Versatile Video Coding technical guidelines

(v1.0-rc)

Media Coding Industry Forum

3855 SW 153rd Drive

Beaverton, OR 97003

https://www.mc-if.org/

1 Notice

2 Motivation and scope

Versatile Video Coding (VVC/H.266) is the state-of-the art video coding standard jointly developed by ISO/IEC JTC1/SC29 Motion Picture Experts Group (MPEG) and ITU-T Q6/16 SG16 Video Coding Experts Group (VCEG). The first version of VVC was completed in July 2020 included profiles for professional and consumer applications. VVC offers best-in-class compression performance with significant gains, in the order of 50%, over its predecessor HEVC/H.265. On top of its unparallel compression performance, VVC was designed to handle a wide variety of applications and use cases including broadcast and streaming of video with 4K and 8K resolutions, native support for High Dynamic Range (HDR) and Wide Color Gamut (WCG), efficient coding of computer graphics imagery and screen-content video, built-in scalability, and multi-layer support. VVC also offers functionality and features for emerging applications such as 360° omnidirectional video or ultra-low latency streaming.

Since the launch of the VVC standard, several application-oriented standards developing organizations and industry fora worldwide have been examining inclusion of VVC-based profiles and corresponding receiver capabilities. The technical guidelines presented in this document aim to:

- cover best practices of VVC configuration in terms of functionality and compression performance for industry relevant VVC-based profiles and interoperability points,
- provide guidance and up to date information on VVC operating bitrate ranges for selected applications,
- provide information on the usage of VVC with accompanying technologies e.g., such as Versatile Supplemental Enhancement Information (VSEI) messages,
- advocate interoperability and seek commonality of VVC usage across transport and application layer standards and relevant industry guidelines,

The initial scope of these technical guidelines document is limited to VVC usage in video broadcast and streaming applications.

3 Tables and figures

3.1 List of figures

3.2 List of tables

Contents

4 Terms and abbreviations

4.1 Terms and abbreviations

The following terms are used are used for the purpose of this document.

2K	Informal reference to an HD resolution, e.g., 1920x1080
4K	Informal reference to a UHD resolution, e.g., 3840x2160
8K	Informal reference to a UHD resolution, e.g., 7680x4320
BD-rate	Bjöntegard-Delta rate - a measure of the bitrate reduction in percentage
GOP	Group of pictures in coded video sequences (CVS) between 2 IRAP pictures
HD	High Definition
HDR	High Dynamic Range
HEVC	High Efficiency Video Coding (Rec. ITU-T H.265 | ISO/IEC 23008-2)
HLG	Hybrid Log-Gamma [14]
PQ	Perceptual Quantizer [14]
UHD	Ultra-High Definition
VVC	Versatile Video Coding [1]
WCG	Wide Color Gamut (ITU-R BT.2020 [12])

4.2 Abbreviations in VVC standard

The following abbreviations are referenced as defined in the VVC standard.

AU	Access Unit
AUD	Access Unit Delimiter
CBR	Constant Bit Rate
CPB	Coded Picture Buffer
CRA	Clean Random Access
CVS	Coded Video Sequence
DPB	Decoded Picture Buffer
DCI	Decoding Capability Information
GCI	General Constraints Information
GDR	Gradual Decoding Refresh
HRD	Hypothetical Reference Decoder
IDR	Instantaneous Decoding Refresh
IRAP	Intra Random Access Point
NAL	Network Abstraction Layer
OLS	Output Layer Set
PH	Picture Header
POC	Picture Order Count
PPS	Picture Parameter Set
RADL	Random Access Decodable Leading (picture)
RASL	Random Access Skipped Leading (picture)
SEI	Supplemental Enhancement Information
SPS	Sequence Parameter Set
VBR	Variable Bit Rate
VCL	Video Coding Layer
VPS	Video Parameter Set

5 Introduction

This document aims to serve as an informative reference on VVC configuration aspects relevant for applications in scope of this document. Provided information on VVC usage, configuration aspects and coding performance was collected from several sources referenced in the document.

MC-IF VVC technical guidelines document is organized as follows:

- Section 6 provides a general overview of VVC standard and its usage and configuration aspects including, pre-processing, encoding and post-decoding processes. This section presents VVC aspects relevant for broadcast and streaming applications from VVC technology's centric viewpoint.
- Section 7 provides more specific information on VVC usage in broadcast and streaming applications. This section reviews VVC configuration aspects from taking an-application centric viewpoint. The scope of this version of the guidelines is restricted to the final distribution stage and applications based on VVC Main 10 profile.
- Annexes provide supplementary information.

Document history:

Date <YYYY.MM.DD>	Version	Information
2023.09.15	1.0rc (current)	Version 1.0rc (release candidate) published at mc-if.org available for a community review.

THIS VERSION IS A RELEASE CANDIDATE VERSION AND IS AVAILABLE FOR COMMUNITY REVIEW. COMMENTS AND ERROR REPORTS AGAINST THE TECHNICAL GUIDELINES CAN BE SUBMITTED AT: vvc-guidelines@lists.mc-if.org

FOR MORE INFORMATION REGARDING THE GUIDELINES, SUCH AS PLANNED RELEASE DATES AND REVIEW PERIOD, PLEASE REFER TO https://www.mc-if.org/broadcast-streaming-guidelines/ /

6 VVC configuration aspects

This section provides a general overview of VVC standard and its usage and configuration aspects including, pre-processing, encoding, and post-decoding processes. Decoding processes, which are normatively specified in the VVC specification and referred application specifications, are not covered in this informative guideline. Presented are VVC aspects relevant for broadcast and streaming applications from VVC technology's centric viewpoint. The section comprises the following sections:

- **6.1 VVC overview** provides background information regarding VVC including description of VVC profiles, bitstream structure, selected tools, and features.
- **6.2 Pre-encoding processes**, **6.3 Encoding processes** and **6.4 Post-decoding processes** provide an overview of VVC configuration choices and how these relate to operational constraints and application's requirements.
- **6.5 VVC coding performance** provides a summary of VVC coding performance reported from several studies including objective performance measurements and subjective video quality assessments.

6.1 VVC overview

Versatile Video Coding (VVC) version 1 was standardized by ITU-T as Recommendation H.266 and in ISO and IEC as International Standard 23090-3 (MPEG-I Part 3) [1]. Version 1 focuses primarily on the video with 8 and 10 bits per sample in both luma and chroma components. A second version was published in 2022 covering 12- and 16-bit sample depth. VVC is the most current video coding standard jointly developed by ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). It was developed by the Joint Video Experts Team (JVET) consisting of experts from VCEG and MPEG.

MPEG and VCEG have a long history of jointly developing very successful video coding standards that have revolutionized the digital video space and its applications and are used in billions of devices ranging from TVs to tablets to cell phones to PCs. Their first joint standard known as MPEG-2/H.262 [3] was developed in early 1990s. That standard ushered in the era of digital video distribution and storage on a world-wide scale. Thereafter, the MPEG-4 AVC/H.264 [4] video coding standard, developed in the early 2000s became the primary engine behind the deployment of High-Definition TV (HDTV), internet based video transmission in applications such as YouTube, and today is most widely deployed codec serving broadcast and streaming applications. High Efficiency Video Coding (HEVC/H.265), formally known as Recommendation H.265 and International Standard 23008-2 [5], developed in 2010s, is at the core of many UHDTV/4K products and services in multiple application spaces, and is poised to surpass H.264 in deployment in the broadcast and streaming fields in the foreseeable future.

All MPEG-2, AVC, and HEVC use the same basic operation principles, which can be characterized as hybrid motion compensated transform-based codecs. The value of a reconstructed sample is calculated based on spatial or temporal neighbors (reference samples) whose spatio-temporal position is coded in the bitstream, and the inverse-transformed and dequantized prediction error. In addition, AVC, HEVC and VVC include other in-loop processing such as in-loop filtering, to remove noise in the reference pixels used for predicting the current pixels. Control information for those loop filters, as well as control information is also part of the bitstream.

At a very high-level, VVC and HEVC codec structures are comparable. The basic concepts, strategies and technologies used in pre-processing, post-processing, rate control, integration with the targeted system and delivery layers, commercial insertion etc. are also similar for these two codecs. The two codecs primarily differ in the signal processing technologies in use—which are responsible for the compression performance gain as well as increases in computational complexity and memory demands. Assuming adequate provisioning for such computational demands, therefore, VVC can relatively easily replace HEVC in an application.

VVC can be deployed with various rate control mechanisms used in different applications, such as CBR (Constant Bit Rate), VBR (Variable Bit Rate), CVBR (Capped VBR), Constant Quality/Constant Rate Factor (CRF). Encoders can be implemented either in the live/real time encoding or offline encoding modes including in faster or slower than real time, multiple passes, or look-ahead based encoding as well as low delay or random-access modes.

VVC can support SDR (Standard Dynamic Range) or HDR (High Dynamic Range) video as well as multiple Chroma resolutions such as Monochrome, 4:2:0, 4:2:2, 4:4:4.

VVC delivers a substantial increase in compression efficiency over its predecessor, HEVC/H.265, achieving around 50% bitrate reduction for the same video quality, at the expense of higher computational complexity. VVC also improves on HEVC by including coding tools designed to support a wide range of video content properties including High Dynamic Range (HDR), Wide Color Gamut (WGC), and computer-generated imagery for gaming and remote screen content sharing. VVC's state-of-the art coding tools, and rich functionality target efficient delivery of established and emerging video formats such as Ultra-High Definition (UHD) with 4K and 8K resolutions, and VR 360º video. Even version 1 includes spatial and SNR scalability support in addition to the single layer bitstream.

Many applications rely on metadata signaled in supplemental enhancement information (SEI) messages, which are specified in VVC itself and in VSEI (Versatile SEI messages for coded video bitstreams) (ITU-T Recommendation H.274 | ISO/IEC 23002-7 [2]). SEI and VUI information relevant only to the VVC syntax are included in the VVC specification, whereas SEI and VUI independent of the VVC syntax, and therefore potentially useful also for other (future) video codecs, are available in the VSEI specification.

The many improvements VVC boasts over HEVC in the source coding and signal processing are not discussed herein in detail, as their adequate implementation and configuration in both encoder and decoder implementation can be assumed in any commercially viable VVC product. Like with older video codecs, also in VVC, the optimization especially of the encoder with respect to these technologies is one of the key differentiators in the competitive landscape, and if we take experience from earlier codecs as a guideline—and there is no reason to believe that this experience shouldn't apply for VVC also—we will see significant improvements in coding efficiency of commercial VVC codec as competitors innovate to get the best performance out of VVC's design and syntax.

A comprehensive overview of VVC is available in [6].

6.1.1 VVC profiles, tiers, levels

Profiles, levels and tiers jointly specify conformance points to facilitate interoperability between implementations, devices and applications implementing VVC. From the outset, a decoder compliant to a particular profile, level and tier must be able to decode any bitstream compliant to that profile, level and tier. However, VVC includes concepts known as sub-profiles and DCI information, which can be optionally present in the bitstream. If such information is present, it can indicate that the bitstream does not exercise certain tools as specified in the sub-profile or DCI bits. Therefore, their implementation is not required for decoding the bitstream.

Profiles

VVC supports a very wide range of applications and allows a large set of coding tools that can be implemented by an encoder to produce VVC compliant bitstreams. Requiring every decoder to support all those coding tools would become onerous and expensive for many applications. Therefore, a subset of the entire bitstream syntax defined in the specification is collected in a profile targeting specific application scenarios. In the VVC specification this selection is expressed negatively—a bitstream in compliance with a given profile includes all tools except those specifically profiled out, usually expressed by forbidding values of certain syntax elements that indicate the use of a tool.

Unlike HEVC, where new functionality and specialized coding tools were added incrementally over time (e.g., 4:2:2 and 4:4:4 chroma formats were added in HEVC edition 2, scalability support in HEVC edition 3, and screen-content coding in HEVC edition 4), VVC includes support for all this functionality already from the first version. Three sets of profiles were included in the first edition of VVC to support a host of video applications:

- Main 10 and Main 10 4:4:4 profiles, where Main 10 is the expected flagship profile and intends to support a wide range of broadcast and streaming applications. Main10 includes support for 8- and 10-bit video, temporal sub-layer functionality, sub-pictures, and reference picture resampling among many other tools. Main 10 4:4:4 includes support for additional chroma sampling formats: 4:2:2 YCbCr chroma typically used in video production and contribution applications and 4:4:4 RGB format for applications operating on content out of graphics frame buffers, e.g., remote desktop.

- Multilayer Main 10 and Multilayer Main 10 4:4:4 extend Main 10 and Main 10 4:4:4 profiles, respectively, with support for multi-layer bitstreams involving spatial or SNR scalability as well as multiview.
- Main 10 Still Picture and Main 10 4:4:4 Still Picture specialize Main 10 and Main 10 4:4:4 profiles, respectively, to support bitstreams comprising a single intra-coded picture.

The second edition of the VVC standard published in April 2022 added new operation range extension profiles supporting bit depths up to 12 bits for YCbCr chroma formats, intra only profiles including up to 16 bits for RGB formats targeting high fidelity content acquisition and studio production applications:

- Main 12, Main 12 Intra, Main 12 Still Picture profiles.
- Main 12 4:4:4, Main 12 4:4:4 Intra, Main 12 4:4:4 Still Picture profiles.
- Main 16 4:4:4, Main 16 4:4:4 Intra, Main 16 4:4:4 Still Picture profiles.

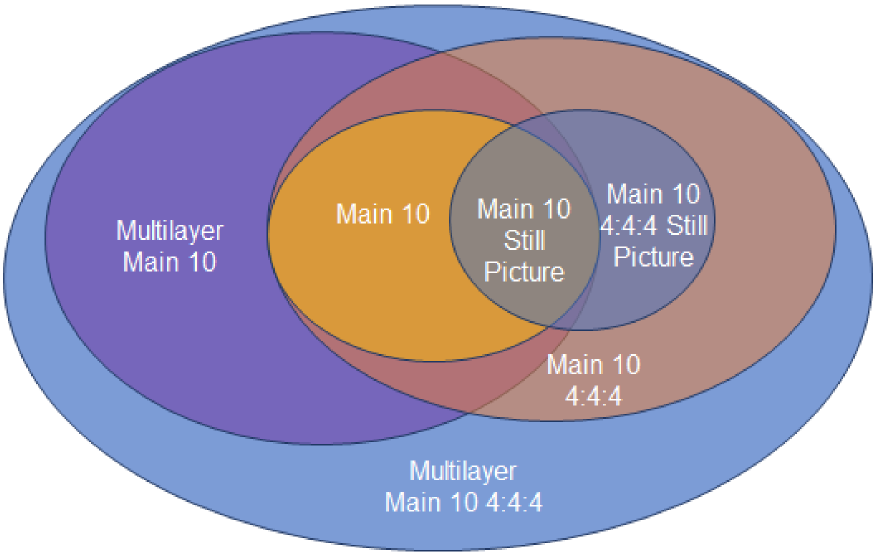

Figure 6-1 VVC version 1 profiles and their relationship (after [7]).

Levels

A level is a defined set of constraints imposed on the parameters in the bitstream that limit the maximum pixel processing rate to be supported at that level. Together with certain other constraints limiting the maximum frame rate, and geometric constraints on the picture, in effect levels limit the maximum picture size and the frame rate to be supported at that level. For example, Level 4.0 allows support for 1280x720 at 60 fps (frames per second) but does not support 4Kx2K (3840x2160) pictures at 60 fps. This removes the burden and the cost associated with implementing the decoder to be able to decode picture sizes and the frame rates higher than that are needed by the application. In the example above, a Level 4.0 compliant decoder is not required to be able to decode 4Kx2Kx60p video.

VVC has specified 14 levels covering vast range of picture sizes from SQCIF (128x96) to 12288x6480 at various picture rates. As some examples, Level 4.0 supports 1280x720x60p video, Level 4.1 supports 1920x1080x60p video and Level 5.1 supports 3840x2160x60p or 4Kx2Kx60p video. An “unlimited” level is available for such applications that handle computational/memory management aspects through technology outside the level concept.

Tiers

For many decoders implementation, the entropy decoding mechanism (known as Context-based Adaptive Binary Arithmetic Coding or CABAC) is the main, if not the only, module that needs to directly respond to the bitrate of a bitstream. CABAC is a complex technology that requires many decisions and branches, and hence is not easily parallelizable. Insofar, its efficient implementation presents a challenge for software but - even more so - for hardware implementations. A high bitrate CABAC engine is a significant cost factor.

The more cost sensitive applications, for example entertainment TV, tend to trade off quality for lower bitrate and operate and comparatively low bitrates for a given resolution. Less cost-sensitive professional applications, on the other hand, tend to use higher bitrates.

Historically, before the tier concept was introduced with HEVC, codec engines had to support entropy decoding at bitrates sufficient for cost-insensitive professional applications even in entertainment devices that would never be exposed to such high bitrate bitstreams, for example for the simple reason that the network connectivity would be insufficient. An implementer had to make the choice to either support the level as mandated in the standard and absorb the related cost or step out of the defined profile/level system.

The tier system resolves that problem, by defining a "Main" and a "High" tier. In both tiers, the maximum bitrate is derived from the maximum resolution (more precisely—pixel processing rate) defined by the level. However, the maximum bitrate is calculated by a different multiplicative factor from that pixel processing rate. If the "Main" tier is signalled, the maximum bitrate of a bitstream is what JVET thinks is appropriate for cost-sensitive entertainment devices. Even at "High" tier, the bitrate is constraint—albeit at a considerably higher number—so to ensure implementability of the decoder.

Sub-Profiles and DCI bits

Sub-Profiles and the DCI information were motivated based on the observation that HEVC and the then forthcoming draft VVC versions include many tools with limited applicability for certain applications. The traditional mechanism to tailor video codecs toolsets towards applications were profiles. Historically, that occasionally has led to the proliferation of dozens of profiles, only very few of which were in practical use. The remaining more optimized profiles were not used as there was no critical mass for chip/software development towards those profiles. As a result, some applications used codec implementations with a large functionality set of which only a small part was exercised, while other applications were crippled through the lack of hardware-supported functionality.

JVET's reaction to this situation was to radically minimize the profiles to only three families: Main, Multilayer, and Still Picture. That, however, led to very powerful profiles with corresponding implementation complexity.

Profiles defined by the video coding standards groups, such as JVET, have historically been fine-tuned by application standards bodies such as DVB, ATSC, 3GPP, and others, for the purpose of removing tools not deemed useful for that particular application. Within the domain of such groups, such informal sub-profiling worked well. However, historically, there was no way to express, in the bitstream and in a normative codepoint, such sub-profiling. VVC solves this problem by providing a normative codepoint. MC-IF offers a registry service for sub-profiles (editorial update pending).

The DCI bits can be used to signal, in a video bitstream, the absence of the use of certain coding tools. More than 50 such bits are currently defined, and an extension mechanism is provided that allows adding more bits as the VVC standard develops. Coding tools whose absence can be signaled can be as broad as any one of VVC's three loop filters, or as comparatively small as allowing only a single slice per coded picture. When a bitstream is labelled as not exercising a certain tool, then a decoder does not need to implement that tool, or if it does, then the related code would not be exercised. As a result, walled garden applications which have control over both encoder and decoder can be selective in what tool they implement and can deploy implementations not exercising a certain tool when commercial circumstances makes that advisable.

6.1.2 VVC bitstream structure

On the syntactical level, a VVC bitstream is made up of a stream of Network Abstraction Layer units (NAL units). Those can be viewed as packets of different types, containing bits associated with the type. Some of the types are related to the decoding engine, such as coded slices. Others relate to control information such as parameter sets. Still others refer to metadata. Some observation about this syntactical level is included below. However, before diving into the syntax, it's necessary to briefly review the bitstream from a source picture perspective. In this view, the bitstream consists of a series of coded pictures with associated control and metadata.

