



# **Ultra HD Forum Guidelines**

## **Blue Book: Ultra HD Production and Post Production**

**Ultra HD Forum**

**8377 Fremont Blvd., Suite 117**

**Fremont, CA 94538**

**UNITED STATES**

**VERSION 3.0.0**

**Apr 7, 2023**



# 1. Foreword

This new version v3 of the Ultra HD Forum Guidelines provides a holistic view of modern media systems, their mechanisms and workflows, and how those are impacted by the latest generation of improvements – the “Ultra HD” technologies, those that take media beyond the limits established at the start of this millennia, characterized in large part by the video resolutions and the dynamic range offered for media in “high definition”, i.e., ITU-R Rec. BT.709. The Forum considers Ultra HD to not only be any UHD media (i.e., 4K resolution, or higher), but also HD-resolution media with enhancements such as High Dynamic Range, Wide Color Gamut, etc. Ultra HD is a constellation of technologies that can provide significant improvements in media quality and audience experience.

This work represents over eight years of collaborative effort. These new books would not have been possible without the leadership of Jim DeFilippis, who represents Fraunhofer and chairs our Guidelines Work Group with invaluable support from the co-chair, Pete Sellar of Xperi as well as technical assistance from Ian Nock of Fairmile West Consulting.

Our gratitude to all the companies listed in the Acknowledgments that have participated in this effort over the years and specifically to Nabajeet Barman (Brightcove), Andrew Cotton (BBC), Jean Louis Diascorn (Harmonic), Richard Doherty (Dolby), Chris Johns (Sky UK), Katy Noland (BBC), Bill Redmann (InterDigital), Chris Seeger (Comcast/NBCUniversal), and Alessandro Travaglini (Fraunhofer).

This document, *Ultra HD Production and Post Production* (Blue Book), is one of a series of books, referred to as the Rainbow Books, that compose the Ultra HD Forum Guidelines. If any of these terms sound unfamiliar, follow the link below to the Black Book. If a particular standard is of interest, links such as the one above are available to take you to the White Book, where references are collected.

The Rainbow Books are, in their entirety:

White Book [Guidelines Index and References](#)

Red Book [Introduction to Ultra HD](#)



---

Orange Book	<a href="#">Foundational Technologies for Ultra HD</a>
Yellow Book	<a href="#">Beyond Foundational Technologies</a>
Green Book	<a href="#">Ultra HD Distribution</a>
<b>Blue Book</b>	<b><a href="#">Ultra HD Production and Post Production</a></b>
Indigo Book	<a href="#">Ultra HD Technology Implementations</a>
Violet Book	<a href="#">Real World Ultra HD</a>
Black Book	<a href="#">Terms and Acronyms</a>

Updates in this new version of the Ultra HD Forum Guidelines are described on the following page.

I hope you will enjoy reading today.

If you want to know more about Ultra HD, and join our discussions on how it can be deployed, I invite you to join the Ultra HD Forum. You can start by visiting our website:  
[www.ultrahdforum.org](http://www.ultrahdforum.org).

Nandhu Nandhakumar, President, Ultra HD Forum  
April 2023



---

## 1.1 Changes from version 2.6 to 3.0

What's new in the Spring 2023 version of the UHDF Guidelines Blue Book, *Ultra HD Production and Post Production* (v3.0), edited by Katy Noland and Chris Seeger.

The *Ultra HD Production and Post Production* is the fifth of the series of Rainbow Books on the Guidelines for Ultra HD. The scope and purpose of this book is to describe the use of Ultra HD technology for production and post production of UHD content. Included are discussion of conversions between HDR formats (HLG and PQ) as well as up-mapping and down-mapping between SDR and HDR as well as mapping between color spaces for HD-SDR color gamut (ITU-R Rec BT.709) and UHD-HDR wide color gamut (ITU-R Rec BT.2020).

While most of the information in this edition is material from the previous version of the Guidelines (v2.6), the information has been updated. In Section 8.4, objective color measurements have been revised to use the  $\Delta E$ -ITP methodology to evaluate the quality of color and luminance conversions. Each section includes references to more detailed information contained in the companion Rainbow Books.

We hope this new format will be helpful in understanding UHD technologies as well as planning for new or expanded Ultra HD services.

Jim DeFilippis and Pete Sellar,

Guidelines Working Group Co-Chairs, Ultra HD Forum, April 2023





## 2. Acknowledgements

We would like to provide the acknowledgement to all the member companies, past and present, of the Ultra HD Forum who have contributed in some small or large part to the body of knowledge that has been contributed to the Guidelines Color Books, including the specific subject of this book.

ARRIS	ATEME	ATT DIRECTV
British Broadcasting Corporation	BBright	Beamr
Brightcove Inc.	Broadcom	B<>COM
Comcast / NBC Universal LLC	Comunicare Digitale	Content Armor
CTOIC	Dolby	DTG
Endeavor Streaming	Eurofins Digital Testing	Fairmile West
Fraunhofer IIS	Harmonic	Huawei Technologies
InterDigital	LG Electronics	Mediakind
MovieLabs	NAB	Nagra, Kudelski Group
NGCodec	Sky UK	Sony Corporation
Xperi	Technicolor SA	Verimatrix Inc.
V-Silicon		



### 3. Notice

The Ultra HD Forum Guidelines are intended to serve the public interest by providing recommendations and procedures that promote uniformity of product, interchangeability and ultimately the long-term reliability of audio/video service transmission. This document shall not in any way preclude any member or nonmember of the Ultra HD Forum from manufacturing or selling products not conforming to such documents, nor shall the existence of such guidelines preclude their voluntary use by those other than Ultra HD Forum members, whether used domestically or internationally.

The Ultra HD Forum assumes no obligations or liability whatsoever to any party who may adopt the guidelines. Such an adopting party assumes all risks associated with adoption of these guidelines and accepts full responsibility for any damage and/or claims arising from the adoption of such guidelines.

Attention is called to the possibility that implementation of the recommendations and procedures described in these guidelines may require the use of subject matter covered by patent rights. By publication of these guidelines, no position is taken with respect to the existence or validity of any patent rights in connection therewith. Ultra HD Forum shall not be responsible for identifying patents for which a license may be required or for conducting inquiries into the legal validity or scope of those patents that are brought to its attention.

Patent holders who believe that they hold patents which are essential to the implementation of the recommendations and procedures described in these guidelines have been requested to provide information about those patents and any related licensing terms and conditions.

All Rights Reserved

© Ultra HD Forum 2022



## 4. Contents

<b>1. Foreword</b>	<b>2</b>
1.1 Changes from version 2.6 to 3.0	4
<b>2. Acknowledgements</b>	<b>5</b>
<b>3. Notice</b>	<b>6</b>
<b>4. Contents</b>	<b>7</b>
<b>5. List of Figures</b>	<b>9</b>
<b>6. List of Tables</b>	<b>11</b>
<b>7. Ultra HD Production and Post Production</b>	<b>12</b>
7.1. HDR and WCG Technologies	13
7.1.1. Foundation Ultra HD HDR Technologies	13
7.1.2. Perceptual Quantization (PQ) and PQ10 HDR	14
7.1.3. Hybrid Log-Gamma (HLG) and HLG 10	15
7.1.4. ITU Recommendation on HDR, ITU-R BT.2100	17
7.2. Static HDR Metadata-SMPTE ST 2086, MaxFALL, MaxCLL	18
7.2.1. HDR10	19
7.2.2. HDR10 Metadata Generation	20
7.2.3. HDR10 Metadata Carriage	20
7.3. Signaling Transfer Function, System Colorimetry and Matrix Coefficients	21
7.4. ICTCP Color Representation	28
7.5. Peak Brightness Considerations	29
7.5.1. Studio Video over IP	30
<b>8. Conversions Between Transfer Functions</b>	<b>32</b>
8.1. Conversion between HDR formats	32
8.1.1. PQ10 to HLG10	32
8.1.2. HLG10 to PQ10	33
8.2. Conversion of SDR to HDR: Up-Mapping	34
8.3. Conversion HDR to SDR: Down-Mapping	39
8.4. Objective Measurement of the Quality of Tone Mapping	45
<b>9. Real-Time Program Service Assembly</b>	<b>52</b>
9.1. Maintaining Dynamic Range and System Colorimetry Parameters	52
9.2. Conversions from SDR/BT.709 to PQ10 and HLG10	53
9.3. Conversion from PQ10/HLG10 to SDR BT.709	56



---

9.4. Avoiding Image Retention on Professional and Consumer Displays	56
9.4.1. Background	56
9.4.2. Definition of Static Images	57
9.4.3. Recommendations	57
<b>10. Production for Pre-Recorded Content</b>	<b>58</b>
10.1. Camera Requirements	59
10.1.1. High Dynamic Range and Wide Color Gamut Considerations	59
10.1.2. Imaging Devices: Resolution, Dynamic Range and Spectral Characteristics	60
10.2. Reference Monitor Considerations	61
10.3. On-Set/ Near-Set Monitoring	62
10.4. Color Grading	62
10.4.1. Grading Workflow	62
10.4.2. Grading Room Configuration	62
10.4.3. Grading Monitor Configuration	63
10.4.4. Grading System Configuration	63
10.5. Channel-based Immersive Audio Post Production	63
10.5.1. Immersive Audio Monitoring Environment	65
10.5.2. Immersive Audio Mastering	65
<b>11. Production for Live Content</b>	<b>66</b>
11.1. Live Production in Trucks or Studio Galleries	66
11.2. HDR Production with encodings other than PQ and HLG	67
11.3. Channel-based Immersive Audio Production	68
<b>12. Additional Ultra HD Technologies beyond Foundation Ultra HD</b>	<b>72</b>
12.1. Dynamic HDR Metadata	72
12.2. Ultra HD-Next Generation Audio (NGA)	74
12.2.1. NGA Post Production	74
12.2.2. NGA Live Production	74
12.2.3. NGA Production tools and workflows	75
12.2.3.1. Introduction	75
12.2.3.2. Dolby Technology Workflows	76
12.2.3.3. MPEG-H Audio Workflows	87



## 5. List of Figures

### List Of Images

- [Figure 1. Content Production and Distribution Workflow](#)
- [Figure 2. X-Rite Color Checker Chip Chart](#)
- [Figure 3. Scene Light Conversion Accuracy, SDR to HLG](#)
- [Figure 4. Display Light Conversion Accuracy, SDR to HLG](#)
- [Figure 5. Down-Mapping HLG to SDR Color Accuracy](#)
- [Figure 6. Sarnoff “Yellow Brick Road”© Test Chart \(SDR\)  
\(courtesy of SRI International\)](#)
- [Figure 7. LUT 1 Conversion from HLG to SDR](#)
- [Figure 8. LUT 2 Conversion from HLG to SDR](#)
- [Figure 9. LUT 3 Conversion from HLG to SDR](#)
- [Figure 10. ICtCp and the Relationship to the HVS](#)
- [Figure 11.  \$\Delta\$ EITP Color Difference \(Source vs Output\)](#)
- [Figure 12.  \$\Delta\$ TPITP Color Difference - Chroma ONLY \(Source vs Output\)](#)
- [Figure 13.  \$\Delta\$ IITP Color Difference - Intensity ONLY \(Source vs Output\)](#)
- [Figure 14. Intensity Tracking \(Source vs Output\)](#)
- [Figure 15.  \$\Delta\$ HITP Hue Difference \(Constant Intensity and Chroma\)](#)
- [Figure 16. Hue Difference \( \$\Delta\$ HITP\) and Hue Angle \( \$\Delta\$ hITP\)](#)
- [Figure 17. ITP Plots of pseudo BT.202 color space transforms using  
the Sarnoff “Yellow Brick Road” Chart](#)
- [Figure 18. CIE 1979 u' v' color spaces](#)
- [Figure 19. Sample Live Workflow with Mixed Format Source Content](#)
- [Figure 20. Pre-recorded Content Production Workflow and Interfaces](#)
- [Figure 21. Channel-based Immersive Audio Post-Production](#)
- [Figure 22. Channel-based Immersive Audio Live Production](#)
- [Figure 23. TVP 4k, Live Euro transmission with Atmos](#)
- [Figure 24. Dynamic Metadata Live Event example workflow](#)
- [Figure 25. Encoding pipeline using an ATEME Titan Live Encoder](#)
- [Figure 26. Setup Screen Captures](#)
- [Figure 27. S-ADM setup in the audio control room](#)
- [Figure 28. Signal Path overview](#)
- [Figure 29. Typical MPEG-H Workflow for Live Productions](#)
- [Figure 30. MPEG-H Audio production workflow based on SMPTE ST 2110  
\(simplified\)](#)
- [Figure 31. Example of MXF based workflow using ADM metadata](#)



- [Figure 32. MPEG-H Audio Post-Production workflow \(simplified\)](#)
- [Figure 33. Export window of the MPEG-H Authoring Plug-in](#)
- [Figure 34. Full Interoperability between MPEG-H Masters and ADM-based workflows](#)
- [Figure 35. Signal flow of an MPEG-H AMAU System](#)
- [Figure 37. Schematic audio production and authoring workflow of the ESC MPEG-H field test](#)
- [Figure 37. Schematic audio production and authoring workflow of the ESC MPEG-H field test](#)
- [Figure 39. Audio Monitoring Room for Immersive Sound Production during ESC 2019](#)
- [Figure 39. Audio Monitoring Room for Immersive Sound Production during ESC 2019](#)
- [Figure 40. Additional Hamasaki Square and Ambience microphones on the arena's ceiling](#)
- [Figure 41. MPEG-H Audio Immersive and Interactive demonstration during the ESC 2019](#)
- [Figure 42. Setup of the UHD FTV Channel using MPEG-H Audio during French Tennis Open](#)
- [Figure 43. MPEG-H Audio during French Tennis Open 2019: Demonstrations at the Philippe Chatrier court](#)
- [Figure 44. MPEG-H Audio during French Tennis Open 2019: Control Room](#)
- [Figure 45. European Athletics Championships 2018 in Berlin: Control Room](#)
- [Figure 46. MPEG-H Audio Transmission setup during European Athletics Championships 2018](#)
- [Figure 47. European Athletics Championships 2018 in Berlin: Control Room](#)



## 6. List of Tables

- [Table 1. Example Display Luminance Values for HLG on Different Displays](#)
- [Table 2. File based Signaling for SDR/BT.709](#)
- [Table 3. File-Based Signaling for SDR/BT.2020](#)
- [Table 4. File-Based Signaling for HDR/BT.2020](#)
- [Table 5. BT.2408 Signal ranges and achievable color volume for PQ and HLG on a 1000 cd/m2 nominal peak luminance display](#)
- [Table 6.  \$\Delta E\$ -ITP Values for the Up-mapping from SDR to HLG](#)
- [Table 7. Pre-recorded Content Interface Descriptions](#)
- [Table 8. Channel Based immersive Audio Workflow Elements](#)
- [Table 9. Description of Live Channel Based Audio Workflow Elements](#)
- [Table 10. Representations Available](#)



## 7. Ultra HD Production and Post Production

The Ultra HD Forum is concerned with establishing viable workflows both for Real-time Program Services and On Demand content that was originally offered live. Real-time Program Services (aka Linear TV) make frequent use of Pre-recorded material, such as edited inserts, interstitials, etc., which involve production and post-production.

Live content has specific requirements and operating practices that are unlike Digital Cinema, Blu-ray™ disc mastering, or other Pre-recorded content practices. Ultra HD workflows and technologies that are designed for these other delivery methods may not apply to Live content production.

Production practices for Foundation Ultra HD audio are similar to those used in current HD content creation. Audio follows multi-channel workflows established for multi-channel 5.1 surround delivery using (as appropriate) [AC-3 \[29\]](#), [DTS-HD \[111\]](#), [E-AC-3+JOC \(an instance of Dolby Atmos®<sup>1</sup>\) \[35\]](#), [HE-AAC \[27\]](#), or [AAC-LC emission \[27\]](#). Although 2.0 stereo sound is possible, Foundation Ultra HD content is considered premium content, and it is therefore recommended that productions provide at least 5.1 channels.

Foundation Ultra HD, production practices for closed captions and subtitles are also similar to those of HD content creation. Closed captions and subtitles follow workflows established for [CTA- 608\[18\]/CTA-708\[19\]](#), [ETSI 300 743\[20\]](#), [ETSI 300 472\[21\]](#), [SCTE-27\[22\]](#), or [IMSC1 \[23\]](#) formats.

The remainder of this section will focus on mechanisms for producing the video components of the content.

As content is produced, it is useful to know in advance for which service mode(s) the content is intended. Equally, service providers planning to deliver Foundation Ultra HD content need to have an understanding of the formats in which the content will be supplied.

The following is a diagram providing an overview of the content production and distribution workflow for Real-time Program Services and potentially capturing Live content for later distribution via an On-Demand Service.

---

<sup>1</sup> Dolby, Dolby Atmos, Dolby Digital and Dolby Vision are trademarks of Dolby Laboratories.



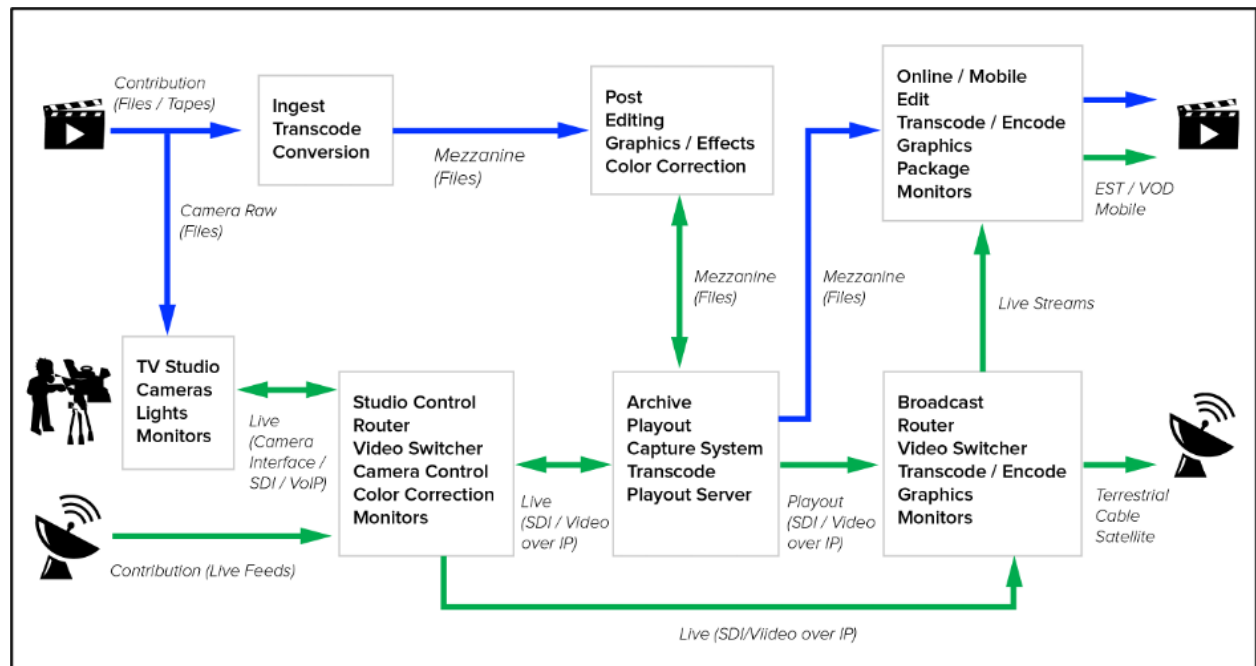


Figure 1. Content Production and Distribution Workflow

## 7.1. HDR and WCG Technologies

There are many terms in use in the field of HDR television. This section explains the existing terminology and how terms are used in these Guidelines.

