



ISO/IEC JTC 1/SC 29/WG 03

MPEG Systems

Convenorship: KATS (Korea, Republic of)

Document type:	Output Document
Title:	Presentation to Metaverse Standards Forum on MPEG-I Scene Description
Status:	Approved
Date of document:	2023-07-20
Source:	ISO/IEC JTC 1/SC 29/WG 03
No. of pages:	35 (with cover page)
Email of Convenor:	young.L @ samsung . com
Committee URL:	https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg3



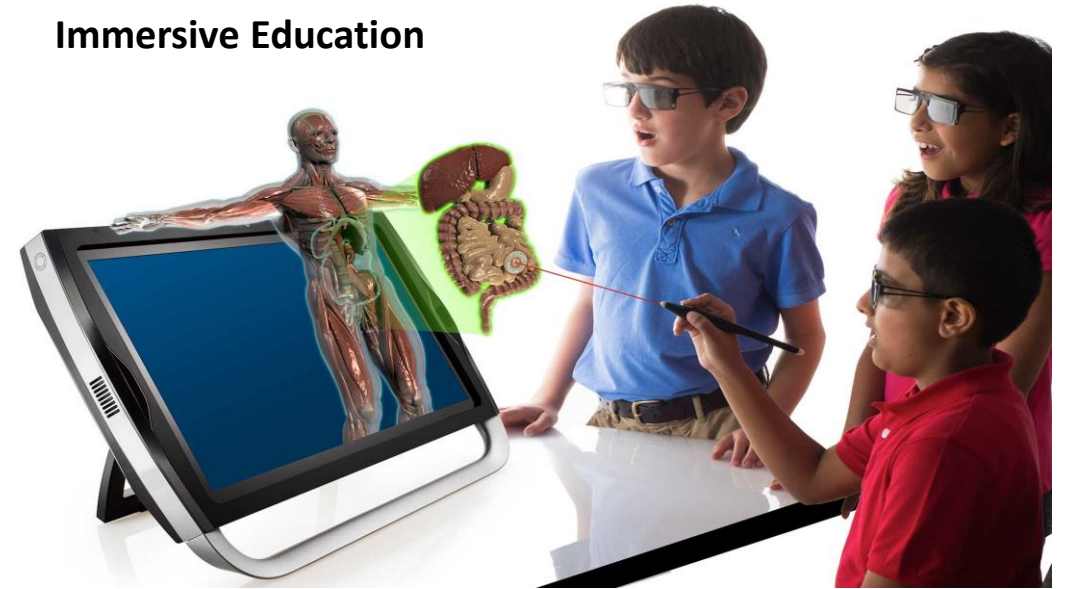
gITF 2.0 Extensions in MPEG and 3GPP for the Metaverse

Imed Bouazizi
Patrice Hirtzlin

Metaverse

- Infinite worlds of immersive experiences
- Diverse Use Cases
- Potential to disrupt today's web surfing experience

Immersive Education



Immersive Entertainment



Immersive Gaming

Relevant Standardization in MPEG and 3GPP



Scene Description

Video Decoding Interface (VDI)

