Immersive Media Standardization - MPEG Overview

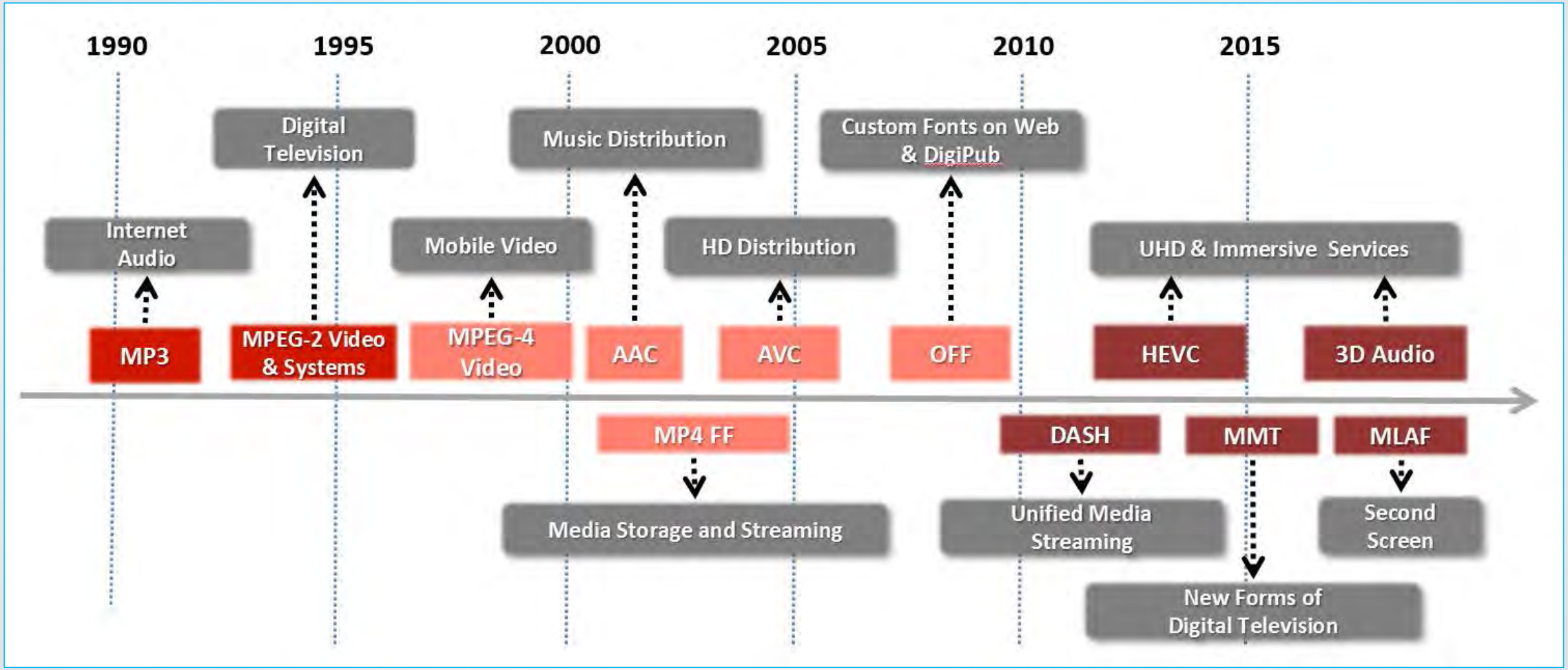
媒体技术实验室，华为中央研究院
2017年10月

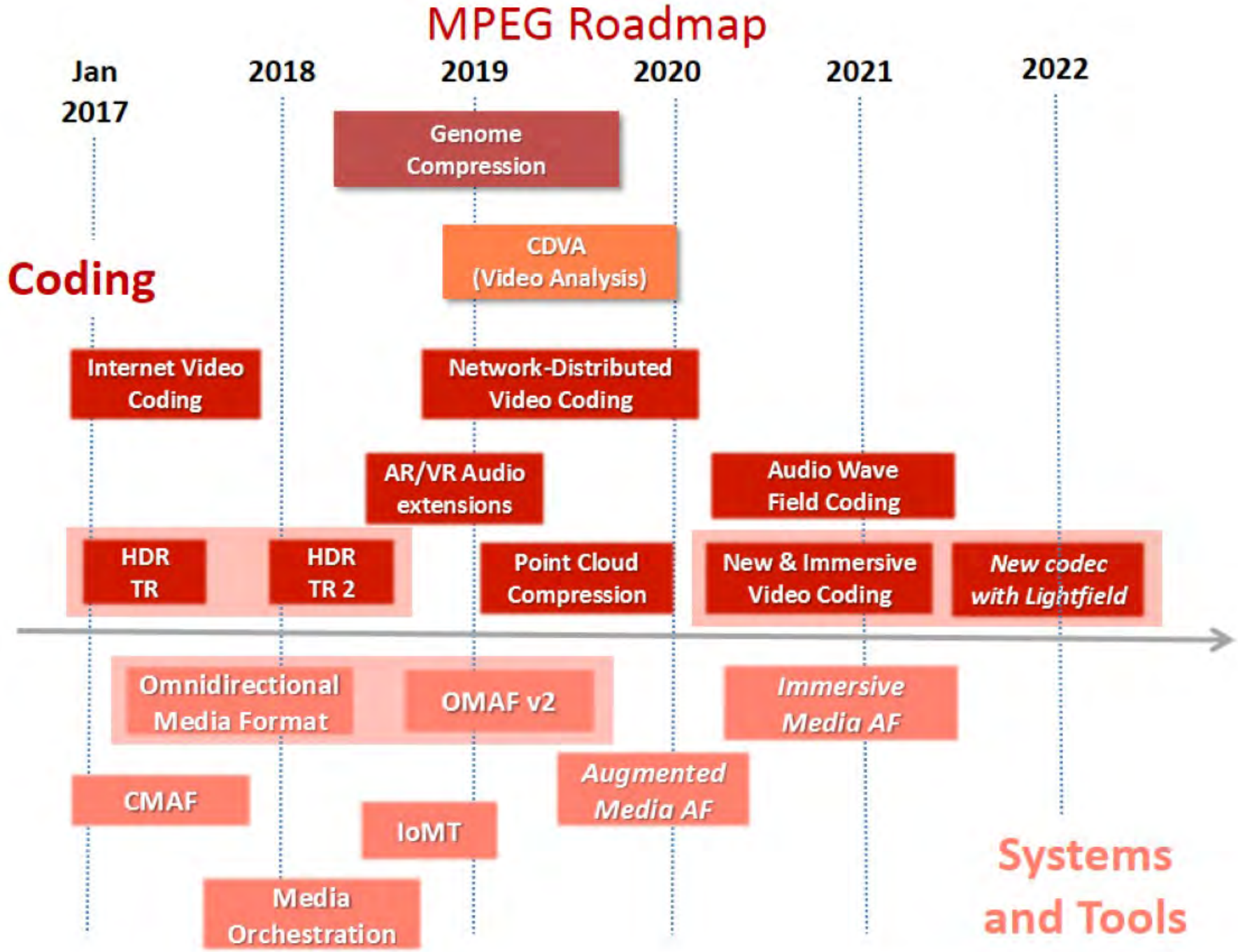
HUAWEI

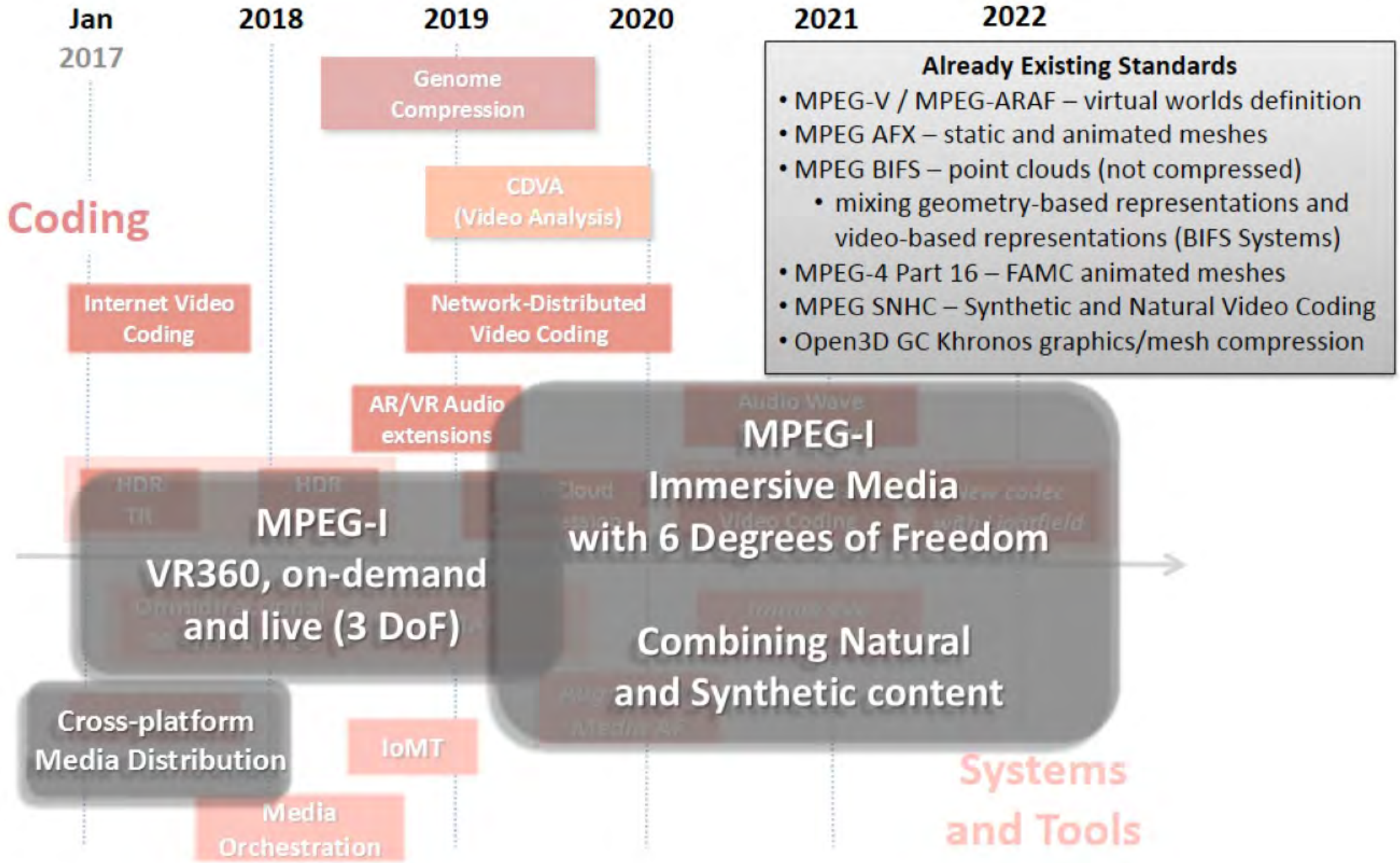
Immersive Standards Organizations Overview

- International Standards Development Organizations
  - ISO / IEC SC 29 (WG11 MPEG, WG1 JPEG)
  - ITU – SG16 VCEG
  - IEEE P2048, P1589
  - 3GPP SA4 Codec
- Industry Development Organizations
  - VRIF – VR Industry Forum
  - GVRA – Global Virtual Reality Association
  - DVB – CM Commercial Module VR Group
  - OpenXR – Khronos Group
  - CTA – Consumer Technology Association AR/VR WG
  - VESA – Video Electronics Standards Association AR/VR SIG
  - ITA – Immersive Technology Alliance
  - AREA – Augmented Reality for Enterprise Alliance
  - SMPTE – Society of Motion Picture and Television Engineers
  - SID – Society for Information Display
- Open Source Alternatives
  - WebVR – Progressive Enhancement WG
  - IETF NETVC – Internet Video Codec
  - AOM – Alliance for Open Media

MPEG Standards




MPEG Roadmap
Jan 2017
2018
2019
2020
2021
2022
Genome Compression
CDVA (Video Analysis)
Coding
Internet Video Coding
Network-Distributed Video Coding
AR/VR Audio extensions
Audio Wave Field Coding
HDR TR
HDR TR 2
Point Cloud Compression
New & Immersive Video Coding
New codec with Lightfield
Omnidirectional Media Format
OMAF v2
Immersive Media AF
Augmented Media AF
CMAF
IoMT
Media Orchestration
Systems and Tools


Jan 2017
2018
2019
2020
2021
2022
Genome Compression
