|  |  |
| --- | --- |
| 3GPP TS 26.265 V1.1.0 (2025-04) | |
| Technical Specification | |
| 3rd Generation Partnership Project;  Technical Specification Group Services and System Aspects;  Media Delivery: Video Capabilities and Operation Points (Release 19) | |
|  | |
|  |  |
|  | |
| The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP. The present document has not been subject to any approval process by the 3GPPOrganizational Partners and shall not be implemented. This Specification is provided for future development work within 3GPPonly. The Organizational Partners accept no liability for any use of this Specification. Specifications and Reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices. | |

|  |
| --- |
|  |
| ***3GPP***  Postal address  3GPP support office address  650 Route des Lucioles - Sophia Antipolis  Valbonne - FRANCE  Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16  Internet  https://www.3gpp.org |
| ***Copyright Notification***  No part may be reproduced except as authorized by written permission. The copyright and the foregoing restriction extend to reproduction in all media.  © 2025, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC).  All rights reserved.  UMTS™ is a Trade Mark of ETSI registered for the benefit of its members  3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners LTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners  GSM® and the GSM logo are registered and owned by the GSM Association |

Contents

Foreword 5

Introduction 6

1 Scope 7

2 References 7

3 Definitions of terms, symbols and abbreviations 8

3.1 Terms 8

3.2 Symbols 8

3.3 Abbreviations 8

4 Context and Definitions 9

4.1 Motivation 9

4.2 Reference architectures and definitions 9

4.3 Capability Specification 11

4.4 Video representation formats 11

4.4.1 Overview 11

4.4.2 Video signal parameters 11

4.4.3 3GPP Video Formats 15

4.4.3.1 Introduction 15

4.4.3.2 High-Definition 15

4.4.3.3 High Dynamic Range 16

4.4.3.4 Stereoscopic format 17

4.5 Common Bitstream Constraints 19

4.5.1 General 19

4.5.2 AVC Bitstreams 19

4.5.3 HEVC Bitstreams 19

4.6 Reference API parameters 21

4.6.1 Introduction 21

4.6.2 Video Decoder API Parameters 21

4.6.3 Video Encoder API Parameters 21

4.6.4 Player API Parameters 21

5 Video Coding Capabilities 22

5.1 Overview 22

5.4 Single-Instance Encoding Capabilities 24

5.5 Multi-Instance Decoding Capabilities 25

5.6 Multi-Instance Encoding Capabilities 25

6 Video Operation Points 26

6.1 Introduction 26

6.2 AVC Video Operation Points 26

6.2.1 Introduction 26

6.3.2 3GPP AVC HD Operation Point 26

6.3.2.1 Introduction 26

6.3 HEVC Video Operation Points 26

6.3.1 Introduction 26

6.3.2 3GPP HEVC HD Operation Point 27

6.3.2.1 Introduction 27

6.3.2.2 Bitstream Requirements 27

6.3.2.3 Receiver Requirements 27

6.3.3 3GPP HEVC HDR Operation Point 28

6.3.3.1 Introduction 28

6.3.3.2 Bitstream Requirements 28

6.3.3.3 Receiver Requirements 28

6.3.4 3GPP HEVC UHD 29

6.3.4.1 Introduction 29

6.3.4.2 Bitstream Requirements 29

6.3.4.3 Receiver Requirements 29

6.3.5 3GPP HEVC Stereo 30

6.3.5.1 Introduction 30

6.3.5.2 Bitstream Requirements 30

6.3.5.3 Receiver Requirements 30

6.3.6 3GPP MVHEVC Stereo 31

6.3.6.1 Introduction 31

6.3.6.2 Bitstream Requirements 31

6.3.6.3 Receiver Requirements 32

7 Common System Integration 32

7.2.1 General 32

7.2.1.1 Summary 32

7.2.1.2 Codec String 33

7.2.1.3 Decoder Configuration 33

7.2.1.4 Random Access Point 33

7.2.1.5 Coded Access Unit 34

7.2.1.6 Random Access CAU 34

7.2.2 AVC 34

7.2.3 HEVC 34

Annex <A> (normative): Registration Information 34

Annex <B> (informative): Mapping of Operation Points to Implementations 35

B.1 Introduction 35

B.2 WebCodecs API 35

B.2.1 Introduction 35

B.2.2 Mapping of Operation Points to Decoder API 36

B.2.3 Mapping of Operation Points to Encoder API 36

Annex <X> (informative): Change history 36

# Foreword

This Technical Specification has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

In the present document, modal verbs have the following meanings:

**shall** indicates a mandatory requirement to do something

**shall not** indicates an interdiction (prohibition) to do something

The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

**should** indicates a recommendation to do something

**should not** indicates a recommendation not to do something

**may** indicates permission to do something

**need not** indicates permission not to do something

The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

**can** indicates that something is possible

**cannot** indicates that something is impossible

The constructions "can" and "cannot" are not substitutes for "may" and "need not".

**will** indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**will not** indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**might** indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

**might not** indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

**is** (or any other verb in the indicative mood) indicates a statement of fact

**is not** (or any other negative verb in the indicative mood) indicates a statement of fact

The constructions "is" and "is not" do not indicate requirements.

# Introduction

The present document defines service-independent video operation points and capabilities. These may be referenced in 3GPP service specifications or in third-party services.

# 1 Scope

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.

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

Editor’s Note: References need to be updated to latest versions and to include ISO/IEC for dual publications.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[bt709] Recommendation ITU-R BT.709-6 (06/2015): "Parameter values for the HDTV standards for production and international programme exchange"

[bt2100] Recommendation ITU-R BT.2100-2 (07/2018): "Image parameter values for high dynamic range television for use in production and international programme exchange"

[h264] Recommendation ITU-T H.264 (08/2021): "Advanced video coding for generic audiovisual services".

[h265] Recommendation ITU-T H.265 (09/2023): "High efficiency video coding".

[h273] Recommendation ITU-T H.273 (09/2023): "Coding-independent code points for video signal type identification".

[h274] Recommendation ITU-T H.274 (09/2023): "Versatile supplemental enhancement information messages for coded video bitstreams".

[CMAF] ISO/IEC 23000-19: "Information Technology Multimedia Application Format (MPEG-A) – Part 19: Common Media Application Format (CMAF) for segmented media".

[CENC] ISO/IEC 23001-7: "MPEG systems technologies - Part 7: Common encryption in ISO base media file format files".

[DPC] CTA-5003-B: "Web Application Video Ecosystem (WAVE): Device Playback Capabilities Specification", available at https://shop.cta.tech/products/web-application-video-ecosystem-device-playback-capabilities-cta-5003-b .

[6381] IETF RFC 6381: The 'Codecs' and 'Profiles' Parameters for "Bucket" Media Types.

[MSE] 3GPP TR 26.857, "5G Media Service Enablers"

[3dtv] A. Quested and B. Zegel, "3D-TV production standards - first report of the ITU-R Rapporteurs", EBU Technical Review, 2011 Q2, https://tech.ebu.ch/publications/trev\_2011-Q2\_3dtv\_quested

# 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].

**Access Unit:** Smallest individually accessible portion of data within a Bitstream to which unique timing information can be attributed.

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

**Coded Video Sequence:** A sequence of access units that consists of a series of coded frames and any associated metadata (required for decoder and rendering initialization and operations) and conforms to a specific video encoding format and aligns with a certain Operation Point, as defined in this document The first access unit of a CVS is a random access point.

**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*.

**Coded Video Layer:** A sequence of coded pictures within a Coded Video Sequence that can be identified by an unique identifier within the CVS, referred to as layer ID, and that represents one or more video signal components.

**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.

**Random Access Point:** see below (add)

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

**Representation Format:**

**Video Layer sub-bitstream**: The *sub-bitstream* generated by extracting one or more CVLs from a source *bitstream*.

**Video Signal**: to be added

**Video Signal Component**: to be added

## 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

CVL Coded Video Layer

CLVS Coded layer-wise video sequence

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

MSE Media Source Extension

MV-HEVC MultiView extensions of HEVC

NAL Network Abstraction Layer

RAP Random access point

SDR Standard Dynamic Range

UHD Ultra-High Definition

VCL Video Coding Layer

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 signal as well as associated metadata to a rendering and display process. The video signal follows a *representation format*. The video signal can be composed of one or more video signal components, for example a video signal can include multiple views. The representation format defines the signal components and each of its properties.

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 (CVSs).* A CVS is a sequence of access units that consists of a series of coded frames and any associated metadata (required for decoder and rendering initialization and operations). The first access unit of a CVS is a *random access point (RAP)*.

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

The sequence of CVSs is referred to as *Bitstream*. In the context of this specification, Bitstreams conform to a specific video coding format and a specific representation format. The combination of video coding format and a specific representation format is referred to as Operation Point.

Receivers conforming to an Operation Point can decode the bitstream and render the included video signal together with the provided metadata. In the decoding process, 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.

In an extension to the Figure 4.2-1 presented in Figure 4.2-2, a video signal 1 may include another video signal 2 (for example a lower resolution, a hero eye signal), and the video encoder may generate a Video Bitstream such that:

- A receiver conforming to operation point 2 is able to decode the entire video bitstream and supports rendering of the included video signal 2.

