megapixel

A SMPTE ST 2110

Virtual Production Cookbook

2025-03-13

Contents

Contents

About Megapixel

Megapixel is the unrivaled authority in cutting-edge professional display technology and video processing for top-tier brands, creatives and design-led homes. Led by Jeremy Hochman and Keith Harrison, our unrivaled team of engineers and designers consistently delivers the most unique and breakthrough LED solutions to market, helping our visionary clients bring their ideas to life in ways that inspire a sense of wonder and make the seemingly impossible possible. With over 300 patents and award wins from Live Design, the Emmys, and the Oscars, we endeavor to always be at the forefront of digital displays and technology for which we set the bar as the industry standard.

We are the unrivaled authority in cutting-edge professional visual technology and video processing for top-tier brands, creatives and design-led homes.

Megapixel
28316 Constellation Road, Valencia, CA 91355

About NVIDIA

Since its founding in 1993, NVIDIA has been a pioneer in accelerated computing. The company's invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined computer graphics, ignited the era of modern AI and is fueling industrial digitalization across markets. NVIDIA is now a full-stack computing company with data-center-scale offerings that are reshaping industry.

1

Introduction

Live media production continues to embrace software-defined SMPTE ST 2110-based workflows built upon Commercial-Off-The-Shelf (COTS) based hardware systems that combine the Input/Output capabilities of Network Interface Controllers (NICs) with the media processing power of one or more Graphics Processing Units (GPUs). One such application is in-camera visual effects (ICVFX) in virtual production, the technique of capturing GPU-rendered visual effects composited with live action in-camera within an LED volume as shown below. This paper outlines the hardware and software required to build a virtual production stage today and how these technologies work together to make virtual production possible.

A significantly better and more efficient virtual production stage can be built around the SMPTE ST2110 set of standards.

Virtual Production Volume
Image Courtesy of Universal Pictures


megapixel

2

On-Set Virtual Production

On-Set Virtual Production (OSVP) has its roots in virtual sets. In a virtual set, the live action and on-set physical objects are placed in front of a green or blue background screen and captured by a camera with compositing with the background plate performed via a chroma-key operation taking place downstream in the workflow from the camera. Virtual sets are widely used in television and movie production to reduce the cost of producing real sets.

In contrast, in OSVP the actors are placed into a virtual world that is displayed on an LED screen or LED "volume" that surrounds the physical set. As a result the virtual world is an integral part of the physical set and the live action captured by the camera, even down to providing the illumination of the scene. This allows actors and on-set objects to be truly immersed in the virtual world. This is known as In-Camera Virtual Effects (ICVFX) and allows the camera to capture natural subtle cues such as lens distortion, depth of field effects and lens flares.

