Immersive media & 5G

Key takeaways from the Ecosystem & Standards workshop (April 2019)

© 3GPP 2019

Disclaimer

- The content of this presentation has been produced based on publicly available material from 3GPP contributions, including the presentations made during the workshop Ecosystems & Standards “Immersive Media meets 5G” jointly organized by 3GPP and the VR Industry Forum.
- Those workshop presentations are available at the following link:
  - https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/Joint%203GPP%20SA4%20-%20VRIF-AIS%20Workshop%202019/Docs/

© 3GPP 2019

Immersive media & 5G

- A few definitions to start
  - XR, DoF, 5G

- Standardization landscape for immersive media
  - MPEG, Khronos, ETSI, VRIF
  - Focus on 3GPP SA4 activities

- The workshop on Immersive Media meets 5G

© 3GPP 2019

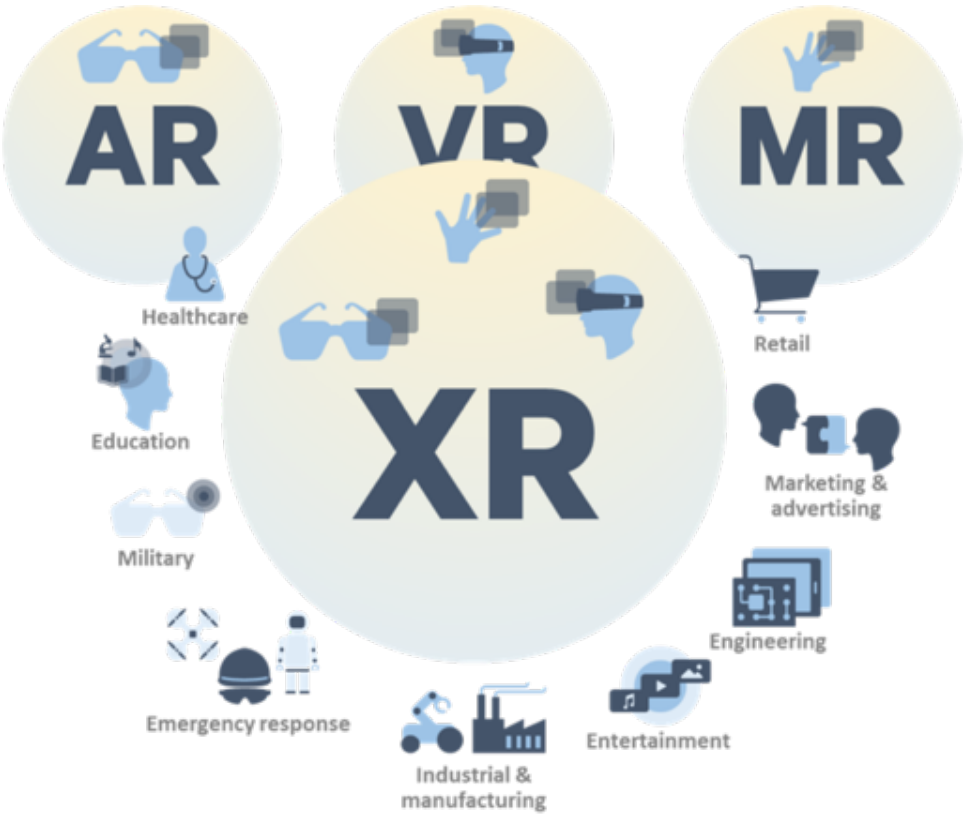
A few definitions

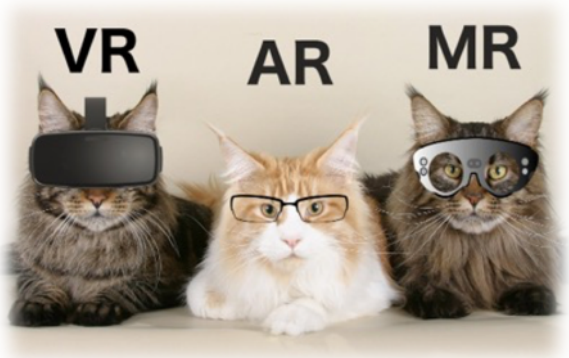
It’s so easy to get mixed up…

© 3GPP 2019

What is meant by XR?

