1 Scope

This document presents recommended practices regarding the TV 3.0 Video Coding, defined in ABNT NBR 25603.

2 References

The following documents are cited in the text in such a way that their contents, in whole or in part, constitute requirements for this document. For dated references, only the editions cited apply. For undated references, the most recent editions of that document (including amendments) apply.

ABNT NBR 25603, *TV 3.0 – Video Coding*

ABNT NBR 15608-2, *Digital terrestrial television – Operational guideline - Part 2: Video coding, audio coding and multiplexing — Guideline for ABNT NBR 15602:2007 implementation*

3 Terms and Definitions

For the purposes of this Document, the following terms and definitions apply.

TV 3.0
Digital Terrestrial Television system defined in the suite of standards ABNT NBR 25601 to ABNT NBR 25609, also known as DTV+.

Ultra-High Definition Television 1
UHDTV-1
Ultra-High Definition with a resolution of 2 160 lines with 3 840 pixels each and a fixed aspect ratio of 16:9, also known as 4K.

Ultra-High Definition Television 2
UHDTV-2
Ultra-High Definition with a resolution of 4 320 lines with 7 680 pixels each and a fixed aspect ratio of 16:9, also known as 8K.

4 Abbreviations

For the purposes of this Document, the following abbreviations apply.

2D	*Two-Dimensional*
3D	*Three-Dimensional*
ASPS	*Atlas Sequence Parameter Set*
AVC	*Advanced Video Coding*
CMAF	*Common Media Application Format*

DASH	*Dynamic Adaptive Streaming over HTTP*
DMI	*Dynamic Mapping Information*
DRE	*Dynamic Resolution Encoding*
GOP	*Group Of Pictures*
HDR	*High Dynamic Range*
HEVC	*High Efficiency Video Coding*
HFR	*High Frame Rate*
HLG	*Hybrid Log-Gamma*
HLS	*HTTP Live Streaming*
IDR	*Instantaneous Decoding Refresh*
IP	*Internet Protocol*
IRD	*Integrated Receiver Decoder*
IRAP	*Intra Random Access Point*
ISOBMFF	*International Organization for Standardization Base Media File Format*
JVET	*Joint Video Experts Team*
LCEVC	*Low Complexity Enhancement Video Coding*
MaxCLL	*Maximum Content Light Level*
MaxFALL	*Maximum Frame Average Light Level*
MIV	*MPEG Immersive Video*
MPI	*Multi-Plane Image*
NAL	*Network Abstraction Layer*
NR	*Nominal Resolution*
OTA	*Over-The-Air*
OTT	*Over-The-Top*
PQ	*Perceptual Quantizer*
PQ10	*Perceptual Quantizer, 10-bit quantization*

PPS	*Picture Parameter Set*
RPR	*Reference Picture Resampling*
SDR	*Standard Dynamic Range*
SEI	*Supplemental Enhancement Information*
SL-HDR	*Single Layer HDR*
SPS	*Sequence Parameter Set*
UHDTV	*Ultra-High Definition Television*
V3C	*Visual Volumetric Video-based Coding*
VCL	*Video Coding Layer*
V-PCC	*Video-based Point Cloud Compression*
VUI	*Video Usability Information*
VVC	*Versatile Video Coding*

5 Overview

The operational guidelines corresponding to the technologies used in the TV 3.0 Video Coding are contained in the following Annexes:

- Annex A contains the Versatile Video Coding (VVC) guidelines;
- Annex B contains the High Efficiency Video Coding (HEVC) guidelines;
- Annex C contains the Advanced Video Coding (AVC) guidelines;
- Annex D contains the Dynamic Resolution Encoding (DRE) guidelines;
- Annex E contains the Low Complexity Enhancement Video Coding (LCEVC) guidelines;
- Annex F contains the HDR10 Media Profile (HDR10) guidelines;
- Annex G contains the Dynamic Metadata for Color Volume Transform - Application #1 (Dolby Vision) guidelines;
- Annex H contains the Dynamic Metadata for Color Volume Transform - Application #4 (HDR10+) guidelines;
- Annex I contains the Single Layer HDR system Part 2 (SL-HDR2) guidelines;
- Annex J contains the Single Layer HDR system Part 1 (SL-HDR1) guidelines;
- Annex K contains the Hybrid Log-Gamma (HLG) guidelines; and
- Annex L contains the Visual Volumetric Video-based Coding (V3C) guidelines.

Annex A

Versatile Video Coding (VVC)

A.1 Scope

This Annex provides Operational Guidelines for the use of Versatile Video Coding (VVC). VVC is also known as ITU-T H.266. The information presented can be found in the Media Coding Industry Forum (MC-IF), VVC Technical Guidelines (v1.0).

A.2 MC-IF VVC Technical Guidelines

The MC-IF VVC technical guidelines are a reference for VVC configuration choices to address operational and interoperability needs while achieving best compression performance. The SBTVD TV3.0 VVC operational guidelines described in this annex reference the MC-IF VVC Technical Guidelines and indicate chapter by chapter which information is relevant for TV3.0.

Section 1 is the disclaimer.

Section 2 is motivation and scope.

Section 3 contains tables and figures.

Section 4 contains terms and abbreviations.

Section 5 contains the introduction.

Section 6 provides a general overview of the VVC standard and its usage and configuration aspects including pre-processing, encoding and post-decoding processes. The section should be read in mind that TV3.0 has selected the single layer VVC Main 10 profile, HDR10 with 4:2:0 chroma sampling for video coding.

Section 6.1 provides background information regarding VVC including description of VVC profiles, bitstream structure, selected tools and features.

Section 6.2 provides information on how an 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.

Section 6.3 provides information on how a video bitstream can be constructed according to operational constraints. It covers temporal structures within a GOP, open and closed GOP configurations, reference picture resampling, encoding of HDR content and encoding of 8-bit content in a 10-bit container.

Section 6.4 provides information on post-decoding, notably SEI messages for HDR and Film Grain Synthesis. TV3.0 supports the Mastering Display Colour Volume (MDCV) and the Content Light Level Information (CLLI) SEI messages.

Section 6.5 provides objective and subjective performance data.

Section 7 provides guidelines for broadcast and streaming applications. The section should be read in mind that TV3.0 has selected the single layer VVC Main 10 profile, HDR10 with 4:2:0 chroma sampling for video coding and ROUTE/DASH for transport.

Section 7.1 describes video workflows within an end-to-end broadcast and streaming ecosystem, key properties of typical video formats and relevant VVC profiles.

Section 7.2 provides an overview of VVC support in media systems and transport standards and relevant broadcast and streaming application specifications. Sub-sections on MPEG ISO Base Media File Format, MPEG DASH and MPEG CMAF are relevant.

Section 7.3 presents high level signaling for video formats supported in relevant broadcast and streaming application specifications. Information in this section is an example and the signaling specified in TV3.0 Video Coding Technical Standard shall be the reference.

Section 7.4 describes VVC bitstream configuration options and impacts related to random access functionality and channel change delays.

Section 7.6 describes VVC support for dynamic resolution changes in a linear broadcast delivery service and improved coding efficiency in adaptive bitrate streaming.

Section 7.9 presents expected bitrate ranges for broadcast and streaming video delivery using VVC as well as the methodology used to achieve these bitrates.

Sections 7.5, 7.7 and 7.8 are placeholders to cover random access and zapping time, frame rate change and spatial scalability and will be provided in a next version of the guidelines.

Annexes provide supplementary information.

Annex B

High Efficiency Video Coding (HEVC)

B.1 Scope

This Annex provides Operational Guidelines for the use of High Efficiency Video Coding (HEVC). HEVC, High Efficiency Video Coding, is an international video coding standard jointly developed by ISO/IEC MPEG and ITU-T Study Group 16 VCEG as ITU-T H.265 or ISO/IEC 23008-2 and published in June 2013.

For the current version of this document, no further information is provided regarding the usage of HEVC on TV 3.0.

Annex C

Advanced Video Coding (AVC)

C.1 Scope

This Annex provides Operational Guidelines for the use of Advanced Video Coding (AVC). AVC, Advanced Video Coding, is an international video coding standard jointly developed by ISO/IEC MPEG and ITU-T Study Group 16 VCEG as ITU-T H.264 or ISO/IEC 14496-2 and published in August 2004.

For the current version of this document, no further information is provided regarding the usage of AVC on TV 3.0.

Annex D

Dynamic Resolution Encoding (DRE)

D.1 Scope

This Annex provides Operational Guidelines for the video coding when Dynamic Resolution Encoding is used.

D.2 TV30 VVC DRE HDR (HFR) UHDTV-1 Bitstreams

The Broadcast Bitstreams are built as described in Annex A.2 of ABNT NBR 25603. A DRE bitstream is characterized by the possible change of the encoded resolution at DASH segment boundaries. The best resolution is selected in the list of resolutions specified in the Annex A.2.3.1.3 or A.2.3.2.3 of ABNT NBR 25603. The best resolution for encoding a video segment depends on the video content characteristics and the selection process is left free to the encoder manufacturer. Depending on the use-cases and requirements that are foreseen, the selection can be made without encoding all possible resolutions (a-priori selection) or after encoding all possible resolutions (a-posteriori selection).

As the video content can vary a lot from one scene to another or even within a scene when using a camera pan or when big occlusion happens, any restriction introduced on the resolution changes as well as on the frequency of the changes will affect the performance of DRE. Resolution changes may happen at two consecutive segments.

Reference Picture Resampling (RPR) may be used in VVC DRE Bitstreams to take advantage of Open-GOP encoding but is not mandatory.

For the best results in a Live scenario with no extra CPU usage or no latency constraints compared to an encoding with a constant resolution, it is recommended to associate DRE with adaptive duration of segments, to get resolution changes aligned with scene cuts.

Same specificities apply to OTT DRE Bitstreams as the ones described for Broadcast Bitstreams in this section. DRE can be applied on one or multiple video representations.

D.3 TV30 VVC DRE HDR (HFR) UHDTV-2 Bitstreams

The Broadcast Bitstreams are built as described in Annex A.2 of ABNT NBR 25603. The same DRE specificities as the ones described in section D.2 of this Annex apply. The best resolution is selected in the list of resolutions specified in the Annex A.2.3.3.3 or A.2.3.4.3 of ABNT NBR 25603.

Same specificities apply to OTT DRE Bitstreams as the ones described for Broadcast Bitstreams in this section. DRE can be applied on one or multiple video representations.

D.4 TV30 LCEVC VVC DRE HDR (HFR) UHDTV-1 Bitstreams

The Broadcast Bitstreams are built as described in Annex E.2 of ABNT NBR 25603. The two-layers LCEVC VVC DRE bitstreams, for instance, can be built in two different ways:

- the base layer is limited to an HD resolution selected among 1920x1080, or 1280x720 resolutions and LCEVC applies a two-dimensional 2:1 scaling across both dimensions to reconstruct the UHDTV-1 signal up to 3840x2160 resolution,

- no limitation is set for the resolution of the base layer. The base layer resolution is selected among 3840x2160, 2560x1440, 1920x1080 and 1280x720 resolutions and LCEVC applies a two-dimensional 2:1 scaling across both dimensions to reconstruct the UHDTV-1 signal when an HD resolution has been chosen for the base layer and no up-sampling when a UHD resolution has been chosen.

The second scenario can offer a better overall UHD performance by avoiding any strong limitation of the base layer resolution, while the first scenario offers backward compatibility with VVC HDTV only receivers if they exist.

As for the VVC streams in section D.2 of this Annex, the selection process of the best resolution combination for the two-layers is left free to the encoder manufacturer. Depending on the use-cases and requirements that are foreseen, the selection can be made without encoding all possible resolutions (a-priori selection) or after encoding all possible resolutions (a-posteriori selection).

Same specificities apply to OTT DRE Bitstreams as the ones described for Broadcast Bitstreams in this section. DRE can be applied on one or multiple video representations.

D.5 TV30 LCEVC VVC DRE HDR (HFR) UHDTV-2 Bitstreams

The Broadcast Bitstreams are built as described in Annex E.2 of ABNT NBR 25603. There is a high probability that the UHDTV-2 Broadcast Bitstreams are delivered through two transponders. The first transponder delivers a UHD 4K VVC signal and the second transponder delivers the LCEVC UHD 8K enhancement layer.

The bitstream of the first transponder can be built in two different ways:

- the bitstream can be a VVC DRE up to 4K bitstream by selecting the best resolution among 3840x2160, 2560x1440 or 1920x1080 resolutions,

- the bitstream can be a LCEVC VVC DRE up to 4K bitstream, as described in section D.4 of this Annex.

In both cases, the bitstream of the second transponder carrying the LCEVC UHD 8K enhancement layer applies a dyadic upsampling ratio to reconstruct the UHDTV-2 signal up to 7680x4320 resolution.

Same specificities apply to OTT DRE Bitstreams as the ones described for Broadcast Bitstreams in this section. DRE can be applied on one or multiple video representations.

Annex E

Low Complexity Enhancement Video Coding (LCEVC)

E.1 Scope

This Annex provides Operational Guidelines for the use of Low Complexity Enhancement Video Coding (LCEVC). MPEG-5 Part 2 Low Complexity Enhancement Video Coding (LCEVC), ISO/IEC 23094-2, is a video coding technology and official MPEG standard published by ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission).

E.2 Introduction

LCEVC is designed for use in conjunction with existing video codecs, leveraging advanced enhancement tools and encoder-driven upsampling for encoding "residuals," i.e., the difference between the original video and its predicted representation. LCEVC can improve compression efficiency and reduce overall computational complexity for a given resolution and bit-depth, and is designed for integration even with existing encoders and decoders.

This LCEVC Operational Guidelines document is organized as follows:

- Section E.3 is a history of the motivation for and development of LCEVC.
- Section E.4 describes the advantages of implementing LCEVC.
- Section E.5 provides a general overview of how the LCEVC standard works.
- Section E.6 provides technical guidelines for implementation.
- Section E.7 describes broadcast and streaming scenarios: LCEVC for HD, 4K and 8K delivery
- Section E.8 details profile and level configuration.
- Section E.9 provides information about the LCEVC bitstream structure.
- Section E.10 details LCEVC support of Supplementary Enhancement Information (SEI) and Video Usability Information (VUI).
- Section E.11 describes pre-encoding processes.
- Section E.12 describes post-decoding processes.
- Section E.13 provides MPEG-5 Part 2 LCEVC testing data.
- Section E.14 describes best practices for optimizing LCEVC implementation.
- Section E.15 describes carriage of LCEVC in broadcast and streaming applications.