Video, Audio, Haptics

Point Cloud and Mesh
compression



XR Architectures

Media Formats

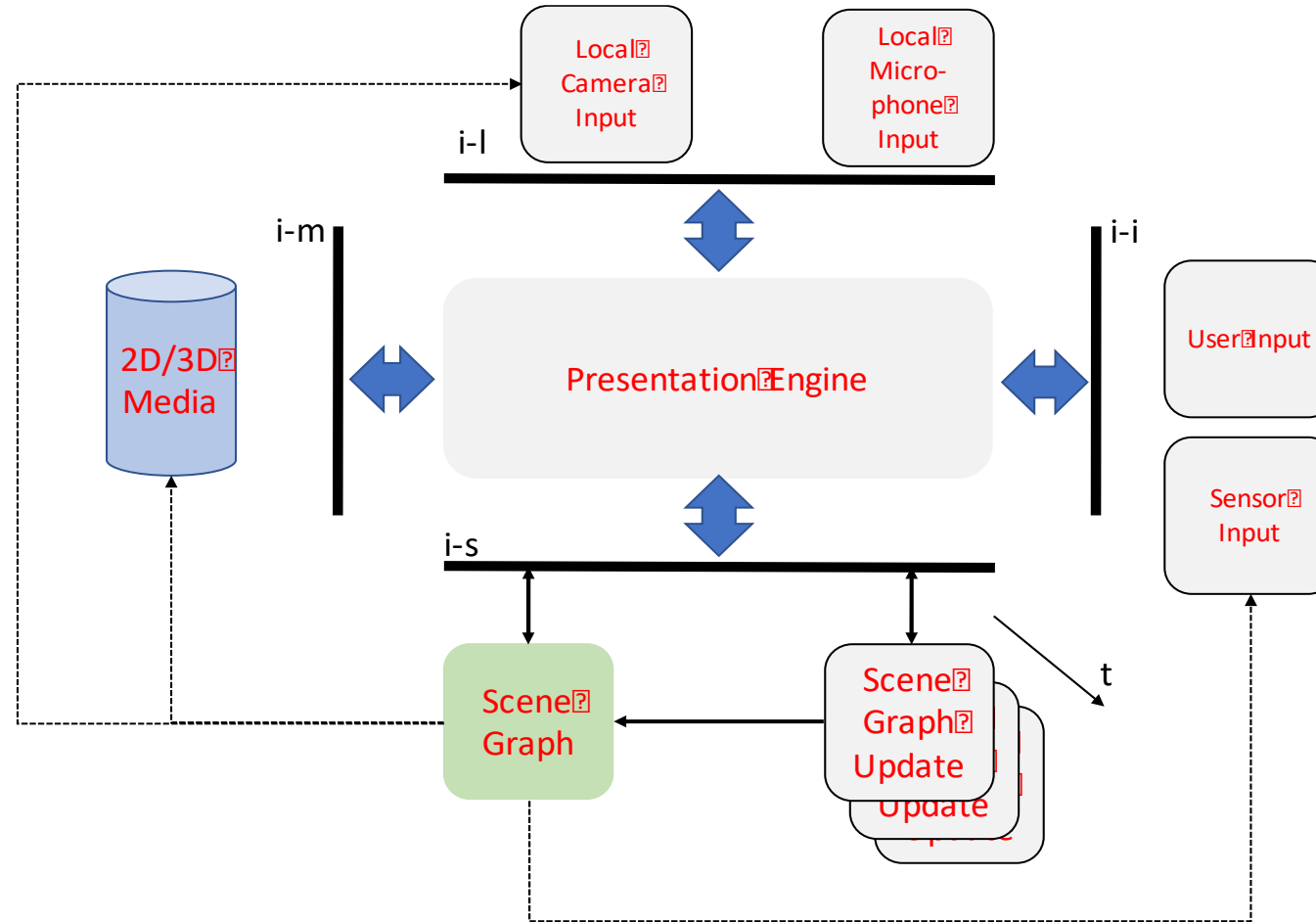
Speech Codecs for XR

Split Rendering

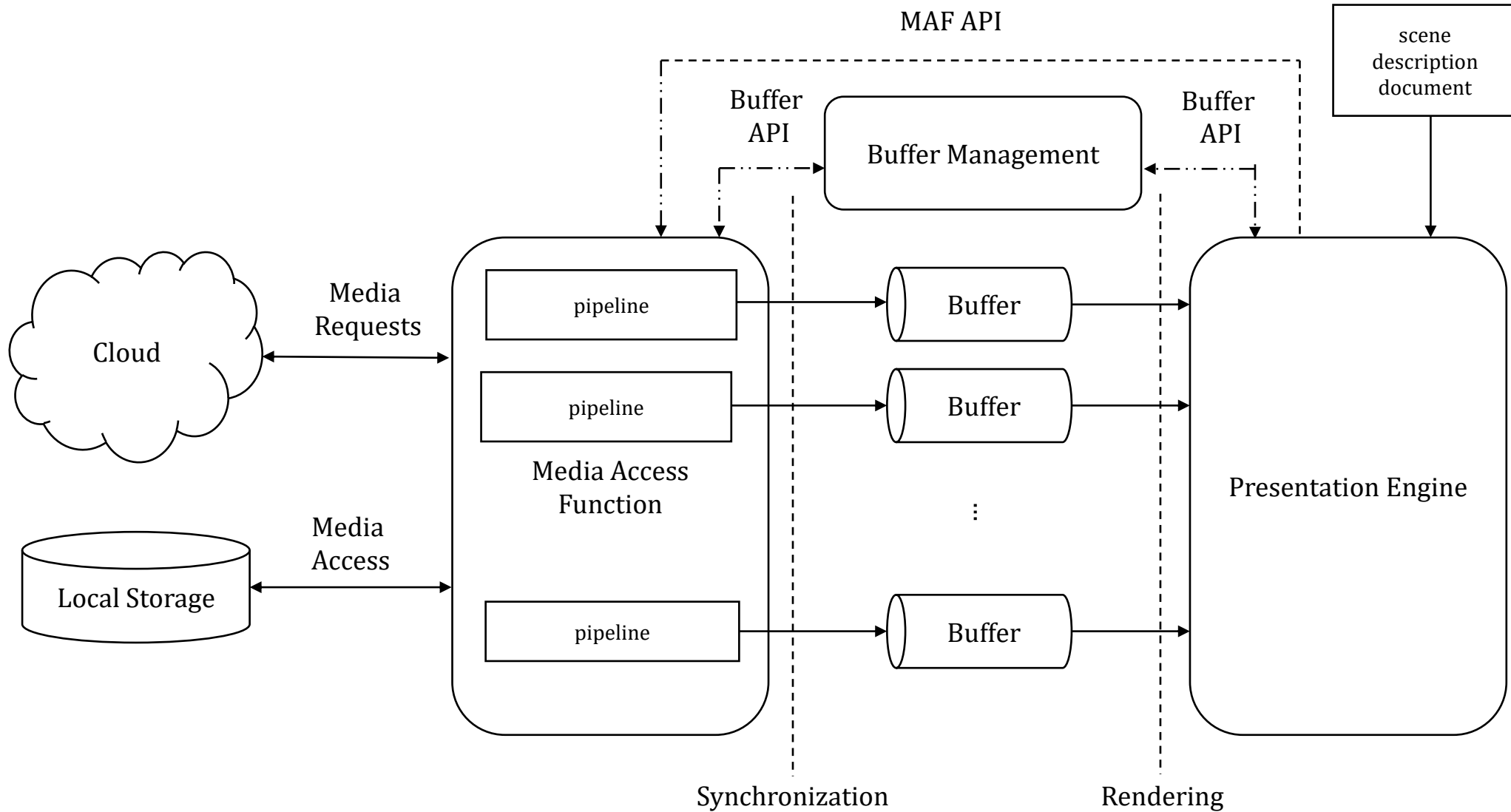
5G system optimizations for XR
QoS and power consumption

