ETSI TS 126 265 V19.2.0 (2026-04)

TECHNICAL SPECIFICATION

5G;
Media Delivery: Video Capabilities and Operation Points
(3GPP TS 26.265 version 19.2.0 Release 19)

Reference

RTS/TSGS-0426265vj20

Keywords

5G

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Important notice

The present document can be downloaded from the
ETSI Search & Browse Standards application.

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format on ETSI deliver repository.

Users should be aware that the present document may be revised or have its status changed,
this information is available in the Milestones listing.

If you find errors in the present document, please send your comments to
the relevant service listed under Committee Support Staff.

If you find a security vulnerability in the present document, please report it through our
Coordinated Vulnerability Disclosure (CVD) program.

Notice of disclaimer & limitation of liability

The information provided in the present deliverable is directed solely to professionals who have the appropriate degree of experience to understand and interpret its content in accordance with generally accepted engineering or other professional standard and applicable regulations.
No recommendation as to products and services or vendors is made or should be implied.
No representation or warranty is made that this deliverable is technically accurate or sufficient or conforms to any law and/or governmental rule and/or regulation and further, no representation or warranty is made of merchantability or fitness for any particular purpose or against infringement of intellectual property rights.
In no event shall ETSI be held liable for loss of profits or any other incidental or consequential damages.

Any software contained in this deliverable is provided "AS IS" with no warranties, express or implied, including but not limited to, the warranties of merchantability, fitness for a particular purpose and non-infringement of intellectual property rights and ETSI shall not be held liable in any event for any damages whatsoever (including, without limitation, damages for loss of profits, business interruption, loss of information, or any other pecuniary loss) arising out of or related to the use of or inability to use the software.

Copyright Notification

No part may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm except as authorized by written permission of ETSI.
The content of the PDF version shall not be modified without the written authorization of ETSI.
The copyright and the foregoing restriction extend to reproduction in all media.

© ETSI 2026.
All rights reserved.

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI IPR online database.

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™**, **LTE™** and **5G™** logo are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Legal Notice

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

The present document may refer to technical specifications or reports using their 3GPP identities. These shall be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between 3GPP and ETSI identities can be found at 3GPP to ETSI numbering cross-referencing.

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Contents

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 Operation 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*.

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

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

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

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

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

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

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

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

[9] 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.

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

[11] 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

[12] SMPTE ST 2086:2018, Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images

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.

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

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

3.2 Symbols

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

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 Operation 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 may 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 using an APIs as shown 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.

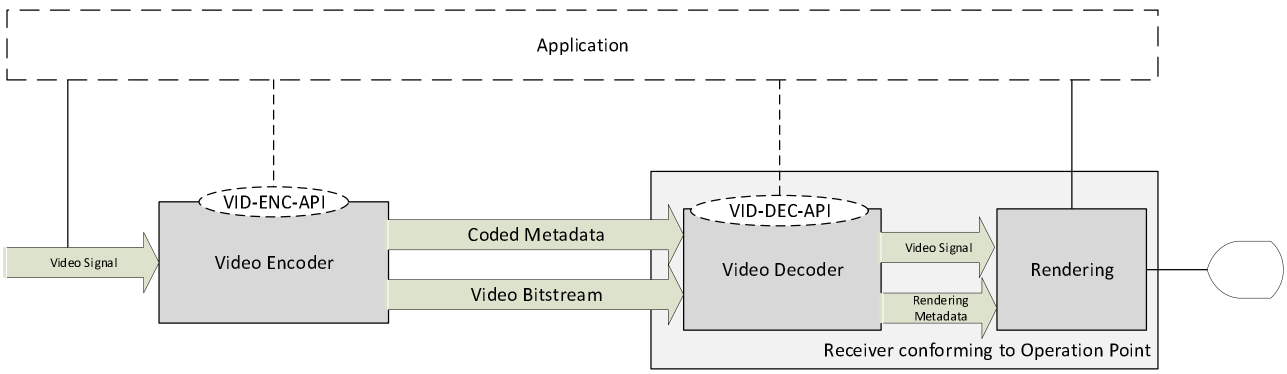

Figure 4.2-1 Reference architecture for video Operation points and capabilities

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 are able to 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 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, or 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 entire video bitstream to decode video signal 1.