- In addition, a receiver conforming to operation point 1 is able to extract the relevant data and access units of the the entire video bitstream to decode video signal 1.



Figure 4.2-2 Extended Reference architecture for video operating points and capabilities with multi-layer Bitstream.

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.

Editor’s Note: The following text needs to be updated. If not it will be removed.

[

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.

]

## 4.3 Capability Specification

This specification defines the following capabilities:

- Video Decoding capability: The capability to decode any video bitstream that conforms to an operating point and provides a conforming output video signal and possibly associated metadata.

- Video Encoding capability: The capability to encode any video signal included in the operating point to a bitstream that is decodable by decoder that conforms to the same operating point.

While not explicitly stated in the capabilities, it is a requirement for decoders and receivers to process the data in real-time. For encoder, real-time encoding is typically also a requirement.

Editor’s Note: This is how far I got.

## 4.4 Video representation formats

### 4.4.1 Overview

This clause defines video representation formats in the context of media delivery in 3GPP. For this purpose, a set of video signal parameters are defined in clause 4.4.2, with the restriction on what is defined in 3GPP media delivery. Based on the defined video signal parameters, clause 4.4.3 defines 3GPP video representation formats.

### 4.4.2 Video signal parameters

Video signals considered in this specification are represented by a sequence of pictures, where a *picture* can represent either an array of *luma* samples in a monochrome format or an array of luma samples and two corresponding arrays of *chroma* samples in a 4:2:0, 4:2:2, or 4:4:4 colour format. Only *progressive* signals are considered. A component refers to an array or single sample from one of the three arrays (luma and two chroma) that compose a picture. The Luma component represents a sample array or single sample representing the monochrome signal related to the primary colours (denoted with the symbol *Y*), and a chroma component represents 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*.

Video signals are typically described by a set of parameters that are required for the proper rendering of the decoded signal. Table 4.4.2-1 documents typical video signal parameters and provides a definition and/or reference.

Table 4.4.2-1 Video Signal Parameters

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Definition | 3GPP restrictions | Service or Application restrictions |
| Spatial Resolution width | The number of active samples per line for the luma component.  Example values are 1280 or 1920 for HD, and 3840 for UHD.  NOTE: The width does not restrict the encoding resolution to fixed values. Cropping parameters can be indicated that prescribe decoders the need to remove spatial video samples in a partially filled coding block that are not intended for presentation. | No restrictions | Restrictions possible |
| Spatial Resolution height | The number of active lines per picture for the luma component.  Example values are 720 or 1080 for HD, and 2160 for UHD.  NOTE: The height does not restrict the encoding resolution to fixed values. Cropping parameters can be indicated that prescribe decoders the need to remove spatial video samples in a partially filled coding block that are not intended for presentation. | No restrictions | Restrictions possible |
| Scan Type | Indicates the source scan type of the pictures as defined in clause 7.3 of Rec. ITU-T H.273.  Typical value is progressive | Progressive only |  |
| Chroma format indicator | Indicates whether the picture has only a luma component or that the picture has three colour components that consist of a luma component and two associated chroma components, such that the width and height of each chroma component are the width and height of the luma component divided by a factor defined by the chroma format as defined in Rec. ITU-T H.274, clause 7.3. | 4:2:0 |  |
| Bit depth | Indicates the bit depth for the samples of the luma component and the samples of the two associated chroma components.  Note that in general, the bit depth of the luma component and of the two associated chroma components may differ.  Typical values are 8 or 10 bits. | 8 or 10 bits  Luma and chroma components shall not differ |  |
| Colour primaries | Indicates the chromaticity coordinates of the source colour primaries as specified in clause 8.1 of Rec. ITU-T H.273.  Typical values are 1 to refer to Rec. ITU-R BT.709-6 [bt709] or 9 to refer to Rec. ITU-R BT.2020-2 and Rec. ITU-R BT.2100-2. | BT.709 or BT.2020/BT.2100 |  |
| Transfer Characteristics | Either indicates the reference opto-electronic transfer characteristic function of the source picture as a function of a source input linear optical intensity input or indicates the inverse of the reference electro-optical transfer characteristic function as a function of an output linear optical intensity as defined in clause 8.2 of Rec. ITU-T H.273.  Typical values are 1 to refer to Rec. ITU-R BT.709-6, 14 to refer to Rec. ITU-R BT.2020-2 (10 bit), 16 to refer to the Rec. ITU-R BT.2100-2 perceptual quantization (PQ) system, or 18 to refer to the Rec. ITU-R BT.2100-2 hybrid log-gamma (HLG) system | BT.709, BT.2020 SDR, BT.2100 PQ, or BT.2100 HLG |  |
| Matrix Coefficients | Describes the matrix coefficients used in deriving the luma and chroma signals from the green, blue, and red primaries. A video full range flag may be supplied with this parameter specifying the scaling and offset values applied in association with the Matrix coefficients. For detailed definition refer to clause 8.2 of Rec. ITU-T H.273.  Typical values are 1 to refer to the non constant luminance YCbCr representation in Rec. ITU-R BT.709-6 or 9 to refer to the non constant luminance YCbCr representations in Rec. ITU-R BT.2020-2 and Rec. ITU-R BT.2100-2. | YCbCr BT.709, YCbCr BT.2020, or YCbCr BT.2100 |  |
| Frame rate | Typical values, using frames per second, are: 120, 120/1.001, 100, 60, 60/1.001, 50, 30, 30/1.001, 25, 24, 24/1.001 | No restrictions | services may only permit a restricted subset |
| Frame packing | Indicates a frame packing arrangement, if present, as defined in clause 8.4 of Rec. ITU-T H.273. | Typically restricted to no frame packing. | Some applications may use frame packing. |
| Projection | Indicates a projection, if present, as defined in Rec. ITU-T H.274, clause 7.3, and typically refers to packing arrangements in clause 8.6 of Rec. ITU-T H.274. | Typically restricted to no projection. | Some applications may use projections. |
| Sample aspect ratio | Indicates width-to-height aspect ratio of the luma samples of the associated pictures as defined in clause 7.3 of Rec. ITU-T H.273.  Typical value is 1 | No specific restrictions, but 1 is expected. |  |
| Chroma sample location type | Specifies the location of the chroma samples relative to the luma samples for frames as defined in Rec. ITU-T H.273, clause 8.7.  Typical values are 0 (chroma samples are horizontally co-sited with and vertically centered between the first luma sample at the top-left corner and the first two luma samples at the top-left corner, respectively) or 2 (chroma samples are co-sited with the luma sample at the top-left corner).  Note that a value of 1 is common for still images. | No specific restrictions, but 0 is expected if not present. For HDR the value is typically set to 2. |  |
| Range | Specifies how luma and chroma samples are represented in digital video as defined in Rec. ITU‑T H.273, clause 8.3 using the parameter VideoFullRangeFlag.  For video applications only the value set to 0 is used, i.e. the video range or restricted range is applied where the luma values range from 16 to 235 in an 8-bit system, and chroma values range from 16 to 240. For 10-bit systems, the values are multiplied by 4.  Note that for still images full range (value set to 1) is commonly used. | No specific restrictions, but 0 is expected if not present. |  |

Certain video experiences are concurrently displaying video signals composed of multiple components. In this case, the video representation format describes each video signal component individually with the parameters defined in Table 4.4.2-1. Additionally, the components of the same video signal are typically jointly described and constrained for properly rendering the video representation.

The video signals made of multiple components can be delivered in either of the following forms:

* As a single encoded video signal using frame packing as defined in Table 4.4.2-1.
* As multiple encoded video signals

Table 4.4.2-2 lists the multi-component video signal parameters.

Table 4.4.2-2 Multi-component Video Signal Parameters

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Definition** | **3GPP restrictions** | **Service or Application restrictions** |
| Stereoscopic Video | Visual media may be stereoscopic, in which case the video signal is composed of two signal components: a view is available to be presented to the left eye and another view is available to be presented simultaneously to the right eye. The presentation of both the left and right views allows for an effect known as stereopsis, which can be defined as "the perception of depth produced by the reception in the brain of visual stimuli from both eyes in combination; binocular vision."  For signal representations, [3dtv] recommends that the Left and Right eyes comply to regular image formats such as Rec. ITU-R BT.709 and any necessary 3D-specific metadata is incorporated with the data. Hence, for stereoscopic video, two synchronized video signals are available, each with identical format parameters (such as the ones defined in this table).  NOTE: When distributing the signal, some systems may use different resolutions for one of the views.  Additional metadata that may be added with stereoscopic video:  - "Hero eye" is the default eye in a stereo (stereoscopic) video pair, often determined by tags set by the cameras used to capture the video. If so signaled, this indicates that the other stereo eye view is derived from the specified stereo eye and may be useful when choosing which eye to use in a monoscopic viewing environment. There is no requirement that either of the two eyes (or views) is tagged as the hero eye, in which case no hero eye tagging may be present. |  |  |

### 4.4.3 3GPP Video Formats

#### 4.4.3.1 Introduction

While a variety of formats may be used based on the video signal parameters defined in clause 4.4.2, for consistent programs and signals, several video formats are defined by a set of restrictions using the video signal parameters in clause 4.4.2. These signals are primarily used to develop interoperability points for TV and movie content distribution.

