysokým stupněm imerze



■ Televizní a filmový průmysl

- vzdálená produkce a postprodukce
- restaurování filmového materiálu

Vzdálená konzole BaseLight v CinePostu (Barrandov)

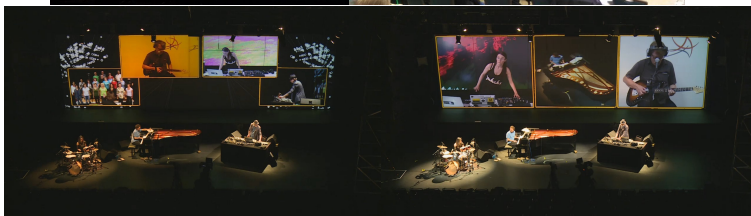
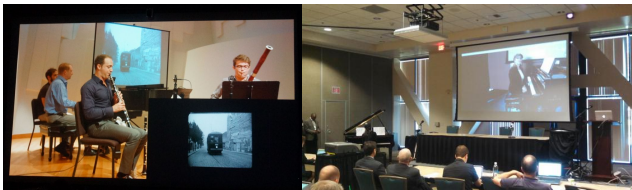


- Aplikace pro spolupráci na dálku
 - sdílení a kombinování různých druhů médií
- Vzdálené ovládání přístrojů
 - Telemikroskopie na University of Southern California
- Řada dalších aplikací

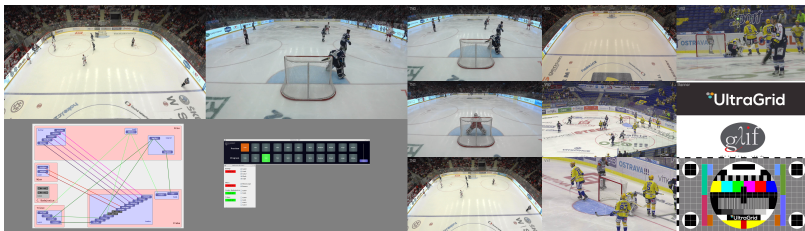
- Ale celé to není jen o velikosti obrazu a jeho rozlišení...

■ Umění

- distribuovaná představení: hudba, tanec, divadlo



- Aplikace s vysokou snímkovou frekvencí



- Aplikace s velkou barevnou hloubkou

- až. 16 b na barevný kanál
- aplikace ve filmovém průmyslu, přenosy HDR videa nebo v např. v patologii

- Co znamená vysoká kvalita přenosů na síti?
 - Datové toky pro nekomprimované video [Gb/s]

Rozlišení	30fps, 8b	60fps, 10b	120fps, 16b
HD - 1080p (1920×1080)	1.4	3.5	11.1
4K - 2160p (3840×2160)	5.6	13.9	44.5
8K - 4320p (7680×4320)	22.2	55.6	177.9

- Sub-sampling barevného prostoru
 - 4:2:2 = 2/3 datového toku
 - 4:1:1 or 4:2:0 - 1/2 datového toku

- Potřebujeme nekomprimované video?
 - Ve většině případů samozřejmě ne!
 - lidský zrak má omezené schopnosti
 - pro archivní účely lze použít bezztrátovou kompresi (ta ale poskytuje jen omezené kompresní poměry)
 - Nízkolatenční aplikace
 - cílem je minimalizovat end-to-end zpoždění při zachování kvality obrazu
 - dostatečná propustnost (de)komprese
 - typicky pouze intra-frame komprese (nebo inter-frame komprese s jednosměrnou predikcí)

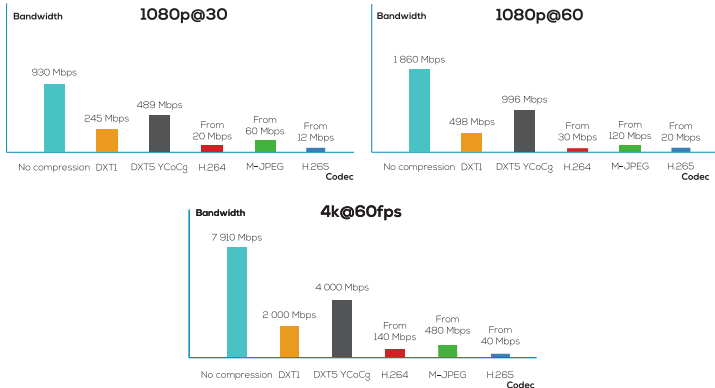
■ Co jsou interaktivní aplikace?

