SMPTE STANDARD

Dynamic Metadata for Color Volume Transform – Core Components

Table of Contents

Approved
May 18, 2016

Foreword

SMPTE (the Society of Motion Picture and Television Engineers) is an internationally-recognized standards developing organization. Headquartered and incorporated in the United States of America, SMPTE has members in over 80 countries on six continents. SMPTE's Engineering Documents, including Standards, Recommended Practices, and Engineering Guidelines, are prepared by SMPTE's Technology Committees. Participation in these Committees is open to all with a bona fide interest in their work. SMPTE cooperates closely with other standards-developing organizations, including ISO, IEC and ITU.

SMPTE Engineering Documents are drafted in accordance with the rules given in its Standards Operations Manual.

SMPTE ST 2094-1 was prepared by Technology Committee 10E.

Intellectual Property

At the time of publication no notice had been received by SMPTE claiming patent rights essential to the implementation of this Engineering Document. However, attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. SMPTE shall not be held responsible for identifying any or all such patent rights.

Introduction

This section is entirely informative and does not form an integral part of this Engineering Document.

When image essence mastered with High Dynamic Range (HDR) and Wider Color Gamut (WCG) is transformed for presentation on a display having a smaller color volume such as a Standard Dynamic Range (SDR) display, the color volume transform process can be optimized through the use of content-dependent, dynamic color volume transform metadata rather than using only static display color volume metadata.

As the image essence characteristics change from scene to scene, the optimal transform that best reproduces the content creator's artistic intent can change. For example, the color volume transform parameters used for a very dark scene could be quite different from those used for a very bright scene.

As an alternative to archiving and managing multiple, transformed masters, one for each targeted system, the transforms can be represented as metadata synchronized with the images of the HDR master. The metadata is captured or generated as a part of the mastering process, during which images are creatively approved, and later applied as color volume transforms to the HDR master, therefore enabling alternate output encodings in the distribution stage.

Provisions are included for supporting different models of color volume transforms with associated metadata sets, herein called applications. Metadata sets from multiple applications and for multiple targeted systems can coexist, but are individually applied to the HDR master to create one or more output signals. Operational details about use and selection of metadata sets are not part of this document.

Figure 1 shows an example workflow using content-dependent, dynamic color volume transform metadata. In this example, the solid arrows represent the flow of complete image data with or without metadata and the dashed arrows represent the flow of metadata and/or image data from which metadata can be extracted. Two approved image versions are provided, one HDR master and one SDR rendition. Some applications use an actual SDR master. Other applications use an interactive SDR rendering, where, during a grading process, the colorist interactively changes a color volume transform while viewing the resulting image on an SDR display. Content-dependent, dynamic metadata is extracted from the approved versions. Mastering display metadata (such as the color volume and dynamic range) is obtained for the HDR and the SDR reference mastering displays. The metadata is carried (here marked as "+m") with the HDR master content for distribution. On playout, the metadata is utilized to transform the HDR content to the color volume and dynamic range of the consumer display device. The metadata can be applied in the distribution facility for different output media, in a consumer playout device, in the consumer display, or not at all, depending on the needs and capabilities of the distribution channel and consumer equipment.