MPEG Immersive Media Architecture

Interfaces

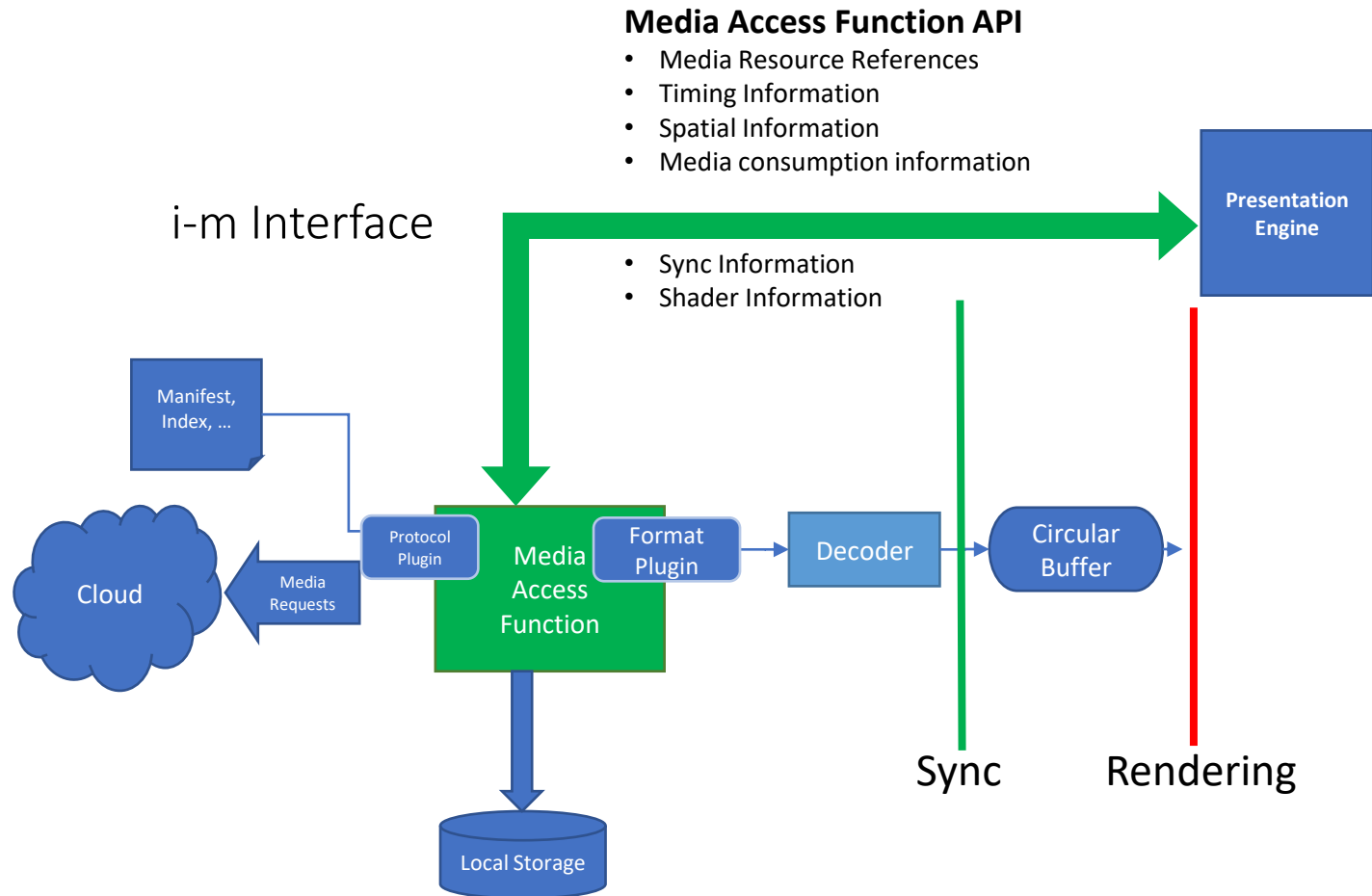


Scene Description Architecture



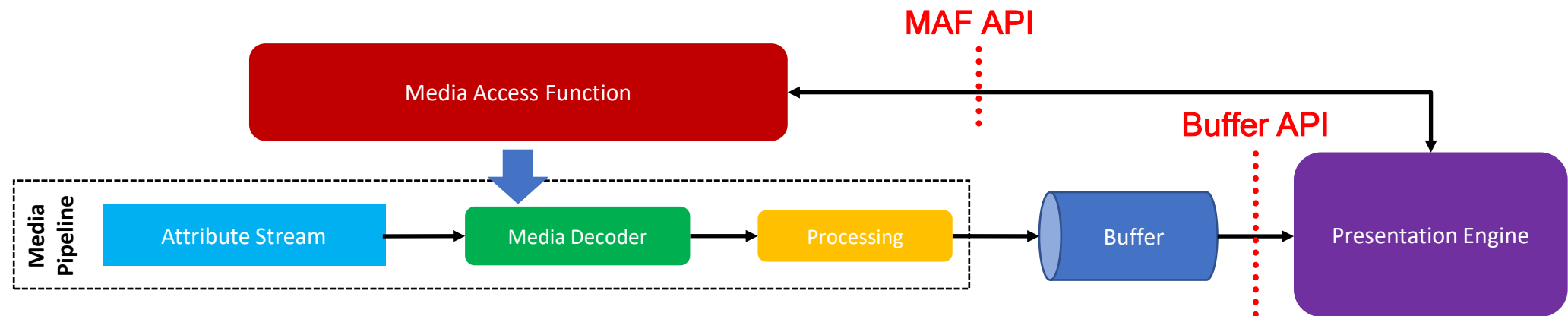
Media Access Function (MAF) API

- Support for wide range of formats through Plugins
- Endpoint for the Media Access Function API
- Optimized Media Fetching
 - Random spatial and temporal access
 - Partial delivery matching Presentation Engine needs
 - Integration with Cloud and Edge media processing



Media Pipelines

- The MAF instantiates and manages Media Pipelines
 - A media pipeline typically handles content of an attribute/component of an object/mesh
 - It produces content in the format indicated by the glTF file
 - The formatted frame is then pushed into the circular buffer
- Media Pipelines are highly optimized and customized for the type and format of media that is being fetched
- Media Pipeline maintains sync information (time and space) and passes that information as buffer metadata



MPEG-I Scene Description

Last Mile vs Interchange

Last Mile

- Low Complexity
- Flat Hierarchy
- Compressed components
- Adaptive and network friendly
- Support for Texture/Light Baking

Interchange

- High fidelity
- Superstructure
 - Hierarchical
 - Distributed
 - Preserves author's intents/choices
 - Documents authoring process
- Lossless
- Preserves asset's metadata/versioning










Last Mile



Interchange

Towards the MPEG of 3D

glTF - The JPEG of 3D!

Audio	Video	Images	3D
			
			New market opportunities for 3D content creation and deployment!