Traditional Virtual Production Stage
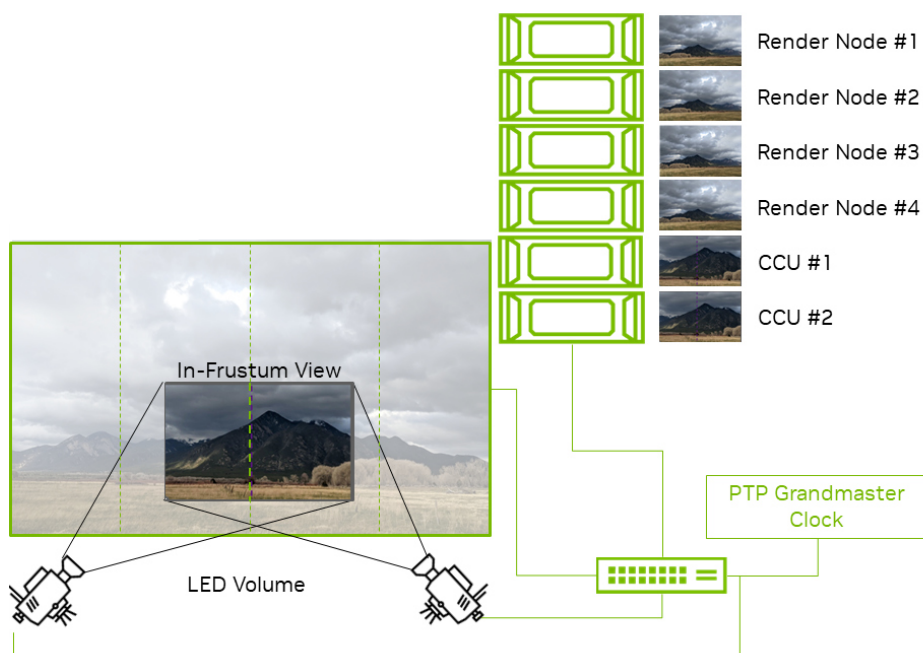Image Courtesy of TnT Productions
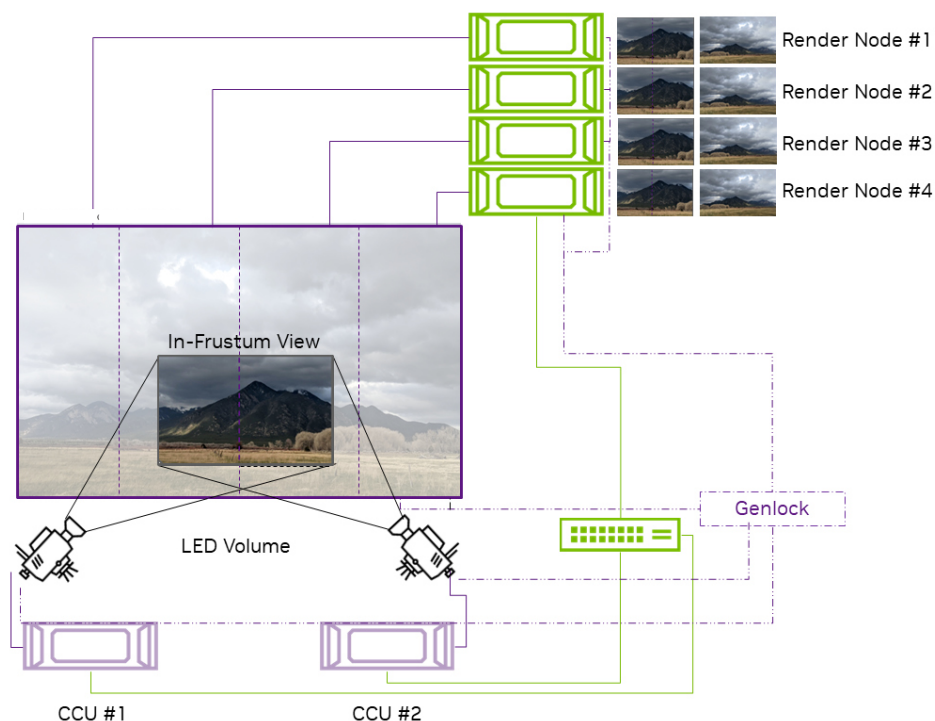
3

SMPTE ST 2110

SMPTE ST 2110 is a family of industry standards for the real-time transmission of media essence streams and accompanying metadata over IP networks. These standards provide an accurate timing model and flow control to prevent network overloading as well as provide the capability for accurate frame-based switching.

Transition from SDI to SMPTE ST 2110

The system architecture of a traditional virtual production stage uses a mix of Serial Digital Interface (SDI) video, Ethernet, HDMI and DisplayPort video transport and display technologies. SDI-based cameras connect to camera control units (CCUs) in addition to position and view tracking systems, so each render node has access to the position and view for all cameras over the network. Each render node requires 2 GPUs, one renders the high resolution inner view frustum for each camera and the second generates the lower resolution background and composites the inner view frustum on top. Each render node then drives its portion of LED volume via the LED processor using HDMI or DisplayPort. This setup is illustrated below.