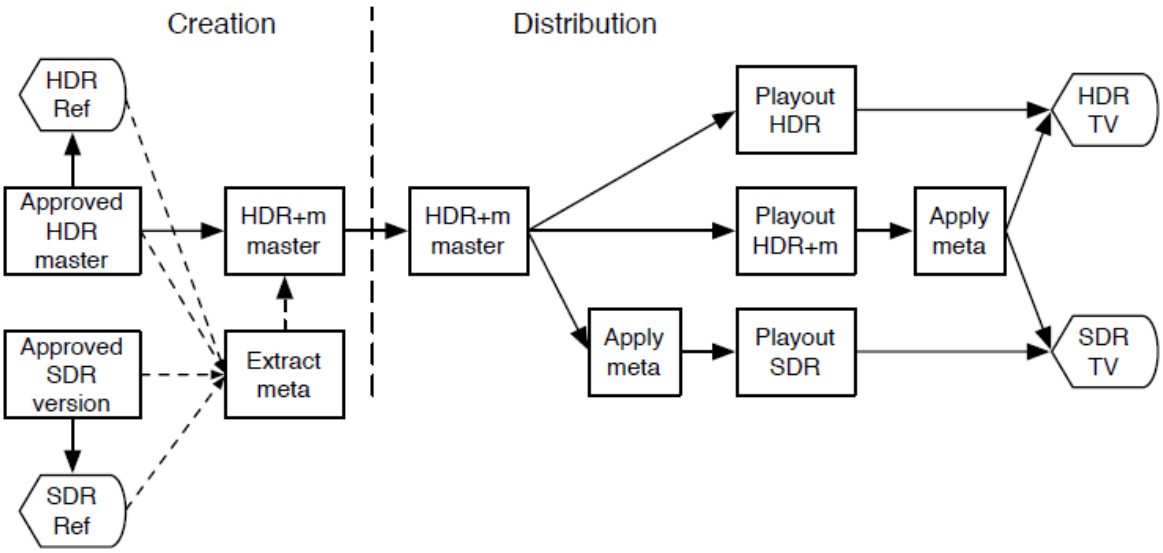

Figure 1 – Example Dynamic Metadata Workflow

1 Scope

The SMPTE ST 2094 suite of documents define metadata for use in color volume transforms of content. The metadata are content-dependent and can vary scene-by-scene or image-by-image. The metadata are intended to transform High Dynamic Range and Wide Color Gamut (HDR/WCG) image essence for presentation on a display having a smaller color volume than that of the mastering display used for mastering the image essence. Multiple applications provide particular color volume transforms.

This standard, a part of the 2094 suite, specifies common definitions, including a transform metadata set, and metadata parameters, for use by other parts of the 2094 suite.

Other parts of the 2094 suite provide representation for carriage and additional application-specific metadata parameters for color volume transforms.

2 Conformance Notation

Normative text is text that describes elements of the design that are indispensable or contains the conformance language keywords: "shall", "should", or "may". Informative text is text that is potentially helpful to the user, but not indispensable, and can be removed, changed, or added editorially without affecting interoperability. Informative text does not contain any conformance keywords.

All text in this document is, by default, normative, except: the Introduction, any section explicitly labeled as "Informative" or individual paragraphs that start with "Note:"

The keywords "shall" and "shall not" indicate requirements strictly to be followed in order to conform to the document and from which no deviation is permitted.

The keywords, "should" and "should not" indicate that, among several possibilities, one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required; or that (in the negative form) a certain possibility or course of action is deprecated but not prohibited.

The keywords "may" and "need not" indicate courses of action permissible within the limits of the document.

The keyword “reserved” indicates a provision that is not defined at this time, shall not be used, and may be defined in the future. The keyword “forbidden” indicates “reserved” and in addition indicates that the provision will never be defined in the future.

A conformant implementation according to this document is one that includes all mandatory provisions ("shall") and, if implemented, all recommended provisions ("should") as described. A conformant implementation need not implement optional provisions ("may") and need not implement them as described.

Unless otherwise specified, the order of precedence of the types of normative information in this document shall be as follows: Normative prose shall be the authoritative definition; Tables shall be next; then formal languages; then figures; and then any other language forms.

3 Normative Reference

The following standard contains provisions which, through reference in this text, constitute provisions of this engineering document. At the time of publication, the edition indicated was valid. All standards are subject to revision, and parties to agreements based on this engineering document are encouraged to investigate the possibility of applying the most recent edition of the standard indicated below.

ISO 11664-3:2012 / CIE S 014-3/E:2011, Colorimetry – Part 3: CIE Tristimulus Values

4 Terms and Definitions

For the purposes of this document, the following terms and definitions apply. Terms are defined here for use in referencing documents throughout the SMPTE ST 2094 suite.

4.1 application

document in the SMPTE ST 2094 suite that define metadata, usage constraints, and additional parameters related to a color volume transform model

4.2 application identifier

integer value identifying an application and its defining document in the SMPTE ST 2094 suite

4.3 application version

integer value specifying the version of the identified application, as given in the application's defining document

4.4 color volume

solid in colorimetric space containing all possible colors a display can produce

4.5 color volume transform

transform, mapping a coordinate in one color volume to a coordinate in another color volume

4.6 dynamic metadata

metadata that can be different for different portions of the image essence

4.7 High Dynamic Range System (HDR System)

system specified and designed for capturing, processing, and reproducing a scene, conveying the full range of perceptible shadow and highlight detail, with sufficient precision and acceptable artifacts, including sufficient separation of diffuse white and specular highlights

4.8 image essence

ordered sequence of rectangular images of same dimensions that can be indexed and the first image has the index of zero

Note: Image as defined here can be a field or a group of images.