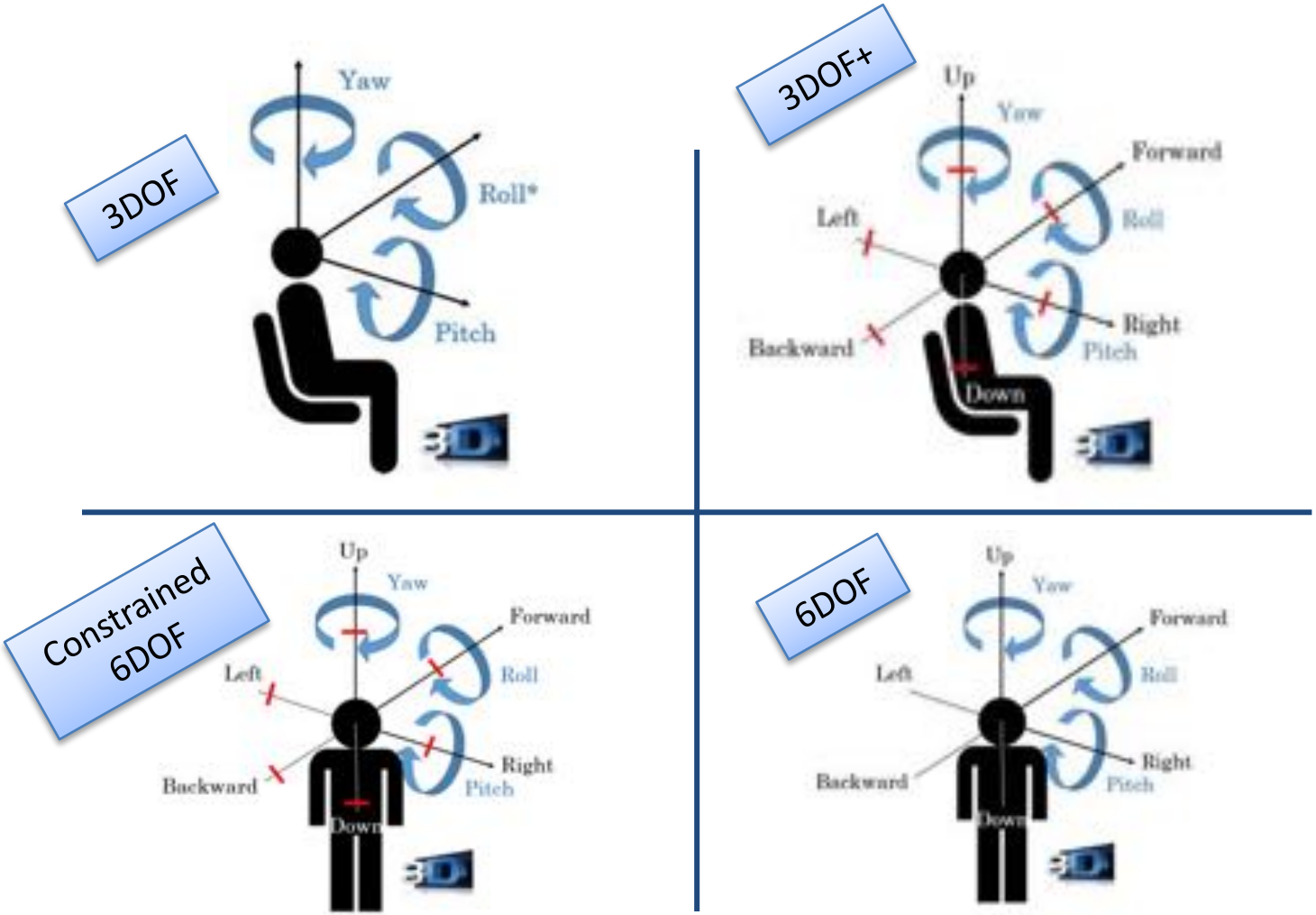
- ***Virtual reality* (VR)** is a rendered version of a delivered audio-visual scene providing the feeling of being physically and spatially located in the virtual environment.
- ***Augmented reality* (AR)** is when a user is provided with additional information or artificially generated items or content overlaid upon their current environment.
- ***Mixed reality* (MR)** is an advanced form of AR where some virtual elements are inserted into the physical scene with the intent to provide the illusion that these elements are part of the real scene.
- ***Extended reality* (XR)** refers to all real-and-virtual combined environments and human-machine interactions generated by computer technology and wearables.

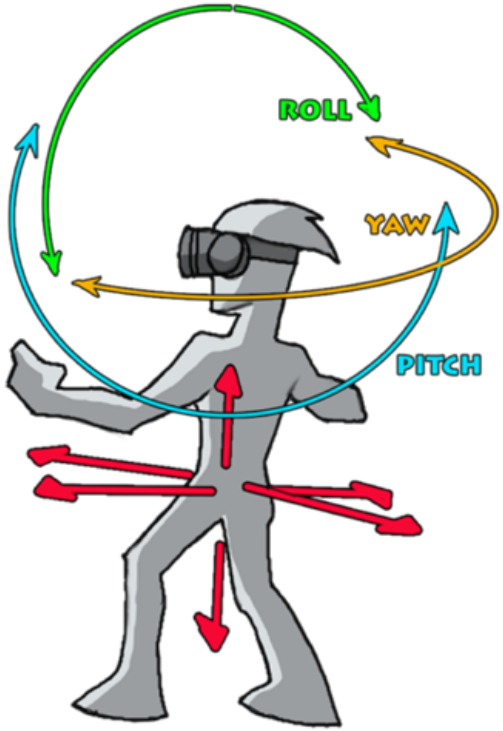




© 3GPP 2019

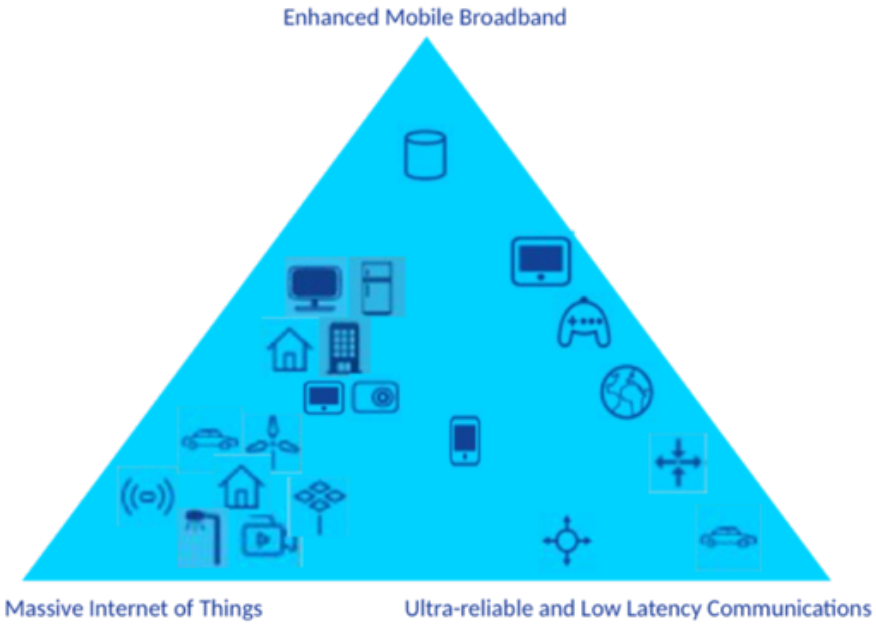
Degrees of freedom (DoF)