To be specific, a VVC bitstream includes one or more coded video sequences (CVSs), each of which contain one or more access units (AUs). A CVS can be decoded independently from any other CVS.

A VVC bitstream contains one or more layers. There must be at least one independent layer in the bitstream, whereas the other layers, if present, can be additional independent layers or dependent layers, which take advantage of inter-layer prediction. An additional independent layer could represent an alpha plane video or a depth map sequence, for instance, and dependent layers may be used for spatial scalability or multiview coding, for example.

Within a layer, a coded layer video sequence (CLVS) can be decoded independently from any other CLVS of the same layer. CLVS boundaries between layers are not required to be aligned except at the beginning of a CVS. A good analogue for a CLVS, originating from MPEG-2 deployments and still in use today, is a Group Of Pictures or GOP.

An AU consists of one or more picture units (PUs) of different layers. A PU contains a single coded picture and associated non video coding layer (non-VCL) data. Layers need not have the same picture rate and consequently it is not required that each AU has a PU at each layer that is present in the bitstream. If we ignore layers, an AU is similar to what a coded picture was in MPEG-2.

A bitstream can be viewed as a sequence of CVSs. Each CVS starts with an IRAP picture and associated control and metadata. The MPEG-2 equivalent would be an I-picture, the AVC equivalent would be an IDR picture and associated metadata. The IRAP picture is followed by any number of PUs (or coded pictures), each of which can be associated with metadata, which are typically predictively coded relative to the IRAP picture and each other. Insofar, a CVS, absent layering, is comparable to GOP. The prediction mechanisms within a GOP can be quite complex and some will be introduced below. The maximum length of the GOP is usually dictated by application demands, such as tune-in time, and in many cases will not exceed one second.

Armed with this picture-centric understanding, and going back to the syntax level where the VVC bitstream can be defined as a sequence of network abstraction layer (NAL) units, following paragraphs take a deeper dive into that NAL unit stream. Each NAL unit has a header and a payload, where the two-byte NAL unit header contains a NAL unit type, a layer identifier, and a temporal sublayer identifier.

NAL unit types are either non-VCL NAL units, which are summarized in **Table 6-1**, or video coding layer (VCL) NAL units, which are described in the subsequent paragraphs. VCL refers to video coding layer, which is a conceptual layer comprising information and mechanisms related to the decoding at and below the syntactical slice layer. The Network Abstraction Layer comprises all other information.

Table 6-1 Non-VCL NAL unit types

		NAL unit type abbreviation(s)	Description
Operating point information	OPI	OPI_NUT	Used to provide the highest temporal sublayer present in the CVS and/or the output layer set (OLS) represented by the CVS to the decoder.
Decoding capability information	DCI	DCI_NUT	Used to provide a set of profile-tier-level syntax structures to the decoder. Each OLS is each CVS must conform to at least one of the given profile-tier-level syntax structures.
Video parameter set	VPS	VPS_NUT	CVS-level syntax elements related to multi-layer operation
Sequence parameter set	SPS	SPS_NUT	CLVS-level syntax elements
Picture parameter set	PPS	PPS_NUT	Syntax elements common to one or more coded pictures
Adaptation parameter set	APS	PREFIX_APS_NUT SUFFIX_APS_NUT	Syntax elements controlling in-loop filtering
Picture header	PH	PH_NUT	Syntax elements common for all slices of a coded picture
AU delimiter	AUD	AUD_NUT	Used to indicate the start of an AU, whether the AU contains RAP pictures in all layers of the CVS, and the type of slices present in the AU
End of sequence	EOS	EOS_NUT	Indicates the end of CVS for the present layer and all dependent layers of the present layer
End of bitstream	EOB	EOB_NUT	Indicates the end of the bitstream
Supplemental enhancement information	SEI	PREFIX_SEI_NUT SUFFIX_SEI_NUT	See section **6.1.5**
Filler data	FD	FD_NUT	Used to control the bitrate

VCL NAL units are coded slice NAL units of different picture or subpicture types. Note that these technical guidelines only touch on the use of subpictures and hence the subsequent description of VCL NAL unit types is simplified so that subpictures are mostly excluded from consideration. Picture types can be classified as random access point (RAP) and non-RAP picture types, further described below and summarized in **Table 6-2**.

RAP pictures may be intra random access point (IRAP) pictures or gradual decoding refresh (GDR) pictures. IRAP pictures in an independent layer contain only intra-coded data, whereas an IRAP picture in a dependent layer may utilize inter-layer prediction but shall not use temporal inter prediction. Pictures that follow an IRAP picture in both decoding order and output order are correctly decodable when the decoding is started from the IRAP picture. Decoding can be started from an GDR picture, but the decoded pictures are correct in content only after an indicated period of pictures. A GDR picture may be used to avoid bitrate peaks caused by intra-coded pictures, which may be important for achieving very low end-to-end delay.

There may be up to seven temporal sublayers in the bitstream. The temporal sublayers indicate a temporal prediction hierarchy so that a picture at a particular temporal sublayer N cannot be predicted from any picture at temporal sublayer greater than N. Consequently, a pruned bitstream can be formed from an original bitstream by excluding any number of the highest sublayers.

Table 6-2 VCL NAL unit types

		NAL unit type abbreviation(s)	Description
RAP pictures			
Instantaneous decoding refresh	IDR	IDR_W_RADL IDR_N_LP	An IRAP picture starting a closed group of pictures (GOP) with RADL pictures (IDR_W_RADL) or without leading pictures (IDR_N_LP).
Clean random access	CRA	CRA_NUT	An IRAP picture starting an open GOP.
Gradual decoding refresh	GDR	GDR_NUT	A RAP picture starting a period of pictures after which all subsequent decoded pictures are correct in content.
Non-RAP pictures			
Step-wise temporal sublayer access	STSA	STSA_NUT	A picture enabling the decoder to increase the number of temporal sublayers being decoded.
Random access decodable leading	RADL	RADL_NUT	A picture that is correctly decodable when the decoding is started from the associated IRAP picture, follows the IRAP picture in decoding order, and precedes the IRAP picture in output order.
Random access skipped leading	RASL	RASL_NUT	A picture that is not correctly decodable when the decoding is started from the associated IRAP picture, follows the IRAP picture in decoding order, and precedes the IRAP picture in output order.
Trailing	TRAIL	TRAIL_NUT	A non-RAP picture that cannot be classified as STSA, RADL, or RASL picture.

Non-RAP pictures may appear in any temporal sublayer. A step-wise temporal sublayer access (STSA) picture at temporal sublayer N enables the decoder to start the decoding of temporal sublayer N if the decoder previously decoded pictures only up to but excluding temporal sublayer N. A leading picture is such a picture that precedes an associated IRAP picture in output order but follows it in decoding order. There are two types of leading pictures, differentiated by whether or not the leading picture can be correctly decoded when the decoding is started from the associated IRAP picture, and referred to as a random access decodable leading (RADL) picture or a random access skipped leading (RASL) picture, respectively. A RASL picture is predicted from one or more pictures preceding the associated IRAP picture in decoding order or from one or more other RASL pictures associated with the same IRAP picture, whereas RADL pictures may only be predicted from the associated IRAP picture or other RADL pictures associated with the same IRAP picture. When a non-RAP picture is not categorized as an STSA, RADL, or RASL picture, it is a trailing picture, which follows the associated IRAP picture in both decoding order and output order.

6.1.3 Picture partitioning

The VVC standard specifies a particular handling of the input video material. Each picture of the input video sequence is divided into one or more spatial regions, each of which may be processed sequentially or independently, depending on the region type.

VVC picture partitioning is, in large parts, inherited from HEVC. Some types of picture partitioning are defined so to meet the needs of specific applications. For example, the newly introduced subpicture is included for frame packing and AR/VR applications. Other picture partitioning concepts are included to provide certain flexibility for computational resource management and implementation, such as tiles and slices, and some are defined as basic processing units for picture coding.

Within the VCL, and below the slice layer, the equivalents of a macroblock are known as CTU, CTB and blocks. All are fundamental elements of the VVC compression. Each coded picture in processed as a sequence of one or more Coding Tree Units (CTU). Each CTU consists of a coding tree block (CTB) of luma samples and, if picture consists of three color componets, two corresponding CTBs of chroma samples. Each CTB is further partitioned into a set of blocks for intra and inter prediction and transform coding. One or more CTUs, and a slice header, form a coded slice.

A coded slice is the typical payload for a VCL NAL unit. CTUs can be accumulated into a slice in raster scan order, or in a rectangular shape. Slices are among the oldest picture partitioning mechanism and were originally proposed (in slight deviated form in H.261, and later in MPEG-1) as an error resilience and Maximum Transfer Unit (MTU) size matching mechanism.

Formally, a VVC slice is defined as an integer number of complete tiles (see below) or an integer number of consecutive complete CTU rows within a tile of a picture. By the definition, a vertical boundary of a slice is always a vertical tile boundary, however horizontal boundaries of slice and tiles do not always overlap, in particular, if a tile is split into multiple slices.

A tile is a type of supported partitioning regions was originally introduced in HEVC for purposes of parallel processing but was also found useful for 360° omnidirectional video applications. VVC tile consists of a sequence of CTUs and covers a rectangular region of a picture. The CTUs in a tile are scanned in raster scan order within that tile. Each picture may be partitioning into one or more tile rows and one or more tile columns. Each tile may be coded/decoded independently by preventing cross-tile prediction and entropy dependencies.

Finally, a VVC subpicture is a newly introduced picture partitioning, analogous to the motion-constrained tile sets (MCTSs) in HEVC. Each subpicture contains one or more slices that collectively cover a rectangular region of a picture. Depending on the selected partitioning, a subpicture and tiles are constraints with one (or both) of the following conditions: all CTUs in a subpicture belong to the same, or/and tile all CTUs in a tile belong to the same subpicture. Subpicture allow independent coding and extraction of a rectangular subset of a sequence of coded picture and intendent to facilitate ROI and viewport-dependent applications.

6.1.4 Application specific coding tools

VVC improves on HEVC by including a combination of efficient coding tools designed from the start to support a wider range of video content properties including High Dynamic Range (HDR), Wide Color Gamut (WGC), and computer-generated imagery for gaming and remote screen content sharing. VVC’s efficient coding tools and rich functionality target efficient delivery of established and emerging video formats such as Ultra-High Definition (UHD) with 4K and 8K resolutions, 360º omnidirectional video, and includes built-in scalability support already at a single layer bitstream level. While not fine-tuning its coding tools to narrow requirements of a particular application, VVC was designed on a premise that its versatility in coding performance and functionality ultimately benefits several applications. Significant advancements were introduced across all categories of coding tools including (see section II.B in [6] for more details): block partitioning, intra-picture prediction, inter-picture prediction, transforms and quantization, entropy coding and in loop-filters.

Some application specific advancements are reviewed below.

Coding tools for screen content/computer generated content

Like HEVC, VVC also includes the coding tools that are highly efficient in compressing the content generated by computers, e.g., screen content, video games, computer graphics etc. However, unlike in HEVC, in VVC those tools are part of the Main 10 profile – the profile that is expected to be very widely deployed in VVC applications. In HEVC those types of coding tools were standardized after the Main and Main 10 Profiles was standardized. So, they are included in Screen Content coding extensions profiles (e.g., Screen-Extended Main and Screen-Extended Main 10 profiles) but are not part of the Main or Main 10 Profiles. Main 10 profile is the most widely deployed HEVC profile. Hence, majority of HEVC codecs and systems cannot take advantage of those coding tools. As, in VVC those tools are part of the Main 10 profile, which is expected to be the most widely deployed profile of VVC, the VVC standard can provide significantly higher coding efficiency for a wider range of applications, systems, and implementations than HEVC.

Coding tools for immersive/360° video

Immersive video-based applications are one of the targeted applications of VVC. It includes 360° omnidirectional video where a captured scene is spherical. To be able to efficiently use 2-dimensional video coding schemes for compression, the spherical video is projected onto 2-dimensional rectangular pictures using various projection formats, e.g., equirectangular projection (ERP) or cube-map projection (CMP). VVC developed various tools, like horizontal wrap around motion compensation, bitstream extraction and merging, sub-pictures, and virtual boundaries to allow a codec to adapt to the characteristics of such video and compress it efficiently.

Coding tools for streaming video, video conferencing

In some applications, e.g., adaptive bit rate (ABR) streaming or video conferencing, the network bandwidth can change dynamically. This requires insertion of Intra Random Access Pictures (IRAPs), e.g., Instantaneous Decoding Refresh (IDR) pictures, to change the resolution of the pictures according to the variations in the network capacity. This can cause loss of coding efficiency and increased delay or discontinuity in displaying the pictures.

VVC allows the spatial resolution to be changed at the inter-coded pictures. It is done via reference picture resampling (RPR) technique. In this approach, when such a resolution change occurs, the decoding process of a picture may refer to one or more previous reference pictures that have a different spatial resolution for inter-picture prediction. RPR allows either increase or decrease in the picture resolution via upsampling or downsampling of a reference picture to predict a current picture having a different resolution. By allowing inter-picture prediction from reference pictures of different resolutions improves coding efficiency and minimizes the jumps in the bit rates associated with the insertion of IRAPs.

Spatial Scalability

VVC also supports Spatial Scalability in its Multilayer Profiles. This can allow backward compatible migration from lower spatial resolution services to high resolution services, e.g., from HDTV to UHDTV/4K or UHDTV/4K to UHDTV/8K etc. This way, while newer systems/decoders produce higher resolutions pictures, the older systems/decoders can still decode a part of the bitstreams to produce lower resolution pictures, and they do not become non-operational as one migrates to providing services with higher spatial resolution pictures.

6.1.5 VSEI specification

Many applications require certain metadata to be useful, and that metadata historically has been included in the video coding standards in the form of Supplementary Enhancement Information (SEI) messages or Video Usability Information (VUI). Observing that many SEI messages and VUI codepoints are independent of the details of the video coding specification, and historically were duplicated across specifications, JVET decided to restructure its specifications so to collect SEI and VUI information that is codec independent in a dedicated specification, and keep only those SEI/VUI information in the VVC specification that are tightly coupled to the VVC syntax and hence not applicable to other (future) standards. The result of that process was the creation of the Versatile Supplementary Enhancement Information (VSEI) standard [2].

VUI parameters in VSEI provide information for the correct display of coded video. For example, VUI parameters indicate that the coded video is progressive or interlaced; SDR, HDR HLG, or HDR PQ; that the source video has a particular aspect ratio; and any other parameters necessary for correct interpretation of the coded video.

SEI messages provide additional information that can assist decoders, displays, and other video receivers perform as desired by the content producer. For example, the display of omnidirectional 360º-video can be assisted by signalling one of the omnidirectional video specific SEI messages in the VVC bitstream. As another example, the annotated regions SEI message can be used to assist video analysis applications by signalling the size and location of objects identified in a video. SEI messages such as the mastering display colour volume (MDCV) and content light level information (CLLI) SEI messages are used extensively to indicate essential properties of HDR video.

6.2 Pre-encoding processes

Pre-encoding processes refers to how incoming video signal may be transformed to make it suitable for encoding such that its properties reflect the intended interpretation of a decoded bitstream and associated meta-data.

6.2.1 Chroma subsampling

The chroma subsampling operation is an integral processing step performed at the interface to the distribution and final emission stage typically served by 8 or 10-bit 4:2:0 codec profile, i.e., Main 10 profile for VVC. The processing aims to reduce the required bandwidth for chroma at the final content distribution stages by reducing resolution of chroma components to a quarter of the corresponding luma signal due to different perceptual impact. Chroma downsampling is performed on either 4:4:4 formats used in studio production or 4:2:2 typically used in program contribution and exchange, or for transport of uncompressed video. In VVC SPS the **sps_chroma_format_idc** syntax element in the Sequence Parameter Set (SPS is used to signal the chroma format (see **Table 7-3** and **Table 7-8**). Subsampling of chroma may result in different alignment of chroma samples in reference to luma samples. For progressive video with 4:2:0 chroma subsampling, there are six potential locations of chroma samples relative to luma sample (see [2] and [28]) which may depend on phase offsets applied during filtering and downsampling operation. It is important to signal the locations to avoid quality degradation in subsequent processing stages.

Availability of chroma location information (**vui_chroma_loc_info_present_flag**) and chroma location type of alignment (**vui_chroma_sample_loc_type_frame**, **vui_chroma_sample_loc_type_top_field**, **vui_chroma_sample_loc_type_bottom_field)** is caried by Video usability information [2]. For Narrow Colour Gamut content, such as represented by ITU-R BT.709 [11], the chroma location type should typically be equal to 0 (vertical interstitial). For Wide Colour Gamut content, such as represented by ITU-R BT.2020 [12] or ITU-R BT.2100 [14], the chroma location type should typically be equal to 2 (co-sited). Respective signaling is provided in **Table 7-8.** The chroma location type is only signaled for chroma 4:2:0 formats. The reader is referred to [28] for more details regarding chroma signalling options. Typically, chroma subsampling is performed with separable low pass FIR filters. While design of the filters is not in scope of the document, it is advised to apply caution when performing downsampling process for HDR video with Perceptual Quantizer (PQ) transfer function due to highly non-linear properties of the PQ transfer function. Potential issues and different chroma downsampling approaches are discussed in [26] (Rec. ITU-T H Suppl. 15 | ISO/IEC TR 23008-14). While the report also discusses coding practices specific to AVC and HEVC codecs, considerations for pre-encoding processes such as chroma downsampling can be applicable to VVC.

6.3 Encoding processes

Encoding processes refers to how a video bitstream can be constructed according to operational constraints.

6.3.1 Reference pictures and temporal structures within a GOP

As previous generation video coding standards, VVC supports picture reordering within a GOP i.e., change order of incoming pictures to exploit bi-directional inter-picture prediction which achieves higher compression efficiency than a uni-directional prediction. Picture reordering operations, performed in the encoder, is reversed in the decoder so that pictures are output from the decoder in the same order as the original source video. Since picture reordering requires additional buffering both encoder and decoder, it is used in applications which are not latency sensitive (such as interactive or conversational services).

Figure 6-2, **Figure 6-3** and **Figure 6-4** show examples of picture referencing structures for different number of temporal layers and reordering pictures. L0 structure does not introduce picture reordering and is often referred to as low-delay due to its relevance for low delay applications. L1 to L5 structures include reordered pictures. The presented structures are referred to as hierarchical GOP structures since reference pictures are used in a hierarchical fashion. The examples show which reference pictures (highlighted in orange) need to be maintained in CPB/DPB in order to be available for prediction of a currently processed picture (highlighted in green). In addition, VVC allows to signal such hierarchical structures with pictures assigned to different temporal layers where pictures from higher layers cannot be used as reference pictures for lower layer pictures. This property as well as the fact that non-reference B pictures (i.e., pictures in the highest temporal layer, which are not used as reference pictures) constitute a large percentage of all pictures allows decoders in applications such as low latency streaming to discard these pictures to alter decoding speed (e.g., see [54]).

For a typical self-similar content, large GOP structures offer higher compression efficiency. Bitrate savings of 3% and 8% on JVET SDR and HDR test data content respectively were reported with GOP-32 over the GOP-16 structure [42]. The trade-off is a higher reordering delay which impacts random access and zapping (channel change) time (see section **7.4** for more details). That said, reordering delay is only one of contributing factors to glass-to-glass delay when considering aspects such as IRAP frequency (necessary for fast tune-in or fast random access within/into existing sessions), and the size of IRAP picture related to constant bitrate transmission channel capacity.

