

# Delivering IMMERSIVE VIDEO Experiences on Intel GPUs

Jill Boyce and Basel Salahieh

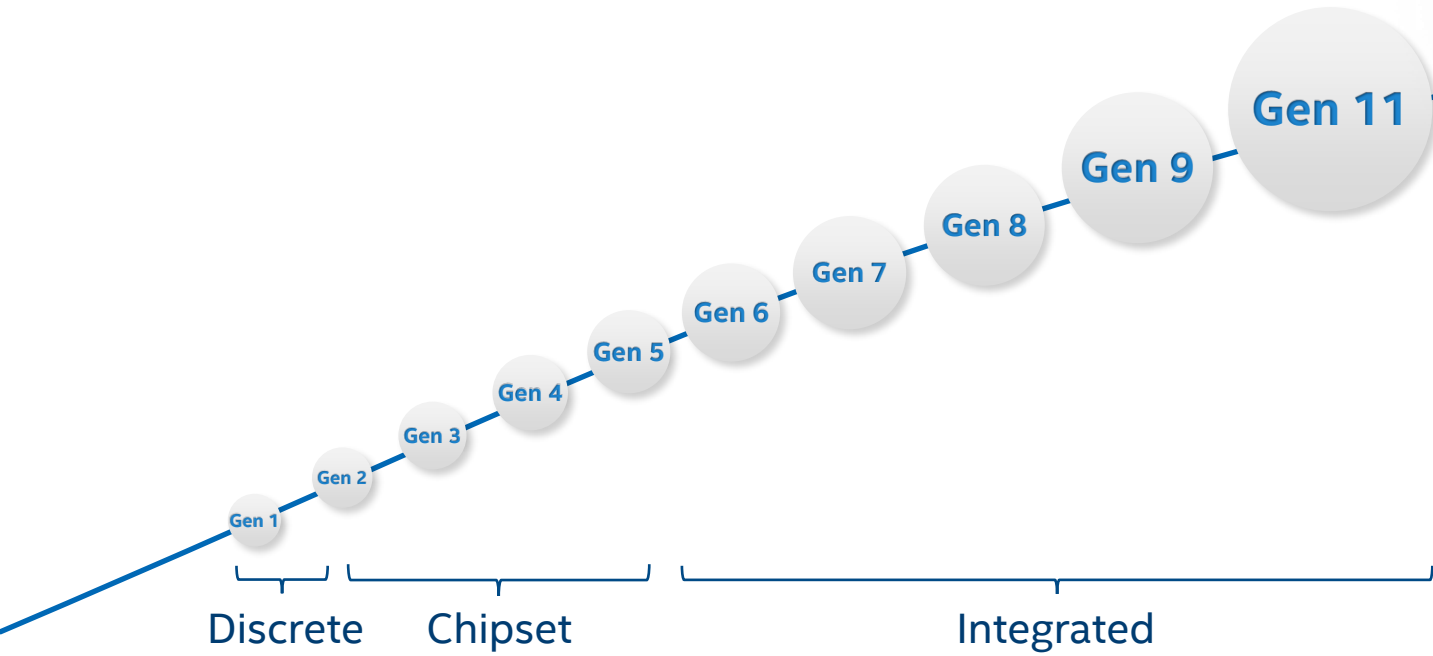
Contributors :

Mengyu Chen, Max Dmitrichenko



intel<sup>®</sup>

# The Road to X<sup>e</sup>





**TIGERLAKE**  
LEADERSHIP INTEGRATED GRAPHICS

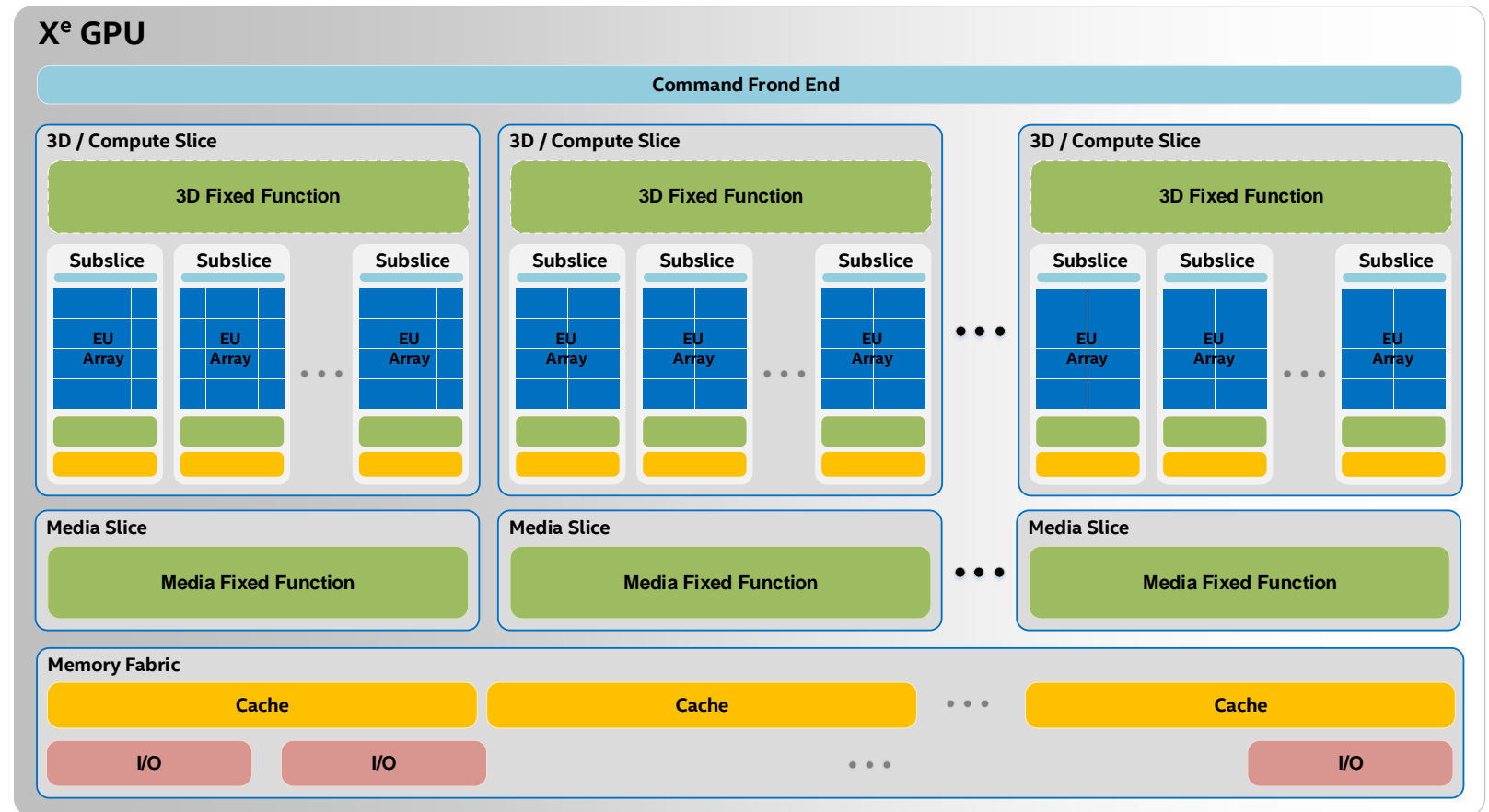
**DG1 "Iris Xe Max"**  
GPU FOR MOBILE CREATORS

**SG1 "Server GPU"**  
VISUAL CLOUD FOR GPU STREAMING

# High Level X<sup>e</sup> Architecture



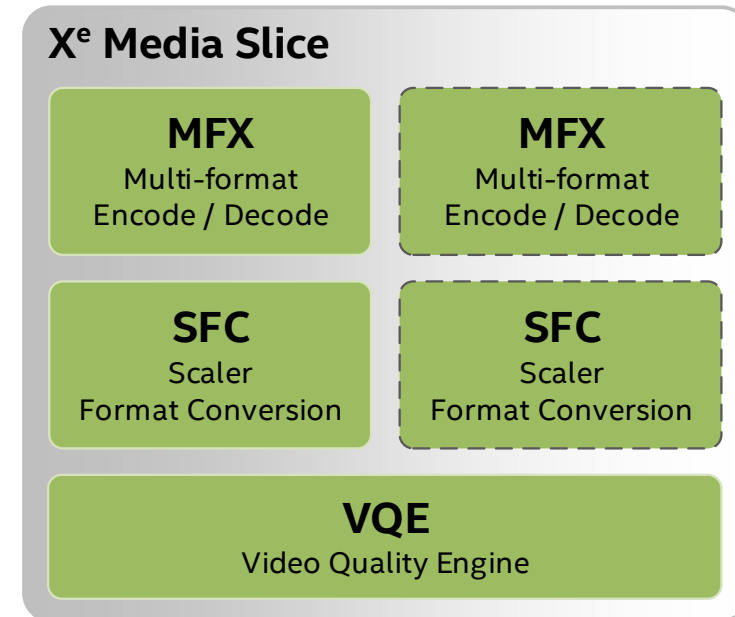
- 3D / Compute Slice
- Media Slice
- Memory Fabric / Cache