© 3GPP 2019

What is 5G? First a vision

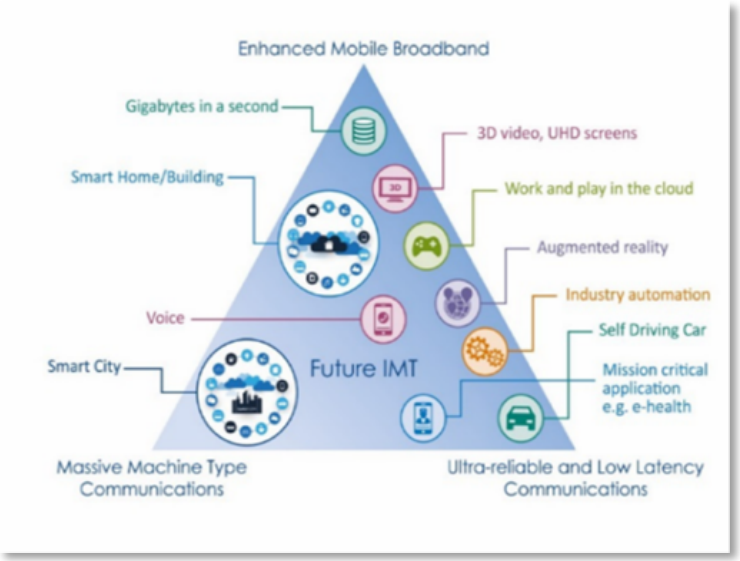


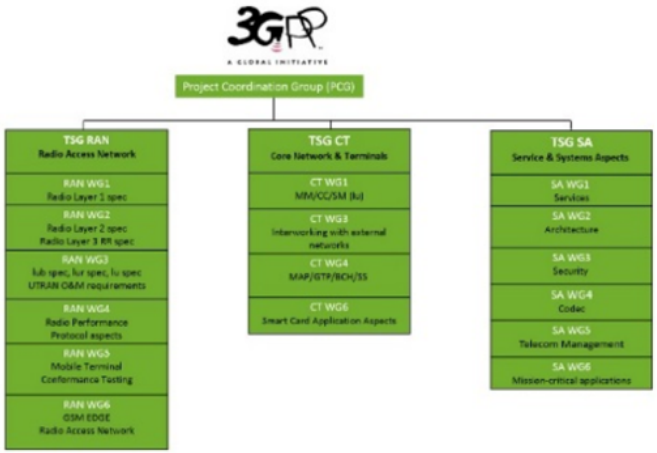
5G / IMT-2020 VISION

- Address demands and business contexts of 2020 and beyond.
- Enable a fully mobile and connected society.
- Empower socio-economic transformations in countless ways.

© 3GPP 2019

What is 5G? From the vision to standards





3GPP Specifications and Reports:	
Requirements	21 series
Service aspects ("stage 1")	22 series
Technical realization ("stage 2")	23 series
Signalling protocols ("stage 3") - user equipment to network	24 series
Radio aspects	25 series
CODECs	**26 series**
Data	27 series
Signalling protocols ("stage 3") -(RSS-CN) and OAM&P and Charging (overflow from 32.-range)	28 series
Signalling protocols ("stage 3") - intra-fixed-network	29 series
Programme management	30 series
Subscriber Identity Module (SIM / USIM), IC Cards. Test specs.	31 series
OAM&P and Charging	32 series
Security aspects	33 series
UE and (U)SIM test specifications	34 series
Security algorithms	35 series
LTE (Evolved UTRA), LTE-Advanced, LTE-Advanced Pro radio technology	36 series
Multiple radio access technology aspects	37 series
Radio technology beyond LTE	38 series

3GPP SA4 addresses the media distribution and codecs aspects such as audiovisuals and conversational services

© 3GPP 2019

Standardization landscape for immersive media

Although not a comprehensive list…

© 3GPP 2019

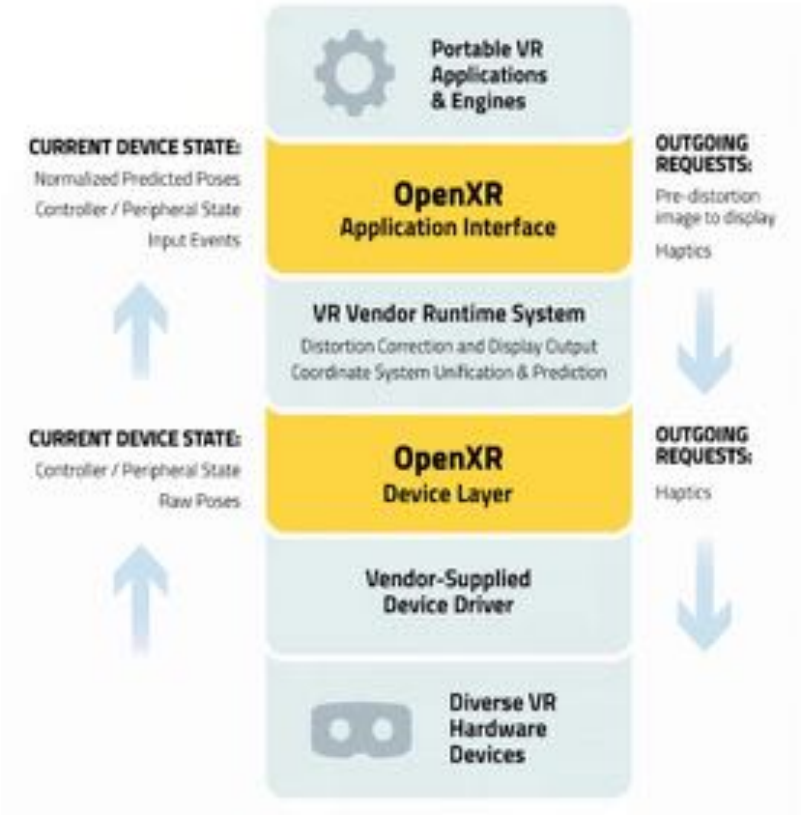
Standardization landscape of XR: MPEG-I

- Video-based point cloud compression **V-PCC**
  - A compression format for 3D objects represented by point clouds wherein any video codec can be used
- Omnidirectional Application format **OMAF**
  - 3DoF & 3DoF+: allowing a certain amount of lateral head movement
  - Interactivity and Overlays
  - Better viewport-adaptive streaming
- Network-based media processing **NBMP**
  - Framework allowing network/service providers to deploy/control media processing operations in the network/cloud (such as edge computing)
- Next generation video codec **VVC**
  - The next generation video coding standard is expected to be ready by mid-2020. It promises a compression efficiency that reduces the bitrate by 40-50% compared to HEVC at the same quality. VVC is considered useful for ever-larger media, including point clouds.

© 3GPP 2019

Standardization landscape of XR: Khronos/OpenXR

- OpenXR is a cross-platform standard defining 2 levels of API that AR/VR runtimes can use to access the XR ecosystem:
  - Apps and engines use standardized interfaces to interrogate and drive devices. Devices can self-integrate to a standardized driver interface.
  - Standardized hardware/software interfaces reduce fragmentation while leaving implementation details open to encourage industry innovation.