Time Dimension

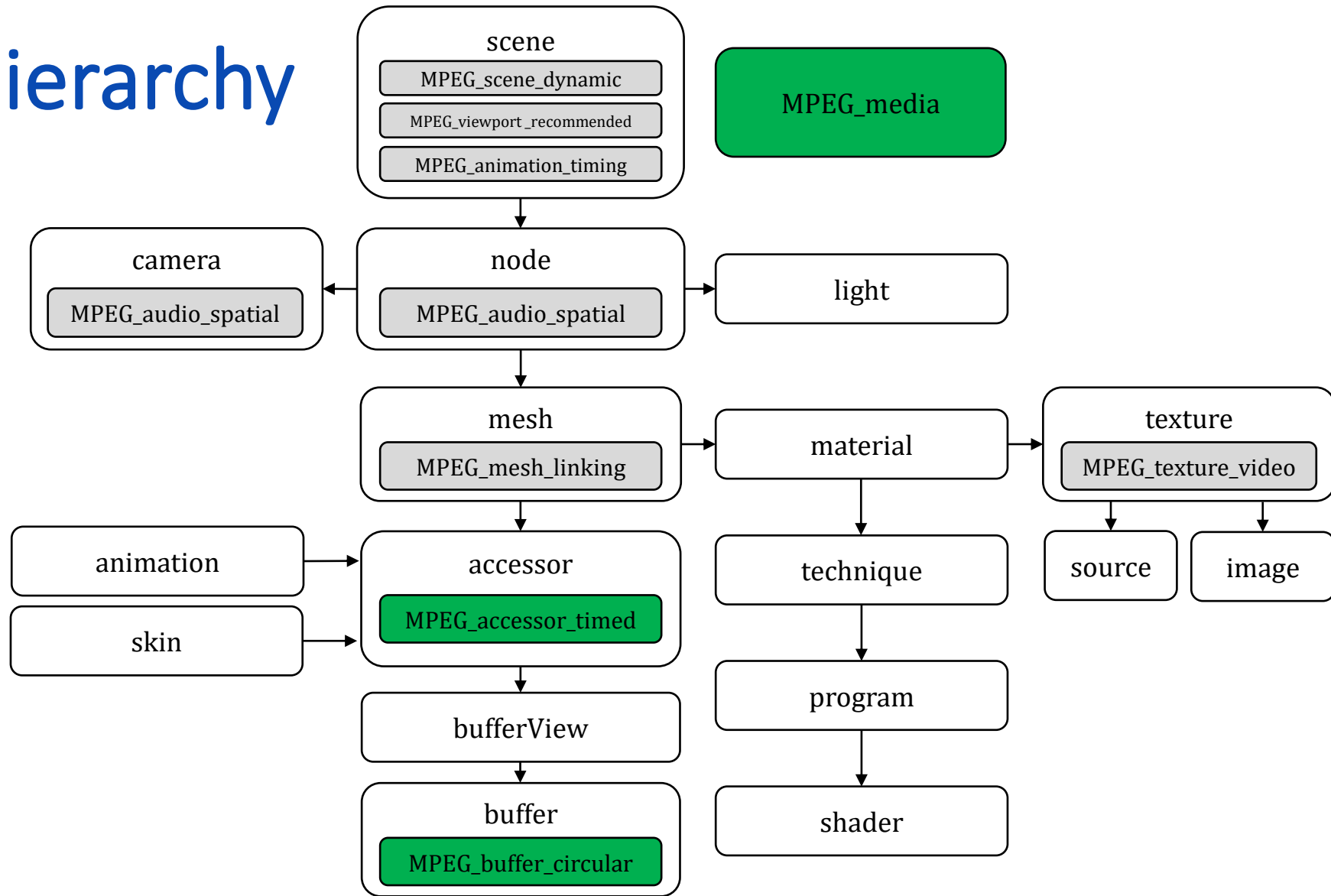


Dynamic 6DoF Scenes



Node Hierarchy

 Core Extensions



External Media References

MPEG_media extension

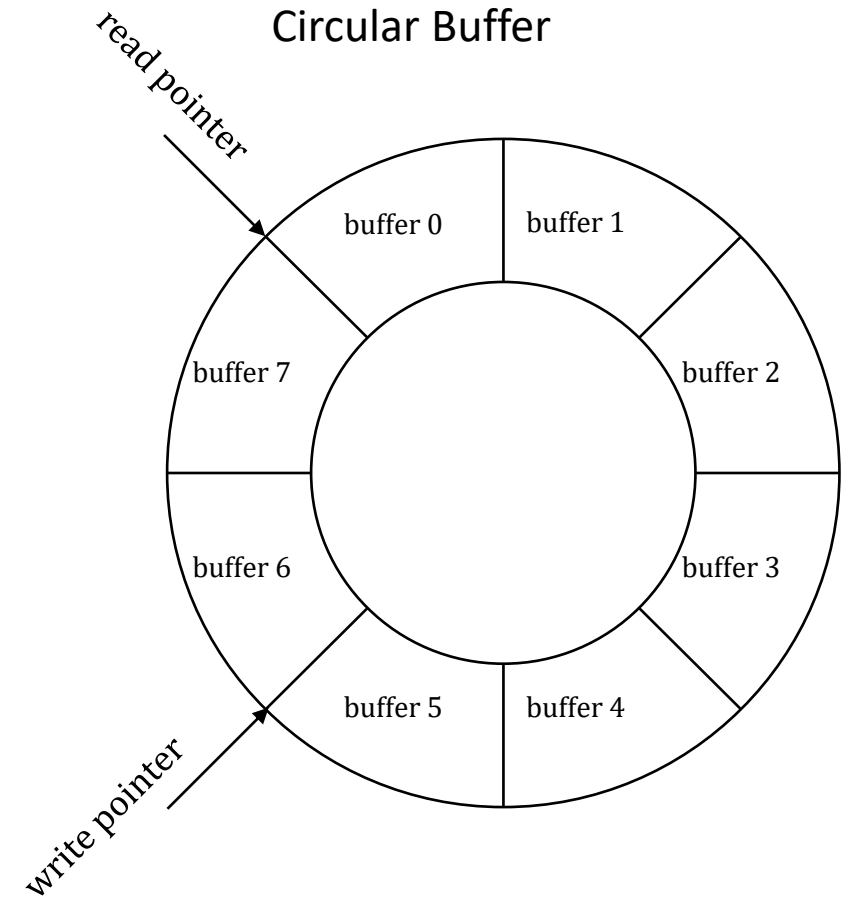
- Top-level extension to glTF 2.0
- Allows referencing all types of media
 - Timed and non-timed
 - Compressed and non-compressed
 - MPEG and non-MPEG
- It supports different types of delivery
 - DASH & CMAF
 - WebRTC
 - HLS & CMAF
 - Local Storage (ISO BMFF, MP4)
- Orthogonal Functions: encryption, etc.
- This extension decouples Media Access Function from Presentation Engine in the Scene Graph

```
{
  "extensions": {
    "MPEG_media": {
      media: [
        {
          "name": "source 0",
          "renderingRate": 25.0,
          "timeOffset": 0.0,
          "autoplay": "true",
          "loop": "true",
          "alternatives": [
            {
              "mimeType": "application/dash+xml",
              "uri": "https://www.foo.com/manifest.mpd",
              "tracks": [
                {
                  "track": "#track=1"
                }
              ]
            }
          ]
        }
      ]
    }
  }
}
```