CDVA (Video Analysis)
Coding
Internet Video Coding
Network-Distributed Video Coding
AR/VR Audio extensions
Already Existing Standards
• MPEG-V / MPEG-ARAF – virtual worlds definition
• MPEG AFX – static and animated meshes
• MPEG BIFS – point clouds (not compressed)
• mixing geometry-based representations and video-based representations (BIFS Systems)
• MPEG-4 Part 16 – FAMC animated meshes
• MPEG SNHC – Synthetic and Natural Video Coding
• Open3D GC Khronos graphics/mesh compression
MPEG-I
VR360, on-demand and live (3 DoF)
MPEG-I
Immersive Media with 6 Degrees of Freedom
Combining Natural and Synthetic content
Cross-platform Media Distribution
IoMT
Media Orchestration
Systems and Tools

MPEG-I Overview

ISO/IEC 23090 – Coded Representation of Immersive Media

MPEG-I Part 1: Architectures and Nomenclature for Immersive Media

MPEG-I Part 2: Omnidirectional Media Format (FDIS)

MPEG-I Part 3: Coded Representation of Immersive Video

MPEG-I Part 4: Coded Representation of Immersive Audio

MPEG-I Part 5: Coded Representation of Point Clouds

MPEG-I Part 6: Metadata

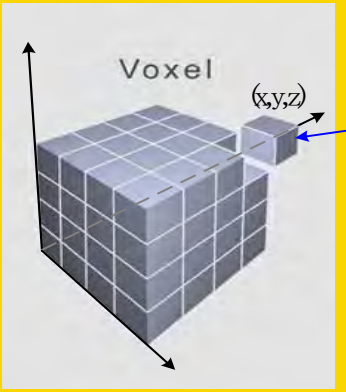
MPEG-I Part 7: Metrics

* MPEG has a history of creating media formats, compression technologies, system delivery technologies

Data Representation for Immersive Media

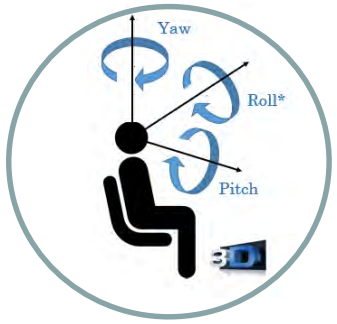
Immersive Media	2.5D	3D
Static	Panoramic Image	3D Image （ Model ）
Dynamic	Panoramic Video	Volumetric Video

Pixel = (R,G,B)|(x,y)

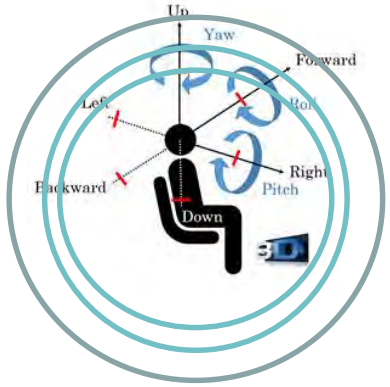


Voxel = (R,G,B…)|(x,y,z,V_x,V_y,V_z)