Note that currently some UHD displays are capable of accepting [BT.2020 \[3\]](#) content, but as of this publication, no direct view display is available that is capable of rendering the full gamut of colors in the BT.2020 system colorimetry. It is assumed that in these cases, the device employs “best effort” gamut mapping tailored to its particular display characteristics, and thus these devices are considered BT.2020 compatible.

### 7.1.1. Foundation Ultra HD HDR Technologies

The following HDR systems are recommended for Foundation Ultra HD due to their conformance with the criteria listed in Section 4 above:

- HLG10: HDR systems or content employing Hybrid Log-Gamma (HLG), the wide color gamut specified in [BT.2100 \[5\]](#) and 10-bit depth



- PQ10: HDR systems or content employing Perceptual Quantization (PQ), the wide color gamut specified in BT.2100, and 10-bit depth
- HDR10: HDR systems or content employing PQ10 and further including or capable of providing SMPTE ST [2086 \[10\]](#), MaxFALL, and MaxCLL static metadata

It should be noted that Real-time Program Services are typically comprised of both live and pre-recorded content, and it is not recommended that service providers alternate between SDR and HDR signal formats or mix different HDR formats. See Section 8.1 for details.

### 7.1.2. Perceptual Quantization (PQ) and PQ10 HDR

One HDR transfer function set defined for use in television is the “Perceptual Quantization” (PQ) developed by Dolby. PQ is defined as a reference display referred transfer function, or EOTF. The transfer function is designed to minimize visibility of banding on a display between the brightness ranges of 0 to 10,000 cd/m<sup>2</sup>. SMPTE standardized the PQ EOTF in [ST 2084 \[9\]](#) and the ITU standardized the EOTF, OETF, and OOTF transfer functions in [ITU-R BT.2100 \[5\]](#). ST 2084 specifies the shape of the curve over a range of 0 to 1 but does not specify SDI code values for the 0 and 1 values. Table 9 in ITU-R BT.2100 describes the code mapping for both narrow range and full range code values<sup>2</sup>.

The Ultra HD Forum has defined the term PQ10 to refer to an HDR system employing the PQ transfer functions and wide gamut colorimetry specified in BT.2100 and 10-bit depth. PQ10 content or systems may or may not be metadata-capable. For example, HDR10 is a Foundation Ultra HD PQ10-based format that has the capability of including certain static metadata (see Section on HDR 10).

The PQ signal is “display-referred”, meaning that the pixel-encoded values represent specific values of luminance for displayed pixels. The intent is that only the luminance values near the minimum or maximum luminance capability of a display are necessarily adjusted to utilize the available dynamic range of the display. Some implementations may apply a “knee” at the compensation points in order to provide a smoother transition from the coded values to the display capabilities; e.g., to avoid “clipping”.

---

<sup>2</sup> [CTA 861-I \[31\]](#) offers additional information about code ranges.



When default display settings are engaged, PQ enables pixel values in the mid-range, including skin tones, to be rendered on a display at the same (absolute) luminance level that was determined at production.

For example, if a scene was graded on a 1000 cd/m<sup>2</sup> grading monitor and then displayed on a 4000 cd/m<sup>2</sup> display, the skin tones can be rendered at the same luminance values on the 4000-nit display as on the 1000-nit monitor per the grader's intent, while the speculars and darker tones can be smoothly extended to take full advantage of the 4000-nit display. See also [Section 7.5](#).

The Ultra HD Forum InterOp testing has shown that consumer displays vary significantly, especially when different viewing “modes” are selected (e.g., “Sports”, “Vivid”, or “Cinema/Movie” picture modes). Research has also shown that ambient light in the viewing room can have an impact on humans' perception of a displayed image. Displays used in non-reference viewing environments may employ adjustments to the PQ curve in order to provide some compensation for the difference between the actual viewing environment and the reference viewing environment.

PQ content requires a down-mapping step in order to provide acceptable SDR quality. See [Section 8.3](#) for a deeper discussion of backward compatibility, including the pros, cons and open questions that apply to various possible methods.

### 7.1.3. Hybrid Log-Gamma (HLG) and HLG 10

Another HDR transfer function defined for use in television is “Hybrid Log-Gamma” (HLG) developed by the BBC and NHK. This is defined as a camera capture transfer curve, or OETF. This curve was designed to provide HDR while maintaining a degree of backward compatibility with SDR/BT.2020 displays. The HLG curve was first specified [in ARIB STD-B67 \[105\]](#)<sup>3</sup>, and later in [ITU-R BT.2100 \[5\]](#). The two OETF specifications are identical. BT.2100 additionally specifies the display EOTF and the OOTF, which are not defined in the ARIB standard.

---

<sup>3</sup> The original v1.0 ARIB specification referred to the breakpoint between the logarithmic and gamma portions of the HLG curve as the “reference white level”, which some manufacturers incorrectly assumed to be the signal level for “HDR Reference White”. Later versions of the ARIB specification have removed the term “reference white level” to avoid any such confusion. The “HDR Reference White” level for both HLG and PQ production is defined in ITU-R BT. 2408 [8].



The Ultra HD Forum has defined the term HLG10 to refer to an HDR system employing HLG, wide gamut colorimetry specified in BT.2100 and 10-bit depth with black at code 64, and nominal peak at code 940<sup>4</sup>.

HLG is a “scene-referenced” HDR technology that uses pixel-encoded values that are intended to represent picture luminance levels relative to each other in a given scene. The intent is that all the displayed pixel luminance values may be adjusted in a defined manner to compensate for specific display capabilities (e.g., peak luminance) and viewing environments in such a way that the values retain their perceptual appearance relative to one another.

As the eye’s sensitivity to light intensity is approximately logarithmic, a power (or “gamma”) law is applied to the HLG relative scene-light pixel values, to scale them to span the luminance range of the display; thereby approximating the same relative subjective brightness values that were determined at production. Since no one luminance range is “fixed”, the (absolute) displayed luminance of mid-range values, including skin tones and reference white, will increase or decrease to scale along with all other values in order to better preserve the appearance of the relative brightness values of the pixels in the scene.

For example, if a scene was graded on a 1000 cd/m<sup>2</sup> grading monitor and then displayed on a 2000 cd/m<sup>2</sup> display, the skin tones and other scene elements will be brighter on the 2000 cd/m<sup>2</sup> display than on the 1000 cd/m<sup>2</sup> monitor. However, given that all pixel values are adjusted through the “gamma law”, the overall image remains perceptually similar to the original “look” selected by the grader. See also Section 6.1.10<sup>5</sup>.

As mentioned above, Ultra HD Forum InterOp testing has shown that consumer displays vary significantly, especially when different viewing “modes” are selected (e.g., “Sports”, “Vivid” or “Cinema/Movie” modes). Research has also shown that ambient light in the viewing room can have an impact on humans’ perception of a displayed image. Section 6.2 of ITU-R report [BT.2390 \[6\]](#) describes how the display gamma may be adjusted to provide some compensation.

---

<sup>4</sup> CTA 861-I [31] offers additional information about code ranges.

<sup>5</sup> It is well known that changes in brightness of a display may affect the perception of image color ([Hunt Effect](#)) [100] as well as image contrast ([Stevens](#)) [101]. Research on the Color Appearance Model (CAM), where brightness exceeds 700 cd/m<sup>2</sup>, agrees with Stevens observations; however as of 2020 the BBC has found no evidence of significant changes to the perception of color or contrast in their research using HLG with displays of different peak brightness ([Katz](#)) [102]. additional research is on-going on these effects with respect to HDR imagery.



HLG is, by design, a relative system, meaning that the displayed luminance of reference white, skin tones, gray levels, and all other signal levels will vary depending on the peak luminance of the display. Displayed luminance values will also differ from SDR displays. A variable gamma accounts for viewer adaptation. [Table 1](#) below illustrates the relative nature of HLG, for HDR Reference White (75% HLG) and a mid-gray level of 38% HLG.

**Table 1. Example Display Luminance Values for HLG on Different Displays**

Display peak luminance (cd/m <sup>2</sup> )	400	600	1000	2000	4000
Reference gamma (see ITU-R BT.2100 Note 5f)	1.03	1.11	1.20	1.33	1.45
Displayed luminance for HDR Reference White, 75% HLG (cd/m <sup>2</sup> )	101	138	203	343	581
Displayed luminance for mid-gray at 38% HLG (cd/m <sup>2</sup> )	17.4	20.9	26.2	35.8	48.7

Content produced using HLG can be displayed on SDR/WCG(BT.2020) devices with a degree of compatibility that may be judged acceptable for programs and services according to Report ITU-R BT.2390, and subjective tests performed by the EBU, RAI, IRT and [Orange Labs \[88\]](#). Backward-compatible HLG is only intended to support SDR/BT.2020 (WCG) displays and not intended for displays which only support SDR/BT.709. See Section 0 below for a deeper discussion of backward compatibility, including the pros, cons and open questions that apply to various possible methods.

#### 7.1.4. ITU Recommendation on HDR, ITU-R BT.2100

In July 2016 ITU-R Study Group 6 approved the publication of a Recommendation on HDR for use in production and international program exchange and is known as [ITU-R BT.2100 \[5\]](#). This recommendation includes the following specifications:



- Spatial resolutions: 1080p, 2160p, 4320p
- Frame rates: 24/1.001, 24, 25, 30/1.001, 30, 50, 60/1.001, 60, 100, 120/1.001, 120
- System Colorimetry: Same as [BT.2020 \[3\]](#)
- Reference Viewing Environment for critical viewing: 5 cd/m<sup>2</sup> luminance of surrounds and less than or equal to 5 cd/m<sup>2</sup> for luminance of the periphery (the remaining area outside of the surrounds which affect eye adaption and cause reflections on the display)
- Reference non-linear transfer functions EOTF, OOTF, OETF: PQ and HLG
- Signal Format: Y'C<sub>B</sub>'C<sub>R</sub>', IC<sub>T</sub>C<sub>P</sub>', and RGB.
- Color sub-sampling: same alignment as specified in BT.2020
- Bit depth value ranges:
  - 10 bit, Narrow (64-940) and Full (0-1023) ranges
  - 12 bit, Narrow (256-3760) and Full (0-4095)
  - Floating point signal representation: Linear RGB, 16-bit floating point

BT.2100 is the primary reference document on HDR for use in production and international program exchange. Note that a full signal specification will need to include the following attributes: spatial resolution, frame rate, transfer function (PQ or HLG), color signal format, integer (10 or 12 bits, narrow or full range) or floating point.

Not all of the parameter values listed above have been deployed as yet, but are included as informative details.

## 7.2. Static HDR Metadata-SMPTE ST 2086, MaxFALL, MaxCLL

SMPTE has specified a set of static metadata in the SMPTE [ST 2086 \[10\]](#) Mastering Display Color Volume Metadata (MDCV) Supporting High Luminance and Wide Color Gamut Images standard. Parameters included indicate the characteristics of the mastering display monitor. The mastering display metadata indicates that the creative intent was established on a monitor having the described characteristics. If provided, the implication is that the artistic intent of the content is within the subset of the overall container per the metadata values. The mastering display characteristics include the display primaries and white point as x,y chromaticity coordinates, and the maximum and minimum display luminance. For example, the metadata may indicate that the system colorimetry of the mastering display uses the DCI-P3 color gamut in a [BT.2020 \[3\]](#) container, and the luminance range is a subset of the 0 to 10,000 cd/m<sup>2</sup> representable in PQ.



---

The Blu-ray Disc Association and DECE groups have defined carriage of two additional metadata items:

- MaxFALL – Maximum Frame Average Light Level; this is the largest average pixel light value of any video frame in the program
- MaxCLL – Maximum Content Light Level: this is the largest individual pixel light value of any video frame in the program

Static metadata may be used by displays to control color volume transforms of the received program to better reproduce the creative intent as shown on the mastering display, given the capabilities of the display device. However, for MaxFALL and MaxCLL static metadata specifically, there are limitations for use with live broadcasts since it is difficult to determine a program's maximum pixel light values in a program that is on-going and has no determined endpoint. According to the UHD Alliance, today's mastering practices may generate outlier values unintentionally, causing the content's associated MaxCLL value to be higher than expected. As a response to that, some content providers use statistical analysis to calculate a MaxCLL value that is more representative of the statistically significant brightest pixels contained in the image sequence. The Ultra HD Forum further suggests that such statistical methodology may also apply to MaxFALL.

It is worth noting that code levels below minimum mastering display luminance and code levels above maximum mastering display luminance were likely clipped on a professional reference monitor and therefore any detail in pictures utilizing code levels outside this range were likely not seen in the content production process.

### 7.2.1. HDR10

HDR10 is a PQ10-based HDR format. The term "HDR10" is in widespread use and has been formally and consistently defined by several groups, including DECE and BDA, as:

- Transfer function: [Rec. ITU-R BT.2100 \[5\]](#) (PQ)
- System Colorimetry: BT.2100
- Bit depth: 10 bits
- Metadata: SMPTE [ST 2086 \[10\]](#), MaxFALL, MaxCLL

Several delivery formats (e.g. Ultra HD Blu-ray™ and CTA's HDR10 profile) have specified delivery using the above parameters with the metadata mandatory but are still considered to be using HDR10 for the purposes of this document. While ATSC A/341 does not use the term "HDR10," when the PQ transfer function is used, the static metadata video parameters described therein are consistent with HDR10 as defined herein.



Note that some displays ignore some or all static metadata (i.e., ST 2086 (MDCV), MaxFALL, and MaxCLL); however, HDR10 distribution systems must deliver the static metadata, when present.

### 7.2.2. HDR10 Metadata Generation

HDR10 includes the static metadata described in Section 6.1.5. MaxFALL and MaxCLL metadata could be generated by the color grading software or other video analysis software. In Live content production, MaxFALL or MaxCLL metadata is not generated during the production process. By definition, it is not possible to generate MaxFALL or MaxCLL for a Live program because these cannot be known until the entire program is produced (i.e., after the program is over). A special value of '0' (meaning, "unknown") is allowed for MaxFALL and MaxCLL. It may be possible to set limits on the output and thus pre-determine MaxFALL and MaxCLL even for Live content production. For example, if MaxFALL and MaxCLL metadata values are provided, a video processor could be inserted in order to reduce brightness of any video frames that would exceed the indicated values (similar to the way audio processors are often inserted to enforce audio loudness and peak levels).

If it is desired to use HDR10 in Live content production, but the production facility does not support carriage of the metadata, then default values for [ST 2086 \[10\]](#), MaxFALL and MaxCLL should be determined and entered directly into the encoder via a UI or a file. SMPTE 2086 metadata could be set to values that represent the monitors used for grading during production of the content.

### 7.2.3. HDR10 Metadata Carriage

SMPTE ST 2086 metadata is carried with Mastering Display Color Volume (MDCV) metadata.

In HEVC and AVC bitstreams, MDCV metadata is carried via SEI message, ([see DASH-IF Guidelines for Interoperability Points. ISO/IEC 23008-2 \[16\]](#), section D.2.28 (Syntax) and D.3.28 (Semantics). Complementary metadata that often can accompany MDCV metadata are MaxFALL/MaxCLL or "Content Light Level Information" (CLLI) which is carried in a static SEI message (sections D.2.35 (Syntax) and D.3.35 (Semantics), [\[26\]](#)). It is expected that the decoder/display will employ "best effort" to tone map the content optimally using this metadata. Where the metadata doesn't exist or is dropped, the end-to-end chain will operate in a manner indistinguishable from PQ10 content (i.e. PQ without metadata).





Both MDCV metadata exists to optimize both tone and color mapping on target displays that differ from the display that created the content. CLLI metadata exists so that a target display can optimize the tone mapping based on characteristics of the content itself.

Both the MDCV and CLLI (MaxCLL/MaxFall) SEI messages can be represented by their corresponding ANC messages as specified in [SMPTE ST 2108-1 \[48\]](#) and can accompany each video frame to which they apply. An SMPTE ST 2108-1 ANC message is a simple, bit-accurate encapsulation of the corresponding SEI message, to ease handling by an encoder or decoder; as a result, SMPTE ST 2108-1 is also able to support certain dynamic metadata technologies (see Section 12). This same metadata can be embedded in a SMPTE ST 2110 IP data stream ([SMPTE ST 2110-40 \[47\]](#)). Additional dynamic metadata systems may also travel down this pipeline (see Section 6.1.12)

Both MDCV and CLLI can be embedded in QuickTime, MP4 and MXF. In QuickTime and MP4 they are embedded in separate MDCV and CLLI atoms(QuickTime) or boxes(MP4). In MXF Image Essence Picture Descriptors are used to embed both MDCV and CLLI static metadata (UL values are defined in [SMPTE ST 2067-21:2016 \[38\]](#)).

Note that in the metadata pipeline there may be multiple points where the metadata could get lost so the values are often not added until the very end of the production process. In transmission, there may be multiple encoders in the end-to-end broadcast chain. In the event that HDR10 is used and the metadata does not reach the decoder, default values are often used.

## 7.3. Signaling Transfer Function, System Colorimetry and Matrix Coefficients

The system colorimetry, transfer function (SDR or HDR), and matrix coefficients must be known to downstream equipment ingesting or rendering content in order to preserve display intent. This is true for file transfers in file-based workflows and in linear content streams in linear workflows.

In file-based workflows, mezzanine file formats such as IMF/MXF (stored in picture essence descriptors), QuickTime, MP4 (stored in NCLC atoms) are often used.

In linear production workflows, these values are transmitted via VPID (Video Payload Identifier) within HD-SD, [SMPTE ST 292 \[99\]](#). In linear compressed transmission workflows, these values are typically signaled in the VUI of an H.265/HEVC or H.264/MPEG-4 AVC bitstream. Details on SEI and VUI messaging are available in the HEVC specification [26], in particular, Appendix D (SEI) and Appendix E (VUI).



---

The tables below ([Table 2](#), [Table 3 a](#) and [Table 4](#)) summarize HEVC Main10 Profile bitstream SDR, PQ and HLG indicators. (In HEVC and AVC specifications, the bitstream elements are bolded and italicized to distinguish them from temporary variables and labels). As shown in [Table 4](#) there are two methods of signaling the HLG transfer function.