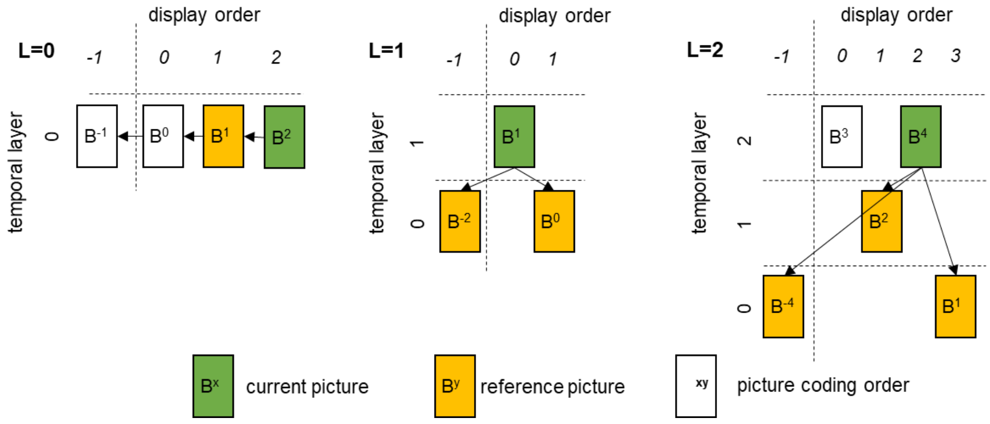

Figure 6-2 GOP structure without picture reordering (left) and hierarchical GOP structures GOP-1 (middle) and GOP-4 (right). Pictures in temporal layer 0 can be of type I, P or B.

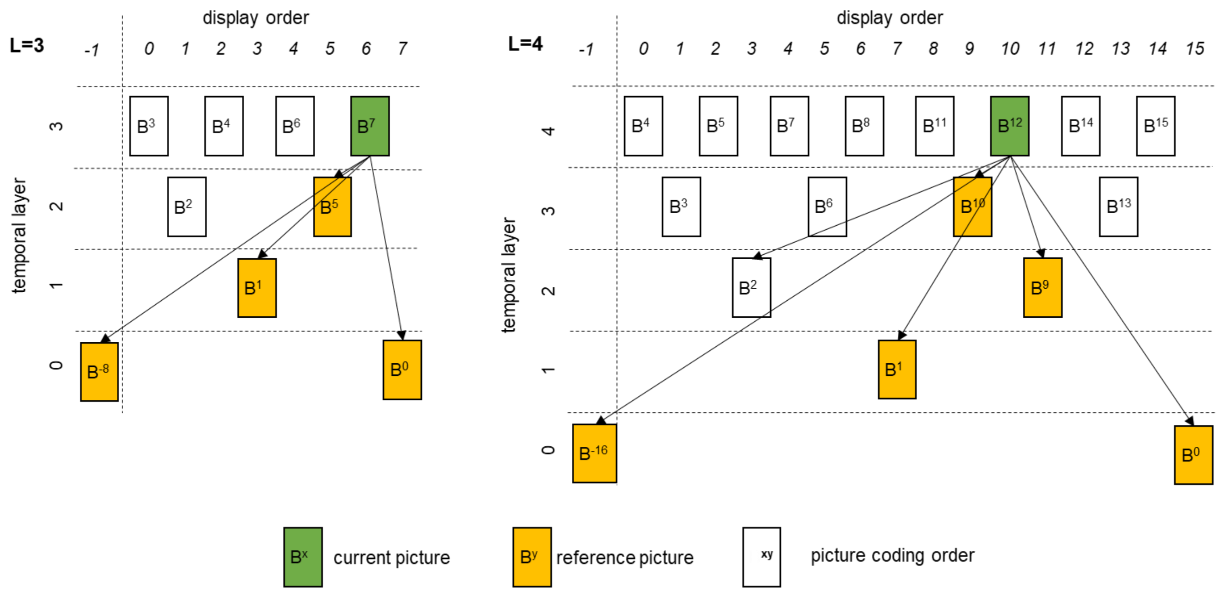

Figure 6-3 Hierarchical GOP structures GOP-8 (left) and GOP-16 (right). Pictures in temporal layer 0 can be of type I, P or B.

Large hierarchical GOP structures also require several reference pictures to be used in order to enable a complete picture referencing structure. This is illustrated in **Figure 6-2**, **Figure 6-3** and **Figure 6-4** with highlighted reference pictures (in orange). In VVC the limit of number of reference pictures that can be stored in DPB (**MaxDpbSize**) was increased compared with HEVC and is set to 8. This enables the use of GOP-32 structure such as illustrated in **Figure 6-4**.

In the example presented, there are six reference pictures which available for prediction (not all have to be actively used) when picture B^{13} is coded. Since DBP limit includes currently decoded pictures, this results in 7 pictures that need to be stored in DPB.

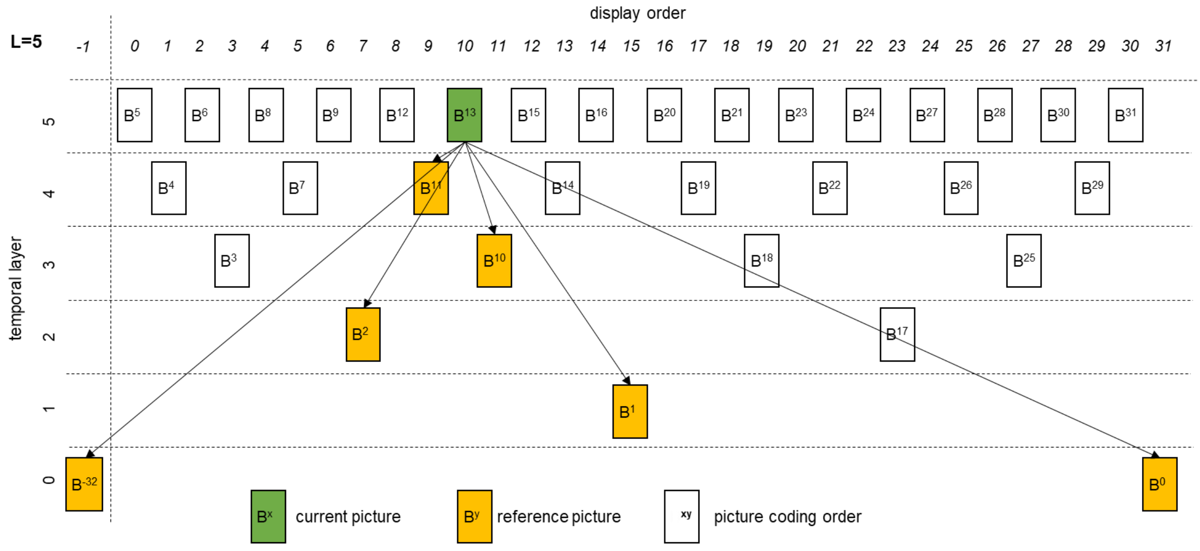

Figure 6-4 Hierarchical GOP structure – GOP-32. Pictures in temporal layer 0 can be of type I, P or B.

6.3.1.1 Additional coding optimizations

Pictures in higher temporal layers are more efficiently coded due to availability of several reference pictures from lower temporal layers, their comparatively closer distance (in the time domain) to these reference pictures, and the resulting larger temporal correlation. Because pictures at lower temporal layers are used as references, their quality impacts the quality of inter-prediction for pictures at higher temporal layers. Therefore, a varying level of quantization error for coded residual may be introduced across temporal layers to balance quality and bitrate of pictures across the hierarchy of a GOP structure. Typically, this would mean that pictures in lower temporal layers are coded with lower average QP (Quantization parameter) values compared with pictures at higher temporal layers. Such a varying QP mechanism can also be applied in cases where there are only few or even a single temporal layer, e.g., such as in low delay configuration (L=0, see **Figure 6-2**). However, in low-delay applications, large variations in picture quality/bitrate may be difficult to accommodate if rate control runs with a constrained buffer.

Another coding optimization, affecting both objective compression performance and subjective quality, is to align a denoising process to the applied GOP structure. Camera captured video content typically contains noise which increases bitrate needed to compress pictures at a certain quality. The application of hierarchical GOP coding structure may result in varying level of noise being retained across pictures at different temporal layers during encoding process, e.g., some noise may be left in lower temporal layer pictures while being completely removed in higher layer temporal pictures. Such an effect may cause visible artefact such as picture pulsing effect. A denoising filter aligned with GOP structure works on a premise that reference pictures which retain residual noise after encoding/decoding do not provide accurate inter picture prediction to other noisy pictures. Therefore, the filter is applied differently across temporal layers such that only pictures which are used as reference pictures are filtered. On the other hand, pictures seldomly referenced or never referenced to predict other pictures (in high temporal layers) are not filtered.

This approach was reported to reduce the bitrate overhead for encoding noise as part of high-quality encodings and provide objective coding gain of approximately 5% (measured as BD-rate PSNR) as well as subjective benefit [41]. An implementation of this method was included in the VTM reference software and is referred to in section **6.5** as MCTF.

6.3.2 Open and closed GOP configurations

Broadcast and streaming applications require regular and frequent access to the video bitstream. Such access point pictures are referred to as Random Access Point (RAP) pictures. VVC provides a few types of RAP pictures which have different properties and therefore a different operational impact. These picture types are discussed in section **6.1.2** summarized in **Table 6-2**. RAP picture and non-RAP pictures that follow the RAP picture in coding order until the next RAP picture occurs are typically referred to as Group of pictures (GOP).

Non-RAP pictures which are coded between RAP pictures are typically organized in a bi-predictive hierarchical structures in order to achieve high compression performance. While such bi-predictive hierarchical structures incur processing delay at both encoder and decoder due to picture reordering, it is typically within the latency budget for broadcast and streaming applications. The hierarchical structures are very effective at exploiting temporal redundancy due to use of several reference pictures.

The type of RAP picture used affects how non-RAP pictures can reference other pictures in the same GOP as well as across GOPs. Therefore, GOPs can be either referred to as “closed” or “open”. In a closed GOP picture referencing is restricted to be within the GOP. In an open GOP picture referencing to the previous GOP is also allowed for some pictures. Further, there are two types of closed GOP depending on the type of IDR picture type. **Figure 6-5** illustrates closed GOP for IDR type (IDR_N_LP) and **Figure 6-6** illustrates closed GOP for IDR type (IDR_W_RADL).

Both closed GOP structures maintain decoding independency of pictures in the GOP which has important functional role in applications such for switching segments (resolution switching in adaptive bitrate streaming systems) or for splice points. The key difference is location of the IRAP picture in display order, for closed GOP type 2, IRAP picture is preceded by RADL pictures in display order. This on one hand impacts coding efficiency of RADL pictures since some of them (see pictures with display order 0 and 1 in **Figure 6-5**) cannot benefit form bi-predictive referencing using future and past references. On the other hand, closed GOP type 1 has extra Tid=0 frame and in addition there are two high quality (Tid=0) pictures at the GOP boundary which may produce a visible “pulsing” effect.

Open GOP structure, illustrated in **Figure 6-7,** is enabled with CRA_NUT IRAP picture. The coding structure is similar to closed GOP type 2 with the difference that it allows RASL pictures (display order 0, 1 and 2), preceding the CRA picture in display order but follow it in decoding order, which may reference pictures from the previous GOP. This way a receiver does not flush its decoder picture buffer at the CRA picture allowing references previous to the CRA to be used by RASL pictures. In case a receiver random accesses a bitstream from a CRA, i.e., it starts decoding bitstream at a CRA picture it may not display RASL pictures as they may not be decodable due to missing reference pictures, hence skippable.

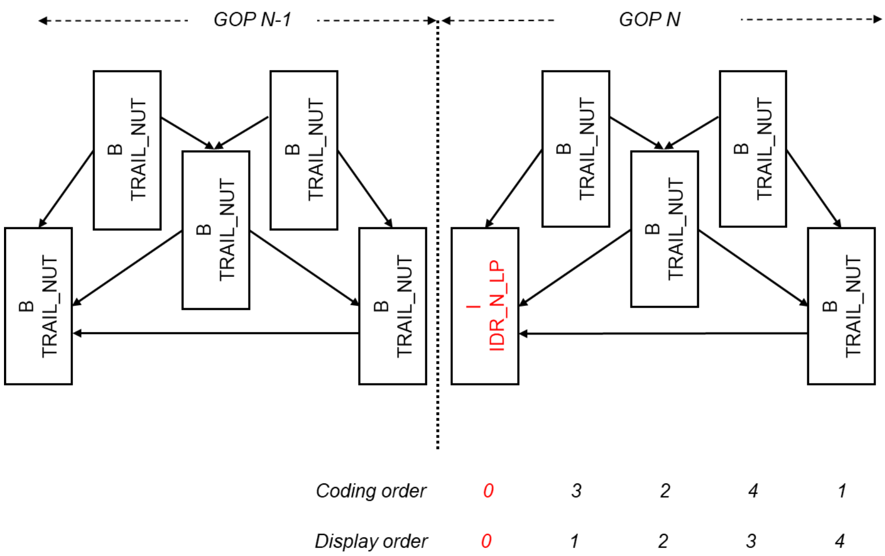

Figure 6-5 Closed GOP structure - type 1 with IDR_N_LP picture.

In terms of coding performance, there is a substantial performance penalty from closed GOP compared to open GOP. This is especially the case for short RAP periods (e.g., 1 sec) which is typically used in broadcast or for short segments used in low delay streaming cases (e.g., 2 sec). The left part of **Table 6-3** reports coding performance gain obtained with open GOP configuration compared with closed GOP type 2 (after [64] Table I). The right part of **Table 6-3** reports results obtained with constrained open GOP coding described in more detail in section **7.6.2**.

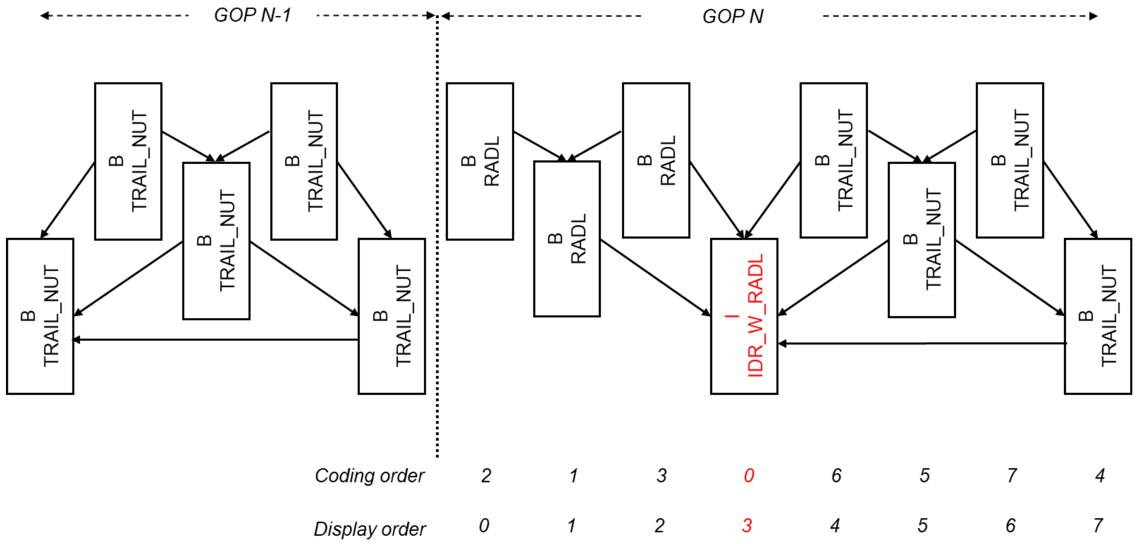

Figure 6-6 Closed GOP structure – type 2 with IDR_W_RADL picture.

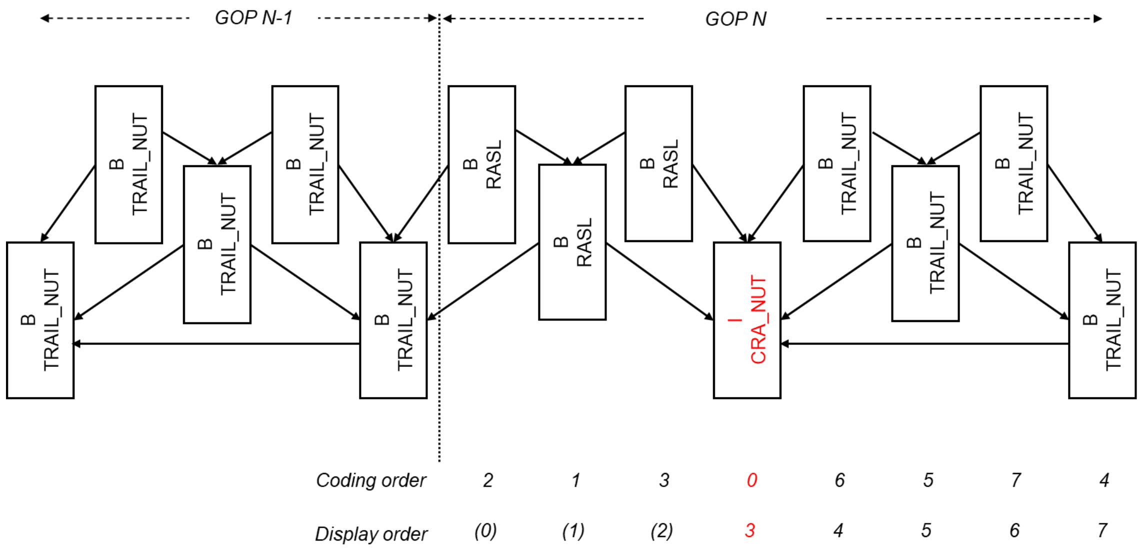

Figure 6-7 Open GOP structure with CRA_NUT and RASL pictures.

Table 6-3 Bitrate saving (BD-rate PSNR) of open GOP compared with closed GOP type 2 (after [64] Table I).

GOP structure	Open GOP			Open GOP with constrained RASL		
	GOP length			GOP length		
	~ 1 sec	~ 2 sec	~ 4 sec	~ 1 sec	~ 2 sec	~ 4 sec
GOP-16	-5.2%	-2.5%	-1.3%	-4.7%	-2.1%	-0.9%
GOP-32	-9.2%	-4.5%	-2.4%	-8.6%	-4.0%	-2.0%

6.3.3 Reference picture resampling

VVC includes a built-in reference picture resampling (RPR) functionality which allows changing resolution at any inter-coded picture. In contrast to spatial scalability functionality in SHVC (scalable extensions of HEVC), VVC reference picture resampling is performed on a block level at the same processing stage as motion compensation. At the motion compensation stage, a scaling ratio, derived based on signalled picture width, height, and scaling offsets (for currently decoded and a respective reference picture), is used together with motion prediction information.

Resolution change using RPR can be applied to any picture. The scaling ratios are restricted so that downsampling ratio is limited to 2x while upsampling ratio is limited to 8x. The VVC specification puts constraints on the use of low-level coding tools when RPR is used. Disallowed are decoder motion vector refinement (DMVR), bi-directional optical flow (BDOF) and prediction refinement with optical flow (PROF), (Sub-Block) Temporal Motion Vector Prediction (SB-TMVP) and reference picture wraparound, i.e., horizontal wrap-around motion compensation.

This functionality is supported in the Main 10 profile. Multilayer profiles use RPR to enable traditional spatial enhancement layers through efficient high level signaling without requiring additional inter-layer signal processing [6]. The use of the RPR feature is further discussed in section **7.6**.

6.3.4 Encoding HDR content

HDR video is natively supported in VVC. All VVC coding tools are applicable to both SDR and HDR video. The distinction between SDR and HDR is made only in high level syntax. Specifically, the colour primaries and optical-electronic transfer characteristics functions that distinguish SDR, hybrid-log gamma HDR (HDR HLG), and perceptual quantizer HDR (HDR PQ) are signaled in the VVC bitstream using video usability information (VUI) metadata specified in VSEI.