# X<sup>e</sup> Media Slice

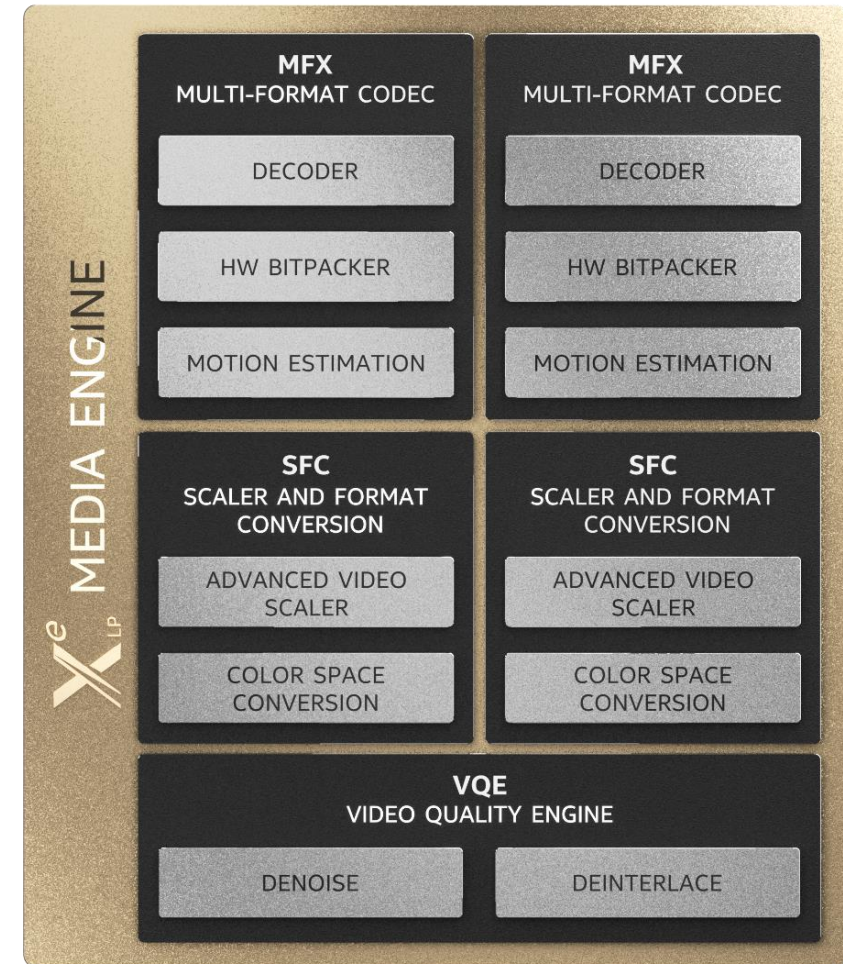


- Fixed function units:
  - MFX - encode / decode / transcode
  - SFC - scaling and format conversion
  - VQE - video quality engine
- Media slices are independent
- Software can distribute a high-resolution stream across multiple slices



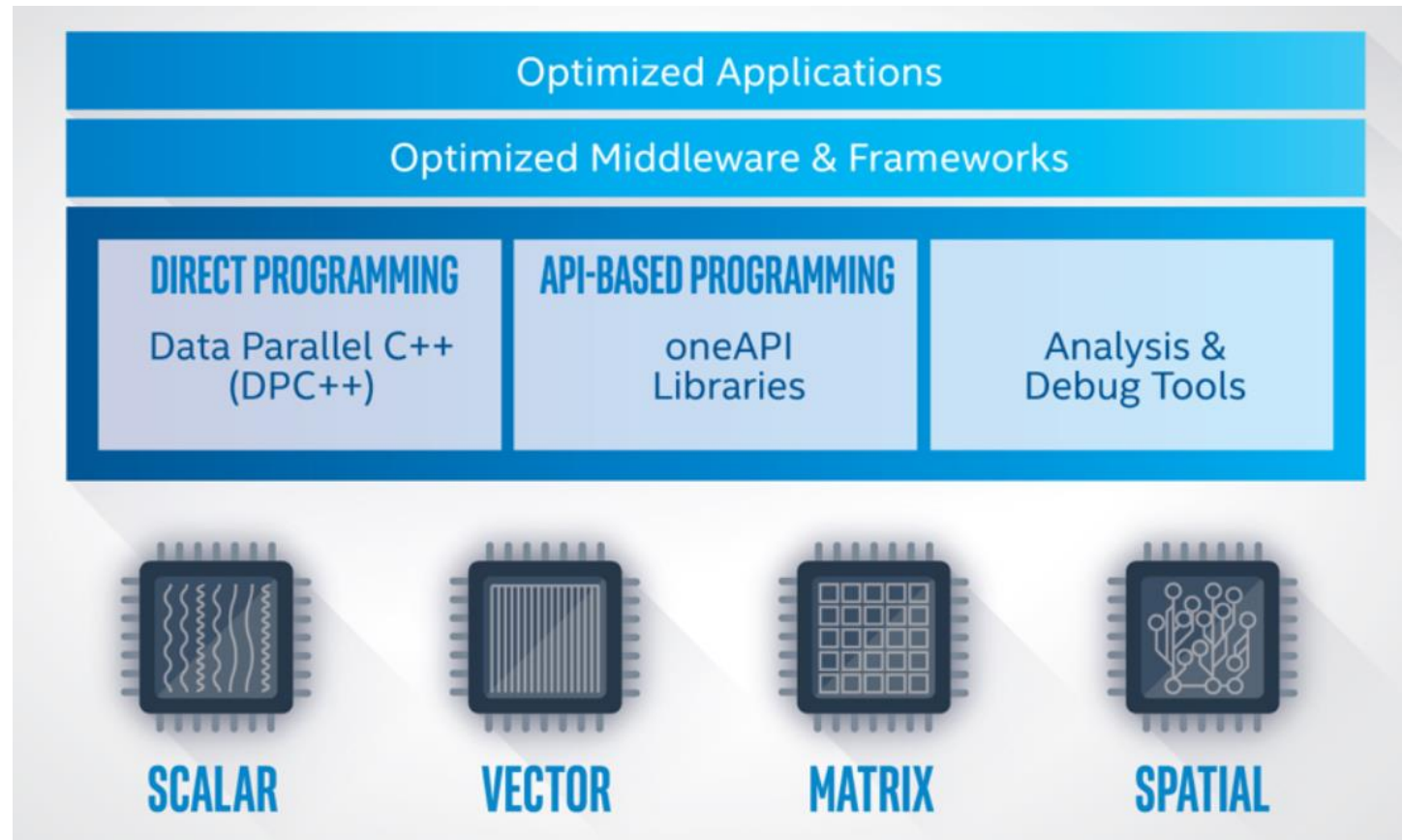
# X<sub>LP</sub><sup>e</sup> Media Engine

- Up to 2x encode/decode throughput
- AV1 decode acceleration
- HEVC screen content coding support (encode/decode)
- Higher quality HEVC encoding
- 8K/4K 60fps playback
- HDR10/Dolby Vision playback



# oneAPI Software Architecture

- oneAPI provides a common developer interface for accelerators, including Media HW through the oneVPL (Video Processing Library)



# Immersive Experience

- An illusory environment that completely surrounds you as if you are part of it.
- Captured as moving images of real people and/or virtual objects that can be viewed from any angle, position, and at any moment in time.