4.9 luminance

luminous intensity of a surface in a given direction, divided by the projected area of the surface element as viewed from that direction in units of candela per square meter (cd/m^2)

4.10 matrix

two-dimensional array of numbers

4.11 maxRGB

maximum of color component values {R, G, B} of a pixel

4.12 metadata group

named collection of metadata items or metadata groups

4.13 metadata item

named value of a given type, along with an associated range of valid values of that type

4.14 metadata set

collection of metadata items or metadata groups

4.15 pixel coordinate

coordinate in a two-dimensional Cartesian coordinate system having signed integer coordinates with x coordinates increasing from left to right and y coordinates increasing from top to bottom, and (0,0) representing the top-left pixel

4.16 processing window

window for selecting image essence pixels for color volume transform (see window)

4.17 sampled function

function $y = f(x)$ represented as a list of { x_i, y_i } input/output pairs

4.18 Standard Dynamic Range System (SDR System)

system having a reference reproduction using a luminance range constrained by Recommendation ITU-R BT.2035 Section 3.2 for video applications, or SMPTE RP 431 for cinema applications

4.19 targeted system

image reproduction system addressed by a color volume transform

4.20 two-input sampled function

function $z = f(x, y)$ represented as a two-dimensional array of output values

4.21 vector

ordered list of numbers

4.22 Wide Color Gamut (WCG)

chromaticity gamut larger than the chromaticity gamut defined by Recommendation ITU-R BT.709

4.23 window

axis-parallel rectangular region in pixel space specified by the pixel coordinates of two opposing corners, (x_{min}, y_{min}) and (x_{max}, y_{max}), and including all pixels with coordinates (x, y), where $x_{min} \leq x \leq x_{max}$, and $y_{min} \leq y \leq y_{max}$

4.24 xy chromaticity coordinates

CIE 1931 chromaticity coordinates x and y, as specified in ISO 11664-3

5 **Overview** (Informative)

This standard together with application-specific documents in the SMPTE ST 2094 suite defines metadata for color volume transforms.

This document defines metadata groups, a core set of metadata items, and data structures.

In application-specific documents, an application defines one or more application-specific metadata sets. A metadata set incorporates a time interval, a window, metadata describing the targeted system display and the color volume transform.

- Metadata for a time interval define a portion of the image essence to be processed by a specified color volume transform.
- Metadata for a window define the part of an image to be processed by a specified color volume transform.
- Metadata for a targeted system display define key metadata items of the targeted system display for which the color volume transform metadata is intended. A receiver of the metadata can use the targeted system display metadata to select the metadata sets that are most applicable to the actual output device. A receiver of the metadata can use the targeted system display metadata to adjust the output of the color volume transform from the specified output device to the actual output device.
- Metadata for a color volume transform define parameters of the color volume transform to apply to the image.

There can be one or more metadata sets per image essence. The number of metadata sets associated with an image essence can be large. For an extreme example, an image essence spanning 24 hours with 5,000 time intervals per hour has 120,000 time intervals. With each time interval containing 3 metadata sets for each of 10 targeted system displays and 15 applications, the total number of metadata set is 54 million. In practice, fewer metadata sets are used.

6 Metadata Set

Names of metadata items and groups are in **bold** (such as **WindowNumber**).

All metadata sets are defined in applications. Each metadata set contains exactly one of each of the following:

- an **ApplicationIdentifier**
- an **ApplicationVersion**
- a **TimeInterval**
- an application-specific **ProcessingWindow**
- an application-specific **TargetedSystemDisplay**
- an application-specific **ColorVolumeTransform**

The above describes the logical structure of a metadata set. The physical representation of one or several metadata sets can be defined in other documents using other structures. For example, a **TimeInterval** can be implicitly defined by coupling metadata sets to individual images; Metadata sets can be grouped by **TargetedSystemDisplay**.

7 Application Identifier and Version

ApplicationIdentifier shall be an integer in the range [1, 31].