When encoding HDR video, the VUI parameter values indicated in **Table 7-8** for HDR PQ or HDR HLG should be signalled to the decoder.

Two new coding tools in VVC were initially introduced to optimize encoding of HDR video. Those coding tools are luma mapping with chroma scaling (LMCS) and luma-adaptive deblocking filtering.

The luma mapping component of LMCS can be used during encoding to redistribute luma values to make fuller use of the entire range of allowed values to improve compression efficiency. During decoding, the luma values are remapped to the original range. The chroma scaling component of LMCS compensates for the interactions between the luma and chroma signal, particularly for WCG video.

The luma-adaptive deblocking filter can be used to reduce the visibility of block distortions. The relationship between the strength of the filter and the local luma values are determined by parameters signalled in the SPS. Typically, stronger deblocking should be applied to areas of higher luminance than to areas of lower luminance.

Example encoding algorithms to derive LMCS and luma adaptive deblocking filter parameters are implemented in the VVC reference software (VTM) [65] with specific optimizations for SDR, HDR HLG, and HDR PQ video data. For more details see [66] (sections 3.7.2 and 3.7.3).

In general, different objective quality metrics are applicable to SDR and HDR video. HDR-sensitive objective quality metrics that could aid in optimization of HDR encoders are described in section **6.5.1.1** and in more detail in [41].

As companions to the coding tools specified in VVC, several HDR-related SEI messages are specified in VSEI. The most widely used are the mastering display colour volume (MDCV) SEI message and the content light level information (CLLI) SEI message (see section **6.4.1** for more detail). MDCV and CLLI SEI messages convey static HDR metadata to enable video to be optimized for the receiving display system.

6.3.5 Encoding 8-bit content in 10-bit container

VVC does not have an 8-bit only profile as was the case with AVC and HEVC. Instead, all VVC profiles are at least 10-bit capable while supporting coding with 8-bit bit depths. While there is a substantial legacy of 8-bit video distribution mainly through deployed AVC Main and High profiles, and HEVC Main profile, video production and contribution applications have been utilizing 10-bit video coding already with AVC High 4:2:2 profile. Several benefits of coding with 10-bit depth, both in terms of objective metrics gains and subjective quality improvements, were demonstrated for SDR content, ranging from legacy interlaced formats to UHD HFR [40][41]. Importantly, the 10-bit coding is required for compression of UHD and HDR video content (see [32]).

Although VVC does not have 8-bit only profile, it can code 8-bit content with 8-bit depth precision. If such a functionality is required, it can be signaled through **sps_bitdepth_minus8** syntax parameter (see **Table 7-3** and

Table 7-8). Additional constraint flag can be signaled through General Constraint Information **gci_sixteen_minus_max_bitdepth_constraint_idc** syntax parameter.

However, even if input content is in 8-bit format, it is likely that coding it with 10-bit precision would be more beneficial due to better compression efficiency. This phenomenon was studied for previous generations of codecs which supported 10-bit coding precision such as AVC (e.g., AVC High 4:2:2 profile [40]) and HEVC (HEVC Main 10 profile [41]). Comparison of 10-bit coding with 8-bit coding of 8-bit content generally shows an increase in objective metrics as well as several subjective benefits, especially for plain surfaces and shallow textures areas (e.g., decreased banding and contouring artefacts).

Since VVC has more advanced coding tools than AVC or HEVC, such benefits needed to be re-evaluated. JVET experts conducted a study comparing 8-bit and 10-bit coding on 8-bit content originating from different genres including camera-captured and computer-generated imagery content [44]. The study showed that VVC retains the benefits of 10-bit coding seen for previous generation codecs. Objective BD-rate gains reported for PSNR and VMAF metrics were respectively 2.5% and 3.3% for a random-access GOP structure and 4.6% and 4.8% for a low-delay GOP structure. A subjective comparison was also conducted with expert viewers. It was reported that the difference between 10-bit and 8-bit was only slightly visible, 10-bit processing showed slightly better or no worse visual quality than internal 8-bit for 8-bit input contents. In particular, the benefits of 10-bit internal processing over 8-bit internal processing were more visible at low bitrates, on computer generated imagery content or in a low-delay configuration. None of the viewers preferred the 8-bit internal processing.

Use of 10-bit coding is generally beneficial since it allows codec to utilize higher bit-depth precision for internal processing, which results in increased fidelity even if incoming video is scaled to 10-bit with a simple left-shift operation. Objective coding gain is consistent and of the order provided by some coding tools and subjective benefits can also be observed in low-textured areas.

Some low complexity encoder architectures may have a benefit of reducing computational complexity by using 8-bit only compression, but it is recommended that the benefits of such a trade-off are carefully studied.

6.4 Post-decoding processes

Post-decoding processing refers to how the output of a conforming video decoder is intended to be altered before being displayed to a viewer (or ingested by some downstream analyzer or other receiver).

Post-decoding processes are controlled by metadata signaled by VUI and SEI message parameters. Typically, VUI parameters provide information for the correct display of coded video, i.e., VUI parameters indicate attributes of video such as colour primaries, matrix coefficients, and transfer characteristics. SEI messages typically provide either additional descriptive information about coded video or parameters used as inputs to post-decode processes that alter the decoded video in a useful way.

6.4.1 High Dynamic Range SEI messages

Several SEI messages are particularly relevant for display of HDR video and for display of SDR video on modern high-luminance consumer displays. These SEI messages convey information that can be used by a receiver to optimally adapt video content to different consumer displays and viewing environments.

6.4.1.1 Mastering display colour volume (MDCV) SEI message

The MDCV SEI message provides information about how source video looked to creative professionals when the content was created (i.e., mastered). Specifically, the MDCV SEI message identifies the colour primaries, white point, and minimum and maximum luminance of the display used by the creative professional. The metadata carried in the MDCV SEI message corresponds to metadata specified in SMPTE 2086 [67]. Consumer displays can use the information in the MDCV SEI message to adjust how coded video is displayed so that it matches the source look as closely as possible.

6.4.1.2 Content light level information (CLLI) SEI message

The CLLI SEI message indicates how bright the source video was when mastered. Specifically, the CLLI SEI message identifies the maximum light level and average light level, in cd/m^2, over the entire source content represented in 4:4:4 red, green, and blue channels. Consumer displays can use the information in the CLLI SEI message to adapt coded video to optimize displayed video in terms of both visual quality and energy consumption [68].

6.4.1.3 Content colour volume (CCV) SEI message

The CCV SEI message provides information about the characteristics of video content when it was mastered. The CCV SEI indicates colour primaries and the minimum, maximum, and average luminance of the nominal colour volume of the mastered content. Consumer displays could use the information in the CCV SEI message to adapt coded video to optimize displayed video.

6.4.1.4 Ambient viewing environment (AVE) SEI message

The AVE SEI message indicates the viewing conditions assumed when the source content was mastered. Specifically, the AVE SEI message conveys the ambient illuminance, in lux, and the chromaticity coordinates of the assumed viewing environment. When the actual ambient viewing environment is different from the assumed ambient viewing environment, consumer displays could use the information in the AVE SEI message to adapt decoded video so that it is displayed in a way that matches as closely as possible to how it would appear in the assumed viewing environment.

6.4.1.5 Alternative transfer characteristics (ATC) SEI message

Typically, the ATC SEI message is used to support backwards-compatible display of HLG HDR video on SDR televisions. For example, source HLG content can be coded in a bitstream carrying VUI metadata indicating SDR transfer characteristics to support SDR displays. The ATC SEI message can be used so that the same coded bitstream is displayed correctly on an HLG HDR television, i.e., an HLG-capable consumer display can use HLG transfer characteristics conveyed by the ATC SEI message to override the SDR transfer characteristics conveyed in VUI metadata.

6.4.1.6 Dynamic HDR metadata carried in Rec. ITU-T T.35 SEI messages

Dynamic HDR metadata refers to metadata that could change during a video sequence. In contrast, the MDCV, CLLI, CCV, AVE, and ACT SEI messages described above provide static HDR metadata that are constant for an entire video sequence.

Neither dynamic HDR metadata nor dynamic HDR mapping are explicitly specified in VVC or VSEI. However, VSEI does specify a "user data registered by Rec. ITU-T T.35 SEI messages” that can be used to carry dynamic HDR metadata and mapping information. Examples of the specification and use of user data registered by Rec. ITU-T T.35 SEI messages to convey HDR metadata for use with VVC can be found in ETSI TS 101 154 [69] for metadata specified in SMPTE ST 2094-10 [70], SMPTE ST 2094-40 [71], and SL-HDR2 (ETSI TS 103 433-2) [72].

6.4.2 Film grain synthesis

Film grain synthesis refers to the addition of random visual spatial temporal texture to decoded pictures. There are several motivations for use of film grain synthesis. The original and still relevant use case was to reduce bitrate needed for cinematic content that has visible film grain in the source content. Such cinematic content can be denoised by the encoder to improve compression efficiency and an approximation of the removed noise can be reapplied after decoding to restore the visual film grain look [73]. A more recently emerging use case is to add visual texture to mask artefacts such as contours or increase apparent sharpness [74][75][76] of video softened by compression or by resizing in adaptive streaming.

VSEI specifies a film grain characteristics (FGC) SEI message that conveys content-adaptive parameters determined by the encoder to a film grain synthesis model at the decoder. The FGC SEI message in VSEI is functionally the same as the FGC SEI messages specified in the HEVC and AVC specifications.

The FGC SEI message supports 2 types of luma-dependent film grain synthesis processes: one is based on frequency filtered random noise and the other is based on auto-regressive filtered random noise. Both types of process can produce a similar effect.

The luma-related metadata signalled in the FGC SEI message enables the strength and sharpness of the film grain effect to be sculpted differently in different intensity intervals of the decoded video. Luma-based control of the film grain effect can be particularly useful in simulating actual source film grain which inherently has different statistics for different intensity intervals.

More details on use of the FGC SEI message and film grain synthesis are provided in the draft ISO/IEC Technical report on film grain synthesis technology for video applications [77]. Example software implementations for denoising source content, determination of film grain parameters to be signalled in the FGC SEI message, and film grain synthesis processes are available in the VVC reference software VTM [65]. The VTM reference software also includes a facility to edit VVC bitstreams to modify FGC SEI message parameters quickly without re-encoding. Additional open-source software for film grain synthesis is available at [78] and [79].

6.5 VVC coding performance

This section provides objective and subjective performance data obtained through formal testing in the context of the codec characterization in MPEG.

Objective compression performance tests use a variety of computer-executed metrics that compare the reconstructed samples with the source samples. While quick, automatizable, and therefore cost efficient, objective performance metrics have the disadvantage that their results and averaged human perception are not perfectly correlated. Older metrics such as Peak Signal to Noise Ratio (PSNR) are outright inadequate in this regard, whereas more modern metrics may offer a somewhat better performance. However, it’s industry consensus at this time that properly conducted subjective studies offer more reliable results than any objective metric.

Subjective video testing, when conducted properly, can be highly reliable. However, the requirements of having a control test setup (a roomful of equipment), and many human test subjects, ideally from different cultural backgrounds involved for hours, makes them expensive and time consuming for routine testing. Herein, reported are both objective and subjective results.

6.5.1 Objective coding performance

Compression performance can be efficiently estimated with the use of objective measurements of quality degradation obtained in decoded video signal compared with original source content. There are several distortion metrics available to perform such objective measurements. The key advantages of objective measurements are their relatively low computational cost and reproducibility of results (as compared with subjective assessment of picture quality - see **6.5.2**). On the other hand, objective metrics may not always respond well to perceptual optimizations in encoders and may not correlate well with subjective quality as experienced by end users. However, metrics such as MS-SSIM or VMAF are claimed to provide much closer correlation with subjective quality compared with metrics such as PSNR.

This section provides an overview of selected VVC performance tests. Reader can find many more performance tests based on objective metrics reported in literature. Tests reported below used VVC and HEVC reference test model software, VTM and HM respectively. These reference test models are well aligned with each other and only include limited encoder optimizations, used otherwise in commercial encoder offerings. This means that performance gains demonstrated between the two test models are expected to indicate the underlying coding gain offered by VVC over HEVC rather than performance gains that could stem from different levels of encoder optimizations.

6.5.1.1 Objective performance metrics and BD-rate performance

Results are reported in following clauses are expressed in different quality metrics which are summarized in this section. For SDR video, PSNR and PSNR-Y (PSNR calculated on luma component only) metrics, documented in [61], are reported. VMAF (Video Multimethod Assessment Fusion) metric is documented in [62]. Objective quality metrics used during HEVC and VVC development for HDR video data have been documented in [61]. It was determined that calculating the PSNR for the luma and chroma codewords, as it is typically done for SDR video, is still appropriate for some HDR HLG video formats. For HDR PQ video, additional metrics were determined to be more suitable. These metrics include: PSNRL100, wPSNR and DE100, as described below and in more detail in [63]:

- PSNRL100: This metric is calculated using the luminance values rather than the luma codewords. It is based on the CIELab colour space representation of the input and output sample values of the codec.
- wPSNR: This is a PSNR-like metric calculated from the codewords that attempts to compensate for the more significant distribution of luma codewords to the darker regions by performing a weighting of the codewords before calculating the PSNR.
- DE100: This is a metric based on the CIELab colour space representation of the input and output sample values of the codec. It is specifically targeted at chrominance fidelity.

Coding performance gains can be estimated with Bjöntegard-Delta (BD)-rate (see [61] for more details). BD-rate interpolates (bit)rate-distortion measurements calculated for each operating point to obtain rate-distortion curves. Performance gain is measured as an area between the two R-D curves obtained for tested codecs. The BD-rate gain (expressed as a negative number) corresponds to a bitrate savings obtained at the same quality level as expressed by the metric used in the test.

6.5.1.2 UHDTV 4K

Table 6-4 includes reported VVC performance results for UHDTV 4K content obtained on the following UHDTV 4K video data sets:

- JVET Common Test Conditions test set [36].
- MPEG VVC verification test sets: UHDTV 4K SDR [35] and UHDTV 4K HDR [33].
- 3GPP SA4 5G Video Codec Characteristics study test set [23].

Table 6-4 Selected VVC performance results based on objective measurements for UHDTV 4K (VVC Main 10 Profile, MT/L5.1)

Source	Test details	Performance gains BD-rate [metric]	VVC software [release date]	HEVC software [release date]	VVC configuration details
[36]	2160p BT.709 fps: 30, 50, 60	- 42% [PSNR-Y]	VTM-20.0 [04/23]	HM-17.0 [01/23]	IRAP=CRA GOP length ≈ 1 sec GOP structure=GOP-32 DPB size (pictures) = 8 MCTF=On
	2160p BT.2100 PQ fps: 24, 25, 50	- 37% [wPSNR-Y] - 40% [DE100] - 38% [PSNR-L100]			
	2160p BT.2100 HLG fps: 60	- 32% [PSNR]			
[35]	2160p BT.709 fps: 30, 60	- 36% [PSNR]	VTM-10.0 [08/20]	HM-16.22 [07/20]	IRAP=CRA GOP length ≈ 1 sec GOP structure=GOP-32 DPB size (pictures) = 8 MCTF=On
[33]	2160p BT.2100 PQ fps: 60	-35% [PSNR-Y]	VTM-12.0 [02/21]	HM-16.23 [03/21]	IRAP=CRA GOP length ≈ 1 sec GOP structure=GOP-32 DPB size (pictures) =8 MCTF=On
	2160p BT.2100 HLG fps: 60	- 31% [PSNR-Y]			
[23]	2160p BT.709 fps: 23.98, 50, 59.94, 60	- 33% [PSNR-Y] - 38% [VMAF]	VTM-16.0 [03/22]	HM-16.24 [10/21]	IRAP=CRA GOP length ≈ 1 sec GOP structure=GOP-32 DPB size (pictures) = 8 MCTF=On
	2160p BT.2100 PQ fps: 24, 59.94, 60	-33% [wPSNR-Y] -56% [DE100] -35% [PSNR-L100]			

6.5.1.3 UHDTV 8K

Table 6-5 includes reported VVC performance results for UHDTV 8K content on the following 8K video data sets:

- The Explorers™ 8K data set [80] (results reported in [37]).
- ITE UHDTV 8K HDR data set [81] (results reported in [38] and [39]).
- Fraunhofer HHI 8K Berlin Test Sequences [82] (results reported in [38]).

Table 6-5 Selected VVC performance results based on objective measurements for UHD 8K (VVC Main 10 Profile, MT/L6.1)

Source	Test content	Performance gains BD-rate [metric]	VVC software [release date]	HEVC software [release date]	VVC configuration details
[37]	4320p BT.2100 PQ fps:50	-33% [PSNR-Y]	VTM-8.0 [02/20]	HM-16.20 [09/18]	N/A
[38]	4320p BT.2020 4320p BT.2100 HLG fps: 60	-31% [PSNR-Y] -35% [VMAF]	VTM-11.0 [11/20]	HM-16.20 [09/18]	IRAP=CRA GOP length ≈ 1 sec GOP structure=GOP-16
[39]	4320p BT.2100 HLG fps: 59.94	-28% [PSNR-Y]	VTM-6.1 [10/19]	HM-16.20 [09/18]	IRAP=CRA GOP length ≈ 0.5 sec GOP structure=GOP-16

6.5.1.4 UHDTV 2K and sub-HD resolutions

Table 6-6 includes reported VVC performance results for UHDTV 2K content on the following data sets:

- JVET Common Test Conditions test set [36].
- 3GPP SA4 5G Video Codec Characteristics study test set [23].
- MPEG VVC verification test set: HDTV SDR [35].

Table 6-6 Selected VVC performance results based on objective measurements for HD and sub-HD resolution content (VVC Main 10 Profile, MT/L4.1)

Source	Test content	Performance gains BD-rate [metric]	VVC software [release date]	HEVC software [release date]	VVC configuration details
[36]	1080p BT.709 fps: 50, 60	-37% [PSNR-Y]	VTM-20.0 [04/23]	HM-17.0 [01/23]	IRAP=CRA GOP length ≈ 1 sec GOP structure=GOP-32 DPB size (pictures) = 8 MCTF=On
	480p BT.709 fps: 30, 50, 60	-34% [PSNR-Y]			
	240p BT.709 SDR fps: 30, 50, 60	-32% [PSNR-Y]			
[23]	1080p BT.709 fps: 23.98, 50, 59.94, 60	-31% [PSNR-Y] -36% [VMAF]	VTM-16.0 [03/22]	HM-16.24 [10/21]	IRAP=CRA GOP length ≈ 1 sec GOP structure=GOP-32 DPB size (pictures) = 8 MCTF=On
	1080p BT.2100 PQ fps: 24, 59.94, 60	-32% [wPSNR-Y] -58% [DE100] -34% [PSNRL100]			
[35]	1080p BT.709 fps: 60	-38% [PSNY-YUV] -37% [MS-SSIM]	VTM-11.0 [11/20]	HM-16.22 [07/20]	IRAP=CRA GOP length ≈ 1 sec GOP structure=GOP-32 DPB size (pictures) = 8 MCTF=On

6.5.2 Subjective coding performance

Objective metrics are very efficient and an effective tool to compare video coding technologies. However, it is also well known that all objective metrics suffer from limitations in terms their ability to predict video quality as experienced by end-users. Therefore, subjective assessment tests of compression performance offered by new coding technologies may of great value to the industry as they help benchmark perceived quality delivered with new codecs in relation to legacy technologies. MPEG Visual Quality Assessment Advisory Group (AG5) is responsible for conducting such verification tests for MPEG codecs. Guidelines for conducting formal verification tests, based on established industry practices [9][10], are maintained by the MPEG AG5 WG and are available in [60].