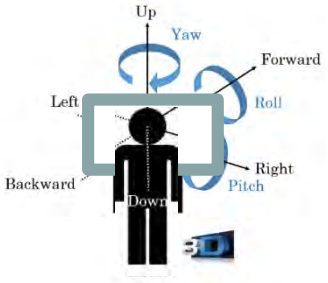
MPEG-I: Stages of Immersion



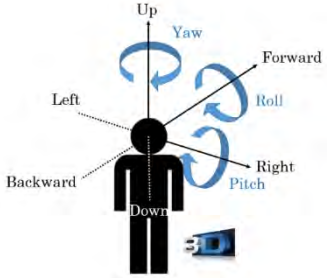
3DoF



3DoF+

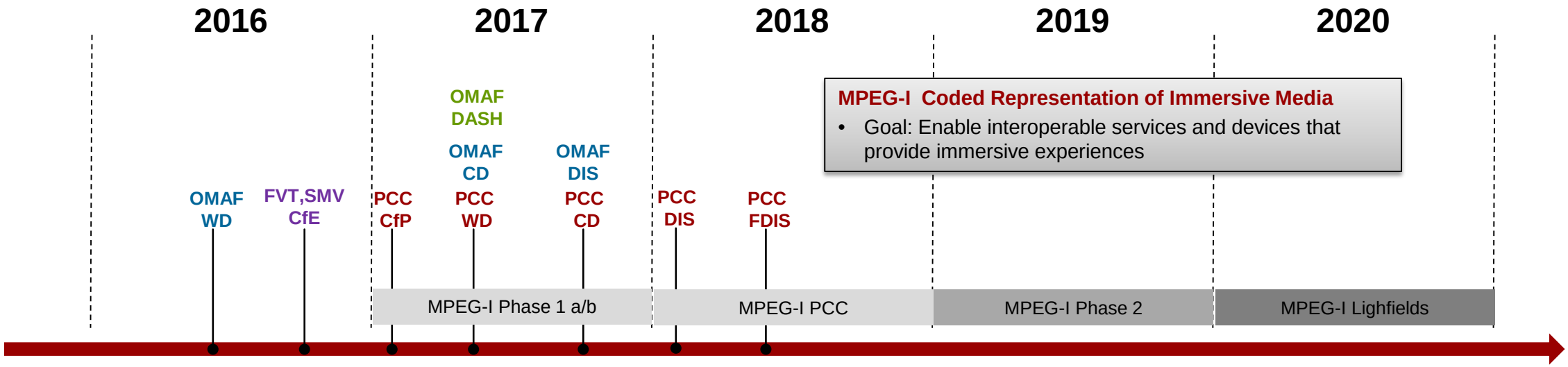


windowed 6DoF



6DoF

MPEG-I — Coded Representation of Immersive Media



MPEG-I Standardization

- Part 1: Technical Report on Immersive Media
- Part 2: Application Format for Omnidirectional Media
- Part 3: Immersive Video
- Part 4: Immersive Audio
- Part 5: Point Cloud Compression

Planned Work for MPEG-I

- Phase 1a
  - 3DoF 360 VR during **2017**
  - Use HEVC initially
  - Use new MPEG-H audio profile
- Phase 1b
  - Point Cloud Compression (PCC)
    - PCC Technical development **2017**
    - PCC Standardization starts in **2018**
  - Optimization of 3DoF 360 VR ongoing to **2020**
- Phase 2
  - Native VR, 6DoF Video
  - Audio, FVC Video, Graphics

MPEG Point Cloud Short Term Work Summary

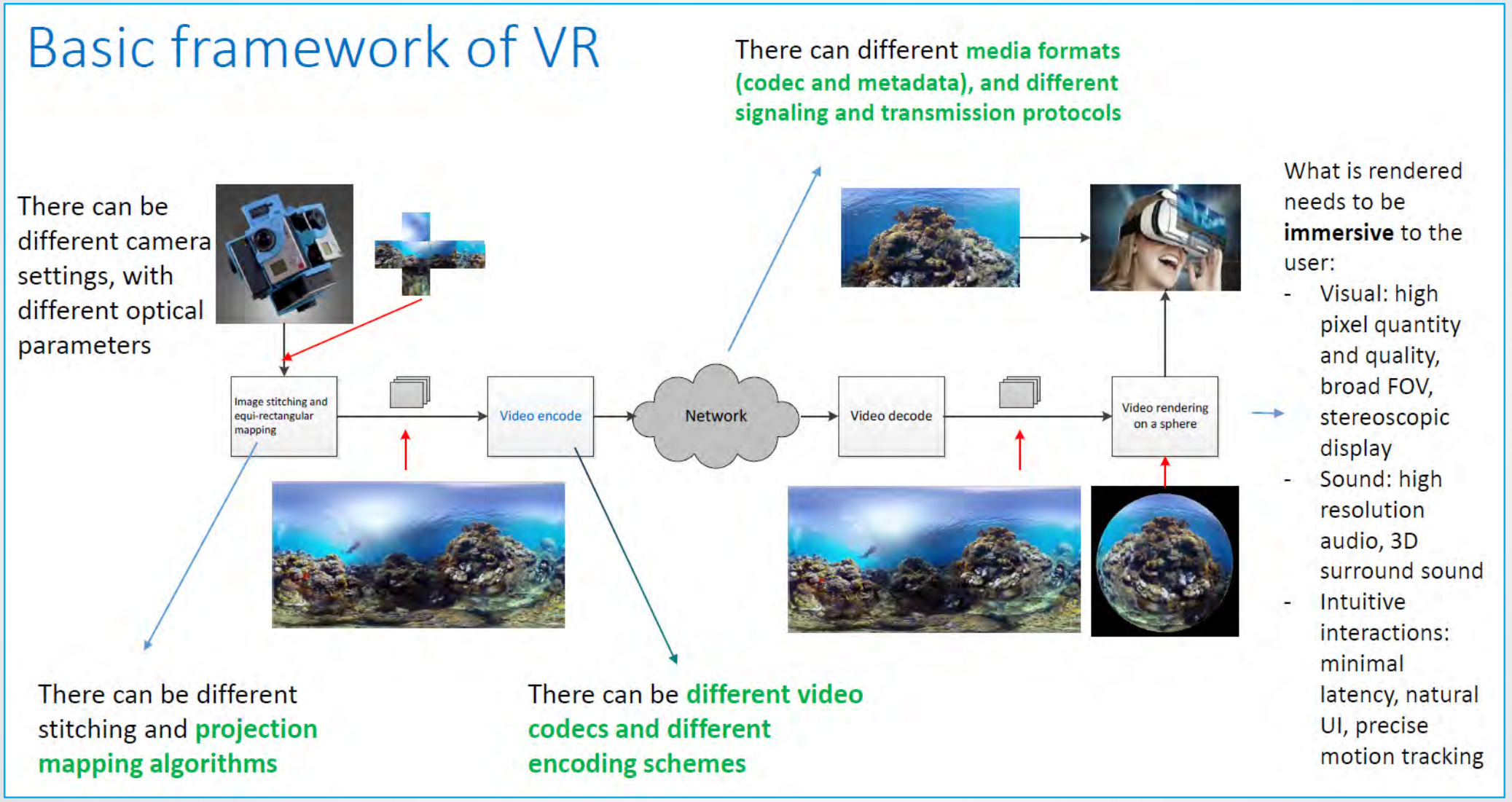
- Virtual objects with 360 Degree video
- Sensors (car, multimedia)
- Compression: Coordinates plus attributes
  - **Set** versus list
  - Issue for mesh compression
- Work on test data, anchor and evaluation criteria
- Draft CfP; CfP at Geneva (**Jan 2017**); Results Hobart

MPEG Lightfield Short Term Work Summary

- Call for test material Dense camera arrays
  - Lightfield cameras
  - Camera moved with robot
  - Data to be shared with JPEG
- Workplan towards 6DoF lightfield and VR

