INTERNATIONAL ORGANISATION FOR STANDARDISATION

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 MPEG2014/N14510**

**April 2014, Valencia, Spain**

|  |  |
| --- | --- |
| **Source:** | Requirements |
| **Status:** | Approved |
| **Title:** | Draft Requirements and Explorations for HDR/WCG Content Distribution and Storage |
| **Editor(s):** | Ajay Luthra, Edouard François, Walt Husak |

**Abstract**

Currently, TVs providing Standard Dynamic Range (SDR) typically support content with brightness in the range of the order of 0.1 to 100 nits. However, that range is significantly smaller than the range that the human eye can see in real life. Future TVs and various video distribution environments are expected to give a viewing experience that is closer to a real life experience and to provide to a user the sense of “being there”. This requires supporting significantly higher dynamic ranges as well as broader color gamut. This document provides requirements, use cases and test conditions associated with it.

# Introduction

Currently, TVs providing Standard Dynamic Range (SDR) typically support brightness in the range of the order of 0.1 to 100 nits. However, that range is significantly smaller than the range encountered in real life. For example, a light bulb can have much more than 10,000 nits, surfaces lit in the sunlight can have brightness upwards of 100s of thousands of nits, while the night sky can be as low as 0.005 nits (or lower).

One of the key aspects of Ultra High Definition TVs (UHDTVs) is to provide a user a sense of “being there” and “realness”. This requires creating, capturing and displaying the content that has much higher peak brightness and much larger contrast values than today’s TVs. Simply increasing resolution is not significant enough to attain this goal.

In addition, this also requires providing colors that are significantly richer than the ones provided by today’s standard, e.g. BT. 709. Thus, the new content will not only have several orders of magnitude larger brightness and contrast, but also significantly wider color gamut (for example, BT. 2020 or even wider than that in the future).

The content may consist of computer-generated animation and camera captured video or images. It may be distributed to a consumer electronically (i.e. via a network connection) or physically (on optical media, flash memory, magnetic storage, etc.).

As both the dynamic ranges and the volumes of the color gamut associated with various capturing and displaying devices are expected to increase significantly, it is not clear at this stage if existing MPEG video coding standards are able to efficiently support the needs of future TV as well as other content distribution environments. This document pulls together the needs of future high quality content distribution systems, including future 4k UHDTV, that support content with much higher dynamic ranges and wider color gamut. In this process, a more complete picture that involves the end-to-end chain, from video generation to final destination (in places where MPEG video coding standards are used) is considered. That chain includes the stages of creation (studios), capture, intermediate (mezzanine) level distribution, and final distribution to the home (see Annex A).

# Definitions

# Higher Dynamic Range

Overall, the dynamic range of a scene can be described as the ratio of the maximum light intensity to the minimum light intensity. In digital cameras, the most commonly used unit for measuring dynamic range is in terms of f-stop, which describes total light range by powers of 2. The current ad hoc use of the term f-stop, refers to the following dynamic ranges:

10 f-stops = a difference of 210 = 1024: 1 contrast ratio.

14 f-stops = a difference of 214 = 16,384: 1 contrast ratio.

16 f-stops = a difference of 216 = 65,536: 1 contrast ratio. It is normally regarded as 100,000:1; approximately what the eye can see in a scene with no adaptation. *[Ref?]*

20 f-stops = a difference of 220 = 1,048,576: 1 contrast ratio. It is normally regarded as 1,000,000:1; approximately what the eye can see in a scene with minimal (no noticeable) adaptation. *[Ref?]*

In the ad hoc categorization of the dynamic ranges, the following definitions are typical:

* Standard Dynamic Range (SDR) is ≤ 10 f-stops
* Enhanced Dynamic Range (EDR) is > 10 f-stops and ≤ 16 f-stops
* High Dynamic Range is (HDR) > 16 f-stops

# Scene Referred and Display Referred pictures

Scene referred pictures relate to the real luminance captured from the scene. It corresponds to how an image appeared when it was originally captured. Scene Referred values are linearly related to the amount of light in the depicted scene. In a Scene Referred pipeline the processed image is not directly viewable.