ST 2110 Virtual Production Stage
Image Courtesy of TnT Productions

Render Node #1
Render Node #2
Render Node #3
Render Node #4
CCU #1
CCU #2

In-Frustum View

LED Volume

PTP Grandmaster Clock

The problem with this architecture is that it relies on a collection of different technologies and doesn't scale. It works OK for one camera but to support two cameras typically the quality of the inner frustum render must typically be reduced as one GPU must now render two inner frustum views.

A significantly better and more efficient virtual production stage can be built around the SMPTE ST2110 set of standards, all the components connect through the network switch. With this approach the CCUs become another instance of a Render Node built on COTS hardware and can be configured through software defined functionality.

The cameras connect to their CCU render nodes via ST2110 and each renders the inner view frustum for the connected camera. The rendered inner view frustums are then multicast via ST2110 over Ethernet to Render Nodes driving each LED region, which composite the inner view frustums onto the background and then use ST2110 to send the composited result to the LED processor for that region. Synchronization between all cameras, render nodes and LED processors is achieved by using a Precision Time Protocol (PTP) Grand Master Clock. Current Unreal Engine 5.2 implementation supports SMPTE ST 2110 multicast emission of the inner frustum view to all render nodes while nDisplay render nodes still leverages standard DisplayPort or SDI to drive the displays. Future versions will allow nDisplay render nodes to feed LED walls with SMPTE ST 2110 synchronized output as shown in the diagram below. In this case, SMPTE ST 2110 replaces all legacy SDI, HDMI and DisplayPort cabling simplifying the cabling and removing the complexity of fiber optical extenders to overcome length restrictions of HDMI/DP.

The remainder of this paper will dive into the technical details on the hardware and software required to build a ST2110-based virtual production stage.

4

Hardware Systems

COTS Servers

SMPTE ST 2110 converged infrastructure builds upon COTS servers which offer many advantages in terms of performance and flexibility over traditional bespoke video equipment. A key demonstration of both this flexibility and performance is Direct Memory Access (DMA) which is a high-bandwidth, low-latency transfer operation between the key components inside the system using the Peripheral Component Interconnect Express (PCIe) architecture. The bandwidth of the PCIe interconnect far exceeds previous bespoke interconnects. As mentioned above these COTS servers are used for both render and camera nodes within the overall system architecture and allow full software-defined configuration and control.

GPU

Each render and camera node contains one or more NVIDIA RTX professional GPUs for video image processing, 3D rendering, real-time ray tracing, AI inference and visual simulation. Based upon the new NVIDIA Ada Lovelace GPU architecture, the latest RTX GPUs such as the NVIDIA RTX™ 6000 Ada Generation GPU deliver unparalleled levels of real-time rendering performance and realism to the virtual production stage. Multiple GPUs are connected via an NVIDIA Quadro Sync® II card for synchronization across multiple connected display outputs.

**NVIDIA RTX 6000 Ada Generation GPU**

NIC

Modern software-defined workflows utilize media processing servers built using COTS NICs such as the NVIDIA ConnectX-6 or ConnectX-7 SmartNICs. Similar to PCIe interconnect bandwidth, NIC hardware bandwidths far exceed bespoke system interconnects and currently offer anywhere from 100 Gb/s to 400 Gb/s between high-performance servers. To put this in perspective this is between 8 to 32 times more bandwidth than afforded by 12G SDI links and clearly offers tremendous benefit for media processing workflows. As bandwidths have improved exponentially, so has the processing power of the NIC, which has also driven the emergence of DPUs.

DPU

The data processing unit is a new class of processor that is optimized for data center infrastructure workloads. NVIDIA BlueField-2 DPU enhances data center performance and efficiency by offloading, accelerating, and isolating data center infrastructure services at speeds up to 200Gb/s. Supporting PCIe 4.0 with full backward compatibility, BlueField-2 is designed to fit into most high-volume enterprise servers, and provide accelerated, programmable network connectivity. The use of BlueField-2 in virtual production is specifically interesting thanks its unique in-hardware precision timing capabilities, supporting SMPTE ST 2110 as detailed later in this document.