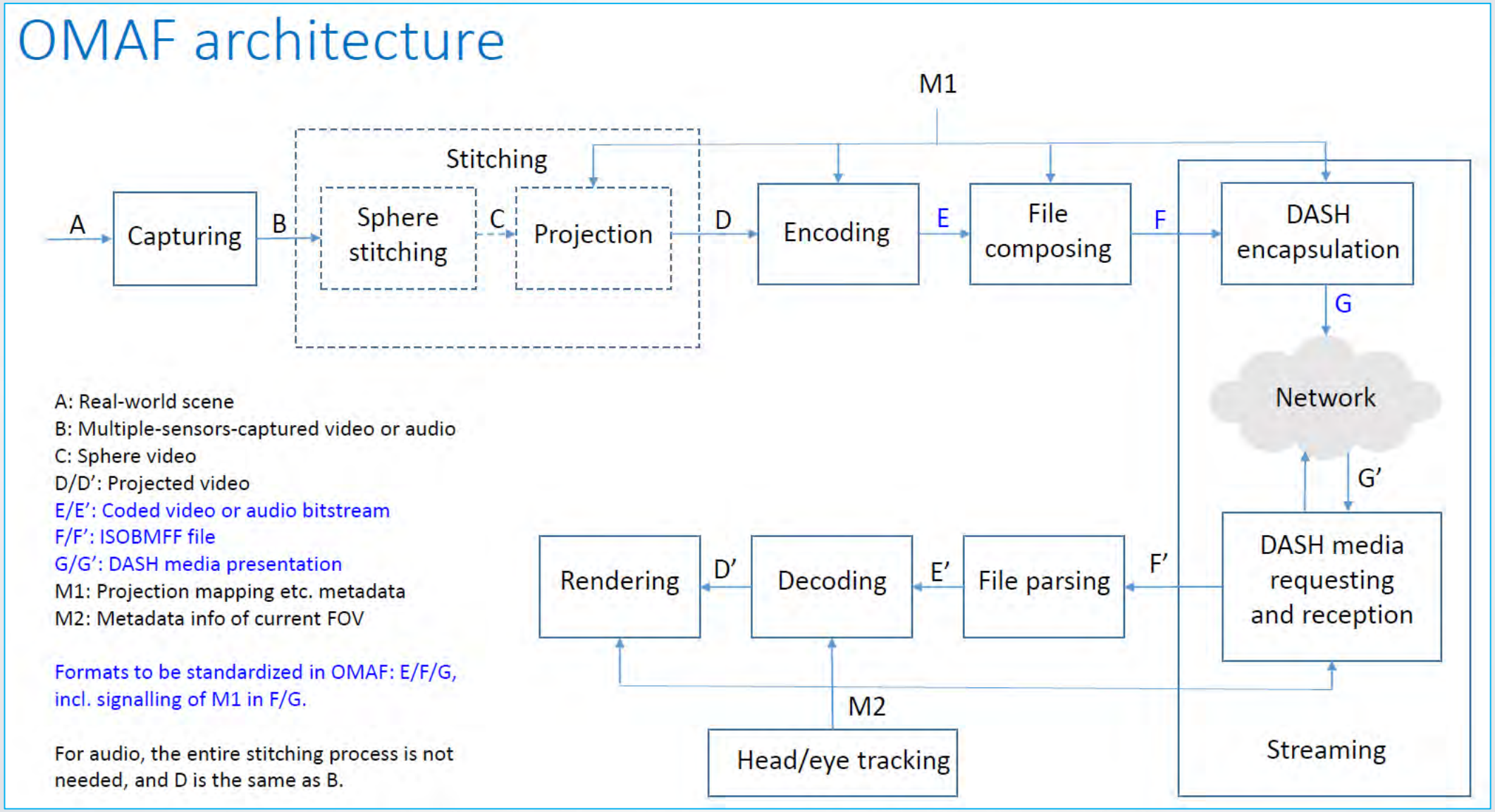
MPEG-I Part 2: OMAF
Omnidirectional Media Format

MPEG-I Basic Framework



Source: 2nd ITU-T Mini-Workshop on Immersive Live Experience (ILE), Geneva, Jan 2017

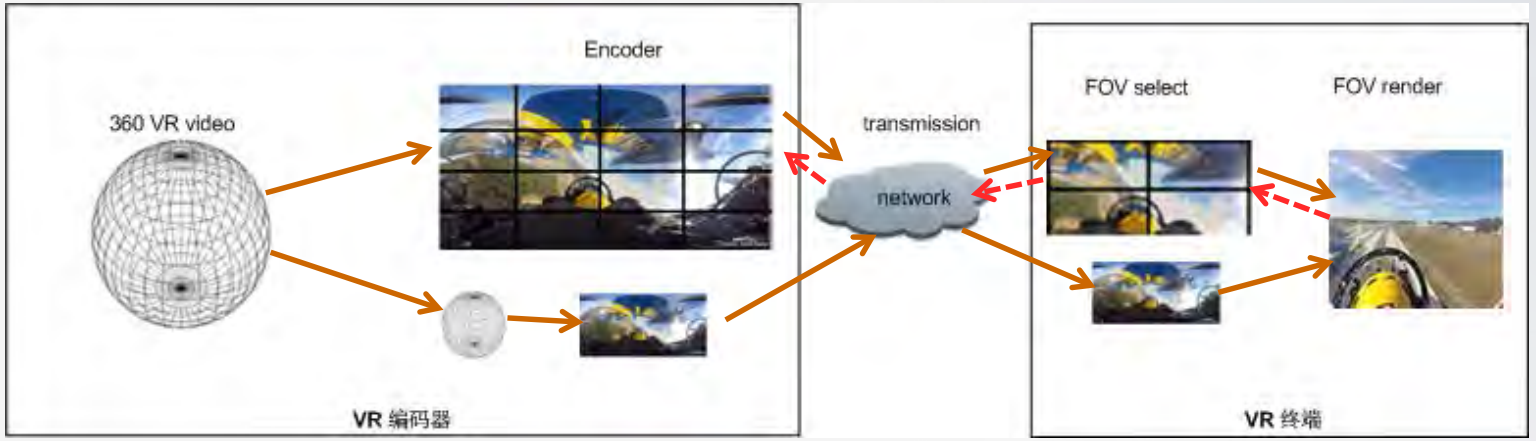
OMAF Architecture

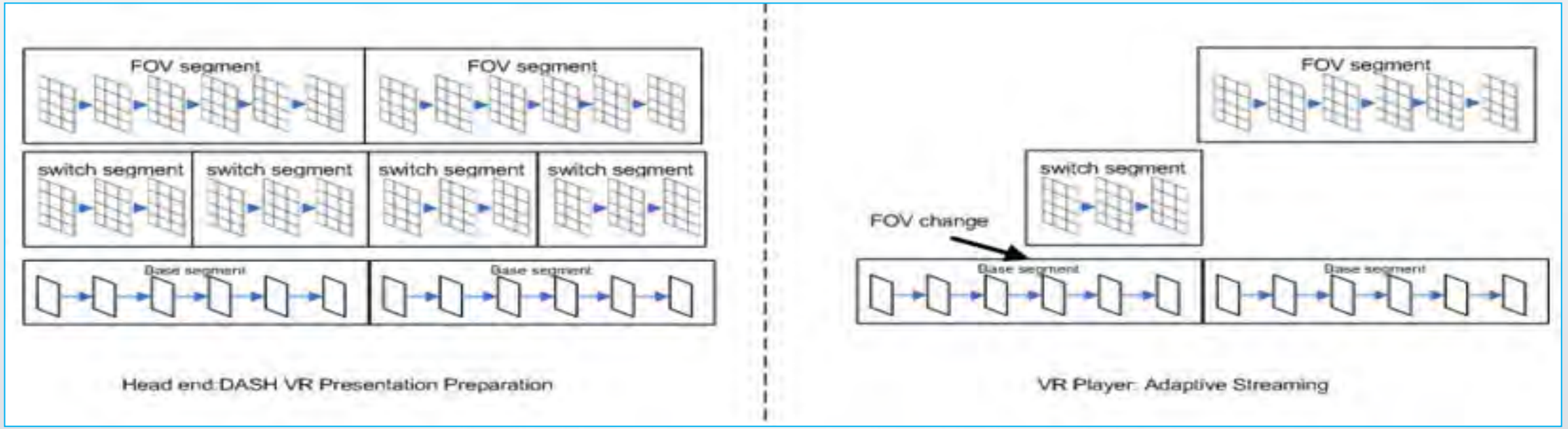


Source: 2nd ITU-T Mini-Workshop on Immersive Live Experience (ILE), Geneva, Jan 2017

DASH-VR Huawei Solution

- Coding all the big sphere content will generate a big bit stream, but user view dependent coding will require a strict e2e delay. The two requirements will not be matched in short time.
- Sending a small sphere bit steam and a user dependent view bit steam simultaneously will be a good compromise solution.
- Perception based Streaming is the key point.