Display Referred corresponds to how an image gets rendered on a specific display. The pixel sample values in the captured scene or associated graded sequence may get modified to match the capabilities of the display. For example, if a scene is captured or represented in BT.2020 color space with colors out of the display gamut (e.g. Rec.709 display) then at least the out-of-gamut colors get modified to match the display capabilities. Similarly, the luminance/luma values in the captured or graded scene may get modified to match the dynamic range and the peak luminance of the display.

# Video Level Considerations and Requirements

# Higher Dynamic Range

The dynamic ranges of the content and the displays should be decoupled.

The achievable and desired brightness and the dynamic ranges of various displays may be significantly different than those on the capturing and creating ends. For example, a content creation system may be able to create or capture content with contrast of 1,000,000:1 but it may be neither desirable nor feasible to have displays with that range. So, there may be a need to do some display dependent mapping of the content’s dynamic range. That mapping may be done at the encoding end or the receiving end. This may also be a function of the distribution mechanism, e.g. point-to-point communication or broadcasting.

Therefore, it is important that the standard be largely future proof in terms of maximum brightness and contrast range and it should be developed with flexibility in their values. Specific limits may be specified in the context of profiles and levels specifications. The standard should be able to migrate from where the content capturing and displaying ranges are today to where they will be in the medium to the long terms.

The standard should allow signaling of meta-data to assist in the conversion of the bit-stream’s dynamic range to the display’s dynamic range.

# Content and Input Types and Bit Depths

All three types of content should be supported:

* Computer generated (animation)
* Camera captured video
* Images

The standard shall support integer (10, 12 and 16 bits) and half-floating point (IEEE 754) input video data formats. (Comment: Investigate/Experiment with 8 and 10 bits inputs)

(*Comments: One or several compression internal image integer formats can be defined. As the standard is likely to operate internally using fixed point arithmetic, mechanisms should be provided that would allow an encoder to map floating point formats to the appropriate integer formats that the encoder considers to be the most efficient.)*

A mechanism to indicate the mapping used to create the input integer values provided to an encoder shall be provided.

Mechanism should be provided that would allow a receiver to map the decoded video format to the one needed for display.

Note: Negative values of the signal (e.g. XYZ, RGB, YCbCr) shall be allowed. *(To be further discussed. Also, do we need to have a cap on the negative values?)*

Coded EDR and HDR video for consumer distribution may be non-full range.

*Note: As an example from SDI    [0,4\*D-1] and*

*[219\*D,  2^N-1] (luma)*

*[224\*D, 2^N-1] (chroma)*

*where D = (1<<(BitDepth-8))*

# Electro Optic Transfer Function (EOTF)

Input video to the encoder may or may not have an EOTF applied to it. This applies to both integer and floating point representations. Multiple EOTFs should also be supported. Mechanism to map the input EOTF (linear or non-linear) to another function, which may be more efficient from a coding efficiency point of view or appropriate from display point of view, should also be provided.

Luma and Chroma may have different EOTFs.

# Colour Sampling and Formats

The standard should support 4:2:0, 4:2:2 and 4:4:4 chroma formats. Chroma sub-sampling and the domain in which it gets performed may have a visual quality and coding efficiency impact and should be studied.

# Colour Spaces

The standard should support multiple color spaces. Key color spaces of interest include (but not limited to):

1. CIE 1931 XYZ
2. Recommendation ITU-R BT.2020
3. DCI-P3 (SMPTE ST 428-1:2006)
4. Recommendation ITU-R BT.709
5. CIE Luv (CIE 1976)

*Comment: White points need to be clarified.*

Higher Dynamic Range (EDR and HDR) can be combined with any of the color spaces, for example, BT. 709, BT. 2020 or XYZ or others. The standard should allow signaling of meta-data to assist in the conversion of the bit-stream color space to a display color space. Impact of compression noise and interaction of the compression noise with those mapping schemes should be studied. More generally, the impact of potential interactions between the transfer functions, color space conversions, chroma sub/up-sampling and compression should be studied.