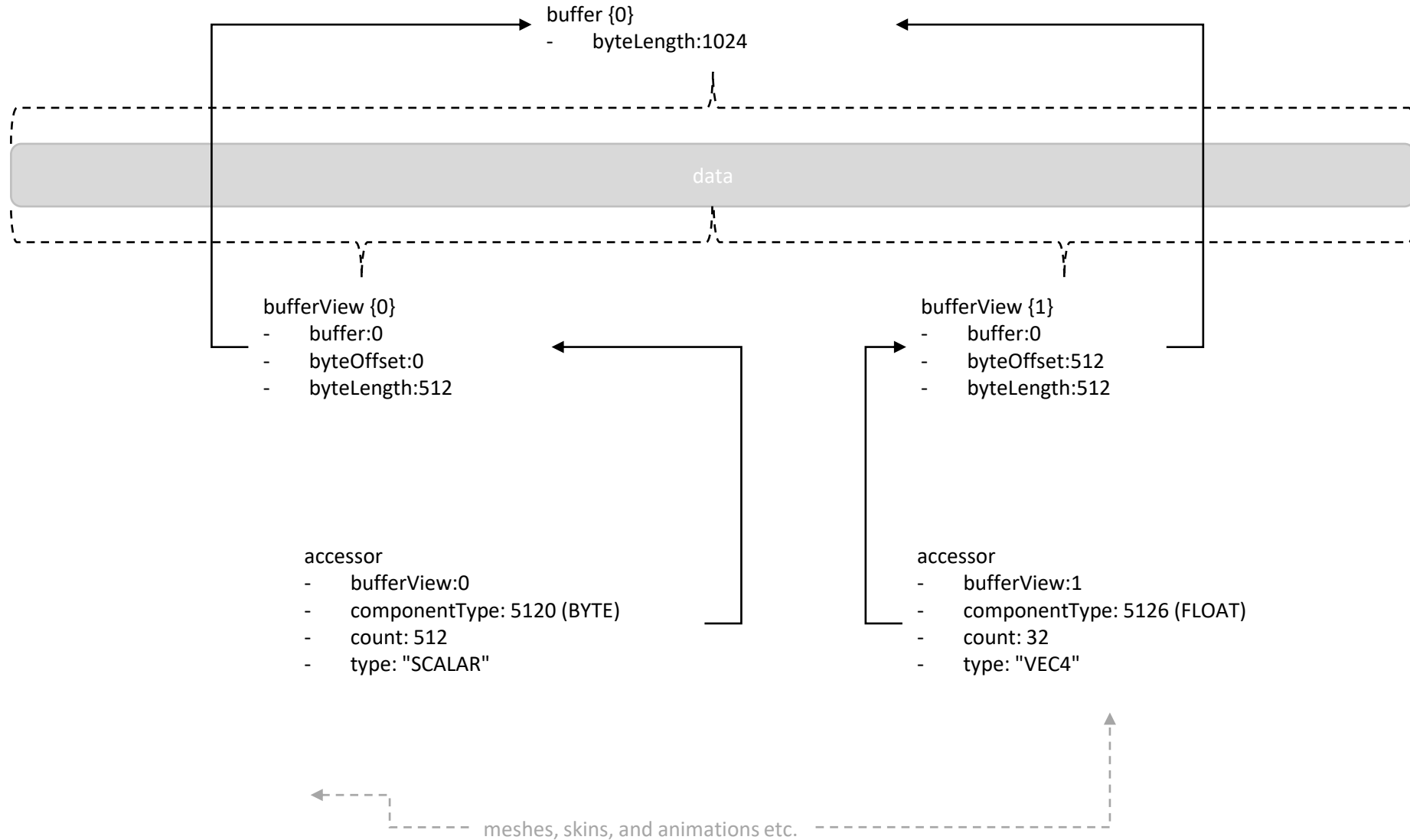
Describing how to access data

MPEG_accessor_timed and MPEG_buffer_circular extensions

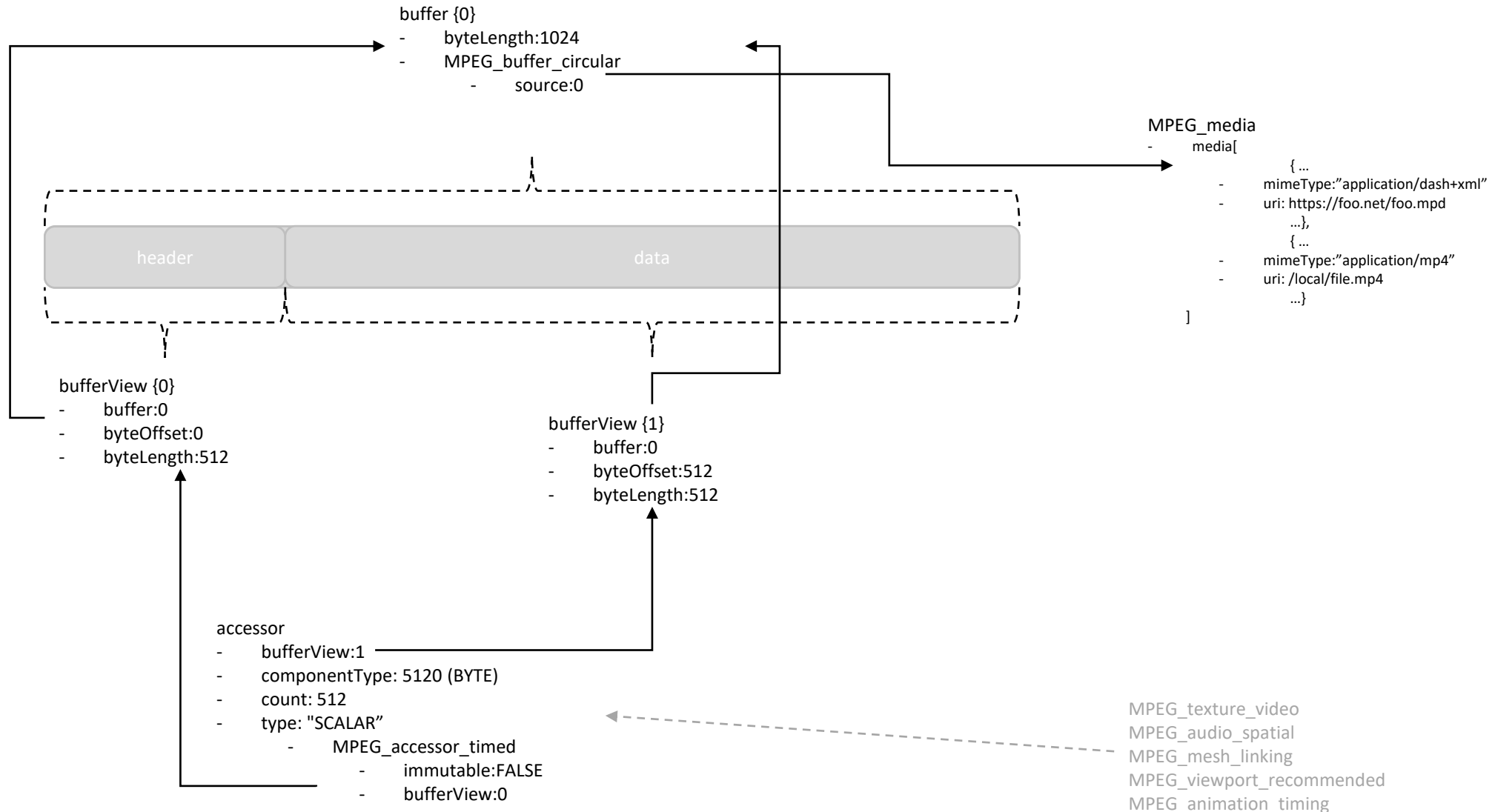
- glTF accesses data through accessors
 - They define the components of the data and their data types (e.g. a VEC3 of floats)
 - Semantics are provided by the referencing attribute/property (e.g. position)
 - The **accessor** points into a **bufferView**, which defines how the data is packed in the referenced **buffer**
 - No support for timed data
- MPEG_accessor_timed
 - Extension to accessor
 - Used to access all types of dynamic and timed media (audio, visual, volumetric, ...)
 - Backwards compatible: in case of no support, fallback to static data
- MPEG_buffer_circular
 - Extension to buffer
 - Dynamic variable-size swap chain buffer for exchange of media data for rendering
 - Acts as the interface between the Presentation Engine and MAF. All requested data through MAF API is delivered through a Buffer or Circular Buffer
 - Header is used to propagate metadata such as timestamps
 - Circular Buffer references MPEG_media



Static Buffer Usage



Dynamic Buffer Usage



Timed accessor header information

- Mutable Information in Buffer View and Accessor
- Accessor information that may change over time
 - componentType
 - bufferView
 - type
 - normalized
 - byteOffset
 - count
 - max
 - min
- bufferView information that may change over time
 - bufferViewByteOffset
 - bufferViewByteLength
 - bufferViewByteStride

Video Textures

MPEG_texture_video extension

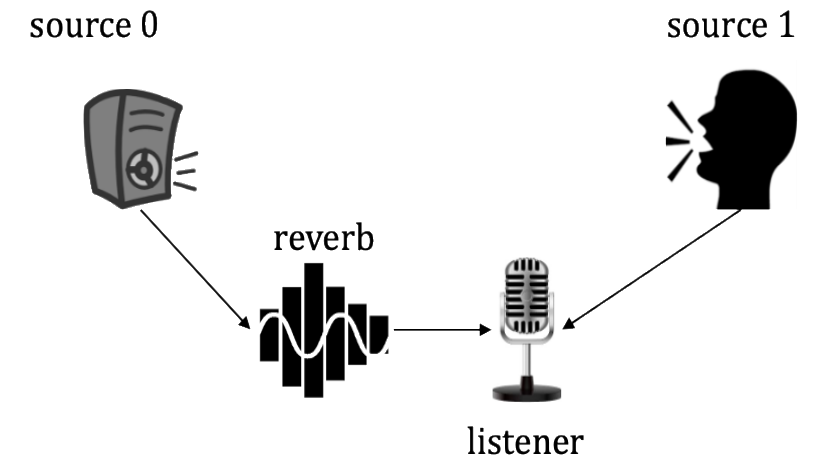
- Materials in the scene may make use of textures
- A texture in glTF 2.0 only supports references to images of format JPEG or PNG
- MPEG_texture_video adds support for dynamic textures such as atlases
 - Point into a timed accessor instead of an image
 - Keep the image pointer as fallback
 - To support dynamic atlases, texture coordinates themselves are dynamic and fed through a timed accessor as well

```
{
  "textures": [
    {
      "sampler": 0,
      "source": 1,
      "extensions": {
        "MPEG_texture_video": {
          "accessor": 2,
          "width": 2048,
          "height": 2048,
          "format": "RGB"
        }
      }
    }
  ]
}
```

Spatial Audio

MPEG_audio_spatial extension

- glTF has no support for audio
- The MPEG_audio_spatial extension:
 - Audio Sources can be coupled to visual nodes to share the same transformations
 - Supports 3 types of nodes:
 - Audio Source: emits audio signals. Simple mono and HOA sources are supported
 - Audio Effect: a reverb zone effect is currently supported
 - Audio Listener: provides the position of the listener
 - The Audio Listener may be linked to the scene camera to allow for an immersive spatial experience. The listener will move together with the camera.
 - Actual rendering is not defined.
 - It is up to the Audio Rendering Engine to convert the signals that are received at the audio listener into a format that matches the actual speaker setup.
 - For example, binauralization is done for users wearing an HMD.



