**3GPP TSG SA WG4#127-bis-e S4-240XXX**

**online, Teams, 8 - 12 April 2024**

**Agenda item:** 9.8

**Source:** Qualcomm Incorporated, others

**Title:** [FS\_Beyond2D] Scenarios – keep it simple

**Document for Discussion and** Agreement

Note that this is an early draft and just documents on how we view the scenarios and how this can align with TR 26.955 framework.

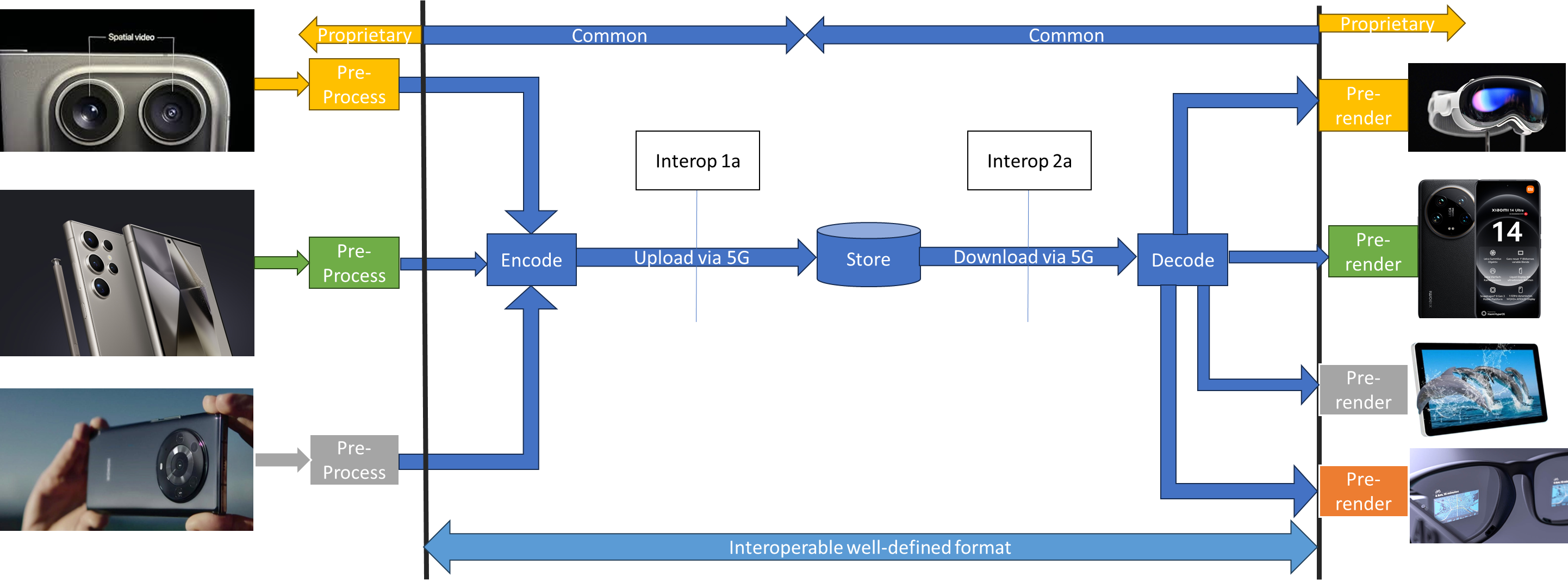
# Introduction

During SA4#127-bis-e several scenarios for the Beyond 2D study have been proposed. We observe that all of the scenarios are quite complex and attempt to address very specific deployment scenarios. We propose to simplify initial scenarios and also to address format convergence.

# Scenario 1: UE-to-UE format format exchange

In a first scenario and like the most important one, existing and emerging capturing systems on mobile devices are used to capture video formats that go beyond regular 2D video formats. The captured scene is recorded and shared with another mobile device, or it is consumed on an HMD, or it consumed on AR glass. Sharing may be done by different means (upload and download, stream, communication). The different means may create different “sub-scenarios” eventually.