Multiple color differencing scheme should be supported. (*Comment: Need to provide more details, especially for CIE 1931 XYZ and YDzDx.*)

# Picture Formats

# Spatial Resolutions

The standard should focus on a set of rectangular picture formats that would include all commonly used picture sizes. Key spatial resolutions of interest include (but not limited to):

High Definition: 1920 × 1080 (Progressively scanned) or 1280 × 720p

Ultra High Definition: 3840 × 2160

Full 4k: 4096 × 2160

# Frame Rates

Typical frame rates of 24, 25, 50 and 60 fps should be supported. The standard should support commonly used non-integer frame rates (for example, 29.97). Standard should allow support of higher than 60 Hz (e.g. 100, 120 Hz) frame rates.

*Comment: Do we want to support non-integer frame rates for > HDTV resolution. Get feedback from other SDOs and Revisit.*

Variable frame rate should be supported.

# Compression Performance

Impact of compression noise and interaction of the compression noise with various dynamic range mapping schemes should be studied.

Visually lossless compression should be supported.

*(Comment: is there a need for mathematically lossless compression? This may have implication for floating point inputs.)*

Appropriate method(s) of objectively measuring compression efficiency and visual quality needs to be developed. See Annex B for further details.

(Compression efficiency – To be further discussed)

# Multiple Generations of Encoding/Transcoding

The quality of picture should not degrade significantly as it goes through multiple generations of encoding. In studios, that is done as a part of video processing. In the consumer space, a content stored or sent to a set top may be transcoded for further distribution in home or outside the home. In a head end, content may be transcoded to multiple different bit rates and resolutions for Adaptive Bit Rate streaming using, for example, DASH.

*Comment: Some capture or mastering applications may use full video range, usually a high bit-rates and non-subsampled chroma (e.g. 4:4:4), prior to re-encoding to contribution or distribution.*

# Picture processing

The quality of picture should not degrade significantly as it goes through different kind of picture processing such as frame rate conversion, denoising, chroma keying.

*Note: to be discussed further*

# Overlays processing

The overlays of graphics over the content should not degrade the quality of the content. The overlay may include logos as well as other data. The content may also be displayed side by side with other content that may not be of same dynamic range. The version of the standard that is used in the studios should be such that various special effects (for example, chroma keying, page turns, etc.) are feasible, after multiple encoding and decoding generations, without significant loss in the quality.

The content may be combined with graphics, including closed captioning, and different videos with different ranges and/or color spaces can be combined.

# Backward and Forward Compatibility

As the content creation, capturing and display technologies are continuing to progress at a fast pace, it is very important to have a standard that does not put undue burden on today’s devices but also provides backward compatibility with them as the new technology gets developed and deployed.

Scalability based video coding approach could be considered as means to provide that compatibility. It is recognized though that there could also be a need for non-backward compatibility where the performance of such a scheme is significantly improved because the need to maintain backward compatibility is removed.

Backward compatibility with low bit-depth decoders conforming with low bit-depth (8- or 10-bits) could be considered.

Rendering compatibility with displays of different capabilities could be considered.

Single layer case

* Metadata used in post-processing mechanisms to convert HDR decoded data into a lower dynamic format could be enabled.

Multi-layer case

* A lower dynamic version of the HDR bitstream could be extracted and decoded.
* Metadata used in post-processing mechanisms to convert a lower dynamic version of the HDR version into a higher dynamic range version could be enabled.

The following are some of the scalability schemes that are expected to play a key role in future deployments.

# Temporal Scalability

The scalability extension of the standard should support 2 or more layers with different frame rates in factors of integer multiples, for example from 60 Hz to 120 Hz or higher.

# Colour space (gamut) scalability

To achieve compatibility and inter-operability among various systems that will support different color spaces and to provide good migration paths from narrower to wider color gamut systems, a multiple layer (scalability) extension of the standard may also be developed. The scalability extension should support a base and enhancement layer(s) with different color spaces. Inter-layer prediction methods that predict values in an enhancement layer color space from values in a base layer color space could be considered.