## 6 DEGREES OF FREEDOM



Y AXIS YAW



X AXIS PITCH



Z AXIS ROLL



Y AXIS UP/DOWN



X AXIS LEFT/RIGHT



Z AXIS FRONT/BACK

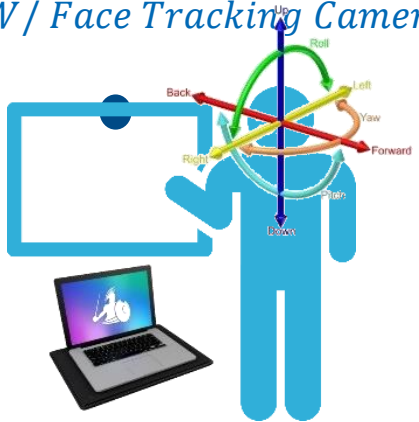


# Consumer Platforms for Immersive Content

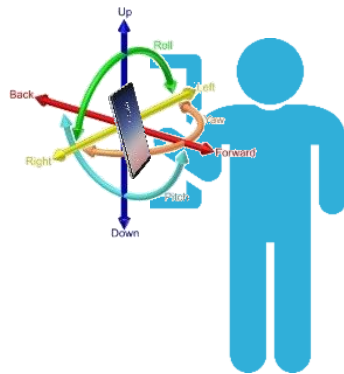
- 6DoF content can be consumed on wide range of devices, many are available today hence, accessible by millions of customers.



*Laptops & Screens  
W/ Face Tracking Camera*



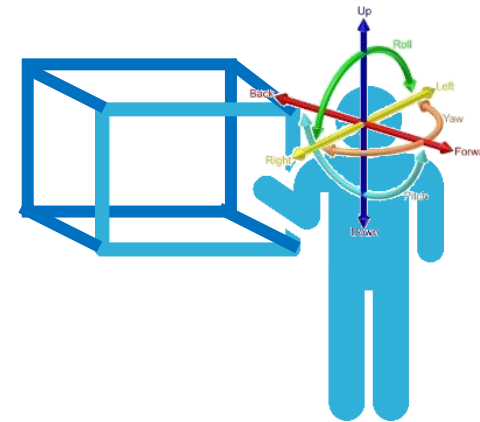
*Phones & Tablets  
W/ Inertial Sensors*



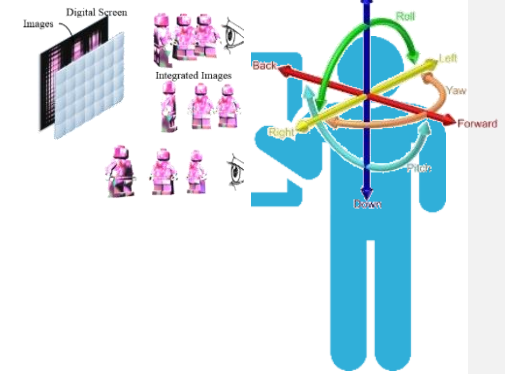
*Head Mounted Display*



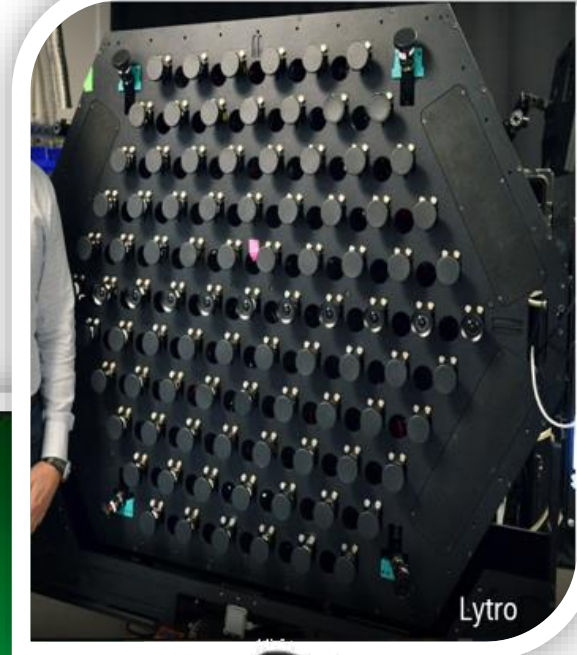
*Volumetric Displays*



*Light Field Displays  
If Dense Sampling*



# Immersive Capture Systems



# Applications

e-learning



Immersive Sport



Immersive meeting



DIY / Instructional videos



Streaming Services



Live events



# Problem @ Hand

- Immersive media requires a lot of data, compute, and bandwidth.
- Currently, client playback of content captured in 6DoF is limited to a rendered 2D video (synthesized along a virtual path) that is streamed & consumed.
- Need to compress the data in order to stream to clients and decode & render to enable 6DoF.



# MPEG I “Immersive” Family of volumetric coding standards

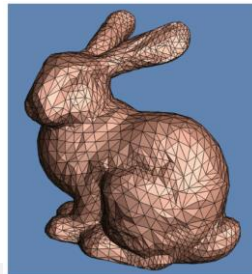
- V3C/V-PCC

- Technical finalization July 2020, FDIS ballot to be submitted
- Encoder inputs point clouds and decoder outputs 3D recon.
- Separation of point cloud specific aspects (V-PCC) from general projection/atlas/patch coding (V3C)



- Future mesh extension for V3C

- In exploration phase, no timeline



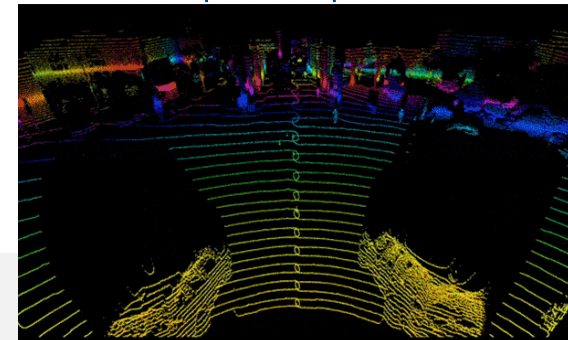
- MIV

- Based on V3C
- Encoder inputs multiple camera views with any capturing configuration (real or virtual) and renderer outputs viewports
- Committee Draft issued July 2020, technical completion mid-2021



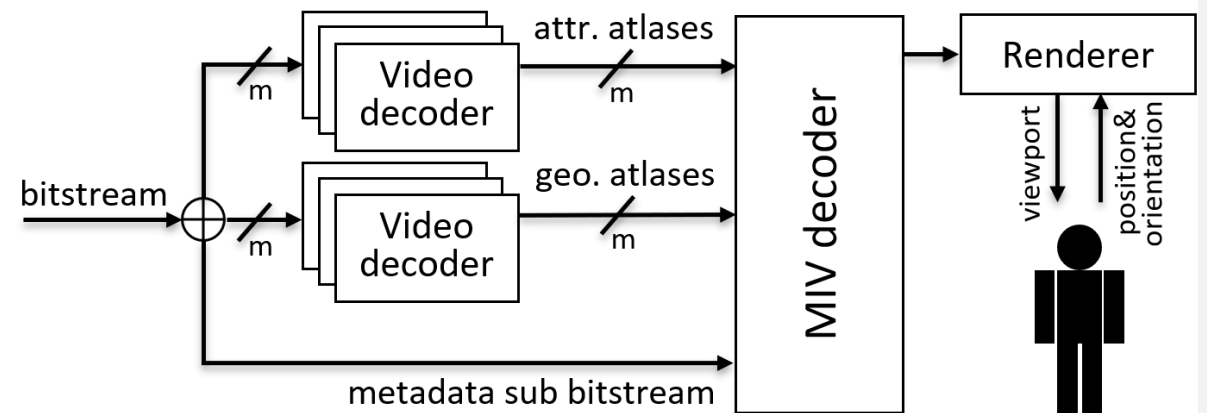
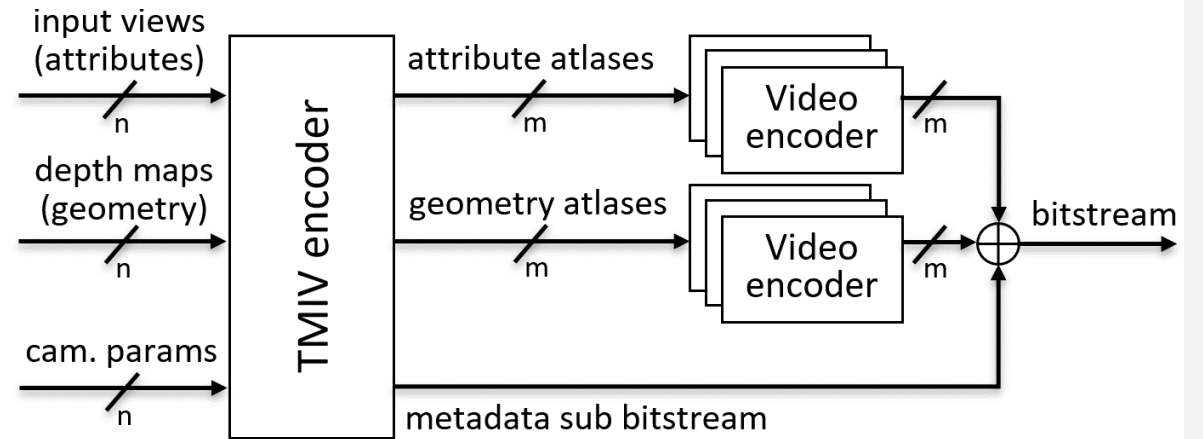
- G-PCC