The present clause describes the signal characteristics of the following 3GPP video formats:

- 3GPP High Definition (HD): is meant to address the distribution of conventional 2D video services including HDTV and other conventional 2D formats.

- 3GPP High Dynamic Range (HDR): enables the distribution of 2D video up to 4K, e.g., for Ultra HD TV, and adds the support of high dynamic range capability on top of the 3GPP HD format.

- 3GPP Stereoscopic: is a format suitable for the video consumption of devices creating a depth perception using 2 images, one for each eye.

#### 4.4.3.2 High-Definition

3GPP High-Definition (HD) formats are defined based on Rec. ITU-R BT-709-6 [bt709]. 3GPP HD formats shall conform to Rec. ITU-R BT-709-6 [bt709] with the following restrictions and extensions:

- Only the following formats are included 24/P, 25/P, 30/P, 50/P and 60/P. Interlace and progressive segmented frame signals are excluded.

- Only the Non-Constant Luminance YCbCr signal format is considered.

- Other aspect ratios than 16:9 may be considered to address different screen sizes and orientations.

An informative summary of the parameters of a 3GPP HD format based on the parameters defined in Table 4.4.2-1 is provided in Table 4.4.3.2-1.

Table 4.4.3.2-1 Video Signal Parameters for 3GPP HD format

|  |  |
| --- | --- |
| Parameter | Restrictions |
| Picture aspect ratio | 16:9 is the only format defined in ITU-R BT-709-6 [bt709].  In 3GPP, to support different applications with different screen sizes and orientations, other picture aspect ratios may be considered including 9:16 and 1:1.  NOTE 1: The display orientation of the pictures in the video signal, for example portrait or landscape mode is implicit to the picture aspect ratio, but may be explicitly signalled.  NOTE 2: The aspect ratio of the encoded pictures may be different from the picture aspect ratio of the video signal. |
| Spatial Resolution width x height | 1920 × 1080 is the only format defined in ITU-R BT-709-6 [bt709].  Other spatial resolutions may be considered to address different aspect ratios, for example 1080 x 1920, 1024 x 1024, 1440 x 1440.  NOTE 1: Down-sampled resolutions may be created for distribution, for example in case of adaptive streaming.  NOTE 2: To accommodate the block coding structure of a given specification, quite often the encoded signal may be padded. In such cases, normative cropping is typically applied to remove spatial samples that are not intended to be presented.  NOTE 3: The width and the height of the encoded pictures may be different from the width and the height of the pictures in the video signal. |
| Scan Type | The source scan type of the pictures as defined in clause 7.3 of Rec. ITU-T H.273 is progressive. |
| Chroma format indicator | The chroma format indicator is 4:2:0. |
| Bit depth | The permitted values are 8 or 10 bit. The bit depth is the same for all samples. |
| Colour primaries | Only the value 1, as defined in clause 8.2 of Rec. ITU-T H.273, is permitted. |
| Transfer Characteristics | Only the value 1, as defined in clause 8.2 of Rec. ITU-T H.273 is permitted. |
| Matrix Coefficients | Only the value 1, as defined in clause 8.2 of Rec. ITU-T H.273, is permitted. |
| Frame rates | The permitted values are 60, 60/1.001, 50, 30, 30/1.001, 25, 24, 24/1.001 fps. |
| Frame packing | No frame packing is applied. |
| Projection | No projection is used. |
| Sample aspect ratio | The pixel aspect ratio is 1 (square pixel), i.e. only the value 1 as defined in clause 7.3 of Rec. ITU-T H.273 is permitted. |
| Chroma sample location type | The location of the chroma samples relative to the luma samples for progressive frames as defined in Rec. ITU-T H.273, clause 8.7, is set to 0 (chroma samples are horizontally co-sited with and vertically centered between the first luma sample at the top-left corner and the first two luma samples at the top-left corner, respectively). |
| Range | The restricted video range is used. |

#### 4.4.3.3 High Dynamic Range

3GPP High Dynamic Range (HDR) TV formats are defined based on Rec. ITU-R BT-2100-2 [bt2100]. 3GPP HDR TV formats shall conform to ITU-R BT-2100-2 [bt2100] with the following restrictions and extensions:

- Only 4:2:0 colour subsampling is considered

- Only the Non-Constant Luminance YCbCr signal format is considered

- Only 10-bit representations are considered

- Other aspect ratios than 16:9 may be considered in order to address different screen sizes and orientations.

An informative summary of the parameters of a 3GPP HDR format based on the parameters defined in Table 4.4.2‑1 is provided in Table 4.4.3.3-1.

Table 4.4.3.3-1 Video Signal Parameters for 3GPP HDR format

|  |  |
| --- | --- |
| Parameter | Restrictions |
| Picture aspect ratio | 16:9 is the only format defined in ITU-R BT-2100-2 [bt2100].  In 3GPP, to support different applications with different screen sizes and orientations, other picture aspect ratios may be considered including 9:16 and 1:1.  NOTE 1: The display orientation of the pictures in the video signal, for example portrait or landscape mode is implicit to the picture aspect ratio, but may be explicitly signalled.  NOTE 2: The aspect ratio of the encoded pictures may be different from the picture aspect ratio of the video signal. |
| Spatial Resolution width x height | 7680 × 4320, 3840 × 2160, 1920 × 1080 are the only formats supported in ITU-R BT-2100-2 [bt2100].  Other spatial resolutions may be considered to address different aspect ratios, for example 1080 x 1920, 1024 x 1024, 1440 x 1440.  NOTE 1: Down-sampled resolutions may be created for distribution, for example in case of adaptive streaming.  NOTE 2: To accommodate the block coding structure of a given specification, quite often the encoded signal may be padded. In such cases, normative cropping is typically applied to remove spatial samples that are not intended to be presented.  NOTE 3: The width and the height of the encoded pictures may be different from the with and the height of the pictures in the video signal. |
| Scan Type | the source scan type of the pictures as defined in clause 7.3 of Rec. ITU-T H.273 is progressive |
| Chroma format indicator | The chroma format indicator is 4:2:0. |
| Bit depth | The permitted value is 10 bit. |
| Colour primaries | Only the value 9 as defined in clause 8.2 of Rec. ITU-T H.273 is permitted. |
| Transfer Characteristics | Only the values 14 (for SDR with WCG), 16 (for PQ) and 18 (for HLG) as defined in clause 8.2 of Rec. ITU-T H.273 are permitted. |
| Matrix Coefficients | Only the value 9 as defined in clause 8.2 of Rec. ITU-T H.273 is permitted. |
| Frame rates | The permitted values are 120, 120/1.001,100, 60, 60/1.001, 50, 30, 30/1.001, 25, 24, 24/1.001 fps. |
| Frame packing | No frame packing is applied. |
| Projection | No projection is used. |
| Sample aspect ratio | The pixel aspect ratio is 1 (square pixel), i.e. only the value 1 as defined in clause 7.3 of Rec. ITU-T H.273 is permitted. |
| Chroma sample location type | the location of chroma samples relative to the luma samples for progressive frames as defined in Rec. ITU-T H.273, clause 8.7 is set to 2 (chroma samples are co-sited with the luma samples at the top-left corner). |
| Range | The restricted video range is used. |

#### 4.4.3.4 Stereoscopic format

The 3GPP Stereoscopic format uses a two-component video signal, one component for the left eye and another component for the right eye as defined in Table 4.4.3-1. The components for each eye closely follow the specifications of the 3GPP HDR format, but there are some restrictions and extensions, namely:

- Only 4:2:0 colour subsampling is considered.

- Frame rates include high frame rate for movies, namely 48 fps.

- the spatial resolution for each component is restricted to a maximum value of 4K (3840 × 2160).

- Only the Non-Constant Luminance YCbCr signal format is considered.

- Square picture aspect ratios are supported for different screen sizes.

An informative summary of the parameters of a 3GPP Stereoscopic format based on the parameters defined in Table 4.4.2-1 is provided in Table 4.4.3.4-1.

Table 4.4.3.4-1 Video Signal Parameters for 3GPP Stereoscopic format for each component

|  |  |
| --- | --- |
| Parameter | Restrictions |
| Picture aspect ratio | 16:9, 1:1. |
| Spatial Resolution width x height | 3840 × 2160, 1920 × 1080, 2048 × 2048, 1024 × 1024.  NOTE 1: Down-sampled resolutions may be created for distribution, for example in case of adaptive streaming.  NOTE 2: To accommodate the block coding structure of a given specification, quite often the encoded signal may be padded. In such cases, normative cropping is typically applied to remove spatial samples that are not intended to be presented. |
| Scan Type | The source scan type of the pictures as defined in clause 7.3 of Rec. ITU-T H.273 is progressive |
| Chroma format indicator | The chroma format indicator is 4:2:0. |
| Bit depth | The permitted values are 8 or 10 bit. 8 bit is only permitted for SDR. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

## 4.5 Common Bitstream Constraints

### 4.5.1 General

This clause defines common definitions for bitstreams that are used in capability definitions in the remainder of this document.