*Comment: one intended application for CGS is to enhance an SDR backward compatible bitstream (e.g. content display referred to BT.709)  with side information to either better predict a WCG construct or serve as a metadata side stream tone “up” mapping*

# SDR to HDR Scalability

To achieve compatibility and inter-operability among various systems that will support different dynamic ranges and to provide good migration paths from SDR to higher dynamic range systems, a multiple layer (scalability) extension of the standard may also be needed. Some systems may also need different frame rates, as well as different color gamut, for SDR and HDR layers.

# Metadata

Metadata may be optionally added by the content publisher to describe the content or to help post-processing steps to be applied before delivery or display.

The content publisher may optionally provide metadata to describe the color space of the reference display that the content was mastered on. The metadata may refer by name to a predefined display or consist of the relevant parameters for a general display model (for example, color primaries, white point, peak luminance and black level). This metadata is strictly informative and does not alter the intended presentation.

Optionally, to aid in the conversion of content for delivery to a display with a different color volume, a standardized set of title-specific conversion metadata may be provided by the content publisher. As an example, this metadata might be composed of mathematical operations (such as a 3x3 matrix) in conjunction with 1D or 3D Look-Up Tables.

The metadata should allow a feasible implementation of real-time decoding and post processing with computational and memory power available to devices at the consumer electronics level for B2C use case.

# Complexity

The complexity should allow for feasible implementation of encoders and decoders within the constraints of the available technology at the expected time of usage. It should be capable of trading-off complexity, compression efficiency and visual quality, by having:

* One or more operating points with similar complexity compared to HEVC v1 and RExt but with better visual quality,
* One or more operating points with increase complexity and commensurate increase in visual quality.

The complexity of the metadata and of the processes using these metadata should be studied.

*Note: Complexity includes: Power consumption, computational power, memory bandwidth etc.*

# Profiles and Levels

To be defined for various subsets of applications.

# System Level Considerations

Coding technology should specify efficient and easy ways to carry the bitstreams over widely used protocols and networks. The compressed video layers will get transported using MPEG-2 Systems, MPEG DASH and various Internet Protocols. Coding technology should specify efficient and easy ways to store the bitstreams over widely used file formats (for example, MP4 …).

**ANNEX A**

**Use Cases and Associated Attributes**

# A.1 Contribution

# - Multiple Generations/Cycles of encoding and decoding

# - Editable and low delay (Very short GOPs, including Intra-only)

- Mixing of sequences and chroma keying

- Very High Fidelity

- Relatively very high bit rates (lower compression ratios)

- …

# A.2 Mezzanine (a.k.a B2B)

(Note: Includes delivery from the studio to the distribution channel)

# - High fidelity

# - Relatively high bit rates

- Long GOPs (and Short GOPs? Intra only?)

- Transcoding to various formats (e.g. Lower bit rates, SDR, narrower color gamut, lower resolution etc.)

- …

# A.3 Distribution to consumers/end-users (a.k.a B2C)

(Note: Includes delivery from the distribution channel to the consumer)

# - Acceptable fidelity

# - Wide range of bit rates

- Wide range of GOP size

- Transcoding to various formats (e.g. Lower bit rates, SDR, narrower color gamut, lower resolution etc.) for in-home as well as outside the home distribution.

- …

**ANNEX B**

**Test Conditions and Test Sequences**

# Purpose

The purpose of this Annex is to describe evaluation test conditions to potentially start the work associated with the development of HDR and WCG coding standards. In this document the focus is only on 2D video.

# Tentative timeline

Start exploration and defining Exploratory Experiments (EEs) in April 2014 (see Annex C).

Results of EEs are expected by Oct 2014 /Feb 2015. Key results to enable the decision to proceed to CfP are targeted for October 2014.

TBD

# Test conditions

# Test material under consideration

# File Exchange Formats - OpenEXR and TIFF

The filenames are specified as follows: Name\_Resolution\_Fps\_Format\_Primaries\_xxx.yyy

with

* Name: sequence name
* Resolution: picture size (e.g. 1920x1080p)
* Fps: frame rate in frame per second
* Format: format of the samples (e.g. ff for 32-bit floating point, hf for half-float 16-bit floating point, 10b for 10-bit integer)
* Primaries: color primaries (e.g. 709, 2020, P3)
* xxx: frame number

Add text file with each clip stating:

 - how video was color timed / graded.