- Technical finalization 1<sup>st</sup> half of 2021
- Intended for sparse point clouds

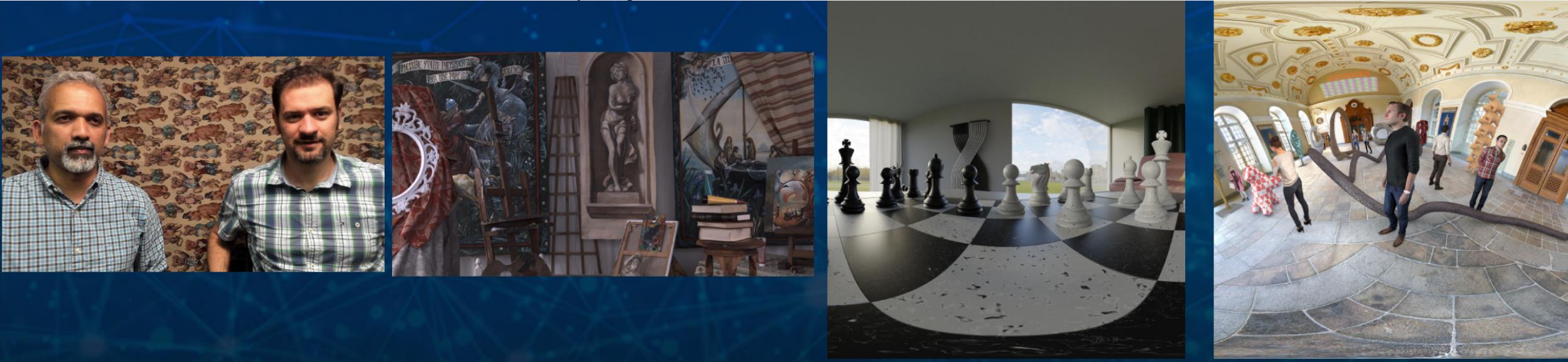
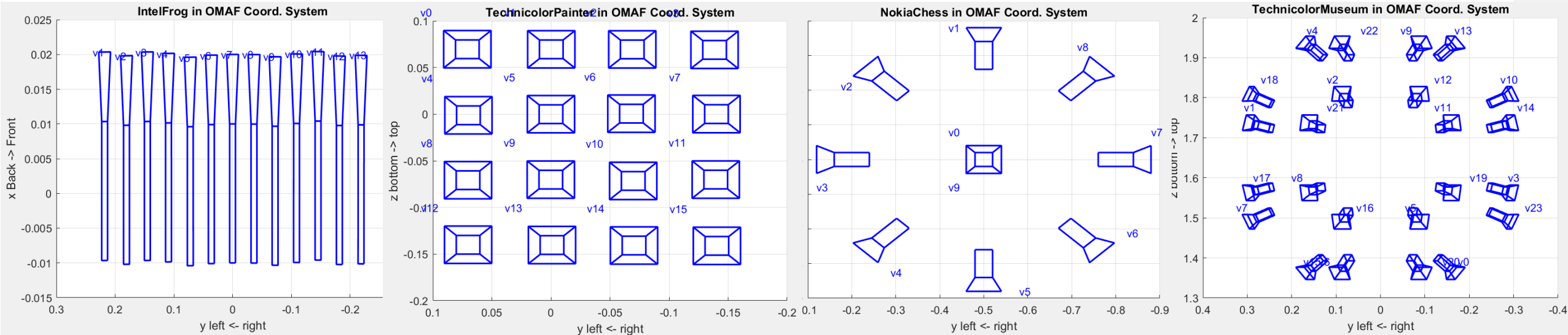


# MPEG Immersive Video (MIV)

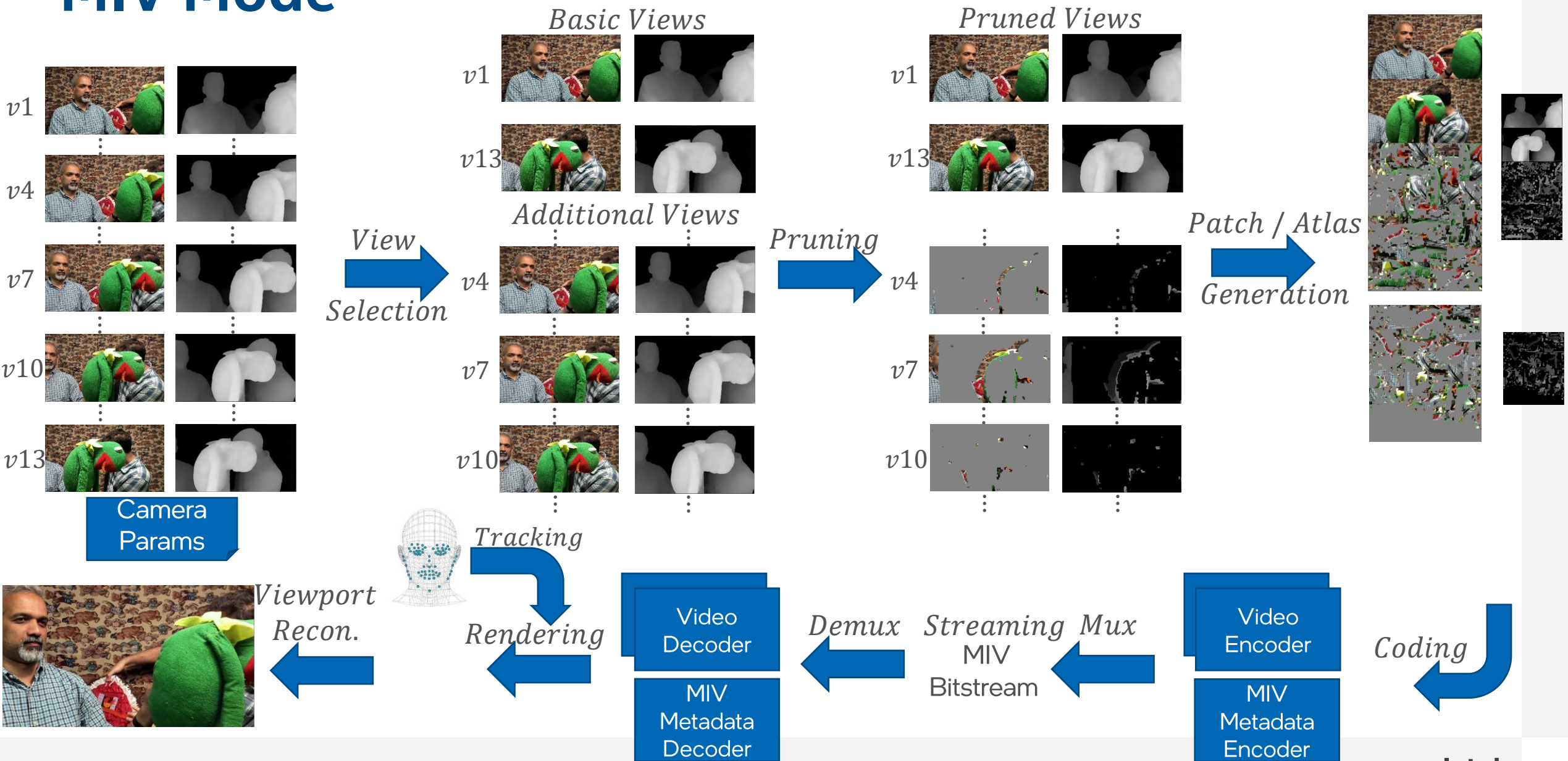
- An industrial effort aiming to provide data access and delivery standard for the emerging immersive ecosystem by the 2<sup>nd</sup> half of 2021.
- Based on Visual Volumetric Video-based Coding (V3C) which is also used for Video-based Point Cloud Compression (V-PCC).
- Encoder inputs any number of camera views with any capturing configuration and renderer outputs viewpoints.
- Can be used with any video codec: HEVC, HEVC SCC, AV1, VVC, AVC
- High-level syntax to indicate alignment of atlases with views, to enable viewport dependent streaming, & substream access for decoding & rendering.