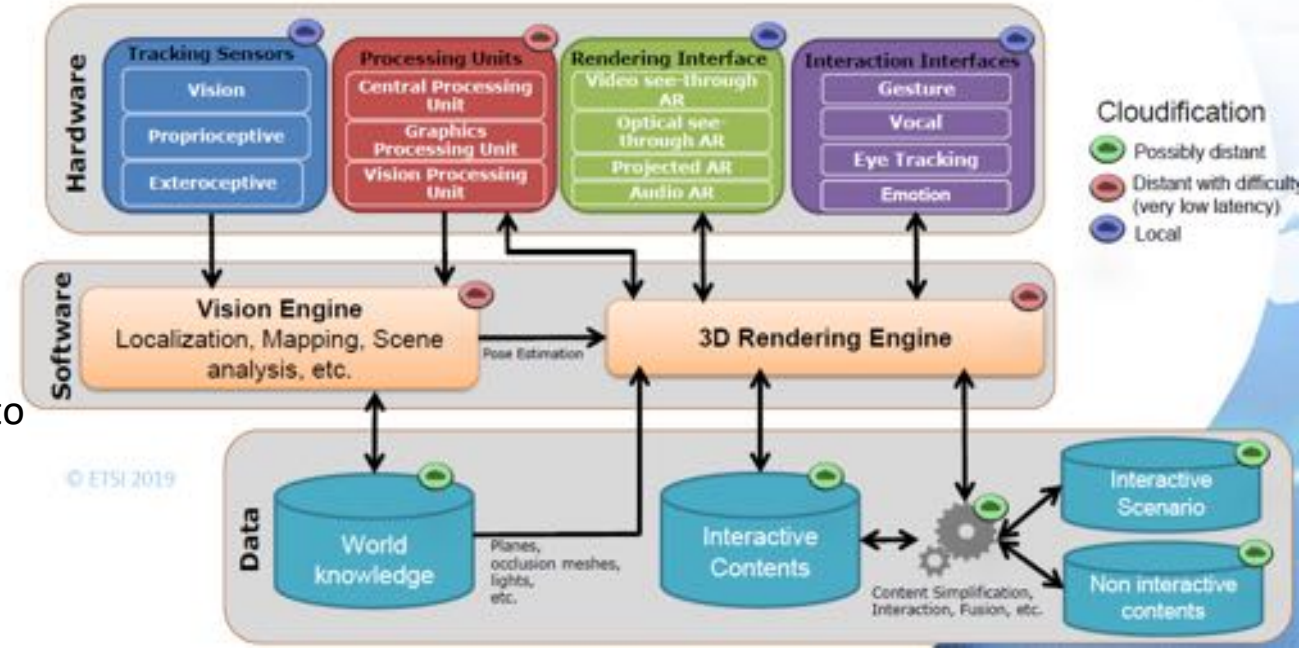
KHRONOS GROUP

© 3GPP 2019

Standardization landscape of XR: ETSI ARF

- Augmented Reality Framework Industry Specification Group (ARF ISG).”
- The overall objective if the ISG ARF is define a framework for the interoperability of AR applications and services, in order to prevent market fragmentation and enable providers to offer part of an overall AR solution
  - To define a modular architecture and achieve interoperability at the interface of building blocks
    - The focus is on interfaces, no standardization of core technologies
    - The building blocks can be open source, proprietary or standardized technologies

- Achieving Interoperability:
  - Will give end-users confidence to invest in and use (future proof) AR solutions
  - Can prevent lock-in situation with a single vendor
  - Can ease technology swapping, e.g. to improve applications performance and services



Standardization landscape of XR: VRIF Guidelines

- VR Industry forum defines guidelines on immersive media from content production to end device decoder including security aspects

CES 2018

The initial release of the VRIF Guidelines focused on the delivery ecosystem of **360°** video and incorporates:

- Documentation of cross-industry interoperability points, based on ISO MPEG's Omnidirectional Media Format (OMAF)
- Best industry practices for production of VR360 content, with an emphasis on human factors such as motion sickness
- Security considerations for VR360 streaming, focusing on content protection but also looking at user privacy.

April 2019

The draft VR distribution guidelines 2.0 was published end of April 2019

It contains the following additions: ***Live VR*** *services ; Text & Fonts in VR experiences ; Watermarking 2D frames ; FOV emphasis for viewport independent content*