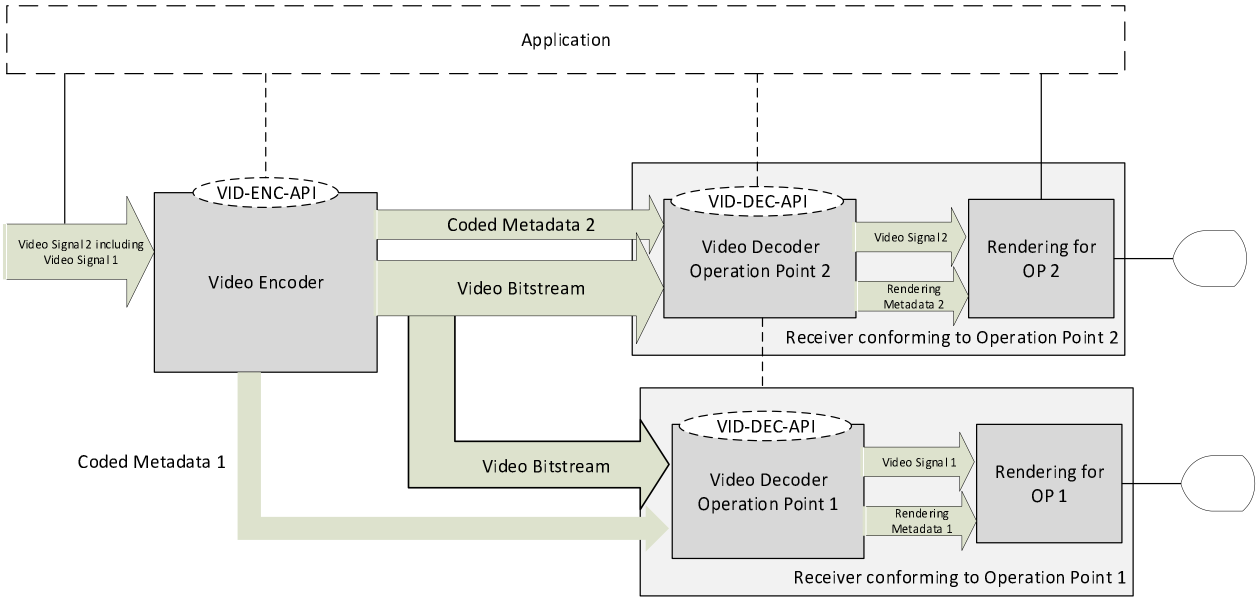

Figure 4.2-2 Extended Reference architecture for video Operation points and capabilities with multi-layer Bitstream.

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

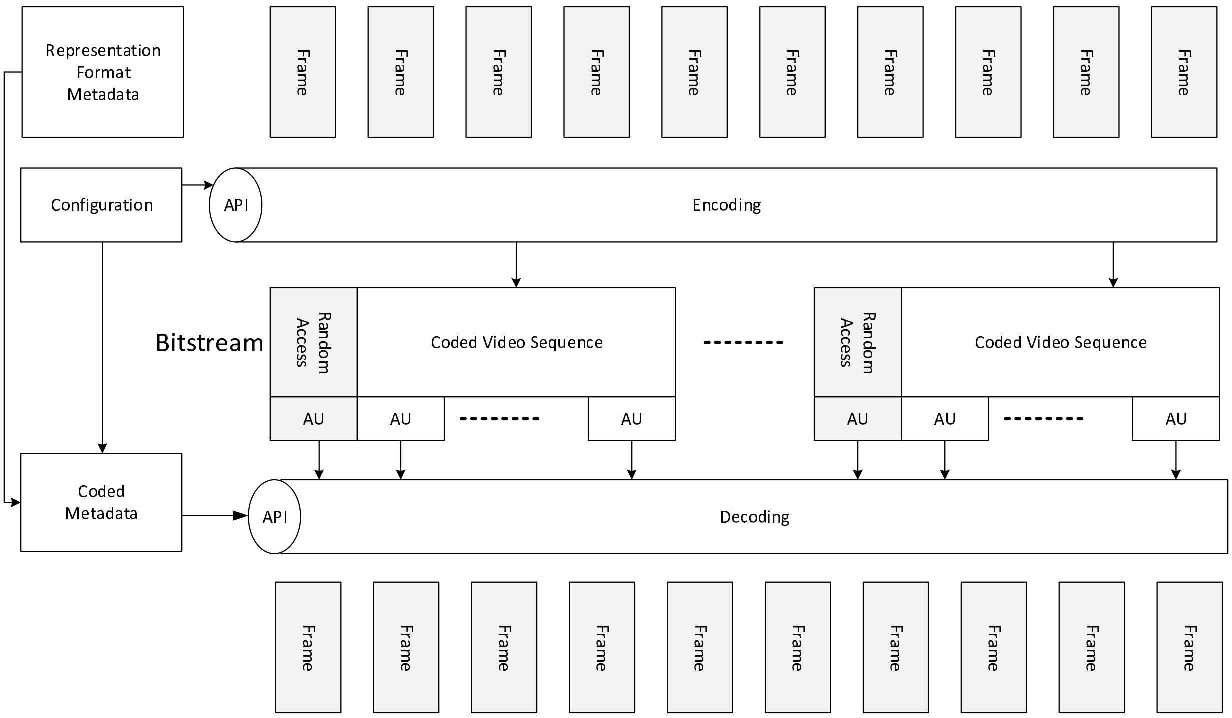

Figure 4.2-2 Informative Data model for illustration purposes

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.

Based on this introduction, the following terms are defined:

- **Operation Point:** A combination of video signal restrictions 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 Operation Points.
  - comprised by access units that serve as units to be provided to decoders for regenerating frames.
- **Conforming Receiver**: A function that can decode and render a Bitstream conforming to an Operation Point.

4.3 Capability Specification

This specification defines the following capabilities:

- Video Decoding capability: The capability to decode any video bitstream that conforms to an Operation 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 Operation point to a bitstream that is decodable by decoder that conforms to the same Operation point.

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

4.4 Video representation formats

4.4.1 Overview

This clause defines video representation formats in the context of media delivery in 3GPP. Video Representation Formats are restricted and well-defined video signals to be used within typical 3GPP service constraints.

In order to define video representation formats, video signal parameters are defined in clause 4.4.2.