E.3 Why LCEVC? Market needs and requirements

E.3.1 Market needs

In October 2018, 28 industry signatories presented a request to MPEG outlining market needs and requirements for a software-based capability extension on top of existing and future video codecs (MPEG membership required to view).

They claimed that “despite surging demand for video, it is often difficult – or prohibitively costly – to deliver the high video quality that most end users expect.” For instance, in most countries bandwidth availability is often insufficient for reliable delivery of OTT video content. Video sessions on mobile devices and in prime time “internet rush hours” are characterized by worse bandwidth availabilities.

They added that the combination of legacy devices and long replacement cycles “makes it difficult to upgrade video services to higher resolutions and frame rates (e.g., 1080p60, 4K, 8K) without either ignoring customers with legacy video devices or creating duplicate services for new devices. The consequent low availability of higher resolution services reduces the demand for newer decoder devices, in a vicious cycle.”

Signatories were industry leaders active in live and VoD streaming, videoconferencing, VR, broadcast video, and real-time video feeds for industrial applications.

LCEVC was driven by several commercial needs put forward to MPEG by many leading industry experts from various areas of the video delivery chain, from vendors to traditional broadcasters, from satellite providers to over-the-top (OTT) service providers and social media [4].

Service providers work with complex ecosystems. They make choices on codecs based on various factors, including maximum compatibility with their existing ecosystems, costs of deploying the technology (including royalty rates), etc. Sometimes they are forced to make certain choices. Whichever is the case, changing codecs cannot be done without relevant up-front investments and large amounts of time. Accordingly, having the possibility to upgrade an ecosystem without the need to replace it completely and still having the freedom to select a base codec of their choice is an important option that operators need to have.

Further, service operators, small and big alike, are increasingly concerned about the cost of delivering a growing number of services, often using decentralized infrastructures such as cloud-based systems or battery-powered edge devices. The need to increase the overall efficiency of video delivery systems must also be balanced with the seemingly conflicting needs to upgrade video resolutions and consume less power.

Finally, the “softwarization” of solutions across the technological spectrum has brought up the need to also have codec solutions which do not necessarily require a bespoke dedicated hardware for operating efficiently, but rather can operate as a software layer on top of existing infrastructures and deliver the required performances.

LCEVC seeks to solve the above issues by providing a solution that is compatible with existing (and future) ecosystems whilst delivering it at a lower computational cost than it would be otherwise possible with a tout-court upgrade.

Aside from rapidly improving the efficiency of legacy workflows, LCEVC can also improve the business case for the adoption of next-generation codecs, by combining their superior coding efficiency with.

E.3.2 MPEG requirements

In response to that request, MPEG issued a set of requirements for a new video coding standard and a Call for Proposals for those companies and organizations that have developed video compression technology that they believe address the requirements.

The Requirements define a codec that allows a full resolution encoded/decoded stream formed from enhancing a stream encoded/decoded with a hardware codec and a data stream which, when added to the coded/decoded stream, would bring the video to the full resolution.

In particular, the key performance requirements were defined as follows:

- When enhancing an n-th generation codec (e.g., AVC), compression efficiency for the aggregate stream is appreciably higher than that of the n-th generation MPEG codec used at full resolution and as close as possible to that of the (n+1)-th generation codec (e.g., HEVC) used at full resolution, at bandwidths and operating conditions relevant to mass market distribution; and

- Encoding and decoding complexity for the aggregate full resolution video (i.e., base plus enhancement) shall be comparable with that of the base encoder or decoder, respectively, when used alone at full resolution.

The key implementation and non-technical requirements for the video coding project were:

- The video stream should be decodable without specific firmware or OS support by all devices capable to decode the base codec, with equivalent resource utilization (e.g., processing power, battery consumption, etc.) as the base decoder at full resolution decoded in hardware;

- All web browsers should be able to decode high resolution video without plug-ins and/or browser upgrade, e.g., via HTML5 javascript;

- The additional data stream should be compatible with the existing ecosystem, e.g., ad insertion, metadata management, CDNs, DRM/CA and network protocols such as DASH, HLS, and MMT;

- The overall processing power requirement to encode a video stream should be comparable with that of the base codec when used alone at full resolution.

Multiple independent tests and cross-checks conducted during the MPEG collaborative phase of standardization demonstrated that LCEVC successfully satisfies its requirements (see Sections E.14 below and technical documentation and performance evaluations in the Resources section of the LCEVC website for a summary of tests performed and LCEVC performance results).

The LCEVC standard, (ISO/IEC 23094-2), was published in October 2021. Below a summary of the standardization timeline for LCEVC,

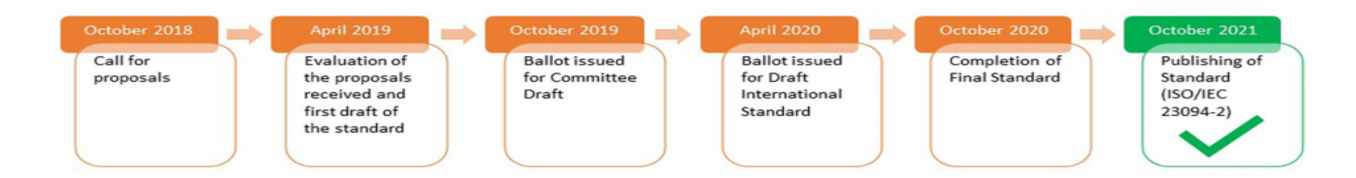

Figure E.1 – LCEVC Standardisation timeline

E.4 LCEVC's Technical and Commercial Benefits

LCEVC is an enhancement of other codecs, as opposed to a replacement, and as such, provides the optimal trade-off between compression and complexity for a given codec and/or application, with the potential in multiple use cases of reducing either bandwidth requirements for a given visual quality, or coding energy requirements, or both.

This enhancement is achieved by a combination of processing input video at a lower resolution via an existing single-layer codec, and using a simple, small set of highly specialized tools to correct impairments, upscale, and add details to the processed video. LCEVC's low-complexity requirement means that the tool definition process accounts for the availability of hardware acceleration for graphics processing available in existing chipsets, and it is especially amenable to optimized software implementations with low power consumption (e.g., using SIMD CPU, GPU, and heterogeneous parallel processing), as well as small-footprint silicon implementations.

LCEVC also enables a stream to signal adaptive dithering post-processing, which reduces banding and aliasing impairments, as well as providing a platform for the inclusion of user data within the bitstream at transform block level. This allows LCEVC to support the adoption of evolving image manipulation techniques within the standard, while still offering an efficient method of encoding residual data and providing up to mathematically lossless picture reconstruction. In addition, tiling options for the residual sub-layers enable parallel execution of entropy coding, striped ultra-low-latency use cases and region of interest decoding of the enhancement.

In terms of overall rate control accuracy, LCEVC data is encoded after the base layer, which allows it to compensate at least in part for accuracy issues in the coded size of the base layer (or base layer stripe), with consequent benefits in real-time, low-latency use cases. LCEVC can efficiently extend the features of a video service, even if some of the target devices do not support LCEVC decoding. LCEVC is codec- and hardware-agnostic. A video service implementing LCEVC can introduce new features (such as UHD, HDR, etc.) without disruption to devices that support only the base codec, and without requiring duplication of video workflows to separately serve new and legacy devices. Furthermore, when retrofitting a portion of existing target devices with LCEVC, the video service benefits even more from the upgrade to service quality, without the need to wait for a full replacement of the installed base of target devices.

Due to its multi-layer structure, LCEVC can also be used as a scalable codec, although it has not been primarily built as a scalable codec. Because of this unique feature, LCEVC can typically be used in two modes:

- Enhancement Mode: to provide an enhanced full resolution video - albeit starting from a lower resolution video which is separately and independently decodable; or

- Scalable Mode: to provide a scalable video, where the same receiver can switch between a lower and a higher resolution.

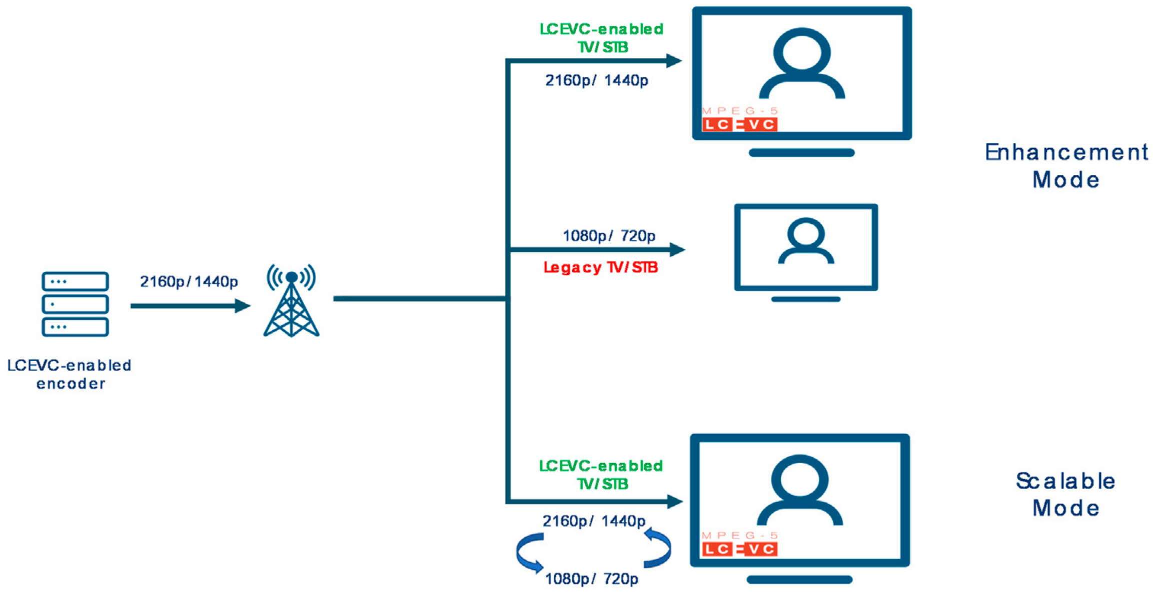

Figure E.2 – Two of the typical deployment modes for LCEVC

Indeed, current commercial deployments use LCEVC to provide an enhanced full resolution video on top of a base codec widely available in the ecosystem (e.g., H.264/AVC, H.265/HEVC or H.266/VVC), knowing that devices which, for one reason or another cannot decode the LCEVC component, can still decode the lower resolution base layer and display a video at a quality which is typically similar or higher than that they would otherwise receive.

This aspect allows for a phased deployment of LCEVC without affecting current services. For example, LCEVC enables deployments where a service provider intends to enhance existing codecs (e.g., H.264/AVC-H.265/HEVC - based services which can be upgraded to LCEVC-over- H.264/AVC-H.265/HEVC based services).

However, LCEVC is also very important for deployments where a service provider intends to enhance new codecs (e.g., LCEVC-over-H.265- H.266/VVC based services). In this case, besides the inherent advantages of using LCEVC as described above, LCEVC enables a more efficient and sustainable use of the new codec, particularly for live scenarios, as well as helps to accelerate the deployment and ease transitioning from existing systems to the new ones.

The low-complexity nature of LCEVC has other deployment advantages. Naturally, like any other coding technology, LCEVC can be deployed in silicon/ASIC.

In contrast to other coding technologies though, efficient LCEVC implementations can also be obtained without need for bespoke hardware, with decoding done using general processing units (CPU, GPU, DSP). This may enable quicker deployment by adding the LCEVC functionality to the decoding stack and, where possible, by software upgrading the existing devices.

LCEVC can reduce the processing cost of these newer compute-intensive codecs by up to 70%. Initial test results show that LCEVC reduces Video-on-Demand transcoding time (i.e., energy consumption) of the full Adaptive Bit Rate (ABR) ladder by 70% - or speeds up transcoding by 3.2x - while at the same time generating a superior video quality.

Key technical LCEVC features and benefits include:

- **SDR (Standard Dynamic Range) and HDR (High Dynamic Range) support:** LCEVC supports bit-depths from 8-bits to 14-bits (with the possibility of producing a higher bit-depth enhancement starting from a lower bit-depth base), as well as multiple Chroma resolutions such as Monochrome, 4:2:0, 4:2:2, 4:4:4.

- **Sparse residual data processing:** The coding scheme processes one or two layers of residual data. This residual data is produced by calculating differences between a reference video frame (e.g., a source video) and a base-decoded upscaled version of the video. The resulting residual data is sparse information, typically edges, dots and details which is then efficiently processed using very simple and small transforms that are designed to deal with sparse information.

- **Efficient use of existing codecs:** The base codec is typically used at a lower resolution. As such, the base codec operates on a smaller number of pixels, thus allowing the codec to use less power, operate at a lower quantisation parameter (QP), and use tools in a more efficient manner. Applying LCEVC to any codec reduces its complexity and can typically improve encoding density by up to 30% (see https://docs.v-nova.com/v-nova/lcevc/lcevc-sdk-overview#lcevc-benefits-include).

- **Resilient and adaptive coding process:** The coding scheme allows the overall coding process to be resilient to the typical coding artifacts introduced by traditional discrete cosine transform (DCT) block-based codecs. The first enhancement sub-layer (L-1 residuals) enables the correction of artifacts introduced by the base codec, whereas the second enhancement sub-layer (L-2 residuals) enables the addition of details and sharpness to the corrected upscaled base for maximum fidelity, up to and including lossless coding. Typically, the lower the quality of the base reconstruction, the more the first layer may contribute to the correction. Conversely, the better the quality of the base reconstruction, the more bitrate can be allocated to the second sub-layer to add the finest details.

- **Agnostic base enhancement:** The coding scheme can enhance any base codec, from existing codecs (VVC, HEVC, AVC, etc.) to future codecs. The enhancement operates on a decoded version of the base codec in the pixel domain; therefore, it can be applied to any format, as it does not require any information on how the base is encoded and/or decoded.

