**3GPP TSG SA WG4 Meeting 131-bis-eS4-250472r02**

**online, , 11th Apr 2025 – 17th Apr 2025 revision of S4aV250023**

**Source: Qualcomm Germany, Tencent**

**Title: Pseudo-CR on [VOPS] Discussion on Terminology**

**Spec: 3GPP TS26.265v1.0.0**

**Agenda item: 9.5**

**Document for: Decision**

**1. Introduction and Discussion**

At the last meeting, the discussion around terminology came up. This document summarizes definition and proposes updates to address consistent definitions.

In H.265, the following terms are defined:

- **bitstream**: A sequence of bits, in the form of a *NAL unit stream* or a *byte stream*, that forms the representation of *coded pictures* and associated data forming one or more *coded video sequences (CVSs)*.

- **byte stream**: A sequence of bytes forming an encapsulation of a NAL unit stream into a format containing start code prefixes and NAL units as specified in Annex B.

- **network abstraction layer (NAL) unit stream**: A sequence of NAL units.

- **coded video sequence (CVS)**: A sequence of access units that consists, in decoding order, of an IRAP access unit with NoRaslOutputFlag equal to 1, followed by zero or more access units that are not IRAP access units with NoRaslOutputFlag equal to 1, including all subsequent access units up to but not including any subsequent access unit that is an IRAP access unit with NoRaslOutputFlag equal to 1.

- **access unit**: A set of NAL units that are associated with each other according to a specified classification rule, are consecutive in decoding order, and contain exactly one coded picture with nuh\_layer\_id equal to 0.

- **random access**: The act of starting the decoding process for a bitstream at a point other than the beginning of the stream.

- **decoding process**: The process specified in this Specification that reads a bitstream and derives decoded pictures from it.

In TR 26.947, which was the baseline for defining TS 26.116, the following was considered in clause 9, conclusions

*This report provides in clause 4 the considered scenario that is of relevance for the TV video profile. The key issues are summarized:*

*- The content provider would like to provide content from a multitude of sources to a multitude of device classes.*

*- It is relevant to define a reduced and constrained amount of content formats that can target a wide variety of device classes, but the device classes are able to decode and render the content formats.*

*Figure 7 provides an overview of the expected specification work for TV Video Profile. Operation Points are defined by a collection of tools that may be used by the one generating the service offering to generate a "bitstream". Note that the term "bitstream" is used, despite the data may be delivered not in a bitstream mode, it may be packetized in segments or in packets. However, the intention is to initially address the use of media coding tools and generating media bitstreams. The system level signaling is a derived aspect and should not influence the definition of the operation points.*

**

*Figure 7: Overview of the expected specification work for TV Video Profile*

*The specification of a TV Profile operation should include the following aspects:*

*- Definition of an Operation Point: A collection of discrete combinations of different content formats including spatial and temporal resolutions, color mapping, transfer functions, etc. and the encoding format.*

*- The bitstream: If a bitstream conforms to a certain operation point, it includes features that are included in the operation point. There should be a signalling on delivery system level that indicates to what operation point the bitstream is conforming.*

*- The receiver: If a receiver conforms to a certain operation point, it includes all necessary tools to decode and render a bitstream that conforms to an operation point.*

*- The bitstream and client requirements handling includes usage VUI messages.*

*- The specification should document the requirements for the bitstream and the receiver.*

*- A bitstream may conform to one or multiple operation points.*

*- A receiver may conform to one or multiple operation points.*

*- The specification needs to be extensible to new operation points for future releases and to other media types.*

*- The number of operation points should be small and be justified by different deployment cases and/or device classes.*

*- Operation Points should have catchy names.*

*- The specification should be written such that operation points can be referred from outside.*

*- Operation Points should provide an overview on compatibility to other operation points.*

*- Operation Points should include the relevant aspects identified in the present document.*

*- An operation point may be delivered by a delivery system, for example DASH. For each operation point, the delivery specific aspects for the sender and the receiver should be defined. Specific aspects such as DASH signalling should be included.*

In TS 26.116, the following is defined:

**- Bitstream:** A media bitstream that conforms to a video encoding format and certain Operation Point.

**- Operation Point:** A collection of discrete combinations of different content formats including spatial and temporal resolutions, colour mapping, transfer functions, etc. and the encoding format.