# Samples of MIV Content

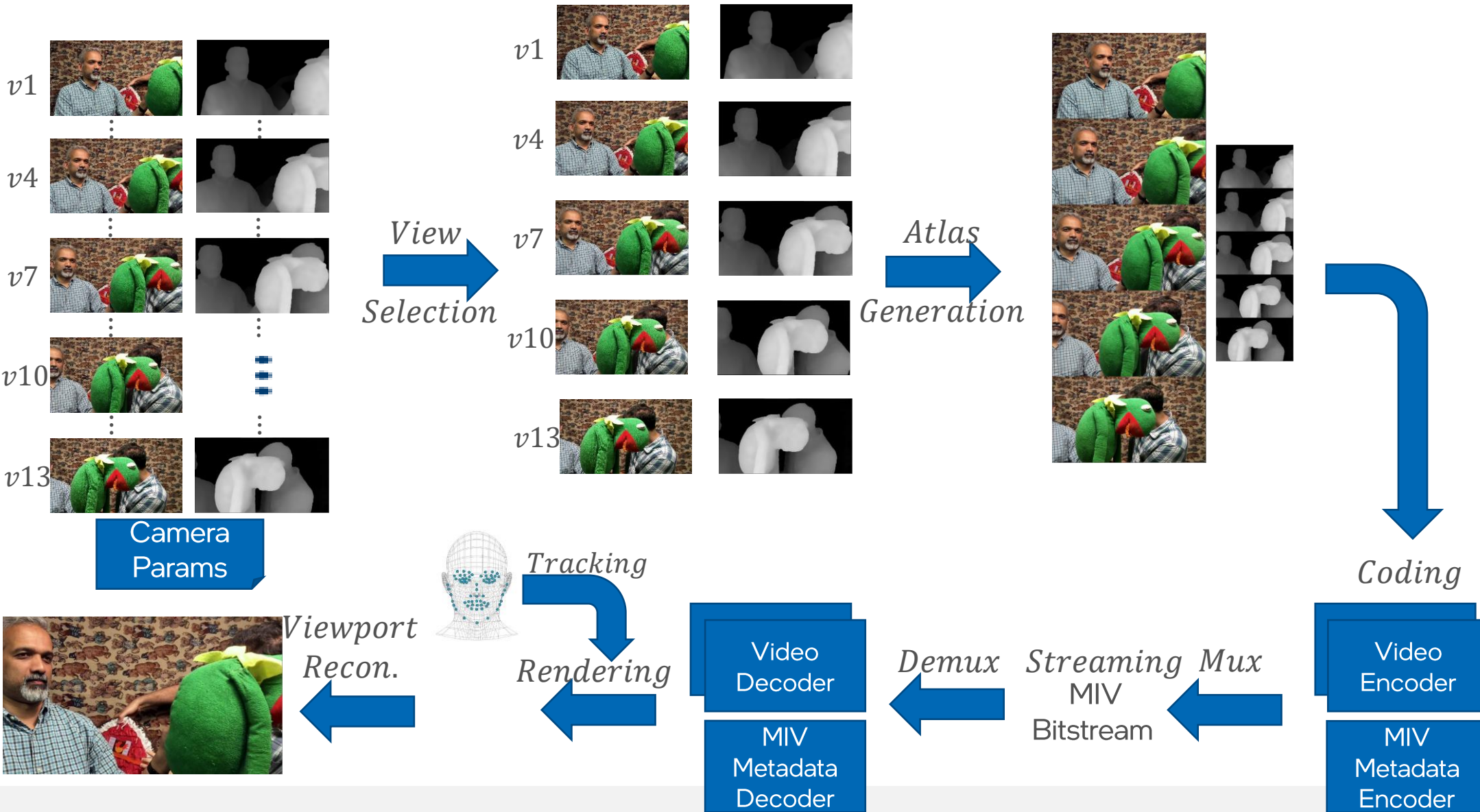


# MIV Mode



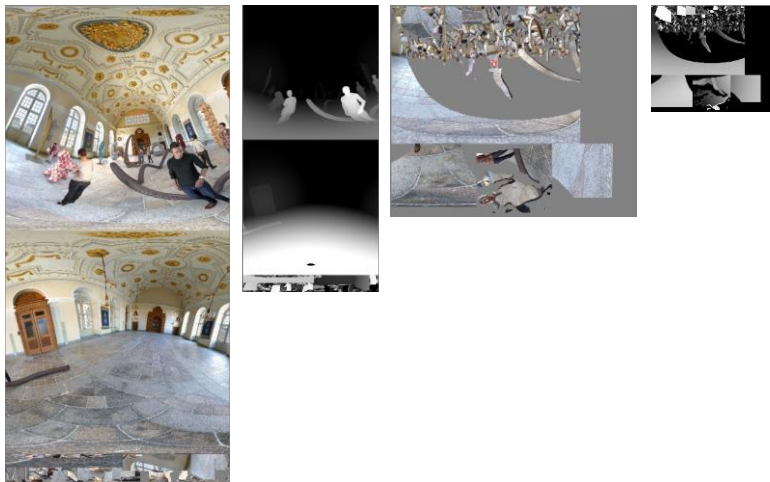


# MIV View Mode

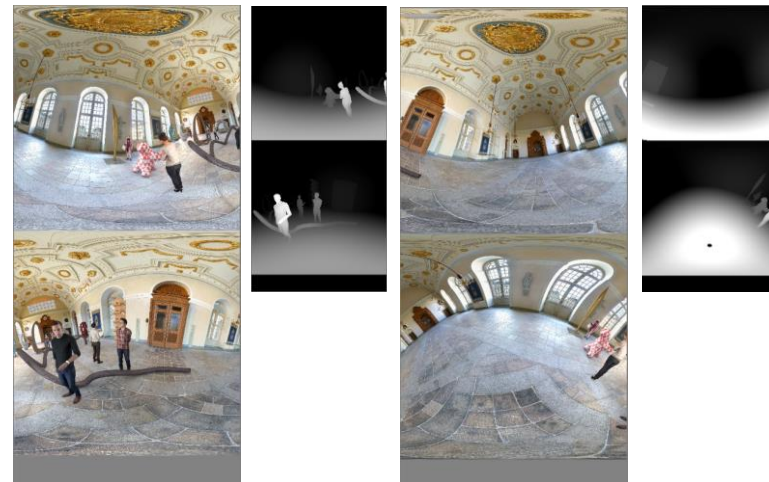


# MIV Operating Modes

*MIV Atlas*



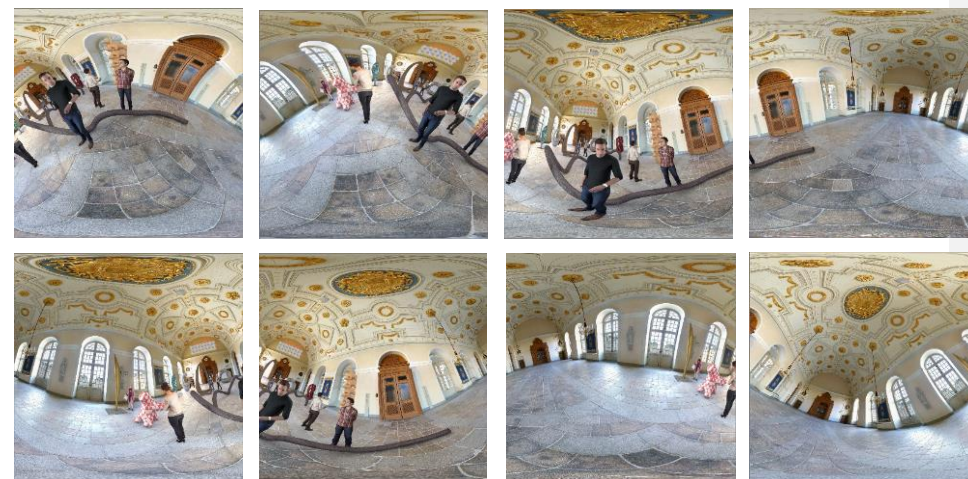
*MIV View*



*MIV Entity*

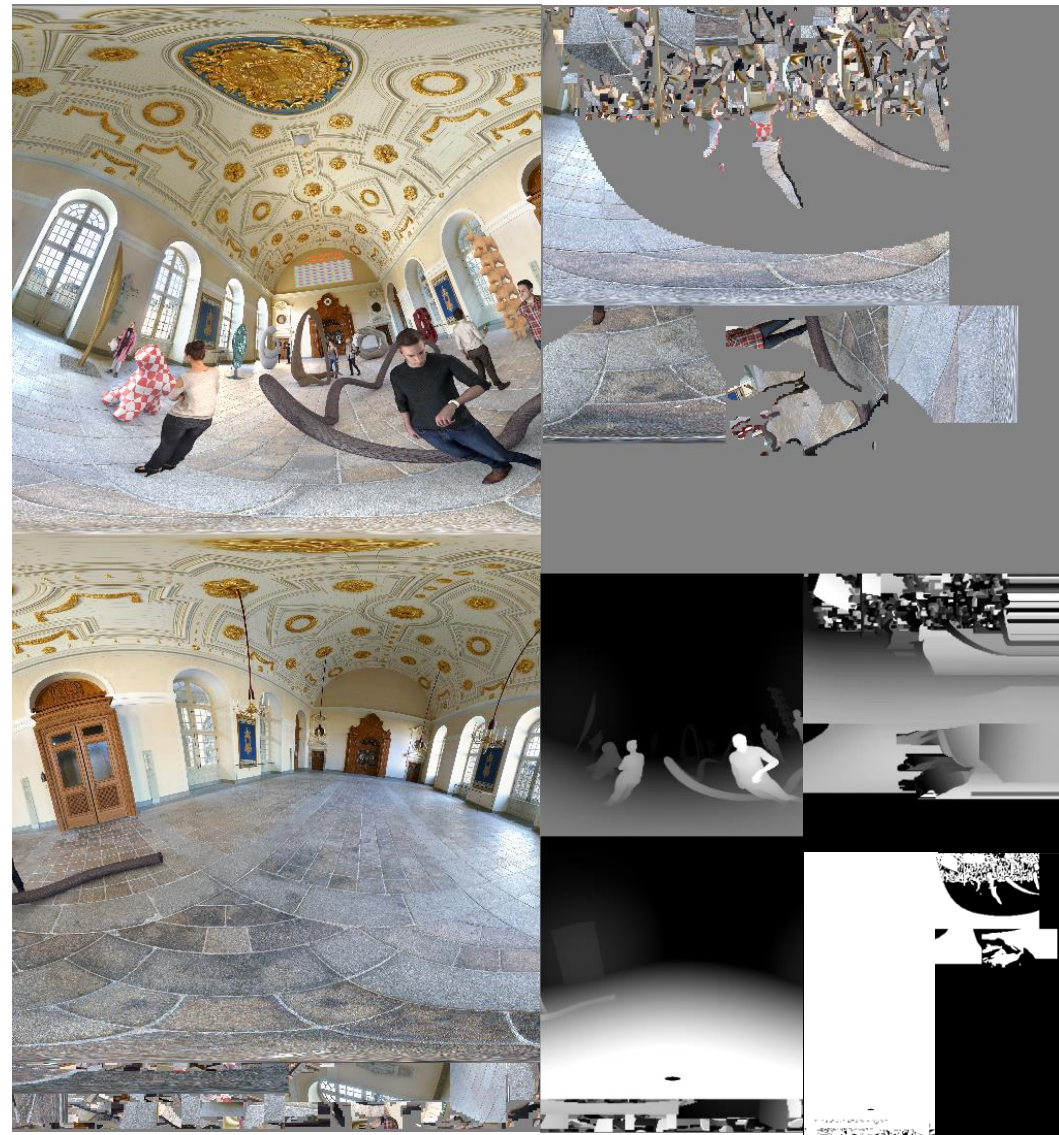
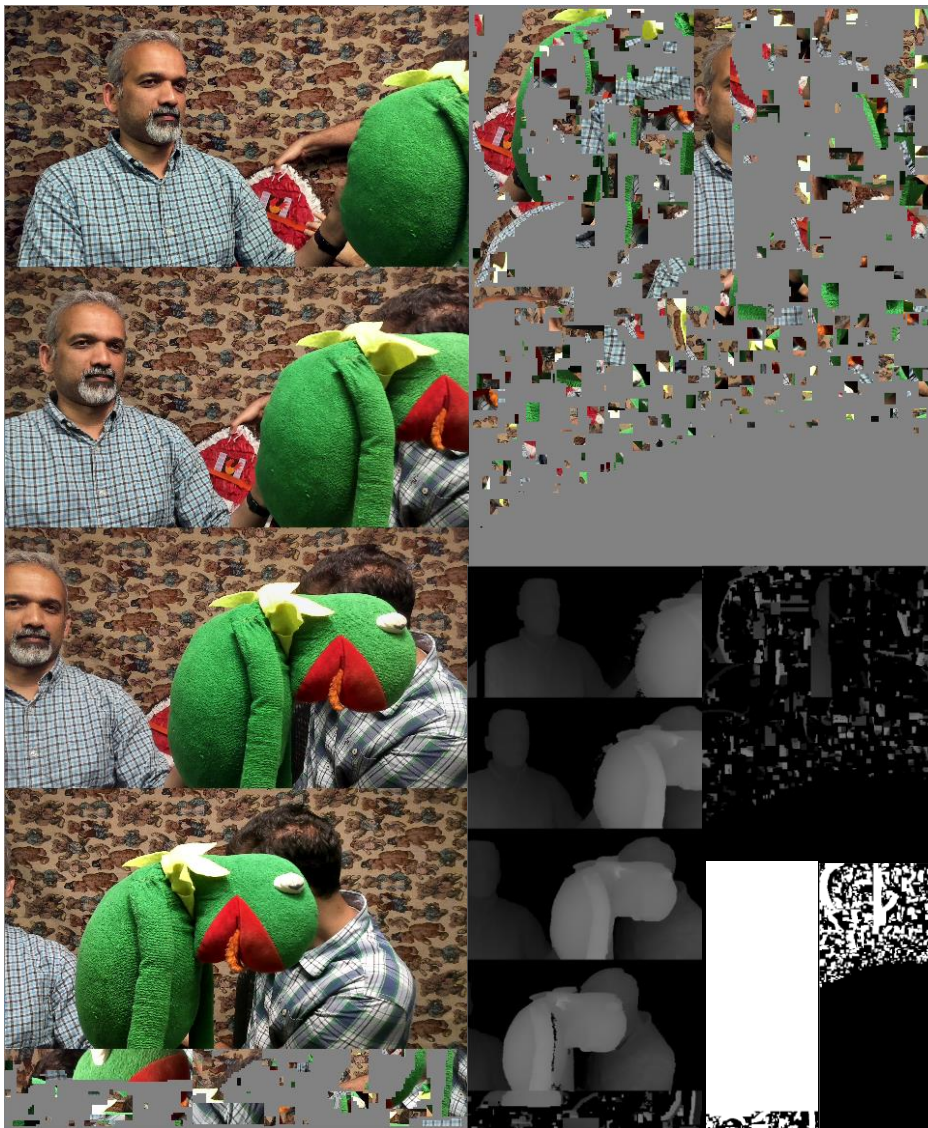


*MIV Geometry Absent*



All modes can either embed occupancy in depth or signaled occupancy explicitly

# MIV Frame Packing



MIV can combine texture, geometry/depth, and occupancy (from multiple cameras) in same packed frame, to enable single video stream

# V3C/MIV metadata SEI in a Single HEVC Stream

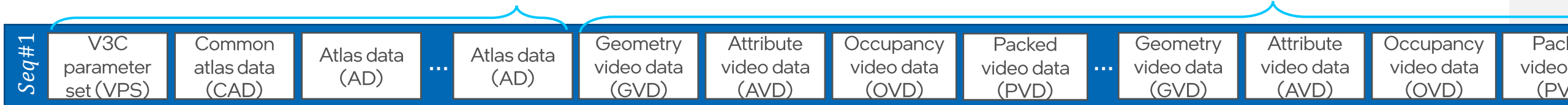
## ▪ Motivations:

- Single layer video codes are widely optimized in media HW of GPUs and supported by many video players.
- Pack all MIV atlases into a single video data to simplify encoding & decoding operations and avoid sync / buffering issues in case of multiple streams.
- Group all non-video V3C units together in a single Supplemental Enhanced Information (SEI) message that can be embedded in a regular HEVC bitstream.
- MIV content in this format can use a legacy streaming server without any changes required on the server end.