### 4.5.2 AVC Bitstreams

For an AVC/ITU-T H.264 [h264] bitstream, *motion-vector constraints* are defined that the bitstream does neither include horizontal motion vector component values that exceed the range from −2048 to 2047, inclusive, nor does have vertical motion vector component values that exceed the range from −512 to 511, inclusive, in units of ¼ luma sample displacement.

NOTE: This constraint should be indicated by using values of log2\_max\_mv\_length\_horizontal less than or equal to 11 and values of log2\_max\_mv\_length\_vertical less than or equal to 9.

For an AVC/ITU-T H.264 [h264] bitstream, *rate constraints* are defined that the for the bitstream,

- the maximum VCL Bit Rate is constrained to be 120 Mbps with cpbBrVclFactor and cpbBrNalFactor being fixed to be 1250 and 1500, respectively; and

- the bitstream does not contain more than 16 slices per picture.

### 4.5.3 HEVC Bitstreams

The following definitions are provided for HEVC/ITU-T H.265 [h265] bitstreams.

For an HEVC/ITU-T H.265 [h265] bitstream, *progressive constraints* are defined that the following flags in the active Sequence Parameter Set (SPS):

- general\_progressive\_source\_flag shall be set to 1,

- general interlaced\_source\_flag shall be set to 0,

- general\_non\_packed\_constraint\_flag shall be set to 1, and

- general\_frame\_only\_constraint\_flag shall be set to 1.

For an HEVC/ITU-T H.265 [h265] bitstream, *VUI constraints* are defined:

- Video Parameter Sets (VPS) NAL units as defined in Recommendation ITU-T H.265 / ISO/IEC 23008-2 [h265] may be present, but the Bitstream shall be valid if the Receiver ignores the VPS.

- The Video Usability Information (VUI) is present in the active Sequence Parameter Set, i.e. the vui\_parameters\_present\_flag shall be set to 1.

- In the VUI,

- the aspect ratio information is present, i.e. the aspect\_ratio\_info\_present\_flag value shall be set to 1,

- the colour parameter information is present, i.e. video\_signal\_type\_present\_flag value shall be set to 1 and the colour\_description\_present\_flag value shall be set to 1.

- only video range signals are used, i.e. the video\_full\_range\_flag shall be set to 0,

- no overscan signalling is present, i.e. the overscan\_info\_present\_flag shall be set to 0,

- the chroma location shall be signalled, i.e. chroma\_loc\_info\_present\_flag shall be set to 1,

- the timing information may be present. If the timing information is present, i.e. the value of vui\_timing\_info\_present\_flag is set to 1, then the values of vui\_num\_units\_in\_tick and vui\_time\_scale shall be set according to the frame rates allowed for each operation point. The timing information present in the video Bitstream should be consistent with the timing information signalled at the system level. The frame rate shall not change between two RAPs. fixed\_frame\_rate\_flag value, if present, shall be set to 1.

[For an HEVC/ITU-T H.265 [h265] bitstream, *frame-packing constraints* are defined:

- the following flags in the active Sequence Parameter Set (SPS):

- general\_progressive\_source\_flag shall be set to 1,

- general interlaced\_source\_flag shall be set to 0,

- general\_non\_packed\_constraint\_flag shall be set to 0, and

- general\_frame\_only\_constraint\_flag shall be set to 1.

- The frame packing arrangement SEI message shall be present with the following characteristics:

- The value of frame\_packing\_arrangement\_type shall be set to either the value of 3 for the side-by-side packing arrangement, or the value of 4 for the top-bottom/over-under packing arrangement.

- The value of quincunx\_sampling\_flag shall be set to 0.

- The value of content\_interpretation\_type shall be set to either 1 or 2.

- The value of spatial\_flipping\_flag shall be set to 0.

- The value of frame0\_flipped\_flag shall be set to 0.

- The value of field\_views\_flag shall be set to 0.

- The value of current\_frame\_is\_frame0\_flag shall be set to 0.

- The values of frame0\_grid\_position\_x, frame0\_grid\_position\_y, frame1\_grid\_position\_x, and frame1\_grid\_position\_y, shall remain the same throughout the bitstream.

- The value of upsampled\_aspect\_ratio\_flag shall be set to 0, indicating the presence of full resolution frame packed video and the aspect\_ratio\_idc shall be set to 1.

- All parameters shall remain the same for the entire bitstream.

## 4.6 Reference API parameters

### 4.6.1 Introduction

When media is played back, the decoder and the playback pipeline need to be initialized. For this purpose, certain parameters are required. In CTA-5003 [DPC], a media playback model is described that is aligned with HTML 5.1 and the <video> element, as well as the Media Source Extensions.

### 4.6.2 Video Decoder API Parameters

Video decoders are typically accessed by API parameters. The parameters are used for the following purposes:

- to identify the capability of the device in order to check whether the signal can be played back

- to initialize the decoding and playback platform to allocate the resources for decoding and rendering

Table 4.6.2-1 provide relevant parameters for Video Decoder APIs.

Table 4.6.2-1 Video Decoder API Parameters

|  |  |  |
| --- | --- | --- |
| Parameter | Restrictions | Status |
| media type | Specifies the media type of the component, in this case video. | required |
| codecs | Specifies through a well-defined string the codec parameters which the encoded video signal is compliant to. | required |
| video format parameters | Specifies additional video format parameters as defined in Table 4.4.2.1 to describe the signal and to initialize the encoder. | optional |

Editor’s Note: The capability of such API for decoding and playback of multilayer content, e.g. for stereoscopic content needs to be documented.

### 4.6.3 Video Encoder API Parameters

Video encoder API parameters are for further study.

### 4.6.4 Player API Parameters

Media players are typically configurable via API parameter. The main purpose of the API are:

- For video components, to create one or more display windows to display the decoded video signal

- To bind a media source, possibly remote, to the one or more created display windows.

Table 4.6.2-2 Display Window Object Parameters

|  |  |  |
| --- | --- | --- |
| Parameter | Restrictions | Status |
| width | Specifies the width of a video player window, in pixels | required |
| height | Specifies the width of a video player window, in pixels. | required |
| video format parameters | Specifies additional video format parameters as defined in Table 4.4.2-1 to describe the signal. | optional |

Editor’s Note: The relationship between the width and height in the above table and the spatial resolution of the video signal needs be to be clarified.

# 5 Video Coding Capabilities

## 5.1 Overview

This clause defines video decoding capabilities and video encoding capabilities for 3GPP media delivery.

NOTE: These clause does not specify whether these capabilities are required, recommended or suggested to be supported. This aspect is left specific service specifications or external specifications to refer to the capabilities defined in this clause.

5.2 Codecs, Profiles and Levels

5.2.1 Codec & profile

This specification defines capabilities based on the following video codecs and video codec profiles:

- AVC/H.264 Progressive High Profile [h264],

- HEVC/H.265 Main Profile Main Tier [h265],

- HEVC/H.265 Main-10 Profile Main Tier [h265].

- HEVC/H.265 Multiview Main 10 Main Tier [h265].

[- HEVC/H.265 Multiview Extended 10 Tier [h265].]

5.2.2 Codec & profile & Levels

This specification defines capabilities based on the following video codec profile and levels:

- AVC/H.264 Progressive High Profile Level 3.1,

- AVC/H.264 Progressive High Profile Level 4.0,

- AVC/H.264 Progressive High Profile Level 4.2,

- AVC/H.264 Progressive High Profile Level 5.1,

- AVC/H.264 Progressive High Profile Level 6.1,

- HEVC/H.265 Main Profile Main Tier Level 3.1,

- HEVC/H.265 Main-10 Profile Main Tier Level 4.1,

- HEVC/H.265 Main-10 Profile Main Tier Level 5.1,

- HEVC/H.265 Main 10 Profile Main Tier, Level 5.2,

- HEVC/H.265 Main-10 Profile Main Tier Level 6.0,

- HEVC/H.265 Main-10 Profile Main Tier Level 6.1,

- HEVC/H.265 Multiview Main 10 Profile Main Tier Level 5.1,

[- HEVC/H.265 Multiview Extended 10 Profile Main Tier Level 5.1.]

5.3 Single-Instance Decoding Capabilities

Editor’s Note: Comment from Waqar

* Decoding capabilities should just be the profile-level-tier. Here we could do away with bitstream or VUI constraints.
* All these constraints on bitstream and VUI can then be on the operating point. So these operating points can even refer to 4.5
* Common Bitstream Constraints where needed, or where specific constraints are needed, we can add these there.

5.3.1 AVC Decoding Capabilities

The following decoding capabilities are defined:

**- AVC-FullHD-Dec**: the capability to decode AVC/ITU-T H.264 Progressive High Profile Level 4.0 [h264] bitstreams.

**- AVC-UHD-Dec:** the capability to decode AVC/ITU-T H.264 Progressive High Profile Level 5.1 [h264] bitstreams with *rate constraints* as defined in clause 4.5.2.