NVIDIA BlueField-2 DPU
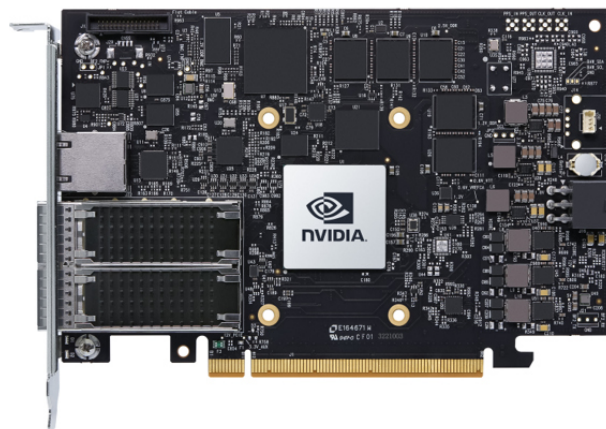
5

LED Processor

LED screens are controlled by an LED Processor that takes the output from a GPU or SmartNIC, either in its native form or via converter if the output of the GPU doesn't match the input of the processor. For example, Megapixel's HELIOS® LED Processing Platform supports SMPTE ST 2110 natively without any additional hardware allowing users to take an input directly from a NVIDIA ConnectX SmartNIC or BlueField-2 DPU. As the workflow is IP-based, the distance limits of legacy formats such as DisplayPort or HDMI do not apply and the HELIOS unit can be located much further from the machine generating the virtual environment. This is beneficial as on most OSVP stages, the output device is often at a control location in direct view of the LED screen, whereas the HELIOS may be located behind the screen (or a rack room) in a separate location.

Megapixel's HELIOS® LED Processing Platform supports SMPTE ST 2110 natively without any additional hardware...

The processor takes each video frame it receives and performs a number of tasks with it. Color correction and mapping are common tasks the processor performs. The video raster that is received by the processor often will not line up with the layout of the panels that comprise the screen and some form of layout or mapping is required to correct this. In addition, content may not be in the correct color space and tweaking of colors may be required on both the input and the panel configured so the content is displayed correctly to the camera.

Once HELIOS has been configured – usually once, at the start of production – it will receive a video input and convert this video input into a stream of packets sent to each individual LED panel with video, color and timing data embedded.