MPEG conducted VVC verification in several application categories: UHDTV 4K SDR video [23], UHDTV 4K HDR video [24], HDTV 2K SDR video, including low latency video with conversational and gaming content [35], and 360° omnidirectional video [25]. Tests details and final reported results are summarized in **Table 6-7**. In addition to MPEG verification tests, results from VVC subjective assessment test performed on UHD 8K content reported in [28] are included.

Table 6-7 VVC Main 10 performance gains over HEVC Main 10 based on subjective assessment tests.

Source	Test content	Performance gains BD-rate [MOS]	VVC software [release date]	HEVC software [release date]	Use case	VVC configuration details
[38]	4320p BT.2020 4320p BT.2100 HLG fps: 60	-41%	VTM-11.0 [11/20]	HM-16.20 [09/18]	Broadcast and streaming	IRAP=CRA GOP length ≈ 1 sec GOP structure=GOP-16
[35]	8192 x 4096p 6144 x 3072p 4320 x 2160p	-50%	VTM-11.0 [11/20]	HM-16.22 [07/20]	Immersive 360° video	Equirectangular projection
		-56%				Cube-map projection
[35]	2160p BT.709 fps: 30, 60	-43%	VTM-10.0 [08/20]	HM-16.22 [07/20]	Broadcast and streaming	IRAP=CRA GOP length ≈ 1 sec GOP structure=GOP-32 DPB size (pictures) = 8 MCTF=On
		-49%	VVenC-0.1.0 [09/20]			preset=medium perceptual AQP=ON rate control=OFF
[33]	2160p BT.2100 PQ fps: 60	-49%	VTM-12.0 [02/21]	HM-16.23 [03/21]	Broadcast and streaming	IRAP=CRA GOP length ≈ 1 sec GOP structure=GOP-32 DPB size (pictures) =8 MCTF=On
	2160p BT.2100 HLG fps: 60	-52%				
[35]	1080p BT.709 fps: 60	-49%	VTM-11.0 [11/20]	HM-16.22 [07/20]	Broadcast and streaming	IRAP=CRA GOP length ≈ 1 sec GOP structure=GOP-32 DPB size (pictures) = 8 MCTF=On
		-51%	VVenC-0.3 [03/21]			preset=medium perceptual AQP=ON rate control=OFF
	1080p BT.709 fps: 30, 50	-35%	VTM-11.0 [11/20]	HM-16.22 [07/20]	Conversational	IRAP=IDR GOP length=Inf GOP structure=LD (low delay)
	1080p BT.709 fps: 60	-38%			Gaming	

6.5.3 VVC coding performance – summary

Reported results demonstrate consistent VVC gains over HEVC across video formats and with variety of test content. **Figure 6-8** provides a summary of subjective and objective gains obtained with VVC over HEVC for various application scenarios. The results show that average bitrate savings provided by tend to be larger when assessed in subjective tests by viewers (bitrate savings in the range of 35%-55%) compared with objective measurements (bitrate savings in the range of 30%-40%).

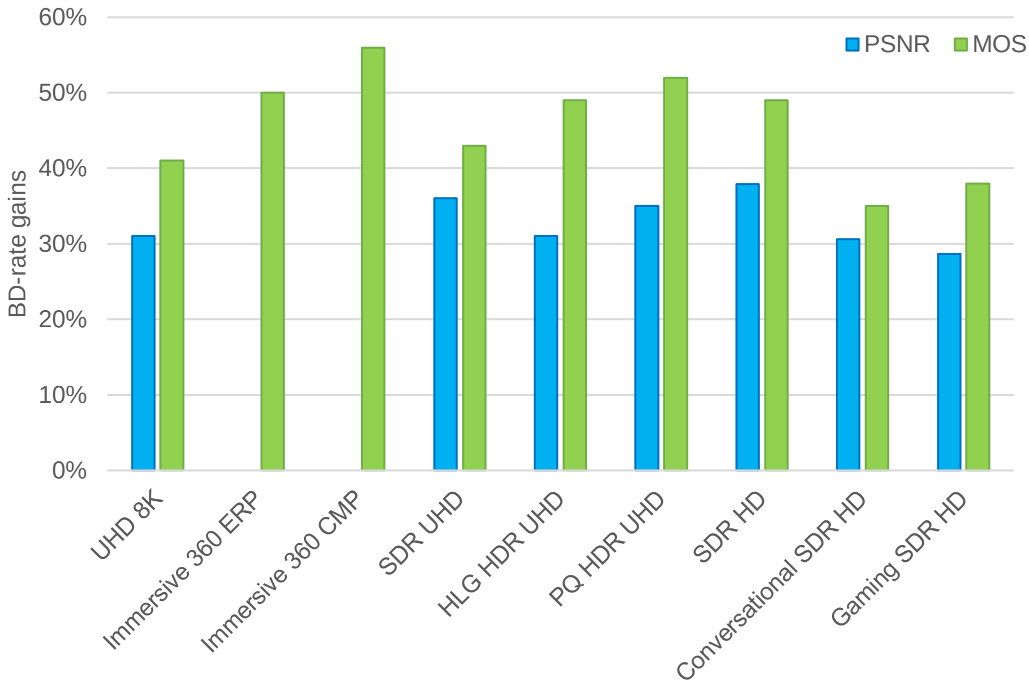

Figure 6-8 Summary of reported VVC coding gains (objective – PSNR and subjective – MOS) over HEVC for various applications.

7 VVC in broadcasting and streaming applications

This section presents some details and specifics regarding VVC use in broadcast and streaming applications. This section reviews VVC configuration aspects taking application centric viewpoint. The section comprises the following sections:

- **7.1 Overview of end-to-end broadcast and streaming ecosystem** describes video workflows withing end-to-end broadcast and streaming ecosystem, key properties of typical video formats and relevant VVC profiles.
- **7.2 VVC support in broadcast and streaming systems** provides an overview of VVC support in media systems and transport standards and relevant broadcast and streaming application specifications.
- **7.3 VVC high-level signaling for distribution video formats** presents high level signalling for video formats supported in relevant broadcast and streaming application specifications.
- **7.4 Random access and zapping time** describes VVC bitstream configuration options and impacts related to random access functionality and zapping (channel change) delays.
- **7.5 Low latency considerations** is currently in draft.
- **7.6 Resolution change** describes VVC support for dynamic resolution changes in a linear broadcast delivery service and improved coding efficiency in adaptive bitrate streaming.
- **7.7 Frame rate change** is currently in draft.
- **7.8 Spatial scalability** is currently in draft.
- **7.9 Operating bitrate ranges** presents expected bitrate ranges for broadcast and streaming video delivery using VVC.

7.1 Overview of end-to-end broadcast and streaming ecosystem

Video content is subject to several stages of processing and transformation within the end-to-end video delivery ecosystem. Some key interfaces can be identified (see UHD Forum's Guidelines [32]) as: content capture and production, content aggregation and distribution and content consumption by end users. **Figure 7-1** shows the interfaces in an end-to-end video production and delivery workflow.

Typically, requirements regarding video content fidelity and quality are gradually being relaxed throughout the delivery chain which is reflected in typical video formats used in respective interfaces. For content capture and production, high fidelity video formats with bit depths between 10-16 bits and chroma 4:2:2 or 4:4:4 RGB formats are used. For such applications VVC profiles, Main 10 4:4:4, Main 12 4:4:4 or Main 16 4:4:4, including Intra-only profiles, are relevant due to ability encode 4:2:2 YCbCr or 4:4:4 RGB chroma formats. Content contribution and primary distribution may require bit depth reduction to 10-12 bits per sample and chroma downsampling to 4:2:2 or 4:2:0. This allows application of profiles supporting 4:2:0 YCbCr formats, such as Main 10 or Main 12 profiles.

Final emission to end users involves further reduction to bit depths 8 – 10 bits per sample and chroma subsampling to 4:2:0. The use of 8 bits per sample format is becoming constrained to SDR-only operating points. However, even in these cases 10 bits per sample coding is recommended if allowed in a given operating point (see section **6.3.5**). Most of applications requirements in the final emission stage, as defined by application specifications (see **7.2**), including enhanced video quality features such as HDR, WCG, HFR can be addressed with the Main 10 profile.

For applications requiring service scalability through multi-layer delivery, Multilayer Main 10 profile is used.

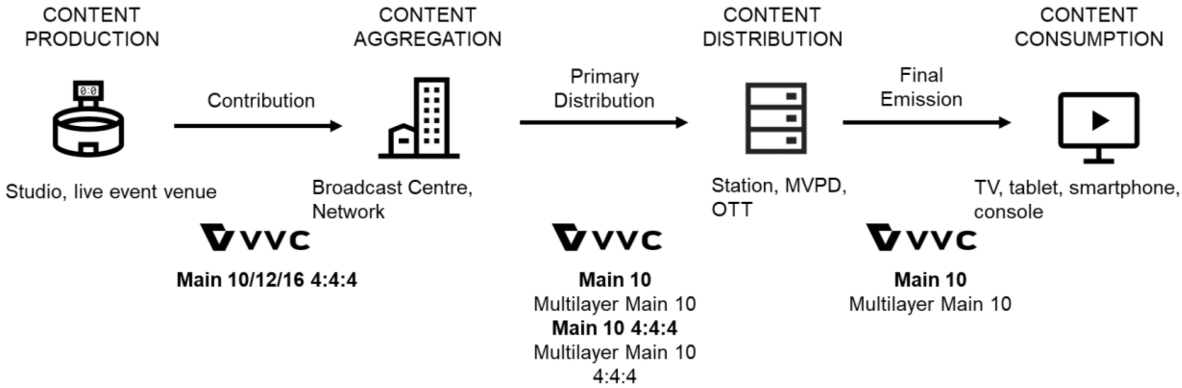

Figure 7-1 End-to-end video content transmission workflow.

7.2 VVC support in broadcast and streaming systems

7.2.1 Systems and transport standards

MPEG has developed and maintains several key media systems and transport standards used for deployment of broadcast and streaming services.

7.2.1.1 MPEG-2 TS (ISO/IEC 13818-1 | Rec. ITU-T H.222.0)

MPEG-2 Transport Stream (MPEG-2 TS) specified in ISO/IEC 13818-1 and Rec. ITU-T H.222.0 is a container format used for encapsulation packetized elementary stream. MPEG-2 TS is a foundation of several digital video broadcast and broadband standards. VVC carriage in MPEG-2 Transport Stream was added in the 8th edition of ISO/IEC 13818-1 [8].

7.2.1.2 MPEG ISO Base Media File Format (ISO/IEC 14496-15)

MPEG ISO Base Media File Format (ISO BMFF) standardized as ISO/IEC 14496-12, is a container for storage and carriage of timed media [17]. VVC, in the same way as its predecessors AVC and HEVC is built on a network abstraction layer (NAL) structuring interface, and the data structures carried by this interface are referred to as NAL units (see section **6.1.2**). ISO/IEC 14496-15 standard (Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format) specifies how VVC (as well as AVC and HEVC) are stored in file formats based on ISO BMFF. VVC support was added in Amendment 2 of ISO/IEC 14496-15 [19]. Storage of both single-layer and multi-layer VVC bitstreams is defined. The standard also defines samples and sub-samples reflecting the high-level bitstream structure and independently decodable units such as rectangular regions comprising of VVC subpictures or slices, e.g., for immersive applications. Signalling and extraction of a specific operating point with a combination of video layers is also supported.

7.2.1.3 MPEG DASH (ISO/IEC 23009-1)

MPEG-DASH is a media delivery protocol for Dynamic Adaptive Streaming over HTTP standardized by MPEG as ISO/IEC 23009-1 in 2012 [20] and has been subsequently extended to cover additional features. MPEG-DASH is codec agnostic and can be used with media codecs contained in MPEG ISO BMFF (ISO/IEC 14496-12) or MPEG-2 TS (ISO/IEC 13818-1). Integration of media codecs can be specified externally defined interoperability points, e.g., through DASH-IF (see paragraph **7.2.2.4**).

7.2.1.4 MPEG CMAF (ISO/IEC 23000-19)

MPEG Common Media Application Format (CMAF) standardized in ISO/IEC 23000-19 [15] is a common streaming format driven by the convergence between MPEG-DASH and HTTP Live Streaming (HLS) [16]. VVC support was added in the 3rd edition of ISO/IEC 23000-19 in 2022. VVC is supported in MPEG CMAF through two media profiles. The baseline VVC media profile targets single layer bitstreams conforming to VVC

Main 10 profile and Main Tier. The multilayer VVC media profile targets multilayer bitstreams conforming to VVC Multilayer Main 10 profile and Main Tier.

7.2.1.5 MPEG HEIF (ISO/IEC 23008-12)

The High Efficiency Image File Format (HEIF) specified in ISO/IEC 23008-12 is a standard developed by MPEG for the storage of images and image sequences [21]. VVC support was added in the second edition of the standard in 2022.

7.2.1.6 MPEG OMAF (ISO/IEC 23090-2)

The Omnidirectional Media Format (OMAF) is s a storage and streaming format for omnidirectional media, including 360◦ omnidirectional video specified in ISO/IEC 23090-2 [22]. The following VVC profiles have been added in the third edition of the standard: VVC-based simple tiling OMAF video profile, VC-based viewport-independent OMAF video profile, OMAF VVC image profile, and CMAF media profile for the VVC-based viewport-independent OMAF video profile.

7.2.1.7 IETF

While not a broadcast standard setting organization as the previously mentioned groups, the IETF has created, or is in the process to create, certain Requests For Comments (RFCs) that can and are used in the broadcast and streaming fields. In the broadest sense, protocols such as IP4 and IPv6 are used on the ISO/OSI routing layer, and TCP and UDP on the transport layer. On the application layer, technologies like MPEG-DASH build on http, and in development are specifications and proof of concept implementations of media transmission over QUIC. All mentioned IETF technologies are agnostic to the media coding technology used to code the media bitstreams, and hence not further discussed herein.

However, the IETF also specifies a protocol known as Real-Time Transport Protocol (RTP) in RFC3550. While RTP itself is also media coding agnostic, it relies on RTP payload formats that enable media specific optimizations in the transport, as well as capability exchange/negotiation using other IETF and non-IETF media agnostic protocol suites such as WebRTC, SIP, or ITU-T H.323. Of these, WebRTC - originally intended as video conferencing for browsers - is increasingly in used for IP-based contribution.

To support VVC, the IETF developed the RTP Payload Format for Versatile Video Coding as RFC 9328 [47]. This payload format, on the media plane implements all commonly used features and packetization modes known from payload formats of previous ITU/MPEG NAL-unit-based video codecs, such as H.264/AVC and H.265/HEVC. Signaling H.266 support using system standards utilizing the RTP payload format's Media Type registration necessarily will need adaption, but other than that, the integration of VVC into a system design supporting, for example, H.265/HEVC should be straightforward.

7.2.2 Application specifications

7.2.2.1 ARIB ISBD

ARIB ISBD has been investigating VVC Main 10 and Multilayer Main 10 profiles for its next generation digital video broadcasting system [47].

7.2.2.2 ATSC

The Advanced Television Systems Committee (ATSC) has noted their intention to include VVC in the ATSC 3.0 standard. A report on ATSC 3.0 and Global Convergence mentions explicitly that "ATSC is currently specifying Versatile Video Coding (VVC) for inclusion in the ATSC 3.0 suite of standards" [49].

7.2.2.3 CTA Wave

CTA Wave added VVC profile to its Web Application Video Ecosystem Content Specification (CTA-5001-D) in 2021 [50].

7.2.2.4 DASH-IF

DASH Industry Forum (DASH-IF) added VVC profile to its DASH-IF Interoperability Points guidelines in 2022 [51].

7.2.2.5 DVB

The DVB project endorsed commercial requirements for next generation video codecs for advanced 4K and 8K services in July 2021. Commercial requirements for prospective codecs included [47]:

- support for 8K video with resolution up to 7680x4320 pixels, with support for high dynamic range (HDR) and high frame rates (HFR);
- the ability to delivery 8K video over legacy broadcast multiplexes at excellent quality;
- the ability to enable five 4K services in a 40 Mbit/s DVB-T2 multiplex (compared with three 4K services expected with HEVC);
- provide at least 27% more efficient live broadcast encoding than HEVC, and over 30% performance gains for 4K streaming use cases while maintaining performance gains for sub 4K resolutions.

The DVB project added VVC to its core specification TS 101 154 in February 2022. The revision was published in [53]. An update to the DVB-DASH specification (TS 103 285) including VVC was published in [54].

The DVB project office published VVC transport streams and DASH packages for verification and validation of VVC video services conforming to updated DVB specifications in March 2023 [55].

ETSI’s version of TS 101 154 including VVC was published in June 2023 [56].

7.2.2.6 Fórum SBTVD

Sistema Brasileiro de Televisão Digital (aka Forum SBTVD) is the organization charged with creation of broadcast and hybrid broadcast-broadband television standards in Brazil. SBTVD has been working toward a new standard for Brazil known as TV 3.0.

Forum SBTVD issued call for proposals (CfP) in in July 2020 [83]. Use cases and requirements in the CfP included:

- delivery of 4K OTA and 8K via OTT
- native HDR/WCG support, HFR and
- a reduced-resolution portrait-mode closed second video service for sign language purposes.

VVC was selected as the Video Base Layer Codec, for both OTA and OTT delivery.

7.2.2.7 SCTE

SCTE adopted VVC into its standards, SCTE 281-1 [57] and 281-2 [58] in March 2023. The standards specify the carriage of VVC for cable video services. VVC carriage based on MPEG-2 transport stream for linear delivery and based on SCTE 214-1 for adaptive bitrate streaming is specified.

7.2.3 VVC operating points in application specifications

Table 7-1 provides a summary of specified VVC operating points across different application specifications. **Table 7-2** provides an overview of supported HDR metadata for VVC specified operating points.

Table 7-1 VVC constraints in application specifications

Specification	VVC profile tier level	Max resolution and frame rate	Colour coding and dynamic range	Progressive only	8-bit coding allowed
CTA Wave CTA-5001-D	Main 10 MT L6.1 Multilayer Main 10 L6.1	7680 x 4320 60 fps	SDR HDR PQ	Yes	Yes
DVB TS 101 154 DVB-DASH TS 103 285	Main 10 MT 4 operating points: L5.1, L5.2, L6.1, L6.2	7680 x 4320 120 fps (Main 10 MT L6.2)	SDR SDR with WCG SDR with HLG HDR PQ HDR HLG	Yes	Yes No
SBTVD *(in draft)*	*Main 10 MT L6.2*	*7680 x 4320* *120fps*	*HDR PQ*	*Yes*	*No*
SCTE SCTE 281-1 2023	Up to Main 10 MT L6.2	7680x4320 120 fps	SDR SDR with WCG HDR PQ	No Yes	Yes No

Table 7-2 VVC and associated HDR metadata in application specifications

Application specification	**HDR metadata and SEI payload type**						
	MDCV	CLLI	ATC	SMPTE ST 2094-10	SL-HDR1	SL-HDR2	SMPTE ST 2094-40
	137	144	147	4 (ITU-T T.35)			
CTA Wave CTA-5001-D	HDR PQ (optional)	-	-	-	-	-	-
DVB TS 101 154 DVB-DASH TS 103 285	HDR PQ (optional)	HDR PQ (optional)	SDR with HLG	HDR PQ (optional)	-	HDR PQ (optional)	HDR PQ (optional)
SBTVD TV3.0 *(in draft)*	*HDR PQ (optional)*	-	-	*HDR PQ (optional)*	-	*HDR PQ (optional)*	*HDR PQ (optional)*
SCTE SCTE 281-1 2023	SDR with WCG HDR PQ	HDR PQ (optional)	-	HDR PQ	SDR with WCG HDR PQ	-	HDR PQ

7.3 VVC high-level signaling for distribution video formats

Paragraph **7.2.3** provides an overview of video formats and VVC operating points defined in broadcast and streaming application specifications. The coding-independent code points (CICP) specification for video (standardized as Rec. ITU-T H.273 | ISO/IEC 23091-2) [28] defines code points and fields that identify properties of video signals. Technical report [29] (Rec. ITU-T H Suppl. 19 | ISO/IEC TR 23091-4) provides information on value ranges for code points commonly used in video production and distribution workflows. CICP specification does not cover VVC- and VSEI- specific signalling values which are covered in this section for the relevant formats and operating points.