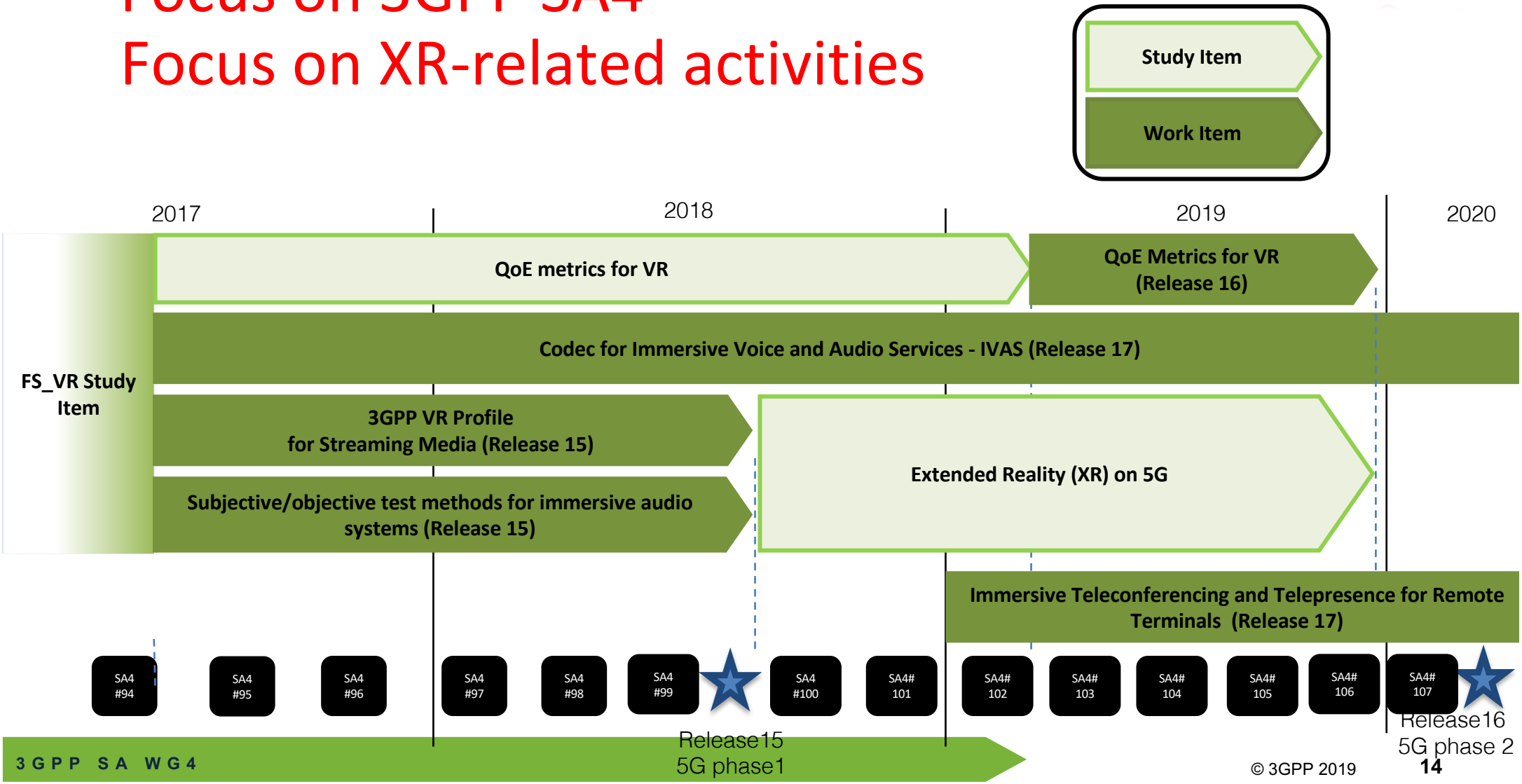
Future work

The next work packages are as follows:

- **Volumetric Capture**: Three-dimensional elements
- Cloud AR/VR: Using 5G to powering the next level of immersive experience
- **Augmented Reality**
- Securing XR assets
- **Social VR**: Removing the VR isolation factor
- New industry verticals

© 3GPP 2019

Focus on 3GPP SA4
Focus on XR-related activities

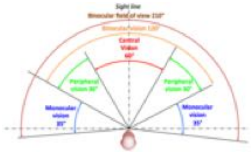


© 3GPP 2019

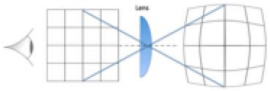
Focus on 3GPP SA4 Checkpoint on VR (360)

- April 2016 – June 2017: Study on Virtual reality

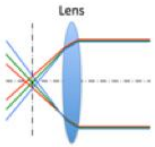
Video systems



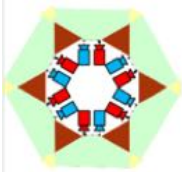
- Human factors
- FOV and lenses
- Optical aberrations

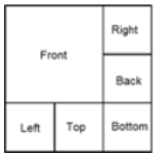


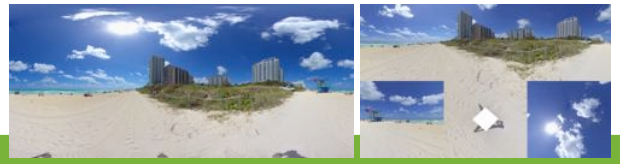
VR Video Workflow



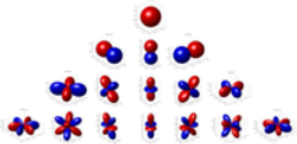
- Capture
- Stitching
- Projection
- Packing
- Encoding/decoding
- Rendering







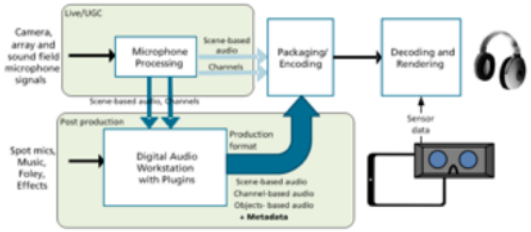
Audio systems



- Channel based
- Object based
- Scene based

VR Audio Workflow

- Content production
- Audio production formats
- Rendering systems
- Data exchange
- Ambisonics analysis
- Rendering



3GPP TR 26.918

Virtual Reality (VR) media services over 3GPP

© 3GPP 2019

Focus on 3GPP SA4 Checkpoint on VR (360)

- Release 15 Technical specification for streaming services
  - Definition of client architecture and API for VR streaming services
  - Set of operating points covering the large range of device capabilities from Carboards to high-end tethered HMDs.
  - Definition of Media profiles: mapping of operating points to DASH delivery
  - System metadata is added to support rendering of 360 experiences on 2D screens, including the aspects of rendering without pose information