**- AVC-8K-Dec:** the capability to decode AVC/ITU-T H.264 Progressive High Profile Level 6.1 [h264] bitstreams with the with *motion-vector* constraints and *rate constraints* as defined in clause 4.5.2.

5.3.2 HEVC Decoding Capabilities

The following decoding capabilities are defined:

ONE EXAMPLE FROM APPLE 830

- **HEVC-HD-Dec**: the capability to decode

A bitstream containing a single] sub-bitstream conforming to HEVC/ITU-T H.265 Main Profile, Main Tier, Level 3.1 [h265] with *progressive* constraints as defined in clause 4.5.3, or

A bitstream containing multiple layers where the base layer sub-bitstream conforms to HEVC/ITU-T H.265 Main Profile, Main Tier, Level 3.1 [h265] with *progressive* constraints as defined in clause 4.5.3.

END EXAMPLE FROM APPLE 830

- **HEVC-HD-Dec**: the capability to decode bitstreams conforming to both, HEVC/ITU-T H.265 Main Profile, Main Tier, Level 3.1 [h265] bitstreams with *progressive* constraints as defined in clause 4.5.3.

- **HEVC-FullHD-Dec**: the capability to decode bitstreams conforming to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 4.1 [h265] bitstreams with *progressive* and *VUI* constraints as defined in clause 4.5.3.

- **HEVC-UHD-Dec**: the capability to decode bitstreams conforming to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 5.1 [h265] bitstreams with *progressive* and *VUI* constraints as defined in clause 4.5.3.

- **HEVC-8K-Dec**: the capability to decode bitstreams conforming to HEVC/ITU-T H.265 Main10 Profile, Main Tier, Level 6.1 [h265] bitstreams with *progressive* and *VUI* constraints as defined in clause 4.5.3 and further constraints:

- the bitstream does not exceed the maximum luma picture size in samples of 33,554,432,

- the maximum VCL Bit Rate is constrained to be 80 Mbps with CpbVclFactor and CpbNalFactor being fixed to be 1000 and 1100, respectively.

**- MV-HEVC-Dual-layers-UHD420-Dec**: the capability to decode bitstreams with

- an HEVC/ITU-T H.265 Main 10 Profile base layer (layer\_id=0),

- and a single enhancement layer that is tagged either:

- as an HEVC/ITU-T H.265 Multiview Main 10 layer, or

- as an HEVC/ITU-T H.265 Multiview Extended 10 layer (layer\_id=1) [h265].

- where each layer conforms to Main Tier, Level 5.1.

NOTE: Both layers are in 4:2:0 format and inter-layer prediction is possible.

NOTE: For this decoding capability the Multiview Main 10 and Multiview Extended 10 profiles are functionally equivalent.

The device should be capable of supporting single layer decoding of HEVC/ITU-T H.265 Main 10 Profile bitstreams at Main Tier, Level 5.2. All layers shall follow the *progressive* and *VUI* constraints as defined in clause 4.5.3.

Editor’s Note: The removal of brackets for Extended 10 is subject to verification that we can playback such content on receivers. For this purpose, we recommend check using the VET-AM1008-v1 with direct http link to the test streams: <https://www.itu.int/wftp3/av-arch/jvet-site/bitstream_exchange/HEVCMultiview/under_test/>.

- **HEVC-Frame-Packed-Stereo-Dec**: the capability to decode bitstreams conforming to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 6.0 [h265] bitstreams with *frame-packing* and *VUI* *constraints* as defined in clause 4.5.3

NOTE: The increase from Level 5.2 for MV-HEVC-UHD-Dec to Level 6.0 in HEVC-Frame-Packed-Stereo-Dec is only to handle larger buffers per frame. There is no increase in the pixels/second between the two capabilities.

## 5.4 Single-Instance Encoding Capabilities

The following encoding capabilities are defined:

**- AVC-FullHD-Enc:** the capability to encode a video signal to a bitstream that is decodable by a decoder that is *AVC-FullHD-Dec* capable as defined in clause 5.3 with the following additional constraints:

- up to 245,760 macroblocks per second;

- up to a frame size of 8,192 macroblocks;

- up to 240 frames per second;

- the chroma format being 4:2:0; and

- the bit depth being 8 bit;

NOTE 1: The 3GPP HDTV format if restricted to 8 bit as defined in clause 4.4.3.2 may be encoded with an **AVC-FullHD-Enc** capable encoder.

- **HEVC-HD-Enc**: the capability to encode a video signal with

- up to 33,177,600 luma samples per second;

- up to a luma picture size of 983,040 samples;

- up to 120 frames per second;

- the chroma format being 4:2:0; and

- the bit depth being 8 bit;

to a bitstream that is decodable by a decoder that is **HEVC-HD-Dec** capable as defined in clause 5.3.

NOTE 2: A restricted version of the 3GPP HDTV format as defined in clause 4.4.3.2 may be encoded with an HEVC-HD-Enc capable encoder.

**- HEVC-FullHD-Enc:** the capability to encode a video signal to a bitstream that is decodable by a decoder that is *HEVC-FullHD-Dec* capable as defined in clause 5.3 with the following additional constraints:

- up to 133,693,440 luma samples per second;

- up to a luma picture size of 2,228,224 samples;

- up to 240 frames per second;

- the chroma format being 4:2:0; and

- the bit depth being either 8 or 10 bit;

NOTE 3: The 3GPP HDTV format as defined in clause 4.4.3.2 may be encoded with an ***HEVC-FullHD-Enc*** capable encoder. A restricted version of the 3GPP HDR TV format as defined in clause 4.4.3.3 may be encoded with an HEVC-FullHD-Enc capable encoder.

**- HEVC-UHD-Enc:** the capability to encode a video signal to a bitstream that is decodable by a decoder that is *HEVC-UHD-Dec* capable as defined in clause 5.3 with the following additional constraints:

- up to 534,773,760 luma samples per second;

- up to a luma picture size of 8,912,896 samples;

- up to 480 frames per second;

- the chroma format being 4:2:0; and

- the bit depth being either 8 or 10 bit;

NOTE 4: The 3GPP HDTV format as defined in clause 4.4.3.2 may be encoded with an ***HEVC-FullHD-Enc*** capable encoder. A restricted version of the 3GPP HDR TV format as defined in clause 4.4.3.3 may be encoded with an ***HEVC-FullHD-Enc*** capable encoder.

## 5.5 Multi-Instance Decoding Capabilities

The following multi-instance decoding capabilities are defined:

**- AVC-FullHD-Dec-2**: The capability of supporting up to two (*N*=2) concurrent decoder instances with the aggregate capabilities of *AVC-FullHD-Dec* as defined in clause 5.4.

**- AVC-UHD-Dec-4**: The capability of supporting up to four (*N*=4) concurrent decoder instances with the aggregate capabilities of *AVC-UHD-Dec* as defined in clause 5.4.

**- HEVC-UHD-Dec-4:** The capability of supporting up to four (*N*=4) concurrent decoder instances with the aggregate capabilities of *HEVC-UHD-Dec* as defined in clause 5.4.

**- UHD-Dec-4**: The capability supporting up to four (*N*=4) concurrent decoder instances with either:

- the aggregate capabilities of *AVC-UHD-Dec-4* as defined in this clause,

- the aggregate capabilities of *HEVC-UHD-Dec-4* as defined in this clause, or,

- the capability of decoding up to 4 bitstreams for which each bitstream does not exceed the capability of being decodable either with *AVC-FullHD-Dec* or *HEVC-FullHD-Dec* as defined in clause 5.4.

**- AVC-8K-Dec-8:** The capability of supporting up to eight (*N*=8)concurrent decoder instances with the aggregate capabilities of *AVC-8K-Dec* as defined in clause 5.4.

**- HEVC-8K-Dec-8:** The capability of supporting up to eight (*N*=8)concurrent decoder instances with the aggregate capabilities of *HEVC-8K-Dec* as defined in clause 5.4.

**- 8K-Dec-8**: The capability supporting up to eight (*N*=8)concurrent decoder instances with either:

- the aggregate capabilities of *AVC-8K-Dec-8* as defined in this clause,

- the aggregate capabilities of *HEVC-8K-Dec-8* as defined in this clause, or,

- the capability of decoding up to:

- eight bitstreams for which each bitstream does not exceed the capability of being decodable either with *AVC-FullHD-Dec* or *HEVC-FullHD-Dec* as defined in clause 5.4; or,

- four bitstreams for which each bitstream does not exceed the capability of being decodable either with *AVC-UHD-Dec* or *HEVC-UHD-Dec* as defined in clause 5.4.

## 5.6 Multi-Instance Encoding Capabilities

This specification does not define multi-instance encoding capabilities.

# 6 Video Operation Points

## 6.1 Introduction

Video operation points define a restricted subset of representation signals and media capabilities. For each Video Operation Point, requirements for the Bitstream and for the Receiver are defined.

Table 6.1-1 provides an overview of defined video operation points.