ApplicationVersion shall be an integer in the range [0, 7].

8 Time Interval

8.1 Introduction

The **TimeInterval** group contains the metadata items defined in Sections 8.2 and 8.3, which describe the start point and duration for which the metadata set is applicable.

8.2 Time Interval Start

The **TimeIntervalStart** shall identify the start point of the portion of image essence for which the metadata set is specified, as an index. The **TimeIntervalStart** shall be a non-negative integer in the range [0, 10000000].

8.3 Time Interval Duration

The **TimeIntervalDuration** shall be the duration of the portion of image essence for which the metadata set is specified, as a count of index positions. The **TimeintervalDuration** shall be a positive integer in the range [1, 10000000].

9 Processing Window

9.1 Introduction

The **ProcessingWindow** group contains the metadata items defined in Sections 9.2, 9.3 and 9.4, which describe two corners of the processing window, in pixel coordinates, and its window number.

9.2 Upper Left Corner

The **UpperLeftCorner** shall be two integers (x_{min}, y_{min}) specifying the pixel coordinate of the top-left pixel of the Processing Window, constrained as specified in Section 9.3 Lower Right Corner. The default value shall be (0, 0).

9.3 Lower Right Corner

The **LowerRightCorner** shall be two integers (x_{max}, y_{max}) specifying the pixel coordinate of the bottom-right pixel of the Processing Window. The default value shall be the pixel coordinate of the bottom-right pixel of the image. The **UpperLeftCorner** and **LowerRightCorner** values shall be constrained such that $0 \leq x_{min} \leq x_{max} <$ the width of the image in pixels and $0 \leq y_{min} \leq y_{max} <$ the height of the image in pixels. The values of x_{max} and y_{max} shall not exceed 65535.

Example: (1919,1079)

9.4 Window Number

The **WindowNumber** shall be the identifier of the Processing Window. The value shall be an unsigned integer in the range [0-15].

The **WindowNumber** shall be unique within any group of metadata sets for the same image, targeted system display metadata values and application identifier.

The default **WindowNumber** shall be 0.

10 Targeted System Display

10.1 Introduction

The **TargetedSystemDisplay** group can include the metadata items defined in Sections 10.2, 10.3, 10.4 and 10.5, which describe the characteristics of the targeted system display. Each application specifies which metadata items to include from this metadata group, and can define additional metadata items for this group. The group shall include no more than one instance of each item.

10.2 Targeted System Display Primaries

The **TargetedSystemDisplayPrimaries** shall be the three primaries of the display. Each primary shall be represented as an xy chromaticity coordinate. Each xy chromaticity coordinate shall be a pair of numbers in the range [0, 1] and in multiples of 0.0001.

Note: The three primaries can be specified in the red, green, and blue order by first specifying the instance with the largest x chromaticity coordinate, followed by the instance with the largest y chromaticity coordinate, and then followed by the instance with neither the largest y nor the largest x chromaticity coordinate.

10.3 Targeted System Display White Point Chromaticity

The **TargetedSystemDisplayWhitePointChromaticity** shall be the chromaticity of the display's white point, represented as an xy chromaticity coordinate. The xy chromaticity coordinate shall be a pair of numbers in the range [0, 1] and in multiples of 0.0001.

10.4 Targeted System Display Maximum Luminance

The **TargetedSystemDisplayMaximumLuminance** shall be the maximum luminance of the targeted system's display, specified in units of candela per square meter (cd/m^2).

The **TargetedSystemDisplayMaximumLuminance** value shall be in the range [1, 10000] and in multiples of 0.01.

10.5 Targeted System Display Minimum Luminance

The **TargetedSystemDisplayMinimumLuminance** shall be the minimum luminance of the targeted system's display, specified in units of candela per square meter (cd/m^2).

The **TargetedSystemDisplayMinimumLuminance** shall be in the range [0, 100] and in multiples of 0.0001.

The **TargetedSystemDisplayMinimumLuminance** value shall be less than the **TargetedSystemMaximumLuminance**.

11 Color Volume Transform

The **ColorVolumeTransform** group contains metadata items that are defined in the applications.

12 Matrix Constraints

12.1 Number of Rows

The number of rows shall be three.

12.2 Number of Columns

The number of columns shall be three.