 - exposure

 - how computational photography/videography was applied (if any)

 - whether or not there are any gain factors assumed in the source video

 - rolloff or knee functions applied in the camera or in post

 - camera settings)

* Mastering display meta data

# Test sequences under consideration

The following test sequences, provided by Technicolor, are under consideration for the evaluation tests.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Resolution** | **Primaries** | **EOTF** | **Sequence name** | **Sxx** | **fps** | **Frames** |
| 1920x1080  RGB 4:4:4 | BT.709 | linear | Seine\_1920x1080p\_25\_hf\_709\_xxx.exr  Balloon\_1920x1080p\_25\_hf\_709\_xxx.exr  Fire eater2\_1920x1080p\_25\_hf\_709\_xxx.exr  Tibul2\_1920x1080p\_30\_hf\_709\_xxx.exr  Market3\_1920x1080p\_50\_hf\_709\_xxx.exr | S01  S02  S03  S04  S05 | 25  25  25  30  50 | 0-199  0-199  0-199  0-239  0-399 |

The test sequences are available in the “testsequences/half-float-sequences” directory of the FTP site a”xxxxx”. Further details of access to the FTP site can be obtained from xxxxx.

Movielabs suggested (m32806) to also consider following sequences for the evaluation tests:

|  |  |  |
| --- | --- | --- |
| No. | Brief description | Master frames |
| 1. | StEM Confetti through Juggler | 1784 to 2144 |
| 2. | StEM Magic Hour: Fountain to Table | 3527 to 3887 |
| 3. | StEM Warm Night: Torch to Table (before slomo) | 6280 to 6640 |
| 4. | StEM Cool Night: Fog | 8208 to 8568 |
| 5. | StEM Rain | 9872 - 10232 |
| 6. | Telescope Interior Halo |  |
| 7. | Telescope Hyperspace to Debris to Interior | 5900 to 6260 |
| 8. | Telescope Nebula + Bubble to Interior | 8200 to 8560 |

These are display referred to P3 primaries, 4000 nits max. luminance with 16-bit precision.

(Note: need more details – format, cost, how to get it, legal constraints on the usage etc.)

More sequences may be provided in July/Oct 2014.

Group should decide which ones to use.

# Coding Conditions

Three coding constraint conditions are defined:

* C1: All Intra (AI)
  + All pictures are coded as Intra pictures.
* C2: Low delay (LD)
  + The first picture is an Intra picture, and there are no backward references for inter prediction (bi-prediction may be applied, but with no picture reordering between decoder processing and output).
* C3: Random Access (RA)
  + A structural delay of processing units not larger than 8-picture "groups of pictures (GOPs)" (e.g., dyadic hierarchical B usage with 4 levels), and random access intervals of 1.1 seconds or less will be used (Intra random access picture every 24, 24, 32, 48 and 64 pictures for 24 fps, 25 fps, 30 fps, 50 fps and 60 fps sequences, respectively).

*Comments:*

*are AI and LD required ?*

*Colour formats:*

* *The pro*vided RGB or XYZ sequences shall be processed using the above test conditions and coding constraints.
* The chroma resolution is 4:4:4.

*Comments: do we also consider 4:2:0 ?*

Additional constraints:

1. Only use of pre-processing if it is part of the encoding process and involves a related normative post-processing step in the decoding process.
2. Only use post-processing if it is a normative part of the decoding process. Any such post-processing must be documented.
3. Quantization settings should be kept static. When change of quantization is used it shall be described.
4. Proponents are discouraged from optimizing encoding parameters using non-automatic means.
5. The video coding test set shall not be used as the training set for training large entropy coding tables, VQ codebooks, etc.
6. Usage of multi-pass encoding is limited to the picture level and must be documented.
7. The encoder parameters (QP, lambda, or similar encoder optimizations are allowed to be changed once while coding a sequence, in order to meet the specified file size limits).

# Operational points under consideration

For the C1 (All Intra) scenario, the following operational points are considered.