**- Receiver:** A receiver that can decode and render any bitstream that is conforming to a certain Operation Point.

Now there were some discussions on e-mail related to the Frame Packing Arrangement SEI message and the necessity to include a message in all cases.  This was considered by one colleague: *If you want to indicate that a frame is 2D the only way to do so is to not associate a frame with such an SEI message. Since we want to do this per CVS, requiring that this SEI message is present in all CVSs makes no sense since you will not have the capability to indicate a 2D signal. It is quite common for content to have a mixture of 3D and 2D segments, and thus this requirement would negatively impact distribution.  Our text tries to handle this gracefully by talking about CVSs and not about bitstreams (and we feel that we need to define the term of a CVS/Coded video sequence) as is done in video coding specs. We feel that the only requirement should be that if the frame packing SEI message is present in a CVS then that should apply to all frames (either by that persisting or by that being repeated for all frames) and should not change within a CVS. That is common practice and that was also what was done in the past for Broadcast 3D TV. We are not defining anything different from what has been used in the past.*

Based on this, the following was considered that in SA4, the primary objectives in the video spec is the definition of "bitstreams" that are assigned to one coded video sequence. In TS 26.116 this is called bitstream and this may cause confusion. A few options were discussed.

- We define a Bitstream as a ”coded video sequence that includes a single representation format”

- We change the terminology also to coded video sequence here

- We invent a new term here

A proposal was made as follows:

*It is not essential to align terminology to every possible specification. It would be better to have a concrete terminology on 3GPP’s end and we can even include a statement in the definition of the terms that says that different video specifications may use different terminology to define similar terms. And then keep silent about it. It would be too much work and too confusing if we switched to different terminology, assuming of course we supported different coding specifications. We primarily should care about implementers of video specifications. It was considered to have both the terminology of a bitstream and the coded video sequence and it should be fine to align it with MPEG specifications such as HEVC. The terminology there is quite clear and easy to follow. A recommended definition by taking what is specified in HEVC but by removing the mention of NAL units is provided in the following.*

*- Bitstream: A sequence of bits that forms the representation of any coded pictures and associated data. This sequence of bits is formed by one or more coded video sequences (CVSs).*

*And then for the Coded Video Sequence*

*- Coded video sequence (CVS): A sequence of bits that consists of a series* ***of coded frames*** *and any associated metadata and conforms to a specific video encoding format and aligns with a certain Operation Point, as defined in this document. Such coded video sequence has no decoding dependency from any other prior coded video sequence and consists, in decoding order, of information specifying the characteristics or format of the encoded video data, a single clear intra random access coded frame followed by zero or more dependent, to the intra random access coded frame, coded frames, and a series of associated coded metadata.*

*What may be missing:*

*- a clear definition of a random-access coded frame.*

*We may also add that video specifications may include different definitions for a bitstream or a coded video sequence. However, for the purposes of this document is the functional definition that can then be used to different video codecs as well as to different delivery systems.*

*Afterwards, we can replace most/all instances of bitstream in subsequent sections with a CVS.*

The idea is kind of depicted below



Now the metadata is what is typically needed for identifying whether the decoder/receiver is capable of decoding the CVS, and also provides the decoder configuration information to initialize decoding and rendering.

Note also that a Bitstream in the context of 3GPP would be a sequence of coded video sequences with the same metadata and configuration, i.e. it would allow random access, but would not change the parameters.

**2. Reason for Change**

Based on all of these discussions it is necessary to update the definitions. This is a starting point to kick off the discussions.

**3. Conclusions**

Let’s discuss and have proper resolution by the April meeting.

**4. Proposal**

It is proposed to agree the following changes to 3GPP TS26.265v1.0.0 at SA4-131-bis-e.

**5. Revision**

**Prior to meeting**

|  |  |  |  |
| --- | --- | --- | --- |
| [**S4aV250023**](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/3GPP_SA4_AHOC_MTGs/SA4_VIDEO/Docs/S4aV250023.zip) | [VOPS] Discussion on Terminology | Qualcomm Germany | Thomas Stockhammer |

**Online Discussion**: (March 18, 2025)

* No comments.