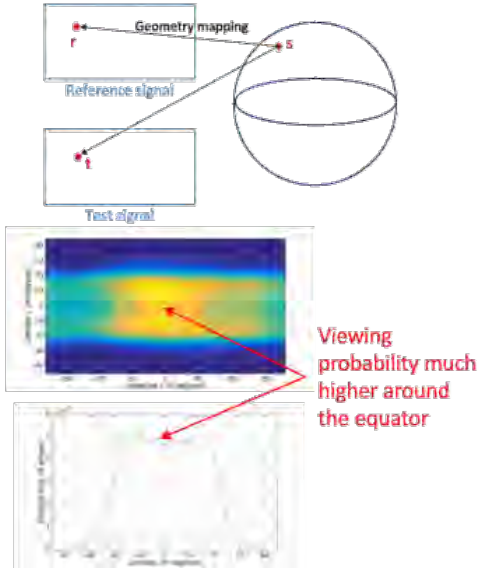




VR – Quality Evaluation

Spherical PSNR, etc

- SPSNR
  - A set of points (s) uniformly sampled on a unit sphere are mapped onto the test and reference videos
  - Error at these mapped coordinates (r and t) is computed
  - Set has 655362 uniformly sampled positions
- Weighted SPSNR (W-SPSNR)
  - Weigh the error at each position s according to its viewing probability
  - Map of point viewing probability collected with 10 viewers and 10 videos (10-second duration)
- Latitude weighted SPSNR (L-SPSNR)
  - Weights along the latitudes obtained by marginalizing along the longitudes
  - Latitude-based weights are applied on the errors



Area weighted PSNR (AW-PSNR)

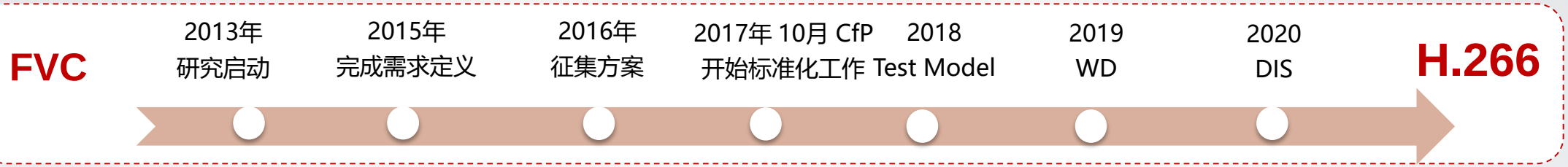
- Proposed by InterDigital and supported in PCT360
- SPSNR samples the sphere very sparsely
  - Less than 10% (655362/3840x1920 = 8.9%) of the projected video samples are used to calculate SPSNR
  - Interpolation will involve more sample positions but still not sufficient, and interpolation introduces other factors into the process
- AW-SPSNR:
  - Weigh the error at each position in the 2D plane according to the area it covers on the sphere
  - Applicable to all projection formats, though OHP not yet implemented in PCT360
- Using ERP as an example
  - For each (ue, ve) on a 2D plane with dimension (W, H), its sphere coordinates (φ, θ) are
    $\varphi = (ue/W-0.5)* 2*\pi$
    $\theta = (0.5-ve/H)* \pi$
  - Area covered on the sphere is
    $solid\_angle(\varphi,\theta) = \cos(\theta)*|d\theta|*|d\varphi|$
  - Weight is normalized to be $\cos(\theta)/(4*\pi)$, where $4*\pi$ is the total area of a unit sphere
- Very similar to W-SPSNR in MPEG document no. m38551 from Zhejiang University, but with proper normalization

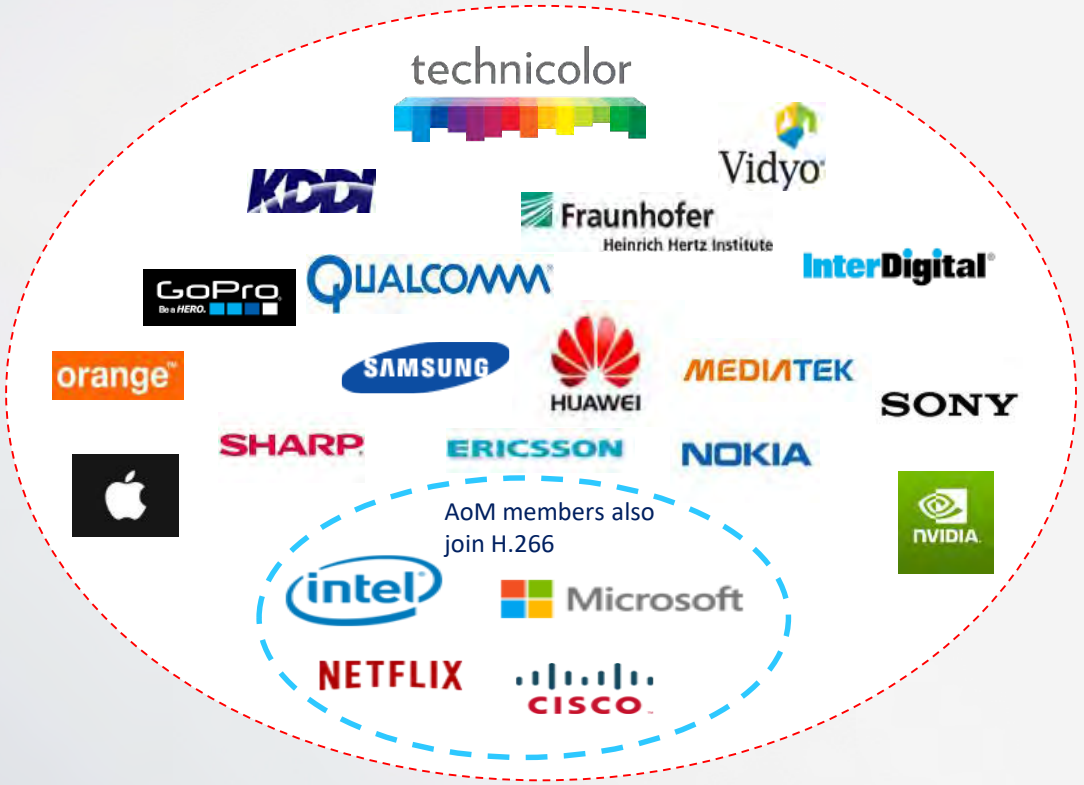
MPEG-I Part 3: Immersive Video Compression

“Today it is not clear whether FVC technologies beyond rectangular video and 3DoF will be hosted by Part 3 or in a new part”