Based on the defined video signal parameters, clause 4.4.3 defines several 3GPP video representation formats providing a subset of well-defined 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 common 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 [6].	Progressive	
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 [7], 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 use the same bit-depth	
Colour primaries	Indicates the chromaticity coordinates of the source colour primaries as specified in clause 8.1 of Rec. ITU-T H.273 [6]. Typical values are 1 to refer to Rec. ITU-R BT.709-6 [2] or 9 to refer to Rec. ITU-R BT.2020-2 [bt2020] and Rec. ITU-R BT.2100-2 [3].	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 [6]. 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 [6]. 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	Frame rate of the video signal. 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 [6].	No, SbS, TaB	Some applications may use frame packing.
Projection	Indicates a projection, if present, as defined in Rec. ITU-T H.274 [7], clause 7.3, and typically refers to packing arrangements in clause 8.6 of Rec. ITU-T H.274 [7].	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 [6]. 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 [6], 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 [6], 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.	

HDR static metadata	Optional information together with BT.2100 PQ, this can include information - Mastering Display Color Volume: Describes the display used during mastering (e.g., peak brightness, color primaries as defined in SMPTE ST 2086 [12], and - Content Light Level Information: Includes MaxCLL (Maximum Content Light Level) and MaxFALL (Maximum Frame Average Light Level) as defined in SMPTE ST 2086 [12]	No specific restrictions at this stage	

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 packaged 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 independently encoded video signals
- As a bitstream including an independently encoded signals and one or multiple dependent 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, [11] 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 time-aligned video signals are available, each with identical format parameters (such as the ones defined in Table 4.4.2-1).
	
	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.
	- optionally, reference display parameters for stereoscopic (3D) video content to support decoders and rendering systems on how the video should be displayed on 3D-capable devices, ensuring consistent and accurate depth perception across different viewing environments. It includes:
	  - reference display width,
	  - optionally, a reference viewing distance,
	  - optionally, a sample shift values to adjust stereo alignment.		

4.4.3 3GPP Video Representation 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 experiences, several 3GPP video representation formats are defined by a set of restrictions using the video signal parameters in clause 4.4.2. These signals are typically used to develop interoperability points for TV and movie content distribution but also have application for user-generated content.

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

The 3GPP High-Definition (HD) video representation format is defined based on Rec. ITU-R BT-709-6 [2].

3GPP HD formats shall conform to Rec. ITU-R BT-709-6 [2] 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 included.
- Other aspect ratios than 16:9 may be considered to address different screen sizes and orientations.

The definition of the 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	Settings
Picture aspect ratio	16:9 should be used as it is the only format defined in ITU-R BT-709-6 [2]. However, to support different applications with different screen sizes and orientations, other picture aspect ratios may be used 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 should be used as it is the only format defined in ITU-R BT-709-6 [2]. However, to support different applications, other spatial resolutions may be used, 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 shall be progressive.
Chroma format indicator	The chroma format indicator shall be 4:2:0.
Bit depth	The values shall be either 8 or 10 bit. The bit depth shall be the same for all samples.
Colour primaries	Only the value 1, as defined in clause 8.2 of Rec. ITU-T H.273 [6], is permitted.
Transfer Characteristics	Only the value 1, as defined in clause 8.2 of Rec. ITU-T H.273 [6] is permitted.
Matrix Coefficients	Only the value 1, as defined in clause 8.2 of Rec. ITU-T H.273 [6], 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 shall be applied.
Projection	No projection shall be used.
Sample aspect ratio	The pixel aspect ratio shall be 1 (square pixel), i.e. only the value 1 as defined in clause 7.3 of Rec. ITU-T H.273 [6] 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 [6], clause 8.7, shall be 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 shall be used.

4.4.3.3 High Dynamic Range

The 3GPP High Dynamic Range (HDR) format is defined based on Rec. ITU-R BT-2100-2 [3].

3GPP HDR TV formats shall conform to ITU-R BT-2100-2 [3] with the following restrictions and extensions:

- Only 4:2:0 colour subsampling is used
- Only the Non-Constant Luminance YCbCr signal format is used
- Only 10-bit representations are used
- Other aspect ratios than 16:9 may be used in order to address different screen sizes and orientations.

The definition of the 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	Settings
Picture aspect ratio	16:9 should be used as it is the only format defined in ITU-R BT-2100-2 [3]. However, to support different applications with different screen sizes and orientations, other picture aspect ratios may be used 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 [3] and should therefore be used. Other spatial resolutions may be used 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 [6] is progressive
Chroma format indicator	The chroma format indicator shall be 4:2:0.
Bit depth	The permitted value shall be 10 bit.
Colour primaries	Only the value 9 as defined in clause 8.2 of Rec. ITU-T H.273 [6] 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 [6] 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 shall be applied.
Projection	No projection shall be 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 [6] 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 [6], clause 8.7 shall be set to 2 (chroma samples are co-sited with the luma samples at the top-left corner).
Range	The restricted video range shall be 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 follow the specifications of the 3GPP HDR format, but there are some restrictions and extensions, namely:

- Only 4:2:0 colour subsampling is used.
- 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 used.
- Square picture aspect ratios are supported for different screen sizes.