Table 7-3 Profile, tier, level signaling for Main 10 profile (Main Tier).

Syntax parameter	Values	Description
Sequence parameter set [1]		
sps_chroma_format_idc	1	4:2:0 chroma
sps_bitdepth_minus8	0	8-bit
	2	10-bit
sps_ptl_dpb_hrd_params_present_flag	1	
general_profile_idc	1	Main 10 profile
general_tier_flag	0	Main tier
general_level_idc	-	see **Table 7-4**

Table 7-4 Signaling of general_level_idc values.

	Spatial resolution			
Frame rate [fps]	**1280x720**	**1920x1080**	**3840x2160**	**7680x4320**
23.97	51	64	80	96
24	51	64	80	96
25	51	64	80	96
29.97	51	64	80	96
30	51	64	80	96
50	64	67	83	99
59.94	64	67	83	99
60	64	67	83	99
100	67	80	86	102
119.88	67	80	86	102
120	67	80	86	102

Table 7-5 Spatial resolution signaling.

Spatial resolution	***Max(8, MinCbSizeY)***	**sps_log2_ctu_size_minus5**	***CtbSizeY***	**sps_field_seq_flag**	**sps_pic_width_max_in_luma_samples**	**sps_pic_height_max_in_luma_samples**	**sps_conformance_window_flag**	**sps_conf_win_left_offset**	**sps_conf_win_right_offset**	**sps_conf_win_top_offset**	**sps_conf_win_bottom_offset**
1920x1080	8			0	1920	1080	0	0	0	0	0
		2	128	0	1920	1152	1	0	0	0	72
		1	64	0	1920	1088	0	0	0	0	8
3840x2160	8			0	3840	2160	0	0	0	0	0
		2	128	0	3840	2176	1	0	0	0	16
		1	64	0	3840	2176	1	0	0	0	16
7680x4320	8			0	7680	4320	0	0	0	0	0
		2	128	0	7680	4352	1	0	0	0	32
		1	64	0	7680	4352	1	0	0	0	32

Table 7-6 Frame rate signaling.

Syntax parameter	Values	Description
Sequence parameter set [1]		
sps_field_seq_flag	0	CLVS conveys pictures that represent frames
sps_ptl_dpb_hrd_params_present_flag	1	
sps_timing_hrd_params_present_flag	1	
num_units_in_tick		See **Table 7-7**
time_scale		See **Table 7-7**

Table 7-7 num_units_in_tick and time_scale signaling.

Frame rate [Hz]	num_units_in_tick	time_scale
23.97	24 000	1 001
24	24	1
25	25	1
29.97	30 000	1 001
30	30	1
50	50	1
59.94	60 000	1 001
60	60	1
100	100	1
119.88	120 000	1 001
120	120	1

Table 7-8 Color and dynamic range signaling (for chroma 4:2:0 formats)

<table>
<tr><td rowspan="4">Syntax parameter</td><td colspan="7">Colour coding and dynamic range tag</td></tr>
<tr><td>Tag</td><td colspan="2">SDR</td><td>SDR+WCG</td><td>SDR+HLG</td><td>HDR PQ</td><td>HDR HLG</td></tr>
<tr><td>Colour primaries</td><td colspan="2">BT.709</td><td>BT.2020</td><td>BT.2020</td><td>BT.2100</td><td>BT.2100</td></tr>
<tr><td>Transfer function</td><td colspan="2">BT.709</td><td>BT.2020</td><td>BT.2020</td><td>BT.2100 PQ</td><td>BT.2100 HLG</td></tr>
<tr><td colspan="8">Sequence parameter set [1]</td></tr>
<tr><td colspan="2">sps_bitdepth_minus8</td><td>0</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td></tr>
<tr><td colspan="2">sps_chroma_format_idc</td><td colspan="2">1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td colspan="2">sps_field_seq_flag</td><td colspan="2">0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td colspan="2">sps_vui_parameters_present_flag</td><td colspan="2">1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td colspan="8">Video usability information [2]</td></tr>
<tr><td colspan="2">vui_progressive_source_flag</td><td colspan="2">1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td colspan="2">vui_interlace_source_flag</td><td colspan="2">-</td><td>-</td><td>-</td><td>1</td><td>1</td></tr>
<tr><td colspan="2">vui_colour_description_present_flag</td><td colspan="2">1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td colspan="2">vui_colour_primaries</td><td colspan="2">1</td><td>9</td><td>9</td><td>9</td><td>9</td></tr>
<tr><td colspan="2">vui_transfer_characteristics</td><td colspan="2">1</td><td>14</td><td>14</td><td>16</td><td>18</td></tr>
<tr><td colspan="2">vui_matrix_coeffs</td><td colspan="2">1</td><td>9</td><td>9</td><td>9</td><td>9</td></tr>
<tr><td colspan="2">vui_full_range_flag</td><td colspan="2">0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td colspan="2">vui_chroma_loc_info_present_flag</td><td colspan="2">1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td colspan="2">vui_chroma_sample_loc_type_frame</td><td colspan="2">0</td><td>2</td><td>2</td><td>2</td><td>2</td></tr>
<tr><td colspan="2">Alternative transfer characteristics information SEI [2]</td><td colspan="2">-</td><td>-</td><td>1</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">preferred_transfer_characteristics</td><td colspan="2">-</td><td>-</td><td>18</td><td>-</td><td>-</td></tr>
</table>

7.4 Random access and zapping time

Random access points (RAP) mark entry points to a video bitstream. RAPs enable receivers to tune in to a video bitstream and start decoding from any such entry point. RAPs enable several functionalities such as starting video consumption from an arbitrary point in time (subject to zapping time delay), seeking, fast-forward/trick-play modes, channel switching for broadcast and representation switching for broadband services and content splicing. In broadcast and streaming applications use of regular and frequent RAPs is necessary to provide these functionalities.

In VVC, RAPs using IDR and CRA NAL unit types are coded using intra prediction only and therefore referred to as IRAP. VVC also mandates RAP with GDR NAL unit type, where GDR itself may include inter-predicted parts of a picture but at the expense of extending decoder refresh across potentially several pictures. **Table 7-9** provides a summary of RAP support across application standards including the mapping of VVC NAL units types against Stream Access Point types as defined in ISO BMFF [17].

Table 7-9 Supported random access point types.

	IDR_W_RADL **SAP type 2**	**IDR_N_LP** **SAP type 1**	**CRA_NUT** **SAP type 3**	**GDR_NUT** **SAP type 4**
DVB TS 101 154	Supported	Supported	Supported	Not supported
MPEG-DASH	Supported	Supported	Supported	Supported
MPEG-CMAF CTA Wave CTA-5001-D DVB-DASH TS 103 285	Supported	Supported	Supported for chunk random access only	Not supported
SCTE 281-1 2023	Supported	Supported	Supported	Not supported

7.4.1 GOP length

Recommended use of RAPs intervals (i.e., GOP length) is typically between 1-2 seconds. Frequent and regular RAPs may improve zapping time change and trick play modes. However, shorter time interval between RAPs (short GOPs) incurs bitrate increase due to less efficient compression. This stems from:

- lower compression efficiency of I pictures compared with L0-B or P pictures;
- disruptions to inter-prediction structure depending on the used RAP NAL unit type (as discussed in section 6.3.2).

It should be noted that intra-coded pictures are not used solely for enabling RAPs and a typical video bitstream may contain more frequent I pictures. For example, in typical video content scene changes or digital video effects create natural disruptions to inter-prediction structures and encoders may benefit from using additional I pictures.

7.4.2 Zapping time

Zapping (channel change) time refers to delay which is incurred when receiver starts up decoding a new programme. Since decoding can start from RAPs, distance between RAPs is one of the determining factors. Other factors are related to picture reordering delay which is related to the employed GOP structure (see section **6.3.1**), the type of RAP used (see section **6.3.2**) and video content frame rate. **Figure 7-2**, **Figure 7-3** and **Figure 7-4** illustrate relationship between Decoding Time Stamp (DTS) and Presentation Time Stamp (PTS) for GOP-4 hierarchical structure (L2 structure in **Figure 6-2**) and closed-GOP type 1, closed-GOP type 2 and open-GOP respectively.

The PTS(IRAP) - DTS(IRAP) offset increases with larger GOP structure (relative to video content frame rate). Closed-GOP type 1 results in smaller offset than closed-GOP type 2 or open-GOP. Display delay for both types of closed-GOP is the same.

DVB specification TS 101 154 recommends PTS(IRAP)-DTS(IRAP) offset to be less than or equal to 0.67 seconds.

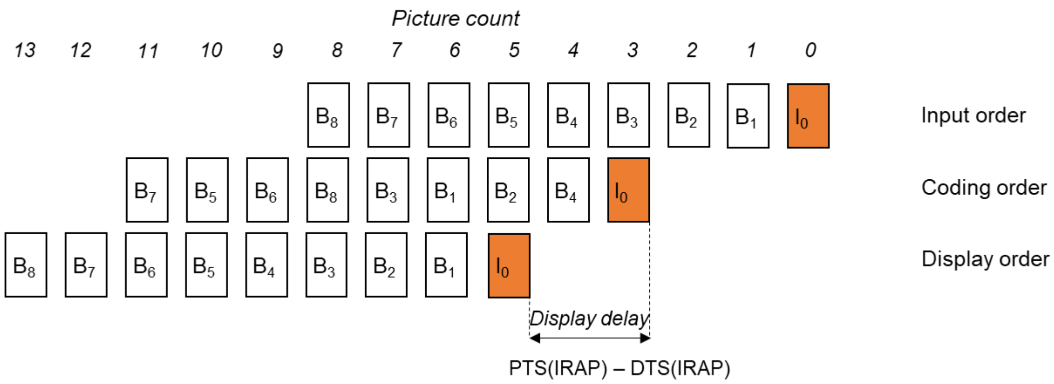

Figure 7-2 Example of input, coding, and display order for IRAP with IDR_N_LP (closed-GOP type 1) when decoding starts at IRAP picture (I_0).

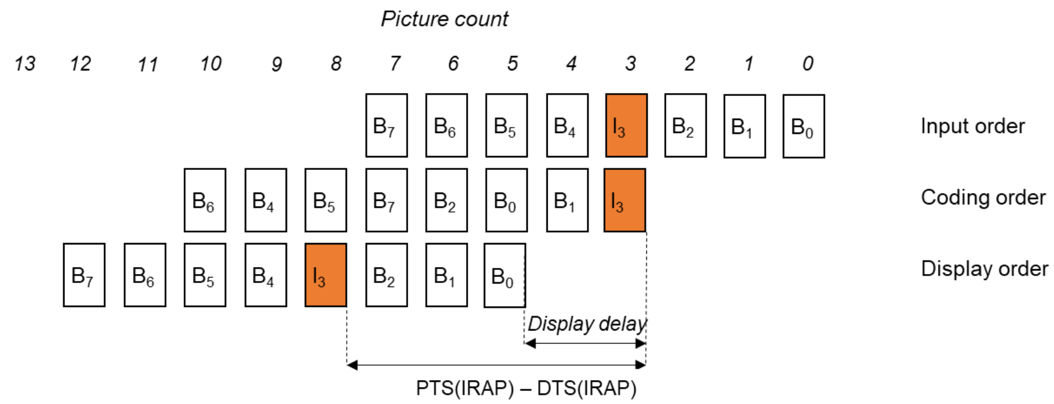

Figure 7-3 Example of input, coding, and display order for IRAP with IDR_W_RADL (closed-GOP type 2) when decoding starts at IRAP picture (I_3).

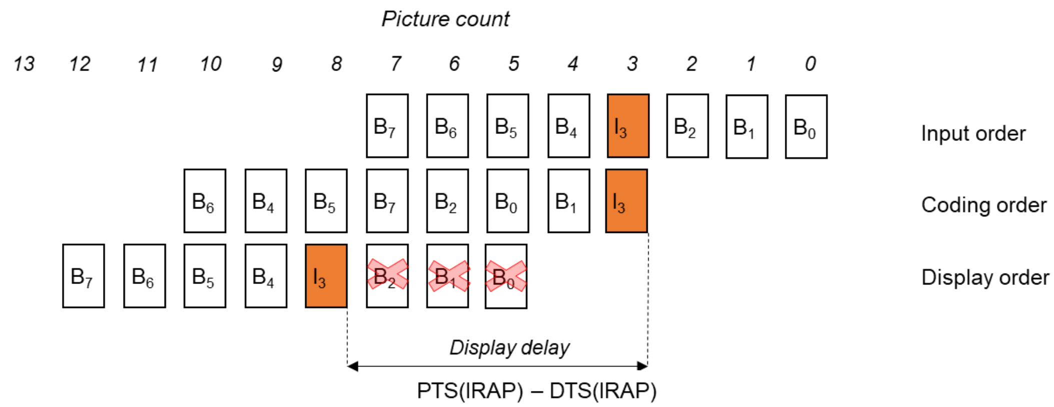

Figure 7-4 Example of input, coding, and display order for IRAP with CRA_NUT (open-GOP) when decoding starts at IRAP picture (I_3).

7.5 Low latency considerations

This section is currently in draft.

7.6 Resolution change

Delivery of video services by the means of providing multiple renditions of a video content coded at different resolutions and bitrates has been the backbone of HTTP adaptive streaming (HAS) for broadband delivery (e.g., DASH or HLS). Initially, designed to target widely varying end-device display formats and variable available bandwidth, techniques such as “Per-title encoding” or Content-adaptive encoding (CAE) have been used to optimize resolutions and bitrates towards complexity of the video content. Independent decoding of short segments of video bitstream with different bitrates and resolution can be achieved with AVC or HEVC with the use of closed-GOP structure (see section **6.3.2**). However, switching resolutions while using more efficient open-GOP structures has not been possible.

For broadcast delivery, such adaptation to video content complexity has been managed from the head-end and dynamic allocation of bitrates between video channels in a statmux group. While the same mechanism as employed for HAS streaming should be possible with AVC and HEVC (i.e., performing resolution change at the IDR pictures), interoperability tests with DVB-T2 TVs demonstrated that seamless switching was not achieved [84].

VVC RPR feature extends resolution change functionality in both linear broadcast delivery by allowing dynamic resolution changes within a single video representation and in HAS streaming by enabling open-GOP switching. The use cases are described in detail below.

7.6.1 Resolution change in a broadcast delivery service

VVC offers ability to change resolution within a single video representation at CVS boundaries or within a CVS. The first approach is to signal new resolution for all AUs in CVS at an IRAP picture. Depending on the type of IRAP picture used, RPR functionality is either used (CRA type of picture) or not (with IDR).

7.6.1.1 Resolution change at IRAP picture

Resolution change at the start of a new GOP with an IDR picture is illustrated in **Figure 7-5**. In this example, resolution is changed from 4K (3840 x 2160) to 2K (1980 x 1080).

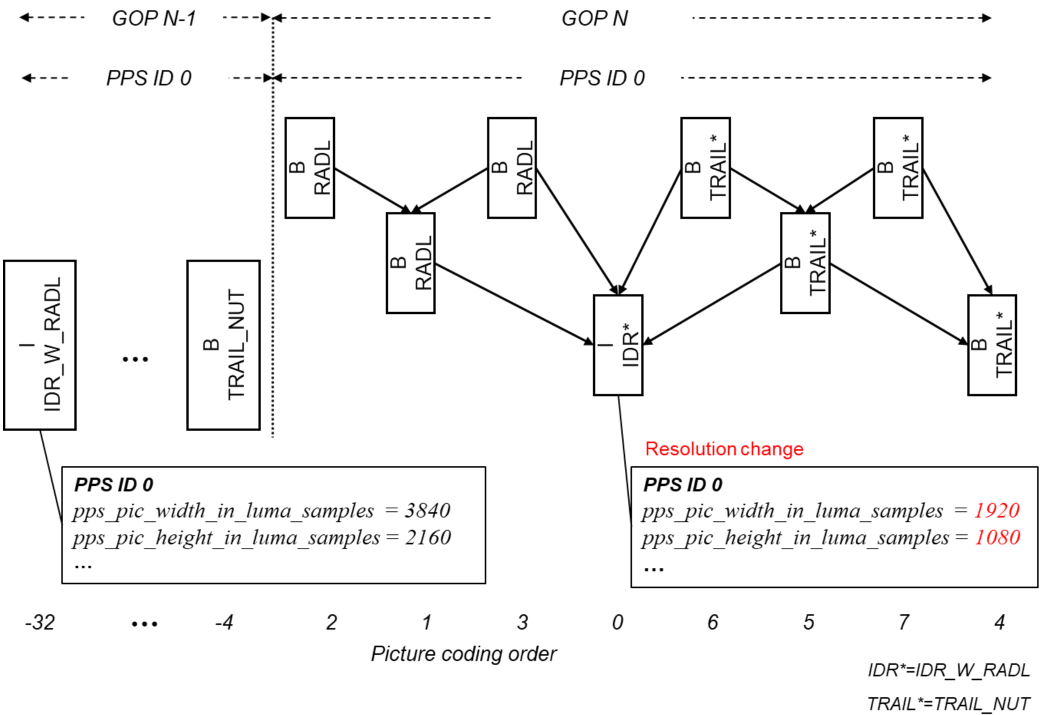

Figure 7-5 Example of a resolution change at IRAP picture (IDR_W_RADL).

Resolution change at IRAP with CRA picture is illustrated in **Figure 7-6**. In this case RASL pictures may reference pictures from a previous CVS which are codec at a different resolution to the CRA picture (due to resolution change at the CRA picture). This is enabled by setting **sps_ref_pic_resampling_enabled_flag** equal to 1.

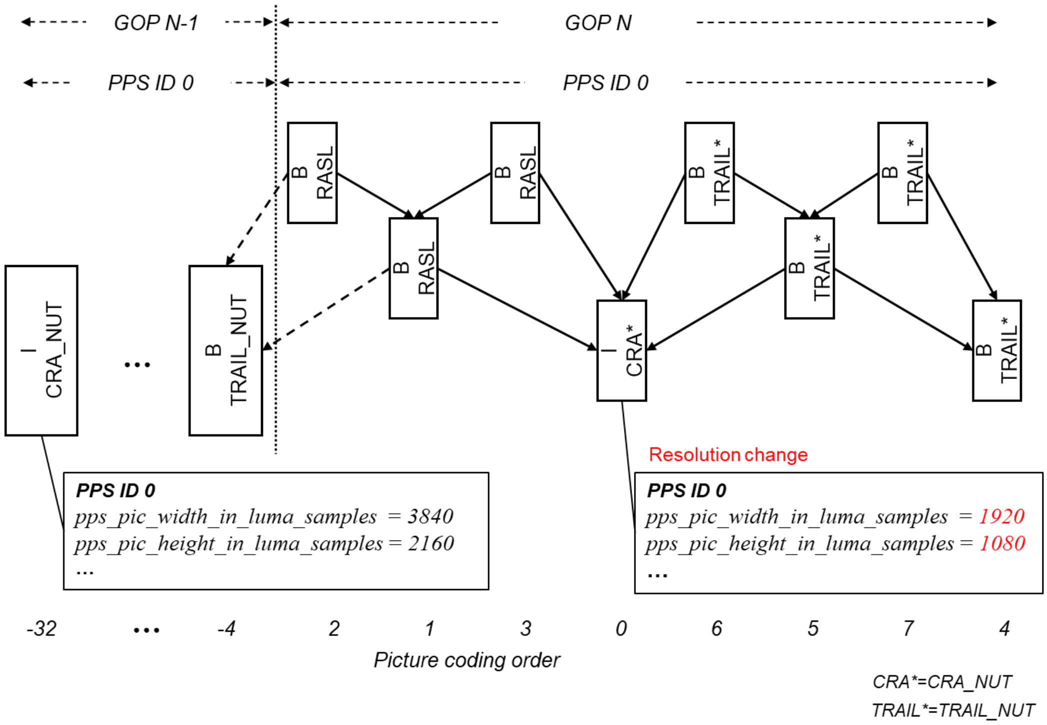

Figure 7-6 Example of a resolution change at IRAP picture (CRA_NUT). Dashed lines indicate required reference picture resampling.