L. Chiariglione – May 2017

MPEG/ITU FVC/H.266 Standard

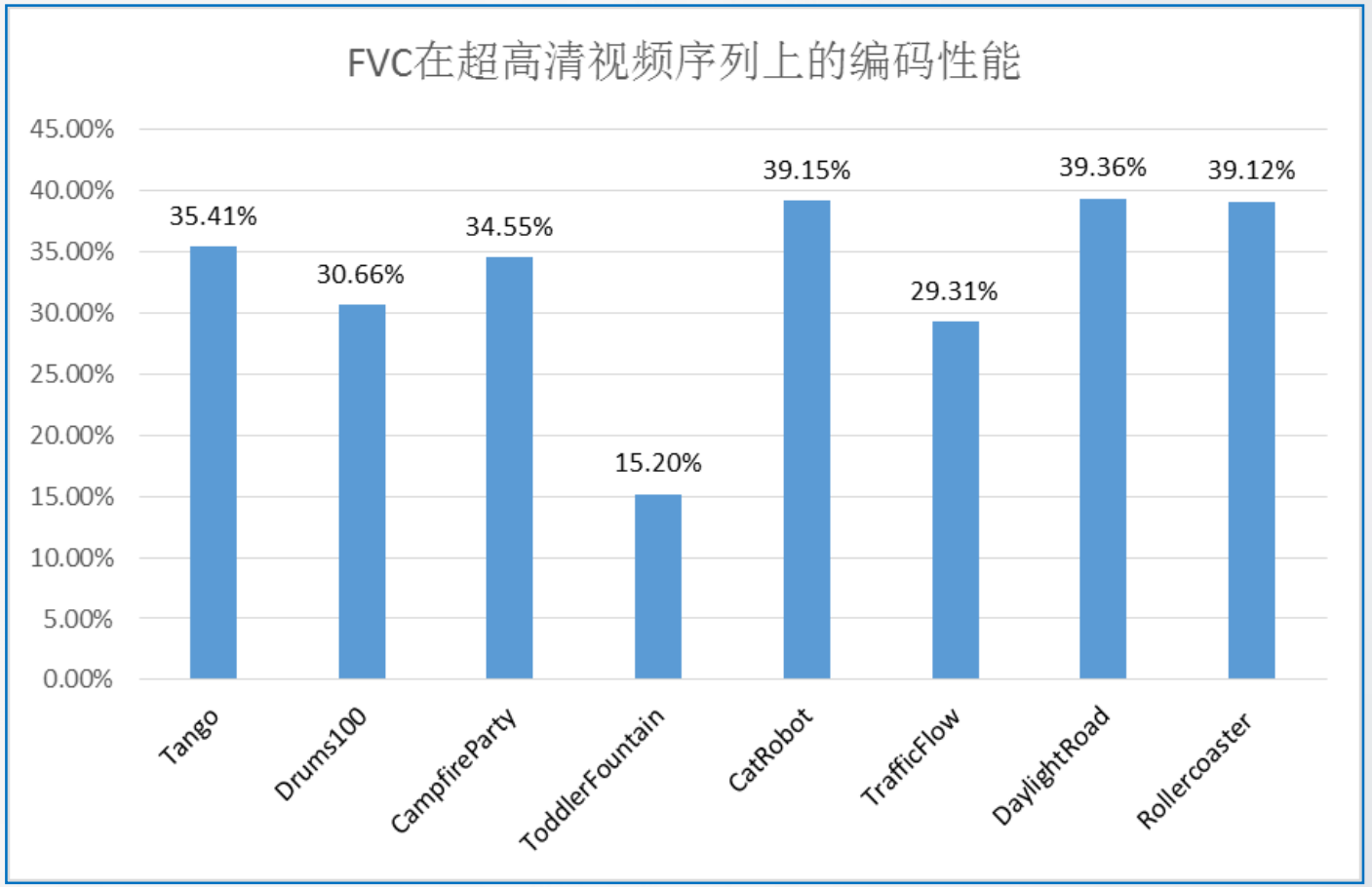




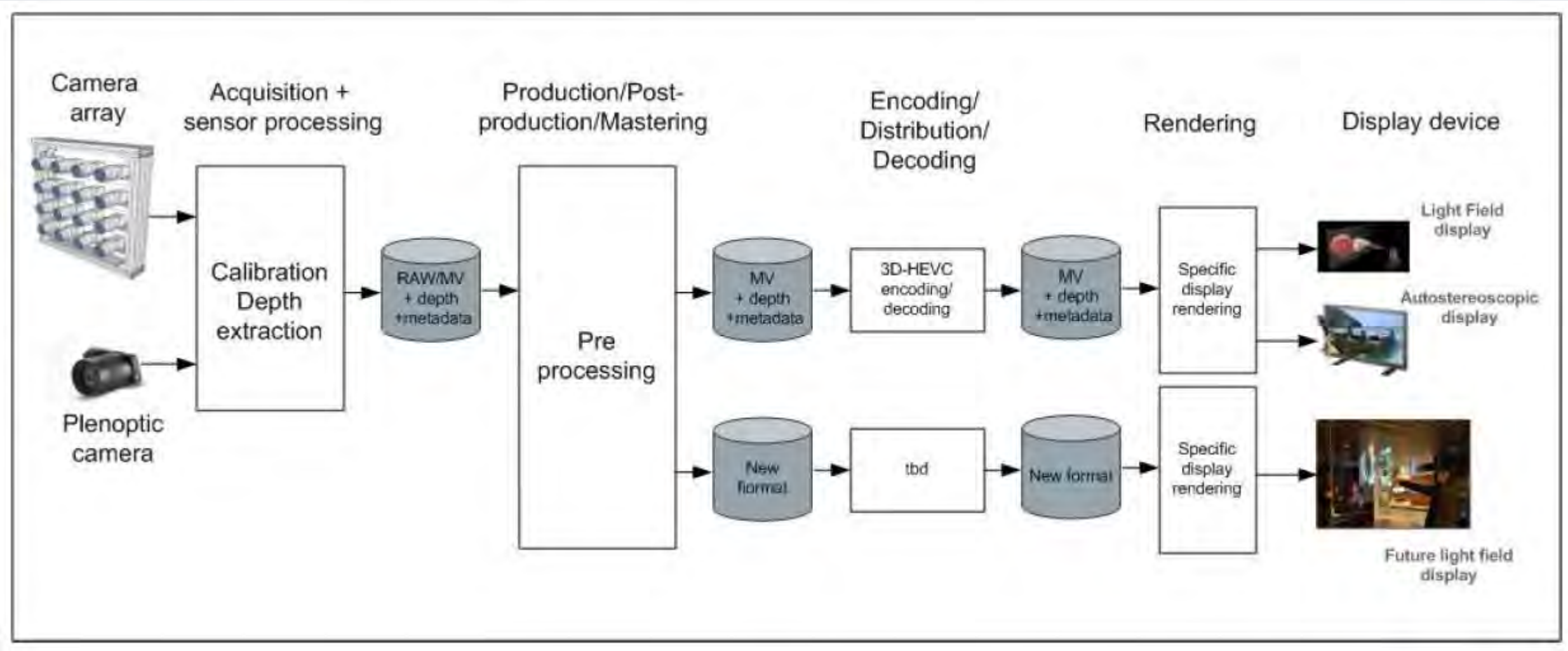
- **视频编码仍然是最受重视的MPEG标准，传统2D视频编码效率的提升，仍然具有重要的价值。**
- 传统娱乐视频之外的需求更为广泛，包括VR，UGC以及监控等场景
- 不仅是高分辨率，HDR，WCG对高体验同样重要
- **传统上每十年提升一倍压缩效率的节奏有加快的趋势**
  由JVET开发的JEM参考软件，相对HEVC已经取得30%的编码效率提升，业界的准备度与H.265同期时代相比可能比较接近。

FVC相对于HEVC的性能提升

- 2015年成立Joint Video Exploration Team（JVET），开发Joint Exploration Test Model（JEM），相对HEVC性能提升达到30%，关键技术包括：
  - 灵活的图像块划分技术
  - 解码端运动矢量推导
  - 基于仿射变换模型帧间预测技术
  - 基于维纳滤波器的自适应环路滤波
  - ……
- 目前JEM复杂度明显高于HEVC（8-10X），但暂时不是重点考虑的问题
- 基于机器学习的视频编码越来越受到重视，但短期内可能难以颠覆传统的视频编码技术框架



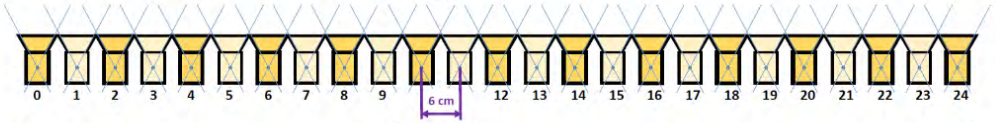
Beyond VR Streaming: Light Field Processing