12.3 Indexing

Each number in the matrix shall be referred to by row and column integer indices 0, 1 and 2. Denoting a number in a matrix as M_{ij} with i as the row index and j as the column index, the first number in the matrix is M_{00} and the last number in the matrix is M_{22}.

12.4 Numerical Representation

Each number in the matrix shall be in the range [0, 1] and in multiples of 0.00001 unless otherwise specified.

13 Sampled Function Constraints

13.1 Introduction

An application using sampled functions can define the range of each sampled function and the numerical representation of the x_i and y_i values.

13.2 Uniqueness of x_i Values

Each x_i value in the list of $\{ x_i , y_i \}$ pairs shall be unique.

13.3 $\{ x_i, y_i \}$ Pair Ordering

The $\{ x_i, y_i \}$ pairs shall be ordered by ascending x_i in the list.

13.4 Input Domain

The sampled function shall have the input domain [0, 1] with the first and last x_i values in the list of $\{ x_i, y_i \}$ pairs being 0.0 and 1.0, respectively.

13.5 Range

The range of the sampled function shall be [0, 1] unless otherwise specified.

13.6 Numerical Representation

The x_i and y_i values shall be in multiples of 0.00001 unless otherwise specified.

13.7 Interpolation Between Samples

The $y = f(x)$ values that are not in the list of $\{ x_i, y_i \}$ pairs shall be obtained through interpolation. The default interpolation shall be linear interpolation.

The default processing of $y = f(x)$ is thus:

$$y = y[i] + (y[i+1] - y[i]) \times \frac{x - x[i]}{(x[i+1] - x[i])}$$

where:

y = output value

x = input value

$y[i]$ = element y_i in the list

$x[i]$ = element x_i in the list

i = index into the list, such that $x_i \leq x < x_{i+1}$

13.8 Default Sampled Function

The default sampled function shall be the identity function $y = x$.

14 Two-input Sampled Function Constraints

14.1 Introduction

An application using two-input sampled functions can define the range of each two-input sampled function and the numerical representation of the z values.

14.2 Data Structure

The two-input sampled function shall be represented as a two-dimensional array of output values (z) having M rows and N values in each row. The array is regularly spaced, independently in each dimension. M and N shall be at least 2. Each row shall correspond to a first input value (x), and each value within a row shall correspond to a second input value (y).

14.3 Value Ordering

The output values shall be ordered in rows by ascending x, and values within each row shall be ordered by ascending y.

Example: { {0.0, 0.3333}, {0.3333, 0.6667}, {0.6667,1.0} } representing $z = (2x + y)/3$ for input values x = {0.0, 0.5, 1.0} and y = {0.0, 1.0}

14.4 Input Domain

The function shall have the input domain [0, 1] for both x and y.

14.5 Mapping between sample points and input values

The first and last row in the data structure shall correspond to the first and last values in the input domain of x. Intermediate rows shall map to intermediate and equidistant x values.

The first and last position in each row shall correspond to the first and last values in the input domain of y. Intermediate positions in a row shall map to intermediate and equidistant y values.

14.6 Range

The range of the function shall be [0, 1] unless otherwise specified.

14.7 Numerical Representation

The z values shall be in multiples of 0.0001 unless otherwise specified.

14.8 Interpolation Between Samples

The $z = f(x, y)$ values that do not have corresponding output values shall be obtained through interpolation. The default interpolation shall be bilinear interpolation.

The default processing of $z = f(x, y)$ is thus:

$$z = a \times \{z_{11}(x_2 - x)(y_2 - y) + z_{21}(x - x_1)(y_2 - y) + z_{12}(x_2 - x)(y - y_1) + z_{22}(x - x_1)(y - y_1)\}$$

where:

$$a = \frac{1}{(x_2 - x_1)(y_2 - y_1)}$$

$z_{11} = z[i][j]$

$z_{12} = z[i][j + 1]$

$z_{21} = z[i + 1][j]$

$z_{22} = z[i + 1][j + 1]$

$x_1 = x_i$

$x_2 = x_{i+1}$

$y_1 = y_i$

$y_2 = y_{i+1}$

$z\,[i][j]$ = element in position j in row i

i, j the indices into the list of z values so that $x_i \leq x < x_{i+1}$ and $y_j \leq y < y_{j+1}$

Annex A Dynamic Metadata Usage in a Generalized Color Volume Transform Model (Informative)

The diagram in Figure A.1 presents a generalized model for the color volume transform of image essence for presentation on a display having a smaller color volume than the original mastering display. The generalized Color Volume Transform Model consists of application-specific processing block(s), which could include such processing functions as Gamut Shaping, Tone Mapping, application of a Color Remapping Matrix, etc. This diagram depicts the color volume transform operations at the pixel level. Specific constraints in the selection of pixels by means of windows as well as by other criteria are covered in the applications.