- **Parallelisation:** The coding scheme does not employ inter-block prediction. The image is processed by applying small (2x2 or 4x4) independent transform kernels over the layers of residual data. Since no prediction is made between blocks, each 2x2 or 4x4 block can be processed independently and in a parallel manner. Moreover, each layer is processed separately, thus allowing the decoding of the blocks and decoding of the layers to be done in a largely parallel manner.

- **Flexibility and extensibility:** LCEVC is not a standalone codec but requires a base codec to operate. As such, the base codec does not need to fulfill any requirements in order to be used in conjunction with LCEVC. This allows the use of any codec as a base codec, current and future, making it possible for LCEVC to be used with any existing and future codec without the need to revisit the LCEVC standardization. In addition, LCEVC offers the flexibility to customize most of the decoding tools.

    - **Quantisation is customisable:** The dequantization uses a quantisation matrix that contains the actual quantisation step widths to be used to decode each coefficient group. For the calculation of this matrix, default coefficients are preset in LCEVC. In addition, custom coefficients can be signaled in the bitstream, creating the opportunity to modify the quantisation process on a frame-by-frame basis to better work for specific sequences. The custom coefficients can be signaled either for only one or for both enhancement sub-layers.

    - **Upscalers are customisable:** There are four preset low-complexity upscalers (up to 4-tap) which can be selected by the encoder, ranging from nearest to a modified cubic interpolation. In addition, a fifth 4-tap upscaler can be used with customisable kernel coefficients. This allows LCEVC to adapt the upscaling process to improve the quality of the upscaled reconstruction for sequences where the adaptive cubic interpolation is beneficial.

    - **Inclusion of additional operations and extensions:** LCEVC incorporates the ability to signal to the decoder post-processing operations such as statistical dithering, which can be useful to reduce banding/aliasing impairments. The in-loop and out-of-loop functionality of LCEVC can be further extended with the insertion of local user data as LCEVC allows the efficient signaling of a specified number of bits at transform block level. The bits are embedded in one of the coefficient groups of enhancement sub-layer 1. User data bits, ignored by a reference decoder, can be optionally processed by a media player to control additional parameters in a system, allowing for backward-compatible enhancement extensions.

From a commercial standpoint, LCEVC provides numerous benefits that decrease cost and increase the ability to serve a wider customer base with a high-quality product:

- **Reduces Churn:** Better video quality and increased resolutions at the same bitrates drive up QoE and customer satisfaction, increasing viewing times for improved ad monetisation and significantly reducing subscriber churn.

- **Increases Reach:** LCEVC enables stable video delivery at lower bitrates (as low as 100 kbps), meaning that viewers can be reached over the most limited networks.

- **Reduces Delivery Costs:** Storage and CDN delivery costs can be reduced by up to 50%.

- **Compatibility:** LCEVC is delivered in industry-standard formats such as HLS and MPEG-DASH, meaning that it is transparent to existing infrastructure (CDNs, DRM, ad insertion, etc.) and even plays back on players that are not LCEVC-enabled.

- **Easy to Deploy:** LCEVC has been designed so that it can be implemented without necessarily requiring dedicated hardware acceleration and/or reusing available hardware blocks whenever possible. Depending on the application, it can be for example deployed at all levels of a software/firmware stack, e.g. from application level to operating system.

- **Broad Device Support and Leverage of Available Hardware:** LCEVC is compatible with the underlying codec that it enhances, and is supported by a broad range of encoding and player vendors. Native base codec decoding can leverage the existing hardware for that codec, and LCEVC leverages available hardware acceleration for graphics and scaling.

- **Reduced Demands on Resources:** Decoding is extremely lightweight, often freeing up resources and matching or reducing battery power consumption vs. native base codec decoding.

E.5 How LCEVC Works

E.5.1 Innovative multi-layer technology

LCEVC brings an innovative approach to video compression, employing a multi-layer process in which any base codec (e.g., H.266/VVC, H.265/HEVC, H.264/AVC, etc.), is enhanced via an additional low-bitrate stream. The data stream structure, as illustrated in Figure E.3, is defined by two component streams:

- a base stream conformant to a base codec (e.g., H.266/VVC, H.265/HEVC, H.264/AVC) and decodable by a standard hardware decoder; and,

- an enhancement stream conformant to the LCEVC standard and consisting of one or two enhancement layers suitable for software processing implementation with sustainable power consumption. The enhancement provides improved compression efficiency to existing codecs, and reduces encoding and decoding complexity for on-demand and live streaming applications.

The two component streams are usually encapsulated within a transport layer (e.g., using DASH, CMAF, MPEG-2 TS, etc.), demuxed at the decoder to produce the two separate component streams which are then fed to their respective decoders and processed to produce a final full-resolution output video. More details are provided in E.5.4.

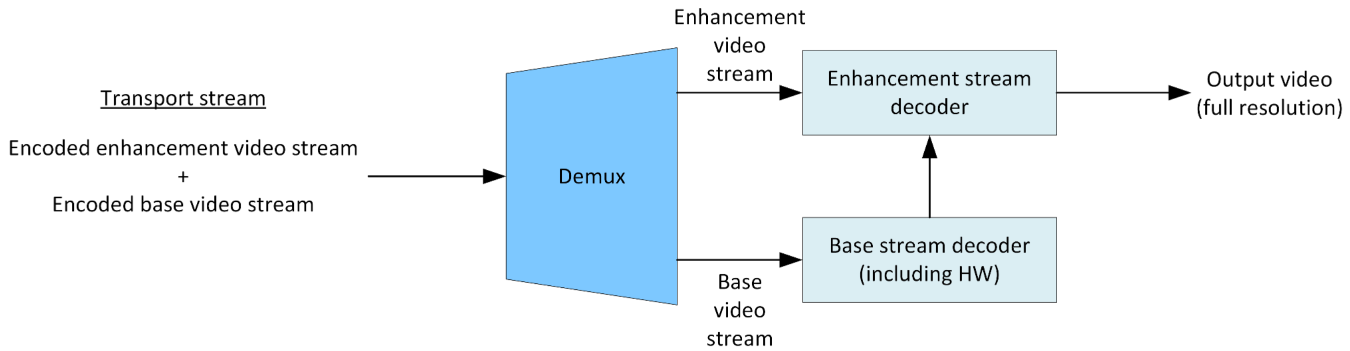

Figure E.3 – Schematic diagram of decoding of full resolution video using LCEVC

LCEVC employs a highly efficient and low complexity enhancement that provides up to two layers of encoded residuals that correct artifacts produced by the base video codec, and adds detail and sharpness to reconstruct the final output video, resulting in improved video compression efficiency while reducing the overall encoding complexity. Figure E.4 shows how the decoding process produces the final output picture.

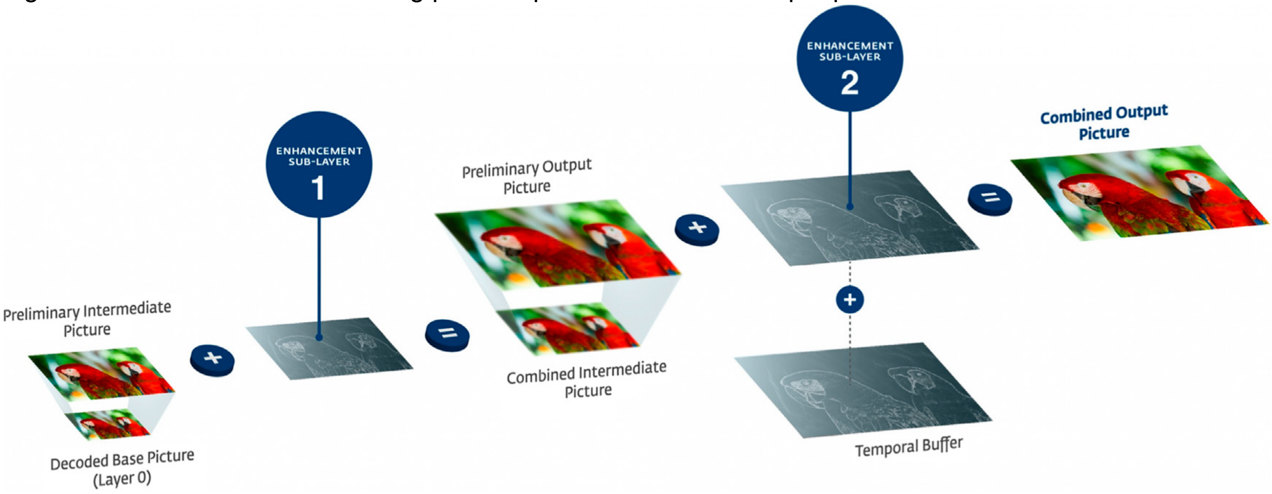

Figure E.4 – How LCEVC works: Two sub-layers of residual data

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

E.5.2Encoding Process

E.5.2.1 Encoder

The encoding process to create an LCEVC conformant bitstream encompasses three major steps, as shown in Figure E.5.

E.5.2.2 Base codec

Firstly, an input sequence is processed in order to be fed to a base codec. In LCEVC there are two scaling processes available. A first scaling process (set by the variable *scaling_mode_level2*) determines whether a first scaling is applied to the input sequence and, if so, whether across both dimensions or across the horizontal dimension only. A second scaling process (set by the variable *scaling_mode_level1*) determines whether a second scaling is applied to the sequence scaled via the first scaling process and, if so, whether across both dimensions or across the horizontal dimension only. In typical settings, only the first scaling process is applied. Depending on the selected scaling, the processed input sequence is fed to the base codec which produces a base bitstream conformant to the specification of the base codec.

E.5.2.3 Enhancement sub-layer 1

The reconstructed base picture may be upscaled depending on the value of the variable *scaling_mode_level1* and then subtracted from the first-order downscaled input sequence in order to generate sub-layer 1 (L-1) residuals. These residuals form the starting point of the encoding process of the first enhancement sub-layer. A number of coding tools, which will be described further in the following section, process the input and generate entropy encoded quantised transform coefficients (L-1 Coefficient Layers).

E.5.2.4 Enhancement sub-layer 2

As a last step of the encoding process, the enhancement data for sub-layer 2 (L-2) needs to be generated. In order to create the residuals, the coefficients from sub-layer 1 are processed by an in-loop LCEVC decoder to achieve the corresponding reconstructed picture. Depending on the value of the variable *scaling_mode_level2*, the reconstructed picture is processed by an upscaler to produce the preliminary output picture. Finally, the residuals are calculated by a subtraction of the input sequence and the preliminary output picture.

Similar to sub-layer 1, the samples are processed by a few coding tools. In addition, a temporal prediction can be applied on the transform coefficients in order to achieve a better removal of redundant information.

The entropy encoded quantised transform coefficients of sub-layer 2 (L-2 Coefficient Layers), the L-1 Coefficient Layers, a Temporal Layer specifying the use of the temporal prediction on a block basis, and the Headers are all included to form the LCEVC bitstream.

Figure E.5 illustrates the LCEVC encoding process.

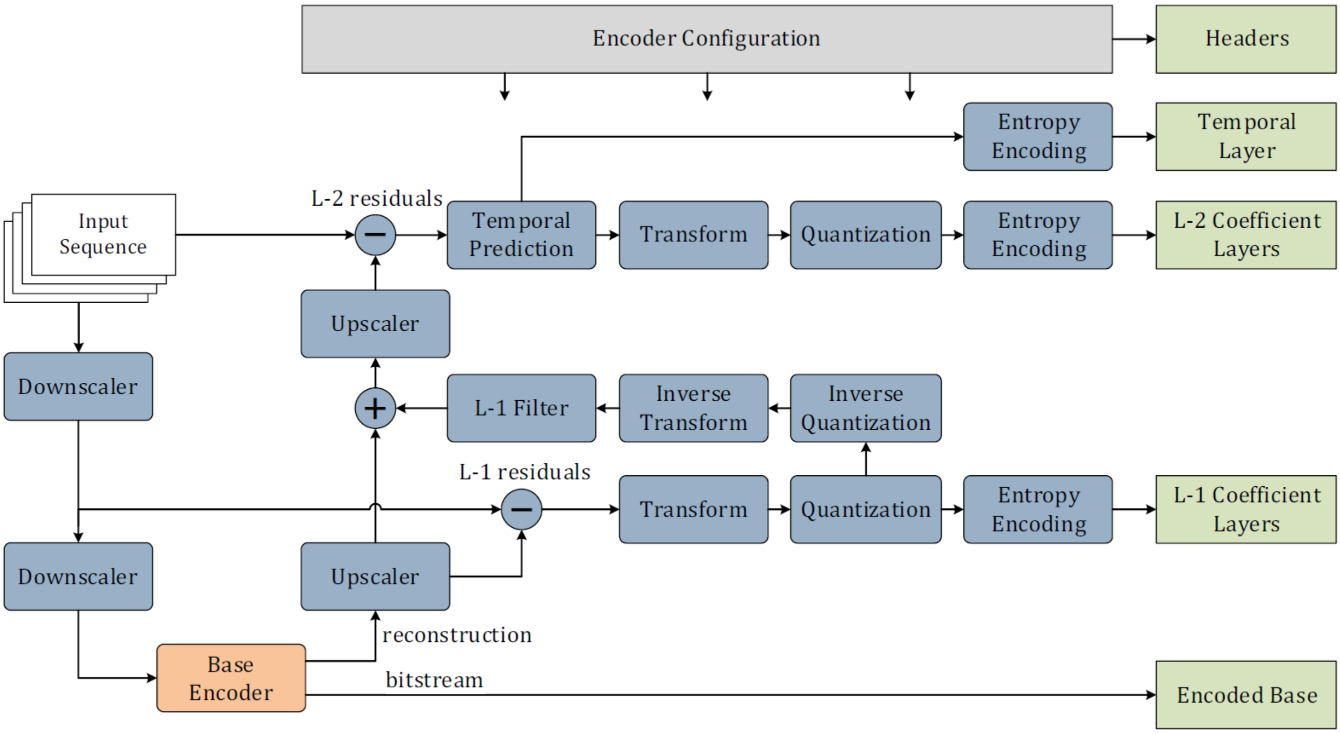

Figure E.5 – LCEVC Encoder Structure

E.5.3 Encoding Tools

E.5.3.1 Overview

This section provides a brief overview of the coding tools which are available in accordance with the LCEVC specification. For simplicity, the description is mostly focussed on an encoder.