7.6.1.2 Resolution change at non-IRAP picture

The second approach is to perform resolution change within a GOP (CVS) with the use of RPR. This is enabled by setting both **sps_ref_pic_resampling_enabled_flag** and **sps_res_change_in_clvs_allowed_flag** equal to 1. VVC specification allows change of resolution of any number of pictures at any time. However, it is recommended that referred resolutions in PPSs are signaled at IRAP picture. This imposes a constraint on encoder to pre-determine resolutions within the whole CVS already at the start of encoding (IRAP picture). Practically, however, this may not be an issue due to other operational requirements on ensuring frequent RAP as well as other application specific constraints (see section **7.6.1.4**).

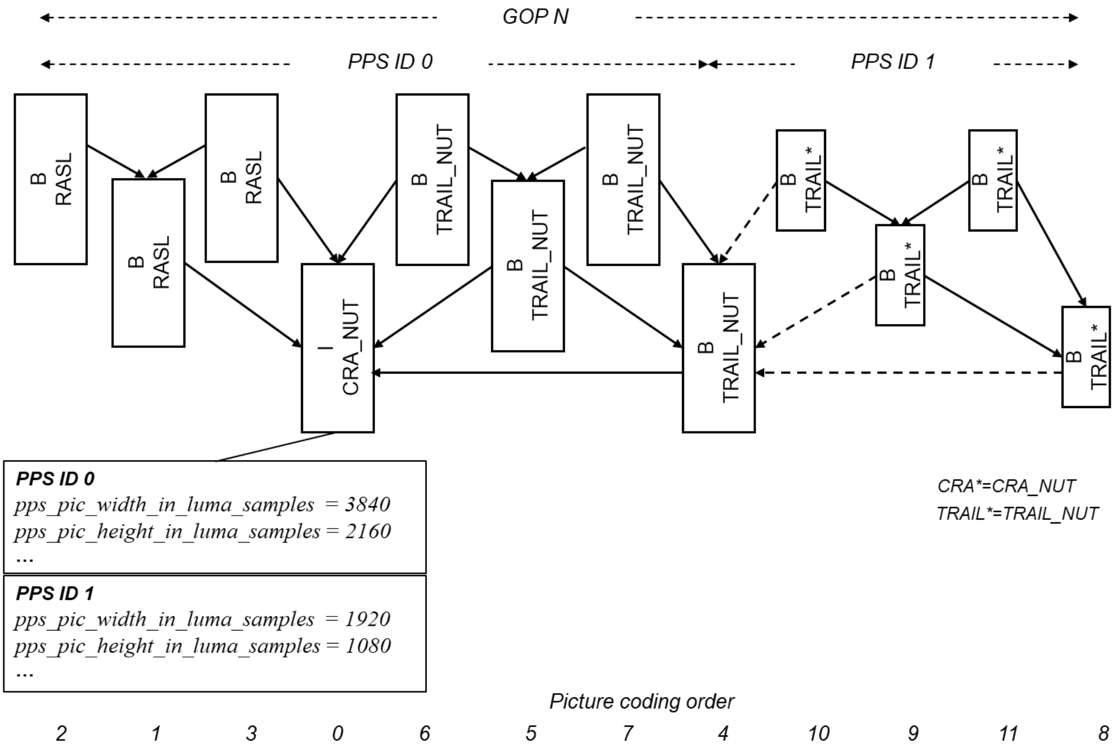

Figure 7-7 Example of resolution change at non-IRAP picture. Additional PPS (PPS ID 1) with intended new resolution is signalled at the IRAP picture but only referenced from a resolution change point.

7.6.1.3 Benefits of resolution change in a broadcast delivery service

Use of variable resolution encoding using VVC Main 10 profile in a broadcast delivery service allows encoders to choose optimal trade-offs with regards to coded resolution and available bandwidth. Therefore, resolution change can be seen as an extension of rate control functionality of encoders. An example control algorithm based on GOP-structure resolution determination was introduced in VTM reference software [85] and was reported to provide additional 10% objective performance gains. Variable resolution encoding with VVC was further studied in [86] and objective gains up to 22% were reported. Subjective evaluation showed that compression performance benefits were clearly visible at low bitrates. Overall, resolution changes were considered to be seamless.

7.6.1.4 Constraints in application specifications

Constraints related to the use of RPR in VVC, described in section 6.3.3, concern the restrictions regarding scaling factors and the use of selected VVC's coding tools. Otherwise, VVC allows resolution change on any picture. Additional constraints may be imposed by application specification.

DVB's TS 101 154 [53] puts additional restrictions such that resolution change can only happen between specified resolutions. Additional constraint in DVB's specification is related to the frequency of resolution change.

7.6.2 Resolution change in a streaming service

As discussed in section **7.6.1** resolution change can happen either by starting a new CVS, i.e., at IDR boundaries or within a CVS either at any picture or at IRAP boundaries, i.e., at CRAs. Resolution change within a CVS at any picture could be performed as part of the encoders rate distortion optimization but is not discussed in the following. The most relevant aspect for streaming services based on adaptive HTTP streaming is resolution change as a result of switching at segment boundaries.

Using IDRs for resolution change is a typical practice and has been widely used in the past. However, as already discussed, the compression performance of IDRs compared to using CRAs (i.e., open GOP structures) is associated with an inferior compression performance (see **Table 6-3**). Switching resolution in a streaming service is typically the result of a rate adaptation algorithm on the receiver side as a response of throughput variations. It switches from the segment of one bitstream (produced by one encoder) to a segment of another bitstream (produced by another encoder). Therefore, the encoding of the content needs to be carefully prepared as discussed in the following.

In case of segment switching in adaptive HTTP streaming scenarios, the change in resolution is not determined by the encoder/sender but happens as a result of "manipulation" of a bitstream. In such scenarios, typically, several bitstreams exist with GOPs that have the same resolution within a particular bitstream, but different resolutions among different bitstreams. Therefore, when GOPs of different bitstream are concatenated (e.g., as a result of switching to segments of a different bitstream) the decoder/receiver is fed with a bitstream that has GOPs with different resolutions. When open GOP structures are used, RPR is active at switching points in the mixed bitstream. If the encoder had used the tools that are automatically de-activated when RPR is used for pictures for which the decoder de-activates such tools, visual artifacts would be visible during such switches. Therefore, to allow bitstreams to be manipulated as described above, at least some of the RPR constraints described in section **6.3.3** need to be activated for RASL pictures associated with an CRA of an open GOP structure.

In particular, the tools that have been identified as clearly contributing to artifacts when they were not constrained at the encoder side are: DMVR, (SB)TMVP and Reference picture wraparound. Therefore, these tools need to be de-activated at the encoder side for RASL pictures when they have active reference pictures prior to their associated IRAP in decoding order. Not de-activating BDOF and PROF at the encoder for RASL pictures has not shown to cause any severe drift at the decoder side when such resolution switches happen and therefore could be used at the encoder side without causing any problem.

In addition to the aforementioned constraints, Cross Component Linear Model (CCLM) introduces also noticeable artefacts in the color components of reconstructed samples when performing a bitstream manipulation as described above, by concatenating GOPs of different bitstreams that use open GOP structures, e.g., bitrate or resolution switching in streaming scenarios. In order to avoid such problems, CCLM needs to be constrained for RASL pictures so that B and P slices do not contain any CU that has **cclm_mode_flag** equal to 1. Constraining an encoder to disable the aforementioned tools for RASL pictures associated with a CRA pictures reduces the coding efficiency. However, the open GOP coding approach still provides coding efficiency benefits compared to closed GOP when applying these constraints as can be seen on the right side of **Table 6-3**.

Indication that a bitstream follows the discussed constraints for RASL picture associated with a CRA picture is important for system integrations and is carried out by including the constrained RASL encoding indication (CREI) SEI message at each CVS of the bitstream and can be achieved by constraining the RASL pictures of the bitstream as follows:

— The PH syntax structure has **ph_dmvr_disabled_flag** equal to 1.

— The PPS referred to by the RASL picture has **pps_ref_wraparound_enabled_flag** equal to 0.

— No CU in a slice with sh_slice_type equal to 0 (B) or 1 (P) has **cclm_mode_flag** equal to 1.

— No collocated reference picture precedes the CRA picture associated with the RASL picture in decoding order.

A further aspect that needs to be considered, when resolution switching is performed with open GOP structures, is that the IRAP at which the resolution switch happens (i.e., the CRA picture) does not start a new CVS and therefore the SPS NAL unit does not change. This means that the encodings of the different bitstreams need to be to some extent coordinated, e.g., the same luma coding tree block size shall be used, or same tools shall be enabled. Most of the syntax elements in the SPS NAL units of the different bitstreams need to be the same, with some exceptions. For instance, syntax elements indicating the maximum picture size are typically different indicating the picture size of the different bitstreams (**sps_pic_width_max_in_luma_samples** and **sps_pic_height_max_in_luma_samples**). Also, as indicated in **Table 7-5**, the conformance window might be different for different resolutions, leading to SPS NAL units having for instance different values for **sps_conf_win_bottom_offset**. The conformance window in VVC is of particular interest for open GOP resolution switching, since as discussed when resolution changes with open GOP structures, the SPS NAL unit does not change and a GOP after a switch can have different conformance window. VVC allows to indicate the conformance window in the PPS NAL unit in addition to the SPS NAL unit. In order to allow for resolution switch with open GOP when not all bitstreams have the same conformance window PPS NAL units are required to set the value of **pps_conformance_window_flag** to 1 and indicate the conformance window.

Finally, since resolution switching might happen, even though each original bitstream does not use RPR, the value of **sps_ref_pic_resampling_enabled_flag** and **sps_res_change_in_clvs_allowed_flag** in the SPS shall be set to 1, with the associated constraints that subpictures and virtual boundaries cannot be enabled, i.e., **sps_subpic_info_present_flag** and **sps_virtual_boundaries_present_flag** shall be equal to 0.

7.7 Frame rate change

This section is currently in draft.

7.8 Spatial scalability

This section is currently in draft.

7.9 Operating bitrate ranges

Operating bitrates can be widely dependent on several factors including:

- distribution formats (resolution and frame rates)
- content complexity
- video quality requirements
- CBR vs. VBR (statmux)
- encoder's complexity and performance (expected to improve over time)
- real-time vs. offline encoding.

7.9.1 VVC bitrates in linear video applications

The following sections look at estimating operating VVC bitrate ranges for different operating points based on two methodologies. The first methodology surveys reported and recommended bitrate ranges for HEVC deployments and applies bitrate savings for VVC over HEVC as reported in section **6.5**. Data from the following reports and recommendations were used:

- Recommendation ITU-R BT.2073-2 [30] provides maximum bitrate values for high quality broadcast (CBR) of critical test sequences as assessed by experts.
- Report ITU-R BT.2343-8 [31] provides a collection of HEVC emission bitrates from field trials of UHDTV over DTTB networks.
- UHD Forum guidelines [32] provides recommended HEVC bitrate ranges for UHD services.
- 3GPP TR 26.925 [24] provides information on typical media traffic characteristics on 3GPP networks.
- EBU report TR 036 [25] provides an analysis of TV programme accommodation in a DVB-T2 multiplex.

Table 7-10 Reported HEVC operating bitrates.

Profile, tier, level	Spatial resolution	Max. frame rate	Bitrate(s) [Mbit/s]	Source
UHDTV 8K				
HEVC Main 10 MT L6.2	7840x4320	120	90-120	[30]
			85-110	[45]
HEVC Main 10 MT L6.1	7840x4320	60	80-100	[30]
			28-113	[31]
			80-100	[46]
			25-80	[24]
UHDTV 4K				
HEVC Main 10 MT L5.2	3840x2160	120	35-50	[30]
			30	[31]
HEVC Main 10 MT L5.1	3840x2160	60	10-40	[32]
			30-40	[30]
			9-36	[31]
			10.4-22.5	[25]
			8-16	[24]
UHDTV 2K				
HEVC Main 10 MT L4.1	1920x1080	60	5-18	[32]
			10-15	[30]
			2.5-3	[31]
			5-7	[24]

The second methodology looks at data from available subjective tests and derives VVC bitrates based on expected picture quality. Data from early VVC deployments and trials is also provided.

7.9.1.1 UHDTV 4K applications - expected bitrates

Methodology 1: VVC bitrates estimated from HEVC bitrates

Bitrate ranges reported for HEVC in UHDTV 4K applications roughly align in the range of 10-40 Mbit/s. Assuming 40%-50% expected performance gains provided by VVC, an operating bitrate range for VVC can be estimated to be between 5-6 Mbit/s for the low-end bitrates and 20-24 Mbit/s for the high-end bitrates.

Methodology 2: VVC bitrates estimated from subjective assessment tests data

The data presented in **Figure 8-1** and **Figure 8-2** broadly aligns with the 10-40 Mbit/s operating range for HEVC and 5-20 Mbit/s operating range for VVC at the same visual quality. It should be noted that commercial encoders may not employ all coding tools in the same exhaustive manner as reference encoders used in the tests, at the very least, in the initial deployments. On the other hand, reference encoders typically do not include perceptual optimizations such as the ones provided in commercial encoders. Such perceptual optimizations are known to have a substantial impact on reducing bitrates while maintaining the same visual quality (note, that such optimizations may be codec agnostic and not necessarily affect compression performance difference between two tested codecs with similar level of optimizations). Such an effect can be seen in results reported for VVenC encoder (see **Table 6-7**) where the difference between VVenC and VTM was due to employed perceptual optimizations such as adaptive quantization in the former.

7.9.1.2 UHDTV 2K applications - expected bitrates

Methodology 1: VVC bitrates estimated from HEVC bitrates

Bitrate ranges reported for HEVC in UHDTV 2K applications roughly align in the range of 3-15 Mbit/s. Assuming 40%-50% expected performance gains provided by VVC, an operating bitrate range for VVC can be estimated to be between 1.5-2 Mbit/s for the low-end bitrates and 7.5-9 Mbit/s for the high-end bitrates.

Methodology 2: VVC bitrates estimated from subjective assessment tests data

Test sample is smaller than one available for UHDTV 4K dataset, yet the results show that for relatively easy content VVC can obtain very good quality already at bitrates <1 Mbit/s and for medium to high complexity content at around 3Mbit/s.

7.9.1.3 UHDTV 8K applications – expected bitrates

Methodology 1: VVC bitrates estimated from HEVC bitrates

Range of reported bitrates for UHDTV 8K services in **Table 7-10** is relatively broad and ranges from around 30 Mbit/s to over 100 Mbit/s. This may result from different requirements for 8K services used in deployed services and trials and available data rates offered by employed transmission schemes. While HEVC enables 8K transmission cost in bandwidth may still present prohibitive. Assuming 40% coding performance gains offered by VVC (see **Table 6-7** and [38]) the equivalent bitrate range equates between 20Mbit/s to 60Mbit/s. Such an operating range aligns with data rates offered by legacy transmission schemes (i.e., <40 Mbit/s and 65 Mbit/s). Analysis in [37] reported expected data rates available for 8K transmission to be around 35Mbit/s.

Methodology 2: VVC bitrates estimated from subjective assessment tests data

Subjective assessment tests for UHDTV 8K were reported in [38] (see **Table 6-7**). A different rating methodology than in MPEG VVC verification test was used and authored investigated which of the tested bitrate points offered the same level of quality as uncompressed 8K video. Results showed very high dependency on the complexity of content used in the test. Reported bitrates for the same quality as uncompressed sources varied between 11Mbit/s to 180Mbit/s. The upper bound reported in the test exceeds practical limits for commercial deployments (and for reported HEVC deployments) therefore it could be considered not a valid indication of operational range. The lower bound may be relevant for low complexity content.

7.9.2 VVC bitrate ladder

This section concerns video delivery bitrates used in HTTP adaptive streaming (HAS) systems. In HAS applications video content is available to client devices through several renditions of the same video essence encoded at different resolutions and bitrates such that client device can switch between low- and high-quality renditions based on available bandwidth. Such a set of content resolutions and bitrates is often referred to as bitrate ladder.

One reference point for video bitrate ladder is Apple's HLS content authoring specification [18] which includes a set of recommended resolutions and bitrates for H.264/AVC and H.265/HEVC. This specification is often used as a reference point for alternative ladders proposals including for VVC. **Table 7-11** includes an example of reported VVC bitrate ladder proposal. Other proposals and studies of bitrate ladder for VVC were reported in [87] and [89].

An alternative to a fixed bitrate ladder, i.e., where any content in a particular format is encoded with the same operating bitrate (e.g., 10 Mbit/s for 3840x2160p30 with VVC), is so called Content-adaptive encoding (CAE) or per-title encoding. In the latter approach, bitrate ladder is customized such that renditions bitrates and switching points are set in attempt to provide optimal quality across the whole range of bitrates. Derivation of such customized and optimized CAE bitrate ladders is beyond the scope of this document.

Data from VVC subjective assessment tests (see **6.5.2**) and analysis of operating bitrates in **7.9.1** indeed show that in some cases UHDTV 4K content can be delivered with bitrates < 5Mbit/s. Independent analysis of VVC bitrate ladder in comparison to previous generation codecs based on convex-hull computations reported in [89] and in [90] showed that VVC allows for better use of the higher resolution than the other codecs, by providing a wider range of bitrates that are able to stream that resolution. The benefit for end users the benefit to streaming service providers is twofold: the ability to improve service quality to end users as described above, and additionally to increase scale at the same overall network capacity.

Table 7-11 HLS HEVC bitrate ladder and examples of reported VVC bitrate ladders.

	HEVC HLS bitrate ladder [18]	**Fixed VVC bitrate ladder after [88]**
Resolution	Bitrate [kbit/s]	
3840 x 2160	20 000	10 000
3840 x 2160	13 900	7 000
2560 x 1440	9 700	5 200
1920 x 1080	7 000	4 000
1920 x 1080	5 400	3 100
1280x720	4 080	2 400
1280x720	2 900	1 800
960×540	1 930	1 300
960×540	1 090	790
960×540	730	530
768×432	360	260
640×360	160	130

8 Annex A: Bitrate estimation based on subjective assessment test data

Annex A presents an analysis based on subjective assessment of VVC quality reported in MPEG verification test reports (see **Table 6-7**). UHDTV 4K tests were very extensive and included SDR and both HDR formats, HDR PQ and HDR HLG. MPEG performed VVC verification test using an 11-grade MOS scale [10] which is preferred over 5-grade MOS scale for assessment of low bitrate video codecs [9]. Subjective assessment was conducted with 4 to 5 bitrate points across the whole quality range according to the methodology in Rec. ITU-R BT.500 [10]. However, in order to estimate broadcast quality operating points, a quality range between MOS≥7 (or MOS≥4 in 5-grade MOS scale – see **Table 8-1**) may be most relevant [83]. Bitrates and MOS scores for HEVC and VVC were interpolated using the same cubic interpolation method as used in MPEG verification tests. **Figure 8-1** presents a scatter plot of HEVC and VVC bitrates for MOS=7 and **Figure 8-2** presents a similar plot for MOS=8.

Table 8-1 Mapping between 5-grade MOS scale and 11-grade MOS scale based on [9] (Table B.2).

5-grade MOS scale		11-grade MOS scale		
Quality / Impairment	Score	Score	Impairment	
		10	Imperceptible	
Excellent / Imperceptible	5	9	Slightly perceptible	somewhere
		8		everywhere
Good / Perceptible, but not annoying	4	7	Perceptible	somewhere
		6		everywhere
Fair / Slightly annoying	3	5	Clearly perceptible	somewhere
		4		everywhere
Poor / Annoying	2	3	Annoying	somewhere
		2		everywhere
Bad / Very annoying	1	1	Severely annoying	somewhere
		0		everywhere

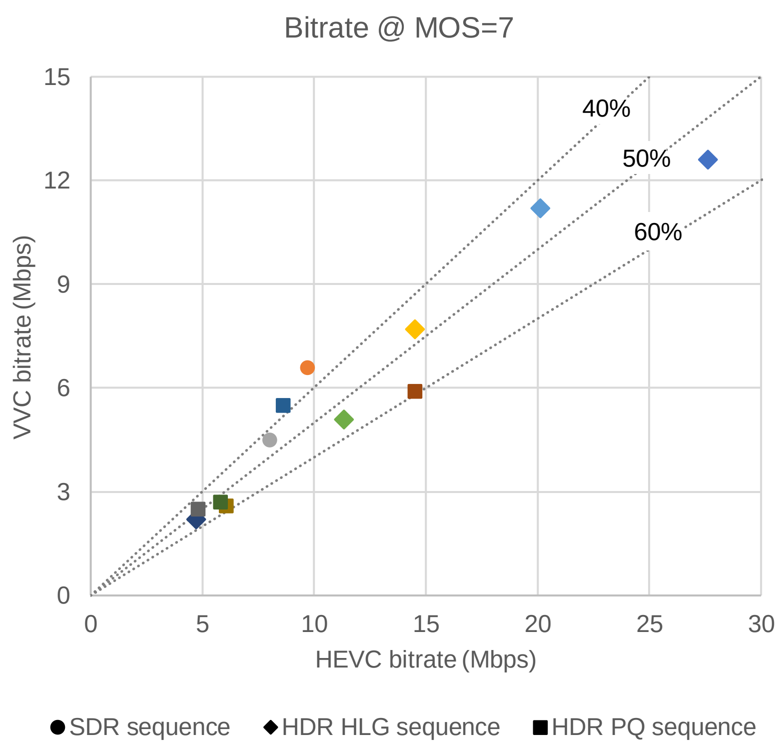

Figure 8-1 Scatter plot of HEVC and VVC bitrates for the same MOS quality (MOS=7). Dashed lines indicate corresponding ranges of bitrate reduction (40%, 50% and 60%).

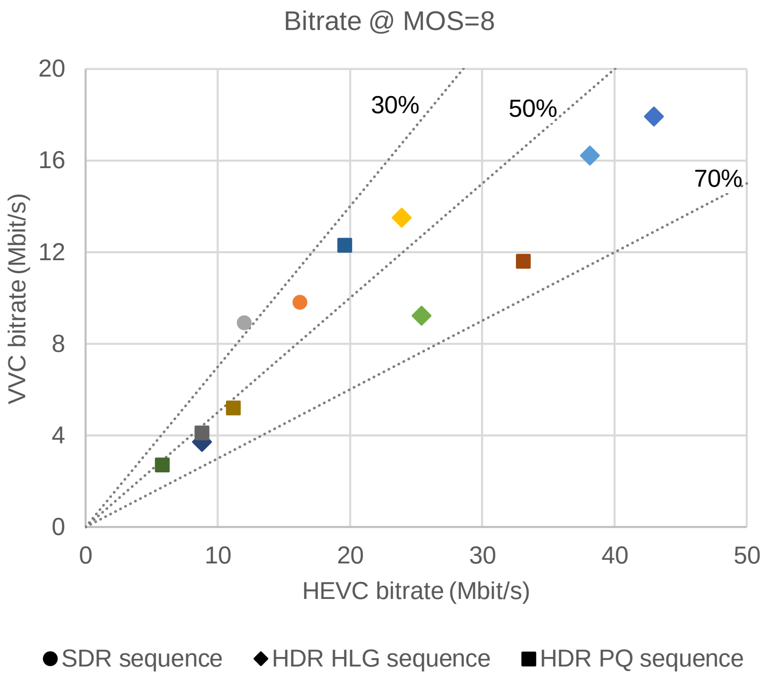

Figure 8-2 Scatter plot of HEVC and VVC bitrates for the same MOS quality (MOS=8). Dashed lines indicate corresponding ranges of bitrate reduction (30%, 50% and 70%).

Table 8-2 reports MOS scores for test content in MPEG VVC HDTV SDR verification test (see **Table 6-7** and [35]).

Table 8-2 HEVC and VVC bitrates for HDTV 2K SDR content at MOS=7 and MOS=8 based on subjective assessment tests reported in [35].

	HEVC bitrate [Mbit/s]		VVC bitrate [Mbit/s]	
Sequence	MOS=7	MOS=8	MOS=7	MOS=8
BarScene	0.8	-	0.4	0.5
DrivingPOV	4.6	7.2	2.2	3.2
Meridian	-	-	0.5	-
Metro	3.2	5.5	1.8	2.8

9 References

[1] ITU-T | ISO/IEC, “Rec. ITU-T H.266 | ISO/IEC 23090-3:2022 “Information technology -- Coded representation of immersive media -- Part 3: Versatile Video Coding”, “2022.

[2] ITU-T | ISO/IEC, “Rec. ITU-T H.274 | ISO/IEC 23002-7:2022 “Versatile supplemental enhancement information messages for coded video bitstreams” ,” 2022.

[3] ITU-T | ISO/IEC, “Rec. ITU-T H.262 | ISO/IEC 13818-2:2013 – Information technology -- Generic coding of moving pictures and associated audio information-- Part 2: Video”,” 2013.

[4] ITU-T | ISO/IEC, “Rec. ITU-T H.264 | ISO/IEC 14496-10:2014, “Information technology -- Coding of audio-visual objects-- Part 10: Advanced Video”, “2014.

[5] ITU-T | ISO/IEC, “Rec. ITU-T H.265 | ISO/ 23008-2:2020, “Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 2: High efficiency video coding”,” 2020.

[6] Bross, B., et al., "Overview of the Versatile Video Coding (VVC) Standard and its Applications,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 10, October 2021 pp. 3736-3764.

[7] ISO/IEC JTC 1/SC 29/AG 3, “White paper on Versatile Video Coding (VVC)”, N0035, 2021.

[8] ITU-T | ISO/IEC, “Rec. ITU-T H.222.0 | ISO/IEC 13818-1:2022 “Information technology -- Generic Coding of moving pictures and associated audio information -- Part1: Systems”,” 2022

[9] ITU-T, “Rec. ITU-T P.910: Subjective video quality assessment methods for multimedia applications,”, 2022.

[10] ITU-R, “Rec. ITU-R BT.500: Methodologies for the subjective assessment of the quality of television images,” 2019.

[11] ITU-R, “Rec. ITU-R BT.709: “Parameter values for the HDTV standards for production and international programme exchange”,” 2015.

[12] ITU-R, “Rec. ITU-R BT.2020: “Parameter values for ultra-high definition television systems for production and international programme exchange”,” 2015.

[13] DVB “Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications,” *DVB Document A001 Rev.21*, November 2022. [replace with ETSI reference]

[14] ITU-R, “Rec. ITU-R BT.2100: “Image parameter values for high dynamic range television for use in production and international programme exchange”,” 2018.

[15] ISO/IEC, “ISO/IEC 23000-19:2019/FDIS: “Information technology -- Multimedia application format (MPEG-A) -- Part 19: Common media application format (CMAF) for segmented media (2nd edition)”,” 2022.

[16] IETF, “HTTP Live Streaming (HLS)”, RFC 8216.

[17] ISO/IEC, “ISO/IEC 14496-12:2022: "Information technology -- Coding of audio-visual objects -- Part 12: ISO base media file format".

[18] Apple, “HTTP Live Streaming (HLS) authoring specification for Apple devices”.

[19] ISO/IEC, “ISO/IEC: 14496-15:2022: “Information technology -- Coding of audio-visual objects -- Part 15: Carriage of network abstraction layer (NAL) unit structured video in ISO base media file format”,” 2022.

[20] ISO/IEC, “ISO/IEC 23009-1:2022: “Information technology -- Dynamic adaptive streaming over HTTP (DASH) -- Part 1: Media presentation description and segment formats”,” 2022.

[21] ISO/IEC, “ISO/IEC 23008-12:2022: “Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 12: Image File Format”,” 2022.

[22] ISO/IEC, “ISO/IEC 23090-2:2023: “Information technology -- Coded representation of immersive media -- Part 2: Omnidirectional media format”,” 2023.

[23] 3GPP, “3GPP TR: 26.955: Video codec characteristics for 5G-based services and applications,” June 2022.

[24] 3GPP, “3GPP TR: 26.925: Typical traffic characteristics of media services on 3GPP networks,”, March 2022.

[25] EBU, “TR 036, TV programme accommodation in a DVB-T2 multiplex for (U)HDTV with HEVC video coding, v1.0, March 2016.

[26] ITU-T | ISO/IEC, “Rec. ITU-T H Suppl. 15 | ISO/IEC TR 23008-14:2018 “Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 14: Conversion and coding practices for HDR/WCG Y'CbCr 4:2:0 video with PQ transfer characteristics”, “2018.

[27] ITU-T | ISO/IEC, “Rec. ITU-T H Suppl. 18 | ISO/IEC TR 23008-15:2018 “Information technology-- High efficiency coding and media delivery in heterogeneous environments-- Part 15: Signalling, backward compatibility and display adaptation for HDR/WCG video”,” 2018.

[28] ITU-T | ISO/IEC, “Rec. ITU-T H.273 | ISO/IEC 23091-2:2021 “Coding-independent code points for video signal type identification” ,” 2021.

[29] ITU-T | ISO/IEC, "Rec. ITU-T H Suppl. 19 | ISO/IEC TR 23091-4:2021 "Information technology -- Coding-independent code points -- Part 4: Usage of video signal type code points"," 2021.
[30] ITU-R, "Rec. ITU-R BT.2073-2: "Use of high efficiency video coding for UHDTV and HDTV broadcasting applications"," 2022.
[31] ITU-R, "Rec. ITU-R BT.2343-8: "Collection of field trials of UHDTV over DTTB networks"," 2023.
[32] UHD Forum, "Guidelines v3: Ultra HD Distribution", 2023.
[33] ITU-T | ISO/IEC Joint Video Experts Team (JVET), "VVC verification test report for UHD SDR video content," *JVET-T2020,* October 2020.
[34] ITU-T | ISO/IEC Joint Video Experts Team (JVET), "VVC verification test report for high dynamic range video content," *JVET-W2020,* July 2021.
[35] ITU-T | ISO/IEC Joint Video Experts Team (JVET), "VVC verification test report for HD SDR and 360° video content," *JVET-V2020,* April 2021.
[36] Ohm, J-R., "Meeting Report of the 30th Meeting of the Joint Video Experts Team (JVET), by teleconference, 21–28 April 2023", JVET-AD1000-v1, April 2023
[37] Biatek, T, et al., "Future MPEG standards VVC and EVC: 8K broadcast enabler," IBC, 2020.
[38] Bonnineau, C., et al., "Perceptual Quality Assessment of HEVC and VVC Standards for 8K Video. IEEE Transactions on Broadcasting, 2022, 68 (1), pp.246-253.
[39] Nemoto, S., "[AHG13] Compression performance analysis for 8K HLG sequences", *JVET-P0616,* October 2019.
[40] Larbier, P., "Using 10-bit AVC/H.264 Encoding with 4:2:2 for Broadcast Contribution", available at: https://www.x266.nl/x264/10bit_01-ateme_pierre_larbier_422_10-bit.pdf
[41] Enhorn, J., et al. "Improvements to a Temporal Filter for Video Coding", 2022 IEEE International Symposium on Circuits and Systems (ISCAS), May. 2022.
[42] Andersson K., et al., "Addition of a GOP hierarchy of 32 for random access configuration for VTM", JVET-S0180, June 2020.
[43] Vanne, J., et al., "Comparative study of 8 and 10-bit HEVC encoders," 2014 IEEE Visual Communications and Image Processing Conference, Valletta, Malta, 2014, pp. 542-54
[44] Jung, J., et al., "Quality evaluation of internal 10-bit versus 8-bit processing with 8-bit source content" JVET-Z0107, April 2022.
[45] Sugito, Y., et al., "A Study on the Required Video Bit-rate for 8K 120-Hz HEVC/H.265 Temporal Scalable Coding," 2018 Picture Coding Symposium (PCS), San Francisco, CA, USA, 2018, pp. 106-110
[46] Ichigaya, A. and Nishida, Y., "Required bit rates analysis for a new broadcasting service using HEVC/H.265," IEEE Transactions on Broadcasting, vol. 65, no. 2, 2016, pp.417-425.
[47] IETF, "RTP Payload Format for Versatile Video Coding (VVC)", RFC 9328, 2022.
[48] Broadcasting System Committee, "Report of the Terrestrial Digital Broadcasting System Advancement Working Group", https://www.soumu.go.jp/main_content/000870940.pdf, March 2023.
[49] ATSC, "ATSC Planning Team Report: ATSC 3.0 and Global Convergence", https://www.atsc.org/atsc-documents/pt-6-report-atsc-3-0-and-global-convergence/ , December 2022.
[50] CTA Specification, Web Application Video Ecosystem – Content Specification, CTA-5001-D," 2021.
[51] DASH Industry Forum, "DASH-IF Interoperability Points; Part 7: Video," May 2022.
[52] Power, J., "New video codecs for DVB services: fast-forward to 2024…," *DVB Scene,* Issue 58, September 2021.
[53] DVB, "Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications," *DVB* BlueBook A001r21 (Interim draft TS 101 154 V2.8.1), November 2022.
[54] DVB, "DVB MPEG-DASH Profile for Transport of ISO BMFF Based DVB Services over IP Based Networks," DVB BlueBook A168r6 (Interim draft TS 103 285 V1.4.1), October 2022.
[55] DVB, "VVC test content", https://dvb.org/specifications/verification-validation/vvc-test-content/, March 2023.
[56] ETSI, "Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications", ETSI TS 101 154 V2.8.1, 2023
[57] SCTE, "VVC Video Constraints for Cable Television Part 1- Coding", SCTE 281-1 2023, 2023.
[58] SCTE, "VVC Video Constraints for Cable Television Part 2- Transport", SCTE 281-2 2023, 2023.
[59] Fautier, T. and Mazieres, E., France Television Lab blog post "8K EXPERIMENT AT ROLAND GARROS 2019" (https://www.francetelevisions.fr/lab/projets/8K-Experiment-at-Roland-Garros-2019)
[60] ISO/IEC JTC 1/SC 29/AG 5, "Guidelines for Verification Testing of Visual Media Specifications", N35, October 2021.
[61] ITU-T | ISO/IEC, "ITU-T Technical Paper HSTP-VID-WPOM | ISO/IEC TR 23002-8:2021" Working practices using objective metrics for evaluation of video coding efficiency experiments"," 2020.
[62] Netflix, "VMAF – Video Multi-Method Assessment Fusion," version 1.3.5, repository, Aug. 2020. https://github.com/Netflix/vmaf/tags

[63] ITU-T | ISO/IEC Joint Video Experts Team (JVET), "VTM and HM common test conditions and evaluation procedures for HDR/WCG video," *JVET-AC2011*, January 2023.
[64] Skupin, R., et al., "Open GOP resolution switching in HTTP adaptive streaming with VVC," *Proc. 35th Picture Coding Symp. (PCS)*, June 2021, pp. 1–5.
[65] ITU-T H.266.2 Reference software for ITU-T H.266 versatile video coding.
[66] ITU-T | ISO/IEC Joint Video Experts Team (JVET), "Algorithm description for Versatile Video Coding and Test Model 20 (VTM20)," JVET-AD2002, April 2023
[67] SMPTE ST 2086:2018: Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images
[68] Zink, M. and Smith, M. D., "Managing HDR content production and display device capabilities", Proc. International Broadcasting Convention 2015 (IBC 2015), Amsterdam, NL, Sep. 2015
[69] ETSI 101 154: Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications.
[70] SMPTE ST 2094-10:2016: Dynamic Metadata for Color Volume Transform – Application #1
[71] SMPTE ST 2094-40:2020: Dynamic Metadata for Color Volume Transform - Application #4
[72] ETSI TS 103 433-2 (SL-HDR2)
[73] SMPTE RDD 5:2006 - SMPTE Registered Disclosure Doc - Film Grain Technology — Specifications for H.264 | MPEG-4 AVC Bitstreams
[74] Kurihara, T., et al., "Analysis of sharpness increase by image noise," Proc. SPIE 7240, Human Vision and Electronic Imaging XIV, 724014 (10 February 2009)
[75] Wan, X., et al., "Improvement in perception of image sharpness through the addition of noise and its relationship with memory texture," in Human Vision and Electronic Imaging XX, SPIE, Mar. 2015
[76] Johnson, G. M. and Fairchild, M. D., "Sharpness Rules," in Color and Imaging Conference, Jan. 2000
[77] ITU-T | ISO/IEC Joint Video Experts Team (JVET), "Film grain synthesis technology for video applications (Draft 4)", JVET-AC2020, Jan 2023
[78] Versatile film grain synthesis (VFGS) GitHub, https://github.com/InterDigitalInc/VersatileFilmGrain
[79] Film Grain Synthesis Library, https://github.com/ittiam-systems/libfgs
[80] The Explorers, https://theexplorers.com/landing?culture=en
[81] ITE, https://www.ite.or.jp/content/test-materials/
[82] Fraunhofer HHI, 8K Berlin Test Sequences, https://www.hhi.fraunhofer.de/en/departments/vca/research-groups/video-coding-systems/8k-sequences.html
[83] SBTVD Forum, TV3.0 Call for proposals, https://forumsbtvd.org.br/wp-content/uploads/2020/07/SBTVDTV-3-0-CfP.pdf
[84] Ducloux, X., Diascorn, J-L., Fautier, T., "Exploring the benefits of dynamic resolution encoding and support in DVB standards," IBC, 2022.
[85] Andersson, K., et al., "AHG10: GOP-based RPR encoder control", *JVET-AB0080-v2*, October 2022.
[86] Guionnet, T., et al., "Perceptual quality evaluation of variable resolution VVC video coding using RPR", in Proceedings of the 2nd Mile-High Video Conference (MHV '23). Association for Computing Machinery, New York, NY, USA, 14–20, 2023.
[87] Baraghi, T., et al., "Multi-codec ultra high definition 8K MPEG-DASH dataset", in proceedings of the 13th ACM Multimedia Systems Conference (MMSys '22), pp. 216–220, 2022.
[88] Iwamura, S., et al., "Verification test plan for VVC multilayer coding" JVET-AD2021, April 2022.
[89] Kaafarani, R., et al., "Evaluation Of Bitrate Ladders For Versatile Video Coder". In VCIP 2021 - IEEE Visual Communications and Image Processing, Dec 2021, Munich, Germany. pp.1-5.
[90] Wieckowski, A., et al., "VVC in the cloud and browser playback: it works", In Proceedings of the 1st Mile-High Video Conference (MHV '22). Association for Computing Machinery, New York, NY, USA, 19–24.