Table 2. File based Signaling for SDR/BT.709

	<u>System Identifier</u>	<u>BT. 709 YCC</u>	<u>BT. 709 RGB</u>	<u>Full-Range 709 RGB</u>	<u>BT. 601 525</u>	<u>BT. 601 625</u>
Color properties	Color primaries	BT.709	BT.709	BT.709	BT.601	BT.601
	Transfer Characteristics	BT.709	BT.709	BT.709	BT.709	BT.709
	Signal Format	Y'CbCr	R'G'B'	R'G'B'	Y'CbCr	Y'CbCr
Other	Full/narrow range	Narrow	Narrow	Full	Narrow	Narrow
	4:2:0 chroma sample location alignment	Interstitial	N/A	N/A	Interstitial	Interstitial
CICP parameters Rec. ITU-T H.273 ISO/IEC 23091-2 [113]  (QuickTime/HE/AVC) <sup>1</sup>	ColorPrimaries	1	1	1	6	5
	TransferCharacteristics	1	1	1	6	6
	MatrixCoefficients	1	0	0	6	5
	VideoFullRangeFlag	0	0	1	0	0
SMPTE MXF parameters SMPTE ST 2067-21 <sup>2</sup>	Color primaries	06.0E.2B.34.04.01.01.06.04.01.01.01.03.03.00.00			06.0E.2B.34.04.01.01.06.04.01.01.01.03.01.00.00	06.0E.2B.34.04.01.01.06.04.01.01.01.03.02.00.00
	Transfer Characteristic	06.0E.2B.34.04.01.01.01.04.01.01.01.01.01.02.00.00				
	Coding Equations	06.0E.2B.34.04.01.01.01.04.01.01.01.01.02.02.00.00	N/R	N/R	06.0E.2B.34.04.01.01.01.04.01.01.01.02.01.00.00	
	Full/narrow level range	Inferred (indicated in black reference level, white reference level, Color range)				
	4:2:0 chroma sample location alignment	Inferred (Chroma LocType=0)	N/A	N/A	Inferred (Chroma LocType=0)	Inferred (Chroma LocType=0)

Table 2 Notes:



- 
1. QuickTime/MP4/HEVC/AVC: ITU-T Series H.Supplement 19, Usage of Video Signal Type Code Points. <https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=13895>
  2. SMPTE UL's available: [https://registry.smpite-ra.org/view/published/labels\\_view.html](https://registry.smpite-ra.org/view/published/labels_view.html)


**Table 3. File-Based Signaling for SDR/BT.2020**

Color properties	Color Primaries	BT.2020	BT.2020
	Transfer Characteristics	BT.2020	BT.2020
	Signal Format	Y'CbCr	R'G'B'
Other	Full/narrow range	Narrow	Narrow
	4:2:0 chroma sample location alignment	Co-sited	N/A
CICP parameters Rec. ITU-T H.273   ISO/IEC 12091-2 [113]  (QuickTime/ HEVC/AVC) <sup>1</sup>	Color Primaries	9	9
	Transfer Characteristics	14	14
	Matrix Coefficients	9	0
	Video Full Range Flag	0	0
SMPTE MXF parameters SMPTE ST 2067-21 <sup>2</sup>	Color Primaries	06.0E.2B.34.04.01.01.0D.04.01.01.01.03.04.00.0	
	Transfer Characteristic	06.0E.2B.34.04.01.01.0E.04.01.01.01.01.09.00.00	
	Coding Equations	06.0E.2B.34.04.01.01.0D.04.01.01.01.02.06.00.00	N/R
	Full/narrow level range	Inferred (indicated in black reference level, white reference level, color range)	
	4:2:0 chroma sample location alignment	Inferred (ChromaLocType = 2)	N/A

**Table 3 Notes:**

1. QuickTime/MP4/HEVC/AVC: ITU-T Series H. Supplement 19, Usage of Video Signal Type Code Points. <https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=13895>

2. SMPTE UL's available: [https://registry.smpite-ra.org/view/published/labels\\_view.html](https://registry.smpite-ra.org/view/published/labels_view.html)


**Table 4. File-Based Signaling for HDR/BT.2020**

	<u>System Identifier</u>	<u>BT.2100 PQ</u> <u>YCC</u>	<u>BT. 2100</u> <u>HLG</u> <u>YCC</u>	<u>BT.2100 PQ</u> <u>IC<sub>T</sub>C<sub>P</sub></u>	<u>BT.2100 PQ</u> <u>RGB</u>	<u>BT.2100</u> <u>HLG</u> <u>RGB</u>
Color properties	Color primaries	BT.2020 / BT.2100	BT.2020 / BT.2100	BT.2100	BT.2020 / BT.2100	BT.2020 / BT.2100
	Transfer Characteristics	BT.2100 PQ	BT.2100 HLG	BT.2100 PQ	BT.2100 PQ	BT.2100 HLG
	Signal Format	Y'CbCr	Y'CbCr	IC <sub>T</sub> C <sub>P</sub>	R'G'B'	R'G'B'
Other	Full/narrow range	Narrow	Narrow	Narrow	Narrow	Narrow
	4:2:0 chroma sample location alignment	Co-sited	Co-sited	Co-sited	N/A	N/A
CICP parameters Rec. ITU-T H.273 ISO/IEC 23091-2 [113]  (QuickTime/HEVC/AVC) 1	Color Primaries	9	9	9	9	9
	Transfer Characteristics	16	18	16	16	18
	MatrixCoefficients	9	9	14	0	0
	Video Full-Range Flag	0	0	0	0	0
SMPTE MXF parameters SMPTE	Color Primaries	06.0E.2B.34.04.01.01.0D.04.01.01.01.03.04.00.00				
	Transfer Characteristic	06.0E.2B.34.04.01.01.0D.04.01.01.01.01.0A.00.00	06.0E.2B.34.04.01.01.0D.04.01.01.01.0B.00.00	06.0E.2B.34.04.01.01.0D.04.01.01.01.0A.00.00	06.0E.2B.34.04.01.01.0D.04.01.01.01.0A.00.00	06.0E.2B.34.04.01.01.0D.04.01.01.01.0B.00.00



ST 2067- 21 <sup>2</sup>	Coding Equations	06.0E.2B.34.04.01.01.0D.04.01.0 1.01.02.06.00.00		06.0E.2B.34.04. 01.01.0D.04.01.0 1.01.02.07.00.00	N/R	N/R
	Full/narrow level range	Inferred (indicated in black reference level, white reference level, Color range)				
	4:2:0 chroma sample location alignment	Inferred (ChromaLocType = 2)	Inferred (ChromaLocType = 2)	unknown	N/A	N/A

**Table 4 Notes:**

1. QuickTime/MP4/HEVC/AVC: ITU-T Series H.Suppment 19, Usage of Video Signal Type Code Points. <https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=13895>
2. SMPTE UL's available: [https://registry.smpite-ra.org/view/published/labels\\_view.html](https://registry.smpite-ra.org/view/published/labels_view.html)

In one method, the SDR transfer function indicator is signaled in the VUI and the HLG transfer function indicator is transmitted using an alternative transfer characteristics SEI message embedded in the bitstream. In this way, an “HLG aware” STB or decoder/display would recognize that the bitstream refers to content coded with HLG (since it is indicated by the preferred\_transfer\_characteristics syntax element of the SEI). If an “HLG aware” STB is connected to a TV that does not support HLG, the STB would transmit the SDR indicator over HDMI to the TV. If it is connected to a TV that supports HLG, the STB would copy the transfer function value in the SEI (to indicate HLG) and transmit this over HDMI to the TV.

In the other method, the HLG transfer function indicator is directly signaled in the VUI in the same way PQ or SDR would be signaled.

In theory it is possible to achieve a lossless conversion between the two methods of signaling HLG by flipping the VUI transfer function characteristics indicator value and inserting or removing the alternative transfer characteristic SEI.

Using the first method (i.e., including the SDR transfer function indicator in the VUI and the HLG transfer function indicator in the SEI) enables some level of backward compatibility with SDR/WCG displays. Service providers may also deem results to be acceptable on SDR/WCG displays using the second method (i.e., including the HLG transfer function indicator in the VUI). Service providers may wish to test both methods.

QuickTime and MP4 use MPEG CICC (Coding Independent Code Points) values for signaling. MXF Image Essence Descriptors values are used to signal these same values.



As described in [Section 11.1](#), service providers should convert or remap all content into a single, consistent system colorimetry and transfer function. Setting the initial values in the encoder should be adequate assuming the encoder is dedicated to a single format.

As of 2023, while there are methods for signaling system colorimetry, transfer function, matrix coefficients and HDR-related metadata through the end-to-end supply chains, however due to legacy devices with limited support for these capabilities it is extremely difficult to successfully ensure that the signaling and metadata survive through the entire broadcast linear production pipeline. Because gaps exist with legacy devices using SDI or HDMI interfaces and software that doesn't consistently support file-based signaling in the mezzanine file wrapper, verification of proper signaling and accurate video passthrough is recommended. The Ultra HD Forum observes that standards bodies have attempted to address these issues and provided documentation, much of which is referenced herein, but some of which is still under development.

## 7.4. $IC_T C_P$ Color Representation

The expanding range of display technologies, from various color primaries to increasing dynamic range, is creating a marketplace where color management is becoming increasingly important if artistic intent is to be maintained. In many applications, traditional color transformations may not be possible due to limitations in bandwidth, speed, or processing power. In these cases, image processing such as blending, resizing, and color volume transform must be performed on the incoming signal. With growing color volumes and the increasing need for color processing, distortions already known to be caused by non-uniformity of standard dynamic range (SDR) non-constant-luminance (NCL)  $Y'C'_B C'_R$  (hue linearity and constant luminance) will become more prevalent and objectionable.

The  $IC_T C_P$  color representation is a more perceptually uniform color format based on the human visual system. The improved decorrelation of saturation, hue, and intensity make  $IC_T C_P$  ideal for the entire imaging chain from scene to screen.  $IC_T C_P$  follows the same operations as NCL  $Y'C'_B C'_R$ , making it a possible drop-in replacement. These color processing improvements are achieved by utilizing aspects of the human visual system and by optimizing for lines of constant hue, uniformity of just-noticeable-difference (JND) ellipses, and constant luminance. The perceptually uniform design of  $IC_T C_P$  allows for complex tasks such as color volume transform to be easily performed on HDR and WCG imagery with minimal error.

$IC_T C_P$  is included in [BT.2100 \[5\]](#) and is being deployed by OTT service providers as well as implemented by numerous consumer TV manufacturers. Another derivative of  $IC_T C_P$ , ITP, is





---

described in [ITU-R BT.2124 \[107\]](#) and is described in [Section 8.4](#) for use as an objective color metric that can be useful in testing the entire broadcast pipeline.

## 7.5. Peak Brightness Considerations

Luminance values can be coded as “absolute values” on a scale of 0-10,000 cd/m<sup>2</sup> (e.g., PQ format) or as “relative values” (e.g., RAW, gamma, S-log or HLG). The optimal coding method depends on the application and other considerations.

Cameras can be capable of capturing a higher contrast range than reference monitors and consumer displays are capable of rendering. When content is graded, the grader makes judgments according to what he or she sees within the luminance range of the reference monitor, which may have a peak brightness of 1,000 or 2,000 cd/m<sup>2</sup>, for example. In theory, if the consumer display characteristics match that of the reference monitor, the consumer will see what the grader intended. In practice of course, consumer HDR displays have varying peak brightness performance, black level and color gamut, so it is up to the display (or source device) to map the color volume of the transmitted content to best match the capabilities of the particular display.

As a PQ signal may carry pixel values as high as 10,000 cd/m<sup>2</sup>, it is helpful in the display to indicate the actual maximum pixel value that are expected and are creatively pertinent as a basis for color volume transforms. HDR10 systems are capable of providing static metadata to assist mapping the transmitted content for the consumer display, enabling a given display to optimize the color volume mapping based on its capability. The use of HDR10 metadata may help to optimize display mapping, but as the metadata may not be practical in Live broadcast content, the use of PQ10 may be preferred for Live content. Although a system objective is to preserve creative intent with limited perceptual alteration all the way to the consumer display, some perceptual alteration will occur because of different display capabilities both in peak brightness and minimum black levels.

HLG does not require tone-mapping metadata, and instead has a specified relationship between overall system gamma (implemented as part of the display EOTF) and nominal peak display luminance. An overall system gamma of 1.2 is specified for HLG when displayed on a 1,000 nit monitor. BT.2390 [6] states that the artistic intent of HLG content can be preserved when displaying that content on screens of different peak luminance (even when that display is brighter than the mastering display), through changing the system gamma. [BT.2100 \[5\]](#) provides a formula to determine overall system gamma based on the desired display peak luminance in the reference viewing environment. Section 6.2 of [BT.2390 \[6\]](#) further specifies how the system



gamma should be adapted for darker or brighter viewing environments. To preserve creative intent, these formulae are recommended.

[ITU-R BT.2408 \[8\]](#) recommends the reference level for “HDR Reference White” be set to 75% HLG (equivalent to 203 cd/m<sup>2</sup> with an HLG display normalized at 1,000 cd/m<sup>2</sup>) or 58% PQ (PQ represents absolute light levels and thus Reference White will be 203 cd/m<sup>2</sup> on a reference display). Reference white can serve as a good anchor point around which the rest of the image can be shaded. 203 cd/m<sup>2</sup> was chosen to provide sufficient headroom for “highlights” while allowing comfortable viewing in HDR/WCG and after conversion to SDR/WCG. Some experts are considering a more flexible range for HDR reference white to maintain an acceptable SDR average picture level (APL) when down mapped. When selectively down mapping the HDR anchor point into SDR, the APL will change. Down mapping HDR reference white to SDR at a lower signal level can allow the preservation of additional highlights in HDR but also a lower SDR APL. Conversely, down mapping HDR reference white into SDR at a higher signal level will cause more clipping with loss of detail and a higher SDR APL. HDR reference white serves as a visible anchor point for conversions to-and-from SDR and ultimately establishes the ability to “roundtrip” through the entire workflow in a “single-master workflow” (SDR->HDR->SDR).

Alternatively, HLG can support almost an additional stop of dynamic range and a larger color volume by utilizing the commonly referred “super-white” signal range (between 10-bit code values 941 and 1019). Professional monitors are available which can display HLG super-white signals. Since not all displays support this range with HLG10, the extended color volume cannot be guaranteed.

Note that future displays may become available with higher peak brightness capability compared with those available today. Content that is expected to be of interest to consumers for many years to come may benefit from retaining an archive copy coded in “absolute values” of light (e.g., PQ) or original camera capture format (e.g., RAW, log) so that future grades of higher luminance range can be produced and delivered to viewers.

### 7.5.1. Studio Video over IP

In 2017, the Society of Motion Picture and Television Engineers (SMPTE) published the first of its SMPTE ST 2110 suite of standards for “Professional Media Over Managed IP Networks”, which exploits IP networking equipment for transport of video, audio, and metadata within television facilities, in addition to or in lieu of traditional serial digital interfaces (SDI). The SMPTE ST 2110 suite is built upon the Real time Transport Protocol (RTP) as described in RFC 3550 [42], with the suite providing sufficient constraints such that best-effort IP connections can



succeed where historically, bandwidth and synchronization was provided by SDI connections, which offered assured point-to-point latencies.

The SMPTE ST 2110 suite currently comprises five documents: [SMPTE ST 2110-10 \[43\]](#) specifies the network interface requirements, overall system timing model, and session description protocols. [SMPTE ST 2110-20 \[44\]](#) specifies the format for uncompressed video essence, while [SMPTE ST 2110-21 \[45\]](#) identifies timing characteristics for these video streams and requirements for compliant senders and receivers. [SMPTE ST 2110-30 \[46\]](#) specifies the format for uncompressed digital audio. Lastly, [SMPTE ST 2110-40 \[47\]](#), published in 2018, specifies carriage of SMPTE ST 291-1 Ancillary Data (i.e., ANC messages). Each of the video, audio, and metadata essences described in SMPTE ST 2110-20, -30, and -40 are individual streams that are synchronized in accordance with SMPTE ST 2110-10.



## 8. Conversions Between Transfer Functions

A receiver may not be able to switch seamlessly between HDR transfer functions; therefore, it is recommended that only one transfer function be used for a given Real-time Program Service in Foundation Ultra HD. This section offers guidelines to service providers to convert HDR content from PQ10 to HLG10 or vice versa in order to maintain a single transfer function in the service. Equations and best-practices for these conversions can be found in [ITU-R BT.2390 \[6\]](#).

### 8.1. Conversion between HDR formats

#### 8.1.1. PQ10 to HLG10

It is possible some providers may receive pre-recorded programs that are produced in PQ10 (a Foundation UHD format) but may require conversion to the HLG10 format for their specific distribution pipeline. Since PQ10 (which uses the PQ transfer function) can contain peaks as high as 10,000 cd/m<sup>2</sup>, [BT.2408 \[8\]](#) section 6.4 suggests that PQ10 content be pre-normalized to a nominal peak luminance of 1000 cd/m<sup>2</sup> prior to conversion. Once PQ10 is normalized to 1000 cd/m<sup>2</sup>, it is possible to perform the required conversion to HLG10 a 3D LUT. Note, due to the conversion, there can be a small loss of color volume in HLG10 if signals are clipped to the HLG nominal signal range ( $x=1.0$ ) as noted in [Table 5](#) below, taken from BT.2408 Section 6.5.

[Table 5](#) illustrates that there are LUTs can be designed to preserve a larger color volume in HLG10, by exploiting the 10-bit code values above nominal video range (code values 941-1019, often referred to as “super-whites”), thus minimizing the color volume differences between PQ and HLG with a 1,000 cd/m<sup>2</sup> peak white. Because the narrow range 10-bit code value for PQ at a white level of 1000 cd/m<sup>2</sup> is approximately 723 (for Y’ of YCbCr), nominal video range usage is all that is necessary for PQ.

As of 2022, it has been noted that some reference displays may not process HLG10 code values beyond narrow/nominal video range (e.g. 10-bit code values 941-1019) when in the factory default mode. Results using these default levels are therefore not assured in the displayed image, which could result in loss of visual detail or minor color volume<sup>6</sup>.

---

<sup>6</sup> Note, this issue may also be evident in some consumer displays.



**Table 5. BT.2408 Signal ranges and achievable color volume for PQ and HLG on a 1000 cd/m<sup>2</sup> nominal peak luminance display**

Color	BT. 2100 PQ Luminance <sup>1</sup> (cd/m <sup>2</sup> )	BT. 2100 HLG Luminance (cd/m <sup>2</sup> )	
Max value of color	x = 0.75	x = 1.0	Max non-linear signal, E', to match PQ luminance
{x,x,x} // Peak white	1,000.0	1,000.0	R'=G'=B'= 1.000
{x,0,0} // Peak red	262.7	201.1	R' = 1.041
{0,x,0} // Peak green	678.0	627.3	G' = 1.012
{0,0,x} // Peak blue	59.3	33.7	B' = 1.086

**Table 5 Notes:**

1. PQ nominal peak white at 1,000 cd/m<sup>2</sup>
2. In practice some displays might not achieve a luminance higher than their nominal peak value.