E.5.3.2 Downscaler and Upscaler

Two non-normative downscalers can be used to downscale the input sequence to a lower resolution. The downscaling can be performed either in both vertical and horizontal directions, only in the horizontal direction, or alternatively is not be applied. Two upscalers are available reconstructing the sequence at a higher resolution. One of four specified four-tap upscaling kernels can be used. Additionally, a customisable four-tap kernel is available whose coefficients can be signaled in the LCEVC bitstream.

After the upscaling process, an additional predicted residuals coding block can be processed. For low bitrates, the tool can improve visual quality by correcting upscaling average errors at the quad level where otherwise the bitrate-to-signal non-zero average coefficients would not be available. For higher bitrates, it reduces the number of average coefficients required to be signaled, reducing the overall achieved bitrate.

E.5.3.3 Transform

LCEVC allows the usage of two different transforms. Both operate with a small kernel of size 2×2 or 4×4. If the upscaling process is performed in the horizontal direction only, the transform kernels are slightly modified to better reflect the preceding 1-dimensional upscaling.

The transforms are chosen to have a small kernel to optimize the coding of sparse residuals. In particular, sharp edges can efficiently be transformed. The use of the Hadamard transform is due to its better energy compaction of sparse input data, which is indeed the case for the enhancement layers prediction residuals. Moreover, the block size is limited to 4 × 4 to allow for parallel encoding implementations, whereby each block can be transformed independently on a multi threaded CPU or GPU architecture. The size used by the encoder is transmitted in the bitstream as coding metadata.

E.5.3.4 Quantisation

The transform coefficients are quantised using a linear quantiser whose quantisation step width is signaled on a picture basis. The linear quantiser may use a dead zone whose size changes relative to the quantisation step. The quantisation can be configured independently for both enhancement sub-layers to give more flexibility to the encoder. Additionally, the level of quantisation can be modified for transform blocks that use temporal prediction compared to intra-coded transform blocks. For chroma planes, the quantisation can be scaled separately from the luma plane.

E.5.3.5 L-1 filter

The L-1 filter can be applied on the sub-layer 1 residuals if the transform with the larger kernel size (4×4) is used. The aim of this filter is to reduce the blocking artifact which the transform applied at a lower resolution can create. The residuals on the outer boundary of the transform block are multiplied with a coefficient between 0 and 1. The value of these coefficients can be signaled independently for edges and corners.

E.5.3.6 Temporal prediction

The temporal prediction uses a zero-motion vector prediction with a temporal buffer which stores residuals from the previous frame only. The decision, where to use temporal prediction, is done on a transform block basis. Additionally, an entire tile of 32×32 residuals can be signaled to be used without temporal prediction, reducing the signaling overhead for, e.g., a fast-moving part of the sequence.

E.5.3.7 Entropy encoding

The two sets of coefficient layers and the temporal layer are processed independently by an entropy encoder before encapsulation into the bitstream. The entropy coding process consists of two components: a Run Length Encoder (RLE) and a Prefix Coding encoder. Additionally, it is possible to use only the RLE for the entire data within a coefficient layer.

Tiling options enable parallel execution of entropy coding, including the possibility to define custom tile sizes for residual sub-layer data. It is also possible to embed user data into some of the coefficients at transform block level.

E.5.4 Decoding process

E.5.4.1 Overview

For the creation of the output sequence, the decoder analyzes the LCEVC conformant bitstream. As can be seen in Figure E.6, the process can again be divided into three parts, and is essentially the inverse of the encoding process.

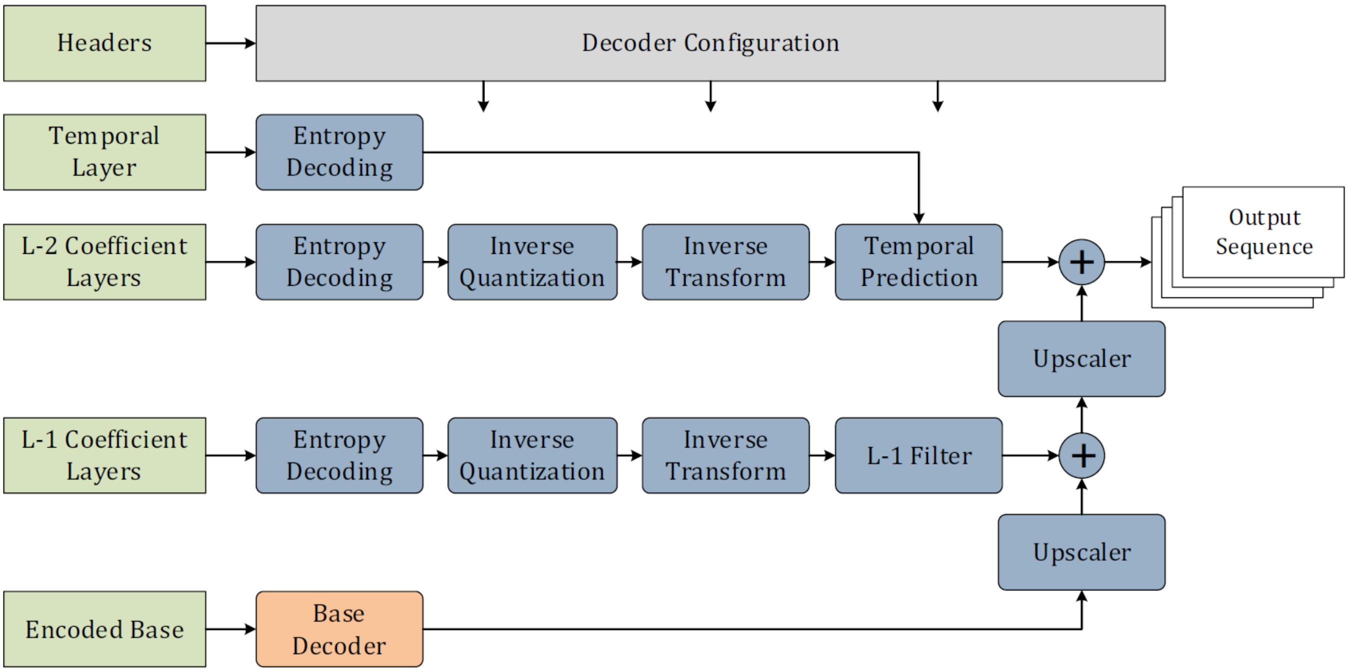

Figure E.6 – LCEVC Decoder Structure

E.5.4.2 Base codec

In order to generate the Decoded Base Picture (Layer 0) the base decoder is fed with the extracted base bitstream. According to the chosen scaling mode in the configuration, this reconstructed picture might be upscaled and is afterwards called Preliminary Intermediate Picture.

E.5.4.3 Enhancement sub-layer 1

Following the base layer, the enhancement part needs to be decoded. Firstly, the coefficients belonging to enhancement sub-layer 1 are decoded using the inverse tools of the encoding process. Additionally, an L-1 filter might be applied in order to smooth the boundaries of a transform block. The output is then referred to as Enhancement Sub-layer 1 and is added to the preliminary intermediate picture which results in the Combined Intermediate Picture. Again, depending on the scaling mode, an upscaler might be applied and the resulting Preliminary Output Picture has the same dimensions as the overall output picture.

E.5.4.4 Enhancement sub-layer 2

As a final step, the second enhancement sub-layer is decoded. According to the temporal layer, a temporal prediction might be applied to the dequantized transform coefficients. This Enhancement Sub-layer 2 is then added to the Preliminary Output Picture to form the Combined Output Picture as a final output of the decoding process.

E.5.5 Differences Between LCEVC and other Scalable Video Coding Technologies

As described above, MPEG-5 LCEVC has some similarities with a scalable codec (i.e. spatial scalability thanks to the upsampler) but it is also substantially different for the following reasons:

- Generally, in a scalable codec, the base layer is encoded with the same standard of the enhancement layer. As specified in the description LCEVC is codec agnostic. The base layer used in LCEVC can be any codec. This particular feature allows LCEVC to be used with any standard, from the older one (i.e. H.264/AVC) to the latest one (H.266/VVC) and also with any others that have not been developed by MPEG (Verification Test Report on the Compression Performance of Low Complexity Enhancement Video Coding, 2021) (Jiménez-Moreno et al., 2022) (Martini). It is worth mentioning here that even the scalable extension of the H.265/HEVC standard (Scalable High efficiency Video Coding, SHVC) allows the base layer to be compliant with any standard or proprietary format. However, SHVC will then use coding tools such as the Context Adaptive Binary Arithmetic Coding (CABAC) which may not be supported on legacy hardware (think about an H.264/AVC receiver compliant with the baseline profile) hence its deployment in application scenarios targeted by LCEVC may not be possible.

- The MPEG-5 LCEVC structure is using simple tools specifically designed for the sparse nature of the residual data which allow to keep the complexity low and limit the overhead associated with the enhancement layers, a common problem of scalable codecs. This makes it possible to have a software version of the MPEG-5 LCEVC that can run on existing hardware and on top of existing base codec with no need to develop a specific hardware for it. As a consequence, the base codec can work more efficiently and faster given the ability of LCEVC to work with a base codec running at a quarter of the resolution. As a demonstration of this, the results of the MPEG verification test are described in [5].

- As distinguished from most scalable codecs, MPEG-5 LCEVC provides two levels of enhancement that can be applied at different stages or resolutions. Each level has its own independent quantisation module and sub-layer of the bitstream that can easily decouple from the other. This also allows bitrate allocation flexibility to cope with different types of content.

- From Figure E.5 and E.6, it may be noted that MPEG-5 LCEVC offers up to two cascade scaling processes in order to further improve the efficiency of the base layer. Each scaler can be user defined, along the following degrees of freedom: kernel size, type of upscaling (i.e., which sub-layer, L-1 or L-2) and kernel values.MPEG-5 LCEVC offers 4 normative upsamplers and one 4 taps user defined kernel. Scalable codecs generally offer only one fixed scaling engine and it is not programmable.

- MPEG-5 LCEVC can handle different bit-depths up to 14 bits per pixel in the main profile. The standard allows the base layer to work on a different bit-depth compared to the input signal one. This operation can effectively enhance a base layer working at a lower bit-depth to a higher one contributing to maintain the fidelity of the input signal. An example of this application is delivering HDR with technologies that cannot deliver more than 8-bit like AVC4.

E.6 LCEVC technical guidelines

The many improvements LCEVC brings in the source coding and signal processing are not discussed in this document in detail, as their adequate implementation and configuration in both encoder and decoder implementation can be assumed in any commercially viable LCEVC product. As with older video codecs, also in LCEVC technology, the optimisation 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 technologies as a guideline—and there is no reason to believe that this experience should not apply for LCEVC also—we will see significant improvements in coding efficiency of commercial LCEVC technology as competitors innovate to get the best performance out of LCEVC's design and syntax.

A comprehensive overview of LCEVC is available in: https://www.mpeg.org/standards/MPEG-5/2/

E.7 Broadcast and streaming scenarios: LCEVC for HD, 4K and 8K delivery

E.7.1 LCEVC for HD, 4K and 8K delivery

LCEVC is used in combination with a base codec to provide a more efficient compression at a lower computational cost. With the advent of larger screens and the need to provide immersive experiences to customers, providing higher resolution is an obvious requirement, and with that a higher level of compression efficiency is also needed. In addition, reduction of computational complexity both at encoder and decoder is key from a sustainability point of view, particularly when producing and consuming higher quality.

LCEVC can be used to deliver any type of content at any type of resolution with the right tradeoff between compression and complexity. In particular, LCEVC can deliver HD, 4K and 8K content which are relevant for TV 3.0 use cases as well as for streaming applications (where also sub-HD resolutions can be delivered very easily).

E.7.2 LCEVC and High Dynamic Range (HDR) encoding

E.7.2.1 Overview

LCEVC’s ability to provide the ability to signal the bit-depth of the base layer independently from that of the enhancement layer is particularly relevant for encoding HDR content. In addition to transmitting the base layer and enhancement layer at the same bit-depth (e.g., 10-bit), the format allows transmission of a lower bit-depth base layer (e.g., 8-bit) and a higher bit-depth enhancement (e.g., 10-bit). This latter feature allows further quality enhancement capabilities, for example, transmission of an SDR base layer with an HDR enhancement layer or, in the future, a 10-bit-depth HDR base layer with a 12-bit-depth HDR enhancement layer.

There are two modes of operation: full HDR and extended HDR. In the full HDR mode, the bit-depth of the base layer and that of the enhancement layer are the same - today this is typically specified as 10-bit to reflect the commercial HDR formats available today, but the bit-depth can extend to up to 14 bit. In the extended HDR mode, the bit-depth of the base layer is lower than that of the enhancement layer, meaning for example that we can encode an HDR source using for the base layer an AVC/H.264 or HEVC/H.265 8-bit encoding and then adding on top an LCEVC 10-bit enhancement layer to bring the full-resolution video to HDR quality, allowing HDR support to devices and workflows which are not designed to deliver HDR quality video. In this way, both devices which are HDR capable and devices that are non-HDR capable can receive the same bitstream and display HDR quality.

Moreover, LCEVC is agnostic to PQ or HLG transfer functions as the input format when a backward-compatible stream is desired. The input sequence can be converted from the original format to the lower bit-depth format and then use this lower bit-depth sequence as input to the base encoder. Then, the enhancement layers will work with the original bit-depth converting back the decoded sequence of the base to the original format

E.7.2.2 MPEG-5 LCEVC features for HDR content

LCEVC already has some specific tools for managing HDR content and gives the possibility of adding new ones to better fit the HDR features.

Bit-depths up to 14-bit: Currently, LCEVC allows up to 14 bits in its main profile data representations, which covers the main HDR formats, i.e. HDR10, HLG, a big part of HDR10+ and Dolby Vision. As such, LCEVC can cover the encoding necessities in the great majority of HDR architectures at this time. Moreover, the great flexibility of the LCEVC encoder allows us the combination of different bit-depths for base layer and enhancement layers. For example, in a full HDR mode a bit-depth of 10 or 12 bits can be used for both base and enhancement layers. In extended HDR mode, an 8-bit base layer with a 10-bit enhancement layer or a 10-bit base layer with a 12-bit enhancement layer can be used.