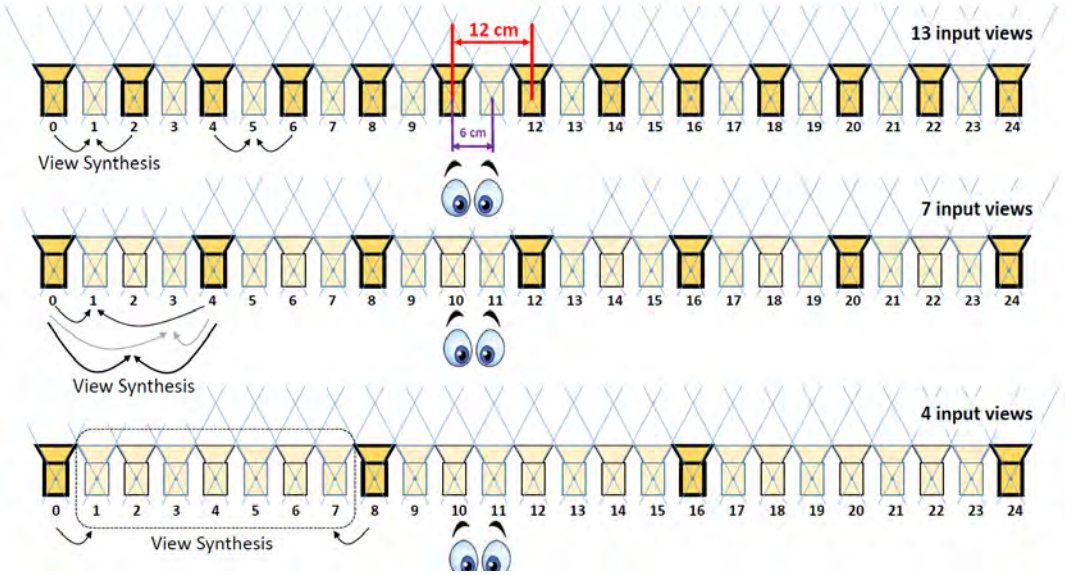
View Synthesis for 6DoF VR

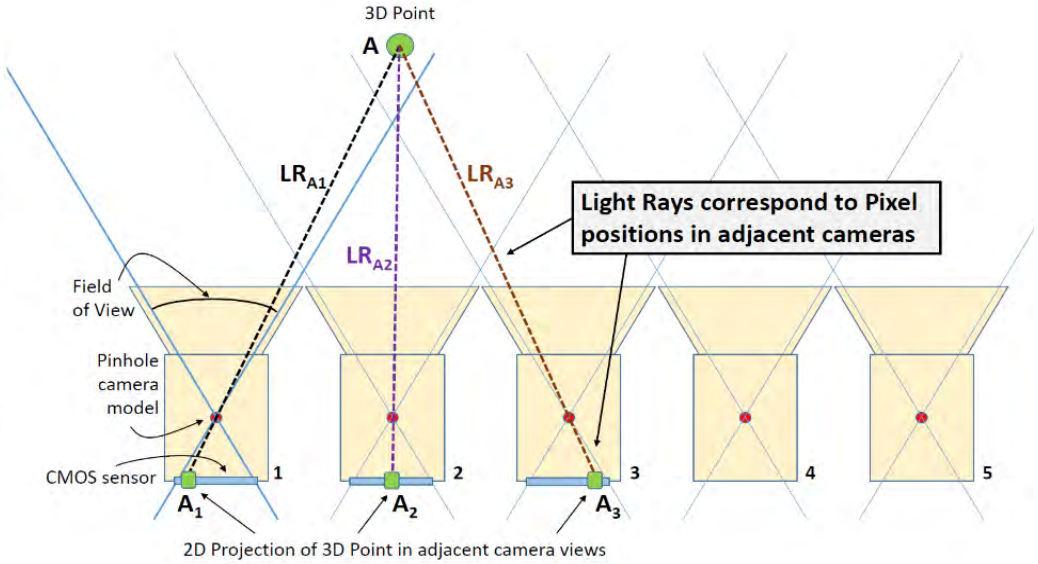
25 cameras =
¼ of 100 cameras

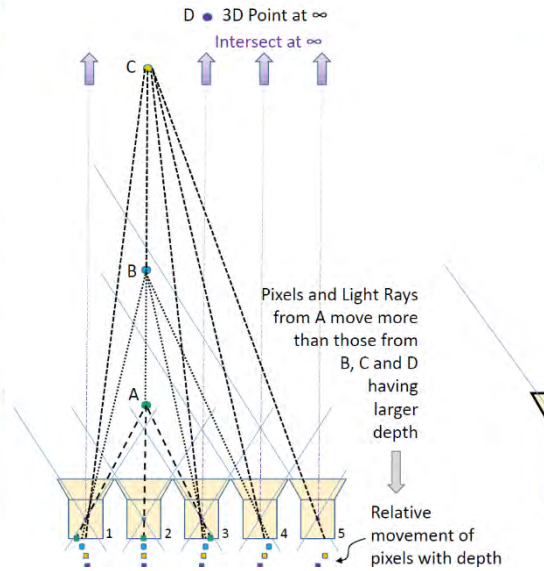
If cameras can be put max. 6 cm apart, they respect the inter-eye distance for stereo viewing (even then transition is discontinuous)

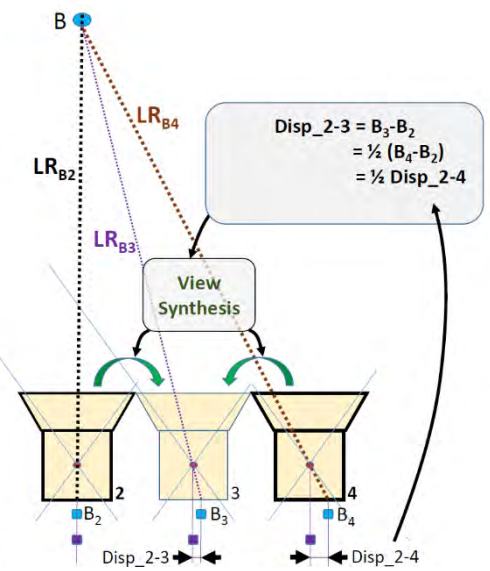


Project all views and smoothen view transitions with diffuser lens









Source: View synthesis for 6DoF with Lightfields, G. Lafruit ULB, Brussels

MPEG-I Part 5: Point Cloud Compression

“This is an area of research (in MPEG) for which scarce results are available”

L. Chiariglione – May 2017

MPEG 3D Graphics Activities

MPEG has been working on Graphics for a long time

MPEG 3DG Technologies

1992/4 — MPEG-1/2 (AV content)

1997 — VRML

1998 — MPEG-4 v.1
- Part 11 - BIFS:
  - -Binarisation of VRML
  - -Extensions for streaming
  - -Extensions for server command
  - -Extensions for 2D graphics
  - - *Real time augmentation with audio & video*
- Part 2 - Visual:
  - - 3D Mesh compression
  - - Face animation

1999 — MPEG-4 v.2
- Part 2 – Visual
  - - Body animation

First form of broadcast signal augmentation

MPEG 3DG Technologies

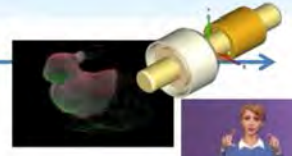
2003 — MPEG-4
- Part 16 - AFX:
  - - A rich set of 3D graphics tools
  - - Compression of geometry, appearance, animation

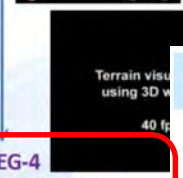
2005 — MPEG-4
- AFX 2nd Edition:
  - - Animation by morphing
  - - Multi-texturing

2007 — MPEG-4
- AFX 3rd Edition
  - - WSS for terrain and cities
  - - Frame based animation

2011 — MPEG-4
- AFX 4th Edition
  - - Scalable complexi mesh coding

A rich set of 3D Graphics representation and compression tools





Frame-based Mesh Compression - FAMC

❑ What if no skeleton is defined, only animated meshes?

❑ FAMC: cluster the vertices and encode cluster animation parameters plus corrections

SPLASH – SPLAts and meSH

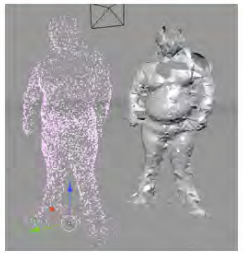
Mesh

Splats + Mesh = SPLASH

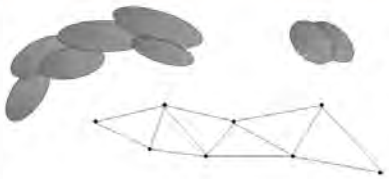
Splats

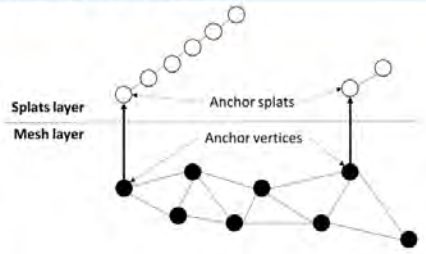
3DGfx Data representation forms closed to Real World capture

- Entire object is expressed as a point cloud
- Some regions of the object are meshed, some other are point clouds
- Both forms shall be animated (including variation of the connectivity)