- Streaming vs. interaktivní aplikace v reálném čase
- Typická omezení jsou daná lidským vnímáním zpoždění
 - ITU-T G.115: 150 ms zpoždění jednosměrné audio komunikace přes telefonní linku
 - v případě některých aplikací lze tolerovat 400ms RTT (experimenty se vzdáleným ovládáním robotů na operačních sálech)
 - jiné aplikace jsou mnohem citlivější, např. vzdálené, distribuované koncerty vyžadují teoreticky zpoždění <35ms
- Otázky synchronizace obrazu a zvuku
 - většina lidí pozná desynchronizaci obrazu a zvuku o 100 ms a více (lip-sync)
 - hudebníci jsou mnohem citlivější

■ Technologie

- dostupná platforma pro vysoce kvalitní, interaktivní přenosy obrazu a zvuku vyvíjená i v laboratoři Sitola na FI MU
- švýcarský nůž pro vysoce kvalitní přenosy videa a audia
- postavená na běžně dostupném (herním) HW
 - podpora Linuxu, MacOS X a Windows
 - běžně dostupné karty pro zahytávání obrazu a zvuku
 - běžné grafické karty
 - 10GE síť je výhodou, ale určitě není nutná
- co nejnižší zpoždění na běžně dostupném HW
- open-source software, BSD (GPL) licence

- Přenos pomocí protokolů UDP/RTP
- Vlastní rozšíření protokolu RTP, které umožňuje například autokonfiguraci přenosu
- Forward Error Correction
 - Low-density Generator Matrix LDGM
 - CPU a GPU implementace
 - CPU (s využitím SSE) používáme kvůli režii přenosu CPU↔GPU
 - packet loss do 10% se dá zvládnout s rozumnou režii
 - umožňuje přežít až 20% packet loss při použití JPEG komprese
 - Reed-Solomon codes

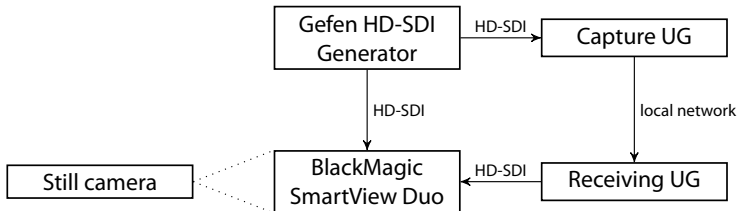


- Propustnost implementace kompresních mechanismů v UltraGridu (Intel® Core™ i7-4960X CPU @ 3.6GHz (6 jader), NVIDIA GeForce GTX 960 GPU)

<i>Komprese</i>	Nejvyšší dosažený výkon
RTDXT:DXT1	4K 60 fps
RTDXT:DXT5	4K 60 fps
GPU DXT	4K 60 fps
GPU JPEG:90	4K 120 fps nebo 8K 60 fps
H.264 (ffmpeg)	4K 30 fps
H.265 (ffmpeg)	4K 24 fps (skoro)
H.264 a H.265 (NVENC)	2× 4K 30 fps

- End-to-end zpoždění (od akvizice po přehrávání) v lokální síti
 - <150 ms pro interaktivitu: ITU-T doporučení G.114
- Video
 - závisí na HW pro akvizici/přehrávání (a kompresi): 2–5.5 snímku videa a snímkové frekvenci → 40–183 ms
 - nejnižší zpoždění přenosu videa **40 ms**
 - nekomprimované 60p video
 - Blackmagic DeckLink 4K Extreme, Desktop Video 1.8.2, Linux
- Audio
 - embedded spolu s videem: synchronizované
 - nezávislý zdroj/přehrávání: nejnižší spolehlivé **15 ms**
 - lze manuálně upravit, tak aby zpoždění audia odpovídalo zpoždění videa

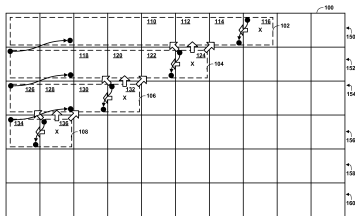
- Metoda měření zpoždění přenosu videa



- Paralelizace enkódování GPU JPEG přes více fyzických GPU
- 2x Nvidia Titan X, výkon 12.1 TFLOPS
- Nvidia M60, výkon až 9.6 TFLOPS
- **Nvidia GTX 1080, výkon až 8.9 TFLOPS**
- Výkon GPU JPEG komprese
 - až 8K, 80 fps