Demo Unity Player

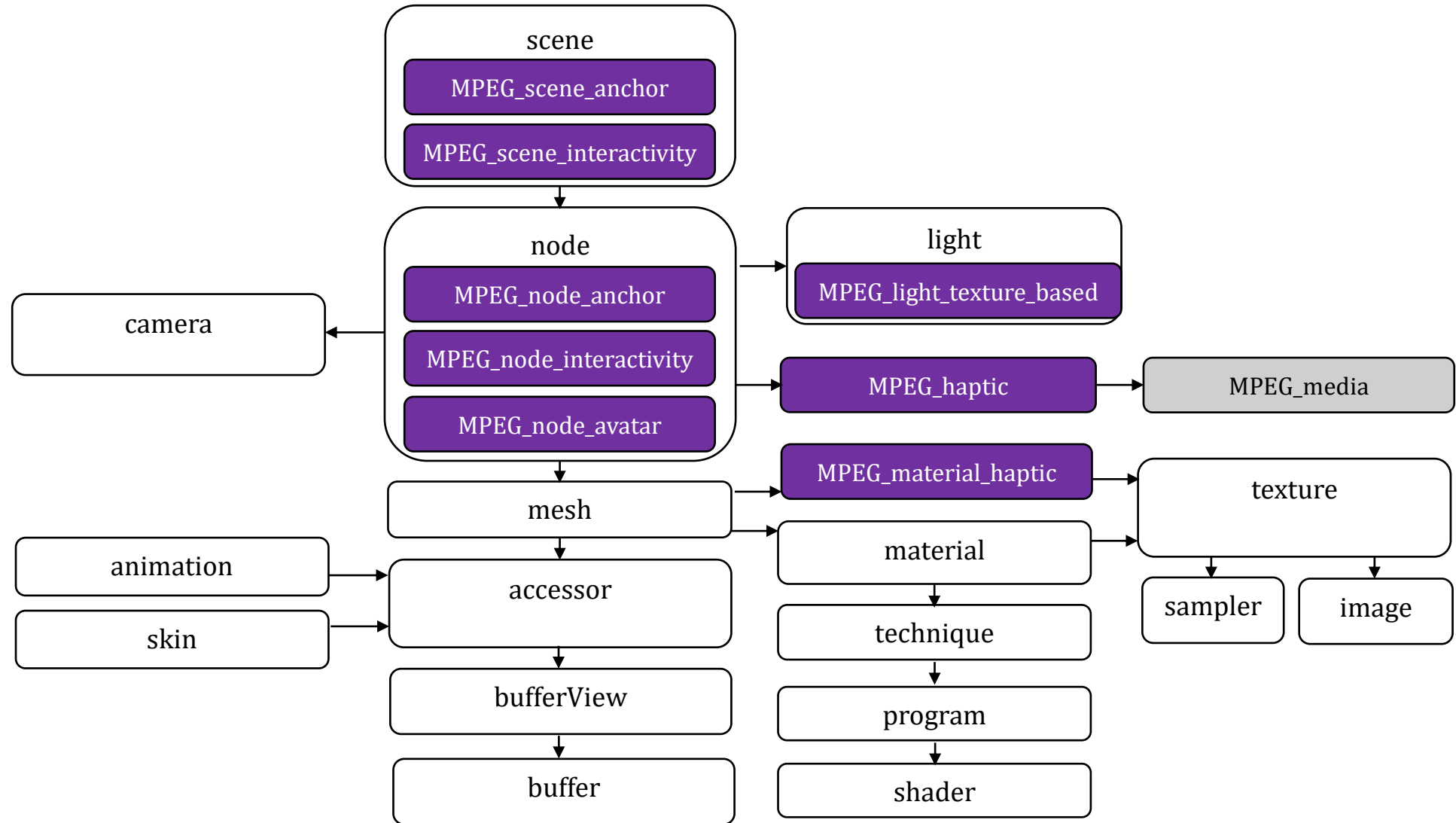
Other extensions

- MPEG_scene_dynamic
 - Provides the possibility to indicate that the scene description document will be updated
 - Updates are provided through JSON patch protocol
 - Patch sample is an atomic update operation (all patch operations part of one transaction)
 - Consistency/Validity of scene after application of a patch is the responsibility of the author
- MPEG_viewport_recommended
 - provides dynamically changing information which includes translation and rotation of the node which includes the camera object, as well as the intrinsic camera parameter of the camera object.
- MPEG_animation_timing
 - provides alignment between MPEG media timelines and animation timeline defined by glTF 2.0

MPEG-I SD

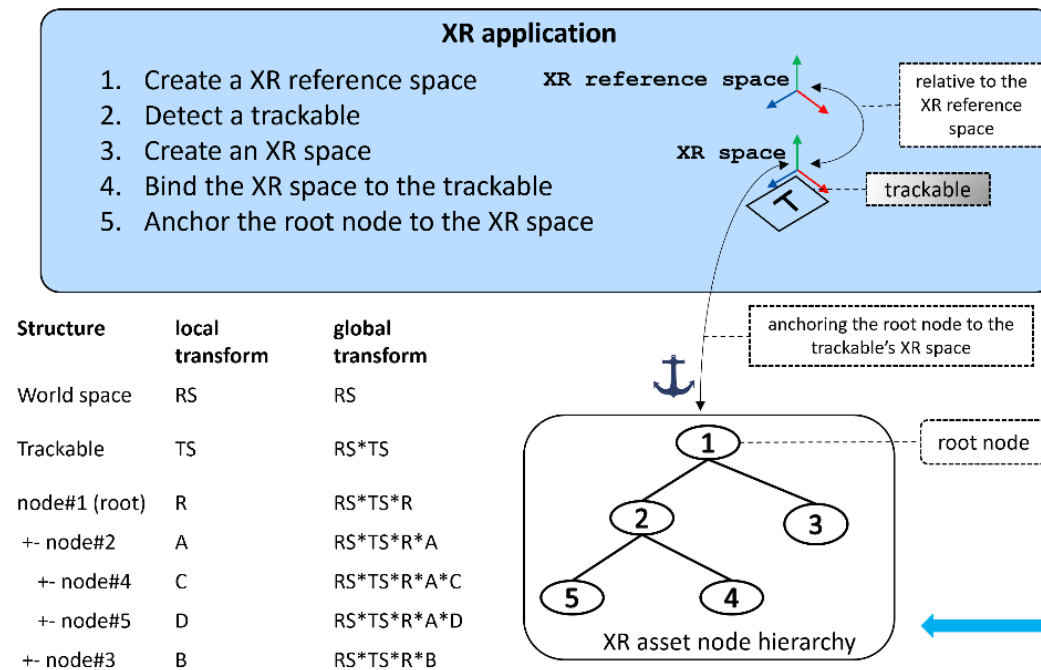
Phase 2

New Phase 2 Extensions



AR Anchoring

- Extension to allow anchoring a scene or a root node to a trackable
- Trackables can be reference trackables or application-defined trackables