Note that delivery of a Ultra HD HDR program, encoded with HEVC, it is possible to encode the PQ-related metadata ([SMPTE ST 2086 \[10\]](#), CLLI) in a SEI message even though HLG is signaled as the transfer function. At this time, that practice is not defined in any current standard, and thus it should be avoided or employed with caution due to unexpected behavior by the consumer displays.

### 8.1.2. HLG10 to PQ10

It is possible that Live programs may be produced in HLG10 (a Foundation Ultra HD format); however, some service providers may favor PQ10 delivery making it necessary to convert HLG10 encoded content into PQ10. This conversion can be accomplished with a static or programmable 3D LUT prior to encoding. Note that this conversion involves a translation from a scene-referred relative signal (HLG10) to a display-referred absolute signal (PQ10). As such, a target peak luminance needs to be established as part of the LUT conversion function. [ITU-R Report BT.2390 \[6\]](#) recommends using a nominal peak luminance of 1,000 cd/m<sup>2</sup>, so that value may be used; i.e., the nominal peak value of the HLG signal (100% HLG) would map to the PQ value of 1,000 cd/m<sup>2</sup>. HLG10 production signal flows should preserve overshoot levels above



---

nominal video range (up to 1810 cd/m<sup>2</sup> white) with an HLG10 overall system gamma of 1.2 ([EBU r103, “Video signal tolerance in Digital Television Systems \[106\].”](#)).

## 8.2. Conversion of SDR to HDR: Up-Mapping

The [Figure 3](#) and [Figure 4](#) below represent the results of up-mapping tests of a camera color chip chart (X-Rite Color Checker® 709-2014, See [Figure 2](#)) between BT.709 SDR and BT.2100 HLG using the Sony® HDRC-4000 HDR Production Converter Unit. Blue circles represent the original color, with a 2x linear scaling applied to adapt the displayed luminance from the usual 100 cd/m<sup>2</sup> SDR nominal peak white to the HDR Reference White level of 203 cd/m<sup>2</sup>. Green triangles represent the converted colors. [Figure 3](#) is a plot of scene-light conversion from the X-Rite ColorChecker® test pattern and [Figure 4](#) is a plot of display-light conversion. These 2D plots show the color accuracy of the chip chart sample conversions. The chip chart provides a limited set of target colors and luminance levels. Other test patterns, such as the Sarnoff “Yellow Brick Road” (see [Figure 6](#)) with larger color and light ranges can provide additional color and luminance information on color volume accuracy.

While the purpose of scene-light conversion is to match SDR/HLG tone-mapping and color presentations with the least error, the color plots show that the match is not perfect due to each format having different OOTF's and color gamuts.



Figure 2. X-Rite Color Checker Chip Chart



## SDR-BT.709 to HLG-BT.2100 Scene-Light Conversion

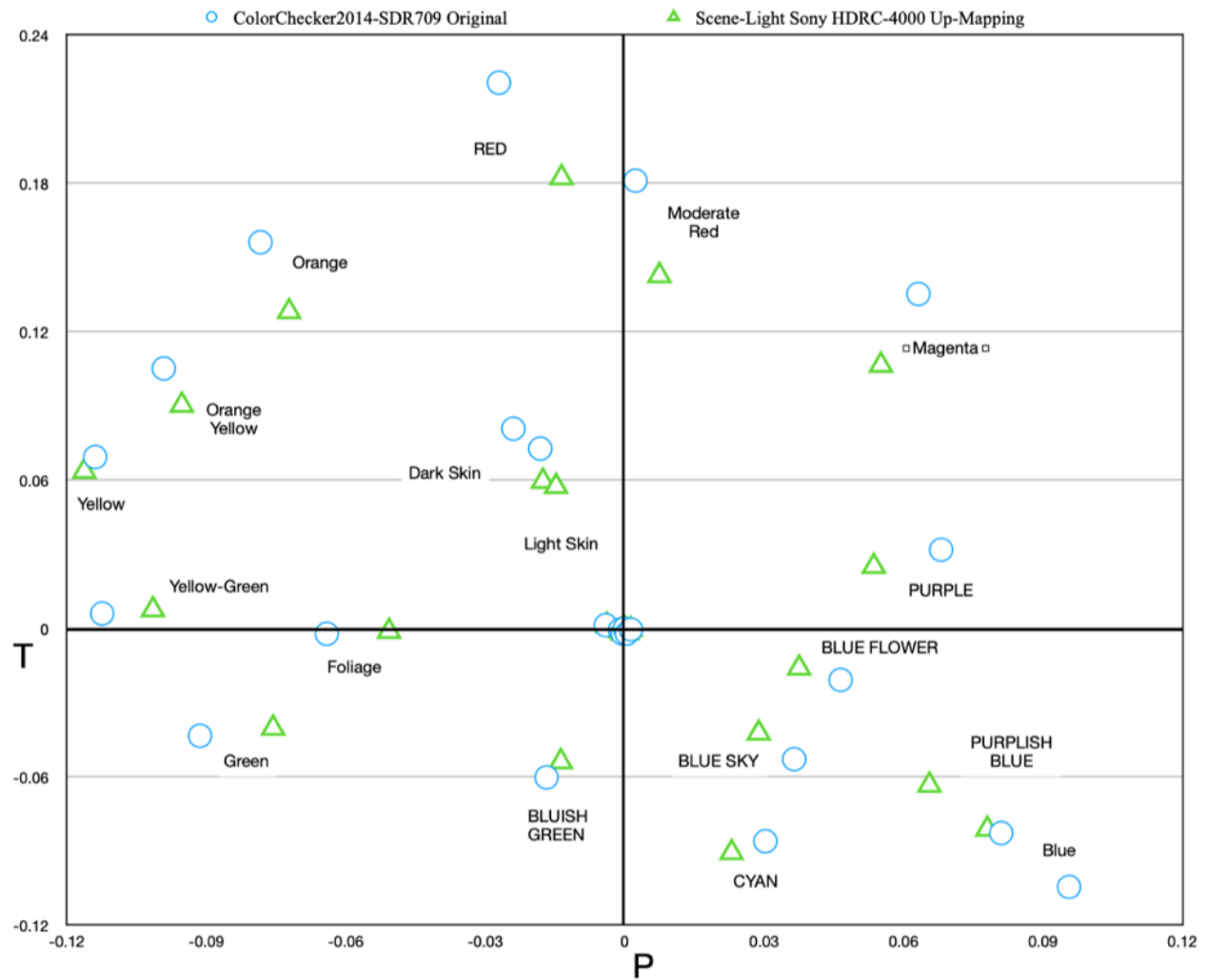


Figure 3. Scene Light Conversion Accuracy, SDR to HLG





## SDR-BT.709 to HLG-BT.2100 Display-Light Conversion

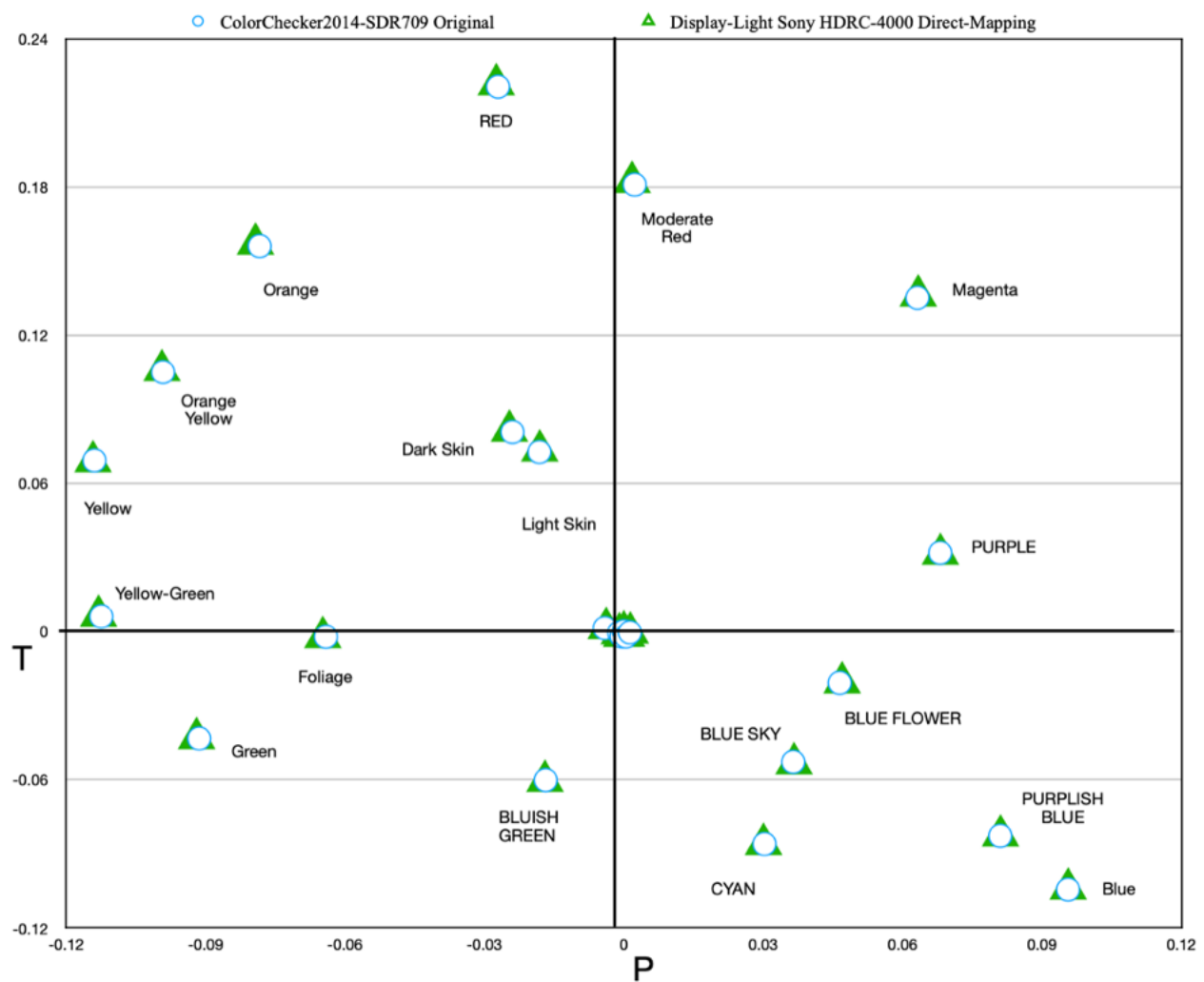


Figure 4. Display Light Conversion Accuracy, SDR to HLG



**Table 6.  $\Delta E$ -ITP Values for the Up-mapping from SDR to HLG**

ITP Difference Original vs LUT Converters	
Colorchecker CSM Suite HLG 709 colors in 2020 Container	Sony HDC4000 Display Referred SDR->HLG
	$\Delta E$ -ITP
dark skin	0.6107
light skin	0.4276
blue sky	0.5727
foliage	0.4997
blue flower	0.4987
bluish green	0.3500
orange	0.9263
purplish blue	0.4047
moderate red	0.8893
purple	0.3227
yellow green	0.5079
orange yellow	0.5393
blue	0.4771
green	0.5339
red	0.7252
yellow	0.3631
magenta	0.7047
cyan	0.3159
white	0.2364
Gray 2	0.2208
Gray 3	0.2589
Gray 4	0.2547
Gray 5	0.2193
Gray 6	0.1750



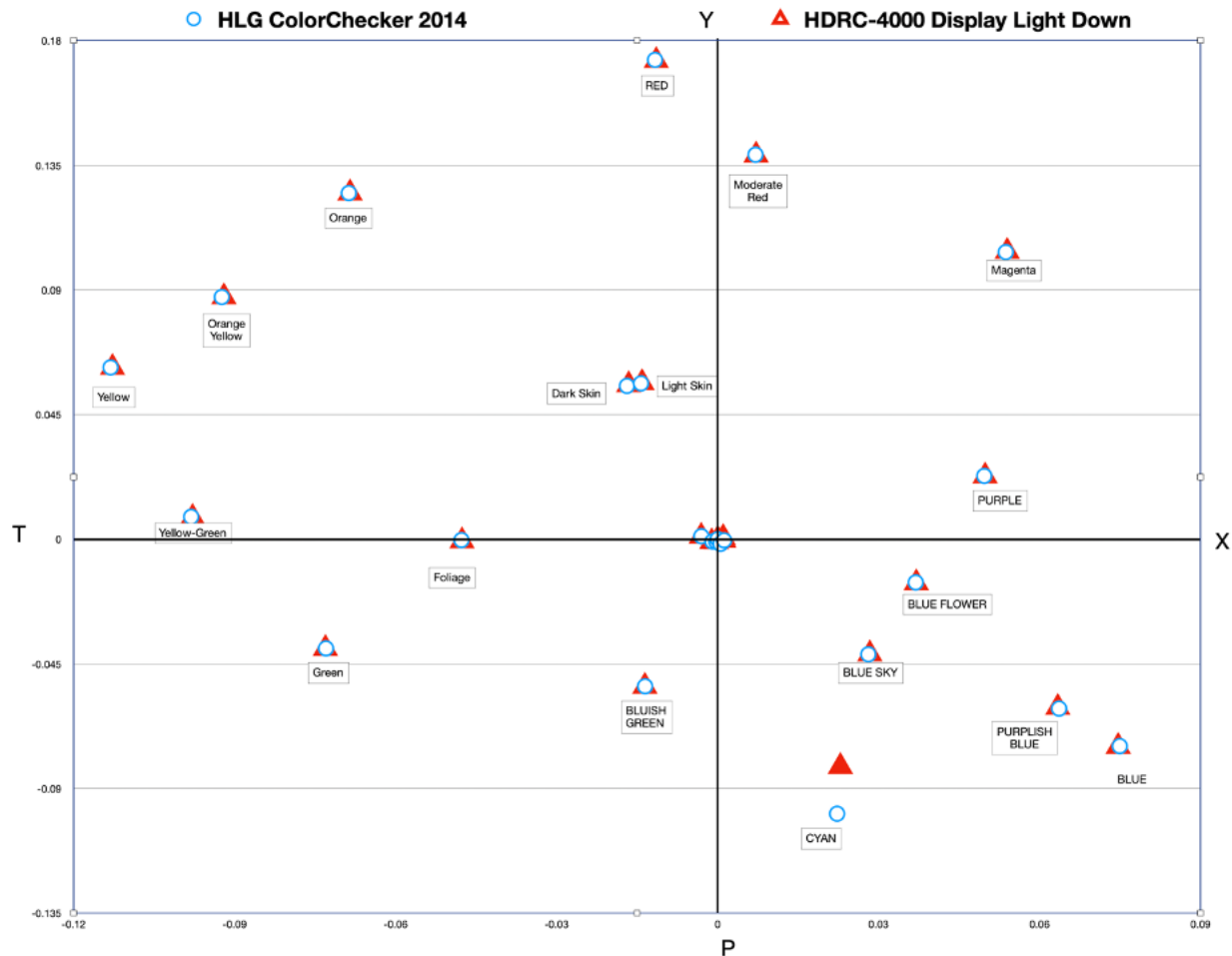
### 8.3. Conversion HDR to SDR: Down-Mapping

Two common approaches have emerged from the HDR to SDR format conversion:

- One approach attempts to maintain the subjective appearance of the HDR content when converted to SDR, and viewed in a reference BT.2035 environment (100 cd/m<sup>2</sup> with a display gamma of 2.4). Examples of this type of converter are described in ITU-R Report BT.2446, and the approach has been adopted by Sky for its live Premier League coverage and the BBC in their licensed conversion LUTs.
- The other approach attempts to maintain the subjective appearance of the HDR content when converted to SDR, and viewed on a 203 cd/m<sup>2</sup> Gamma 2.4 display that conforms to ITU BT.1886. This approach is described in Annex J and has been adopted by NBCUniversal, Fox Sports, and others.

The 200 cd/m<sup>2</sup> shading approach allows SDR and HDR monitors to be used in close proximity in production, so that the content appears well matched between the HDR and SDR displays. It may also provide a better match on SDR consumer displays, which now tend to be brighter than the reference display based on 100 cd/m<sup>2</sup>. However, if 203 cd/m<sup>2</sup> SDR content is viewed on a 100 cd/m<sup>2</sup> display, mid-tones will appear subjectively darker. Conversely, when 100 cd/m<sup>2</sup> SDR content is viewed on a consumer TV the mid-tones may appear subjectively brighter.

Discussion is under-way with industry bodies to address the differences in SDR levels arising from the two approaches, which should be formalized for international program exchange. Techniques are being investigated that should ensure that signals derived using either approach appear the same on consumer TVs.



**Figure 5. Down-Mapping HLG to SDR Color Accuracy**

The [Figure 5](#) represents display-light down-mapping of BT.709 colors to BT.2100 HLG colors using the Sony HDRC-4000. Blue circles represent the original color, red triangles represent the down-mapped colors. These X/Y plots show the chip chart sample color accuracy. BT.709 colors down-mapped to an HLG-BT.2100 container are preserved with the exception of cyan is out-of-gamut because of the 2x gain applied to BT.709 colors within a BT2100 HLG container

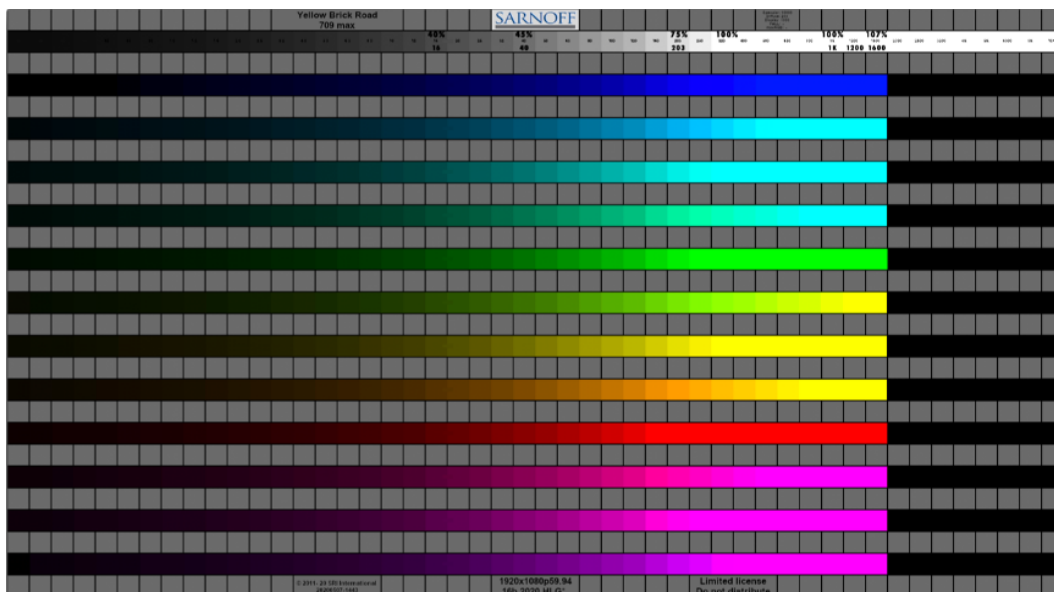
The human visual system response is very complicated and thus the evaluation of a color mapping will be subjective when deciding how to convert from a larger color volume to a smaller one. In some cases, it may be preferable to allow a hue shift in order to preserve saturation. The specific design of 3D LUTs make different choices regarding gamut/luma compression and



when to allow hue shifts. It is important to test the implementation of every LUT and LUT converter for objective accuracy as well as subjective perceptual choices.