Dithering: One of the drawbacks when using a reduced bit-depth for the base encoder is the appearance of visible artifacts that can degrade the visual quality of the output picture. One of the most common artifacts is the banding that appears when a sequence contains a big area with a soft color transition, this artifact also can be easily obtained when using a hard quantisation. To solve this, LCEVC allows the use of a post-processing dithering module in the decoding process. The MPEG-5 LCEVC standard is able to indicate the strength of the dithering to be applied, but the type is not mandated.

Separated quantisation for the chroma: QP adjustment is a common way to improve the quality of the HDR encoded sequences, reducing the artifacts generated when the HDR content is managed by a video encoder devised to deal with SDR sequences. Moreover, the QP adjustment is also recommended in the guidelines for the operation of the AVC or HEVC video coding systems for compressing HDR video [10]. LCEVC allows for the use of different QP values for the luminance and chrominance planes and the different transformed coefficients. This flexibility when selecting the QP, can be used to make a proper adjustment of the QP values for the HDR content. In the current version of the LCEVC reference software, the QPs are adjusted for every type of coefficient after the transformation. Also, LCEVC signals a modifier for the QP of the chrominance planes. There are 8 bits available to specify a value used to scale the QP of L-2 for the chrominance.

User data as additional information: LCEVC gives the possibility of including additional information in the regular bitstream. To do so, at transform level (so, for every 2×2 or 4×4 data block, depending on the encoder configuration) we can use up to 6 bits of the first transform coefficient to signal extra information. These bits will be extracted in the decoder side and used to adjust the decoding process as the user had specified. This information can be used to make an adaptive adjustment of the QP values for luma or chrominance, depending on the brightness or color values, which allows extra flexibility to avoid quantisation artifacts for HDR content without adding extra bits in the bitstream. This signaling is part of the LCEVC standard, but the specific use is not normative. So, a LCEVC compliant decoder must, at least, be able to read and store this extra information. The specific use of it will depend on the decoder implementation.

E.8 LCEVC configuration aspects - profiles and levels

E.8.1 Profiles and levels overview

Profiles and levels specify restrictions on the bitstreams and hence limits on the capabilities needed to decode the bitstreams. Profiles and levels can jointly specify conformance points to facilitate interoperability between implementations, devices and applications implementing LCEVC.

Profiles and levels specify restrictions on the bitstreams and hence limits on the capabilities needed to decode the bitstreams.

Each profile specifies a subset of algorithmic features and limits that are supported by all decoders conforming to that profile. A decoder compliant with a particular profile and level must be able to decode any bitstream compliant with that profile and level.

Each level specifies a set of limits on the values that may be taken by LCEVC syntax elements. The same set of level definitions is used with all profiles, but individual implementations may support a different level for each supported profile. For any given profile, a level generally corresponds to a particular decoder processing load and memory capability.

E.8.2 TV 3.0 Profile

The TV3.0 profile defined for LCEVC is based on the Main profile as defined in ISO/IEC 23094-2 with some further restrictions as specified in Annex E.

E.8.2.1 Profiles Overview

LCEVC supports a wide range of applications and allows the coding tools that can be implemented by an encoder to produce LCEVC-compliant bitstreams.

Contrary to more complex codecs which include multiple computationally expensive coding tools thus requiring different profiles to be defined in order to limit those tools, LCEVC includes a limited number of low-complexity tools and therefore it includes all the LCEVC tools in both profiles.

Each profile specifies a subset of algorithmic features and limits that shall be supported by all decoders conforming to that profile.

NOTE Encoders are not required to make use of any particular subset of features supported in a profile.

Each level specifies a set of limits on the values that may be taken by the syntax elements of this document. The same set of level definitions is used with all profiles, but individual implementations may support a different level for each supported profile. For any given profile, a level generally corresponds to a particular decoder processing load and memory capability.

LCEVC therefore provides only two profiles:

- Main
- Main 4:4:4

The only difference between Main and Main 4:4:4 is in the chroma sampling, which in the case of Main is limited to monochrome and 4:2:0.

NOTE In the TV 3.0 system, LCEVC only uses the Main profile.

The relationship between the profiles is shown in Figure E.7.

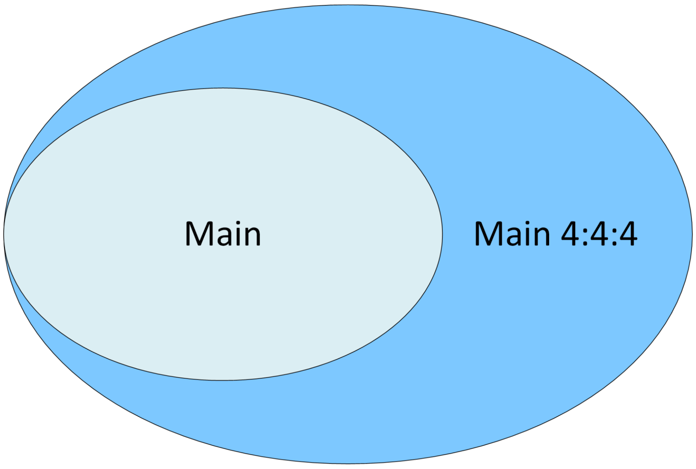

Figure E.7 – LCEVC profiles and their relationship

E.8.2.2 Global configuration parameter sets constraints

All constraints for global configuration parameter sets that are specified are constraints for global configuration parameter sets that are activated when the bitstream is decoded.

E.8.2.3 Main profile

Conformance of a bitstream to the Main profile is indicated by profile_idc equal to 0.

Bitstreams conforming to the Main profile shall have the following constraint:

- Active global configuration data blocks shall have chroma_sampling_type equal to 0 or 1 only.

Decoders conforming to the Main profile at a specific level (identified by a specific value of level_idc) shall be capable of decoding all bitstreams and sub-layer representations for which all of the following conditions apply:

- The bitstream is indicated to conform to the Main profile.

- The bitstream or sub-layer representation is indicated to conform to a level that is lower than or equal to the specified level.

E.8.3 Levels

A level is a defined set of constraints imposed on the parameters in the bitstream that limit the maximum pixel processing rate 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 supported at that level. For example, Level 2 provides support for up to 1920 × 1080 at 60 fps (frames per second) but does not support 3840 × 2160 or 7680 × 4320 at 60 fps.

LCEVC provides seven levels:

- Levels 1 through 5 each provide two sublevels.
- Levels 6 and 7 each provide three sublevels.

The LCEVC levels cover a vast range of picture sizes from 1280 × 720 (30 fps) to 7680 × 4320 (240 fps) and 15360 × 8640 (120 fps).

Levels are defined based on the following two parameters:

- Count of luma samples of output picture in time (i.e., the output sample rate), and
- Maximum input bitrate for the coded picture buffer LCEVC (CPBL).

Both sample rate and bitrate are based on observation periods of one second.

The levels and their specifications supported in TV 3.0 are listed in Table E.1.

Table E.1 - General level limits

Level	Sublevel	Maximum output sample rate	Maximum Coded Picture Buffer LCEVC (CPBL) bitrate (bit per second per thousand output samples	Example resolution frame rate	Maximum number of tiles
1	0	29,410,000	4	1280 × 720 (30 fps)	16
1	1	29,410,000	40	1280 × 720 (30 fps)	16
2	0	124,560,000	4	1920 × 1080 (60 fps)	32
2	1	124,560,000	40	1920 × 1080 (60 fps)	32
3	0	527,650,000	4	3840 × 2160 (60 fps)	128
3	1	527,650,000	40	3840 × 2160 (60 fps)	128
4	0	2,235,160,000	4	7680 × 4320 (60 fps)	256
4	1	2,235,160,000	40	7680 × 4320 (60 fps)	256
5	0	3,981,312,000	4	7680 × 4320 (120 fps)	256
5	1	3,981,312,000	40	7680 × 4320 (120 fps)	256

E.9 LCEVC bitstream structure

The LCEVC bitstream contains a base layer, which may be at a lower resolution, and an enhancement layer consisting of up to two sub-layers. This subsection briefly explains the structure of this bitstream and how the information can be extracted.

While the base layer can be created using any video encoder and is not specified further in the LCEVC specification, the enhancement layer must follow the structure as specified. Similar to other MPEG codecs, the syntax elements are encapsulated in network abstraction layer (NAL) units which also help synchronize the enhancement layer information with the base layer decoded information. Depending on the position of the frame within a group of pictures (GOP), additional data specifying the global configuration and controlling the decoder may be present.

The data of one enhancement picture is encoded into several chunks. These data chunks are hierarchically organized as shown in Figure E.8. For each processed plane (nPlanes), up to two enhancement sub-layers (nLevels) are extracted. Each of them again unfolds into numerous coefficient groups of entropy encoded transform coefficients. The amount depends on the chosen type of transform (nLayers). Additionally, if the temporal prediction is used, for each processed plane an additional chunk with temporal data for Enhancement sub-layer 2 is present.

On the syntactical level, an LCEVC bitstream is made up of a sequence of Network Abstraction Layer units (NAL units), which can be viewed as packets of different types (see Table E.2), containing bits associated with the type.

- Each LCEVC NAL unit corresponds to one picture and contains a payload.
- Some NAL unit types are related to the decoding engine, such as coded slices.
- A NAL unit which includes encoded data comprises at least two data blocks:
    - A picture configuration data block, and
    - An encoded (tiled) data block.
- Other NAL units relate to control information such as parameter sets or metadata.

There are two types of NAL units:

- Instantaneous Decoder Refresh (IDR): this contains an IDR segment. NOTE: An IDR NAL unit must be present if the corresponding base layer picture is an IDR picture.
- Non-IDR: this contains a non-IDR segment.

Observations about this syntactical level are included below. However, before providing details about the syntax, it is 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.

An LCEVC bitstream includes one or more coded video sequences (CVSs), each of which contains one or more access units (AUs). A CVS can be decoded independently from any other CVS.

An LCEVC bitstream contains one or more sets of coefficient layers. There are two enhancement sub-layers (L-1 and L-2) as well as a temporal layer. Each of the enhancement sub-layers can contain either 4 or 16 coefficient layers, depending on the transform used (4 if a 2x2 transform is used; 16 if a 4x4 transform is used).

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 coded picture. A coded picture is a coded representation of a picture containing all transform units (TUs) of the picture. Layers need not have the same picture rate and consequently it is not required that each AU has a coded picture at each layer that is present in the bitstream. If we ignore layers, then an AU is similar to what a coded picture was in MPEG-2.

Payload types include, for example, global_configuration sets that set configurations for the entire GOP, while the picture_configuration payload type applies only to the current picture. Payload types, payload content, and minimum frequency are specified in Table E.2.

Table E.2 – Content of payload and minimum frequency of appearance of such content within a bitstream

payload_type	Content of payload	Description	Minimum frequency
0	process_payload_sequence_config()	Specifies sequence configuration.	At least with first IDR in sequence
1	process_payload_global_config():	Specifies configuration that applies to the entire GOP.	Per IDR picture
2	process_payload_picture_config()	Specifies configuration that applies to only the current picture.	Per picture
3	process_payload_encoded_data()	Contains encoded data.	Per picture (if no_enhancement_Bit_flag == 0)
4	process_payload_encoded_data_tiled()	Contains encoded tiled data.	Not specified (optional)
5	process_payload_additional_info()	Specifies additional information.	Not specified (optional)
6	process_payload_filler()	Contains filler payload	Not specified (optional)
7 .. 30	Reserved	Reserved	Reserved
31	Unspecified	Unspecified	Unspecified

A video coding layer (VCL) 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.

NAL unit types are either VCL or non-VCL NAL units, and are specified in Table E.3. which are described in the subsequent paragraphs. NAL unit types are either non-VCL NAL units, which are summarized in Table E.3, or video coding layer (VCL) NAL units, which are described in the subsequent paragraphs. VCL refers to the 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 E.3 – NAL unit type codes and NAL unit type classes

nal_unit_type	Name of nal_unit_type	Content of NAL unit and RBSP syntax structure	NAL unit type class
0..27	UNSPEC0…UNSPEC 27	Unspecified	Non-VCL
28	LCEVC_NON_IDR	Non-IDR segment	VCL/Non-VCL
29	LCEVC_IDR	IDR segment	VCL/Non-VCL
30	LCEVC_RSV	Reserved	VCL/Non-VCL
31	UNSPEC_31	Unspecified	Non-VCL

Armed with this picture-centric understanding, and returning to the syntax level where the LCEVC bitstream can be defined as a sequence of network abstraction layer (NAL) units.

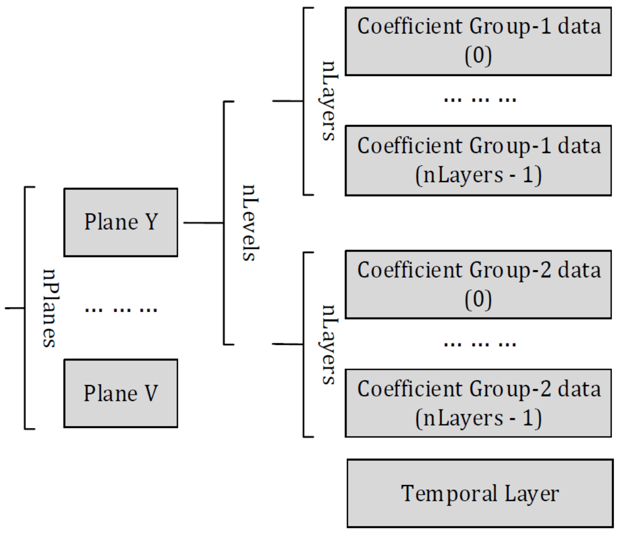

Figure E.8 – Encoded enhancement picture data chunk structure

E.10 LCEVC Support of Supplementary Enhancement Information (SEI) and Video Usability Information (VUI)

E.10.1 SEI/VUI Overview

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). SEI and VUI are specified in Annex D and E of LCEVC specification.

LCEVC natively supports SDR and HDR content, and all LCEVC coding tools are applicable to both SDR and HDR video. LCEVC can follow the VUI parameters of the base layer, or carry its own set of SDR or HDR parameters.

Many applications rely on metadata signaled in supplemental enhancement information (SEI) messages, which are specified in LCEVC. SEI and VUI information relevant only to the LCEVC syntax are included in the LCEVC specification.