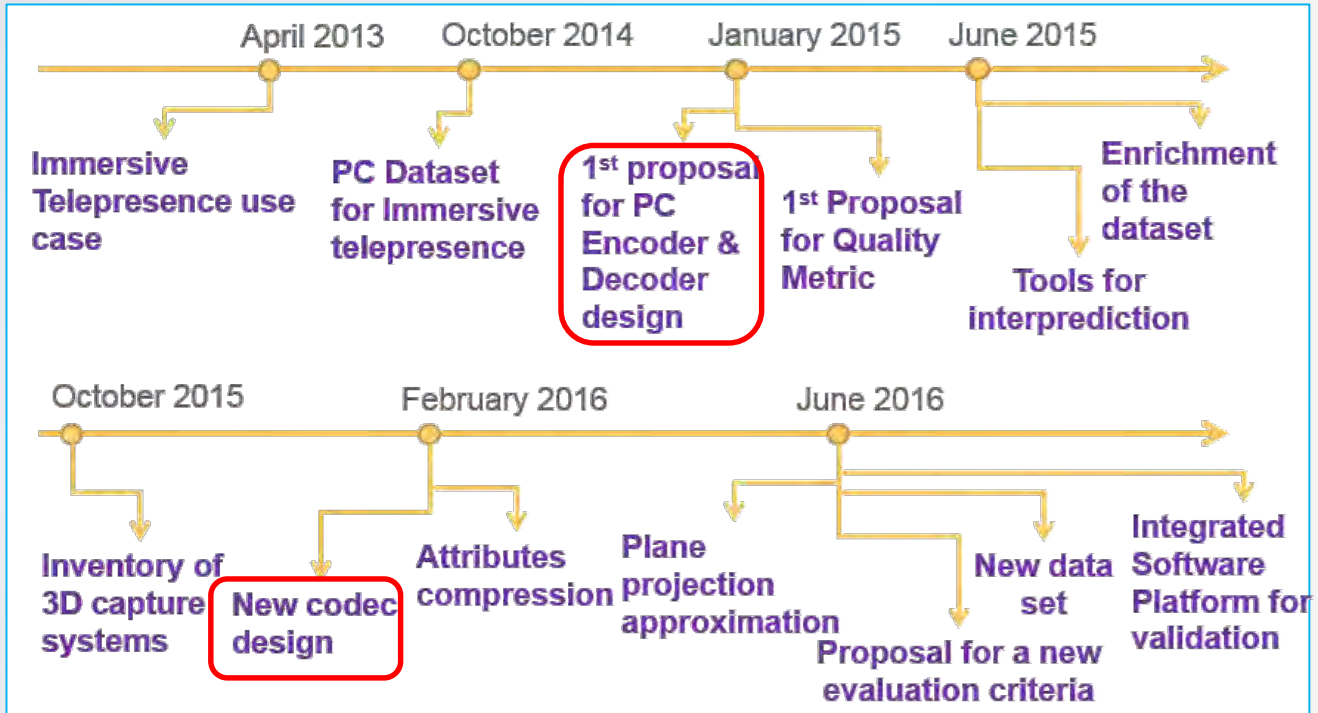
SPLASH – SPLAts and meSH





MPEG Point Cloud Compression

MPEG 3DG Point Cloud Compression Timeline



Preliminary Development Plan:

Year	Month	Day	MPEG meeting	City	Country	Stage
2017	01	20	117	Geneva	CH	Approval of CfP
2017	10	27	120	Macao	CN	Approval of WD
2018	04	20	122	San Diego	US	Approval of CD
	07	20	123	Ljubljana	SI	Approval of DIS
	10		124	??	CA	Approval of FDIS

PCC - Requirements

- 3D Point Cloud Representation
  - X, Y, Z
  - multiple attributes being associated with each 3D position including color, reflectance, normal vectors and transparency
  - view-dependent attributes
  - time-varying
- 3D Point Cloud Compression
  - Lossy
  - Lossless
  - Progressive and/or scalable coding
  - View-dependent coding
  - Random Access
  - Error resilience

Point Cloud Compression: Content categories

Test Category	Information to be encoded
Category 1. Static Objects and Scenes	X, Y, Z, R, G, B
Category 2. Dynamic Objects	$(X, Y, Z, R, G, B)_i$
Category 3. Dynamic Acquisition	$(X, Y, Z, R, G, B, I)_i$

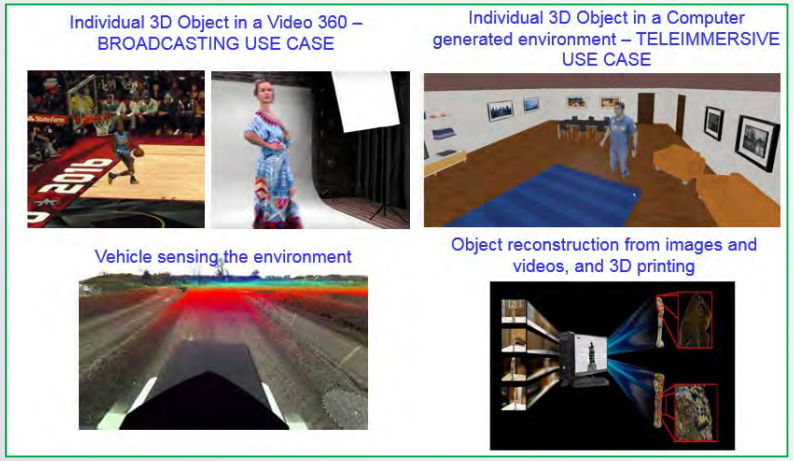
MPEG Point Cloud Compression

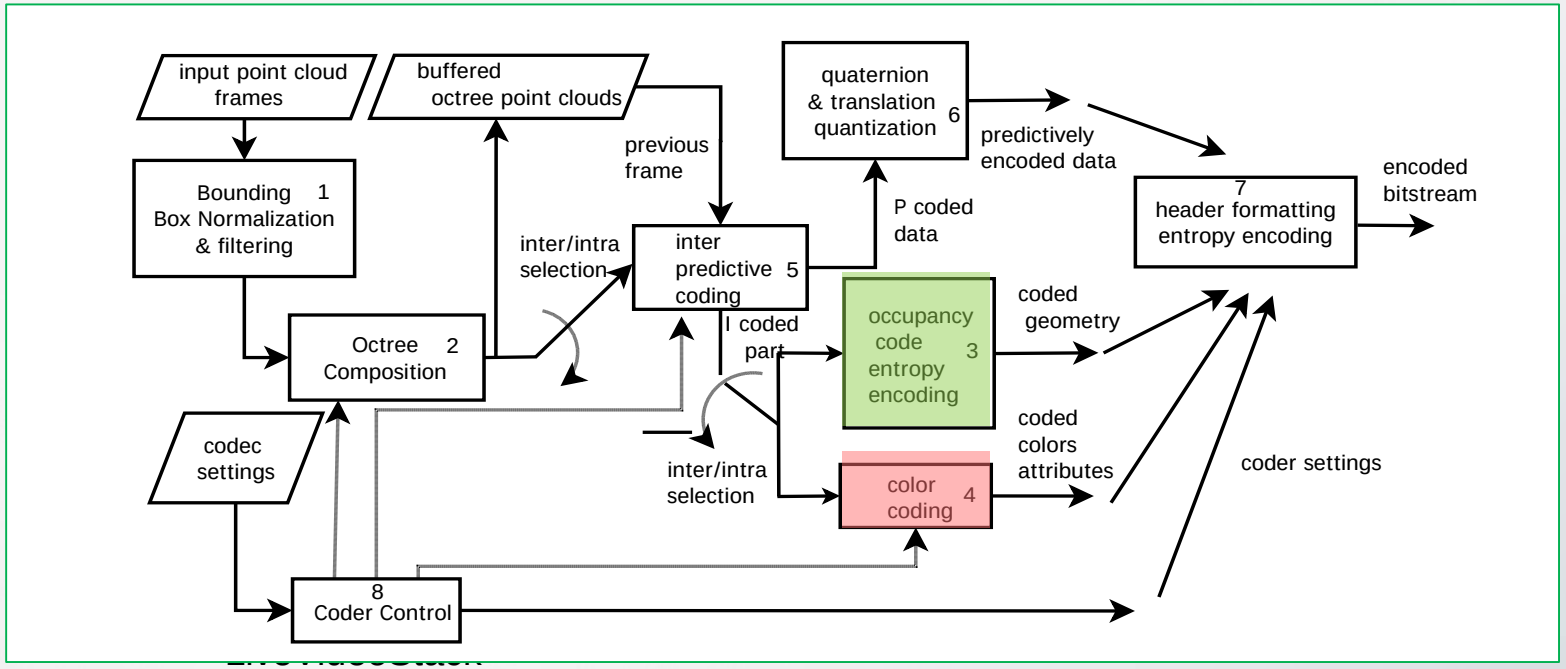
Point Cloud Compression: Category 1

Point Cloud Compression: Category 2

Point Cloud Compression: Category 3

Point Cloud Use Cases





Summary

- Immersive Media is here to stay and will continue to change the way people consume information.
- New immersive experiences generate orders of magnitude more data than today.
- Compression will continue to play a fundamental role for commercial deployment.
- International SDOs are committed to the creation of new immersive standards.
- The first deployment of immersive media is 360 VR and the main standardization effort is MPEG OMAF.
- Today there is industry fragmentation, lack of killer apps and inflated consumer expectations … but that will change as the technology for capture, processing, transmission and display evolves.
- Huawei Media Lab is making key contributions to the normative parts of the standard, including MPEG OMAF standardization. We also actively participate in MPEG-I Point Cloud Compression and Lighfields Compression standardization.

THANK YOU