Figures 6, 7, 8 and 9 use 360 sample points (HLG constrained to 1600 cd/m<sup>2</sup>) from the Sarnoff® Yellow Brick Road Pattern of the “Color Space & Monitor (CSM) Test Pattern Suite” (Figure 6) to compare three different HLG-BT.2100 to SDR-BT.709 down-mapping conversions. Each plot of the conversion results, in ITP space, demonstrate different choices in quality vs color gamut and luma compression, as well as hue shifts. It is important to assess the results of down-mapping using objective measurements as well on a set of reference displays (side-by-side), to fully evaluate how well the down-mapping achieves a perceptual match.

The Figure 7. is the reference pattern used for measurement of the LUT conversions. Figure 7 is an example of a LUT that produces major color inaccuracies. Figure 8, LUT2 Conversion from HLG to SDR, is an example of a LUT that converts BT.2020 colors to follow the absolute chromaticity of BT.709 primaries. In Figure 9, the conversion follows the absolute chromaticity of BT.2020 primaries while “in-gamut” for BT.709 and then perform a hue shift to align with the BT.709 primaries to achieve optimal saturation and preserve original artistic intent as close as possible within the confines of SDR.



**Figure 6. Sarnoff “Yellow Brick Road”© Test Chart (SDR)  
(courtesy of SRI International)**

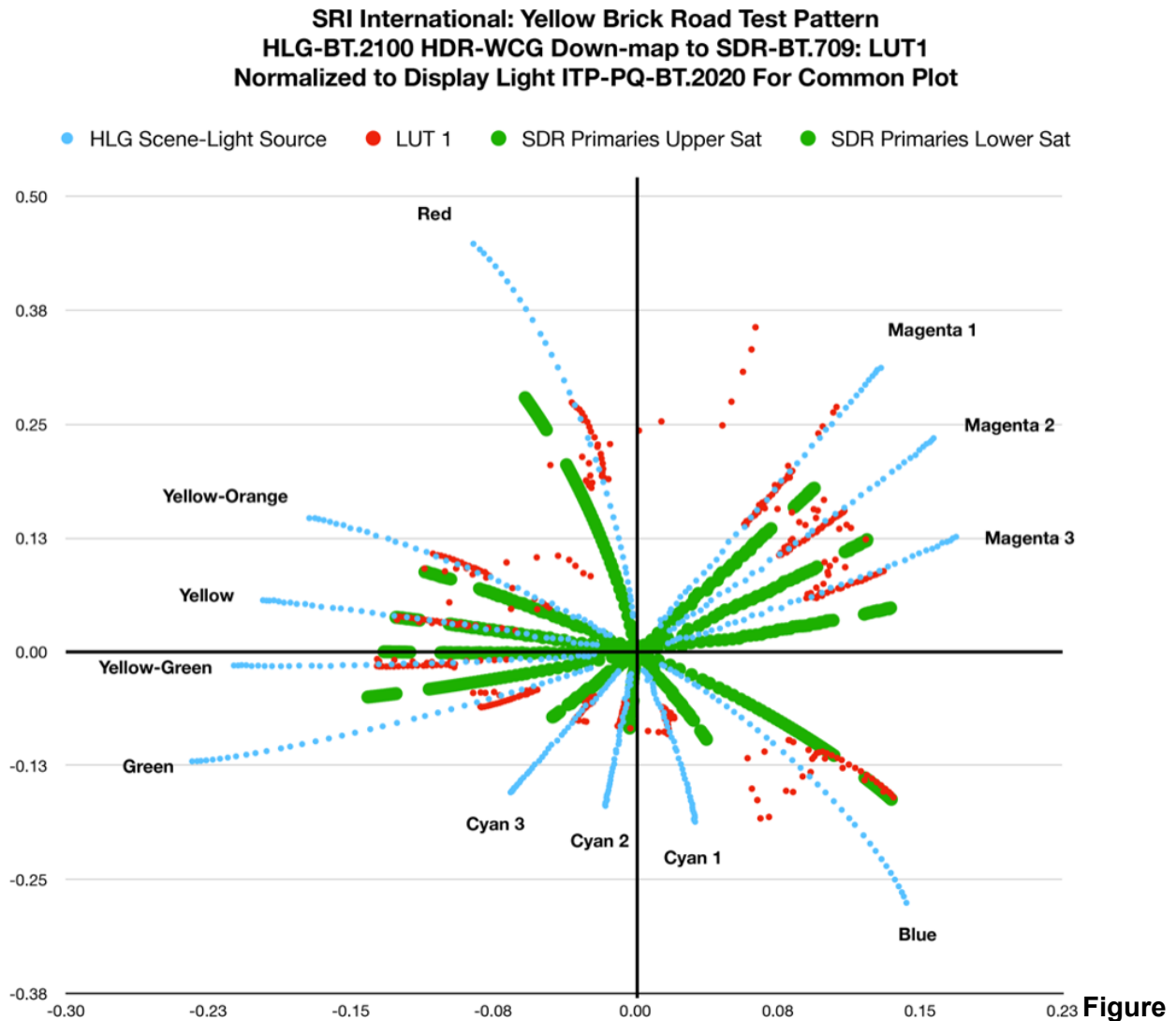
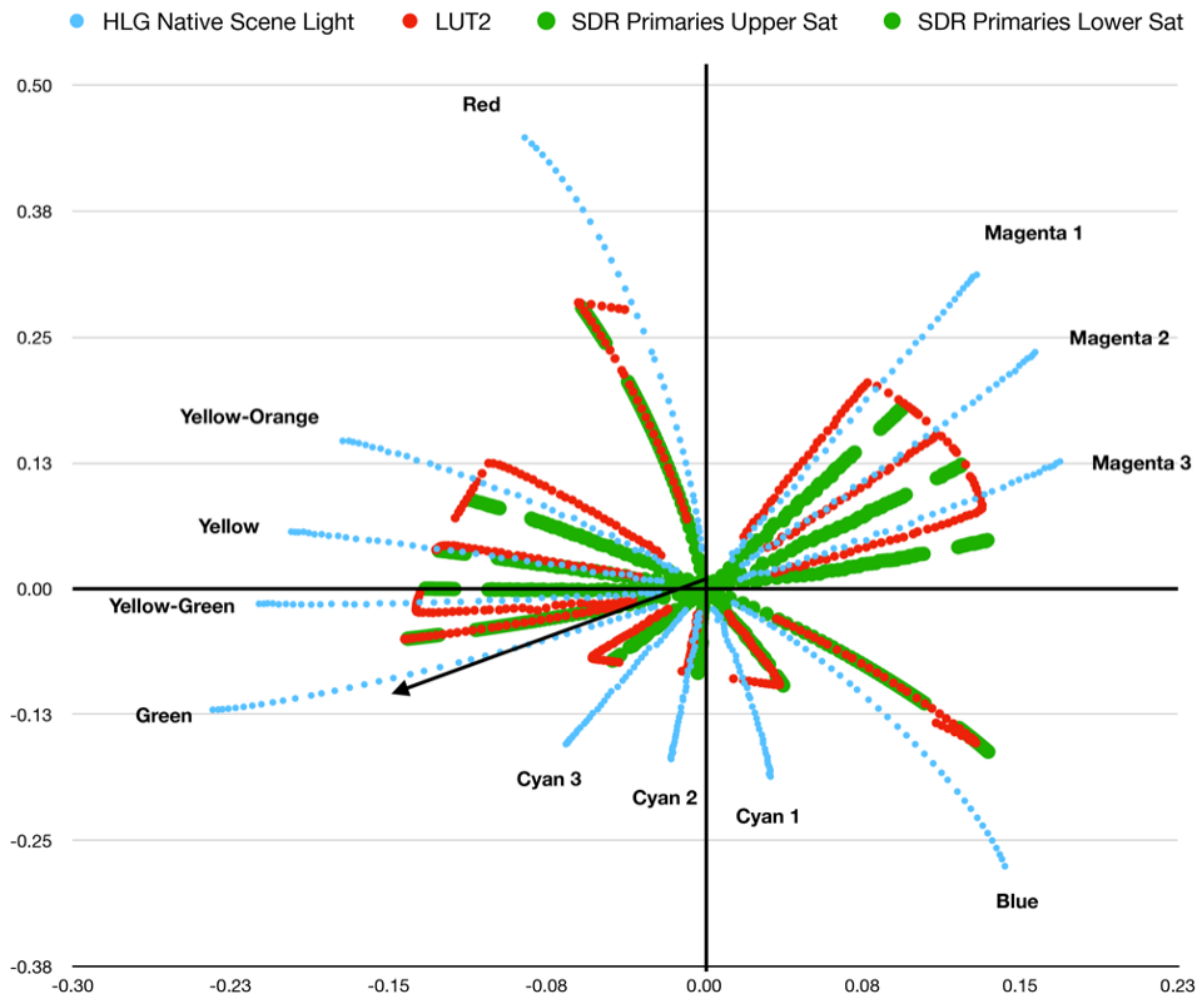


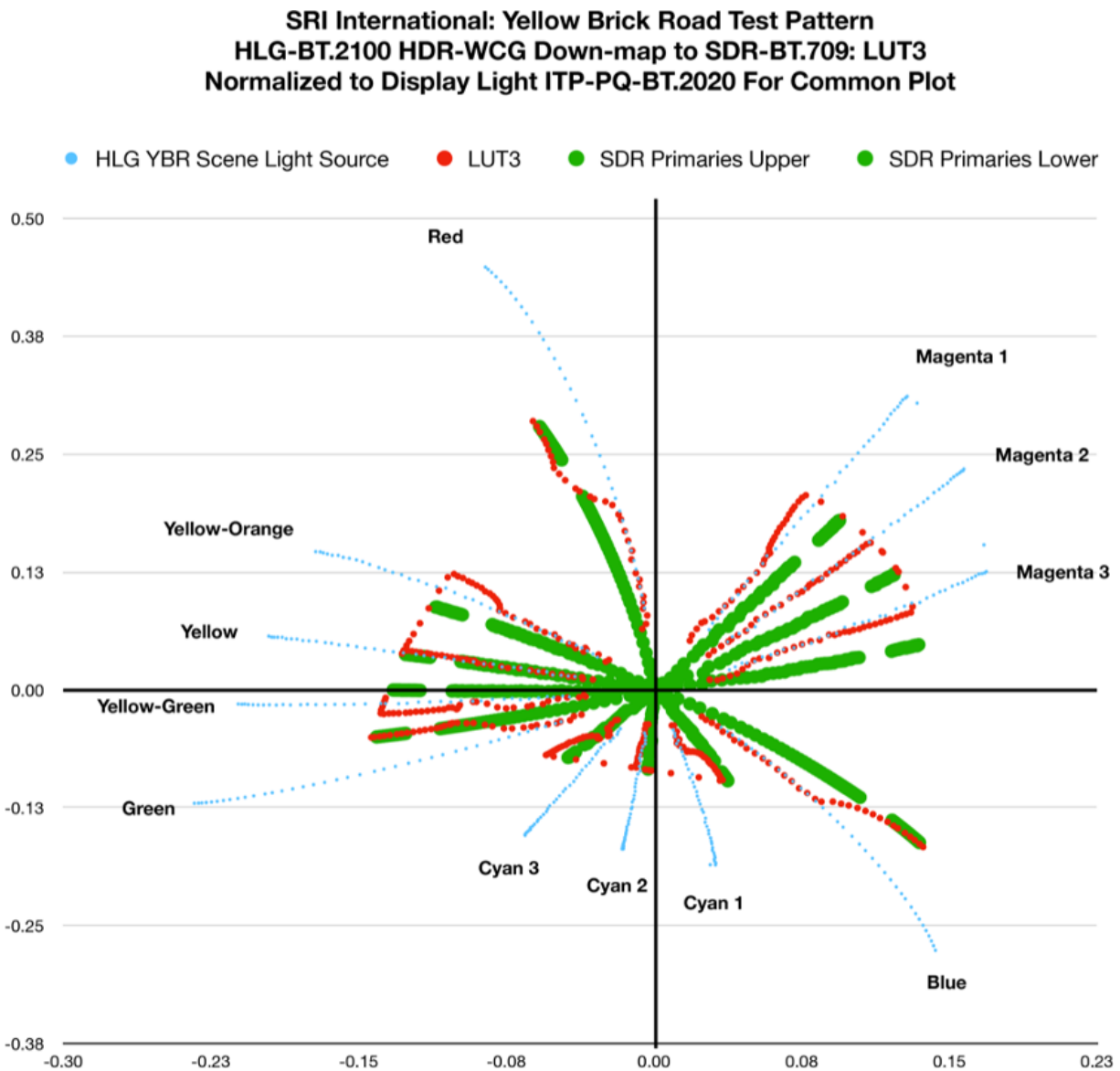
Figure 7. LUT 1 Conversion from HLG to SDR



**SRI International: Yellow Brick Road Test Pattern**  
**HLG-BT.2100 HDR-WCG Down-map to SDR-BT.709: LUT2**  
**Normalized to Display Light ITP-PQ-BT.2020 For Common Plot**



**Figure 8. LUT 2 Conversion from HLG to SDR**

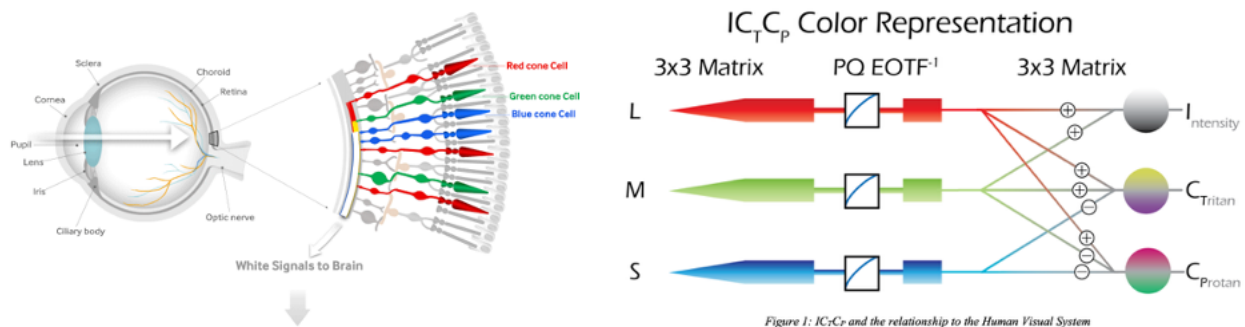






## 8.4. Objective Measurement of the Quality of Tone Mapping

Traditionally video shaders/graders relied on waveform monitors and vector scopes along with subjective viewing on reference displays to establish the appropriate conversion parameters (LUT) between one image format and another. Most often such conversions are provided through a “3D look-up table” (LUT). The ITU has established a new recommendation [BT.2124 \[107\]](#), “Objective metric for the assessment of the potential visibility of color differences in television”, which provides an objective measurement of the color differences between the original image color and the converted image color. BT.2124 uses a  $\Delta E$ -ITP methodology to measure color volume accuracy. A  $\Delta E$ -ITP value of 1 or more indicates a Just-Noticeable-Difference (JND) between the original and the converted image color volume, assuming the most sensitive state of adaptation. The measure is based on  $IC_T C_P$ , specified in [ITU-R Recommendation BT.2100 \[5\]](#) and illustrated in [Figure 10](#). Figure YY is an example of objective plots to measure hue shifts using BT.2124 techniques.



**Figure 10. ICtCp and the Relationship to the HVS**

For measuring hue shifts, we use x/y plots of the chroma components T/P within ITP. T/P components mimic some aspects of the human visual system which allow us to visualize to some degree the “Hunt Effect”, assuming a single adaptation state of the eye. Care should be taken, however, if applying the measure to images displayed at different brightness, as that will affect the adaptation state of the eye and may give a false impression of image colorfulness (re: The Hunt Effect). In which case  $u' v'$  of [CIE Colorimetry-Pt 5: 1976 \[116\]](#) can be used to determine absolute chromaticity (see [Figure 12](#)). T/P plots can, however, still provide a useful indication of hue shifts.



Using  $\Delta E$ -ITP, we can compare a source and UNITY passthrough in a device (without LUT conversion) to verify no alternation of color volume has occurred. Alteration could occur because of poor choices in mathematical precision, using the wrong matrix between YCbCr and RGB (and a number of other possible errors or quality choices). For instance, lower bit-rate codecs will have additional color accuracy errors in many cases.

$\Delta E$ -ITP uses a very simple scale where any number above one (“1”) represents a “just-noticeable-difference” (JND) to the human visual system at its most sensitive state of adaptation. This makes it much easier to get an idea of whether a perceptible issue exists from the plots.

Equations for the measurements<sup>7</sup> that are the basis of the plots in [Figure 11](#):

a)  $\Delta E_{ITP}$  Color Difference (Source vs Output). eq. 8.4(a)

b)  $\Delta TP_{ITP}$  Color Difference - Chroma ONLY (Source vs Output) eq. 8.4(b)

c)  $\Delta I_{ITP}$  Color Difference - Intensity ONLY (Source vs Output) eq. 8.4(c)

d)  $\Delta H_{ITP}$  Hue Difference (Constant Intensity and Chroma) eq. 8.4(d)

Below are the results of video passthrough in a simple frame sync that was being used to realign audio. Different quadrants of a 2160P signal were apparently processed incorrectly which caused a change in color accuracy. The plots below show  $\Delta E$ -ITP,  $\Delta TP_{ITP}$ ,  $\Delta I_{ITP}$ ,  $I_{ITP}$ ,  $\Delta H_{ITP}$  for tracking color volume, Chroma, Intensity, Constant Intensity/Chroma changes and finally identifies hue-shifts using T/P color components from ITP.

<sup>7</sup> A spreadsheet of measurements is available via link here:

[https://github.com/digitaltvguy/NBCU-HDR-SDR-Single-Stream\\_Workflow\\_Recommendation](https://github.com/digitaltvguy/NBCU-HDR-SDR-Single-Stream_Workflow_Recommendation)

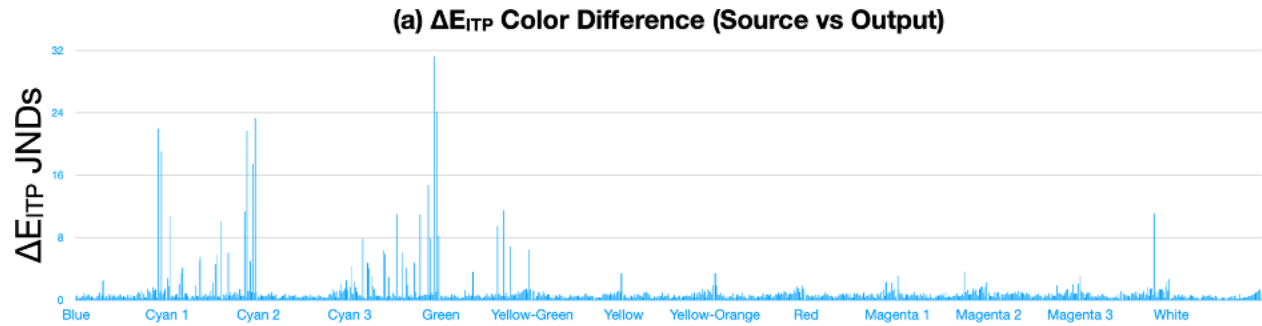


Figure 11.  $\Delta E_{ITP}$  Color Difference (Source vs Output)

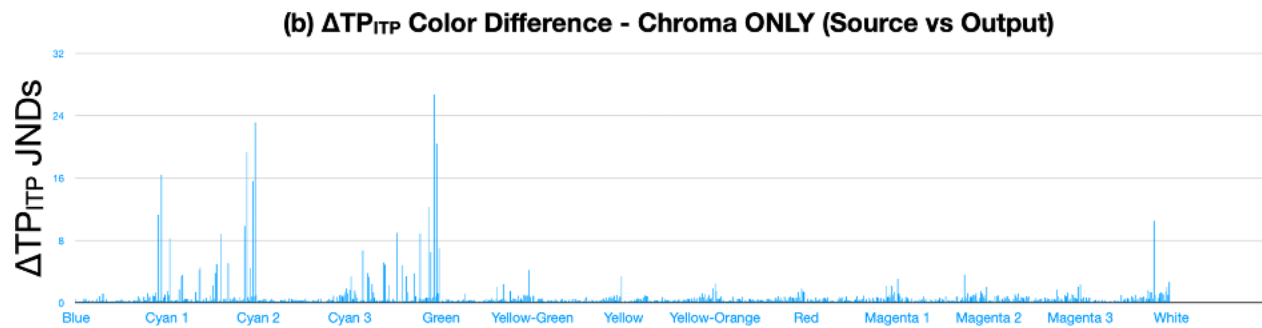


Figure 12.  $\Delta TP_{ITP}$  Color Difference - Chroma ONLY (Source vs Output)

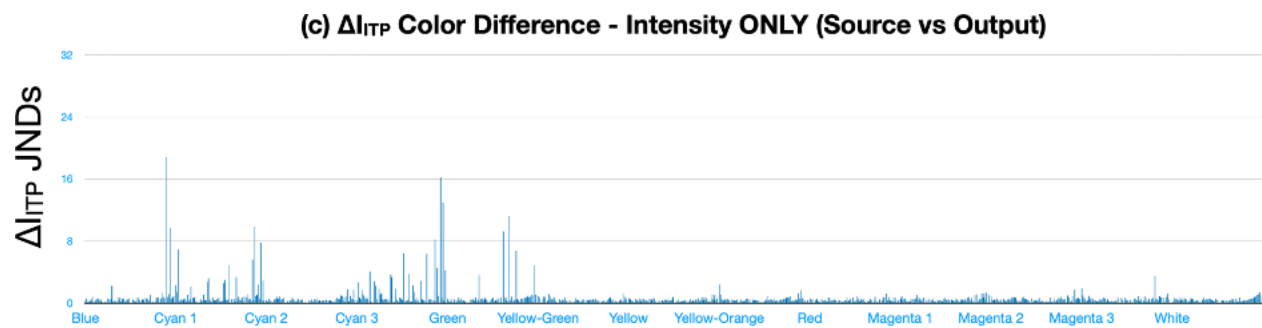


Figure 13.  $\Delta I_{ITP}$  Color Difference - Intensity ONLY (Source vs Output)

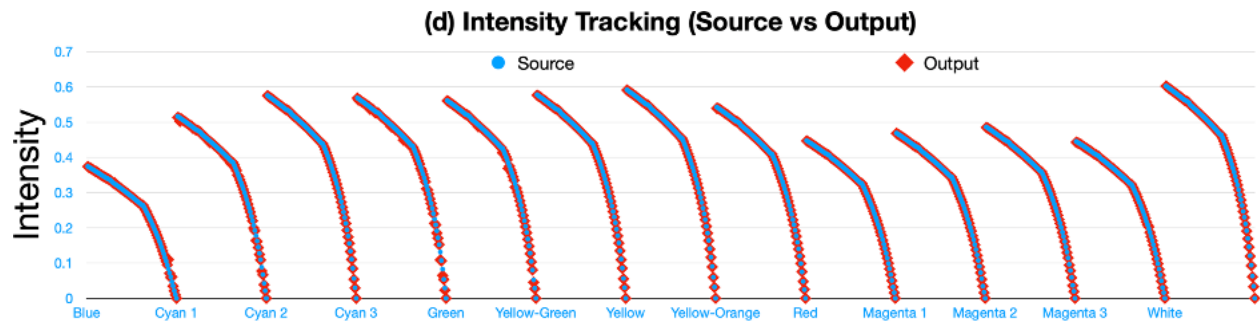
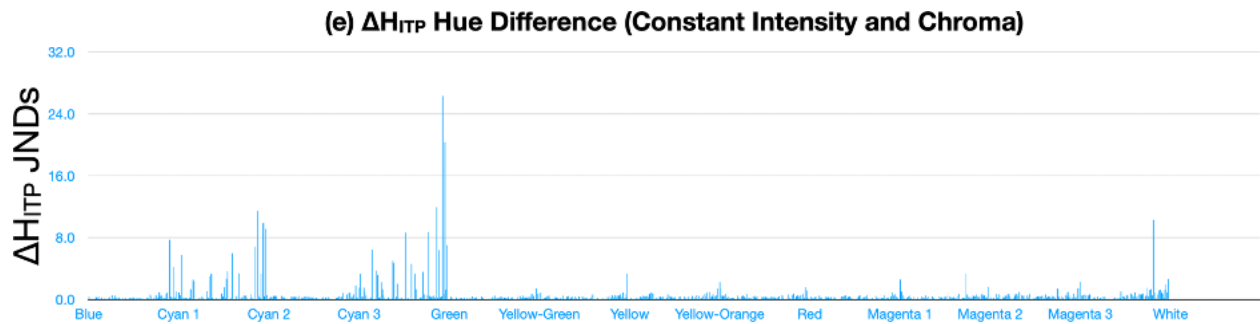
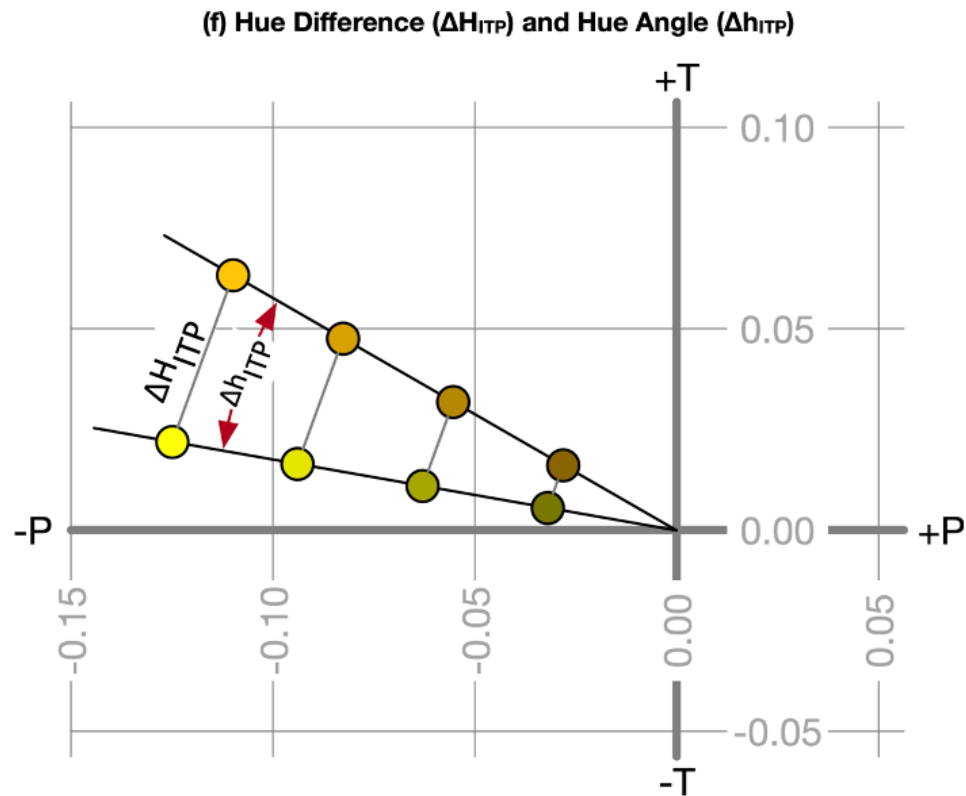


Figure 14. Intensity Tracking (Source vs Output)

Figure 15.  $\Delta H_{ITP}$  Hue Difference (Constant Intensity and Chroma)



The samples in each color difference pair have the same chroma, for being the same distance from the origin. The hue angle difference  $\Delta h_{ITP}$ , in this illustration from yellow to yellow-orange, is constant for these pairs, but the hue difference  $\Delta H_{ITP}$  increases proportionally to a pair's chroma.

**Figure 16. Hue Difference ( $\Delta H_{ITP}$ ) and Hue Angle ( $\Delta h_{ITP}$ )**

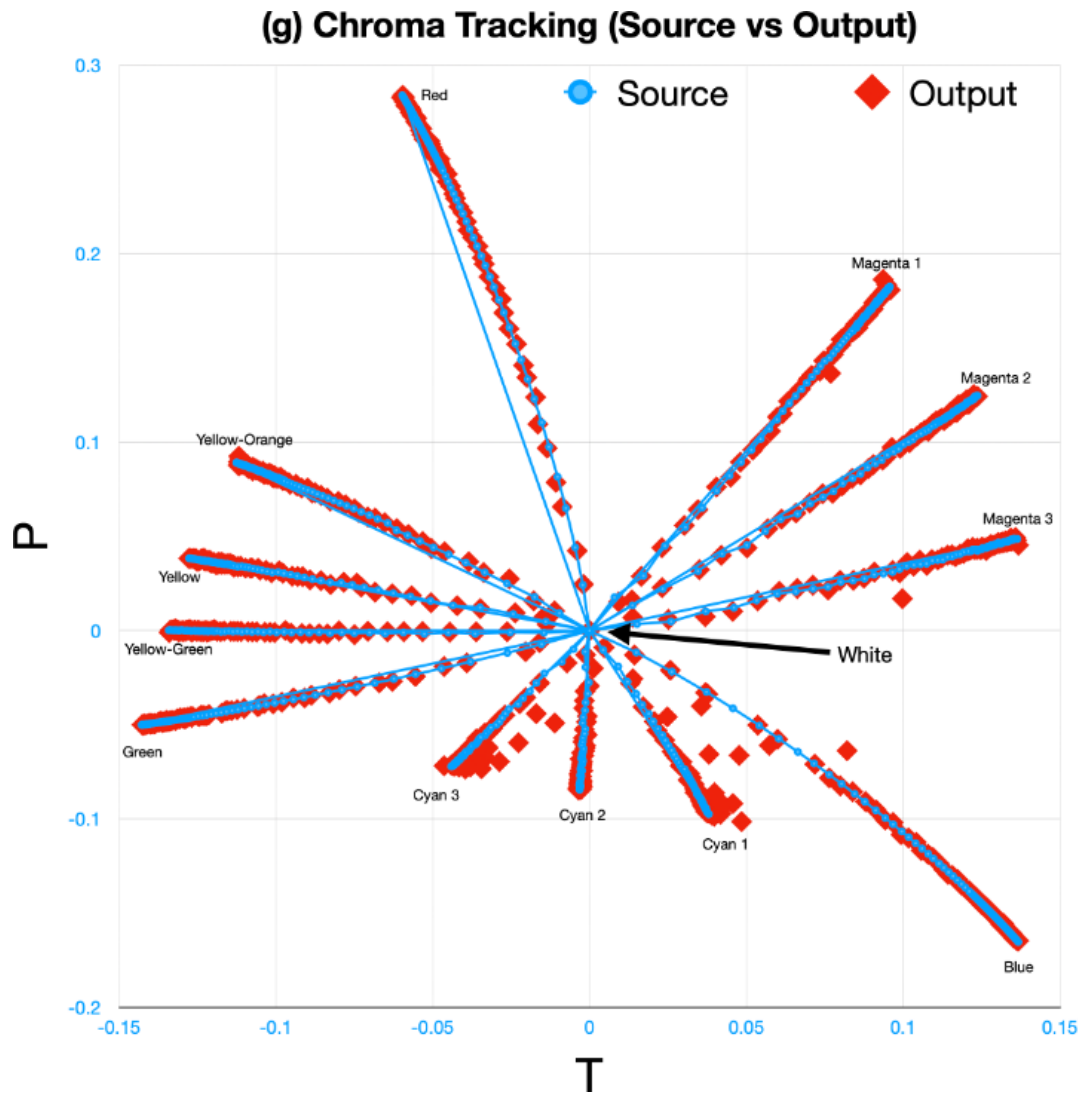


Figure 17. ITP Plots of pseudo BT.202 color space transforms using the Sarnoff “Yellow Brick Road” Chart

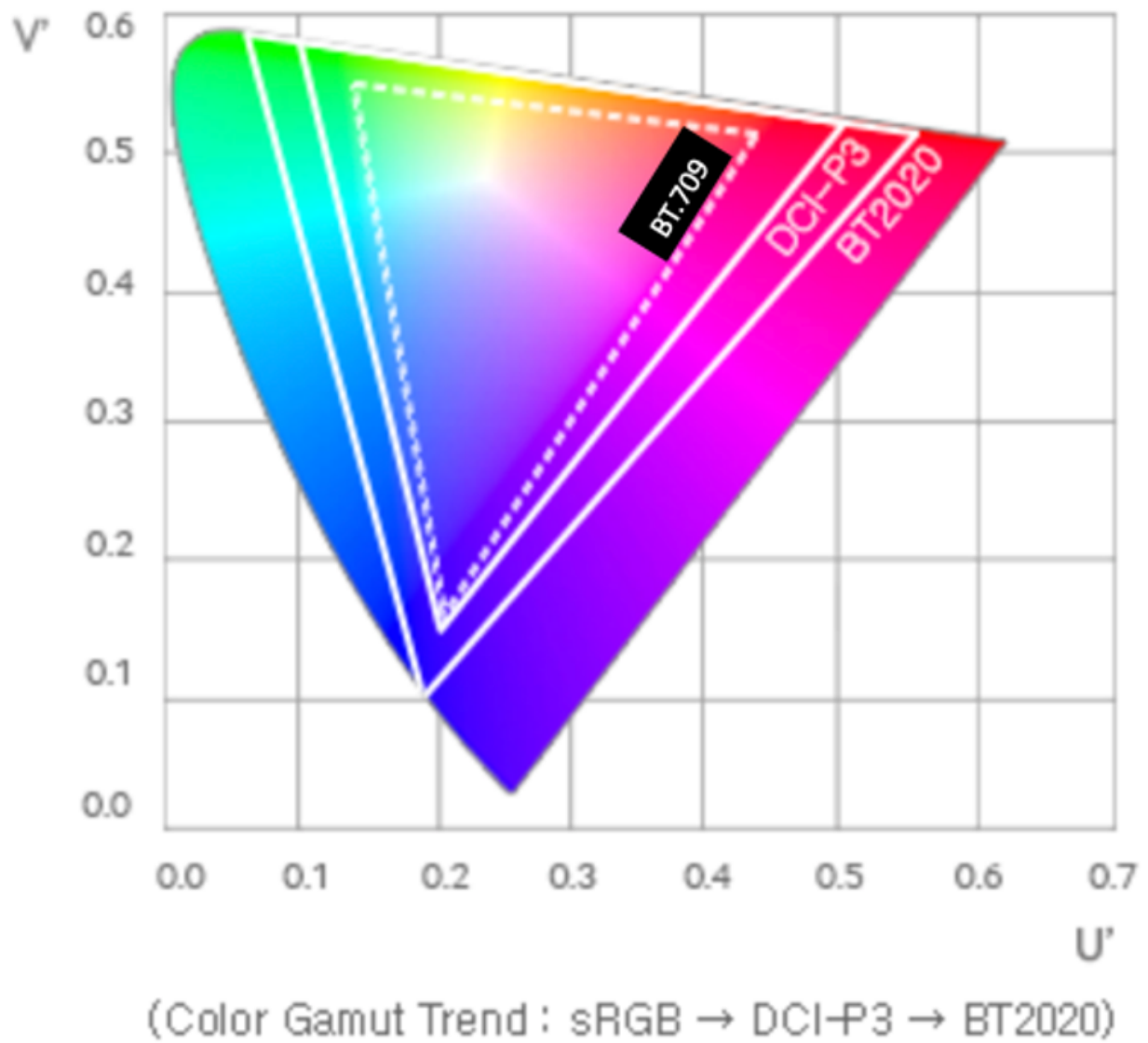


Figure 18. CIE 1979  $u'v'$  color spaces



## 9. Real-Time Program Service Assembly

Real-time Program Services consist of a linear, pre-scheduled stream of content that is assembled in real-time for distribution to consumers such as a broadcast television channel, a cable network, etc. Real-time Program Services consist of Live and/or Pre-recorded content and may also include graphic overlays, such as station logos, emergency text crawls, etc.

A Real-time Program Service may be assembled at one or more locations. At Production, graphics and slow-motion replays may be added to Live content. At the Broadcast Center (see [Green Book-Section 7](#), Figure 1 and Table 1) and service provider, interstitials, logos, news or emergency alert crawls, etc. may be combined to produce the final product delivered to the consumer. It is also possible that assembly can occur at the consumer device, such as for targeted advertisements and rendering closed captions.

### 9.1. Maintaining Dynamic Range and System Colorimetry Parameters

Different dynamic range and system colorimetry parameters should not be mixed in a Real-time Program Service. For example, service operators should not shift between HLG10, PQ10 or SDR/BT.709. Decoders require time to adjust to different encoded data settings – as much as 2 seconds or more has been observed by Ultra HD Forum members – causing a poor consumer experience. OTT providers offering ABR streams must also ensure that the adaptive bitrate streams are all of the same transfer curve and system colorimetry<sup>8</sup>. Similar to the linear progression of a program through time, the progression of rendering successive levels of ABR streams requires quick adjustments on the part of decoders. If a Real-time Program Service must contain a switch point between dynamic range and system colorimetry, it is recommended that such switches be performed overnight or in a maintenance window and black frames be inserted at switch points to allow decoders and viewers to adjust to the new content characteristics.

It is possible to direct map SDR/BT.709 content into HLG10 or PQ10, to up-map SDR/BT.709 content to HLG10 or PQ10 and vice versa, and to convert content from PQ10 to HLG10 or vice versa (see [ITU-R report BT.2408 \[8\]](#)). The following subsections offer guidelines for normalizing content in the headend for this purpose. See also [Section 11.3 in the Violet Book](#) for conversion possibilities in consumer equipment for backward compatibility.

---

<sup>8</sup> See Guidelines for Implementation: [DASH-IF Interoperability Points \[16\]](#) Section 6.2.5, Adaption Set Constraints.





## 9.2. Conversions from SDR/BT.709 to PQ10 and HLG10

In multiple areas of the production chain, there may exist a requirement to “mix” or “switch” SDR/BT.709 sources and PQ10/HLG10 sources when constructing a Real-time Program Service. Mixing of this kind may take many forms as it does today in SDR only environments (using [BT.1886 \[4\]](#) gamma and [BT.709 \[2\]](#) system colorimetry). To make it possible to combine such elements, SDR/BT.709 sources must be converted into the PQ10/HLG10 domain with respect to both an industry compliant transfer function (e.g., HLG or PQ per [BT.2100 \[5\]](#)) and system colorimetry (i.e., BT.709 primaries “mapped” into the [BT.2020 \[3\]](#) container). Such conversions can utilize:

- Direct Mapping: SDR/BT.709 content is decoded and repackaged as PQ10 or HLG10 containers, but while changing the system colorimetry, remapping does not change the color gamut or the dynamic range of the content; the content is simply mapped across to the equivalent color and brightness values.
  - When remapping SDR to HDR for PQ and HLG, the level of reference white should be considered. Reference white in SDR content is typically about 642.5mV (92% of SDR Peak white level. 700mV) leaving very little headroom for speculars. In the case of HLG, the BBC and NHK recommend the reference level for HDR graphics white (aka “reference white”) be set to 75% (equivalent to 203 cd/m<sup>2</sup> on a 1,000 cd/m<sup>2</sup> reference display, or 343 cd/m<sup>2</sup> on a 2000 cd/m<sup>2</sup> reference display). This was chosen as it leaves sufficient headroom for “specular highlights” and allows comfortable viewing when HLG content is shown on HDR/WCG and SDR/WCG displays.
- Up-Mapping: SDR/BT.709 is decoded and then enhanced/modified to emulate PQ10/HLG10 and repackaged as above. Blind (not supervised) up-mapping can lead to undesirable results depending on the technology so care should be used when converting SDR to HDR. Conversion algorithms or equipment should be carefully evaluated prior to their use. [ITU-R BT.2446 \[128\]](#) describes two techniques for SDR to HDR conversion.

Each method has particular advantages and disadvantages, and one or the other may be preferable under different circumstances. If the service provider intends to simulcast the Foundation Ultra HD service in SDR/BT.709 for backward compatibility, then remapping may be preferable, because content that was originally SDR/BT.709 will remain exactly as intended in the legacy version of the service. Conversely, the service provider may prefer that all the segments of the Real-time Program Service look as uniform as possible to ensure a consistent consumer experience, and thus up-mapping may be appropriate. For example, up-mapping may



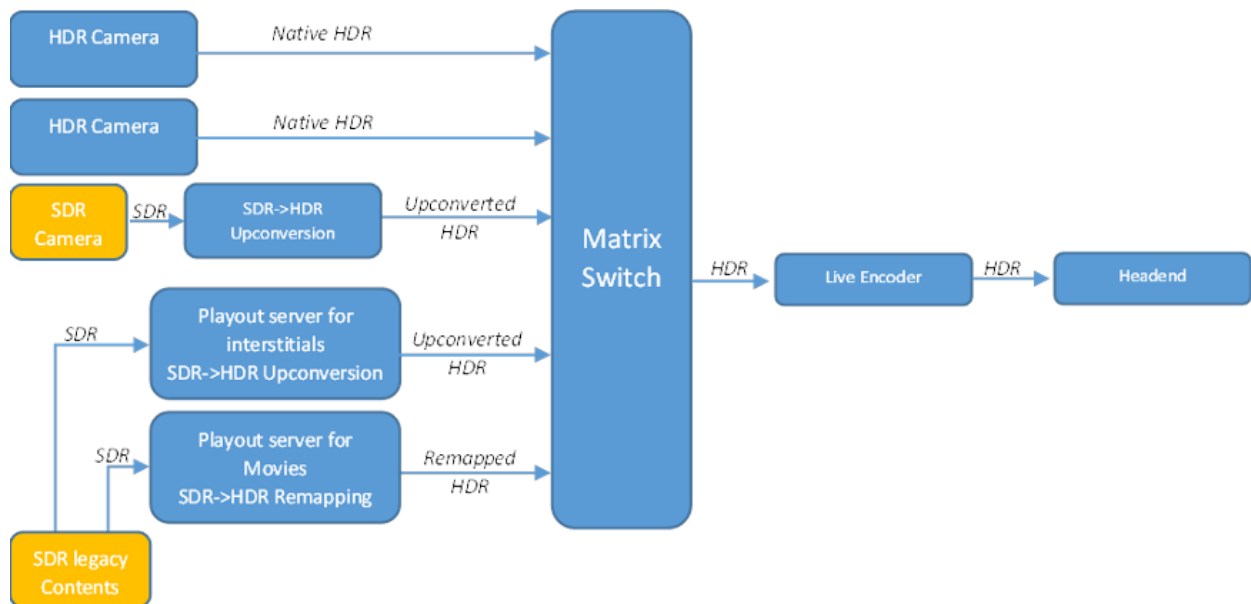
be preferred for Live production mixing SDR/BT.709 and PQ10/HLG10 cameras, high-quality SDR content, and legacy content like advertisements.

Under some circumstances, (e.g. viewing in a darkened room) HDR displays and content can cause discomfort if consideration is not given to the viewer's vision adapting to the average light level on the screen at any given moment. For instance, if a feature program has a low average light level such as a night scene and the picture is abruptly cut to a high average luminance scene in an interstitial, the viewer may experience discomfort similar to that experienced with changing light conditions in nature. When advertisements are inserted into content, consideration should be given with respect to transitions from dark to light.

The described SDR to HDR conversions typically will be performed by dedicated devices using a combination of 1D and 3D LUTs or other appropriate algorithms or technology. Devices of this type may be used for both SDR/BT.709 to PQ10/HLG10 or vice versa as the production and the capability of the equipment in use requires.

A real-time dedicated conversion device is essential for some use cases, which may be encountered throughout the production chain, such as:

- Mix of SDR/BT.709 and PQ10/HLG10 live sources
- Broadcasting live events (typically sports) in PQ10/HLG10 may require a relatively high number of cameras and it is probable that all these cameras will not be HDR-capable. In that situation, SDR/BT.709 cameras can be utilized if the conversion process is implemented either at the output of the SDR/BT.709 camera or at the input of the mixer.
- SDR/BT.709 interstitials in a PQ10/HLG10 program
- With SDR/BT.709 interstitials, the interstitial content will likely come from a playout server. In this case the conversion process has to be implemented either at ingest, at the output of the playout server, or at the input of the mixer.
- Use of SDR/BT.709 content
- Extensive libraries of SDR/BT.709 content may be used, for example a live sports match production that includes archive footage from previous matches; such material needs to be converted to PQ10/HLG10 (using a near real-time file-to-file converter) before entering the video pipeline



**Figure 19. Sample Live Workflow with Mixed Format Source Content**

Converting content to a common set of HDR and WCG technologies may occur at different points in the workflow from Production to the Consumer premises.

At the Production stage, multiple versions of the content may be produced. In the case of Pre-recorded content, producers have the opportunity to create multiple versions of the content applying creative judgment to each version. This is recommended for all Pre-recorded content and especially for interstitial material that may be inserted into a variety of Real-time Program Services, which may be operating with different HDR/WCG configurations. Live content may also be output in multiple formats; however, time constraints may prevent highly detailed artistic input. Automated conversion technologies can be “tuned” by the content creator to make the conversion the best it can be for the given program.

At the Broadcast Center or service provider stage, content may be converted to the provider’s chosen HDR/WCG configuration using automated tools. This may not include program-specific creative input; however, professional equipment may produce results that the service provider deems acceptable.

At the consumer premises, color volume transform may be possible for the purpose of backward compatibility (with limitations, see [Sections 10](#) and [10.4](#) of the Violet Book). Also, display devices may “map” content internally to best suit the display characteristics. Both of those



processes operate on a content stream with one, constant dynamic range and system colorimetry. Real-time Program Service providers should not expect consumer equipment to render seamless transitions between segments of content that have different transfer function or system colorimetry.

### 9.3. Conversion from PQ10/HLG10 to SDR BT.709

This operation is needed for PQ10 and HLG10. This method may be employed at the headend prior to final distribution to provide a backward compatible feed to legacy 2160p/SDR/[BT.709 \[2\]](#) TVs<sup>9</sup> and legacy HD networks. (See also [Section 11 of the Violet Book](#) on Format Interoperability including STB conversions.)

It is possible to do this conversion using a 3D LUT mechanism as described in Section 12.2. Another method is an invertible down-mapping process described in [ETSI TS 103 433 \[33\]](#), in which HDR/WCG content is down-mapped in real time to SDR/BT.709 at or prior to the emission encoder. As a commercially deployed system in 2016, this down-mapping process is considered to be a Foundation Ultra HD technology.

### 9.4. Avoiding Image Retention on Professional and Consumer Displays

#### 9.4.1. Background

As per EBU recommendation [R 129 “Advice to Broadcasters on Avoiding Image Retention on TV Production Displays” \[89\]](#), to minimize the risk of static image retention or premature ageing of displays, broadcasters and other content providers should take note of the following.

Broadcasters encounter ‘static’ images in a number of situations, including on-screen channel identifications, interactive application flags, banner displays, screens displayed when radio services are being received, program guides as well as longer-term text inserts such as sports scores.

---

<sup>9</sup> HLG offers a degree of backward compatibility with SDR/WCG displays that support [BT. 2020 \[3\]](#) color spaces. HLG however, does not offer true backwards compatibility for [BT.709 \[2\]](#) displays without additional processing as described in this section. Both DVB and ARIB require support of BT.2020 system colorimetry in legacy 2160p/SDR TVs, so it may be reasonable to expect that many of these units are BT.2020-compatible, and thus able to render HLG content.



This type of content is likely to remain an editorial feature of many broadcasts. It should also be noted that if image retention issues occur in production or broadcaster environments, it can also occur on viewers' television screens.

### 9.4.2. Definition of Static Images

For the purpose of this document an image is deemed to be static if any part of the screen is occupied by any part of the image for more than a total of six hours in any 12-hour period on more than one occasion in a seven-day period.

If an image is not static the risk of a retained image being formed from it is low. To assist in ensuring that images are not static, certain specific practices might be considered, including:

- Moving the position of images on the screen from time to time in order that the definition of 'static' is not met.
- Instigating a time-out of static images where appropriate.

### 9.4.3. Recommendations

The luminance or fully saturated chrominance (in the case of high color static images) value of any static image should be restricted to a value equal to the average picture level of the screen in order to minimize the risk of forming a retained image.

Two alternative methods of achieving this are:

- To use a technique known as 'Linear-key mixing' that overlays the static image as a partly transparent image over the picture content. The 'added image volume' level that sets the apparent transparency should not be set any higher than a level necessary to make the added image acceptably visible.
- To limit the signal level of the static image to no more than:
  - 40% of peak white for standard dynamic range (SDR) static images
  - 47% of reference white<sup>1</sup> (i.e. 35% of peak white signal level) for Hybrid Log-Gamma (HLG) high dynamic range (HDR) static images
  - 37% of peak white for ST.2084 Perceptual Quantizer (PQ) high dynamic range (HDR) static images

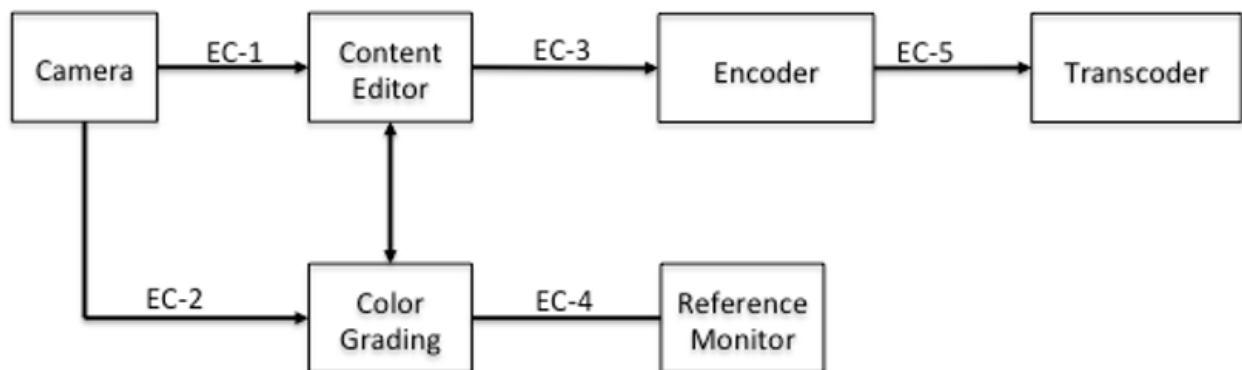
Further, it is recommended that the use of saturated color static images be avoided wherever possible and particularly where one is laid over the other.



## 10. Production for Pre-Recorded Content

This section focuses on the creation and distribution of pre-recorded content intended for inclusion in a Real-time Program Service. This includes all content filmed, captured digitally, or rendered in CGI and delivered via a file format to a service provider. Real-time Program Services may be delivered via MVPD (satellite, Cable, or IPTV), OTT and DTT. See Section 6.2.5 for production of Live content that is recorded for subsequent re-showing.

The following diagram depicts the interfaces and primary functions within the pre-recorded content creation process.



**Figure 20. Pre-recorded Content Production Workflow and Interfaces**

The scope of this guideline includes definition of six functions and five interfaces. The functional descriptions are described within sub-sections below, and the interface descriptions are described at a high level in the following [Table 6](#).

**Table 7. Pre-recorded Content Interface Descriptions**

Reference Point	Content Creation Functions	Reference Point Description
EC-1	Camera – Content Editor	Raw or log camera footage is ingested into the editing program. This interface may optionally contain metadata from the camera used by the editing program to provide guidance to timeline creation parameters.
EC-2	Camera – Color Grading	Raw or log camera footage is ingested into the color grading solution. This interface may optionally provide camera metadata describing the dynamic range resolution or other camera specific information from the captured footage.
EC-3	Content Editor – Encoding	Edited and color graded content is exported to an encoder for master and/or mezzanine level creation. This includes audio, video, essence, timing and CG content. Often the encoding function is included as part of the Content Editor.
EC-4	Color Grading – Professional Reference Monitor	This interface varies based on the content grading environment. It can include SDI, HDMI 2.0, Thunderbolt, DisplayPort and DVI.
EC-5	Encoder – Transcoder	This interface includes all aspects of a finished video delivered in file format for ingest to a transcoder. The transcoder uses this information to create distribution formats. If the transcoder is receiving uncompressed assets, it may also be an encoder.

## 10.1. Camera Requirements

### 10.1.1. High Dynamic Range and Wide Color Gamut Considerations

Material should be acquired at the highest resolution, dynamic range and color gamut of the camera in an operational data format best optimized for the image acquisition application in use.



Although the primary consideration of this document is Foundation Ultra HD, capturing in the native camera format allows for future content grades being used for distribution of additional Ultra HD technologies.

There are numerous production examples that require the use of the camera specific raw or log data format from the camera sensing device. This is due to the requirement to acquire the highest level of picture quality to sustain the levels of post-production image processing usually encountered in high-end movie or a 'made for TV' drama or documentary with use of color correction, visual effects and the expected high levels of overall image quality.

There are equally numerous applications for episodic and live capture that employ camera-specific logarithmic (log) curves designed to optimize the mapping of the dynamic range capability of the camera sensors to the 10-bit digital interface infrastructure of production and TV studio operations. Such log 10-bit signals are often of the 4:2:2 color-sampling form and are either generated internally in the camera head or locally created in the camera control unit. These images are usually recorded utilizing high-quality mastering codecs for further post-production of the captured scenes or, in the case of live transmissions, minimally processed for real-time transmission to viewers.

Camera native log curves are typically designed by camera manufacturers to match the performance characteristics of the imaging sensor, such as sensitivity, noise, native dynamic range, color spectral characteristics, and response to extreme illumination artifacts. It should be noted that data describing the gamut and transfer function characteristics encoded at the camera must be passed down the production chain, particularly in cases where not all cameras used are operated with the same parameters. In post-produced content, such information is typically carried in file metadata, but such metadata is not embedded in the video signal.

### 10.1.2. Imaging Devices: Resolution, Dynamic Range and Spectral Characteristics

In the creation of Foundation Ultra HD content signals, ideally, the image sensing devices should have a sensor resolution and dynamic range equal to or greater than a pixel count commensurate to the signal format.

In the area of spectral characteristics, the more advanced sensing devices will exhibit characteristics approaching the system colorimetry of [BT.2020 \[3\]](#), while more typical devices will produce color performance approximating the DCI-P3 gamut or just beyond the gamut of [BT.709 \[2\]](#).





---

Additional considerations for 2160p capture include:

- Not all lenses are suitable for capturing 2160p resolution and the MTF of the lens must be sufficient to allow 2160p capture on the sensor.
- Nyquist theory applies to camera sensors and care may be needed in selecting cameras that achieve target 2160p spatial resolution performance.
- When transmitting film content, the 16:9 aspect ratio of Foundation Ultra HD does not correspond to the wider aspect ratio of some movies. The two alternative methods of re-formatting (full width or full height) represent the same compromises that exist in HD transmission.
- Content originating from 35mm film will translate to Foundation Ultra HD differently than digital sources, e.g. film grain, achievable resolution.