Legacy vs Converged Infrastructure
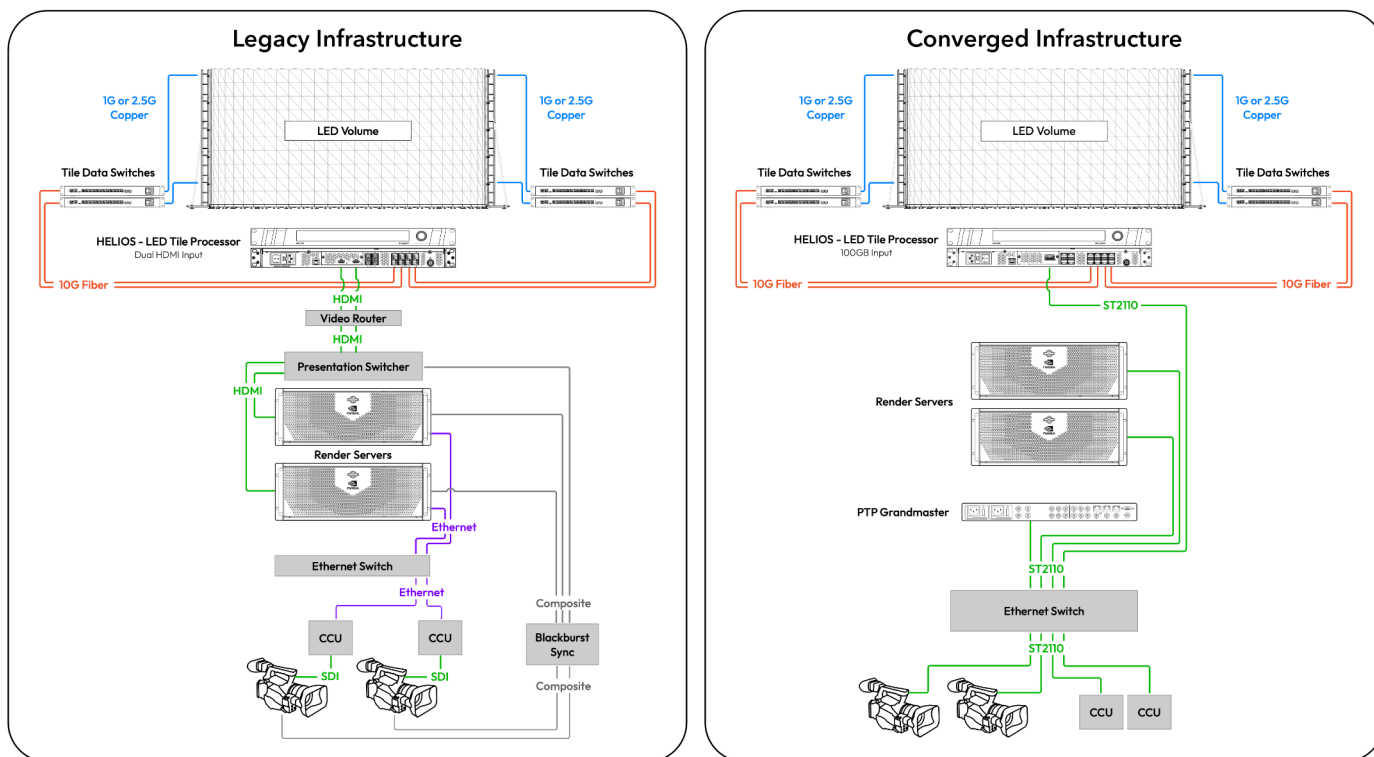
6

LED Wall Panels

Each LED panel has a card inside it that can interpret the data coming from the LED Processor. Receiving cards and processors must match up for the configuration to work. For example, the Megapixel PX1 cards only work with a Megapixel HELIOS Processor. When the panels are purchased, customers must specify to the manufacturer which processing platform they intend to use so the panels can be built with the correct hardware inside. Generally speaking, you cannot mix and match receiver cards with LED Processors.

Most panels communicate with each other and the processor via Ethernet protocol. Most are proprietary to the processor manufacturer. The panels communicate with the processor at a much faster refresh rate than the video content being displayed, allowing for extremely accurate timing of when photons actually exit the LED itself and therefore, when the camera captures this light as an image. This "round robin" communication is critical to get right for virtual production workflows. If the timing of video frame creation, the LED processor passing the frame to the individual panels, and the camera capturing the image are out of sync, the workflow fails showing tearing and blending artifacts. Close and synchronized cooperation between all of the pieces of equipment in the chain is required to pull off what looks like a seamless workflow.

7

Optimized Media Streaming

Rivermax

NVIDIA® Rivermax® in combination with EPIC Unreal Engine, enables studios to enhance flexibility and performance for virtual productions, so creative teams can deliver high-quality, synchronized content to multiple displays.

Rivermax is a unique internet protocol (IP)-based solution for media and data streaming. Rivermax leverages BlueField-2 DPU adapters to deliver media essence streams compliant with the strict timing and traffic flow requirements of the SMPTE ST 2110-21 specification. With the Rivermax plug-in for Unreal Engine, users can now deliver synchronized video over IP and improve manageability of hardware resources.

Rivermax for On-Set Virtual Production delivers three key benefits:

- Low-latency, ultra-high-bandwidth, industry-standard interconnection between render nodes and external video devices.

- An optimized stage architecture to provide the highest visual quality productions while minimizing latency.

- Direct data transfers to and from the GPU for high-throughput low-latency streaming.