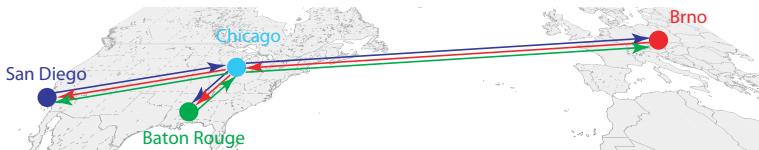
- nástupce H.264/AVC, libx265/ffmpeg
- vysoká míra paralelismu na CPU (Wavefront parallelism a frame parallelism)



- komprese více snímků naráz na úkor zpoždění
- 2 × Intel Xeon CPU E5-2660 v2 @ 2.20GHz
- 1080p – 60 fps nebo 4K – 24 fps
- potřeba akcelarovat na GPU

- Akcelerace enkódování H.264/H.265 na GPU
- GPU Nvidia s jádry (Kepler), **Maxwell** a **Pascal**
- NVidia GTX 9xx – 2x 4K, 60 fps - uměle limitovaný výkon
 - V praxi spíše 2x 4K, 30 fps
 - Limitující faktor jsou paměťové přenosy
- Dobrý bitrate handling

- Poprvé v roce 2005 Brno → Baton Rouge a San Diego
- 1080i60 video, 8bpp, 4:2:2 subsampling = ≈ 1.485 Gbps
- M:N scénář, replikace streamů ve Starlightu, Chicago



- Jumbo frames 8500B
- Chelsio T110, později Myricom Myri-10GE
 - UDP přenos → vždy na rychlosti odesílajícího rozhraní
 - Výkon Chelsio T110 byl ale pouze cca 6 Gbps!

- Bursty
- Netýká se jen nekomprimovaného videa! Např. 1 frame u H.264 videa.
- Nutný traffic shaping → malý odesílací buffer navíc

- Základní parametr je zpoždění přenosu
 - Maximálně 35 ms end-to-end pro symfonický orchestr
 - Záleží na stylu a tempu hudby (a nutnosti tempo přesně držet)
 - Vážná hudba (do 35 ms) vs jazz (do 100 ms) vs tanec (ne nutně synchronní s hudbou) vs činohra (do 150 ms)
- Zpoždění zásadně ovlivňuje schopnost synchronizace v obou směrech

■ Kde se zpoždění zvuku bere

- Fyzikální zpoždění – 1 m fyzické vzdálenosti = ≈ 3 ms
- Bufferování zvuku při zpracování
 - Nutné zlo, bez toho to vůbec nejde
 - 44100 Hz vzorkovací frekvence = 0,023 ms/vzorek
 - S tím nelze pracovat (např. context switch může trvat třeba 0,01 ms) → musíme zpracovávat více vzorků naráz
 - 128 vzorků = ≈ 3 ms naráz

- Kde se zpoždění zvuku bere – přenos
 - (Alespoň v UltraGridu) pipeline
 - Akvizice → paketizace → rekonstrukce z paketů → playback = >12 ms
- Kde se zpoždění zvuku bere – HW
 - Analogový HW bez zpoždění
 - Digitální HW? DANTE protocol. Virtuální zvuková karta pro DANTE má zpoždění ≈ 10 ms!

- Kde se zpoždění zvuku bere – síť

```
--- www.cesnet.cz ping statistics ---
```

```
100 packets transmitted, 100 packets received, 0.0% packet loss  
round-trip min/avg/max/stddev = 5.660/5.885/6.139/0.082 ms
```

```
--- www.ucsd.edu ping statistics ---
```

```
100 packets transmitted, 100 packets received, 0.0% packet loss  
round-trip min/avg/max/stddev = 164.126/164.385/167.777/0.370 ms
```

- Zurich University of the Arts spolu s Run Run Shaw Creative Media Center na City University Hong Kong
- End-to-end zpoždění zvuku 135 ms síť + 30 ms přenos a zvukový HW v sálu → skutečná hudební interakce není možná



- Posunuté party
- Jedna strana vždy hraje podle toho, kdo vede
 - Zpětná vazba pro vedoucího je extrémně zpožděná a rušivá

- Co znamená 120fps
 - 1080p 120 fps video = $1920 \times 1080 \times 120 \times 16$ (4:2:2, 8b)
= **3.7 Gbps**
 - Tj. totéž jako 4K 30 fps při stejném bpp a subsamplingu
 - Při 120 fps je framerate 8.3 ms – 4× kratší čas pro zpracování snímku než u 30 fps videa
- Slow motion (slowmo) záznam vs. živé 120fps video (přes SDI nebo HDMI rozhraní)