3GPP TS 26.118

Virtual Reality profiles for streaming applications

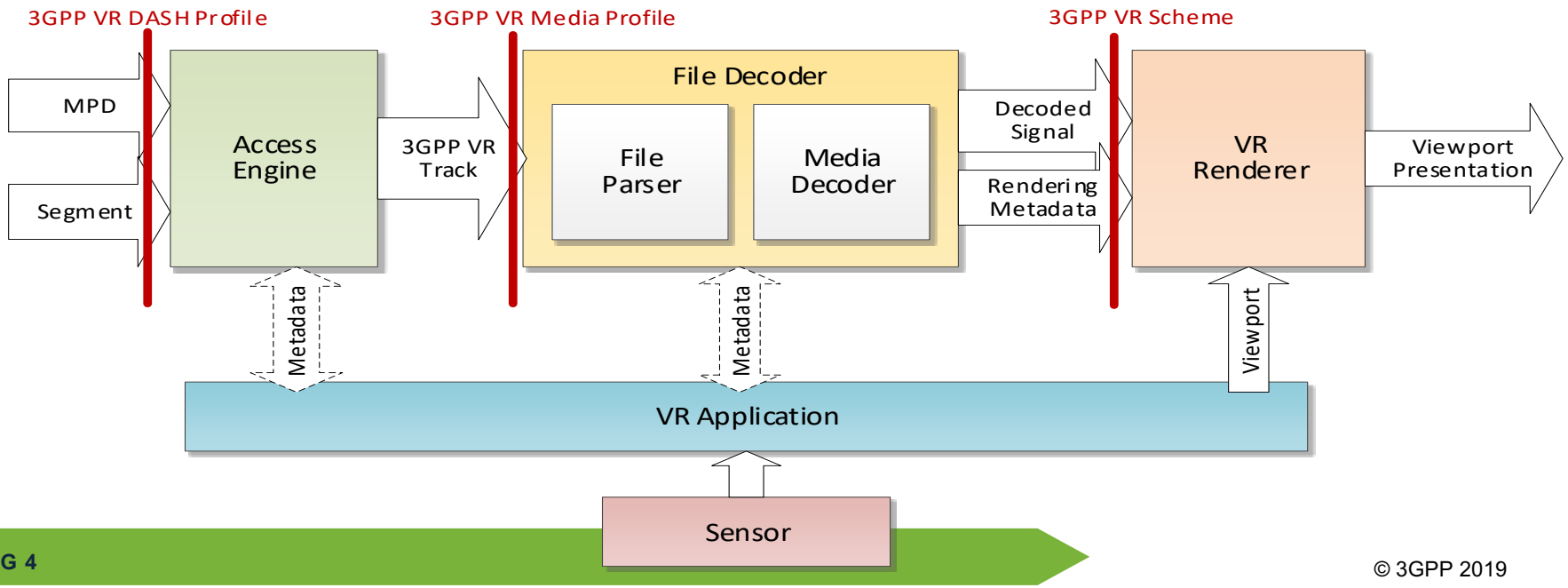
© 3GPP 2019

Focus on 3GPP SA4 VRStream specification

One specification produced:

3GPP TS 26.118 Virtual Reality Profiles for Streaming Media

Reference client architecture defined with interoperability points on media formats and their mapping to DASH delivery.



© 3GPP 2019

Focus on 3GPP SA4 VRStream - Video

3 operating points have been defined:

Basic AVC: AVC High Profile Level 5.1 for mono only, single stream, and reuse of single stream DASH streaming. This profile addresses legacy services and devices.

Main HEVC: HEVC High Profile Level 5.1 allowing mono and stereo, single stream, but either a single or multiple independent Adaptation Sets may be offered, such that a client can choose based on its current pose.

Flexible HEVC: HEVC High Profile Level 5.1, but in addition to the Main Video features, it permits to stream and combine multiple tiles at the receiver for improved quality. Aligned with MPEG OMAF

Operation Point name	Decoder	Bit depth	Typical Original Spatial Resolution	Frame Rate	Colour space format	Transfer Characteristics	Projection	Rotation	RWP	Stereo
Basic H.264/AVC	H.264/AVC HP@L5.1	8	Up to 4k	Up to 60 Hz	BT.709	BT.709	ERP w/o padding	No	No	No
Main H.265/HEVC	H.265/HEVC MP10@L5.1	8, 10	Up to 6k in mono and 3k in stereo	Up to 60 Hz	BT.709 BT.2020	BT.709	ERP w/o padding	No	Yes	Yes
Flexible H.265/HEVC	H.265/HEVC MP10@L5.1	8, 10	Up to 8k in mono and 3k in stereo	Up to 120 Hz	BT.709 BT.2020	BT.709, BT.2100 PQ	ERP w/o padding CMP	No	Yes	Yes

Focus on 3GPP SA4 VRStream - Audio

- 4 solutions were under consideration:

MPEG-H 3D Audio Baseline profile	DTS-UHD
Metadata Assisted EVS Codec (MAEC)	Spatial AAC extension (spAACe)

- Those 4 solutions were characterized with tests conducted on both loudspeaker and headset configurations, testing both the encoder and the renderer performances.
- The test results are documented in a technical report TR 26.818
- One solution specified: **MPEG-H 3D Audio Baseline profile**
  - Enabling the distribution of scene, object and channel-based 3D audio.

© 3GPP 2019

Focus on 3GPP SA4
Extended Reality over 5G (1/4)

- **Study Item launched in October 2018**
- Extended Reality (XR)
  - an envelope that includes
    - VR (Virtual Reality)
    - AR (Augmented Reality)
    - MR (Mixed Reality)
- The study addresses:
  - VR cases in more than 360° navigation
    - 3DOF+ 3 axis rotations + 3 axis translations limited to head movement with fixed body
    - 6DOF Full free navigation (user can walk and look around)
  - AR cases where synthetic objects are overlaid with the real environment (including MR)

© 3GPP 2019

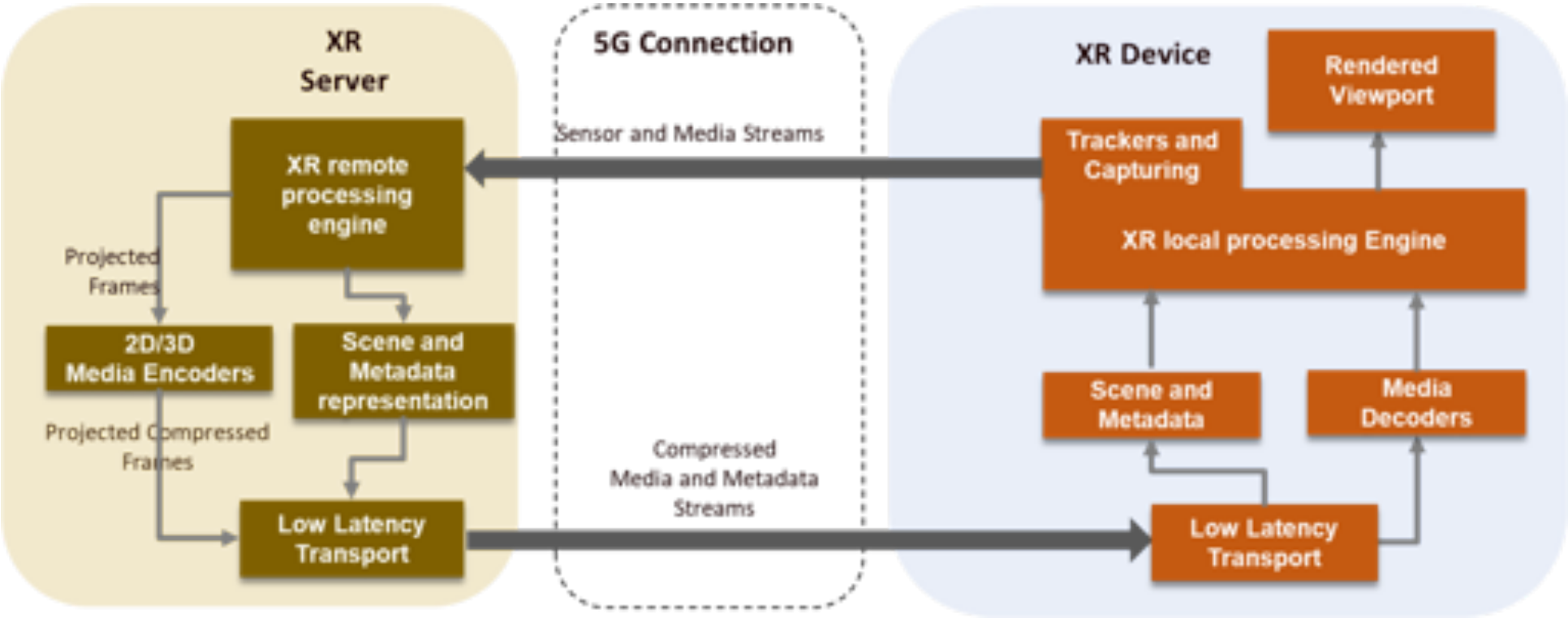
Focus on 3GPP SA4
Extended Reality over 5G (2/4)