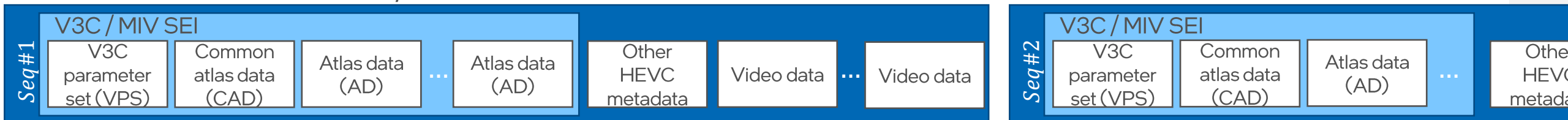
Sample V3C/MIV bitstream

non-video V3C units

video V3C units



HEVC bitstream with V3C/MIV metadata SEI

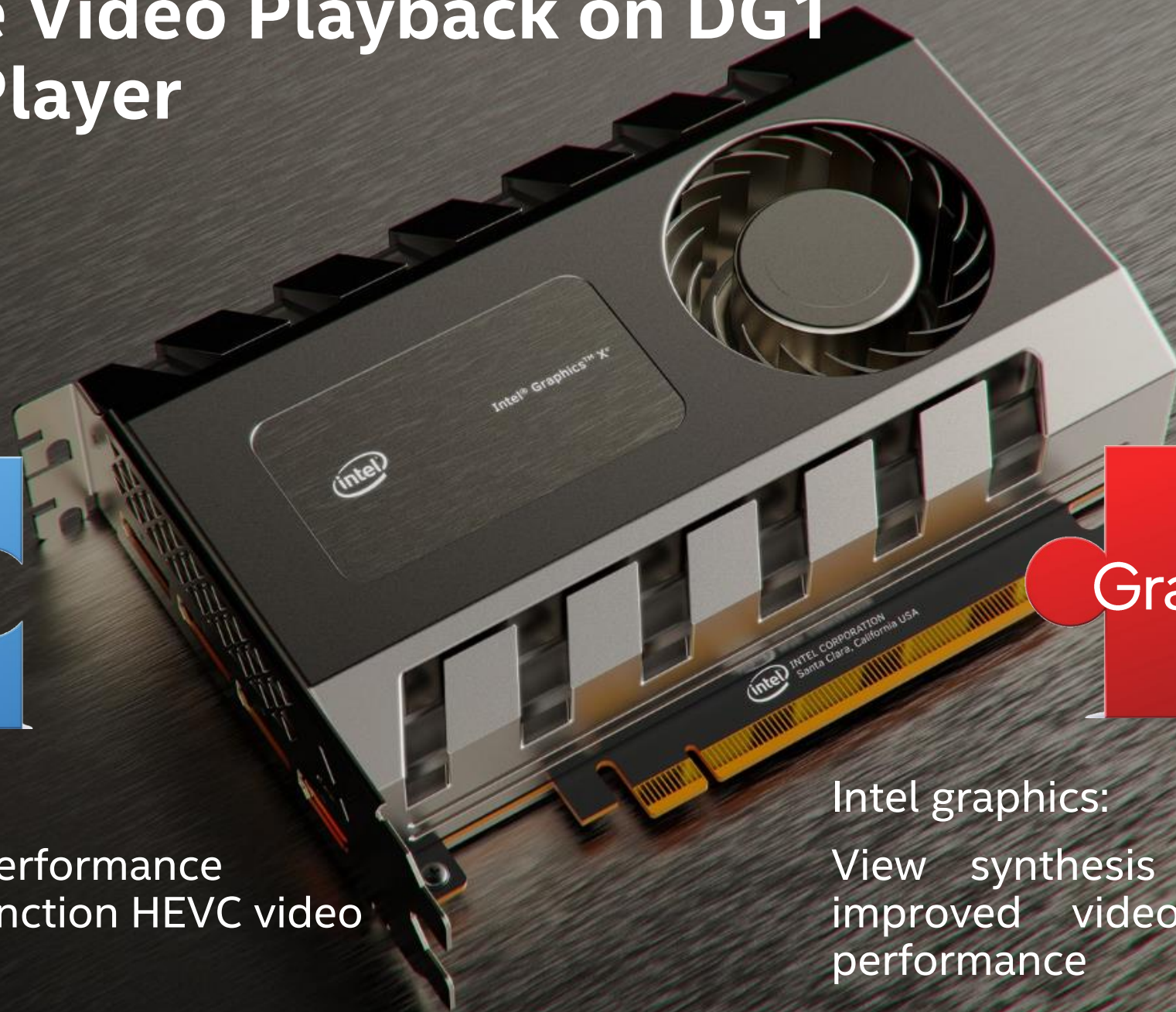


# Immersive Video Playback on DG1 Freeport Player



Intel media:

Low power, high performance  
dedicated fixed-function HEVC video  
decoder



Intel graphics:

View synthesis algorithms for  
improved video quality, high  
performance

# Freeport Player

- Proof of concept done using an open-source VLC player.
- A special video filter has been developed to activate the immersive mode & MIV SEI is extracted from bitstream to operate accordingly.
- HW decoding and rendering are executed on Intel's DG1.
- Rendering stages are implemented as DirectX 11 shaders.
- Face tracking is done using inferencing of a pretrained OpenCV deep neural network module + denoising.

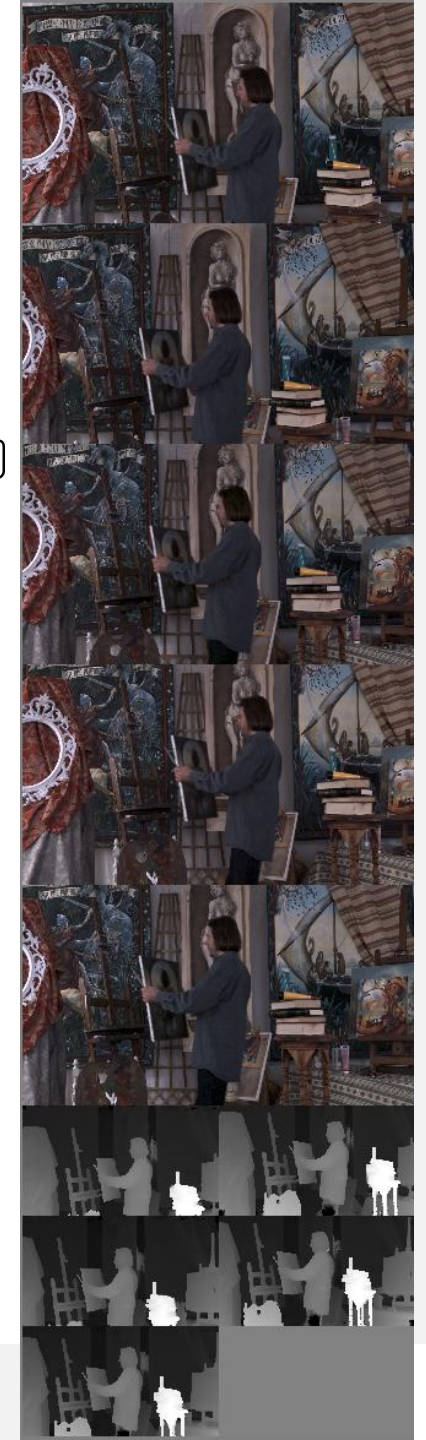
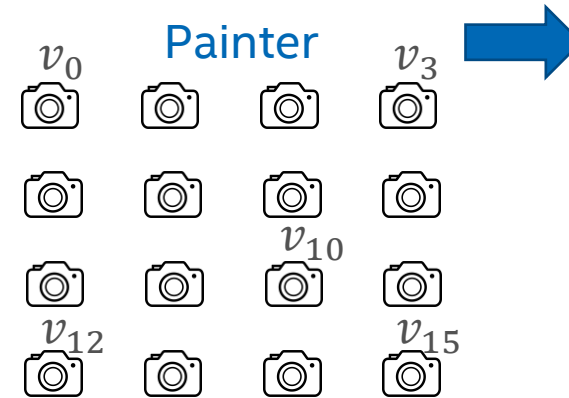
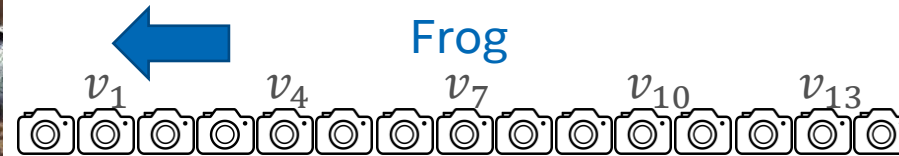