Since LCEVC works in combination with a base codec, an SEI or VUI message could be included either in the LCEVC bitstream or in the base codec bitstream. For those SEI and VUI which are in common between LCEVC and base codec, Annex E of the specification provides useful guideline as to how the decoder would deal with it if the same SEI or VUI is present in both LCEVC and base codec.

E.10.2 SEI

SEI messages assist in processes related to decoding, display or other purposes. However, SEI messages are not required for constructing the luma or chroma samples by the decoding process. Conforming decoders are not required to process this information for output order conformance to the LCEVC standard. Some SEI message information is required to check bitstream conformance and for output timing decoder conformance.

Specification for presence of SEI messages is also satisfied when those messages (or some subset of them) are conveyed to decoders (or to the HRD) by other means not specified in the LCEVC specification. When present in the bitstream, SEI messages shall obey the syntax and semantics specified in the LCEVC standard. When the content of an SEI message is conveyed for the application by some means other than presence within the bitstream, the representation of the content of the SEI message is not required to use the same syntax specified in the LCEVC standard. For the purpose of counting bits, only the appropriate bits that are actually present in the bitstream are counted

SEI messages can assist decoders, displays, and other video receivers to perform as desired by the content producer.

SEI messages such as the mastering display color volume (MDCV) and content light level information (CLLI) SEI messages are used extensively to indicate essential properties of HDR video. User data, both registered by Recommendation ITU-T T.35 SEI and unregistered, can as well be included in SEI messages. LCEVC also provides the reserved SEI message, which consists of data reserved for future backward-compatible use by ISO/IEC.

While SEI messages can enhance decoding, they are not required for constructing the luma or chroma samples by the decoding process. Conforming decoders are not required to process this information for output order conformance to LCEVC. Some SEI message information, however, is required to check bitstream conformance and for output timing decoder conformance.

E.10.3 VUI

VUI parameters provide information for the correct display of coded video. For example, VUI parameters specify progressive or interlaced, format (SDR, HDR HLG, or HDR PQ), aspect ratio, and any other parameters necessary for correct interpretation of the coded video.

VUI parameters are not required for constructing the luma or chroma samples by the decoding process. Conforming decoders are not required to process this information for output order conformance to the LCEVC standard. Some VUI parameters are required to check bitstream conformance and for output timing decoder conformance.

E.11 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.

As pre-encoding processes are not specific to LCEVC, but rather to the manner in which the input sequence is processed, the reader can refer to the VVC Operational Guidelines for information pertaining to Chroma subsampling and Encoding of HDR content.

E.12 Post-decoding processes

E.12.1 Overview

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 color 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.

E.12.2 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.

E.12.3 Mastering display color 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 color 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 ST2086. 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.

E.12.4 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.

E.12.5 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 and CLLI 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 LCEVC. However, LCEVC 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 LCEVC can be found in the SBTVD Forum TV 3.0 Technical Standards, Operational Guidelines, and Conformance Testing documentation.

E.13 MPEG-5 Part 2 LCEVC Test Results

E.13.1 Test results on HD and 4K content

MPEG carried out a formal subjective verification test of the LCEVC Main profile for standard dynamic range (SDR) content testing AVC, HEVC, EVC, and VVC as base and anchor codecs [6]. The purpose of the verification test was to confirm that the coding efficiency objective for the LCEVC standard has been met, achieving a bit-rate reduction at a similar level of subjective visual quality relative to a single-layer video codec. This document reports the results of the verification test to confirm that this goal was achieved and to estimate the magnitude of this achievement.

Verification tests compared full-resolution LCEVC-enhanced encoded sequences both with full-resolution single layer anchors and with half-resolution anchors upsampled to full resolution with Lanczos upsampling. The comparison with half-resolution anchors was used to validate that enhancing a base codec with LCEVC is preferable both with respect to encoding with the base codec alone at full resolution and with respect to encoding with the base codec alone at a lower resolution and then relying on unguided upsampling at the decoder device.

The first set of tests compared full-resolution LCEVC-enhanced encoded sequences with full-resolution single-layer anchors. The average bitrate savings for LCEVC when enhancing AVC were determined to be approximately 46% for UHD and 28% for HD. The average bitrate savings for LCEVC when enhancing HEVC were determined to be approximately 31% for UHD and 24% for HD. Numerical analysis of the average benefit of LCEVC and its statistical significance compared to the corresponding full resolution EVC or VVC codec was more difficult to interpret, due to several test points having overlapping confidence intervals. However, the test results tend to indicate an overall benefit when using LCEVC with these two codecs.

The second set of tests aimed to confirm that LCEVC provided a more efficient means of resolution enhancement of half resolution anchors than unguided upsampling. For these tests, the test sequences were coded using AVC, HEVC, EVC, or VVC at half resolution in both horizontal and vertical directions. For anchor generation, the half-resolution encoded sequences were upsampled with Lanczos filters to full resolution for visual assessment. The same half-resolution encoded sequences were also used as base layers for LCEVC and hence not all curves may overlap as much as would be ideal when calculating a BD-rate. Comparing LCEVC full-resolution encoded sequences with the upsampled half-resolution anchors, the average bitrate savings when using LCEVC with AVC, HEVC, EVC, and VVC were calculated to be approximately 28%, 34%, 38%, and 33% respectively for UHD, and 27%, 26%, 21%, and 21% respectively for HD.

The verification test for LCEVC includes standard dynamic range (SDR) UHD test sequences encoded in random-access configuration and SDR HD test sequences encoded in random-access configuration

Table Table E.4 and Table E.5 show the MOS BD-rate for the sequences in this test.

Table E.4 – MOS BD-rate – LTM vs JM and HM anchors at full resolution

LTM 5.1 vs JM 19.0		BD-rate
UHD	LupoPuppet	−53.98%
	CatRobot	−43.83%
	DrivingPOVLogo	−30.19%
	BoxeLogo	−55.61%
Average		**−45.90%**

LTM 5.1 vs HM 16.20		BD-rate
UHD	LupoPuppet	−31.14%
	CatRobot	−41.88%
	DrivingPOVLogo	−26.01%
	BoxeLogo	−24.44%
Average		**−30.87%**

LTM 5.1 vs JM 19.0		BD-rate
HD	TrafficLogo	−30.18%
	Starcraft	−26.75%
Average		**−28.47%**

LTM 5.1 vs HM 16.20		BD-rate
HD	TrafficLogo	−18.52%
	Starcraft	−29.76%
Average		**−24.14%**

Table E.5 – MOS BD-rate – LTM vs upsampled half-resolution anchors

LTM 5.1 vs JM 19.0		BD-rate
UHD	LupoPuppet	−22.66%
	CatRobot	−26.25%
	DrivingPOVLogo	−26.95%
	BoxeLogo	−34.20%
Average		**−27.52%**

LTM 5.1 vs HM 16.20		BD-rate
UHD	LupoPuppet	−20.64%
	CatRobot	−50.77%
	DrivingPOVLogo	−33.96%
	BoxeLogo	−29.23%
Average		**−33.65%**

LTM 5.1 vs JM 19.0		BD-rate
HD	TrafficLogo	−27.80%
	Starcraft	−26.28%
Average		**−27.04%**

LTM 5.1 vs HM 16.20		BD-rate
HD	TrafficLogo	−22.51%
	Starcraft	−29.10%
Average		**−25.80%**

LTM 5.1 vs ETM 6 rc1		BD-rate
UHD	BarScene	−40.54%
	CatRobot	−39.64%
	DrivingPOVLogo	−46.67%
	BoxeLogo	−24.58%
Average		**−37.86%**

LTM 5.1 vs VTM 11		BD-rate
UHD	Marathon	−30.44%
	MountainBay2	−22.52%
	DrivingPOVLogo	−43.42%
	BoxeLogo	−37.07%
Average		**−33.36%**

LTM 5.1 vs ETM 6 rc1		BD-rate
HD	TrafficLogo	−30.91%
	Starcraft	−11.28%
Average		**−21.09%**

LTM 5.1 vs VTM 11		BD-rate
HD	TrafficLogo	−21.11%
	Starcraft	−19.96%
Average		**−20.53%**

E.13.2 Tests on 8K content

In [7] a test for LCEVC enhancing HEVC against native HEVC is provided. For the tests, we have used real-time implementations of LCEVC and HEVC, namely V-Nova LCEVC SDK and an open source HEVC software encoder implementation (x265). For the test we used the 8K Berlin Test Sequences and used constant rate factor quality configurations for both LCEVC and x265.

The test results are reported in Table E.6. Results are reported using VMAF as an objective metric. As it can be seen, LCEVC enhancing HEVC can provide an average bitrate saving of almost 39% with an average increase of over 6 VMAF points over a corresponding encoding with HEVC.

Table E.6 – 8K test results

	BD-Rate (%)	BD-Distortion (VMAF)
BodeMuseum	–41.99	5.56
OberbaumSpree	–49.74	6.59
NeptuneFountain2	–38.81	7.62
NeptuneFountain3	–36.70	6.87
QuadrigaTree	–36.67	7.47
SubwayTree	–28.92	4.88
TiergartenParkway	–37.91	7.43
Average	–38.68	6.63

E.13.3 Tests on HDR content

Tables E.7 and E.8 report the results of the comparison between LCEVC enhancing VVC and HEVC. In this test the following implementations were used: 1) for LCEVC, a proprietary implementation provided by V-Nova (V-Nova LCEVC SDK); 2) for VVC, an open-source software implementation provided by HHI Fraunhofer (VVenC 1.2.0); and 3) for HEVC, the software test model implementation (HM 16.22).

Two configurations were used. In Configuration A, the proportion of bitrate allocated to base layer and enhancement layer has been set so that the proportion of the overall bitrate allocated to the enhancement layer was between 5% and 20% depending on the operating point. Although not necessarily optimal, the proportions specified in Configuration A are more aligned with those used in the MPEG verification test and are closer to the typical proportions used by LCEVC when encoding the enhancement layers. On the other hand, in Configuration B the proportion of bitrate allocated to base layer and enhancement layer has been set so that the proportion of the overall bitrate allocated to the enhancement layer was kept at 50% for each operating point. Results in Tables E.7 and E.8 are reported in terms of average BD-rate using mean opinion score (MOS) as the quality metric. BD-rate30 represents the bitrate savings provided by a target codec with respect to a reference codec (in the present case, LCEVC-enhancing VVC corresponds to the target codec while HEVC corresponds to the reference codec). A negative BD-rate value corresponds to a saving in bitrate by the same amount. The MOS corresponds to the average score given during subjective tests run according to formal subjective assessments, as defined in Recommendation ITU-R BT.500.31.

Table E.7 – MOS scores and CI for PQ sequences.

Sequence	LCEVC (config A) vs. HM	LCEVC (config B) vs. HM
	BD-rate	BD-rate
vc10-lcevc-1	–45.93%	–71.52%
vc10-lcevc-2	–56.86%	–37.13%
vc10-avs3	–27.51%	10.24%
vc10-dre	–79.54%	–37.37%
vc10-globo	–78.40%	–42.24%
vc10-vvc	–27.23%	–41.73%
Average	–52.58%	–36.63%

Table E.8 – MOS scores and CI for HLG sequences.

Sequence	LCEVC (config A) vs. HM	LCEVC (config B) vs. HM
	BD-rate	BD-rate
vc04-lcevc-1	–54.24%	–82.51%
vc04-lcevc-2	–62.22%	–42.46%
vc04-avs3	–37.46%	–17.90%
vc04-dre	–57.66%	–38.15%
vc04-globo	–72.57%	–54.99%
vc04-vvc	–43.46%	–59.01%
Average	–54.60%	–49.17%

As it may be seen, LCEVC enhancing H.266/VVC provides an average bitrate saving of between 52% and 54% for both PQ and HLG sequences when Configuration A is used, and an average bitrate saving of almost 50% for HLG sequences and 36% for PQ sequences when Configuration B is used instead.

E.14 LCEVC best practices - Optimising results when applying LCEVC enhancement to encoding workflows

There are a number of resources available online about how to configure LCEVC in encoding workflow specifically using the LCEVC SDK software described in Annex A.

In particular, FFmpeg with LCEVC | V-NOVA [12] explains how to encode with LCEVC SDK within FFMPEG, whilst LCEVC Best Practices | V-NOVA [13] provides advice on how to apply LCEVC-enhancement to an encoding workflow.

E.15 Carriage of LCEVC in broadcasting and streaming applications

E.15.1 Storage of LCEVC in NALU File Format

MPEG ISO Base Media File Format (ISO BMFF) standardized as ISO/IEC 14496-12, is a container for storage and carriage of timed media [8]. LCEVC, in the same way as other MPEG codecs, is built on a network abstraction layer (NAL) structuring interface, and the data structures carried by this interface are referred to as NAL units.

Support for storage of LCEVC within NALU File Format is specified in ISO/IEC 14496-15:2022/Amd 1:2023 which was published in October 2023.

The storage mechanism is based on a “dual-track” approach, whereby the base bitstream is identified by one track ID (a “Base ID”), and the LCEVC bitstream refers to one and only one Base ID.

The two main concepts used to link the LCEVC enhancement to its corresponding XVC base (where X stands for "whatever coding format", e.g. AVC, HEVC, EVC, VVC) are:

- Link the LCEVC enhancement to the XVC base by means of an "sbas" box identifying which is the corresponding base;
- Use the same time base for the "stts" boxes in the enhancement, as they are used in the base, to perform picture-to-picture synchronization.

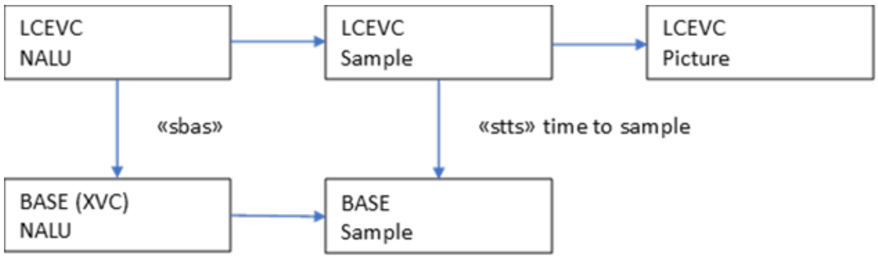

Figure E.11 – LCEVC storage within NALU File Format storage mechanism

E.15.2 CMAF profile for LCEVC

MPEG Common Media Application Format (CMAF) standardized in ISO/IEC 23000-19 [9] is a common streaming format driven by the convergence between MPEG-DASH and HTTP Live Streaming (HLS) [10].