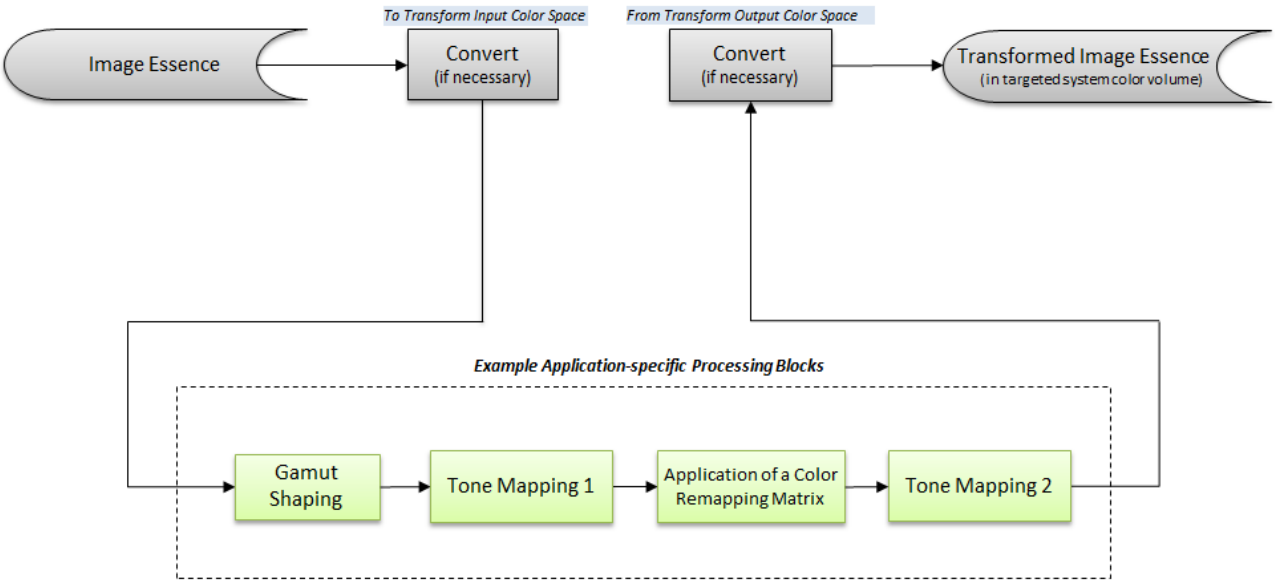

Figure A.1 – Processing blocks of the generalized Color Volume Transform Model

The received signal can differ from the signal used in mastering. For example, the color space might have been re-encoded from Recommendation ITU-R BT.709 to Recommendation ITU-R BT.2020 without altering the content. The input signal can be transformed as needed to an application-specific color space encoding. Likewise, there can be a mismatch between the resulting output format and the signal format supported by the current display, for example, the output of a transform using SMPTE ST 2084 when the device uses Recommendation ITU-R BT.1886. The gray boxes in Figure A.1 indicate such implied transforms.

Annex B Bibliography (Informative)

Recommendation ITU-R BT.709-6 (06/2015), Parameter Values for the HDTV Standards for Production and International Programme Exchange

Recommendation ITU-R BT.1886-0 (03/2011), Reference Electro-Optical Transfer Function for Flat Panel Displays Used in HDTV Studio Production

Recommendation ITU-R BT.2020-2 (10/2015), Parameter Values for Ultra-High Definition Television Systems for Production and International Programme Exchange

Recommendation ITU-R BT.2035-0 (07/2013), A Reference Viewing Environment for Evaluation of HDTV Program Material or Completed Programmes

SMPTE RP 431-2:2011, D-Cinema Quality — Reference Projector and Environment

SMPTE ST 2084:2014, High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays