ATSC Standard:
A/341:2019 Amendment No. 2, ST 2094-40, etc.

Doc. A/341:2019 Amend. No. 2
17 September 2021

Advanced Television Systems Committee
1776 K Street, N.W.
Washington, D.C. 20006
202-872-9160

Note: The user's attention is called to the possibility that compliance with this standard may require use of an invention covered by patent rights. By publication of this standard, no position is taken with respect to the validity of this claim or of any patent rights in connection therewith. One or more patent holders have, however, filed a statement regarding the terms on which such patent holder(s) may be willing to grant a license under these rights to individuals or entities desiring to obtain such a license. Details may be obtained from the ATSC Secretary and the patent holder.

Implementers with feedback, comments, or potential bug reports relating to this document may contact ATSC at https://www.atsc.org/feedback/.

Revision History

Version	**Date**
Candidate Standard approved	31 March 2021
Amendment approved	17 September 2021

ATSC Standard:
A/341:2019 Amendment No. 2, ST 2094-40, etc.

1. OVERVIEW

1.1 Definition

An Amendment is generated to document an enhancement, an addition or a deletion of functionality to previously agreed technical provisions in an existing ATSC document. Amendments shall be published as attachments to the original ATSC document. Distribution by ATSC of existing documents shall include any approved Amendments.

1.2 Scope

This amendment clarifies the term "access unit" and identifies a CMAF Media Profile for ST 2094-10. It also standardizes the emission of dynamic metadata based on SMPTE ST 2094-40:2020 "Dynamic Metadata for Color Volume Transform — Application #4" [1] (henceforth, identified as "ST 2094-40") in the ATSC 3.0 system as specified by CTA-861-H [2]. Some editorial items are improved.

1.3 Rationale for Changes

"Access unit" was not clearly defined. A lack of clarity could be a challenge for implementers and could harm interoperability.

SCTE has recently standardized a CMAF Media Profile for ST 2094-10. This information may be of interest to implementers and it could be helpful if A/341 adopts CMAF in the future.

Dynamic metadata based on SMPTE ST 2094-40 and displays that support the technology exist in the marketplace. This amendment standardizes the emission of ST 2094-40 metadata in the ATSC 3.0 system with the intent of enhancing interoperability.

1.4 Compatibility Considerations

The clarification of "access unit" is informative, rather than normative, so it does not cause compatibility issues. This new clarification potentially improves compatibility.

The identification of a CMAF Media Profile for ST 2094-10 is informative. It adds no normative impact on implementations.

The use of dynamic metadata based on SMPTE ST 2094-40 is optional for emissions and is optional for receivers. The metadata messages are carried by SEI messages in the HEVC bitstream, which are designed to be ignored by receivers that have no interest in the encapsulated data; therefore, no compatibility issues are expected.

2. CHANGE INSTRUCTIONS

Change instructions are given below in *italics*. Unless otherwise noted, inserted text, tables, and drawings are shown in blue; deletions of existing text are shown in ~~red strikeout~~. The text "[ref]" indicates that a cross reference to a cited referenced document should be inserted.

2.1 Normative References

Add the following references.

[1] SMPTE: "Dynamic Metadata for Color Volume Transformation – Application #4," Doc. ST 2094-40:2020, Society of Motion Picture and Television Engineers, White Plains, NY.

[2] CTA: "A DTV Profile for Uncompressed High Speed Digital Interfaces," Doc. CTA-861-H (December 2020), Consumer Technology Association, Arlington, VA.

2.2 Informative References

Add the following reference.

[3] SCTE: "HEVC Video Constraints for Cable Television Part 1-1 HDR," Doc. SCTE 215-1-1 2020b, Society of Cable Telecommunications Engineers, Inc., Exton, PA.

2.3 Modify Section 6.3.1.4

Add the following footnote to the term "access unit" in Section 6.3.1.4.

- The random access picture period of the enhancement layer shall be equal to or greater than the random access picture period of the base layer. When an access unit[4] includes an IRAP picture with nuh_layer_id greater than 0, it shall also include an IRAP picture with nuh_layer_id equal to 0.

[4] The term 'access unit' is defined by the HEVC standard [2].

2.4 Modify Section 6.3.2.1.1

The changes in this section are intended to be purely editorial.

Modify the first paragraph of Section 6.3.2.1.1 as follows. Note the change from bold to non-bold.

~~**SL-HDR1 Metadata**~~SL-HDR1 Metadata messages are an aggregation of parameters, syntactically and semantically specified in [11], Section 6, which may be used in the decoding process to reconstruct HDR video from an SDR encoded video stream.

Modify the second paragraph of Section 6.3.2.1.1 as follows.

An HEVC or SHVC video stream may contain SL-HDR1 Metadata in order to provide both an SDR picture and an HDR picture from the same video stream. When SL-HDR1 Metadata messages are present, they allow reconstructing the HDR video from the received SDR video stream. The reconstructed HDR video can be represented as linear light or using any of the available HDR transfer functions listed in these specifications.

Modify the fourth paragraph of Section 6.3.2.1.1 as follows. Note the change from bold to non-bold.

~~If~~ When present~~ in an HEVC or SHVC bitstream~~, SL-HDR1 Metadata parameters are encapsulated into an ~~HEVC~~ SEI message~~, **SL-HDR Information SEI message**,~~ , named SL-HDR

Information SEI message, specified in [11], Section A.2.2. Mapping SL-HDR Information SEI message syntax elements to SL-HDR1 Metadata is documented in [11], Section A.2.3.

2.5 Modify Section 6.3.2.2.1

Add the word "message" to this sentence in the first paragraph of 6.3.2.2.1.

Furthermore, this metadata message can be used to derive an SDR (ITU-R BT.709 [4]) picture by receiving devices such as an ATSC 3.0 receiver/converter.

Add the following paragraph at the end of Section 6.3.2.2.1.

SCTE 215-1-1 2020b [3] defines an HEVC CMAF Media Profile that includes the carriage of SMPTE 2094-10:2016 dynamic metadata (see DM App #1).

2.6 Update Bullet in Section 6.3.2.2

Update the following bullet in Section 6.3.2.2.

- The bitstream may contain SEI messages with payloadType value equal to 4. This allows for the optional transmission of the ST 2094-10 metadata message described in Section 6.3.2.2.1 and the ST 2094-40 metadata message described in Section 6.3.2.2.2.

2.7 Add a New Subsection under Section 6.3.2.2

Add the following as a new subsection under Section 6.3.2.2 "PQ Transfer Characteristics." The new subsection is titled Section 6.3.2.2.2 "Encoding and Transport of ST 2094-40 Metadata Messages." (For better readability, the text below is not shown in in blueline.)

The HEVC video bitstream may contain ST 2094-40 metadata messages that provide for carriage of metadata elements defined in SMPTE ST 2094-1 [27] and SMPTE ST 2094-40 [1]. ST 2094-40 metadata messages, when present, provide descriptive statistical information and other information that can guide tone mapping in displays, frame by frame.

ST 2094-40 metadata, when present, shall be encoded and transported as User data registered by a Recommendation ITU-T T.35 Supplemental Enhancement Information (SEI) message per CTA-861-H [2] Annex S. The additional constraints specified in CTA-861.H [2] Annex S.4 shall apply. The syntax and semantics of the message are specified in CTA-861-H [2] Annex S.2 and Annex S.3, respectively.

When ST 2094-40 metadata messages are present, the following shall apply:

- The corresponding NAL unit type shall be set equal to PREFIX_SEI_NUT.
- One such message shall be associated with every access unit of the bitstream. If this message is present, it shall only be present once per access unit.

For information about the theory of operation for metadata based on SMPTE ST 2094-40, see Annex F below.

SCTE 215-1-1 2020b [3] defines an HEVC CMAF Media Profile that includes the carriage of SMPTE 2094-40:2020 dynamic metadata (see DM App #4).

2.8 Modify Annex E.1

Change the text related to the semantics of the metadata_refresh_flag as follows:

metadata_refresh_flag – When set equal to 1 cancels the persistence of any previous extended display mapping metadata in output order and indicates that extended display mapping metadata elements follows. The extended display mapping metadata elements persists from the coded picture to which the SEI message containing ST2094-10_data() is associated (inclusive) to the coded picture to which the next SEI message containing ST2094-10_data() and with **metadata_refresh_flag** set equal to 1 in output order is associated (exclusive) or (otherwise) to the last picture in the CVS (inclusive). When set equal to 0 this flag indicates that the extended display mapping metadata elements do ~~does~~ not follow.

2.9 Add a New Annex

Add the following as a new annex. The new annex is titled "Annex F Metadata Based on SMPTE ST 2094-40." (For better readability, the text below is not shown in blueline.)

Annex F Metadata Based on SMPTE ST 2094-40

F.1 THEORY OF OPERATION (INFORMATIVE)

F.1.1 Metadata

The metadata based on SMPTE ST 2094-40 can be considered in three groups. The first is for identification purposes. The second describes a basis OOTF (optical-optical transfer function) for a producer-specified target peak luminance. This basis OOTF can be used to construct the guided OOTF, which defines the adaptation to the presentation display. The third describes statistical characteristics of the video signal. Other elements are constrained so that they are, or effectively are, unused.

Metadata elements associated with identification are not used for image processing. These elements include:

- itu_t_t35_country_code
- itu_t_t35_terminal_provider_code
- itu_t_t35_terminal_provider_oriented_code
- application_identifier
- application_mode

Metadata elements associated with the basis OOTF include:

- targeted_system_display_maximum_luminance
- tone_mapping_flag[w]
- knee_point_x[w]
- knee_point_y[w]

- num_bezier_curve_anchors[w]
- bezier_curve_anchors[w][i]

Metadata elements associated with statistical characteristics include:

- maxscl[w][i]
- average_maxrgb[w]
- num_distributions[w]
- distribution_index[w][i]
- distribution_values[w][i]
- fraction_bright_pixels[w]

F.1.2 Basis OOTF

Prior to emission, the metadata associated with the basis OOTF is set. The basis OOTF is the transfer function for presentation at a single peak luminance point. This point is set within the peak luminance range of television sets that are available to viewers and is represented by targeted_system_display_maximum_luminance.

The basis OOTF curve depends on the content within the video images. In general, the basis OOTF compresses the dynamic range in signal ranges with lower information density and preserves contrast and details in signal ranges with higher information density. A well-constructed basis OOTF is reasonably faithful to the look of the original scene and has no discontinuities or sharp slope changes that might introduce visible artifacts into the images. Specific algorithms for creating a basis OOTF are outside of scope of this document.

The basis OOTF allows those responsible for content to view the baseline tone-mapped results for the target peak luminance level on a mastering monitor for quality control purposes. By providing this same basis OOTF to ST 2094-40-capable displays, these displays receive the same baseline as a common starting point for their individual tone mapping. OOTF curves based on SMPTE ST 2094-40 are composed by a linear part starting from (0, 0) and ending at a knee point (k_x, k_y) and a second part which is the Bezier curve with its anchor points starting with (k_x, k_y). These two parts are linked together as a smooth and continuous curve to avoid banding artifacts. Figure F.1.1 shows an example OOTF.

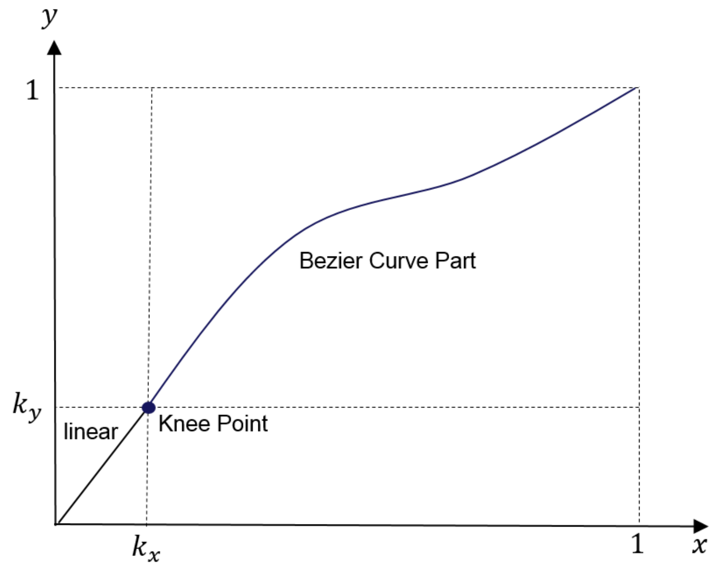

Figure F.1.1 Example OOTF.

See SMPTE ST 2094-40 [1] for more information regarding the basis OOTF.

F.1.3 Reference Method for Receiver-side Tone Mapping using ST 2094-40 Metadata

F.1.3.1. General

Figure F.1.2 shows a block diagram of a tone-mapping function based on ST 2094-40 metadata.

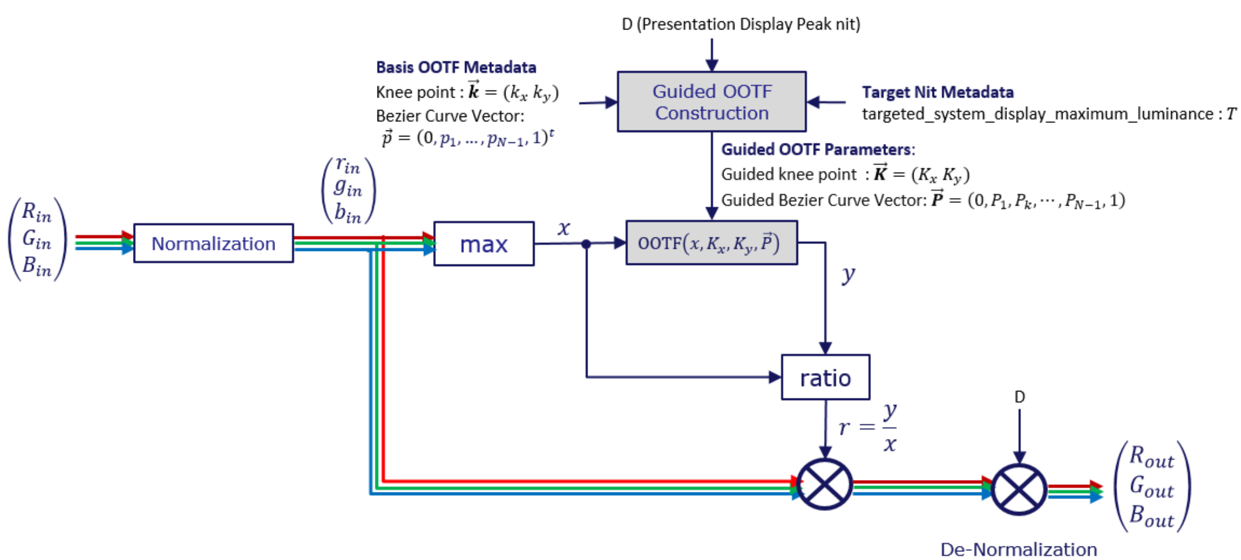

Figure F.1.2 Tone-mapping system of an ST 2094-40-based device.

Normalization is performed as follows:

The linear RGB signals $(R_{in}, G_{in}, B_{in})^t$ fed to the tone-mapping system represent absolute luminance, where t represents the vector transpose operation, and must be converted to a single channel input x before applying the guided OOTF, which processes a single input parameter.

The absolute luminance values $(R_{in}, G_{in}, B_{in})^t$ are normalized into the values between 0 and 1 by

$$\begin{pmatrix} r_{in} \\ g_{in} \\ b_{in} \end{pmatrix} = \begin{pmatrix} \min(1, R_{in}/NORM\ ) \\ \min(1, G_{in}/NORM\ ) \\ \min(1, B_{in}/NORM\ ) \end{pmatrix} \quad (1)$$

where $NORM$ is the normalization factor given by:

$$NORM = \max(D, H_M) \quad (2)$$

in which D is the peak luminance of the presentation display and H_M = **distribution_values[0][8].**

For each pixel, the maximum value of r_{in}, g_{in}, and b_{in} for that pixel is determined, as represented by x. The value of x is applied to the guided OOTF, producing the resultant value, y. The values of r_{in}, g_{in}, and b_{in} are each scaled by the ratio of y/x.

At the end of the process, the signal is de-normalized based on the peak luminance of the presentation display.

F.1.3.2. Guided OOTF Construction

F.1.3.2.1. General

The guided OOTF is based on the peak luminance of the presentation display and is derived from the basis OOTF. There are three cases:

1) The peak luminance of the presentation display (D) is equal to the target peak luminance (T), in which case the basis OOTF can be used directly;
2) the peak luminance of the presentation display (D) is greater than the target peak luminance (T), in which case the basis OOTF is effectively interpolated with a linear transfer function to create the guided OOTF;
3) the peak luminance of the presentation display (D) is less than the target peak luminance (T), in which case the basis OOTF is extrapolated to create the guided OOTF.

In each case, the goals are to create a transfer function appropriate for the presentation display, to be reasonably faithful to the nature of the guided OOTF, and to avoid introducing level or slope discontinuities that might introduce visible artifacts.

A reference method of guided OOTF generation is as follows. This method is shown to produce good results; however, it is possible that other implementations can improve the interpolation and extrapolation processes to produce results with improved detail and faithfulness to the intent of the basis OOTF.

In general, guided OOTF construction is composed of the following two parts with the inputs T (peak luminance of the target display that is obtained with the basis OOTF) and D (peak luminance of the presentation display):

- Guided Knee Point
- Guided Bezier Curve Anchors

They are described in the subsequent sections.

F.1.3.2.2. Guided Knee Point Construction

The construction of the guided knee point can be classified into two cases.

Case I: When $D \le T$

The guided knee point, $\vec{K} = \begin{pmatrix} K_x \\ K_y \end{pmatrix}$, can be obtained by:

$$\vec{K} = (w, \quad 1 - w) \cdot \begin{pmatrix} \vec{k} \\ \vec{K}_0 \end{pmatrix} \tag{3}$$

where · represents the dot product of two vectors, $\vec{k} = \begin{pmatrix} k_x \\ k_y \end{pmatrix}$ is the knee point of the given basis OOTF, $\vec{K}_0$ is a pre-defined constant vector such as $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, and w is the guided knee point mixing parameter as a function of D. There are various ways to design w; however, a linear method is simple and effective. The linear method is as shown in Figure F.1.3, where D_L is a pre-defined low luminance level, and $D_L \le T$.

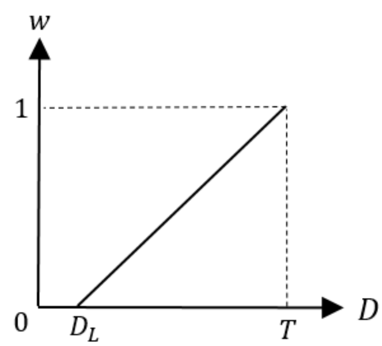

Figure F.1.3 Example of guided knee point mixing parameter function when $D \le T$.

Figure F.1.4 shows a graphical illustration of how the guided knee point is constructed as a function of D as expressed in (3), where the dotted red arrow is the Guide Knee Point trajectory.

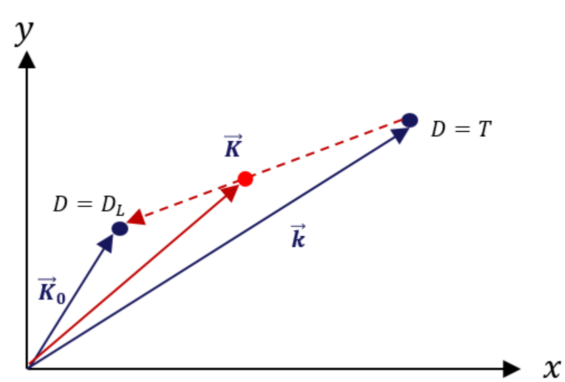

Figure F.1.4 The guided knee point when $D \le T$.

Case II: When $T \leq D$

Similar to $D \leq T$, in the case that $T \leq D$, the guided knee point, $\vec{K} = \binom{K_x}{K_y}$ can be obtained by:

$$\vec{K} = (w, \quad 1 - w) \cdot \begin{pmatrix} \vec{k} \\ \vec{K}_1 \end{pmatrix} \tag{4}$$

where $\vec{K}_1 = \binom{0.5}{0.5}$, and w is the guided knee point mixing parameter as a function of D which can be designed as shown in Figure F.1.5.

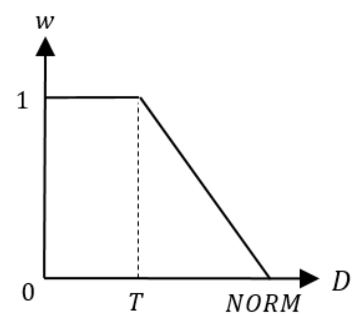

Figure F.1.5 Example of guided knee point mixing parameter function when $T \leq D$.

Figure F.1.6 shows a graphical illustration of how the guided knee point is constructed as a function of D as expressed in (4), where the dotted red arrow is the guide knee point trajectory.

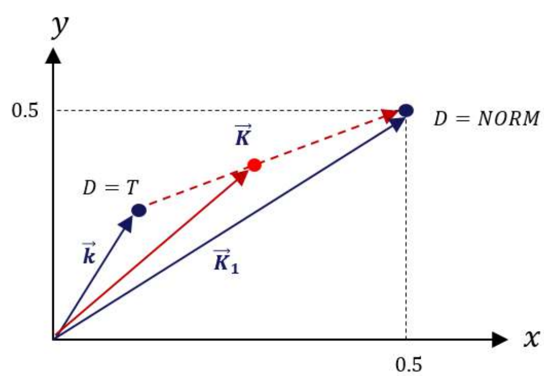

Figure F.1.6 The guided knee points when $T \leq D$.

Note that:

$$\vec{K} = \vec{K}_1 = \binom{0.5}{0.5} \tag{5}$$

when $NORM \leq D$.

F.1.3.2.3. Guided Bezier Curve Vector Construction

The following two properties of the Bezier curve are fundamental to understand the notion behind the guided Bezier curve vector construction.

Property 1: If:

$$\vec{P}_L = \begin{pmatrix} 0 \\ \frac{1}{N} \\ \vdots \\ \frac{N-1}{N} \\ 1 \end{pmatrix} \tag{6}$$

then the normalized explicit Bezier curve becomes an identity line. That is, $B_N(\vec{P}_L, t) = t$ and $\vec{P}_L$ is referred to as the identity Bezier curve vector.

Property 2: Linearly adding two Bezier curves in the same order is equivalent to linearly adding their Bezier curve vectors. That is, $B_N(\vec{\alpha}, t)$ and $B_N(\vec{\beta}, t)$ are two Bezier curves in N$^{\text{th}}$ order characterized by the Bezier curve vectors $\vec{\alpha}$ and $\vec{\beta}$, respectively, such as:

$$\begin{cases} B_N(\vec{\alpha}, t) = \sum_{k=0}^{N} C_N^k \, t^k (1-t)^{N-k} \cdot \alpha_k \\ B_N(\vec{\beta}, t) = \sum_{k=0}^{N} C_N^k \, t^k (1-t)^{N-k} \cdot \beta_k \end{cases}, \tag{7}$$

then the following can be easily shown:

$$(a, b) \cdot \begin{pmatrix} B_N(\vec{\alpha}, t) \\ B_N(\vec{\beta}, t) \end{pmatrix} = B_N(\vec{\gamma}, t) \tag{8}$$

where:

$$\vec{\gamma} = (a, b) \cdot \begin{pmatrix} \vec{\alpha} \\ \vec{\beta} \end{pmatrix} \tag{9}$$

Similar to the guided knee point construction, the guided Bezier curve vector construction from a given basis Bezier curve vector can also be classified into two cases. The basic idea behind guided Bezier curve construction is to interpolate a Bezier curve with the Bezier curve of the basis OOTF and a pre-determined boundary Bezier curve as a function of D.

Case I: When $D \leq T$

In the case of $D \leq T$, the guided Bezier curve, $B_N(\vec{P}, t)$, can be found as:

$$B_N(\vec{P}, t) = (u, 1-u) \cdot \begin{pmatrix} B_N(\vec{p}, t) \\ B_N(\vec{P}_0, t) \end{pmatrix} \tag{10}$$

where $\vec{p} = \vec{p} = (0,\ p_1,\ \ldots,\ p_{N-1},\ 1)^t$ is the Bezier curve vector of the basis OOTF, and $\vec{P}_0$ is a pre-defined Bezier curve vector such as, but not limited to, $(1,1,\cdots,1)^t$, and u is the control parameter as a function of D which can be designed as shown in Figure F.1.7 but not limited to.

By Property 2, the guided Bezier curve vector, $\vec{P}$, of $B_N(\vec{P},t)$ can be calculated as:

$$\vec{P} = (u, \quad 1-u) \cdot \begin{pmatrix} \vec{p} \\ \vec{P}_0 \end{pmatrix} \tag{11}$$

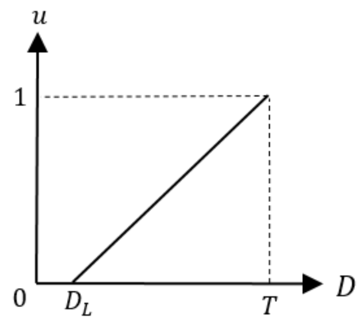

Figure F.1.7 Example of Bezier curve vector mixing coefficient function for $D \leq T$.

Case II: When $T \leq D$

In the case of $T \leq D$, the guided Bezier curve, $B_N(\vec{P},t)$, can be found as:

$$B_N(\vec{P},t) = (u, 1-u) \cdot \begin{pmatrix} B_N(\vec{p},t) \\ B_N(\vec{P}_L,t) \end{pmatrix} \tag{12}$$

where $\vec{p} = \vec{p} = (0,\ p_1,\ \ldots,\ p_{N-1},\ 1)^t$ is the Bezier curve vector of the basis OOTF, $\vec{P}_L$ is the Identity Bezier curve introduced in (6), and u is the mixing parameter as a function of D which can be designed as shown in Figure F.1.8 but not limited to.

By Property 2, the guided Bezier curve vector, $\vec{P}$, of $B_N(\vec{P},t)$ can be calculated as:

$$\vec{P} = (u, \quad 1-u) \cdot \begin{pmatrix} \vec{p} \\ \vec{P}_L \end{pmatrix} \tag{13}$$

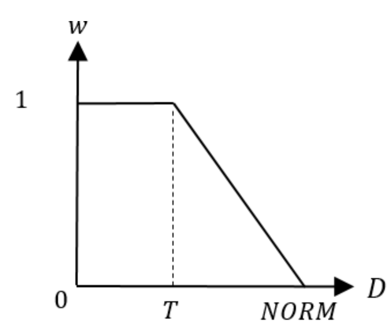

Figure F.1.8 The Bezier curve anchor control determining function for $T \leq D$.

Note that:

$$\vec{P} = \vec{P}_L \quad (14)$$

when $NORM \leq D$.

Note: By merging (5) and (14), the guided OOTF becomes the identity line as depicted in Figure F.1.9 when $NORM \leq D$, which is the case when the dynamic range of the presentation display is larger than that of the incoming HDR scene.

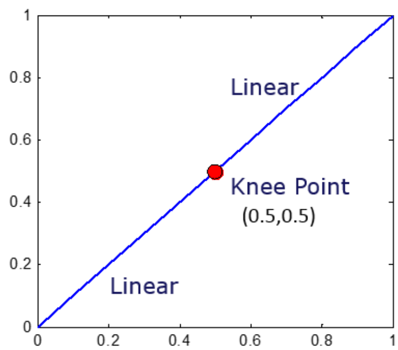

Figure F.1.9 The guided Bezier curve when $NORM \leq D$.

F.1.3.2.4. Slope Continuity Condition at the Knee Point

After the guided Knee point and Bezier curve vector are constructed, one last step is required to adjust the Bezier curve vector to ensure slope continuity at the guided knee point. Note that the linear part and the Bezier curve part of a guided OOTF curve are jointed by the knee point. Discontinuity between the two parts at the knee point may lead to banding artifacts in the tone-mapped signal. In any implementations of the ST 2094-40 tone-mapping system, the condition for slope continuity at the knee point must be satisfied. Namely, the tangent of the Bezier curve part at the knee point should be equal to the slope of the linear part, which can be expressed as:

$$\frac{d}{dx}\left(K_y + \left(1 - K_y\right) \cdot B_N\left(\vec{P}, \frac{x-K_x}{1-K_x}\right)\right)\Bigg|_{x=K_x} = \frac{K_y}{K_x} \quad (15)$$

By solving (15), the condition for slope continuity at the knee point is given as:

$$P_1 = \frac{1}{N} \cdot \frac{K_y}{K_x} \cdot \frac{1-K_x}{1-K_y} \quad (16)$$

Figure F.1.10 shows some examples of various guided OOTF curves that satisfy the condition for slope continuity at the knee point. As can be seen in Figure F.1.10, the OOTF curves are continuous and will not lead to banding artifacts.

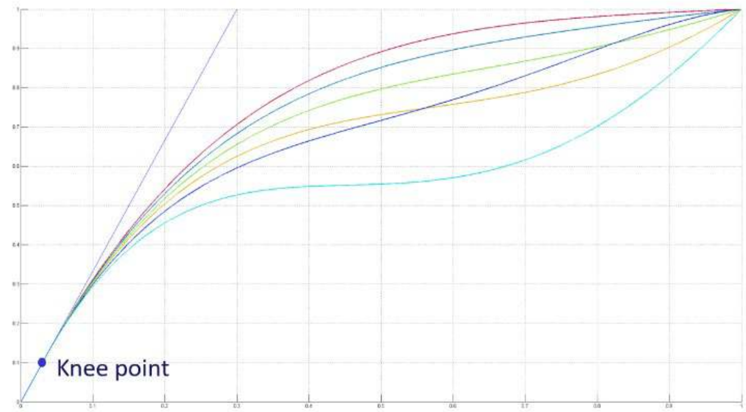

Figure F.1.10 Examples of the continuity of different guided OOTF curves at the knee point.

F.1.4 Statistical Characteristics

Statistical measurements of each video picture utilize straightforward mathematical processes that are performed before emission. See SMPTE ST 2094-40 [1] for details.

While it is possible for a receiver to make these same measurements on a frame-by-frame basis, it is more efficient to perform this process upstream. These video picture-based measurements cannot be completed until the full picture is available, potentially requiring frame buffers if calculated in the receiver. However, when measured before the emission, the calculations do not necessarily add latency. For instance, the measurement can be calculated in parallel with HEVC encoding, which necessarily adds latency of more than one picture of video, and the metadata messages are then inserted after encoding is completed. In addition, the receiver would not be able to make forward-looking, scene-based measurements.

Television sets, even with extremely high peak luminance capabilities, do not necessarily employ a neutral transfer function. The user might have selected a picture preset (Dynamic Mode) with high contrast based on personal preferences. An ambient light sensor might indicate a very bright environment, in which case low- and mid-tones might be lifted in order to ensure that the image is clearly visible. The manufacturer might also employ an adjustment to compensate for characteristics of the physical display device.

The statistical characteristic metadata elements can aid the television set in making the above adjustments. In general, a good implementation would avoid compressing the dynamic range in signal ranges where there is high information density and would instead compress the dynamic range in signal ranges with low information density. How this is achieved is left to the manufacturer and is not specified in this document.

– End of Document –