8

Orchestration, Control and Connections

NMOS

Operational control is provided by the AMWA NMOS suite of specifications allowing seamless connections between any SMPTE ST 2110 device on the network. Each render node exposes an NMOS Sender for its ST2110-20 output stream. HELIOS exposes an NMOS Receiver to configure each incoming ST2110-20 stream. The NMOS Senders and the Receiver are automatically registered with the NMOS Registry via the AMWA IS-04 Registration API. The Registry's Query API enables the NMOS Controller to discover the Senders and the Receiver on the network. Connection requests from the NMOS Controller to the Senders and the Receiver via the AMWA IS-05 Control API allows to connect the render node output streams to the LED wall controller to form the desired video output configuration (illustrated below). Using a simple NMOS controller UI such as the one from Pebble, anyone can simply connect render nodes to the LED processor as if they were simple physical cables.

NMOS Control
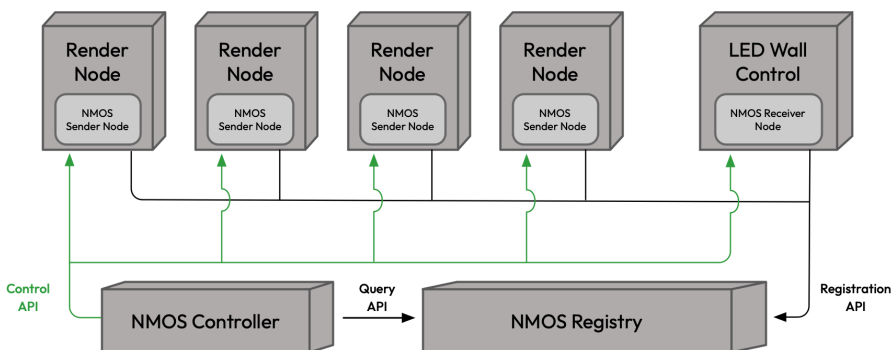
9

Synchronization

Virtual Production would not be possible without synchronization. Ultimately, synchronization ensures the correct photons are emitted from all displays are captured at the correct moment by all cameras. The rendering engine used for the background rendering and the compositing should maintain proper synchronization of both background and camera content that could be split across multiple display tiles. The LED wall finally needs to ensure synchronized display of each of the tiles provided by the rendering engine. If the synchronization breaks at any point of the workflow, tearing artifacts will appear on the LED volume and be visible by the audience watching the content shot on the virtual production setup. We will describe here in detail the multiple places where synchronization occurs and what solution allows hardware PTP accuracy across the entire setup.

PTP Timing as a Service

For proper synchronization, consistent PTP timing is required between all the render nodes. To accomplish this, all render nodes should be within a single PTP domain and use a common PTP profile. However, this still does not guarantee that all PTP followers will behave identically.

Virtual Production would not be possible without synchronization.

PTP Timing as a Service (continued)

This could be due to several elements such as:

● Differences in PTP stack implementations.

● CPU operations that may interrupt the PTP stack thread delaying PTP message processing.

● Misconfiguration of the NIC and CPU resources used by the PTP stack

● Hardware time stamping implementations with different resolutions or rounding methods.

● Differences in the operating system timing capabilities. Specifically, Windows and virtualized environments come with user friendly benefits at the cost of accurate OS timing required for synchronized operations.

For all these reasons, NVIDIA developed the DOCA-Firefly time service, a dedicated software container based on ptp4l Linux software that runs on the ARM cores of the NVIDIA BlueField-2 DPU.

DOCA-Firefly provides the following advantages:

● Accurate timing for Windows, Linux as well as Virtualized environments.

● Operating System agnostic PTP offloading.

● Better tolerance to system configuration, noise and jitter.

● Zero consumption of host resources for PTP.

GPU Synchronization to an External Frame Sync

GPUs implement a pipeline approach to first generate and then display pixels with the rate of the pipeline typically governed by the frame rate of the attached display. In OSVP, a GPU's frame boundary must align with the frame boundaries of other GPUs as well as other devices such as cameras or capture devices. Failure to align frame boundaries results in the GPU updating content while a camera or capture device is recording or capturing that frame creating visual anomalies.