- Objectives
  - **Analysing** the different technologies and equipment in place that provide an Extended Reality experiences.
  - **Collecting** the associated use cases and identifying the 3GPP service(s) they map to.
  - **Identifying**
    - media formats (including audio and video), metadata, accessibility features…
    - client and network architectures and APIs that support XR use cases
    - QoS service parameters and other core network and radio functionalities that would be required or at least beneficial for XR use cases
  - Possibly **conducting** subjective tests so as to estimate the audio and video formats and encoding parameters required for ensuring the quality of experience as considered necessary

© 3GPP 2019

Focus on 3GPP SA4
Extended Reality over 5G (3/4)

- a reference architecture under consideration (example)



© 3GPP 2019

Focus on 3GPP SA4 Extended Reality over 5G (4/4)

- Some Use cases identified so far…
  - **3D messaging**
    - Ability to capture and send 3D models via MMS
  - **Streaming of Immersive 6DoF**
    - Free navigation/ multiple viewpoints in video content
  - **Immersive online gaming and spectator mode**
    - Free navigation in CGI content
  - **Remote assistance in Industry**
    - AR guided assistance for onsite operations
  - **Realtime 3D communication**
    - Immersive conferences including poster sharing
  - **Online shopping with AR**
    - Augmented reality placement of products at home
  - **AR streaming with Localization registry**
    - On-site virtual guides (museums…)

© 3GPP 2019

Ecosystem & Standards Workshop
Immersive media meets 5G

Promote standards and collect industry requirements…

About the workshop

- Jointly organized by 3GPP and the VRIF
- Successful 1st edition held end of 2017 on VR 360.
- This 2nd edition:
  - Themed ***immersive media (XR) meets 5G***
  - Hosted by Sony Pictures studios in L.A. in association with the Advance Imaging Society (Hollywood Lumiere Awards)
- Objective
  - Examine next-generation immersive formats and services, including Virtual, Augmented & Mixed Reality, and investigate 5G business opportunities brought by immersive services
  - Look at their requirements
  - And see if these are met by available / emerging standards and formats

© 3GPP 2019

About the workshop

- 4 sessions organized
  - **Standards**
    - MPEG
    - 3GPP
    - VRIF
    - Khronos
    - ETSI
    - GSMA
  - **Device and network manufacturers**
    - Nvidia
    - Qualcomm
    - Ericsson
    - Intel
    - Dell
  - **Service providers**
    - Orange
    - KPN
    - Verizon
    - CableLabs
    - MobileEdgeX
    - Iconic Engine
  - **Content providers**
    - Sony Pictures
    - Underminer studios
    - 8i
    - Metastage
    - Felix&Paul
    - Atlas V

© 3GPP 2019

Key takeaways from the workshop: Device & network manufacturers – Evolutions

- On devices:
  - Power Savings
  - Eye tracking and Foveated rendering
  - See-through and depth cameras (Inside-out tracking)
  - ARCore: Surface detection, light estimation
  - Displays: 4k per eye, laser projectors
- On Form Factors:
  - AR or VR glasses connected to smartphone via USB-C
  - Power consumption on glasses only 1W
- On Split rendering architecture:
  - Edge cloud: partial rendering (video + meshes)

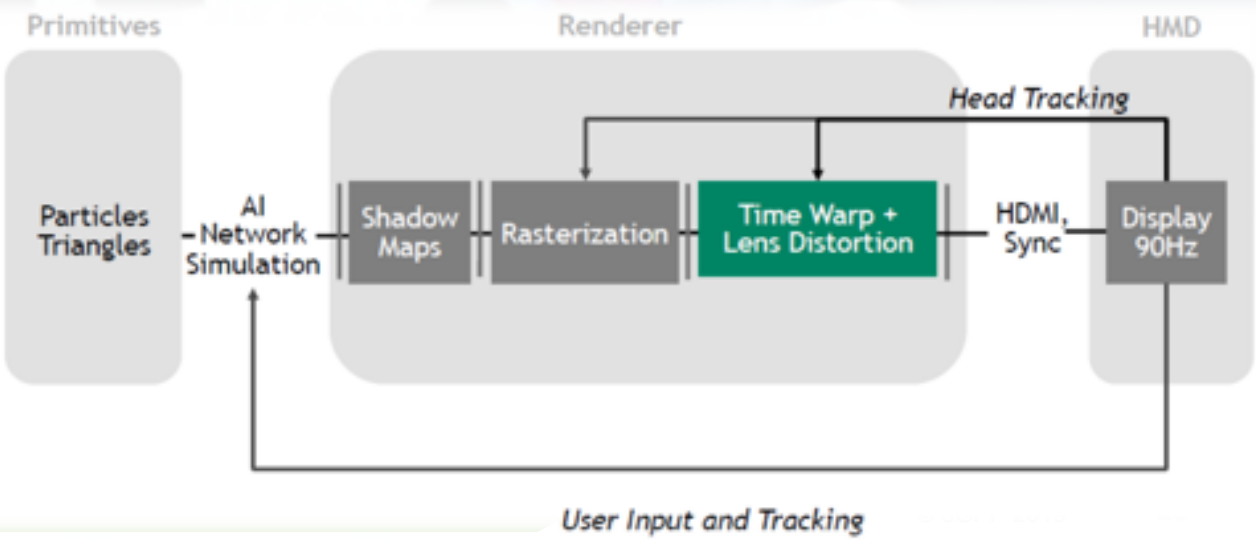
© 3GPP 2019

Key takeaways from the workshop: Device & network manufacturers - Timewarping

More information in Annex

- Interesting concept: Timewarping
  - also known as *Reprojection* is a technique in VR that warps the rendered image before sending it to the display to correct for the head movement occurred after the rendering.
  - Timewarp can reduce latency and increase or maintain frame rate. Additionally, it can reduce judder caused missed frames (when frames take too long to render)

This process takes the already rendered image, modify it with freshly collected positional information from your HMD's sensors, then display it to your screen. Utilizing depth maps (Z Buffers) already present in the engine, Timewarp requires very little computation.