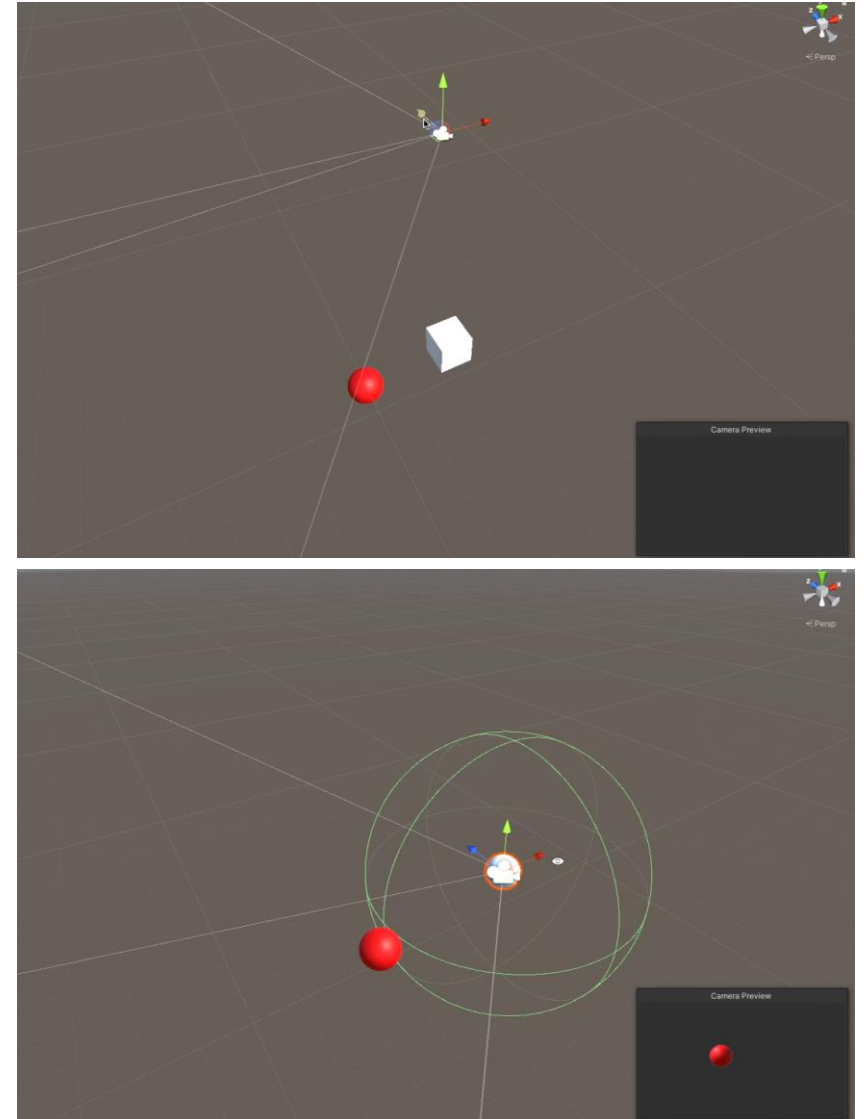


Pseudo glTF2.0

```
"scenes" : [
  {
    "nodes" : [0]
  },
  "nodes" : [ . . .
    {
      "name" : "NODE#1",
      "children" : [1,2],
      "matrix" : "#R"
    }, {
      "name" : "NODE#2",
      "children" : [3,4],
      "matrix" : "#A"
    }, {
      "name" : "NODE#3",
      "matrix" : "#B"
    }, {
      "name" : "NODE#4",
      "matrix" : "#C"
    }, {
      "name" : "NODE#5",
      "matrix" : "#D"
    }
  ]
}
```

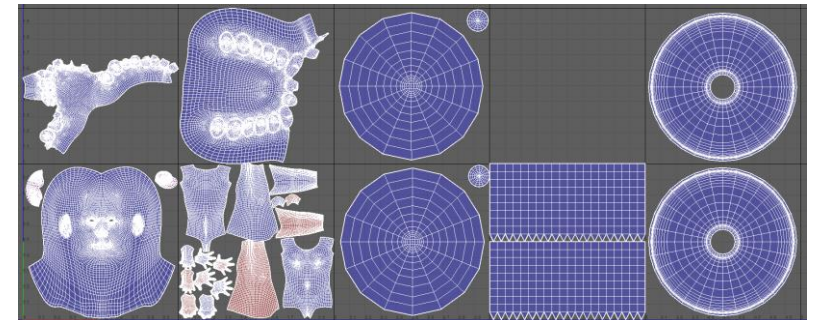
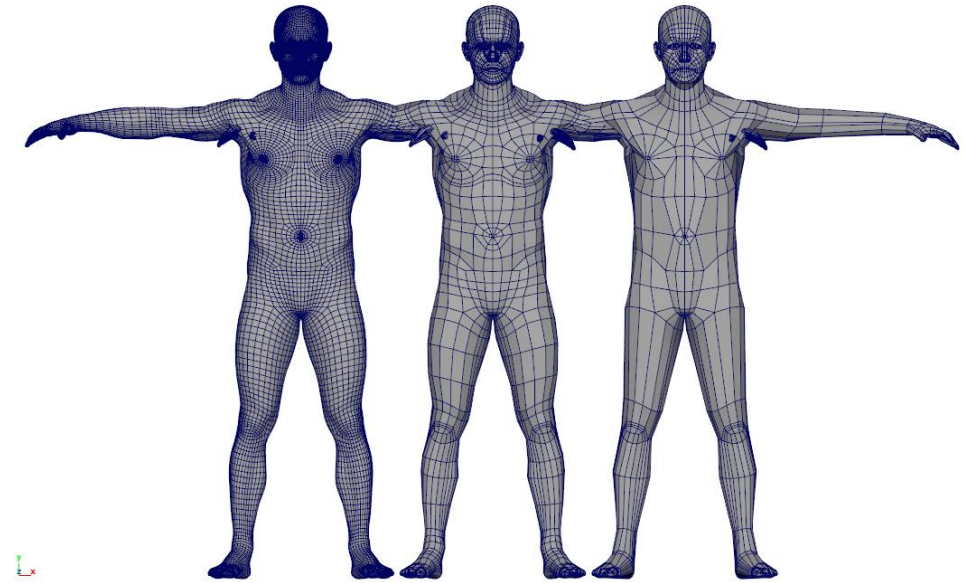
Interactivity

- Simple Trigger-Action interactivity
- Triggers: collision, proximity, user input, visibility
- Actions: activate, transform, block, animation, media, manipulate, set material, set haptic

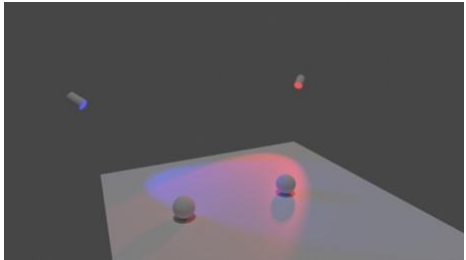


Avatars

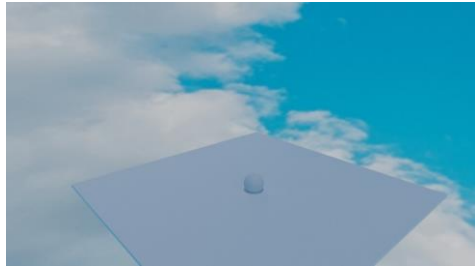
- Reference avatar with UV maps and blendshapes
- Extension to signal nodes that carry Avatars and their breakdown
- Interactivity triggers can be associated with certain segments of the Avatar, e.g. hand
- Avatars may result from real-time animation streams



Light



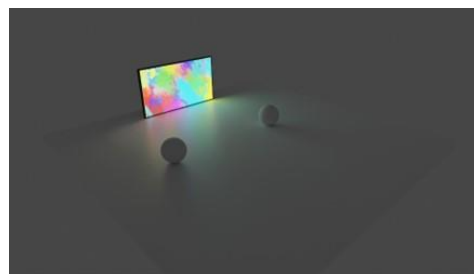
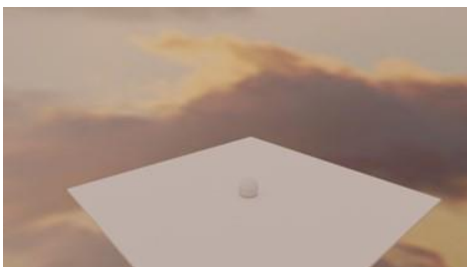
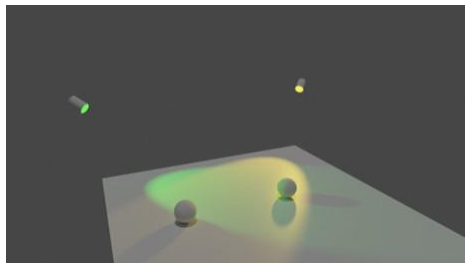
time-evolving
punctual light