Aligning the GPU frame boundary to the frame boundary of other GPUs and devices requires dedicated hardware so that any initial skew between frame boundaries is removed and alignment is maintained over time. This function is performed by the NVIDIA Quadro Sync II which attaches to one or more GPUs and ensures that all their display frame boundaries are aligned. For any connected GPU, the sync board provides the GPU reference clock signal and monitors the GPU's frame boundaries.

NVIDIA Quadro Sync II
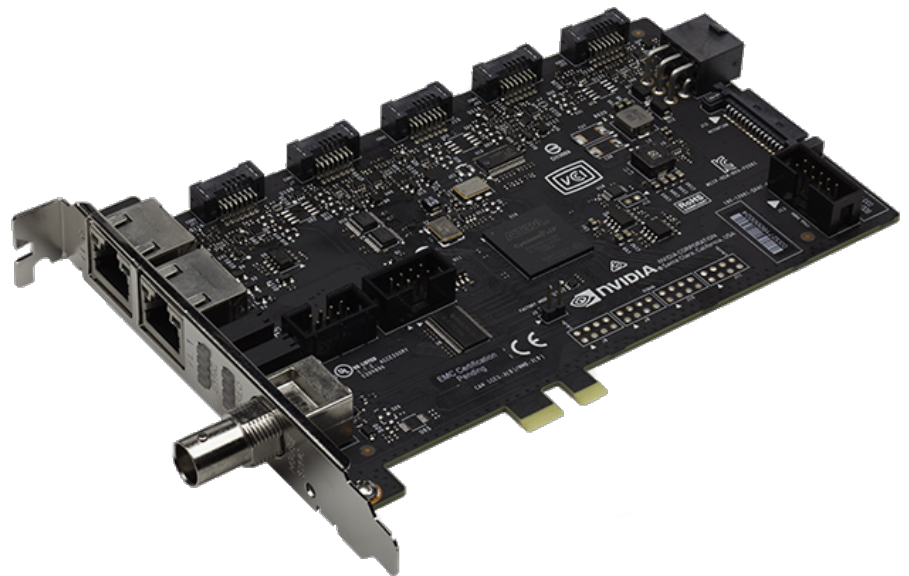
Frame Synchronization

In virtual production the camera shutter and tracking system must all be aligned with the LED tile updates to avoid visual artifacts. This alignment is typically achieved by synchronizing the camera, all the GPUs in the image generators and the LED controllers to the same synchronization source, such as Tri-Level sync. This external synchronization source is distributed throughout a production facility over traditional baseband SDI and is frequently referred to as a house sync. When using SMPTE ST 2110 on converged infrastructure, frame alignment is defined using the Precision Time Protocol (PTP) and embedded in each SMPTE ST 2110 flow.

Temporal Alignment of the LED Tiles via PTP
Location: ARRI Studio London
LED tiles: ROE Visual
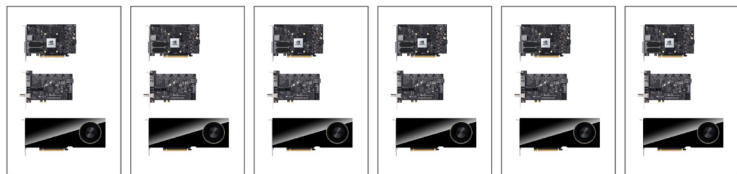Landscape plate: David Noton

LED Display

HELIOS LED Processors

Render Nodes

Network Switch

PTP

Black Burst

PTP Grandmaster

Display Synchronization

A virtual production stage typically comprises many LED tiles grouped into multiple regions and driven by multiple LED processors. Ultimately all tiles must be in temporal alignment so that every tile is showing the same frame and there is no skew or offset between tiles. Likewise all LED tiles must update at exactly the same time to avoid artifacts. To synchronize displays in this way requires that the signal output of all nodes generating pixels that are displayed must be synchronized so that they all deliver the first pixel of any frame to the LED processor for their respective region at exactly the same time. In addition they also must update to the next frame at exactly the same frame boundary across all nodes. This synchronization is achieved using RTX GPUs coupled with Quadro Sync II cards are used in every render node to synchronize the start of the GPU frame to the frame boundaries derived from the PTP time. This technology is a must if you are looking to synchronize physical displays with external synchronization sources such as Tri-Level composite sync as well as synchronize multiple GPU rendering in a single chassis.