The scenario is shown in the below figure. The core issue is that there may be different capturing systems that by pre-processing generate a well-defined format that can be encoded with a beyond 2D video encoder. The produced file asset is uploaded, may be stored and provdied as a message to another device. The other device decodes the format and does some proprietary ways to render the format on the device with a beyond 2D experience.

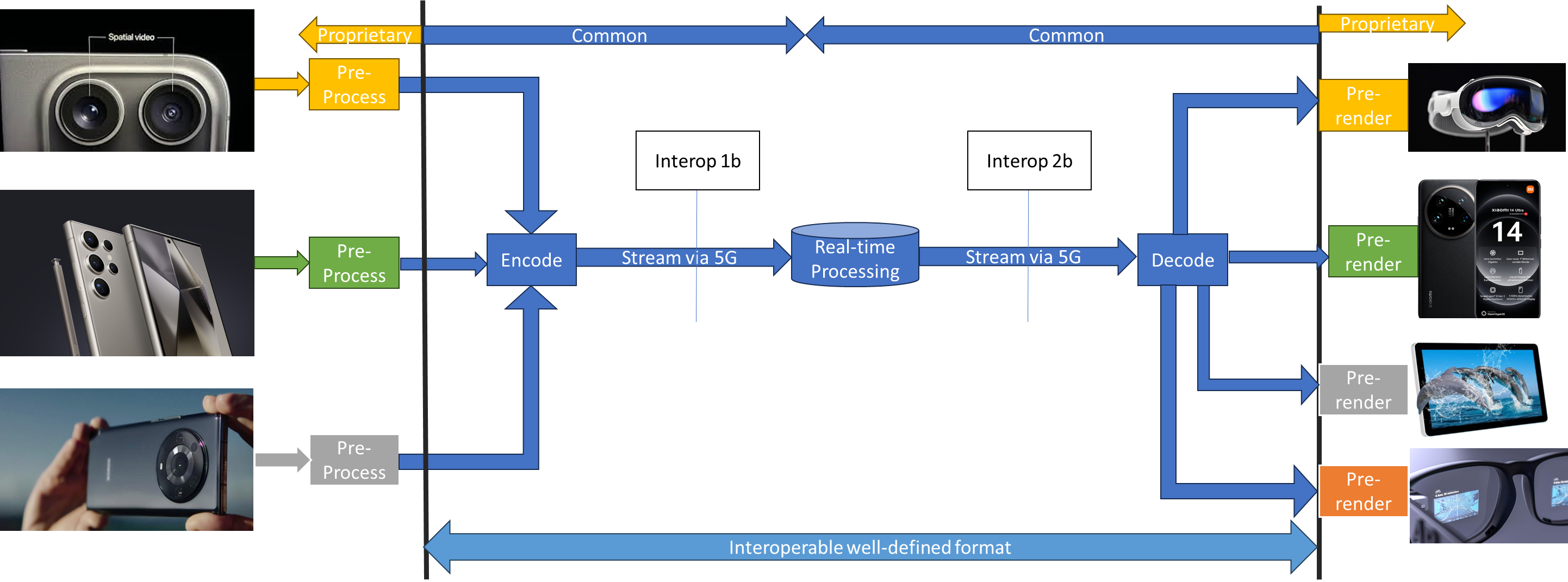


The core questions are:

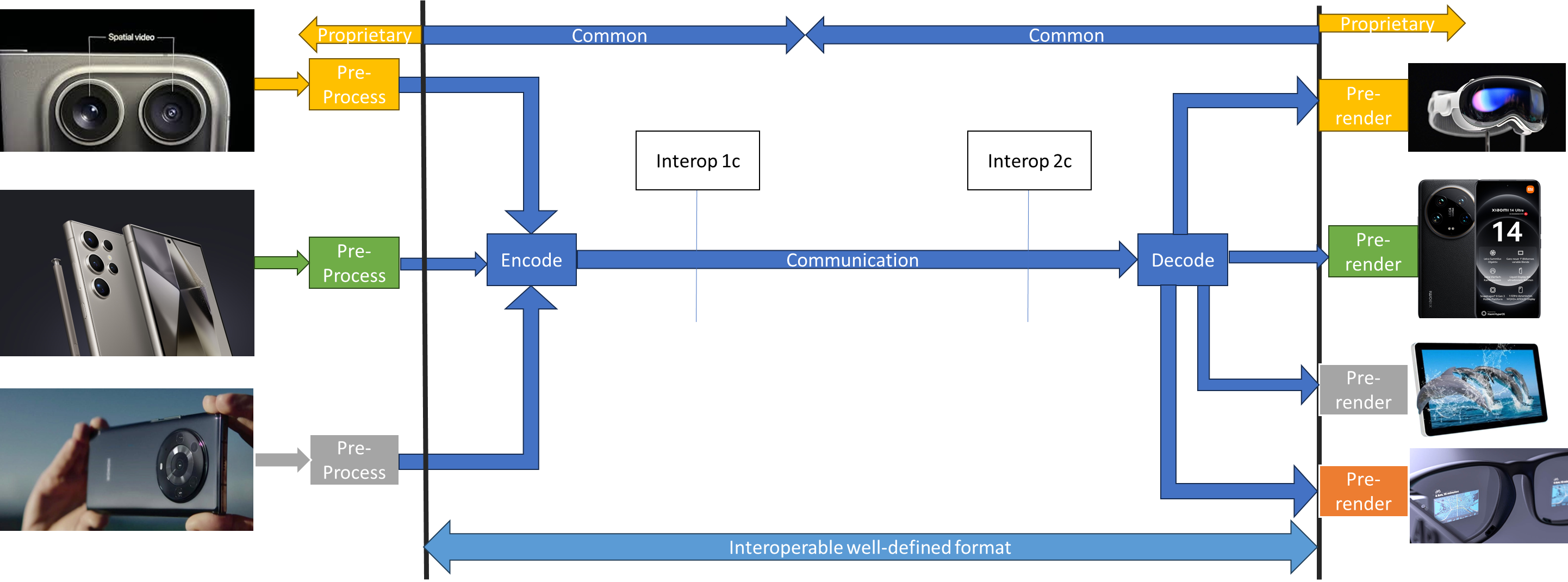
* What are common source formats (multiple views + depth?) of market relevance?
* What beyond 2D experiences can be created on market-relevant devices with such formats?
* Is there a common description of the source formats?
* How can we efficiently compress the source formats?
* What are the resulting bitrates?
* How can we package the bitstreams for delivery?
* What are requirements for each scenario in terms of network support?
* What are suitable technologies for compression? Multiple HEVC streams, MV-HEVC, MIV?

This scenario can be extended to a UE-to-UE streaming and a communication scenario. The differences are small, but in some cases for example random access points are not used. The two scenarios are shown in the next two figures.