# HEVC Bitstreams Used in the Demo

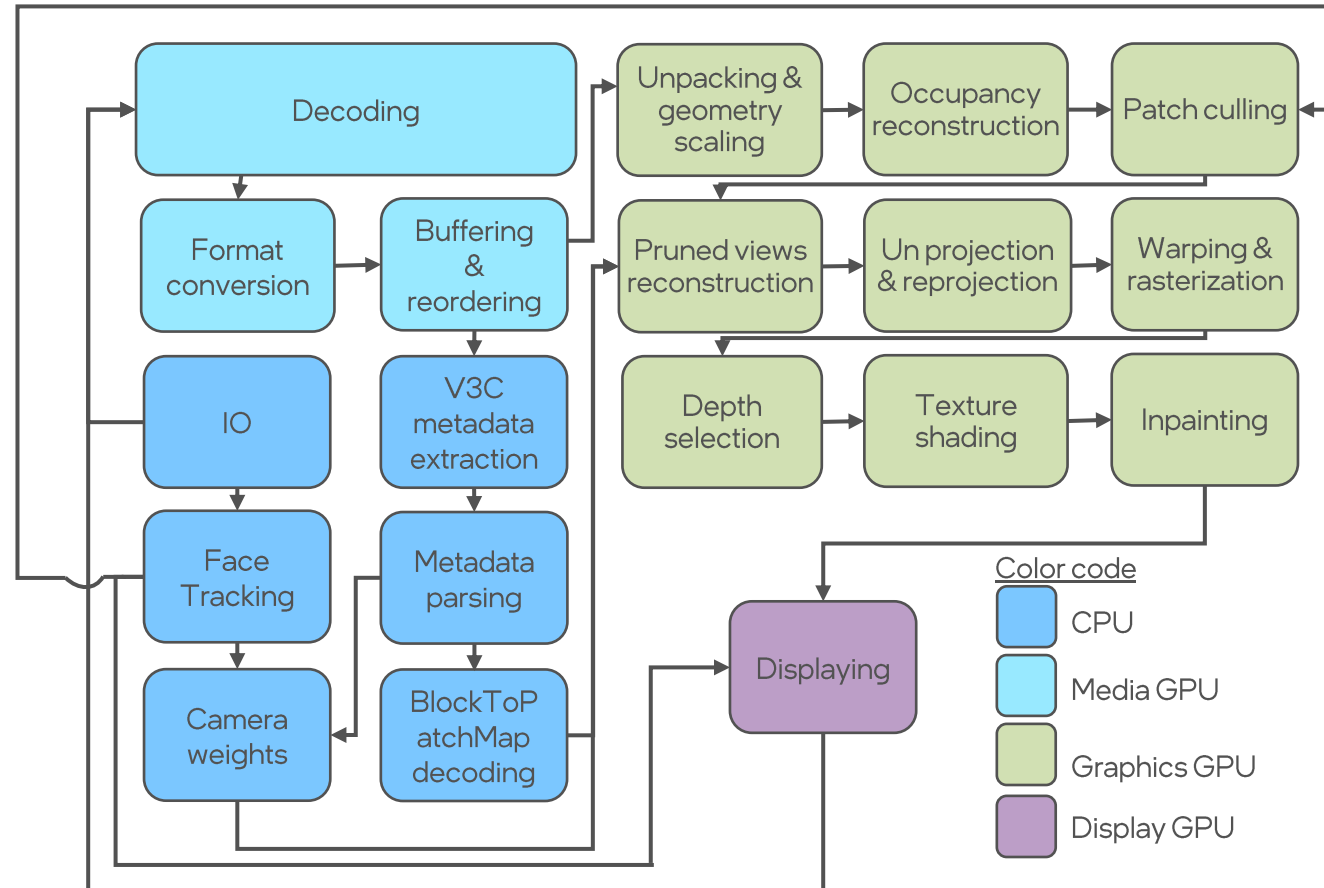
5 selected views each  
@ 1280x720

Encoded in a single video  
@ 1280x4680, 30 fps

MIV SEI carries information about the number of atlases, the resolution & bitdepth, the depth quantization, and the mapping of patches between the atlases and the views, and the packing information.



# Freeport Player – Compute Resources



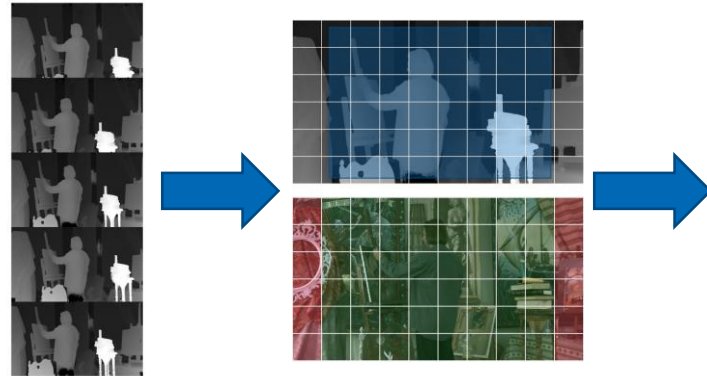


# Rendering – GPU Implementation

Unpacking & geometry scaling



Occupancy recon. & patch culling



Pruned views recon.



Warping & rasterization

Un projection & reprojection

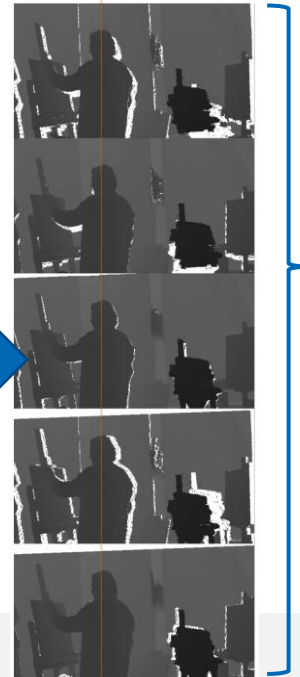
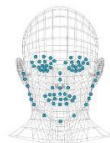
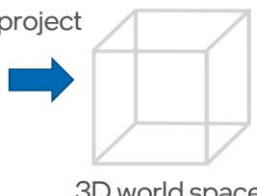
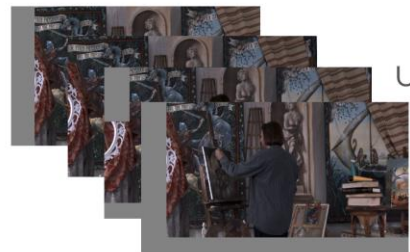
2D image space

Unproject

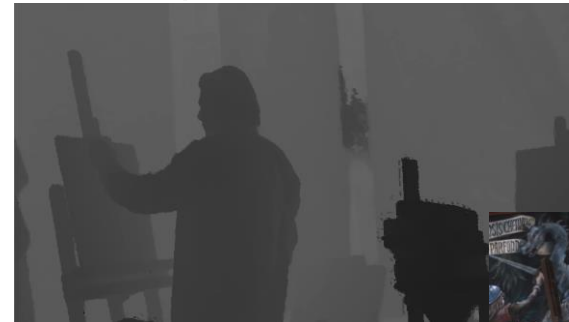
Reproject

3D world space

Desired Target Image Plane



Depth selection



Texture shading & Inpainting



# SUMMARY



- Immersive media gives us a more natural way to consume the content.
- An immersive ecosystem is emerging with compression and delivery infrastructure e.g. MIV being developed.
- Freeport player is a proof of concept of how to interact and navigate immersive content from a single video bitstream with MIV metadata in.
- Content creation is another key piece that we are actively working on!
- Intel's Xe Max GPU is capable of handling immersive workloads.



The Intel logo is centered on a solid blue background. It features the word "intel" in a white, lowercase, sans-serif font. A small blue square is positioned above the letter "i". To the right of the word "intel" is a registered trademark symbol (®).

intel®