All tiles must be in temporal alignment so that every tile is showing the same frame and there is no skew or offset between tiles.

10

Present Barrier

The input resolution of LED processors used in Virtual Production stages is typically 4K, which results in the captured quality by the camera being of a necessary quality. To generate the 4K resolution typically requires powerful GPUs in the image generator nodes to be able to maintain frame rate. The frame rate on every render node is definitely not constant and can vary significantly from frame to frame as the render workload changes with different viewpoints, visible objects, scene complexity etc. This variation in render time between nodes requires what's called a Present Barrier (formerly known as Swap Barrier) to ensure all GPUs move on to the next frame at the same time, despite the fact that some GPUs may have finished rendering significantly sooner than others. Synchronizing with the Present Barrier ensures all down-stream display updates are perfectly synchronized. The image below shows the use of a Present Barrier to synchronize the rendering of 4 GPUs. As a result of the Present Barrier, all GPUs must wait for GPU 3 rendering to complete and because ultimately the GPU 3 render time exceeded the frame time in this case, a frame is missed on the output.

Temporal Alignment of GPU Display Pipelines
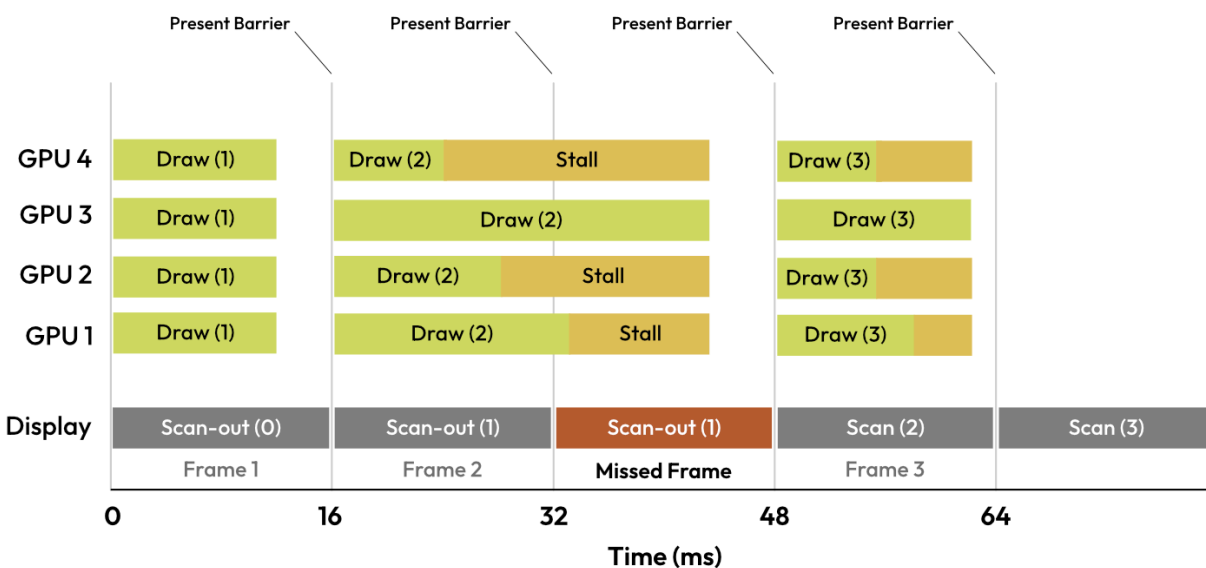
11

Conclusion

The OSVP stage is transitioning to a converged infrastructure based upon COTS hardware and SMPTE ST 2110 network connectivity. Understanding the hardware and software components and how they play together is crucial to a successful deployment.