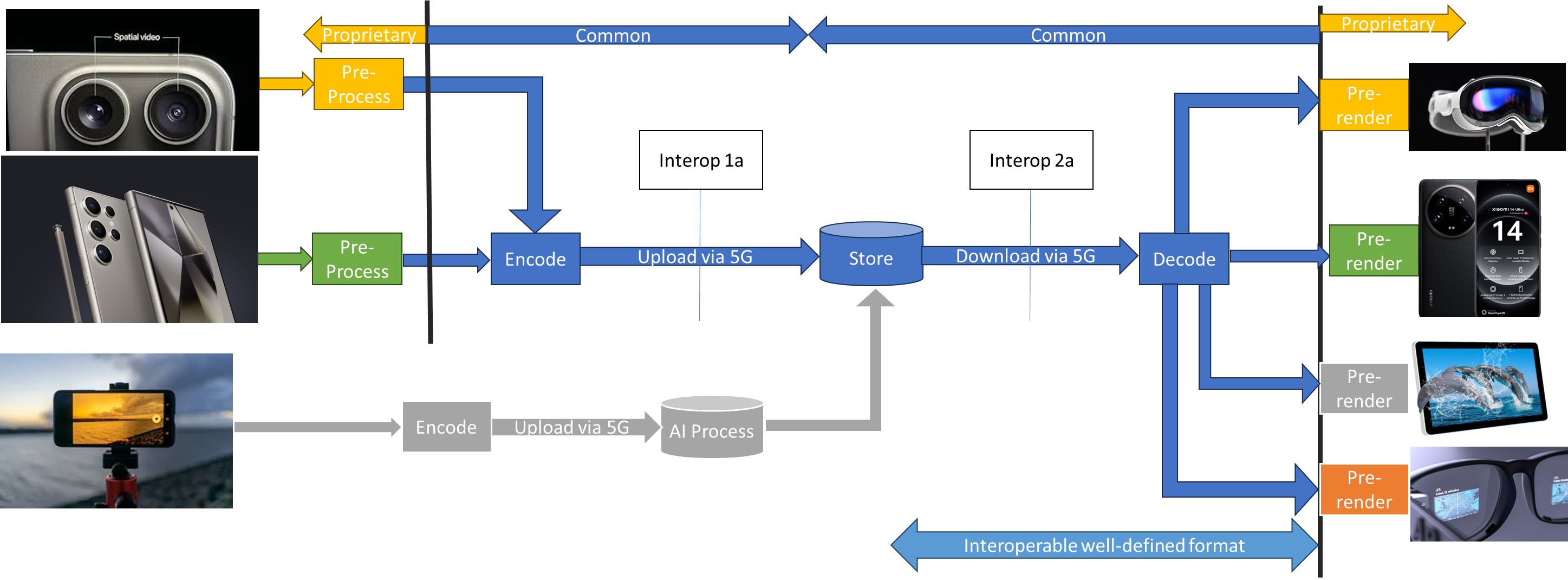
Streaming



Communication



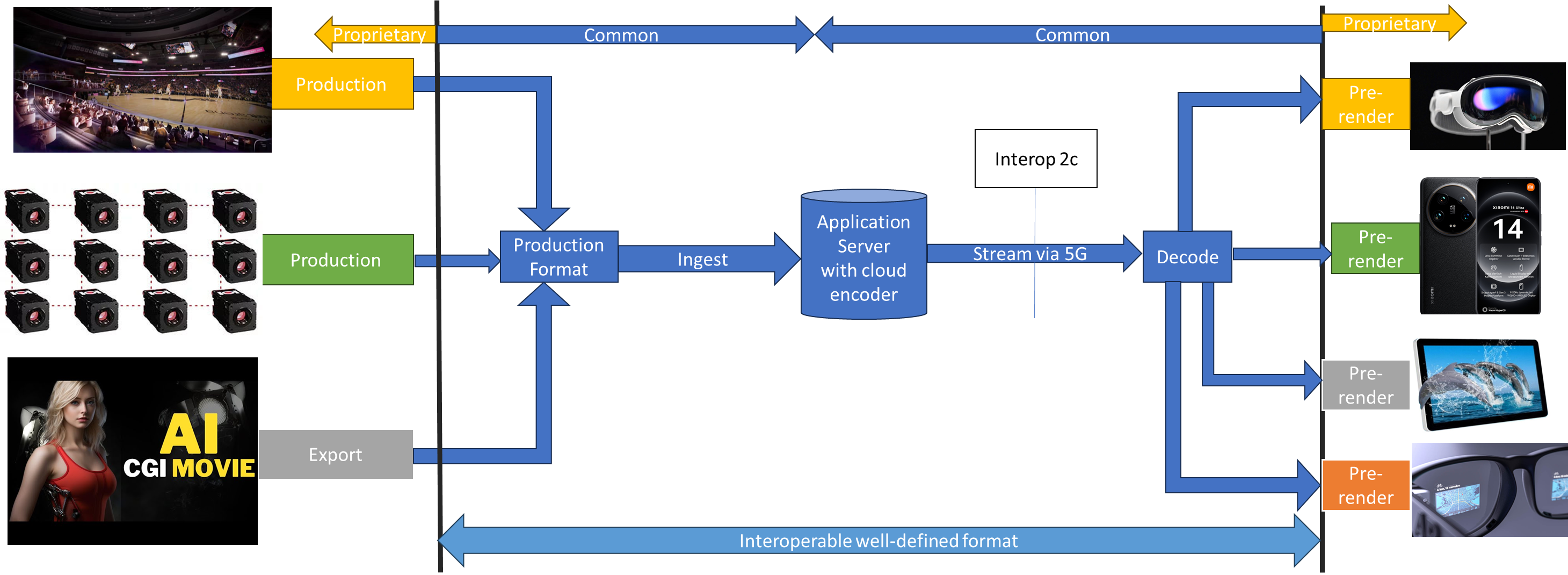
Another scenario may be considered, for which an intermediate processor addresses the conversion of 2D to beyond 2D.



At the end, the scenarios are quite similar, but they address different flavour of the encoding configurations, for example low-latency encoding, addition of RAP points etc.

# Scenario 2: Streaming of Beyond 2D Produced Content

In a second scenario, content is produced professionally, and it can be live, or captured offline or it can be AI/computer generated. The formats are produced against a common production format. The content is then ingest into e.g. a cloud encoder (for example on an Application Server) and this one produces common formats to be played back again on different devices for a beyond 2D experience.



The core questions are:

* What are common production formats (multiple views + depth?) of market relevance?
* What beyond 2D experiences can be created on market-relevant devices with such formats?
* Is there a common description of the source formats?
* How can we efficiently compress the source formats?
* What are the resulting bitrates?
* How can we package the bitstreams for delivery?
* What are requirements for each scenario in terms of network support?
* What are suitable technologies for compression? E.g. multiple HEVC streams, MV-HEVC, MIV?

# Scenario 3: Beyond 2D Content in Split Rendering

In yet another scenario, split rendering servers may use and produce formats beyond 2D to supported better performing split rendering. This is shown in the figure below.

### 

The core questions are:

* What are common split rendering formats of market relevance?
* What beyond 2D experiences can be created on market-relevant devices with such formats?
* Is there a common description of the source formats?
* How can we efficiently compress the source formats?
* What are the resulting bitrates?
* How can we package the bitstreams for delivery?
* What are requirements for each scenario in terms of network support?
* What are suitable technologies for compression? Multiple HEVC streams, MV-HEVC, MIV?

# Devices and Formats

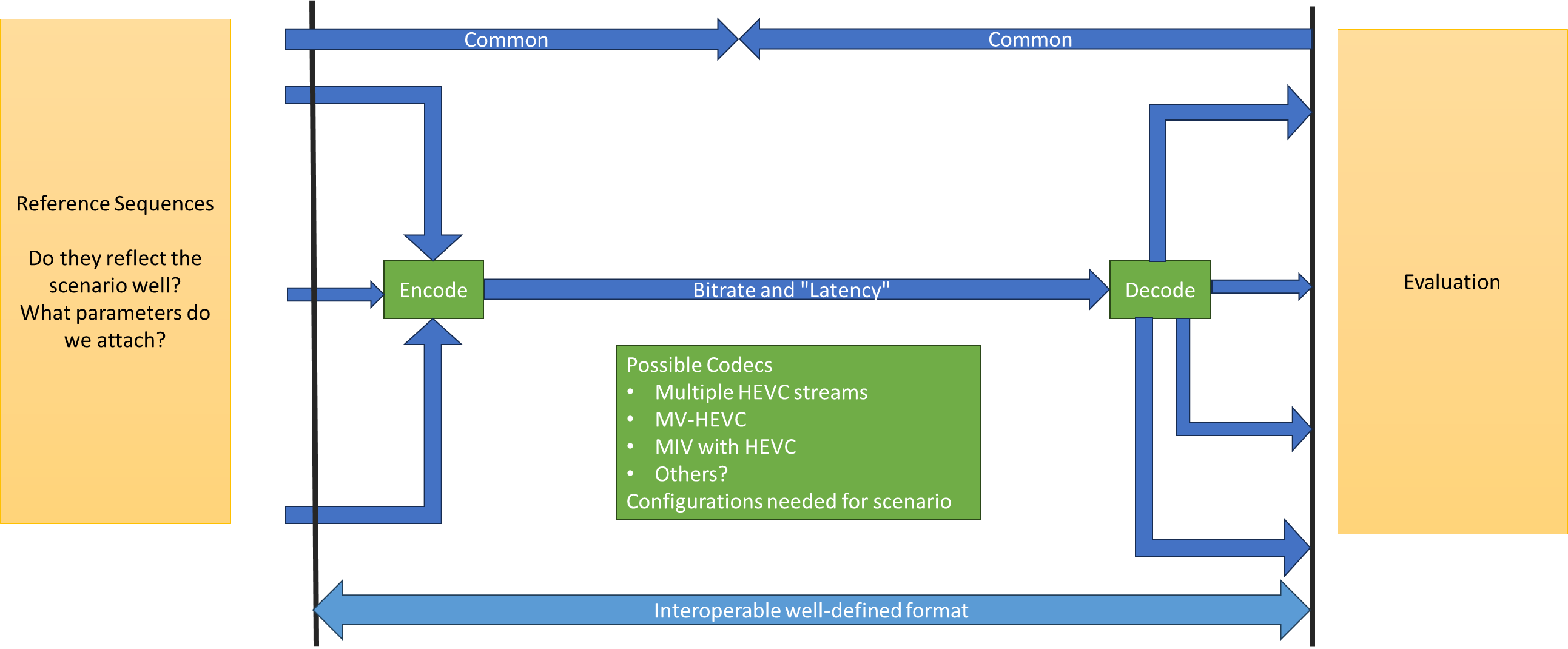
While it is understood that there is currently no harmonized set of formats for production and or device playback, the attempt towards improved format compatiblity and is the core driver for global standards. 3GPP should definitely contribute to this as we already do for the VOPS case with stereo MV-HEVC.

Note that devices are expected to render beyond 2D formats based on their implementation, i.e. there is no need for interoperability or requirements beyond the decoding and the provisioning of metadata. The issue of rendering and presenting should not be part of the discussion. However, it can be checked how existing devices can render such formats in a beyond 2D experiences and invite interested companies to do so and bring data.

# From Scenario to Evalution

However, the evalution should be simplified.

For this purpose, for each of the above scenarios, the input formats should be well defined and reference sequences should be provided – the reference sequences provide real-world example on what is practical today. The evaluation can be done by objective measures using the output of the decoder and comparing with the reference sequences.



Once this framework is in place, we can then define which codecs may be suitable and which configurations for the codecs are to be used for fair evaluation.

Metrics need to be defined, preferably objectively measurable.

It should not be prevented that someone uses the formats and does an evaluaion on actual devices, but this should not be the core subject of the study.

If we apply the principles, we can re-use the TR 26.955 characterization framework.

# Proposal

It is proposed that

* We simplify the scenarios
* We focus initially on scenario 1
* Focus on market-relevant formats for scenario 1 that create end-to beyond 2D opportunities
  + RGBD + depth
  + Stereo + depth
  + More than 2 views?
* Identify candidate technologies to compress the formats
* Define a common evaluation framework align with TR 26.955 characterization frameworks

Note that this is not proposing these scenarios as is right now, but provides an idea on how to address and develop the scenarios.