- K čemu je to dobré?
 - Extrémně věrně zachycené i rychlé pohyby (viz demo)
 - Nízké zpoždění přenosu
 - Snímky videa zpracováváme v rámci pipeline
 - Zpoždění 4 snímků v pipeline = **33.2 ms** zpoždění
 - V tomto případě skutečně **end-to-end zpoždění** od kamery po displej (pokud je displej optimalizovaný pro hry)
- A kde je tedy problém?
 - 2-4× kratší expozice → extrémně málo světla, malý dynamický rozsah

- 8K 60 fps video = $7680 \times 4320 \times 60 \times 16$ (4:2:2, 8b) = **31.85 Gbps**
- 2012 - přenosy 8K videa z předem komprimovaných snímků čtených z diskového pole
- Aktuální výzva
 - akvizice 8K obrazu z existující 8K kamery na komoditním HW
 - Komoditní rozhraní pro akvizici videa
 - aktuální 12G-SDI rozhraní pro přenos videa – 12 Gbps
 - 24G-SDI ve vývoji – 24 Gbps
 - síť
 - architektura PC a zejména PCI-E sběrnice

- **Sony F65 + BPU-8000** od kolegů z PSNC
- 2x **AJA CorVid88** (dohromady 16x 3G-SDI)



- Nvidia M60 accelerator nebo NVidia GTX 1080 (výkon >9 TFLOPS)
- Myricom Myri10GE
- Základní deska vs. fyzická konfigurace PCI-E slotů vs. fyzická velikost karet vs. požadavky karet na počet PCI-E lanes

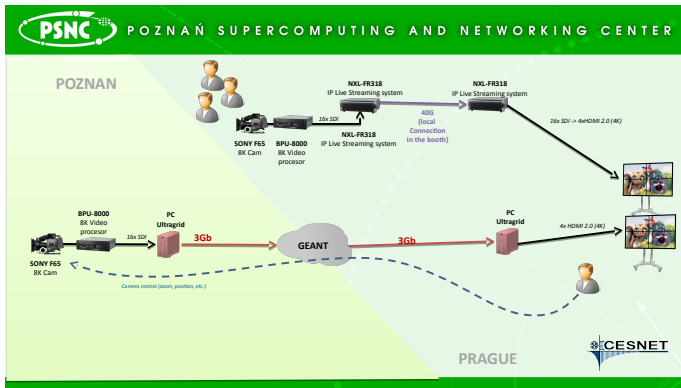


- Akvizice, zpracování a odesílání ve 4 4K, 60 fps dlaždicích
- Paralelní GPU JPEG komprese na Nvidida M60
- Výsledný datový tok cca 3 Gbps

- Akvizice, zpracování a odesílání ve 4 4K, 60 fps dlaždicích
- Paralelní GPU JPEG
- Výsledný datový tok ≈ 3.3 Gbps
- Myricom Myri10GE
- 2x Nvidia Titan X
nebo NVidia GTX 1080
plus NVidia Quadro M6000
- 4 komoditní 4K TV
- Zobrazování 2x4K 60 fps přes OpenGL spotřebuje $>50\%$ výkon GPU



■ Demo na TNC'16



- 9000B jumbo frames!
- Škálovatelnost zobrazování videa
 - Data snímku musí zůstat v paměti GPU (CUDA context) po dekompresi GPUJPEGu a musí být znovu použita v OpenGL contextu
 - CUDA a OpenGL contexty neběží paralelně na architektuře NVidia Maxwell (funguje na architektuře Pascal)

- End-to-end zpoždění přenosu

$$\begin{array}{ccccccc} \underbrace{27.6} & + & \underbrace{50} & + & \underbrace{11.4} & + & \underbrace{2 + 11.6 + 4} & + & \underbrace{6.2} & + & \underbrace{27.3} & = \\ \text{cam+bpu} & & \text{aja} & & \text{enc} & & \text{net} & & \text{dec} & & \text{disp} & \\ \mathbf{140.1\ ms} & & & & & & & & & & & \end{array}$$

Děkuji za pozornost!

<milos.liska@cesnet.cz>
<ultragrid-dev@cesnet.cz>

<http://www.sitola.cz/>
<http://www.ultragrid.cz/>