Key takeaways from the workshop: Device & network manufacturers - Split rendering

- Need to differentiate the different computing configurations including HMD, Client, Edge, and Cloud
  - Client compute
  - File Streaming
  - Command Streaming
  - Video Streaming
- Split rendering can use todays' video codecs, but may benefit from new approaches, both in terms of formats and latency
  - 3D video compression
  - More distributed compression
- 5G may terminate in phone, puck or glass
- Up to 4 GPUs dedicated to one headset, sharing is difficult

© 3GPP 2019

Key takeaways from the workshop: Service providers – B2C use cases

- 2D services in a virtual world - will continue to develop.
- Gaming
  - Multi-Location Multi-Player Video Game
  - E-Sport seen as a promising market

- Immersive Live Event
  - "Be there" from anywhere
  - Live and on-demand
  - VR head set and smartphone
  - Social interactions

- VR/AR need to be social
  - 360 conferences / AR communications / holoportation
  - Need to handle heterogeneous types of clients

© 3GPP 2019

Key takeaways from the workshop: Service providers – B2B use cases

- Education
  - From Virtual Tours to real world training, VR technologies engage students in completely transformative ways; Allowing for deeper learning and higher retention through fun and exciting new possibilities.

- Amusement parks/arcades
  - Location-based content
  - Room-scale experiences
  - Custom haptic seating

- VR for industry
  - Facility tours
  - Product visualization
  - Large scale simulation

© 3GPP 2019

Key takeaways from the workshop: Service providers – B2B use cases

- Onboarding
  - Company training
  - Facility training
  - Products simulation logistics
  - Data visualization
- VR for healthcare
  - Remote/assisted surgery
- Anti-anxiety
  - From Virtual Tours to distraction therapy, VR can create calming experiences that ease a patients anxiety and phobias before ever entering a medical facility.

© 3GPP 2019

Key takeaways from the workshop: Content providers – Capture

- 3D scenes and objects are complex to capture

Full body
- 108 cameras 1 minutes 108 Gb

Head
- 20 cameras 1 minute 20 Gb

8i System
- 30 x 4k Cameras
- RGB Cameras (No IR camera)
- Capture Volume (diameter x height): (2.5m x 2.2m)
- Stage Footprint (width x depth x height): (10m x 10m x 6.1m)

Key takeaways from the workshop: Content providers – Format

- Point clouds vs Mesh: Room for both - depends on the use case
- Production workflow from Metastage

© 3GPP 2019

Key takeaways from the workshop: Content providers – Fidelity

- Use of variable resolution so as to preserve subjective details such as the human face
- AI used for mesh refinement

© 3GPP 2019

Key takeaways from the workshop: Content providers –Requirements

- Acquisitions
  - Camera geometry
  - Timecode limited to 30*fps
  - Recording 4k+
  - Spatialized audio
- Compression
  - Stereo support
  - Multiple video streams in a container
  - Changing formats over time
- Distribution
  - Formats
  - CRM/ Security
  - Discoverability
- Projection Models
  - Equirectangular
  - Octahedron
  - Icosahedron
  - Pyramidal
- Color Management
  - ACES
  - OCIO
  - Display Variation
  - HDR

© 3GPP 2019

Conclusions in brief

- **On devices**
  - Quality is improving but processing requirements are huge for XR.
- **Cloud** (Edge) **rendering** may help. Concept of "split computer vision"
  - Latency is crucial
  - Opportunity for 5G to differentiate with other access networks and create new business models
- **Market**
  - Industrial applications robust market;
  - Consumer more slowly but still growing
- **Formats**
  - Need new low-latency codecs?
    - Do codecs need to be faster than framerates?
  - What is the best approach for Volumetric?
    - Point Cloud / Mesh: There seems to be place for both, standards will need to address the 2.

© 3GPP 2019

Thank you

From 3GPP SA4 Workshop organizing committee:

Frédéric Gabin (Ericsson)	**SA4 Chairman**
Gilles Teniou (Orange)	**SA4 Vice-chair**
Thomas Stockhammer (Qualcomm)	**XR on 5G rapporteur**
Paul Szucs (Sony)	**Workshop Host**

© 3GPP 2019

Annex: Time Warping concept

- How Timewarp Reduces Latency
  - Without Timewarp, your HMD would capture the data about the position of your head, render the image based on this data (correct angle etc.), then display the image when the next scene is due to be on screen. In a 60 fps game, a new scene is displayed once every 16.7 milliseconds. With this process, each image you see is based on the head-tracking data from almost 17 milliseconds ago.
  - With Timewarp, the first 2 parts of the process is the same. your HMD would capture the data about the position of your head and render the image based on the data. Before this image is displayed, your HMD captures the position of your head again. Using this information, the rendered image is modified with a mathematical calculation to fit the latest data. Finally the modified image is displayed on screen. The resulting image is more recent and more accurately depict the position of your head at the time of display than the image initially rendered. Timewarp only works in very short distances and time intervals or the resulting image will look unrealistic or out of place.
- Timewarp allows engines to increase or maintain frame rate when they are otherwise unable to do. It does this by artificially filling in dropped frames. In a game engine limited to 50 frames per second, a new frame is displayed once every 20 milliseconds. To increase the game's frame rate to 60, you need to display a new frame once every 16.7 milliseconds. To increase the fps through timewarp, the last completely rendered frame, not the one that is currently rendering, is updated with the latest data about the position of your head. The modified frame is displayed before your eyes. Theoretically, this method could increase frame rate from 1 to 60. Realistically, timewarp only works in very short distances and time intervals. Long distances or time intervals will make the image appear unrealistic or out of place.

© 3GPP 2019