Table 6.1-1 Overview of Video Operation Points

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Video Format | Decoding Capabilities | Definition |
| 3GPP-AVC-HD | 3GPP-HD (see clause 4.4.3.2) | AVC-FullHD-Dec (see clause 5.4) | 6.2.2 |
| 3GPP-HEVC-HD | 3GPP-HD (see clause 4.4.3.2) | HEVC-FullHD-Dec (see clause 5.4) | 6.3.2 |
| 3GPP-HEVC-HD-HDR | 3GPP-HDR (see clause 4.4.3.3) | HEVC-FullHD-Dec (see clause 5.4) | 6.3.3 |
| 3GPP-HEVC-UHD-HDR | 3GPP-HDR (see clause 4.4.3.3) | HEVC-UHD-Dec (see clause 5.4) | 6.3.4 |
| 3GPP-HEVC-Stereo | 3GPP-Stereo (see clause 4.4.3.4) | HEVC-Frame-Packed-Stereo-Dec (see clause 5.5) | 6.3.5 |
| 3GPP-MVHEVC-Stereo | 3GPP-Stereo (see clause 4.4.3.4) | MV-HEVC-UHD-Dec,(see clause 5.3.2) | 6.3.6 |

## 6.2 AVC Video Operation Points

### 6.2.1 Introduction

The clause defines operation points for AVC. The video Bitstream and Receiver shall conform to Recommendation ITU-T H.264 [h264] with the restrictions described in this clause.

### 6.2.2 3GPP AVC HD Operation Point

#### 6.2.2.1 Introduction

The AVC HD Operation Point permits consistent distribution of HD-based video using AVC. The remainder of this clause 6.2.2 defines the Bitstream and Receiver requirements for the 3GPP-AVC-HD receiver.

### 6.2.2 3GPP AVC HD Operation Point

#### 6.2.2.1 Introduction

The AVC HD Operation Point permits consistent distribution of HD-based video using AVC. The remainder of this clause 6.3.2 defines the Bitstream and Receiver requirements for the 3GPP-AVC-HD receiver.

#### 6.2.2.2 Bitstream Requirements

A 3GPP-AVC-HD Bitstream shall conform to the following requirements

- the Bitstream shall conform to AVC/ITU-T H.264 High Progressive Profile, Level 4.0 [h264] bitstreams with *rate* constraints as defined in clause 4.5.2.

- the Representation Format included in the Bitstream shall conform to the 3GPP-HD Representation format as defined in clause 4.4.3.2.

- the Bitstream shall be decodable by a decoder with **AVC-FullHD-Dec** decoding capabilities.

#### 6.2.2.3 Receiver Requirements

Receivers conforming to the Operation Point 3GPP-AVC-HD shall support decoding and rendering Bitstreams with the restrictions defined in clause 6.2.2.2.

NOTE 1: Rendering includes adherence to the parameters signalled in the bitstream to characterize the distributed Representation format.

## 6.3 HEVC Video Operation Points

### 6.3.1 Introduction

The clause defines operation points for HEVC. The video Bitstream and Receiver shall conform to Recommendation ITU-T H.265 / ISO/IEC 23008-2 [h265] with the restrictions described in this clause.

### 6.3.2 3GPP HEVC HD Operation Point

#### 6.3.2.1 Introduction

The HEVC HD Operation Point permits consistent distribution of HD-based video using HEVC. The remainder of this clause 6.3.2 defines the Bitstream and Receiver requirements for the 3GPP-HEVC-HD receiver.

#### 6.3.2.2 Bitstream Requirements

A 3GPP-HEVC-HD Bitstream shall conform to the following requirements

- the Bitstream shall conform to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 4.1 [h265] bitstreams with *progressive* and *VUI* constraints as defined in clause 4.5.3.

- the Representation Format included in the Bitstream shall conform to the 3GPP-HD Representation format as defined in clause 4.4.3.2.

- the Bitstream shall be decodable by a decoder with **HEVC-FullHD-Dec** decoding capabilities.

Based on this, the following additional restrictions apply

- The chroma sub-sampling shall be 4:2:0 and the value of chroma\_format\_idc shall be set to 1.

- The aspect\_ratio\_idc value shall be set to 1, indicating a square pixel format.

- In the VUI, the values of colour\_primaries, transfer\_characteristics and matrix\_coeffs each shall be set to 1.

- The value of chroma\_sample\_loc\_type\_top\_field shall be set to 0.

The timing information may be present.

- If the timing information is present, i.e. the value of vui\_timing\_info\_present\_flag is set to 1, then the values of vui\_num\_units\_in\_tick and vui\_time\_scale shall be set according to the frame rates allowed for each operation point. The timing information present in the video Bitstream should be consistent with the timing information signalled at the system level.

- The frame rate shall not change between two RAPs. fixed\_frame\_rate\_flag value, if present, shall be set to 1.

#### 6.3.2.3 Receiver Requirements

Receivers conforming to the Operation Point 3GPP-HEVC-HD shall support decoding and rendering Bitstreams with the restrictions defined in clause 6.3.2.2.

NOTE 1: Rendering includes adherence to the parameters signalled in the bitstream to characterize the distributed Representation format.

Receivers should ignore the content of all Video Parameter Sets (VPS) NAL units as defined in Recommendation ITU-T H.265 / ISO/IEC 23008-2 [h265].

NOTE 2: The VPS may be present to address requirements in other Operation Points, but the Bitstream also conforms to this Operation point.

There are no requirements on output timing conformance for H.265/HEVC decoding (Annex C of [6]). The Hypothetical Reference Decoder (HRD) parameters, if present, should be ignored by the Receiver.

### 6.3.3 3GPP HEVC HDR Operation Point

#### 6.3.3.1 Introduction

The HEVC HDR Operation Point permits consistent distribution of High Dynamic Range based video using HEVC. The remainder of this clause 6.3.3 defines the Bitstream and Receiver requirements for the 3GPP-HEVC-HDR receiver.

#### 6.3.3.2 Bitstream Requirements

A 3GPP-HEVC-HDR Bitstream shall conform to the following requirements

- the Bitstream shall conform to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 4.1 [h265] bitstreams with *progressive* and *VUI* constraints as defined in clause 4.5.3.

- the Representation Format included in the Bitstream shall conform to the 3GPP HDR Representation format as defined in clause 4.4.3.3.

- the Bitstream shall be decodable by a decoder with **HEVC-FullHD-Dec** decoding capabilities as defined in clause 5.3.2.

Based on this, the following additional restrictions apply

- The chroma sub-sampling shall be 4:2:0 and the value of chroma\_format\_idc shall be set to 1.

- The aspect\_ratio\_idc value shall be set to 1, indicating a square pixel format.

- In the VUI, the values of colour\_primaries and matrix\_coeffs each shall be set to 9, and the value of transfer\_characteristics shall be set to one of the following values: 14 (for SDR with WCG), 16 (for PQ) and 18 (for HLG).

- The value of the chroma\_sample\_loc\_type\_top\_field shall be set to 2.

The timing information may be present.

- If the timing information is present, i.e. the value of vui\_timing\_info\_present\_flag is set to 1, then the values of vui\_num\_units\_in\_tick and vui\_time\_scale shall be set according to the frame rates allowed for each operation point. The timing information present in the video Bitstream should be consistent with the timing information signalled at the system level.

- The frame rate shall not change between two RAPs. fixed\_frame\_rate\_flag value, if present, shall be set to 1.

#### 6.3.3.3 Receiver Requirements

Receivers conforming to this Operation Point 3GPP-HEVC-HDR shall support decoding and rendering Bitstreams with the restrictions defined in clause 6.3.3.2.

NOTE 1: Rendering includes adherence to the parameters signalled in the bitstream to characterize the distributed Representation format.

Receivers should ignore the content of all Video Parameter Sets (VPS) NAL units as defined in Recommendation ITU-T H.265 / ISO/IEC 23008-2 [h265].

NOTE 2: The VPS may be present to address requirements in other Operation Points, but the Bitstream also conforms to this Operation point.

There are no requirements on output timing conformance for H.265/HEVC decoding (Annex C of [6]). The Hypothetical Reference Decoder (HRD) parameters, if present, should be ignored by the Receiver.

### 6.3.4 3GPP HEVC UHD

#### 6.3.4.1 Introduction

The HEVC UHD Operation Point permits consistent distribution of Ultra-High-definition content using HEVC. The remainder of this clause 6.3.4 defines the Bitstream and Receiver requirements for the 3GPP-HEVC-UHD receiver.

#### 6.3.4.2 Bitstream Requirements

A 3GPP-HEVC-UHD Bitstream shall conform to the following requirements

- the Bitstream shall conform to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 5.1 [h265] bitstreams with *progressive* and *VUI* constraints as defined in clause 4.5.3.

- the Representation Format included in the Bitstream shall conform to the 3GPP HDR Representation format as defined in clause 4.4.3.3.

- the Bitstream shall be decodable by a decoder with **HEVC-UHD-Dec** decoding capabilities as defined in clause 5.3.2.

Based on this, the following additional restrictions apply

- The chroma sub-sampling shall be 4:2:0 and the value of chroma\_format\_idc shall be set to 1.

- The aspect\_ratio\_idc value shall be set to 1, indicating a square pixel format.

- In the VUI, the values of colour\_primaries and matrix\_coeffs each shall be set to 9, and the value of transfer\_characteristics shall be set to one of the following values: 14 (for SDR with WCG), 16 (for PQ) and 18 (for HLG).

- The value of the chroma\_sample\_loc\_type\_top\_field shall be set to 2.

The timing information may be present.

- If the timing information is present, i.e. the value of vui\_timing\_info\_present\_flag is set to 1, then the values of vui\_num\_units\_in\_tick and vui\_time\_scale shall be set according to the frame rates allowed for each operation point. The timing information present in the video Bitstream should be consistent with the timing information signalled at the system level.

- The frame rate shall not change between two RAPs. fixed\_frame\_rate\_flag value, if present, shall be set to 1.

#### 6.3.4.3 Receiver Requirements

Receivers conforming to this Operation Point 3GPP-HEVC-HDR shall support decoding and rendering Bitstreams with the restrictions defined in clause 6.3.4.2.

NOTE 1: Rendering includes adherence to the parameters signalled in the bitstream to characterize the distributed Representation format.

Receivers should ignore the content of all Video Parameter Sets (VPS) NAL units as defined in Recommendation ITU-T H.265 / ISO/IEC 23008-2 [h265].

NOTE 2: The VPS may be present to address requirements in other Operation Points, but the Bitstream also conforms to this Operation point.

There are no requirements on output timing conformance for H.265/HEVC decoding (Annex C of [6]). The Hypothetical Reference Decoder (HRD) parameters, if present, should be ignored by the Receiver.

### 6.3.5 3GPP HEVC Stereo

#### 6.3.5.1 Introduction

The HEVC Stereo Operation Point permits consistent distribution of stereoscopic content using HEVC with frame-packing. The remainder of this clause 6.3.5 defines the Bitstream and Receiver requirements for the 3GPP-HEVC-S receiver.

#### 6.3.5.2 Bitstream Requirements

A 3GPP-HEVC-Stereo Bitstream shall conform to the following requirements

- the Bitstream shall conform to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 5.2 [h265] bitstreams with frame-packing constraints as defined in clause 4.5.3.

- the Representation Format included in the Bitstream shall conform to the 3GPP Stereoscopic format as defined in clause 4.4.3.4.

- the Bitstream shall be decodable by a decoder with **HEVC-Stereo-Dec** decoding capabilities as defined in clause 5.3.2.

Based on this, the following additional restrictions apply

- The chroma sub-sampling shall be 4:2:0 and the value of chroma\_format\_idc shall be set to 1.

- In the VUI, either

- the values of colour\_primaries, transfer\_characteristics and matrix\_coeffs each shall be set to 1.

- The value of chroma\_sample\_loc\_type\_top\_field shall be set to 0.

- or

- the values of colour\_primaries and matrix\_coeffs each shall be set to 9, and the value of transfer\_characteristics shall be set to one of the following values: 14 (for SDR with WCG), 16 (for PQ) and 18 (for HLG).

- The value of the chroma\_sample\_loc\_type\_top\_field shall be set to 2.

The timing information may be present.

- If the timing information is present, i.e. the value of vui\_timing\_info\_present\_flag is set to 1, then the values of vui\_num\_units\_in\_tick and vui\_time\_scale shall be set according to the frame rates allowed for each operation point. The timing information present in the video Bitstream should be consistent with the timing information signalled at the system level.

- The frame rate shall not change between two RAPs. fixed\_frame\_rate\_flag value, if present, shall be set to 1.

Bitstreams not required to be associated with frame packing information for all coded video sequences. It is also possible that such information, when present, may defer from one coded video sequence to another.

#### 6.3.5.3 Receiver Requirements

Receivers conforming to this Operation Point 3GPP-HEVC-3D shall support decoding and rendering Bitstreams with the restrictions defined in clause 6.3.5.2.

NOTE 1: Rendering includes adherence to the parameters signalled in the bitstream to characterize the distributed Representation format.

Receivers should ignore the content of all Video Parameter Sets (VPS) NAL units as defined in Recommendation ITU-T H.265 / ISO/IEC 23008-2 [h265].

NOTE 2: The VPS may be present to address requirements in other Operation Points, but the Bitstream also conforms to this Operation point.

There are no requirements on output timing conformance for H.265/HEVC decoding (Annex C of [6]). The Hypothetical Reference Decoder (HRD) parameters, if present, should be ignored by the Receiver.

### 6.3.6 3GPP MVHEVC Stereo

#### 6.3.6.1 Introduction

The MVHEVC Stereo Operation Point permits consistent distribution of stereoscopic content using MVHEVC. The remainder of this clause 6.3.6 defines the Bitstream and Receiver requirements for the 3GPP-MVHEVC-Stereo receiver.

6.3.6.2 Bitstream Requirements

Editor’s Note: this needs additional signaling:

* Layer dependency is possible, but not needed. Can be two independent layers, i
* 3D reference displays information SEI message

A 3GPP-MV-HEVC-Stereo Bitstream shall conform to the following requirements

- the Representation Format included in the Bitstream shall conform to the 3GPP Stereoscopic format as defined in clause 4.4.3.4.

- The bitstream shall conform to the constraints specified in the **MV-HEVC-UHD** decoding capabilities as defined in clause 5.3.2.

- the Bitstream shall be decodable by

- a decoder with **HEVC-UHD-Dec** decoding capabilities as defined in clause 5.3.2.

- a decoder with **MV-HEVC-UHD** decoding capabilities as defined in clause 5.3.2.

Based on this, the following additional restrictions apply

- The chroma sub-sampling shall be 4:2:0 and the value of chroma\_format\_idc shall be set to 1.

- AuxId[ lId ] shall be equal to 0 in the VPS extension for sub-bitstream with nuh\_layer\_id != 0.

Editor’s Note: this should refer to the bitstream element and not the variable AuxId.

- The aspect\_ratio\_idc value shall be set to 1, indicating a square pixel format.

- In the VUI, either

- the values of colour\_primaries, transfer\_characteristics and matrix\_coeffs each shall be set to 1.

- The value of chroma\_sample\_loc\_type\_top\_field shall be set to 0.

- or

- the values of colour\_primaries and matrix\_coeffs each shall be set to 9, and the value of transfer\_characteristics shall be set to one of the following values: 14 (for SDR with WCG), 16 (for PQ) and 18 (for HLG).

- The value of the chroma\_sample\_loc\_type\_top\_field shall be set to 2.

The timing information may be present.

- If the timing information is present, i.e. the value of vui\_timing\_info\_present\_flag is set to 1, then the values of vui\_num\_units\_in\_tick and vui\_time\_scale shall be set according to the frame rates allowed for each operation point. The timing information present in the video Bitstream should be consistent with the timing information signalled at the system level.

- The frame rate shall not change between two RAPs. fixed\_frame\_rate\_flag value, if present, shall be set to 1.

Bitstreams not required to be associated with frame packing information for all coded video sequences. It is also possible that such information, when present, may differ from one coded video sequence to another.

6.3.6.3 Receiver Requirements

Receivers conforming to this Operation Point 3GPP-MV-HEVC-Stereo shall support decoding and rendering Bitstreams with the restrictions defined in clause 6.3.6.2.

NOTE 1: Rendering includes adherence to the parameters signalled in the bitstream to characterize the distributed Representation format.

There are no requirements on output timing conformance for H.265/HEVC decoding (Annex C of [6]). The Hypothetical Reference Decoder (HRD) parameters, if present, should be ignored by the Receiver.

# 7 Common System Integration

7.1 Introduction

This clause documents general functionalities that are relevant for integration of video codecs into delivery systems to support common APIs on encoders and decoders.

Editor’s NOTE: This text may be removed. Updates are needed.

[

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 CMAF/DASH for 5G Media Streaming, or ISO BMFF for Messaging Services. However, this specification provides mapping principles to delivery protocols.

]

7.2 Functional Definitions

### 7.2.1 General

##### 7.2.1.1 Summary

This clause defines functional definitions for system integration in Table 7.2.1.1-1. The remainder of this

Table 7.2.1.1-1 Functional Definitions

|  |  |  |
| --- | --- | --- |
| Term | Summary | Details |
| Codec String | A single value identifying the codec indicated to render the content in the Bitstream as defined in IETF RFC 6381. | 7.2.1.2 |
| Decoder Configuration | a data structure storing essential parameters needed for decoding and rendering a video stream. | 7.2.1.3 |
| Random Access Point | A byte position in the Bitstream, for which in combination with the Decoder Configuration, the Bitstream can be randomly accessed, i.e. in decoding order the Bitstream carries sufficient information to access the media in the stream. | 7.2.1.4 |
| Access Unit (AU) | See Clause 3.1 |  |
| Coded access unit (CAU) | bits corresponding to an Access Unit | 7.2.1.5 |
| Random Access CAU | A CAU that starts with a random access point | 7.2.1.6 |

##### 7.2.1.2 Codec String

The *Codec String* provides means to identify the codec needed to decode and render the content in the Bitstream. The codec parameters shall also include the profile and level information where applicable. The content of this parameter shall conform to the id-simple production of IETF RFC 6381:2011, subclause 3.2, without the enclosing DQUOTE characters. The codec identifier for the media format, mapped into the name space for codecs as specified in IETF RFC 6381:2011, subclause 3.3, shall be used.

##### 7.2.1.3 Decoder Configuration

The *Decoder Configuration* provides parameters about the Bitstream and shall follow the format defined in ISO/IEC 14496-15 including:

- profile, tier, level

- constraints flags

- chroma format

- bit depth chroma and luma

- frame rates, average or constant

- layering structure

- NAL units

- VPS (Video Parameter Set): Contains parameters that apply to the entire video sequence.

- SPS (Sequence Parameter Set): Contains parameters that apply to a sequence of pictures.

- PPS (Picture Parameter Set): Contains parameters that apply to individual pictures.

- declarative SEI NAL unit, as specified in ISO/IEC 23008-2. When one or more SEI NAL units containing an SEI manifest SEI message and/or an SEI prefix indication SEI message are available, they should be stored as instances of nalUnit.

##### 7.2.1.4 Random Access Point

Different types of Random Access Points are defined as follows:

**- Closed loop RAP (CL-RAP)** is an intra coded picture that can identify a RAP in a bitstream. It can be the first coded picture or can appear later in a bitstream. Each CL-RAP is the first picture in decoding order of a coded video sequence (CVS) but does not need to be an output picture or be the first picture in display order. All frames that follow a CL-RAP in decoding order and belong in the same coded video sequence are decodable and can potentially be all output by the decoder depending on their coding parameters.

**- Open loop RAP (OL-RAP)** is an intra coded frame that can identify a RAP in a bitstream. It can be the first frame in the bitstream in decoding order or can appear later in the bitstream. An OL-RAP does not need to be an output picture or be the first picture in display order. Other pictures that follow the OL-RAP in coding order can refer to an OL-RAP for prediction. However, an OL-RAP, if it is the first picture in the bitstream in decoding order, may also be followed in coding order by some pictures that can refer to pictures that are not present in the bitstream. In that case, these pictures cannot be decoded. These pictures can be referred to as leading pictures. Subsequently, when those pictures are detected, they are not decoded and can be discarded by the decoder.

**- Gradual decoder refresh (GDR) access point** identifies a RAP in a bitstream from where decoding operations can start by a decoder. However, unlike other RAP types, decoding may not be instantaneous and may initially result in decoding errors in the decoded and reconstructed pictures. Nevertheless, these decoding errors are expected to disappear after a certain maximum period, from which point decoding can continue without any further decoding errors.

##### 7.2.1.5 Coded Access Unit

Editor’s Note: This needs to be completed.

##### 7.2.1.6 Random Access CAU

Editor’s Note: This needs to be completed.

##### 7.2.1.7 Coded Video Layer

Different types of Coded Video Layer are defined as follows:

**- Independent CVL** is a CVL which does not depend on any other CVL in the CVS for prediction purposes.

**- Output CVL** is a CVL whose coded pictures are meant to be output after decoding.

**- Base CVL** is an Independent CVL and Output CVL and it is the first CVL in the CVS.

**- Dependent CVL** is a CVL that depends on a Base CVL for prediction purposes.

**- Auxiliary CVL** is an Independent CVL and Output CVL that is not the first CVL of the CVS.

Editor’s Note: This needs to be completed.

### 7.2.2 AVC

Editor’s Note: This needs to be completed.

### 7.2.3 HEVC

Editor’s Note: This needs to be completed.

Annex <A> (normative):  
Registration Information

Editor’s Note: Will collect and registration information such as URNs.

Annex <B> (informative):  
Mapping of Operation Points to Implementations

# B.1 Introduction

This annex provides some background on how to map the reference architectures defined in clause 4 into concrete implementations. The mapping of the capabilities, the configuration of the encoders and decoders through APIs as well as some workflow aspects are provided.

The Annex is not considered to prescribe any implementation but is expected to support implementors to integrate the capabilities and operating points defined in this specification into their workflows.

The Annex also serves as an analysis on what functionalities are available in existing implementations and where there are potential gaps that may be addressed by the owners of the implementation to fully support all features.

# B.2 WebCodecs API

## B.2.1 Introduction

The WebCodecs API [W3CCodecs] specifies a powerful web Application Programming Interface (API) that provides developers with low-level access to the individual samples of media, including frames of a video stream. It is useful for web applications that require full control over the way media is processed, such as video or audio editors, and video conferencing applications. The WebCodecs API uses an asynchronous processing model. Each instance of an encoder or decoder maintains an internal, independent processing queue.

The WebCodecs API provides several video related interfaces:

- VideoDecoder: Decodes EncodedVideoChunk objects.

- VideoEncoder: Encodes VideoFrame objects.

- EncodedVideoChunk: Represents codec-specific encoded video bytes.

- VideoFrame: Represents a frame of unencoded video data.

- VideoColorSpace: Represents the colour space of a video frame.

In order to map a codec to the WebCodecs API, a codec registration procedure for new codecs is defined by W3C in <https://www.w3.org/TR/webcodecs-codec-registry/>.

The registration requirements request the following details:

- A codec string and a specification that provides the details of the codecs string

- The codec string has certain requirements

- Each registration is expected to include

- Recognized codec strings

- EncodedVideoChunk internal data

- VideoDecoderConfig description bytes

- Expectations for EncodedVideoChunk

- Registration may include description of extensions to VideoEncoderConfig dictionaries

- Candidate entries are expected to be announced by filing an issue in the WebCodecs GitHub issue tracker (<https://github.com/w3c/webcodecs/issues/>) so they can be discussed and evaluated for compliance before being added to the registry.

## B.2.2 Mapping of Operation Points to Decoder API

Table B.2.2-1 provides a mapping of operation points to Web Codecs decoder API.

Table B.2.2-1 Mapping of Operation Points to Decoder API

Editor’s Note: This needs to be completed.

|  |  |  |  |
| --- | --- | --- | --- |
| Operating Point | Codecs String | Video Chunk | Video Decoder Config |
| 3GPP-AVC-HD | 'avc1.640029' or 'avc3.640029' | Tbd, see clause 7.2.3 | Tbd, see clause 7.2.3 |
| 3GPP-HEVC-HD | 'hvc1.2.4.L123.B0' or 'hev1.2.4.L123.B0' | Tbd, see clause 7.2.3 | Tbd, see clause 7.2.3 |
| 3GPP-HEVC-HD-HDR | 'hvc1.2.4.L123.B0' or 'hev1.2.4.L123.B0' | Tbd, see clause 7.2.3 | Tbd, see clause 7.2.3 |
| 3GPP-HEVC-UHD-HDR | 'hvc1.2.4.L153.B0' or 'hev1.2.4.L153.B0' | Tbd, see clause 7.2.3 | Tbd, see clause 7.2.3 |
| 3GPP-HEVC-3DTV | tbd | Tbd, see clause 7.2.3 | Tbd, see clause 7.2.3 |
| 3GPP-MVHEVC-3DTV | tbd | Tbd, see clause 7.2.3 | Tbd, see clause 7.2.3 |

## B.2.3 Mapping of Operation Points to Encoder API

Editor’s Note: This subclause needs to be completed.

Annex <X> (informative):  
Change history

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Change history | | | | | | | |
| Date | Meeting | TDoc | CR | Rev | Cat | Subject/Comment | New version |
| 2024-04 | SA4#127bis-e | S4-240616 |  |  |  | Initial version | 0.0.0 |
| 2024-04 | SA4#127bis-e | S4-240758 |  |  |  | Version agreed at SA4#127bis-e | 0.1.0 |
| 2024-05 | SA4#128 | S4-241369 |  |  |  | Version agreed at SA4#128 including S4-240911, S4-241296, S4-241298 | 0.2.2 |
| 2024-08 | SA4#129-e | S4-241669 |  |  |  | Version agreed at SA4#129-e including S4-241479, S4-241705 | 0.3.0 |
| 2024-10 | Post SA4#129-e Video SWG AHG | S4aV240073 |  |  |  | Version agreed during Post SA4#129-e Video SWG AHG October 29, 2024 adding S4aV240060 | 0.3.1 |
| 2024-11 | SA4#130 | S4-241892 |  |  |  | Version submitted for SA4#130 adding agreed S4aV240073 | 0.3.2 |
| 2024-11 | SA4#130 | S4-242064 |  |  |  | Version agreed at SA4#130 including S4-241894, S4-242174, S4-242209, S4-242211 | 0.5.0 |
| 2025-02 | SA4#131 | S4-250031 |  |  |  | Version submitted for SA4#131 | 0.5.1 |
| 2025-02 | SA4#131 | S4-250369 |  |  |  | Version agreed at SA4#131 including S4-250031, S4-250116, S4-250117, S4-250367, S4-2500368, S4-250369, S4-250370 | 0.6.0 |
| 2025-03 | SA#107 | SP-250281 |  |  |  | Version 1.0.0 created by MCC for sending to TSG SA for information | 1.0.0 |
| 2025-04 | SA4#131-bis-e | S4-250710 |  |  |  | Version agreed at SA4#131-bis-e including S4-250640, S4-250679, S4-250704, S4-250706, S4-250707, S4-250741 | 1.1.0 |