time-evolving
environment light



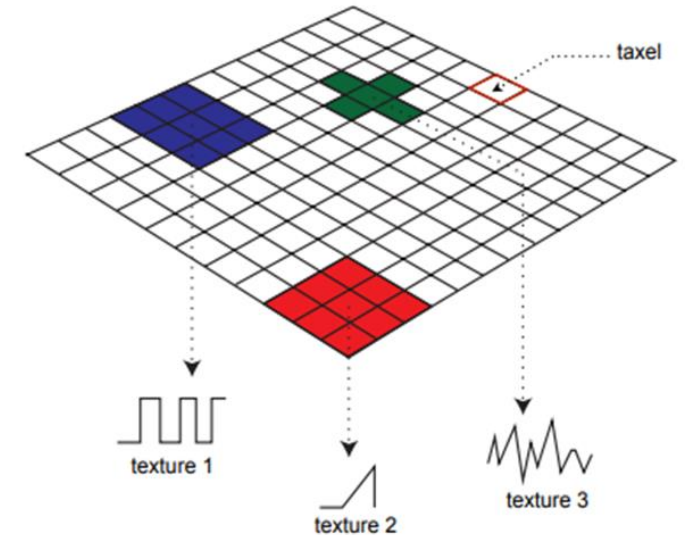
time-evolving
area light



Coherent real/virtual
lighting for AR applications

Haptics

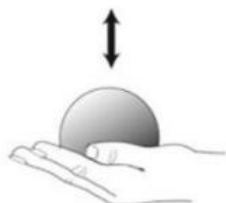
- Extension to define Haptics materials
- Integrates with Interactivity
- Supports: stiffness, friction, vibrotactile, temperature, vibration, and custom haptic maps



**Lateral Motion
(Texture)**



**Unsupported Holding
(Weight)**



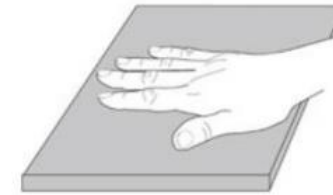
**Pressure
(Hardness)**



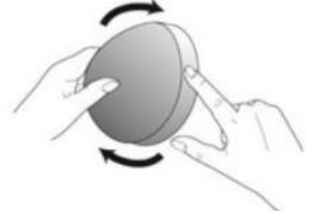
**Enclosure
(Global Shape)
(Volume)**



**Static Contact
(Temperature)**

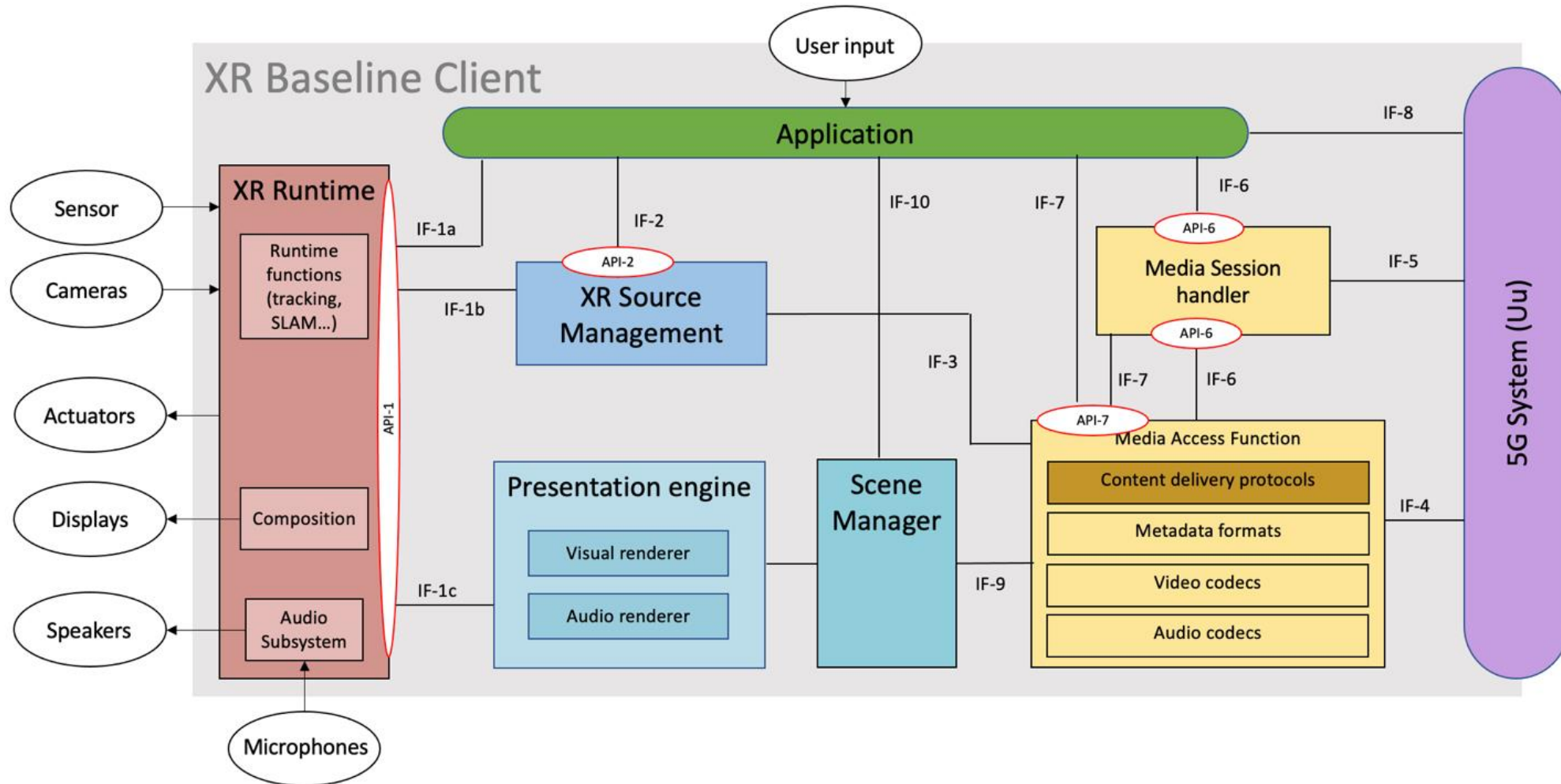


**Contour Following
(Global Shape)
(Exact Shape)**



3GPP Activities

XR Architecture

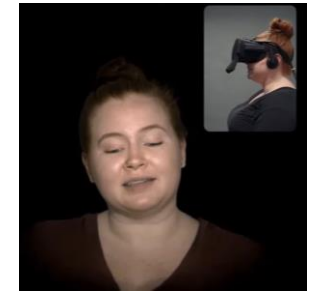


AR Call

- Extend existing IMS and WebRTC-based telephony services
- Multi-party calls and conferencing with AR
- Users join with realistic avatars and contribute 2D/3D content
- Create shared spaces where call participants can interact



High Level Procedure



Update
Loop



Scene of shared space

Scene of shared space

Add nodes to scene

Add nodes to scene

Send scene updates

Send scene updates

Pose and animation updates

Pose and animation updates

Send scene updates

Send scene updates

References

- MPEG Vendor Extensions
 - <https://github.com/KhronosGroup/glTF/tree/main/extensions/2.0/Vendor>
- MPEG-I Scene Description Whitepaper
 - https://www.mpeg.org/wp-content/uploads/mpeg_meetings/140_Mainz/w22138.zip
- Reference Player:
 - Will be available soon
 - <https://www.5g-mag.com/reference-tools>

Questions