A CMAF profile for LCEVC has been finalized as Final Amendment to ISO/IEC FDIS 23000-19 in April 2023 and is expected to be merged with other unrelated pending amendments in the coming months.

The LCEVC profile is built upon (i) the “dual-track” structure as described in E.2 and (ii) the existing CMAF construct of Dependent CMAF Tracks, as defined in Clause H.1 of ISO/IEC 23000-19.

E.15.3 LCEVC in SEI messages

LCEVC could also be transported within a SEI message using the ITU-T T.35 syntax already specified in all existing base codecs. V-Nova was assigned a manufacturer code under the UK Register of Manufacturer Codes for H.221/H.245/T.120 [11] which could be used for this purpose. In particular, with that code, the payload would have sei_payload(0x04, payloadSize) = user_data_registered_itu_t_t35 (payloadSize), with the following format.

Note that the itu_t_t35_country_code is followed by three additional fields specific to the National Register for the UK.

Table E.10 – ITU-T T.35 syntax with the values specified by V-Nova UK manufacturer code

Syntax	Descriptor	Value
user_data_registered_itu_t_t35(payloadSize) {		
itu_t_t35_country_code	b(8)	0xB4
itu_t_t35_payload_header {		
t35_uk_country_code_second_octet	b(8)	0x00
t35_uk_manufacturer_code_first_octet	b(8)	0x50
t35_uk_manufacturer_code_second_octet	b(8)	0x00
}		
i = 4		
do {		
itu_t_t35_payload_byte	b(8)	
i++		
} while(i < payloadSize)		
}		

Annex F

HDR10 Media Profile (HDR10)

F.1 Scope

This Annex provides Operational Guidelines for the use of HDR10. HDR10, formerly known as HDR10 Media Profile, is an open High Dynamic Range video profile created by the Consumer Technology Association and released in August 2015. HDR10 is built on Recommendation ITU-R BT.2100.

For the current version of this document, no further information is provided regarding the usage of HDR10 on TV 3.0.

Annex G

Dynamic Metadata for Color Volume Transform - Application #1 (Dolby Vision)

G.1 Scope

This Annex provides Operational Guidelines for the use of Dolby Vision, formerly known as Dynamic Metadata for Color Volume Transform – Application #1 (SMPTE ST 2094-10).

The information presented can be found in the Ultra HD Forum's Guidelines Technology Implementations Guidelines Section 7.1, also known as the UHDF Guidelines.

G.2 Overview

DM App #1 provides the content creator, distributor and consumer benefits in picture quality across numerous display platforms. DM App #1 allows the content creator an assurance that the creative intent is preserved across multiple devices. Content distributors are offered the flexibility of including dynamic metadata that was created with the content, or they could generate dynamic metadata during the encoding step. Consumers receive benefits since their viewing experience is consistent with other users of DM App #1. Finally, display manufacturers can realize a demonstrable improvement in display performance - even for lower cost displays - that is consistent and closely resembles the content as it was mastered.

The inclusion of DM App #1 metadata in the broadcast content chain is simple for both live linear and streaming applications thereby relieving the broadcaster of complexity while providing a vastly improved audience experience.

Figure 1 in the UHDF Guidelines illustrates the functional block diagram for the generation and bitstream insertion of DM App #1 metadata. Although the picture illustrates an encoding functional block being a HEVC 10/12-bit encoder, in the scope of SBTVD TV 3.0 this functional block is actually a VVC 10-bit encoder. The HPU (HDR Processing Unit) has the option of receiving metadata from external sources or generating metadata when external metadata is not present. The resulting metadata is then multiplexed into the compressed video bitstream.

Figure 2 shows the decoding function and display management for DM App #1. Although the picture illustrates a decoding functional block being a HEVC 10/12-bit decoder, in the scope of SBTVD TV 3.0 this functional block is actually a VVC 10-bit decoder. The metadata is extracted from the bitstream, the video is decoded, and then the resulting HDR imagery is reconstructed. The HDR images have a tone curve applied that is created from the transmitted metadata, display characteristics, and viewing environmental information.

G.3 Encoding

The encoding system used for DM App #1 is described in Section 7.1.4 in the UHDF Guidelines. Figure 4 in the UHDF Guidelines provides a conceptual illustration of the content workflow. Content is mastered and assembled at master control using the master control display as the Mastering Display. DM App #1 (Dolby Vision) is already deployed in most production encoders. Figure 5 in the UHDF Guidelines illustrates a detailed workflow with multiple sources including live cameras, post-produced content, replays, and commercials. The originating content can be in SDR or multiple types of HDR content including PQ, HLG, and SLOG. Figure 5 in the UHDF

Guidelines contains an interim topology where the metadata is created just prior to encoder in the HPU. The same metadata can be used to derive an SDR version saving on duplicate camera and production crews.

Figure 6 in the UHDF Guidelines illustrates the same workflow but with support for SMPTE ST 2110-40 SDI metadata carriage allowing production metadata to replace derived metadata if it is present in the video streams. When present, the metadata is passed through to the encoded bitstream. When metadata is not present, DM App #1 metadata is created at the encoder in the HPU.

G.4 Display Mapping

Figure 2 in the UHDF Guidelines illustrates the broad range of devices that may receive and process DM App #1 Metadata. An example display mapping function can be found in Annex B of SMPTE ST 2094-10.

Annex H

Dynamic Metadata for Color Volume Transform - Application #4 (HDR10+)

H.1 Scope

This Annex provides Operational Guidelines for the use of HDR10+, formerly known as Dynamic Metadata for Color Volume Transform – Application #4 (SMPTE ST 2094-40).

The goal in deploying SMPTE ST 2094-40 dynamic metadata is to improve picture quality without adding workflow complexity that could impact infrastructure or require additional human intervention and effort. Deployment does not complicate quality checking or manipulation of the video at any point in the contribution and delivery chain.

H.2 Approach

The SMPTE ST 2094-40 dynamic metadata is produced automatically in the encoder at the final stage in the signal chain prior to transmission to viewers. This is illustrated in the workflow depicted in Figure H.1.

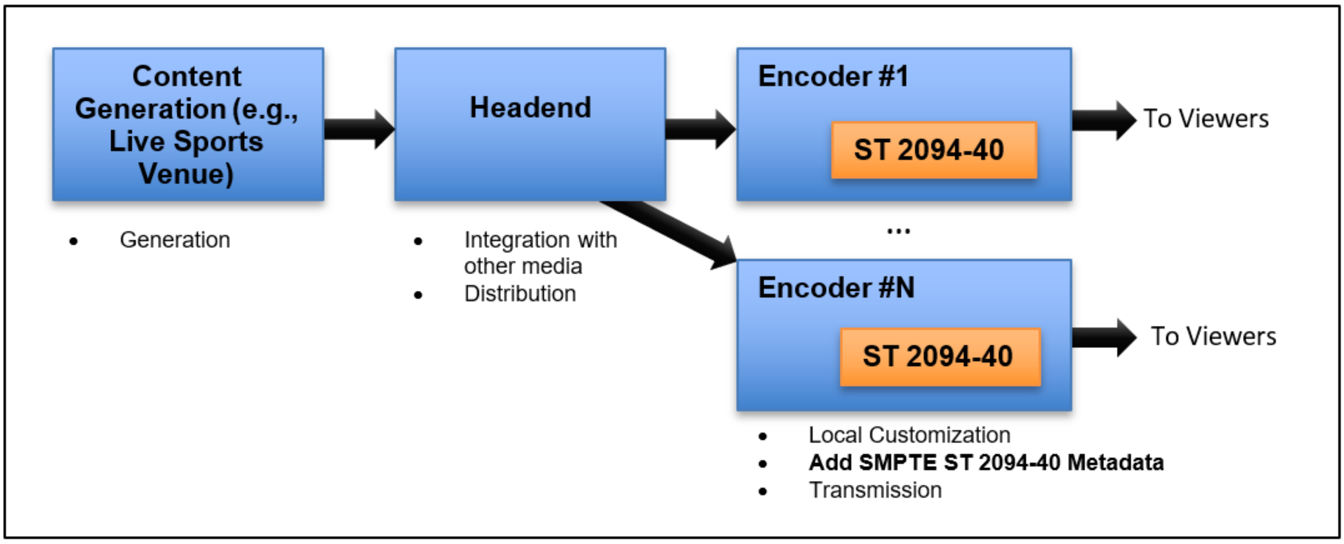

Figure H.1 – Insertion of SMPTE ST 2094-40 Metadata

In this approach, generation of the metadata takes into account all previous processing, including shading, color correction, overlay addition, resizing, and other video effects and processes. From an operational perspective, it simply is a matter of enabling the feature at the level of the encoder operator interface. There is no need to make adjustments or to apply custom settings.

Enabling the feature at the final encoding stage avoids the need to transport, align, or process the metadata within the material infrastructure. Using an automatic process avoids customization activity thereby reducing risk of human error and eliminating incremental burden for staff.

H.3 Basic Deployment Aspects

As shown in Figure H.1, uncompressed video is delivered to the encoder, which compresses the video for broadcast and, if directed by the operator, calculates and inserts the SMPTE ST 2094-40 dynamic metadata. The compressed video codecs include support for HEVC, VVC and LCEVC. The metadata are calculated frame-by-frame and encapsulated in SEI messages associated with the video. The SEI message conveying the metadata is small, comprising about 60 bytes per frame. Other than ensuring that the ‘switch’ to generate the metadata has been set, there is no operator intervention required.

When received, the compressed video is decoded. For devices that support SMPTE ST 2094-40, the SEI messages are processed and used to generate an HDR10 image with SMPTE ST 2094-40 dynamic metadata. For devices that do not support SMPTE ST 2094-40, the SEI messages are ignored and an HDR10 image is generated.

H.4 Inputs to Encoder

Source video may be Standard Dynamic Range or native High Dynamic Range. To prepare for SMPTE ST 2094-40 calculation and insertion, Standard Dynamic Range material should be converted to PQ-based High Dynamic Range (HDR10). Depending on the equipment deployed, this may be done by a separate device, process or internally within the encoder. For native High Dynamic Range material, it may be necessary to convert to HDR10 (e.g., for HLG source). Again, this may be done by a separate device, process or within the encoder.

H.5 Verification Within the Signal Chain

SMPTE ST 2094-40 metadata generation is intended to be done at the final stages of preparation for distribution and consumption. However, encoding normally is done at multiple stages in the signal chain (mezzanine generation by the content provider, contribution, remote distribution from live venues, and distribution from a headend to local broadcasting stations). With suitably equipped encoders, SMPTE ST 2094-40 metadata generation can be performed for verification purposes at any of these stages. Note that any metadata generated within the chain would be discarded before the final encoding for transmission. It would only be added mid-chain for monitoring purposes.

The main reason that one might want SMPTE ST 2094-40 metadata generation within the signal chain is to check the results on SMPTE ST 2094-40 capable displays. An example could be to ensure that camera shading at an HDR live production will look good on lower-performing consumer displays, in addition to on broadcast monitors.

H.6 Monitoring

The visual results can be checked intermittently by staff. Any SMPTE ST 2094-40 capable display can be used for such monitoring

Annex I

Single Layer HDR system Part 2 (SL-HDR2)

I.1 Scope

This Annex provides Operational Guidelines for the use of SL-HDR2, specified in ETSI TS 103 433-2 V1.4.1 (2021-08).

I.2 TV distribution with SL-HDR2 and H.265|HEVC or H.266|VVC video codecs

The Ultra HD Forum produces guidelines for UHD and focuses on end-to-end workflows from the camera to the consumer. Operational guidelines for SL-HDR2 are provided in chapter 8.4 of the Indigo Book of the Ultra HD Forum Guidelines 3.1.0. This document references the Ultra HD Forum Guidelines 3.1.0 and provides additional clarifications in relation to the usage of SL-HDR2 in TV 3.0 video coding.

The first part of chapter 8.4 deals with distribution to consumer devices such as TV sets and STBs.

Figure 15 represents a typical use case of SL-HDR2 being used for distribution of HDR content and addresses both integrated decoder/displays and separate decoder/displays such as a STB connected to a display. Figure 15 illustrates HDR to HDR display mapping and figure 16 illustrates HDR to SDR display mapping.

Figure 17 describes an integrated TV set, where SL-HDR2 is integrated and delivers the display mapped HDR signal directly to the integrated display.

Figure 18 represents the case where a STB decodes the HDR video, performs the HDR to HDR display mapping based on SL-HDR2 dynamic metadata and passes the display mapped HDR video signal via HDMI to the TV set for display. This case does not require SL-HDR support in the TV set.

Figure 19 describes the case where a STB decodes the HDR video and passes the PQ HDR video and the SL-HDR2 dynamic metadata via HDMI to the TV set. The TV set performs display mapping based on SL-HDR2 dynamic metadata and passes the resulting HDR video to the integrated display. This case requires the support of SL-HDR in the TV set.

Figure 20 represents the case where multiple PQ HDR channels including SL-HDR2 dynamic metadata are received and composited in a STB. The composited PQ HDR video and related SL-HDR2 dynamic metadata are passed via HDMI to the TV set. The TV set is the same as in Figure 19.

Descriptions in chapter 8.4 are valid for DASH and HLS based television distribution and all SL-HDR2 supported video codecs including HEVC and VVC.

I.3 TV contribution with SL-HDR2

The second part of chapter 8.4 of the Indigo Book of the Ultra HD Forum Guidelines 3.1.0 deals with using SL-HDR2 as a contribution feed to an HDR facility (Figure 21) and as contribution feed to an SDR facility (Figure 22).

Annex J

Single Layer HDR system Part 1 (SL-HDR1)

J.1 Scope

This Annex provides Operational Guidelines for the use of SL-HDR1, specified in ETSI TS 103 433-1 V1.4.1 (2021-08).

J.2 TV distribution with SL-HDR1 and H.264|AVC codec

Operational guidelines for SL-HDR1 and H.264|AVC video codec are provided in ABNT NBR 15608-2 Annex D. Operational guidelines below in J.3 are also valid for AVC and provide complementary information to ABNT NBR 15608-2 Annex D.