**Decision**: Agreed as basis for further work (not implemented into the draft TS).

[S4aV250023](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/3GPP_SA4_AHOC_MTGs/SA4_VIDEO/Docs/S4aV250023.zip) is **agreed as basis for further work**.

**During Meeting**

[Waqar Zia on Mon, 14 Apr 2025 10:14:34 +0200](https://list.etsi.org/scripts/wa.exe?A2=3GPP_TSG_SA_WG4_VIDEO;cc24d7ba.2504b)

Thanks for the paper and indeed the overall status of definitions has progressed quite a bit. I have some questions though on the figures, specifically the data mode in Figure 4.2-2. For example the relation of block “coded metadata” to the encoder is not clear in this figure and also not to the previous picture where this was output by the encoder.

*I consider that the output of the encoded metadata is provided by the encoder, like decoder config and so on. You add this prior to decoding to the decoding process. Tried to clarify.*

Also, some further clarification on “configuration" can be added. Overall I think these figures are very useful work for understanding, but can we consider e.g. moving Figure 4.2-2 in a more informative Annex?

*Maybe we need some offline drafting during or after meeting. For now I added an editor’s note.*

[Ahmed Hamza on Mon, 14 Apr 2025 08:25:55 +0000](https://list.etsi.org/scripts/wa.exe?A2=3GPP_TSG_SA_WG4_VIDEO;89ded7bf.2504b)

I have uploaded a revision with some edits in the Drafts folder:

<https://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_131-bis-e/Inbox/Drafts/Video/S4-250472-26265-pcr-Terminology-r03-GT_IDCC.docx>

 *Used as a baseline and updated.*

I also have a question on the following definition of a “Bitstream”:

“A sequence of bits that forms the representation of any coded pictures and associated data. This sequence of bits is formed by one or more coded video sequences (CVSs) where each CVS has assigned identical metadata.”

I’m not sure I understand what “identical metadata” means in this definition. Is it identical to other CVSs? Or is this about “sequence” parameter sets, for example? If the latter, then may be “common” metadata would provide more clarity.

 *It is identical in a sense that all CSVs in a Bitstream share identical metadata. I clarified this below.*

\* \* \* First Change \* \* \* \*

# 3 Definitions of terms, symbols and abbreviations

## 3.1 Terms

For the purposes of the present document, the terms given in TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in TR 21.905 [1].

**Bitstream:** A sequence of bits that forms the representation of any coded pictures and associated data. This sequence of bits is formed by one or more coded video sequences (CVSs) where the CVS share identical metadata.

**Coded Video Sequence:** A sequence of bits that consists of a series of coded frames and any associated metadata (required for decoder and rendering initialization) and conforms to a specific video encoding format and aligns with a certain Operation Point, as defined in this document. Such coded video sequence (CVS) has no decoding dependency on any other prior CVS and consists, in decoding order, of information specifying the characteristics or format of the encoded video data, a single intra random access coded frame followed by zero or more dependent, on the intra random access coded frame, coded frames, and a series of associated coded metadata.

**Chroma:** a sample array or single sample representing one of the two colour difference signals related to the primary colours, represented by the symbols *Cb* and *Cr*.

**Hero Eye**: The default eye in a stereo (stereoscopic) video pair, often determined by tags set by the cameras used to capture the video.

**Luma:** a sample array or single sample representing the monochrome signal related to the primary colours (denoted with the symbol *Y*),

**Operation Point:** A collection of discrete combinations of different video representation formats, including spatial and temporal resolutions, colour mapping, transfer functions, and the encoding format.

**Receiver:** A device capable of decoding and rendering any bitstream that is conforming to a certain Operation Point.

## 3.2 Symbols

For the purposes of the present document, the following symbols apply:

Symbol format (EW)

<symbol> <Explanation>

## 3.3 Abbreviations

For the purposes of the present document, the abbreviations given in TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in TR 21.905 [1].

