Performance Optimization of SMPTE 2110 Applications on COTS Hardware

with Network & JPEG XS Offloading to DPUs

Ben Runyan | intoPIX
Gauthier Thieren | intoPIX
Thomas True | NVIDIA

Introduction – the hidden cost of software-driven workflows

New Media Workflows

Move to **COTS hardware:** no purpose-built hardware

Software-based workflows (containerization …)

Deployable regardless of underlying hardware

Limitations

Best-effort design philosophy

Performance issues due to several **bottlenecks**: CPU cores, memory bandwidth, cache hierarchies, and PCIe

Audible artifacts, dropped frames, or **non-compliance** with ST 2110-21 traffic shaping requirements

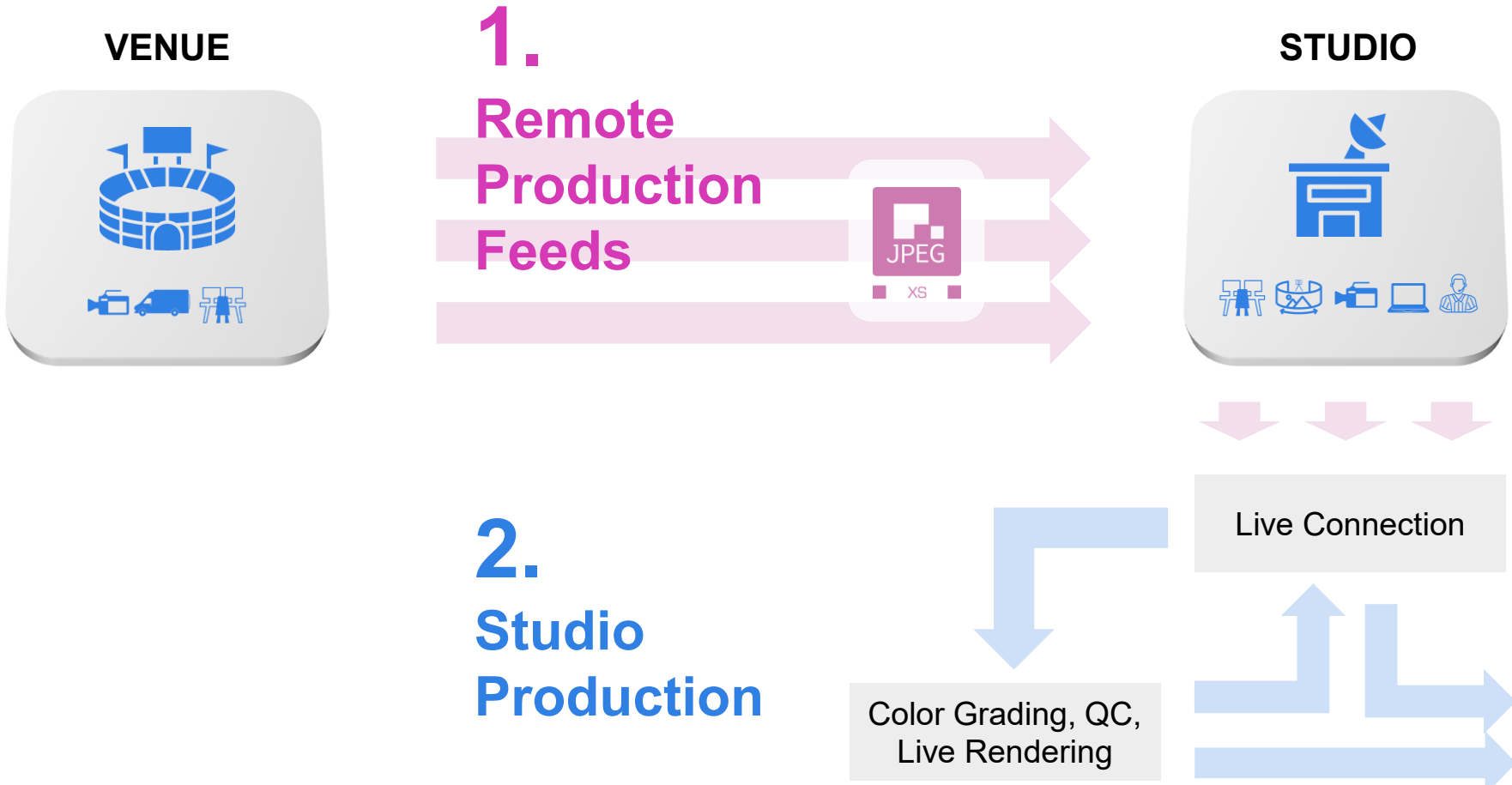
Offload processing

Offload **network-intensive and timing-critical functions** (e.g., ST 2110 & JPEG XS) to DPUs or SmartNICs

Reduce jitter, PCIe traffic, and memory copy **overheads**

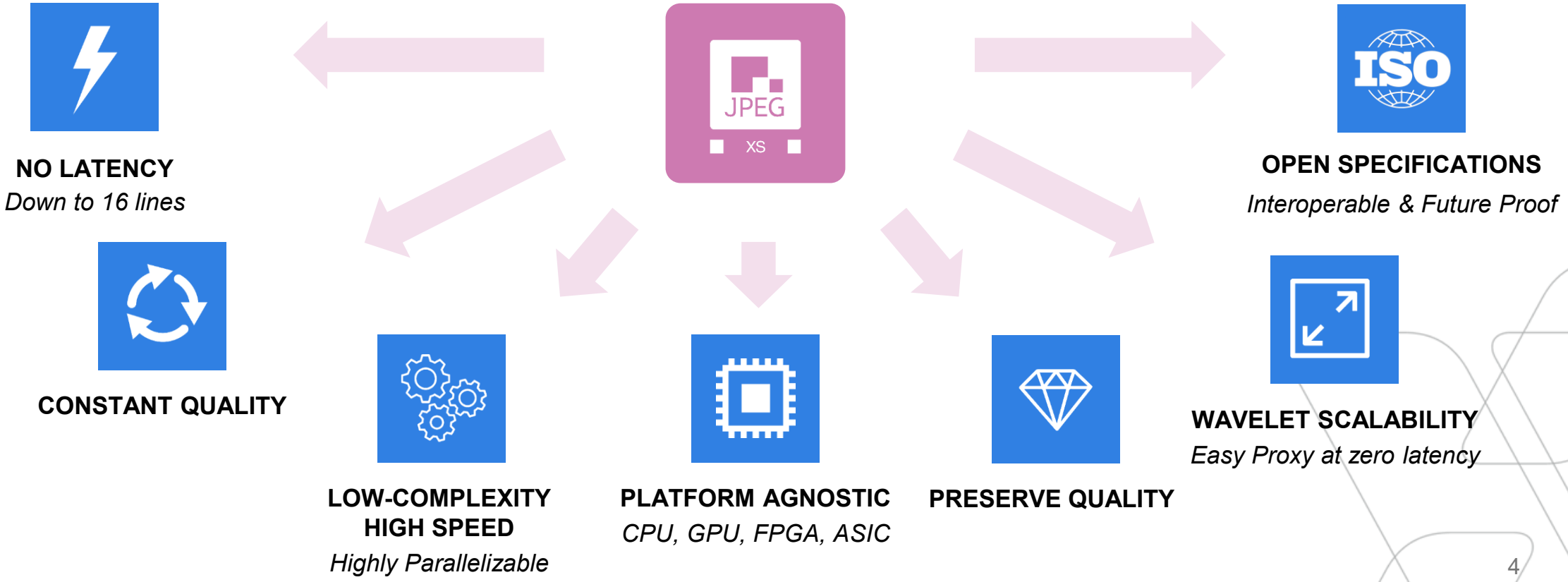
Keep Host & GPU Resources for (AI-driven) media processing tasks

Use-cases – where COTS optimization makes the difference



Why JPEG XS within ST 2110?

The Most Advanced Lightweight Codec to Replace Uncompressed Video



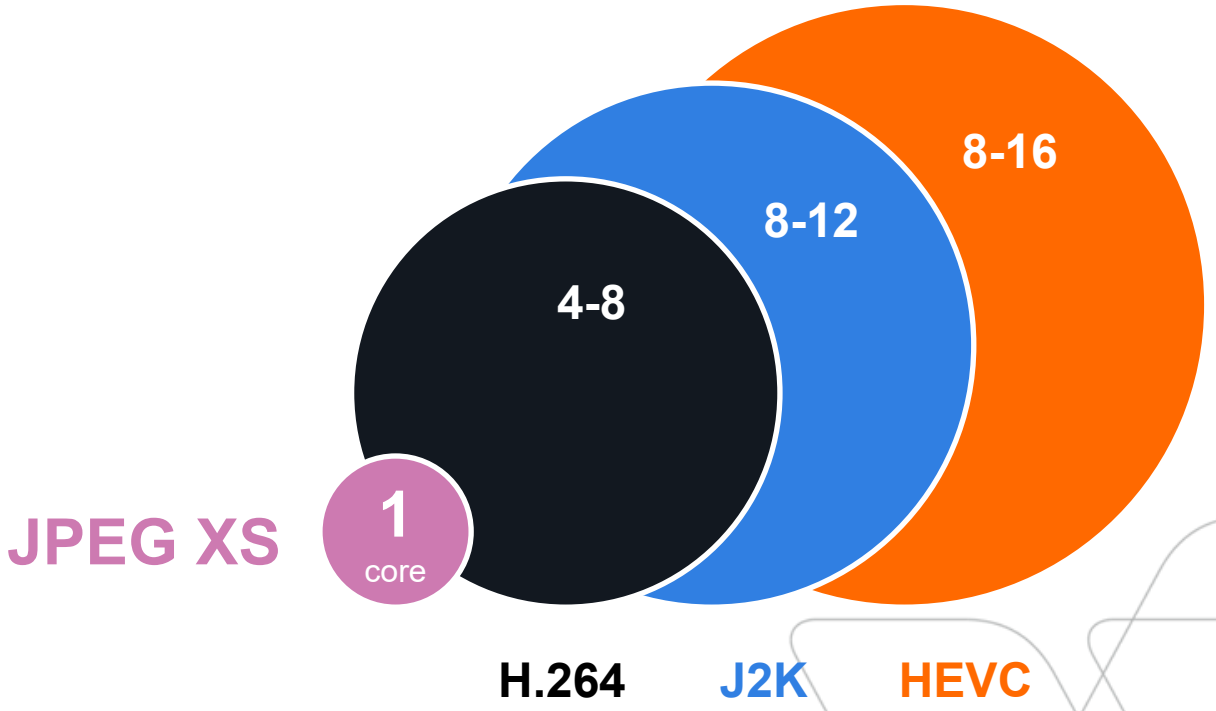
DPU offloading with JPEG XS

Single-Core Encoding

JPEG XS encoding with single-Core performance is hyper efficient – implementation by intoPIX can encode HD video at more than 60fps on a single CPU core.

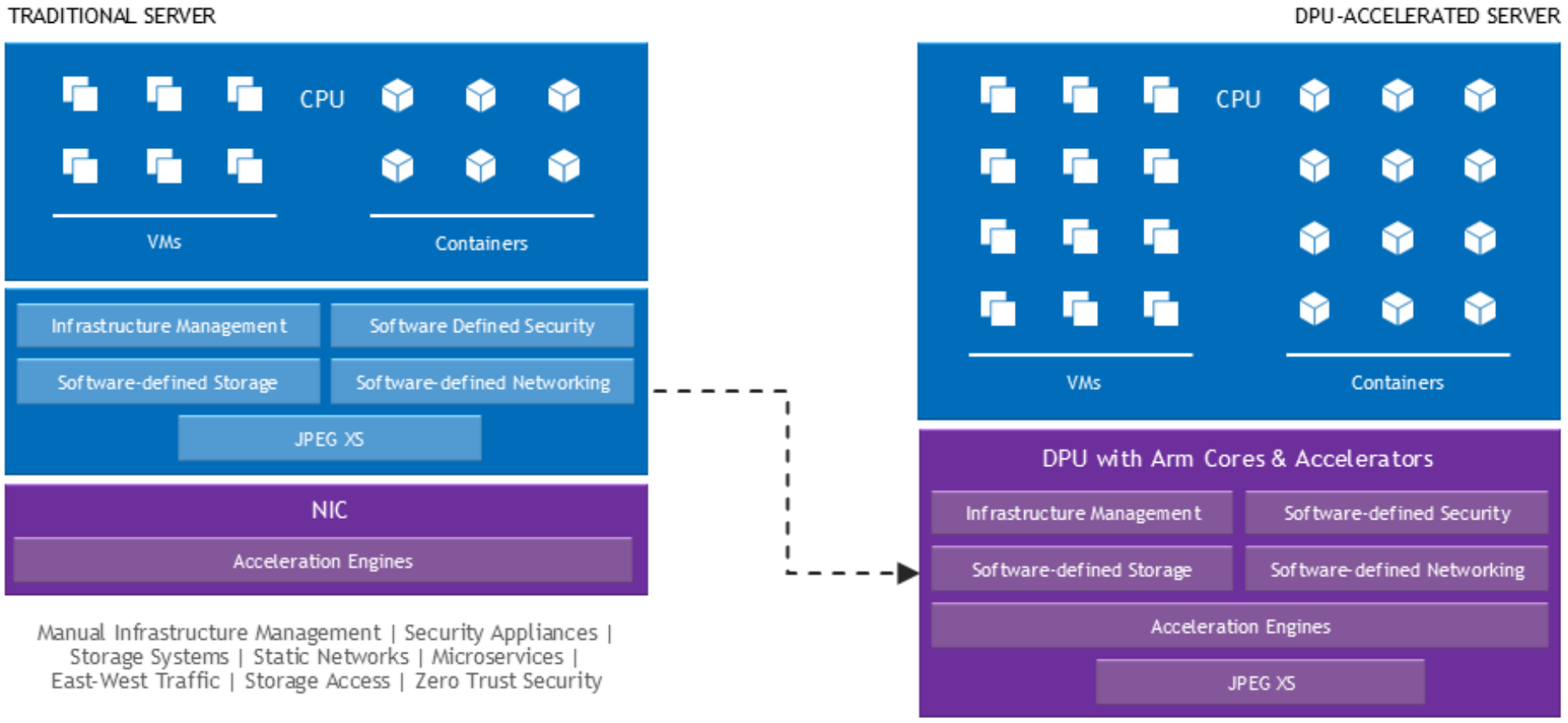
Wavelet Proxy Capability

On top, the wavelet proxy capability bring further CPU consumption benefits at the encoder side or decoder side.



JPEG XS vs. other technologies, for CPU Encoding
Number of Cores required for HD (1080p60) resolution encoding

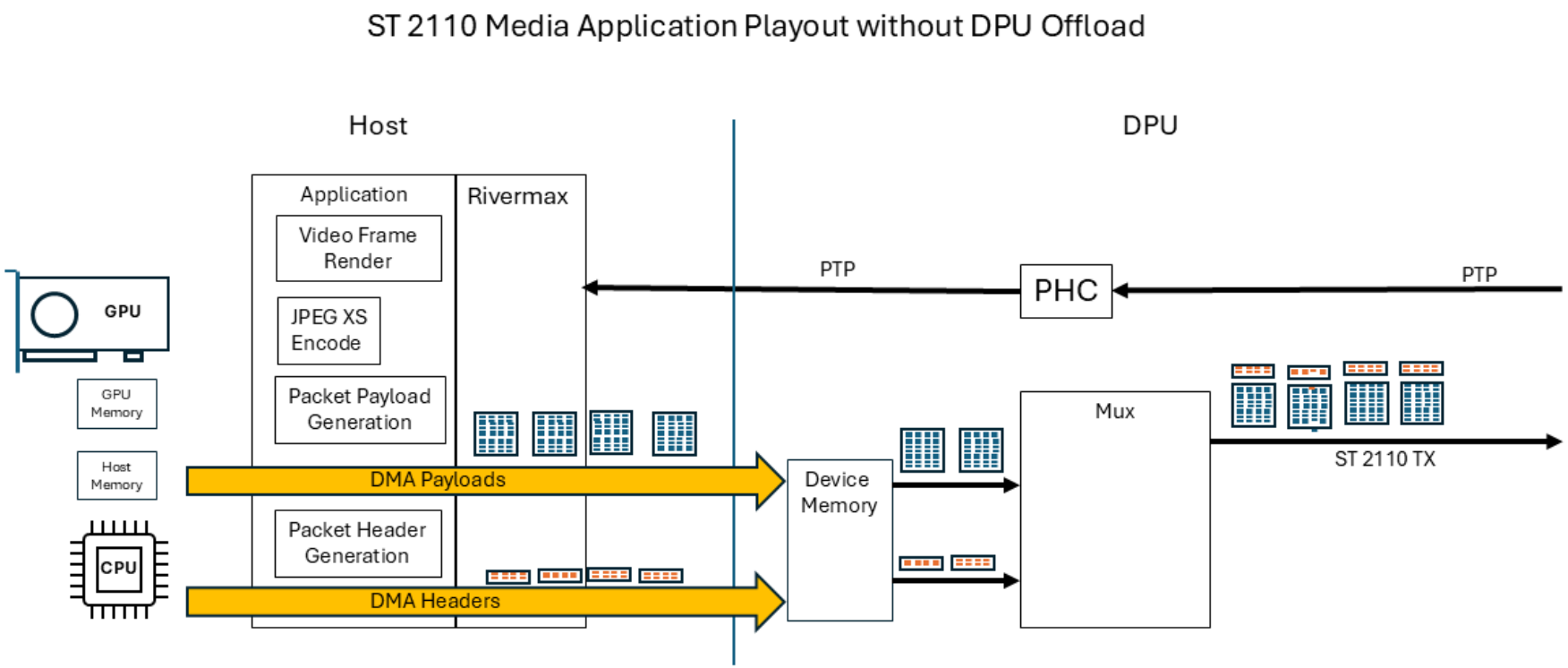
Smart NICs and DPUs



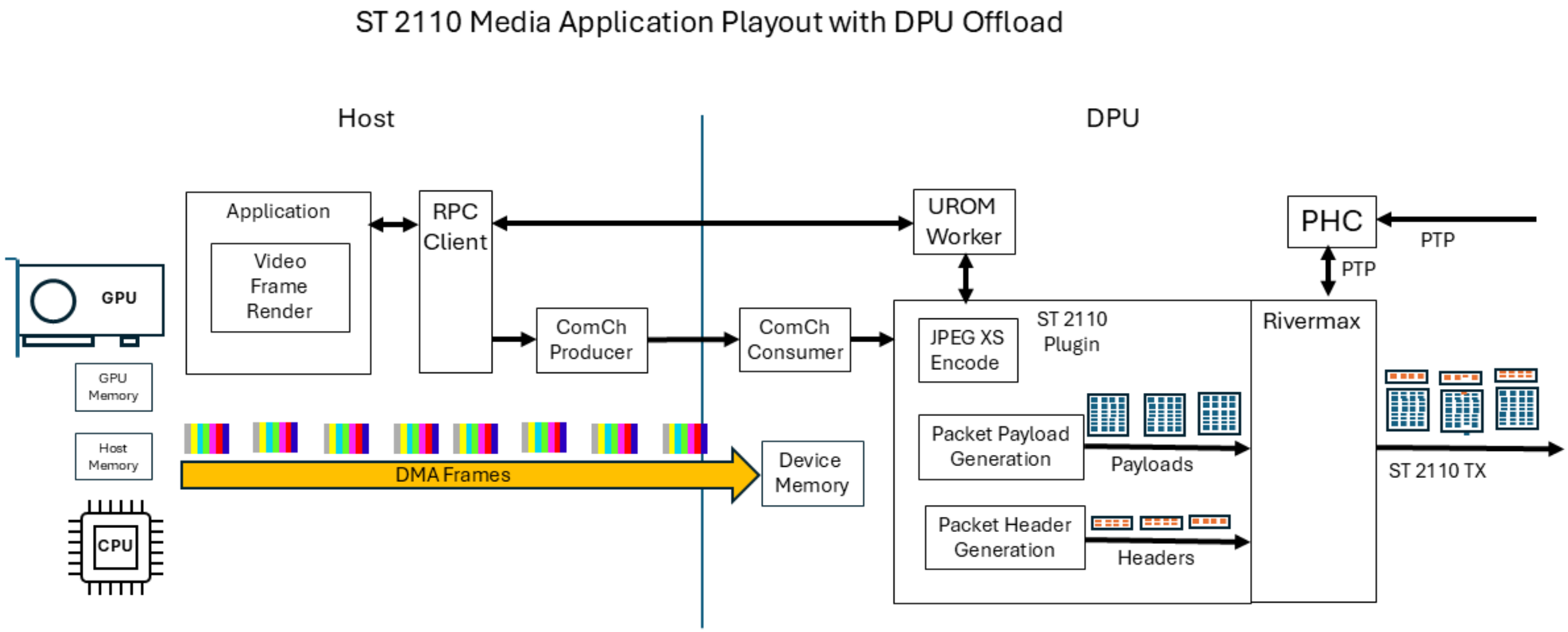
Benefits with DPU Architectural Shift

1.	**More Efficient Data Movement**	Reduction in memory copy operations: **video frames exchanges** instead of smaller groups of network packets
2.	**Reduced and Deterministic Latency**	Constant minimal latency from capture through production to playout: **DPU dedicated resources** (cycles, cache, memory bandwidth).
3.	**Infrastructure Isolation**	**Moving the infrastructure control plane** to the DPU away from the CPU
4.	**CPU Efficiency and Power Usage**	**CPU cycles are freed for application workloads**. Lower both power and cooling requirements.
5.	**GPU Pipeline Efficiencies and Improvements**	**Edge JPEG XS gateway** inside the DPU: ▪ **keep the GPU for high value functions** (graphics rendering, AI inferencing & video processing tasks) ▪ **minimize PCIe bus traffic** using GPU Direct DMA transfers.

Traditional Media Application Playout



Media Application Playout with DPU Offload



First Results

Setup

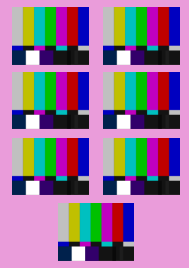
NVIDIA BlueField 3 DPU

- PCIE Gen 5 x16 card
- 16 Arm A78 Hercules cores (2.3 GHz)
- 8 MB L2 cache
- 16 MB LLC system cache
- 32 GB of DDR5
- Ubuntu 24.04 OS

Performance

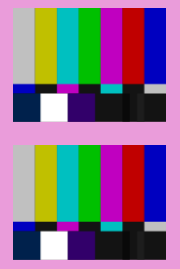
7x 1080p60
10-bit YCbCr / RGB

300Mb Streams
(2 Cores / Stream)



2x 2160p60
10-bit YCbCr / RGB

750Mb Streams
(4 Cores / Stream)



Core Assignment

2 Cores for OS

2 Cores for Each HD Stream:

- Core 1: Frame Transfer from Host + ST 2110-22 TX
- Core 2: intoPIX JPEG XS Encode

The intoPIX SDK performances are regularly improved and new features added. Contact intoPIX to access the latest performance report or to receive updates on the latest releases.

Conclusion

Balance Media and Networking

Keeping **Network & JPEG XS encode as close as possible to the NIC** is the most efficient approach to prioritize high-value media functions.

System Efficiency

Resource saving, Power saving, Cost saving.

Host / GPU Processing	DPU Offloading	DPU Offloading with JPEG XS
Competition for Resources: challenge to balance resources and be hardware-agnostic	**Segmentation of Network/Media functions** Critical requirements for **real-time broadcast applications:** ▪ Data movement reduction ▪ Latency determinism ▪ infrastructure isolation ▪ overall system efficiency	**All advantages with DPU Offloading…** **… with JPEG XS:** ▪ **Easy Integration** in DPU-based architectures ▪ **preserving GPU & CPU** resources ▪ **1/10 bandwidth:** lower Ingress / Egress costs ▪ Free up bandwidth for **server-to-server AI processing**

NAB SHOW

Any Questions?

Thomas True
NVIDIA

Gauthier Thieren
intoPIX

Ben Runyan
intoPIX