The definition of the 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

Parameter		Settings
Stereoscopic Video		A video signal for the left and for the right eye is provided whereby the signals shall have identical parameters for all parameters except for the view identifier and with the restrictions below. The frames are time-synchronized.
Signal parameters for each of the two views with restrictions		
	View identifier	left or right
	Picture aspect ratio	Shall be set to 16:9, 1:1.
	Spatial Resolution width x height	Should be set to 3840 × 2160, 1920 × 1080, 2048 × 2048, 1024 × 1024. However, other resolutions are permitted. 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 [6] shall be progressive
	Chroma format indicator	The chroma format indicator shall 4:2:0.
	Bit depth	The permitted values are 8 or 10 bit. 8 bit is only permitted for SDR.
	Colour primaries Transfer Characteristics Matrix Coefficients	Only the following value combinations are permitted: (1, 1, 1), (9, 14, 9), (9, 16, 9), and (9, 18, 9) for SDR HD, SDR UHD, HDR PQ, and HDR HLG, respectively.
	Frame rates	The permitted values are 60, 60/1.001, 48, 48/1.001, 50, 30, 30/1.001, 25, 24, 24/1.001 fps.
	Projection	No projection is used.
	Sample aspect ratio	The pixel aspect ratio shall be 1 (square pixel), i.e. only the value 1 as defined in clause 7.3 of Rec. ITU-T H.273 [6] is permitted.
	Chroma sample location type	For SDR HD, the location of chroma samples relative to the luma samples for progressive frames as defined in Rec. ITU-T H.273 [6], clause 8.7 shall be set to 0. For SDR UHD, HDR PQ, and HDR HLG, the location of chroma samples relative to the luma samples for progressive frames as defined in Rec. ITU-T H.273 [6], clause 8.7, shall be set to 2.
	Range	The restricted video range shall be used.
Common parameters		
	Hero eye	Left or right If absent, no hero eye is specified.
	Reference display parameter	Parameters include - reference display width - optionally, a reference viewing distance, - optionally, a sample shift values to adjust stereo alignment. If absent, no reference display parameters are specified.

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 [4] 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 [4] 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 [5] bitstreams.

A set of *progressive constraints* is defined as follows:

- In the active Sequence Parameter Set (SPS) of a base layer bitstream (`nuh_layer_id == 0`) the flags:
  - `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`.

A set of *frame-packing constraints* is defined as follows:

- In the active Sequence Parameter Set (SPS) of a base layer bitstream (`nuh_layer_id == 0`) the flags:
  - `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` (left view first) or `2` (right view first).

NOTE: the hero eye, if provided in the representation format, may be indicated with the `content_interpretation_type`. -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 each be set to `0`.
  - The value of `upsampled_aspect_ratio_flag` shall be set to 0, indicating the presence of full resolution frame packed video.
  - All parameters of the frame packing arrangement SEI message shall remain the same for the entire bitstream.

The following *VUI constraints* are defined:

- Video Parameter Sets (VPS) NAL units as defined in Recommendation ITU-T H.265 / ISO/IEC 23008-2 [5] 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 for the base layer bitstream (nuh_layer_id == 0), i.e. the `vui_parameters_present_flag` shall be set to `1`.
- In the VUI,
  - the aspect ratio information should be present, i.e. the `aspect_ratio_info_present_flag` value should 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. When the timing information is present, the values of `vui_num_units_in_tick` and `vui_time_scale` shall be consistent with the frame rates allowed for each operation point and 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.

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 [9], 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

NOTE: The capability of such API for decoding and playback of multilayer content, e.g. for stereoscopic content is for further study.

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

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 [4],
- HEVC/H.265 Main Profile Main Tier [5],
- HEVC/H.265 Main-10 Profile Main Tier [5],
- HEVC/H.265 Multiview Main 10 Main Tier [5],
- HEVC/H.265 Multiview Extended 10 Main Tier [5].

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

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 [4] bitstreams.
- **AVC-UHD-Dec:** the capability to decode AVC/ITU-T H.264 Progressive High Profile Level 5.1 [4] 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 [4] bitstreams with *motion-vector* constraints and *rate constraints* as defined in clause 4.5.2.

5.3.2 Single-layer HEVC Decoding Capabilities

The following single-layer HEVC decoding capabilities are defined:

- **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 [5] 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 [5] with *progressive* constraints as defined in clause 4.5.3.
- **HEVC-FullHD-Dec**: the capability to decode
  - a bitstream containing a single sub-bitstream conforming to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 4.1 [5] 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 10 Profile, Main Tier, Level 4.1 [5] with *progressive* constraints as defined in clause 4.5.3.
- **HEVC-UHD-Dec**: the capability to decode
  - a bitstream containing a single sub-bitstream conforming to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 5.1 [5] 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 10 Profile, Main Tier, Level 5.1 [5] with *progressive* 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 [5] 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.

- **HEVC-Frame-Packed-Stereo-Dec**: the capability to decode a bitstream conforming to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 6.0 [5] bitstreams with *frame-packing* and *VUI constraints* as defined in clause 4.5.3

NOTE: The increase from Level 5.2 for MV-HEVC-Main-Dual-layers-UHD420-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.

5.3.3 Multi-layer HEVC Decoding Capabilities

The following *dual-layer HEVC decoding capabilities* are defined:

- **MV-HEVC-Main-Dual-layers-UHD420-Dec**: the capability
  - to decode bitstreams with an HEVC/ITU-T H.265 Main 10 Profile base layer (`nuh_layer_id==0`), and a single enhancement layer (`nuh_layer_id!=0`) that is tagged as an HEVC/ITU-T H.265 Multiview Main 10 layer [5], and
  - to decode bitstreams with an HEVC/ITU-T H.265 Main Profile base layer (`nuh_layer_id==0`), and a single enhancement layer (`nuh_layer_id!=0`) that is tagged as an HEVC/ITU-T H.265 Multiview Main Profile layer [5], and
  - where each layer conforms to Main Tier, Level 5.1 and where UE should be capable of supporting single layer decoding of HEVC/ITU-T H.265 Main 10 Profile bitstreams at Main Tier, Level 5.2.

NOTE 1: Both layers are in 4:2:0 format and inter-layer prediction is possible.

NOTE 2: HEVC decoders with this decoding capability are also capable to decode bitstreams with a Main Profile base layer as it is a compatible profile according to H.265/HEVC [5], and a single enhancement Multiview Main layer (with `nuh_layer_id!=0`), with the same tier and level restrictions as above, as specified by H.265/HEVC [5].

- **MV-HEVC-Ext-Dual-layers-UHD420-Dec**: the capability
  - to decode bitstreams with an HEVC/ITU-T H.265 Main 10 Profile base layer (`nuh_layer_id==0` and a single enhancement layer (`nuh_layer_id!=0`) that is tagged as an HEVC/ITU-T H.265 Multiview Extended 10 layer [5].
  - to decode bitstreams with an HEVC/ITU-T H.265 Main Profile base layer (`nuh_layer_id==0`), and a single enhancement layer (`nuh_layer_id!=0`) that is tagged as an HEVC/ITU-T H.265 Extended Main Profile layer [5], and
  - where each layer conforms to Main Tier, Level 5.1 and where UE should be capable of supporting single layer decoding of HEVC/ITU-T H.265 Main 10 Profile bitstreams at Main Tier, Level 5.2.

NOTE 3: Both layers are in 4:2:0 format and inter-layer prediction is possible.

NOTE 4: HEVC decoders with this decoding capability can also decode bitstreams with a Main Profile base layer as it is a compatible profile according to H.265/HEVC [5], and a single enhancement Multiview Extended layer (`nuh_layer_id!=0`), with the same tier and level restrictions as above, as specified by H.265/HEVC [5].

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 HD 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 HD 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-MV-HEVC-Main-Stereo	3GPP-Stereo (see clause 4.4.3.4)	MV-HEVC-Main-Dual-layers-UHD420-Dec (see clause 5.3.2)	6.3.6.4
3GPP-MV-HEVC-Ext-Stereo	3GPP-Stereo (see clause 4.4.3.4)	MV-HEVC-Ext-Dual-layers-UHD420-Dec (see clause 5.3.2)	6.3.6.5

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 [4] 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 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 6.2.2 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 [4] 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 [5] 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 [5] 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`.
- A square sample aspect ratio (1:1) of the luma samples shall be used. Hence, the `aspect_ratio_idc` value, if present, shall be set to `1`.
- 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.

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 [5].

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 [5] 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`.
- A square sample aspect ratio (1:1) of the luma samples shall be used. Hence, the `aspect_ratio_idc` value, if present, shall be set to `1`.
- 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.

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 [5].

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 [5] 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`.

- A square sample aspect ratio (1:1) of the luma samples shall be used. Hence, the `aspect_ratio_idc` value, if present, shall be set to `1`.

- 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.

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 [5].

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 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 HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 5.2 [5] bitstreams with *VUI constraints* and *frame-packing constraints* as defined in clause 4.5.3.

- 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.

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.

If the content is rendered in stereoscopic 3D with left and right eye, the receiver shall use the value of `content_interpretation_type` of the frame packing arrangement SEI message to map the views correctly to each eye.

If the content is rendered in 2D, the receiver should render the first view as specified in the `content_interpretation_type` of the frame packing arrangement SEI message.

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 [5].

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 MV-HEVC Stereo

6.3.6.1 Introduction

The MV-HEVC Stereo Operation Points permits consistent distribution of stereoscopic content using MV-HEVC.

The remainder of this clause 6.3.6 defines the Common Bitstream and Receiver requirements as well as two specific operation points based on two *dual-layer MV-HEVC decoding capabilities*, namely **MV-HEVC-Main-Dual-layers-UHD420-Dec** and **MV-HEVC-Ext-Dual-layers-UHD420-Dec** as defined in clause 5.3.2.

6.3.6.2 Common Bitstream Requirements

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

1) the Representation Format included in the Bitstream shall conform to the 3GPP Stereoscopic format as defined in clause 4.4.3.4.
2) The bitstream shall conform to the constraints specified in a *dual-layer MV-HEVC decoding capability* as defined in this clause 5.3.2 and the *VUI* constraints as defined in clause 4.5.3.
3) The base layer shall conform to the *progressive constraints* as defined in clause 4.5.3.

4) The bitstream shall contain a first output layer set containing the layer (`nuh_layer_id = 0`) which follows the constraints specified in the **HEVC-UHD-Dec** decoding capabilities as defined in clause 5.3.2.

NOTE: the hero eye, if provided in the representation format, can be indicated with this first output layer set.

5) The bitstream shall contain a second output layer set containing the layer (`nuh_layer_id` = 0) as output layer and a second layer as output layer which follows the constraints specified in the *dual-layer MV-HEVC decoding capability* as defined in clause 5.3.2. This second layer corresponds to a scalability dimension of type Multiview.

NOTE: Although the operating point allows for layers in the bitstream that are not output layers, the added storage and/or transport capacity needed for such layers need to be taken into account when provisioning a service.

6) In the VPS,

- The value `vps_num_layer_sets_minus1` shall be equal to or greater than 1.
- The value `layer_id_included_flag[ 1 ][ 0 ]` shall be equal to 1 and there shall be a value of `j` with `j` different from 0 for which `layer_id_included_flag[ 1 ][ j ]` is equal to 1.

NOTE: This implements constraint (4) from above into the VPS without an explicit assignment of the layer id to the second output layer.

- The value of `scalability_mask_flag[ 1 ]` shall be equal to 1.
- The value of `ScalabilityId[ 1 ][ 1 ]` shall be derived equal to 1.
- The value of `default_output_layer_idc` shall be equal to 0.

NOTE: These three constraints document that view scalability is used for the second layer.

- The `direct_dependency_flag[ j ][ 0 ]` may either be set to 0 or to `1`.

NOTE: This implies, that layer-dependency is possible, but not needed. The two layers may be independent (if set to 0), or the second layer depends on the base layer (if set to 1).

7) The chroma sub-sampling shall be 4:2:0 and the value of `chroma_format_idc` shall be set to 1.

8) In the VUI,

- A square sample aspect ratio (1:1) of the luma samples shall be used. Hence, the `aspect_ratio_idc` value, if present, shall be set to `1`.
- 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.

9) The Bitstream shall include the `three_dimensional_reference_displays_info` SEI message as specified in Recommendation ITU-T H.265 / ISO/IEC 23008-2 [5] with the following constraints:

- The value of `num_ref_displays_minus1` shall be set to 0.
- The value of the `left_view_id[ 0 ]` shall be set to the corresponding value defined in the `view_id_val` parameter.

- The value of the `right_view_id[ 0 ]` shall be set to the corresponding value defined in the `view_id_val` parameter and shall be different to the `left_view_id[ 0 ]`.

NOTE: This allows to assign right and left eye of a stereo representation signal when using this operation point.

- The remaining parameters may be set based on available reference display parameter as defined in clause 4.4. In the absence of such information, the following parameters may be set:
  - The `prec_ref_display_width` is set to 31.
  - The `ref_viewing_distance_flag` is set to 0.
  - The `exponent_ref_display_width[0]` and `mantissa_ref_display_width[0]` are both set to 0.
  - The `additional_shift_present_flag[0]` and the `three_dimensional_reference_displays_extension_flag` are both set to 0.

VPS NAL units may be present in the bitstream or conveyed by other means. If conveyed by other means, the Recommendation ITU-T H.265 / ISO/IEC 23008-2 [5] requires the VPS NAL units to be available to the decoding process in a timely fashion.

6.3.6.3 Common Receiver Requirements

Receivers conforming to common 3GPP-MV-HEVC-Stereo constraints shall support decoding and rendering Bitstreams with the restrictions defined in clause 6.3.6.2, including the necessary processing of `three_dimensional_reference_displays_info` SEI message as specified in Recommendation ITU-T H.265 / ISO/IEC 23008-2 [5].

If the content is rendered in stereoscopic 3D with left and right eye, the receiver shall use the value of `left_view_id[0]` and `right_view_id[0]` of `three_dimensional_reference_displays_info` SEI message to map the views correctly to each eye.

If the content is rendered in 2D, the receiver should render the view represented by the first output layer set.

NOTE: 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.

6.3.6.4 3GPP MV-HEVC-Main Stereo

This MV-HEVC Stereo Operation Point permits consistent distribution of stereoscopic content using MV-HEVC based on **MV-HEVC-Main-Dual-layers-UHD420-Dec** decoding capabilities as defined in clause 5.3.2.

A 3GPP-MV-HEVC-Main-Stereo Bitstream shall conform to the common 3GPP-MV-HEVC-Stereo bitstream requirements as defined in clause 6.3.6.2 with the *dual-layer MV-HEVC decoding capability* instantiated as **MV-HEVC-Main-Dual-layers-UHD420-Dec** as defined in clause 5.3.2.

Receivers conforming to the 3GPP-MV-HEVC-Main Stereo operation points shall conform to the common receiver constraints in clause 6.3.6.3 for 3GPP-MV-HEVC-Main-Stereo Bitstreams.

6.3.6.5 3GPP MV-HEVC-Ext Stereo

This MV-HEVC Stereo Operation Point permits consistent distribution of stereoscopic content using MV-HEVC based on **MV-HEVC-Ext-Dual-layers-UHD420-Dec** decoding capabilities as defined in clause 5.3.2.

A 3GPP-MV-HEVC-Ext-Stereo Bitstream shall conform to the common 3GPP-MV-HEVC-Stereo bitstream requirements as defined in clause 6.3.6.2 with the *dual-layer MV-HEVC decoding capability* instantiated as **MV-HEVC-Ext-Dual-layers-UHD420-Dec** as defined in clause 5.3.2.

Receivers conforming to the 3GPP-MV-HEVC-Ext Stereo operation points shall conform to the common receiver constraints in clause 6.3.6.3 for 3GPP-MV-HEVC-Ext-Stereo Bitstreams.

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.

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
Coded Video Layer (CVL)	See Clause 3.1	7.2.1.7

7.2.1.2 Codecs Parameter String

The *Codecs Parameter String* provides means to identify the codec needed to decode and render the content in the Bitstream. The codecs parameter string 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

7.2.1.4.1 Definitions

Relevant types of Random Access Points for this specification 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 coded pictures 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 picture that can identify a RAP in a bitstream. It can be the first coded picture 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.

For adaptive streaming applications with CMAF [8], CMAF fragments start with a CL-RAP. More CL-RAP or OL-RAPs may be present within those CMAF fragments.

7.2.1.4.3 Messaging

Content shared with messaging applications starts with a CL-RAP. More CL-RAP or OL-RAPs may be present within the files shared via messaging.

7.2.1.5 Coded Access Unit

Coded access unit definition is for further study.

7.2.1.6 Random Access CAU

Random Access CAU definition is for further study.

7.2.1.7 Coded Video Layer

A CVL represents a component of a video signal (e.g., luma, chroma, auxiliary data).

1. **Layer Identification**: Each CVL is identified with a unique layer ID in the bitstream and decoder configuration.
2. **Dependency Signaling**: CVL dependencies for decoding purpose, if any, is typically declared in the Video Parameter Set (VPS) or SEI messages.
3. **Random Access**: A Random Access Point (RAP) in a CVS shall enable decoding of all CVLs starting from that point.

Different types of Coded Video Layer exist:

- **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 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 another CVL for prediction purposes.
- **Auxiliary CVL** is an Output CVL that is not the first CVL of the CVS.

7.2.2 AVC

The definitions for AVC are for further study.

7.2.3 HEVC

The definitions for HEVC are for further study.

Annex A: Mapping of Operation Points to Implementations

A.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 Operation 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.

A.2 WebCodecs API

A.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.

A.2.2 Mapping of Operation Points to Decoder API

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

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

Operation Point	Codecs String according to IETF RFC 6381 [10]	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	For further study	Tbd, see clause 7.2.3	Tbd, see clause 7.2.3
3GPP-MV-HEVC-3DTV	'desc.usecase=vstereo+codec=hvc1.1.6.L93.B0'(Note 1) or 'hvc1.2.4.L153.B0' or 'hev1.2.4.L153.B0'	Tbd, see clause 7.2.3	Tbd, see clause 7.2.3
Note 1: 'desc' 4CC is used to signal rendering capabilities. 'usecase' specifies the intended use case of the media, here 'vstereo' implying that the resource contains a stereo video pair. 'codec' embeds the codec-specific string.			

A.2.3 Mapping of Operation Points to Encoder API

This work is for further study.

Annex B: Conformance

B.1 Introduction

This annex provides conformance-related information. Figure B.1-1 provides an overview of the conformance functionalities provided in this annex to support interoperable deployment of the operation points defined in this specification.

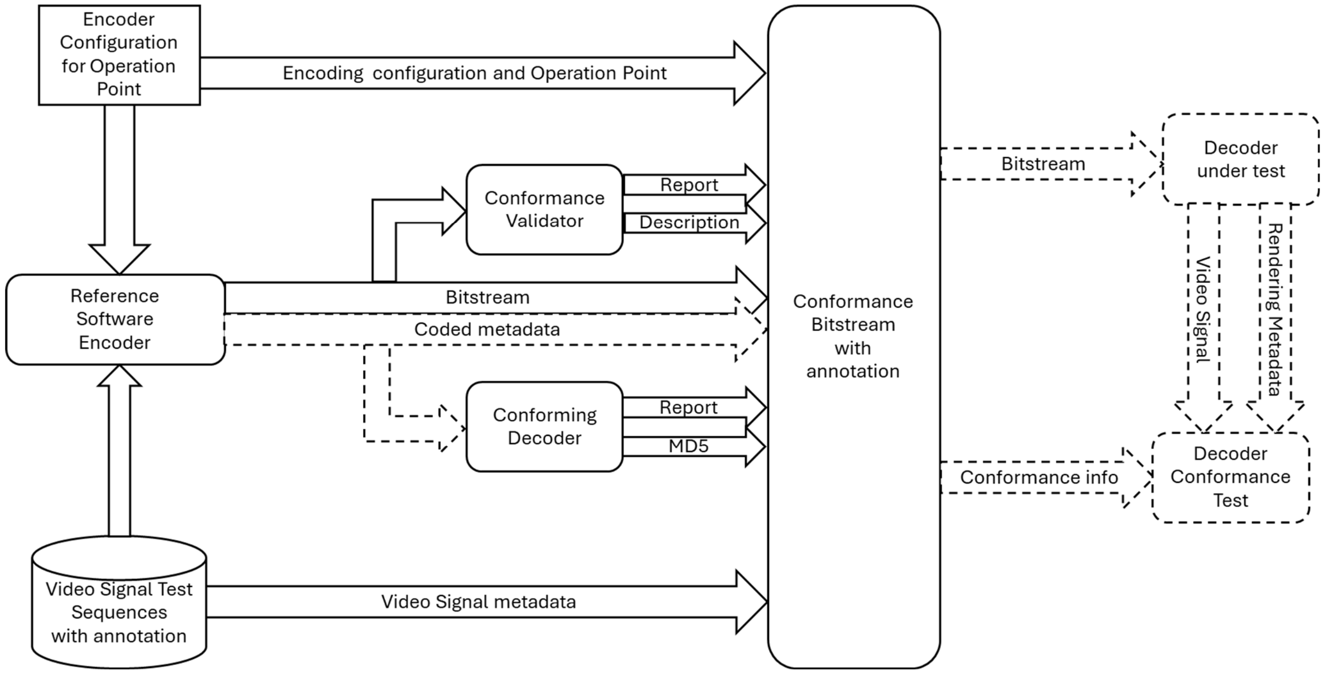

Figure B.1-1 Overview of conformance functionalities provided for this specification

Based on the overview in Figure B.1-1, the following functionalities are provided:

- A set of video signals are provided that conform to one of the representation formats defined in clause 4.4.3 of this specification. In addition, a consistent annotation of the signals using the parameters defined in clause 4.4 are defined. For details refer to Annex B.2.1.

- Workflows including reference software encoders with appropriate configuration are provided to encode the video signals to create a bitstream and potentially coded metadata that conform to the operation point. For details refer to Annex B.2.2.

- A conformance validator is defined that permits to validate the bitstream (and possibly the coded metadata) to be conforming to a specific operation point defined in this specification. For details refer to Annex B.2.3.

- A conforming decoder may be provided for certain operation points that provides a reference implementation to generate the adequate signals based on conforming bitstreams. The conforming decoder may provider information related to the output signal for example an MD5. For details refer to Annex B.2.4.

- Bitstreams that have been validated are provided with collected information including the information of the used video signal, the encoder and its configuration, the report from the validator as well as the coded metadata. In addition, decoding related information may be provided. The bitstreams are referred to as conformance bitstream. For details refer to Annex B.2.5.

Conformance tools are provided at https://forge.3gpp.org/rep/sa4/ts-26.265. These tools and functionalities provided in this Annex may be used for different purposes to support interoperable deployments. Examples for the usage of the tools are provided in Annex B.3.

B.2 Details of the conformance functions

B.2.1 Video Test Signals

Video test signals with proper annotation are provided at https://forge.3gpp.org/rep/sa4/ts-26.265/video-signals.

The video signals are annotated with a JSON document that follows the schema that

- includes the parameters assigned to a video signals as defined in clause 4.4 of this document
- provided at https://forge.3gpp.org/rep/sa4/ts-26.265/video-signals/annotation.json, and
- attached to this specification as `'TS26265_annotation_video_signal_schema.json'`.

B.2.2 Reference encoders and configurations

Workflows, reference encoders and configurations to generate conforming bitstreams for different operation points are provided at https://forge.3gpp.org/rep/sa4/ts-26.265/conformance/conformance-bitstream-generation.

B.2.3 Conformance Validator

A conformance validator for testing bitstreams for conformance against the operation points defined in this specification is provided https://forge.3gpp.org/rep/sa4/ts-26.265/conformance/bitstream-validator.

The validator provides a report about the level of conformance. The detailed semantics and the format of the report is for further study.

The validator also provides a detailed XML-based breakdown of the syntax and feature executed in the bitstream. The detailed semantics and the format of this document is for further study.

B.2.4 Conforming Decoders

A conforming decoder may be provided for certain operation points that provides a reference implementation to generate the adequate signals based on conforming bitstreams. The conforming decoder may provider information related to the output signal for example an MD5. Details are for further study.

B.2.5 Conformance Bitstreams

Bitstreams that have been validated are provided with collected information including the information of the used video signal, the encoder and its configuration, the report from the validator as well as the coded metadata. In addition, decoding related information may be provided. The bitstreams are referred to as conformance bitstream.

Conformance bitstreams are provided here https://forge.3gpp.org/rep/sa4/ts-26.265/conformance/conformance-bitstreams.

A schema for annotation is provided here https://forge.3gpp.org/rep/sa4/ts-26.265/conformance/conformance-bitstreams/annotation.json. Detailed semantics are for further study.

B.3 Usage examples for Conformance Functions

The conformance functions may be used for different purposes to support interoperable deployments:

- the conformance validator may be used to test bitstreams claiming to be conformant to an operation point indeed are conforming, or if not, which test assertions are not fulfilled.

- the conforming bitstreams may be used to test decoders and receivers to be able to decode and process bitstreams conforming to an operation point.

Annex C: 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
2025-05	SA4#132	S4-251137				Version agreed at SA4#132 including S4-251136 which is a merge of S4-250775, S4-250797, S4-250798, S4-250799, S4-250830, S4-250832, S4-250833, S4-250861, S4-250940	1.2.0
2025-07	SA4#133-e	S4-251522				Version agreed at SA4#133-e including S4-251380, S4-251399, S4-251514, S4-251515, S4-251516 S4-251519, S4-251520	1.3.0
2025-09	Post SA4#133-e Video SWG AHG	S4aI25 SP-2509720065				Version agreed during Post SA4#133-e Video SWG AHG September 2, 2025 adding online revisions S4aV240063 and prepared for submission to SA#107 for approval	1.4.0
2025-09	SA#109	SP-250972				Version 2.0.0 created by MCC to be sent to TSG SA for approval	2.0.0
2025-09	SA#109					Version 19.0.0 created by MCC to be published upon approval at TSG SA#109	19.0.0
2026-01	SA#110	SP-251441	0001	1	B	[VOPS] Conformance	19.1.0
2026-01	SA#110	SP-251441	0002	1	F	[VOPS] Essential Corrections	19.1.0
2026-01	SA#110	SP-251441	0003	1	F	[VOPS] Essential corrections	19.1.0
2026-03	SA#111	SP-260127	0004	1		[VOPS] Conformance Updates	19.2.0

History

Version	Date	Status
V19.0.0	October 2025	Publication
V19.1.0	February 2026	Publication
V19.2.0	April 2026	Publication