AVC Advanced Video Coding

CENC Common ENCryption

CMAF Common Media Application Format

CVS Coded Video Sequence

DPC Device Playback Capabilities

FFS For Further Study

HDR High Dynamic Range

HDTV High-Definition TeleVision

HEVC High Efficiency Video Coding

HLG Hybrid Log-Gamma

MSE Media Source Extension

MVHEVC MultiView extensions of HEVC

RAP Random access point

SDR Standard Dynamic Range

UHD Ultra-High Definition

WCG Wide Colour Gamut

# 4 Context and Definitions

## 4.1 Motivation

Video codecs, encoders, and decoders are core components of 3GPP services. At the same time, video encoders and decoders, residing on 3GPP User Equipment (UE) and defined in 3GPP specifications, also provide interoperability points for third-party services. Video capabilities are predominantly independent of the service in use. This specification addresses the definition of video capabilities and operating points such that 3GPP service specifications as well as third-party service providers can refer to the interoperability points defined in this specification.

The present specification makes use some of the concepts recommended in TR 26.857 [2], i.e. the concept of Media Service Enablers.

## 4.2 Reference architectures and definitions

In order to define the normative aspects of this specification, reference architectures are defined. The core architecture is provided in Figure 4.2-1. The workflow addresses the generation of a *video bitstream* from a *video signal* using a *video encoder* as well as the decoding of a video bitstream by a *video decoder* and providing the resulting decoded video as well as associated metadata to a rendering and display process. The video signal can be composed of one or more video signal components, for example a video signal can include multiple views. Video signals follow certain representation formats and can be rendered in a device specific manner.

The video encoder as well as the video decoder may be configured to certain operations indicated by APIs in Figure 4.2-1. These APIs are not normatively specified but serve as an example reference to configure encoders and decoders as documented in Annex [A].



Figure 4.2-1 Reference architecture for video operating points and capabilities

Video encoders produce a sequence of *Coded Video Sequences,* as defined in clause 3.1, and the sequence of CVSs are referred to as *Bitstreams*.

An intra random access coded frame, together with the associated metadata, forms a Random Access Point (RAP) that permits to initialize decoding of the coded video sequence.

The decoder is provided with access units which correspond to pieces of the Bitstream that can be processed by the decoder to regenerate decoded video frames.

Figure 4.2-2 provides an overview of the data model and the definitions in this specification.



Editor’s Note: This figure is for illustrative purposes, informative and may be moved to an Annex.

Figure 4.2-2 Informative Data model for illustration purposes

In this case, configuration information is coded into metadata, that can be provided to the decoder to initialize the decoding of the CSVs included in the Bitstream.

A more system-centric architecture is provided in Figure 4.2-3. The workflow addresses the generation of a *transport stream* from a video signal using a *video encoder* and a *packager*. The package may include for example timing and metadata information. The de-packaging and decoding of the *transport stream* by a de-packager and a *video decoder*, respectively, allows for providing the resulting video signal as well as associated metadata to a rendering and display process. Again, the packager/encoder as well as the de-packager/decoder may be configured to certain operations indicated by APIs in Figure 4.2-2.



Figure 4.2-3 Reference architecture for system operating points and capabilities

Based on this introduction, the following terms are defined:

**- Operating Point:** A collection of different possible video formats including spatial and temporal resolutions, colour mapping, transfer functions, etc. and a video encoding format.

**- Bitstream**: A compressed media representation presented as a sequence of bits

- that forms the representation of any coded pictures and associated metadata data,

- this sequence of bits is formed by one or more CVSs and each CVS has identical metadata

- the sequence of bits conforms to a particular video coding specification/format and one or more Operating Points.

- comprised by access units that serve as units to be provided to decoders for regenerating frames.

**- Receiver**: A device that can ingest and decode any Bitstream that is conforming to a particular video coding specification and Operating Point, and optionally render it.

In addition, on system level the following terms are defined:

**- System Operating Point:** A collection of different possible video formats including spatial and temporal resolutions, colour mapping, transfer functions, etc., a video encoding and a packaging format.

**- Transport Stream:** A packaged media bitstream that conforms to a particular video coding and packaging specification/format and one or more Operating Points.

**- System Receiver:** A receiver that can de-package and decode any system bitstream that is conforming to a particular System Operating Point, and optionally render it.

NOTE: A reference architecture for multiple decoders is for further study.

System Operating Points are not defined in this specification but are left for mappings to specific delivery protocols such as RTP for MTSI, CMAF/DASH for 5G Media Streaming, or ISO BMFF for Messaging Services. However, this specification provides mapping principles to delivery protocols.

\* \* \* End of Changes \* \* \* \*