Table 1: target rate points (Mbit/s) not to be exceeded for C1 configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sequence** | **Rate 1** | **Rate 2** | **Rate 3** | **Rate 4** | **Rate 5** |
| Seine\_1920x1080p\_25\_hf\_709 | 10 | 15 | 20 | 30 | 50 |
| Balloon\_1920x1080p\_25\_hf\_709 | 10 | 15 | 20 | 30 | 50 |
| Fire-eater2\_1920x1080p\_25\_hf\_709 | 10 | 15 | 20 | 30 | 50 |
| Tibul2\_1920x1080p\_30\_hf\_709 | 12 | 18 | 24 | 36 | 60 |
| Market3\_1920x1080p\_50\_hf\_709 | 20 | 30 | 40 | 60 | 100 |
| … | … | … | … | … | … |

For the C2 (Low Delay) and C3 (Random Access) scenarios, the following operational points are considered.

Table 2: target rate points (Mbit/s) not to be exceeded for C2 and C3 configurations.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sequence** | **Rate 1** | **Rate 2** | **Rate 3** | **Rate 4** | **Rate 5** |
| Seine\_1920x1080p\_25\_hf\_709 | 1.5 | 2 | 3 | 5 | 8 |
| Balloon\_1920x1080p\_25\_hf\_709 | 1.5 | 2 | 3 | 5 | 8 |
| Fire-eater2\_1920x1080p\_25\_hf\_709 | 1.5 | 2 | 3 | 5 | 8 |
| Tibul2\_1920x1080p\_30\_hf\_709 | 2 | 2.75 | 4 | 7 | 10 |
| Market3\_1920x1080p\_50\_hf\_709 | 2 | 2.75 | 4 | 7 | 10 |
| … | … | … | … | … | … |

*Comments:*

* *are AI and LD required ?*
* *is it needed to add higher rates to cover a wider range?*
* *Values to be refined further.*
* *AI points roughly correspond to the Intra average bit-rates required for the I-RAP pictures of the RA simulations.*

# Anchors

Anchor bitstreams are available in the “xxxxx” directory of the FTP site “xxxx”. Further details of access to the FTP site can be obtained from xxxx.

Following anchor data will be generated:

# Anchor1 for EEs - Main10 profile

Anchor1 bitstreams are generated using the coding / decoding chain illustrated in Figure 1. The encoding chain comprises two main steps:

* A signal conversion step which comprises the following steps (the inverse conversion being the inverse of this process):
  + Conversion of the input HDR signal into the RGB BT. 2020, XYZ color space,
  + Color transformation from RGB to YCbCr and XYZ to YDzDx,
    - EOTF (PQ) before conversion YCbCr, YDzDx
  + Chroma downsampling from 4:4:4 to 4:2:0
    - Use JCT-VC downsampling method
    - Do it in YCbCr and YDzDx domain
  + Signal quantization to 10 bits integer.
* The encoding step (and inversely decoding step) performed using the HM13.0,
* Use RA configuration (for now)

color

space

conv

transfer

function

color transf

chroma downsamp.

quant 10bits

integer

encoding

HM 4:2:0 10 bits

input HDR video

bit-stream

inverse transfer

function

inv. color trans

chroma upsamp.

inversequant 10bits

output HDR video

bit-stream

decoding

HM 4:2:0 10 bits

inv. color

space

conv

Figure 1: simplified encoding / decoding chains of anchor1.

# Anchor 2 for EEs - Main 12 Profile

Same as Main 10. RA Configuration and settings.

# Anchor 3 (Main 444:12) - TBD

Anchor 3 bitstreams are generated using the coding / decoding chain illustrated in Figure 2. The encoding chain comprises two main steps:

* A signal conversion step which comprises the following steps (the inverse conversion being the inverse of this process):
  + Conversion of the input HDR signal into the xxx color space,
    - (xxx=RGB, XYZ …?)
  + Mapping with the transfer function yyy
    - (yyy=inverse gamma - which value: 2.2, 2.4, …? , PQ-OETF … ?)
  + Color differentiating from xxx to zzz,
    - (zzz =YUV, Ydzdx … ?),
  + Signal quantization to 12 bits integer.
* The encoding step (and inversely decoding step) performed using the HM13.0-RExt6.0, with internal bit-depth of 12bits.