J.3 TV distribution with SL-HDR1 and H.265|HEVC codec

The Ultra HD Forum produces guidelines for UHD and focuses on end-to-end workflows from the camera to the consumer. Operational guidelines for SL-HDR1 are provided in chapter 8.3 of the Indigo Book of the Ultra HD Forum Guidelines 3.1.0. This document references the Ultra HD Forum Guidelines 3.1.0 and provides additional clarifications in relation to the usage of SL-HDR1 in TV 3.0 video coding.

The first part of chapter 8.3 deals with distribution to consumer devices such as TV sets and STBs.

Figure 8 represents a typical use case of SL-HDR1 being used for distribution of HDR content and addresses both integrated decoder/displays and separate decoder/displays such as a STB connected to a display.

Figure 9 describes an integrated TV set, where SL-HDR1 is integrated and delivers the reconstructed HDR signal directly to the integrated display.

Figure 10 represents the case where a STB decodes the SDR video, reconstructs the HDR signal based on SL-HDR1 dynamic metadata and passes the HDR video signal via HDMI to the TV set for display. This case does not require SL-HDR support in the TV set.

Figure 11 describes the case where a STB decodes the SDR video and passes the SDR video and the SL-HDR1 dynamic metadata via HDMI to the TV set. The TV set reconstructs the HDR signal based on SL-HDR1 dynamic metadata and passes the HDR video to the integrated display. This case requires the support of SL-HDR in the TV set.

Figure 12 represents the case where SDR channels including SL-HDR1 dynamic metadata are received and composited in a STB. The composited SDR video and related SL-HDR1 dynamic metadata are passed via HDMI to the TV set. The TV set is the same as in Figure 11.

Descriptions in chapter 8.3 are valid for DASH and HLS based television distribution and all SL-HDR1 supported video codecs including AVC and HEVC.

J.4 TV contribution with SL-HDR1

The second part of chapter 8.3 of the Indigo Book of the Ultra HD Forum Guidelines 3.1.0 deals with using SL-HDR1 as a contribution feed to an HDR facility (Figure 13) and using SL-HDR1 as a contribution feed to an SDR facility (Figure 14).

Annex K

Hybrid Log-Gamma (HLG)

K.1 Scope

This Annex provides Operational Guidelines for the use of the Hybrid Log-Gamma (HLG). The HLG transfer function was jointly developed by the BBC and NHK for high dynamic range (HDR) display. It is backward compatible with the transfer function of SDR (Recommendation ITU-R BT.709), and it was published as ARIB STD-B67 by the Association of Radio Industries and Businesses (ARIB). It is also defined in Recommendation ITU-R BT.2100.

For the current version of this document, no further information is provided regarding the usage of HLG on TV 3.0.

Annex L

Visual Volumetric Video-based Coding (V3C)

L.1 Scope

This Annex provides Operational Guidelines for the use of Visual Volumetric Video-based Coding (V3C). The V3C standard defines a generic volumetric video coding mechanism used in XR immersive videos.

The Moving Picture Experts Group (MPEG) has specified two applications that use V3C: Video-based Point Cloud Compression (V-PCC) (ISO/IEC 23090-5) and MPEG Immersive Video (MIV) (ISO/IEC 23090-12:2023).

L.2 Video-based Point Cloud Compression (V-PCC)

L.2.1 V-PCC profile considerations

Three V-PCC profiles have been identified in the technical specification of the SBTVD TV3.0 standard [Ref1]:

1. “TV30 V-PCC base”
2. “TV30 V-PCC enhanced”
3. “TV30 V-PCC high”

Each profile has been defined as an improvement of the previous one, giving the opportunity to operators to target different applications including performance compromises.

A fairly complete description of the V-PCC codec released by the MPEG 3D Graphics and Haptics coding group is available in [21]. The document describes tools and encoding strategies that may be useful to understand the implementation of the V-PCC standard [19].

A performance analysis done in the SMPTE Motion Imaging Journal [22] provides a comparative study on the combination of different V-PCC tools and gives conclusions in terms of objective and subjective tests while comparing different profiles. Guidance provided in this document is based on [22].

L.2.1.1 “TV30 V-PCC base”

This is the simplest profile in terms of decoding complexity. It is recommended to limit the coding of geometry and attribute data to one map for the current atlas so that it can be easily deployed on mobile devices. Therefore, the rendering of points in the 3D space is straightforward with no reconstruction tools activated.

This profile is suited to devices with low end decoding capabilities and where an operator does not require to preserve creative intent.

L.2.1.2 "TV30 V-PCC enhanced"

This profile is the same as the previous one in terms of coding tools, but it provides reconstruction information in the bitstream that enables the receiver device to improve reconstruction compared to simple reconstruction in 3D space and in addition to preserve creative intent. For obtaining best reconstruction results, it is recommended to use two maps for the current atlas and activate the Occupancy Synthesis with Patch Border Filtering and the Duplicate Point Removal tools.

This profile is suited to devices which support the decoding of at least two maps and where an operator requires to preserve creative intent and keep control of the reconstruction visual quality.

L.2.1.3 "TV30 V-PCC high"

This profile permits the decoding with high quality features, the adding of supplemental orientation projections, Point Local Reconstruction or Occupancy Map Synthesis tools, etc. This profile can be selected for applications requiring a higher fidelity.

Best reconstruction results can be obtained with only one map for the current atlas and by activating the Point Local Reconstruction and the Occupancy Synthesis with Patch Border Filtering tools. In this case, a 1-map profile gives as good results as a 2-map configuration but with less constraints on memory consumption and synchronization requirements due to two frames for geometry and attributes and one frame for occupancy.

In the case of usage of a 2-map configuration, tools such as Duplicate Point Removal mainly brings computation time and memory savings.

As the previous profile, it provides reconstruction information in the bitstream that enables the receiver device to improve reconstruction compared to simple reconstruction in 3D space and in addition to preserve creative intent.

This profile is suited to devices with higher decoding capabilities and where an operator requires to preserve creative intent and keep control of the reconstruction visual quality.

L.3 MPEG immersive video

L.3.1 MIV profile considerations

MIV achieves compression of the input signals through two means:

- Pruning of redundant pixels
- 2D video coding

A decoded MIV access unit (frame) consists of one or more atlases, each with patches that relate the atlas coordinates to projection plane coordinates. Each atlas has occupancy, geometry, texture and/or transparency. It depends on the profiles which components are allowed.

The pruning makes the codec asymmetric because only part of the input pixels is reconstructed. A MIV renderer uses the decoded access unit to render a viewport with 6 degrees of freedom (6DoF) (position and orientation), within a viewing space limited in position and field of view that is indirectly determined by the source views. It is also possible to provide the viewing space directly using an SEI message.

Rendering is non-normative and an implementer has the room to differentiate by providing a better or more efficient rendering. The Hypothetical Reference Renderer ([20] Annex H) describes how to reconstruct samples of a decoded MIV access unit to a point in scene space, and the Test Model for MPEG immersive video [23] includes three example viewport renderers.

The first two profiles are alternatives:

- TV30 MIV multi-view base profile is for having multiple views with depth information.
- TV30 MIV multi-planar base profile is for having a single view with multiple transparent layers.

The final profile is a superset of both:

- TV30 MIV extended profile enables having multiple views with depth and transparency information.

L.3.1.1 TV30 MIV multi-view base profile

This profile is relatively simple to decode with two video components (texture plus geometry with embedded occupancy) and at most two atlases. It enables encoding from capture systems with minimal post processing, thus enabling live encoding. Typical source data may be derived from:

- a capture system consisting of one or more depth cameras
- a capture system comprising multiple 2D cameras and a multiview depth estimator
- a computer graphics model and a ray tracer to synthesize multiple views

It is possible to render directly from the atlas with Z-buffer rendering, but because the appearance of scene objects may be view-dependent, an improved quality is realized when view selection and blending is applied. Two renderers in [23] are suitable for this profile:

- View Weighting Synthesizer
- Additive Synthesizer

The video components of an atlas can be encoded separately resulting in two video sub-bitstreams, or both components can be packed into one video frame, thus reducing the number of decoder instantiations from two to one. There can be one or two atlases, and the packing choice can be made independently for each atlas. This is a trade-off because there are also constraints on the maximum atlas frame size.

L.3.1.2 TV30 MIV multi-planar base profile

This profile is relatively simple to decode with two video components (texture plus transparency) and at most two atlases. It enables back-to-front or front-to-back rendering without Z-buffering or view blending which is more efficient than the TV30 MIV multi-view base profile. However, this complexity is shifted to the encoder, which is required to construct the multi-planar format. Typical source data may be derived from:

- a capture system consisting of one or more depth cameras, plus a plane extractor
- a capture system comprising multiple 2D cameras, a multiview depth estimator, and a plane extractor

- a computer graphics model and a ray tracer to synthesize multiple layers

One renderer in [23] is suitable for this profile:

- MPI Synthesizer

The video components of an atlas can be encoded separately resulting in two video sub-bitstreams, or both components can be packed into one video frame, thus reducing the number of decoder instantiations from two to one. There can be one or two atlases, and the packing choice can be made independently for each atlas. This is a trade-off because there are also constraints on the maximum atlas frame size.

L.3.1.3 TV30 MIV extended profile

This profile is a superset of the other two MIV profiles, and it enables the combination of texture, geometry and transparency components. This additional freedom enables the encoder to more accurately model the scene. The flipside is that the decoder may have an additional video component to decode, and the renderer needs to handle multiple views with geometry and transparency. A renderer may have to be more flexible or choose between different render paths based on the VPS.

The video components of an atlas can be encoded separately resulting in at most three video sub-bitstreams per atlas, or some or all components can be packed together into one video frame, thus reducing the number of decoder instantiations to at least one. This is a trade-off because there are also constraints on the maximum atlas frame size.

L.4 Video codec considerations

HEVC allows fast deployment as at the time of publication this 2D video codec is largely deployed on consumer video devices.

VVC provides substantial gains in compression compared to HEVC from which V3C benefits directly as the V3C encoder relies on 2D video codec to encode its atlases of texture, geometry, occupancy and/or transparency.

Bibliography

[1] ABNT NBR 25603, TV 3.0 – Video Coding

[2] Media Coding Industry Forum (MC-IF), VVC Technical Guidelines (v1.0)

[3] MPEG, 2022-01-25, White paper on Low Complexity Enhancement Video Coding (LCEVC)

[4] "Market needs and requirements for software-based capability extension on top of existing and future video codecs", in ISO/IEC JTC1/SC29/WG11 M44832, Macao, CN, Oct. 2018

[5] "Verification Test Report on the Compression Performance of Low Complexity Enhancement Video Coding, available at https://www.mpeg.org/wp-content/uploads/mpeg_meetings/134_OnLine/w20173.zip"

[6] "Verification Test Report on the Compression Performance of Low Complexity Enhancement Video Coding, available at https://www.mpeg.org/wp-content/uploads/mpeg_meetings/134_OnLine/w20173.zip"

[7] "S. Ferrara et al., "The Next Frontier For MPEG-5 LCEVC: From HDR and Immersive Video to the Metaverse," in IEEE MultiMedia, vol. 29, no. 4, pp. 111-122, 1 Oct.-Dec. 2022"

[8] ISO/IEC 14496-12:2022, Information technology — Coding of audio-visual objects, Part 12: ISO base media file format, available at https://www.iso.org/standard/83102.html

[9] ISO/IEC 23000-19:2024, Information technology — Multimedia application format (MPEG-A), Part 19: Common media application format (CMAF) for segmented media, available at https://www.iso.org/standard/85623.html

[10] RFC 8216, HTTP Live Streaming, available at https://datatracker.ietf.org/doc/html/rfc8216

[11] UK Register of Manufacturer Codes for H.221/H.245/T.120, available at https://www.cix.co.uk/~bpechey/H221/h221code.htm

[12] FFmpeg with LCEVC | V-NOVA website, available at https://docs.v-nova.com/technologies/mpeg-5.lcevc/integrations/ffmpeg

[13] LCEVC Best Practices | V-NOVA website, available at https://docs.v-nova.com/technologies/mpeg-5.lcevc/v-nova-lcevc/lcevc-best-practices

[14] Ultra HD Forum Guidelines 3.1.0: 2023, *INDIGO - Ultra HD Technology Implementations,* available at https://ultrahdforum.org/wp-content/uploads/IndigoBook-Ultra_HD_Technology_Implementations_3.1.0.pdf

[15] SMPTE ST 2110-40: 2018, *Professional Media Over Managed IP Networks: SMPTE ST 291-1 Ancillary Data*

[16] SMPTE ST 2094-10: 2016, *Dynamic Metadata for Color Volume Transform - Application #1*

[17] Recommendation ITU-R BT.2020, *Parameter values for ultra-high definition television systems for production and international programme exchange*

[18] Recommendation ITU-R BT.2100, *Image parameter values for high dynamic range television for use in production and international programme exchange*

[19] ISO/IEC 23090-5:2023(2E), Information technology — Coded Representation of Immersive Media — Part 5: Visual Volumetric Video-based Coding (V3C) and Video-based Point Cloud Compression (V-PCC)

[20] ISO/IEC 23090-12:2023(E), Information technology — Coded Representation of Immersive Media — Part 12: MPEG immersive video

[21] V-PCC codec description, MDS20352, WG 07 MPEG 3D Graphics coding, MPEG134 Online, April 2021, available at https://www.mpeg.org/wp-content/uploads/mpeg_meetings/134_OnLine/w20352.zip

[22] SMPTE, C. Guede, P. Andrivon, J. -E. Marvie, J. Ricard, B. Redmann and J. -C. Chevet, "V-PCC Performance Evaluation of the First MPEG Point Codec," in SMPTE Motion Imaging Journal, vol. 130, no. 4, pp. 36-52, May 2021, doi: 10.5594/JMI.2021.3067962, V-PCC Performance Evaluation of the First MPEG Point Codec | SMPTE Journals & Magazine | IEEE Xplore

[23] Test Model 15 for MPEG immersive video, MPEG 140 Online, October 2022, available at https://www.mpeg.org/wp-content/uploads/mpeg_meetings/140_Mainz/w22173.zip software: https://gitlab.com/mpeg-i-visual/tmiv , document: https://www.mpeg.org/wp-content/uploads/mpeg_meetings/140_Mainz/w22173.zip