*Comments:*

* *xxx, yyy, zzz to be defined*
* *use a more recent HM version if available when evaluation tests are done*
* *do we consider 10 bits instead of 12 bits ?*

color

space

conv

transfer

function

color diff

quant 12bits

integer

encoding

HM-RExt 4:4:4 12 bits

input HDR video

bit-stream

inverse transfer

function

inv. color diff

inversequant 10bits

output HDR video

bit-stream

decoding

HM-RExt

4:4:4 12 bits

inv. color

space

conv

Figure 2: simplified encoding / decoding chains of anchor2.

Corresponding encoding configuration files ‘encoder\_An\_Cm.cfg’ for the two anchors n=1,2 and three coding constraint conditions m=1,2,3 are attached to the document.

# Objective evaluation

The objective quality will be measured using distance metrics between the output of the decoder (after any possible mapping) and the input of the encoder (before any possible mapping). This is illustrated in Figure 3, where the conversion may comprise steps such as mapping function (OETF), color space conversion, chroma subsampling, and the inverse conversion may comprise steps such as inverse mapping function (EOTF), inverse color space conversion, chroma upsampling.

conversion

encoding

input HDR video

decoding

inverse conversion

output HDR video

quality metric evaluation

Figure 3: end-to-end objective quality evaluation.

# Objective quality metrics under consideration

Possible objective metrics for

Preferably, an objective metric providing one single number for the three color components space can be used.

For example, the following objective metrics could be used:

* PSNR (When HDR is converted to multiple exposure pictures, PSNR can be helpful – input contribution for the next meeting). TBD
* *E2000*

The CIE has defined the *E* metric to measure the perceptual distance or difference between two colors. The latest refinement of this metric is the *E2000* [1]. The *E2000* metric may be used to measure the shift in color due to quantization of the color image signals.

mean\_*E2000* is computed as the average over all pictures of the sequence of the *E2000* metric measured for each pixel.

The *E2000* PSNR (PSNR*E*) is computed for each picture as follows:

with *Emean* being the average *E2000* over the picture. In case of half-float representation, *peak*=65504 is the peak value. The PSNR is averaged over the entire sequence.

* + mean\_*E2000*
  + variance *E2000*
  + *E2000* SNR
  + *E2000* PSNR
* Others (to be explored, e.g. Lab Space, modified SSIM etc.)

If possible, proper spatial and temporal pooling of objective metrics could potentially be performed.

Results shall be reported using the Excel spreadsheet template provided in attachment. This template will include metrics such as file size, average bit-rate, objective quality metrics measurements, BD-Rate metrics compared against anchor(s), and encode and decode run-times.

# Subjective evaluation

TBD. See EEs in Annex C.

*Comments: Should we view HDR content on SDR display(s)? If yes then how to do the HDR to SDR conversion? See EE9 in Annex C.*

*Comment: Only when evaluating HDR—>SDR tone mapping (metadata) schemes.*

# Information to be delivered

The following information shall be provided:

1. A technical description of the proposal sufficient for full conceptual understanding and generation of equivalent performance results by experts and for conveying the degree of optimization required to replicate the performance. The description comprises information about:

* single-layer or multi-layer decoding;
* the internal bit-depth of the encoding and decoding processes;
* the internal chroma resolution (e.g. 4:4:4, 4:2:0) of the encoding and decoding processes;
* the usage of metadata: type of information, frequency (per sequence, per gop, per frame …), relative coding cost.

1. Description of the conversion processes applied to the input HDR signal prior to the encoding process, and after the decoding process to generate the output HDR signal, such as:

* mapping transfer function (OETF) and inverse mapping transfer function (EOTF)
* color space conversion and inverse conversion
* chroma subsampling and upsampling

1. The technical description shall state how the proposed technology behaves in terms of random access to any frame within the sequence. For example, a description of the GOP structure and the maximum number of frames that must be decoded to access any frame could be given.
2. The technical description shall specify the expected encoding and decoding delay characteristics of the technology, including structural delay e.g. due to the amount of frame reordering and buffering, the degree of frame-level multi-pass decisions and the degree by which the delay can be minimized by parallel processing.
3. The technical description shall contain information suitable to assess the complexity of the implementation of the technology, including the following:

* Encoding time for each submitted bitstream of the software implementation. Proponents shall provide a description of the platform and methodology used to determine the time. To help interpretation, a description of software and algorithm optimizations undertaken, if any, is welcome.
* Decoding time for each bitstream running the software implementation of the proposal, and for the corresponding constraint case anchor bitstream(s) run on the same platform. Proponents shall provide a description of the platform and methodology used to determine the time. To help interpretation, a description of software optimizations undertaken, if any, is encouraged.
* Expected memory usage of encoder and decoder.
* Complexity of encoder and decoder, in terms of number of operations, dependencies that may affect throughput, etc.
* Complexity characteristics of Motion Estimation (ME) / Motion Compensation (MC): e.g. number of reference pictures, sizes of frame memories (and associated decoder data), sample value word-length, block size, and motion compensation interpolation filter(s), if that information differs from what is already being used in the HM-13.0 anchors.
* Degree of capability for parallel processing.

# References

[1] CIE. Improvement to industrial colour-difference evaluation. Vienna: CIE Publication No. 142-2001, Central Bureau of the CIE; 2001.

[2] Other References

**ANNEX C**

**Exploratory Experiments (EE)**

Note: Members with names in bold (with email addresses) are primary coordinators

**EE1** – Experiment with multiple EOTFs (**Alexis - atourapis@apple.com**, Walt, Chad, Rocco Goris, Herbert, Jim)

Codecs have classically been tagged with a transfer function (OETF/EOTF), color primaries, and matrix coefficients. This EE aims at studying these different components. The evaluation will relate to the models already defined in the HEVC specification, and will consider potential new models.

**EE2** – Experiment with monitors (**Walt - wjh@dolby.com**, Edouard, Alexis)

**EE3** – Subjective Test methodology (**Walt - wjh@dolby.com**, Touradj, Vittorio, Alexis)

**EE4** – Objective Test Methods (**Edouard -** **Edouard.francois@technicolor.com**, Walt, Jacob, Alexis)

Different objective distortion metrics are commonly used for SDR such as PSNR or SSIM. Unfortunately, most of these metrics are suitable for SDR, but are not usable as is for HDR. This experiment aims at identifying distortion metrics that are relevant to evaluate the performance of HDR video coding solutions.

Following topics should be addressed:

* in which domain (linear-light, OETF-mapped),
* in which color space (XYZ, Lab …),
* what metrics (PSNR, DeltaE, Mean Square error, SNR, Mean relative error, …),
* using full range signal or multi-exposure distortions

**EE5** – Subsampling and its impact on coding performance and video quality (**Herbert - herbert.thoma@iis.fraunhofer.de**, Rocco, Peng, Pankaj, Jianle, Alexis)

This experiment aims for a better understanding of how chroma subsampling interacts with different color representations (e.g. YDzDx, CIE u’v’) and how it impacts coding performance and video quality.

It also aims at evaluating how important are 4:2:2 and 4:4:4 chroma subsampling formats for HDR/WCG distribution. In case of 4:2:2 or 4:2:0 format, the EE can also study more intelligent downsampling/upsampling in order to improve the compression performance.

**EE6** – Coded Color spaces (**Alexis - atourapis@apple.com**, Jacob, Edouard, Rocco, Walt, Niko, Pankaj, Jianle)

SDR video is classically coded in a de-correlated YUV or YCbCr space. The aim of this EE is to investigate if this approach is still adapted to HDR video, or if a new representation such as XYZ or a de-correlated version is required. The EE also aims at studying the properties these representations should have, especially for typical consumer applications bitrates.

Impacts of using different color spaces on encoding algorithms should also be studied.

**EE7** – Bit depth (**Peng - pyin@dolby.com**, Rocco, Edouard, Jim, Alexis)

How many bits we need (Main10 vs Main12)

**EE8** – Content Characteristics (Walt - **wjh@dolby.com**, Alexis, TK Tan)

**EE9** – Conversion from HDR to SDR for viewing content on SDR TVs (**Alexis -atourapis@apple.com,** Jacob, Walt)