## 10.2. Reference Monitor Considerations

Use of 2160p, HDR and WCG imply the need for reference monitors that will allow production staff to accurately adjust devices or modify content. Previously in SD and HD TV, there were accepted working practices using professional monitors and standard test signals, such as color bars, PLUGE, sweep, etc. Digital Cinema employs slightly different techniques, using calibrated monitors and LUTs to replicate the viewing characteristics of different consumer viewing conditions.

The recommendation for Foundation Ultra HD is to agree on practical standardized methods for use of Reference Monitors for Real-time Program Service production.

Note that Live content, such as sports events using trucks and Live production switching, do not have a post-production stage; operation is fully real-time. In these environments, unlike post-production suites, there is a limited amount of opportunity for human intervention. Any human intervention happens in real-time by direct interaction with the acquisition device.

For Foundation Ultra HD, a reference monitor can ideally render at least the following: resolutions up to 3840x2160, frame rates up to 60p, [BT.2020 \[3\]](#) system colorimetry (ideally at least the P3 gamut), and HDR (i.e., greater than or equal to the contrast ratio that could be derived from 13 f-stops of dynamic range). It should be noted that as with HD, consumer display technology is likely to progress beyond the capabilities of current generation professional displays. As such, instruments such as waveform monitors and histogram displays are essential tools to ensure optimum Foundation Ultra HD delivery.



## 10.3. On-Set/ Near-Set Monitoring

Viewing conditions for HDR monitoring:

- While it is difficult to establish parameters for viewing conditions for on-set monitoring, it is advisable to follow the recommendations for setup of an on-set monitor as described in [BT.814 \[14\]](#) or the latest version of an equivalent standard. [BT.2100 \[5\]](#) contains some specifications on reference viewing conditions (e.g. 5 nit surround).

Dynamic range of on-set monitor:

- It is recommended to have display devices for on-set monitoring capable of at least 800 cd/m<sup>2</sup> of peak brightness. Some RGB OLED mastering monitors are capable of 1,000 cd/m<sup>2</sup> of peak brightness. Note that future HDR content delivered to the consumer may be intended for substantially greater than 1,000 cd/m<sup>2</sup> peak display brightness.

## 10.4. Color Grading

### 10.4.1. Grading Workflow

Professional color grading should take place in a controlled environment on a professional monitor whose capability is known, stable and can be used to define the parameters of the ST 2086 [10] Mastering Display Color Volume Metadata.

Current industry workflows such as [ACES \[50\]](#) are recommended to be used where the operator grades behind a rendering and HDR viewing transform, viewing the content much as a consumer would. The work would result in a graded master that, when rendered with the same transformation, will result in a deliverable distribution master. (See also [Section 9.1 of the Indigo Book on ACES](#).)

### 10.4.2. Grading Room Configuration

When there is a need to prepare both HDR and SDR video productions, which share the same physical environment and require mixing segments of different dynamic range characteristics in the same master, it is important to ensure the use of equivalent illumination levels encountered in a conventional grading environment. This is because it is important to review both the SDR and HDR rendering of the images to guarantee a level of consistency between them. The black



and dark tones for both the HDR and SDR video pictures are a particular concern as the ability to discriminate the information they contain is highly affected by viewing conditions.

Secondary monitors should be turned off so as to not impact room brightness during grading or the client viewing monitor must be positioned so as to not impact the bias lighting observed by the colorist.

### 10.4.3. Grading Monitor Configuration

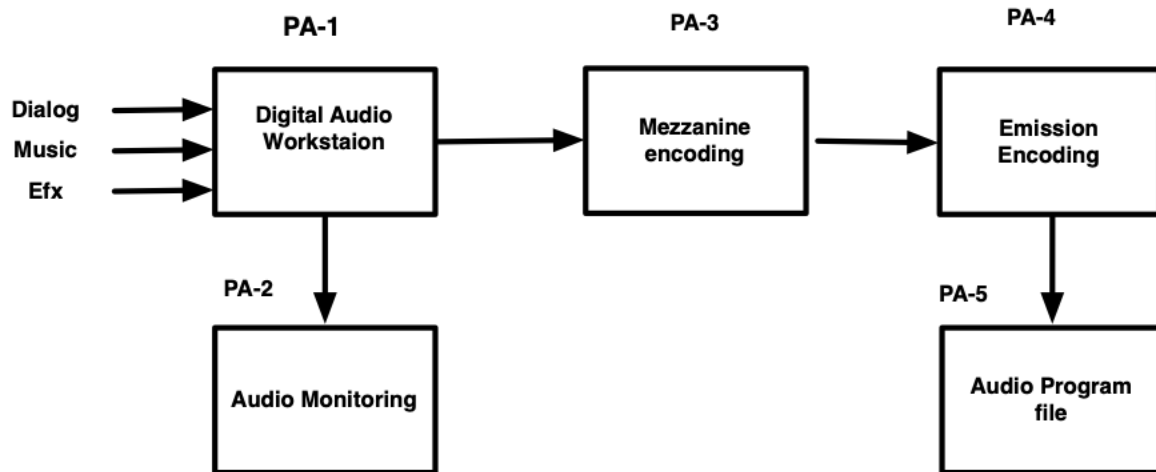
A professional mastering monitor should be specified and configured according to the required deliverable: System Colorimetry, Transfer Matrix, EOTF, White Point and RGB level range (Full or Narrow).

### 10.4.4. Grading System Configuration

As of the publication of this book, no de facto standards have emerged which define best practices in configuring color grading systems for 2160p/HDR/WCG grading and rendering. A full discussion of this topic is beyond the scope of this document and it is recommended that the reader consult with the manufacturer of their specific tool. [Section 13 of the Violet Book](#), Real World Foundation Ultra HD Deployments may provide some insight into this topic.

## 10.5. Channel-based Immersive Audio Post Production

This section describes one example of creating content with channel-based Immersive Audio containing height-based sound elements. This example is for post-produced content (file-based workflow) which can then be used in a real-time program assembly system and distributed in a linear television distribution system. One commercially deployed Immersive Audio system is Dolby Atmos using [E-AC-3+JOC \[35\]](#). A high-level diagram of a post-production mixing environment for this format is shown below:



**Figure 21. Channel-based Immersive Audio Post-Production**

**Table 8. Channel Based immersive Audio Workflow Elements**

Reference Point	Content Creation Functions	Reference Point Description
PA-1	Digital Audio Workstation	A mixer uses a digital audio workstation to create groups of audio tracks consisting dialog, music and effects. Each group consists of channel-based audio beds (e.g. 5.1, 7.1.4).
PA-2	Audio Monitoring	The Immersive Audio mix is rendered to multiple different speaker layouts including 7.1.4 and 5.1 using hardware film-mastering style tools or commercially available software tools
PA-3	Mezzanine Encoding	Immersive Audio and 5.1 channel-based renders are encapsulated into a mezzanine file.
PA-4	Emission Encoding	Immersive Audio is encoded into formats such as E-AC-3+JOC <sup>10</sup> for final emission delivery

<sup>10</sup> See [Violet Book, Section 11.5](#) regarding backward compatibility of E-AC3 w/JoC.



---

PA-5	Audio Program File	<a href="#">SMPTE ST 337 [36]</a> embedded in a delivery file.
------	--------------------	--

### 10.5.1. Immersive Audio Monitoring Environment

Immersive Audio monitoring should take place in a controlled environment consisting of a treated listening room and a calibrated speaker environment. The recommended speaker configuration for immersive mixing and monitoring is 7.1.4 as per [Recommendation ITU-R BS.2051<sup>11</sup>\[129\]](#) (System G minus the Left and Right screen loudspeakers) calibrated to a reference sound pressure level appropriate for the size of the mix room. Typical levels range from 76 dB SPL C-weighted, slow response for a room less than 1,500 cubic feet or smaller (i.e., many outside broadcast trucks) to 85 dB SPL, C-weighted, slow response for rooms 20,000 cubic feet or larger (i.e., many film stages).

### 10.5.2. Immersive Audio Mastering

The channel audio data is recorded during the mastering session to create a 'print master' which includes an immersive audio mix and any other rendered deliverables (such as a 5.1 surround mix). Program loudness is measured and corrected using [ITU-R BS.1770-4 \[37\]](#) methodology.

---

<sup>11</sup> System G minus the Left Screen and Right Screen loudspeakers would be used for 7.1.4, while System D could be chosen for 5.1.4.



## 11. Production for Live Content

The Ultra HD Forum's objectives include understanding the implications of creating and delivering Foundation Ultra HD content all the way through the production chain. It is important that the whole system is understood as technical decisions to deploy a particular Ultra HD production approach upstream may have implications for downstream delivery equipment. The reverse is also true that downstream delivery constraints may affect production techniques.

Live content examples include sports, news, award shows, etc. Pre-recorded content is captured and produced in advance (see [Section 10](#)). Examples include soap operas, sitcoms, drama series, etc. Real-time Program Services may include Live programs and Pre-recorded programs, and Pre-recorded content – such as advertising – may be inserted within Live programs. Real-time Program Services may be delivered via MVPD (Satellite, Cable, or IPTV), OTT and DTT and are delivered on a schedule determined by the service provider.

Unlike Cinema, Ultra HD Blu-ray™ discs or On Demand, Real-time Program Services involve performing complex manipulation of images and audio in real-time at multiple points in the delivery chain, including graphics, captions, virtual studios, Live sound mixing, logos, chroma-keying, DVE moves, transitions between sources, encoding, decoding, transcoding, ad insertion, voice over production, monitoring, conditioning, rendering, closed captioning, standards conversion, etc.

### 11.1. Live Production in Trucks or Studio Galleries

At the live event venue or studio, the action is captured and prepared in near real-time. This section addresses the processes that may be required to assemble the live show prior to compression for distribution to a “headend” or other central facility. Such video processes may include:

- Synchronous live baseband inputs from cameras to production switcher
- Racking of cameras (for chroma/luma/black balance)
- Transitions between cameras (mix, wipe, DVE, etc.)
- Keying (chroma key, linear key, alpha channel, difference key, etc.)
- Overlay of graphics, text, etc.
- Slow motion and freeze frame
- Editing (for action replay/highlights)
- Use of virtual sets



Simultaneous production of both 2160p/HDR and HD/SDR output may be employed, and assets may be shared, e.g., HD graphics or 2160p graphics for both outputs. It is recommended to maintain a single HDR transfer function and system colorimetry. See [Section 8.2](#) and [Section 8.3](#) for details on converting assets from SDR to HDR or vice versa.

Performing the above functions on Foundation Ultra HD content may be different than for HD/SDR content. Foundation Ultra HD content may involve higher spatial resolution, frame rates up to 60p, and HDR/WCG.

- Graphics created for gamma transfer function may need re-mapping for an HDR transfer function depending on creative intent.
- HLG has the characteristic of providing a degree of backward compatibility with legacy SDR devices, both professional and consumer. However, care must be taken to ensure the signal being displayed is encoded with the appropriate system colorimetry preset in the display. For instance, for HLG10, a displayed picture on a legacy display that uses [BT.1886 \[4\]](#) / [BT.709 \[2\]](#) system colorimetry will be incorrect. See [Section 8 of the Yellow Book](#) for details and caveats.

## 11.2. HDR Production with encodings other than PQ and HLG

For episodic television, as with features, colorists typically work in an underlying grading space adopted by the project. Usually, this grading space comprises a log encoding, often with a color gamut native to a principle camera. The content master is created and (ideally) preserved in this project grading space. Conversion to a distribution encoding, e.g., PQ, is performed by an output viewing transform and viewed on a specific mastering display during grading, whereby the grade targets a particular peak luminance.

Similarly, live productions establish a project grading space, often choosing one that is optimized to their cameras and other equipment (e.g., SLog3 with BT.2020 for Sony equipment). Grading and mixing occurs within that selected space.

When contributions having different transfer functions and/or color gamuts are to be combined, there is a concern about cumulative quantization effects. One way to address this is to defer conversion away from the native system colorimetry as long as possible. A more general solution is to transform the various contributions to a common linear encoding as a grading space. In either case, the produced result is converted to the required delivery format, suffering the quantization of conversion only once.



---

In either case, a PQ master can be obtained with a corresponding output viewing transformation. Likewise, to obtain an HLG master, an appropriate output viewing transformation is used. Then the PQ master targets 1000 cd/m<sup>2</sup> and the HLG master is shown on a 1000 cd/m<sup>2</sup> nominal peak luminance display, they are typically expected to visually match each other. ITU-R report BT.2390 [76], Sec 7.3, specifies 1000 cd/m<sup>2</sup> as the reference condition to insure matching between PQ and HLG mastered content. (Note: see [Section 8.1](#) for more information on conversion between PQ and HLG content)

Other workflows may grade in a distribution space, e.g., PQ or HLG, which provides one distribution master directly. Other distribution masters are obtained by performing a conversion to the target space, however such conversions are sometimes difficult to match well. Further, some distribution spaces impose limitations that compromise whether a content archive is future proof.

### 11.3. Channel-based Immersive Audio Production

For a live event, Immersive Audio can be created using existing mixing consoles and microphones used during the event. Using additional console busses, height-based ambience and effects can be added to a traditional 5.1 or 7.1 channel mix which can then be encoded by an ETSI TS 103 420 compliant E-AC-3 encoder. A local confidence decoder can be employed to check typical down mixes, including the backwards compatible 5.1 channel render described in ETSI TS 103 420. During normal mixing, this confidence decoder can serve as a useful continuity check and display (i.e. to make sure the mix is still “on-air”), though due to normal latencies it will likely be found to be impractical to be kept in the monitor path full time. A high-level diagram of a live mixing environment recently used at a major televised event using Dolby Atmos using [E-AC-3+JOC\[35\]](#) is shown in [Figure 22](#) below.



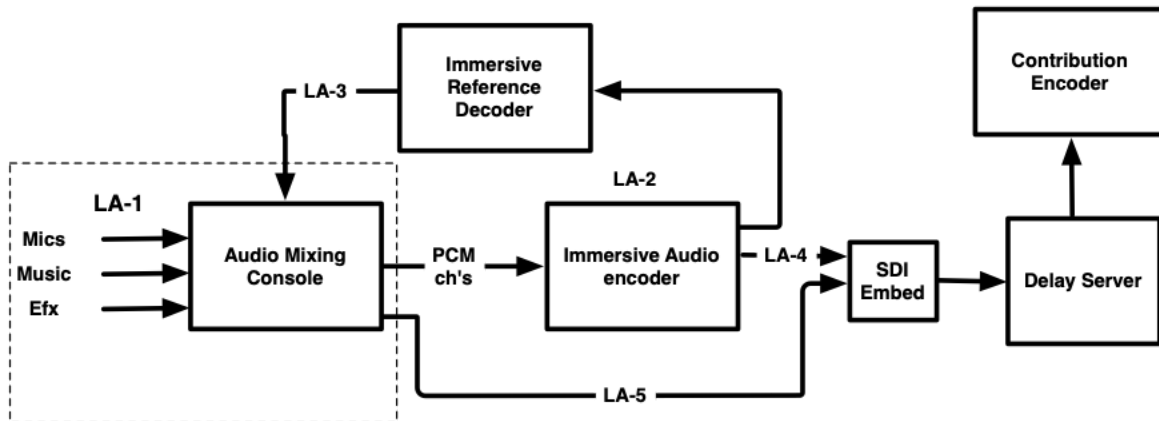


Figure 22. Channel-based Immersive Audio Live Production



Table 9. Description of Live Channel Based Audio Workflow Elements

Reference Point	Content Creation Functions	Reference Point Description
LA-1	Audio Capture and mixing	Microphones are placed throughout the live venue. Microphone feeds are brought into the audio mixing console in the same fashion as 5.1 production. Sound elements are mixed into a final immersive program.
LA-2	Immersive Audio Authoring	Audio groups are created in the mixing console representing the immersive mix as channel-based (e.g. 5.1.2, 5.1.4, 7.1.4) audio <sup>12</sup> .
LA-3	Immersive Audio Monitoring	The Next-Generation Audio Processor renders the audio to 5.1.2, 5.1.4 or 7.1.4 speaker feeds and sends these feeds back to the audio mixing console for monitoring in the live mix environment.
LA-4	Immersive Audio Encoding	Atmos immersive program, encoded as <a href="#">E-AC-3+JOC [35]</a> is delivered as a 5.1 channel bitstream + parametric side-data (steering data) to the contribution encoder <sup>13</sup> transported over either a MADI or SDI link.
LA-5	Legacy Audio Delivery	Stereo or 5.1 complete mixes may be created at the audio mixing console and delivered via traditional means.

<sup>12</sup> The downstream Atmos Channel-Based Immersive emissions encoder, using E-AC-3 + JOC [\[35\]](#) will render a legacy 5.1 audio program. It is recommended to verify the rendered 5.1 audio program using a suitable E-AC-3 decoder in the monitor chain.

<sup>13</sup> See [Section 11.5 of the Violet Book](#) regarding backward compatibility.



In this channel-based immersive audio example, a Dolby Atmos enabled E-AC-3+JOC encoder generates a compressed bitstream containing 5.1 channels of audio that is backwards compatible with legacy E-AC-3 decoders. In parallel, the encoder generates an additional set of parameters ( as specified in [ETSI TS 103 420 \[35\]](#).) that are carried in the bitstream, along with the 5.1 audio, for use by a Dolby Atmos E-AC-3+JOC decoder. The full Atmos decode process reconstructs the original channel-based immersive audio source from the 5.1 backwards compatible representation and the additional parameters. A typical channel-based immersive program can be encoded, for example, at 384-640 kbps total, thus fitting into existing emission scenarios.

For discussion of Next Generation Audio for live production see [Section 12.2.2](#).



## 12. Additional Ultra HD Technologies beyond Foundation Ultra HD

Methods and challenges for adding non-Foundation Ultra HD technologies into real-time program assembly vary from technology to technology. Further, an enhancement Ultra HD technology might be continually incorporated into the program stream (i.e., part of the “house format”) or could be incorporated intermittently. Ultra HD Forum will continue to study these cases and intends to provide further information in future versions of these Guidelines.

### 12.1. Dynamic HDR Metadata

For PQ HDR content<sup>14</sup>, as described in [Section 7.2.5 of the Orange Book](#), HDR10 provides the static metadata elements in a PQ10-based HDR format as specified by SMPTE ST 2086, MaxFALL, and MaxCLL. That section identifies a number of limitations with these particular HDR metadata values, notably the difficulty with setting these values in a live environment, real-world experience suggesting that these values have been set to artificial numbers to force certain looks on consumer displays, and the inability to correctly set these values given limitations of mastering displays.

In addition to these limitations, the values of MaxFALL and MaxCLL are also very limited in that they are only currently specified to provide single values for the entirety of the program. The dynamic range of both narrative and live content can vary dramatically from scene to scene. As a result the static, program-wide metadata values, as strictly defined, are of limited use for a great deal of content that does not have a static, unchanging dynamic range. Interoperability tests show that receivers can recognize changes in the static metadata within the duration of a program; however, it is yet unknown how frequently or quickly such changes can be recognized. For example, it is not expected that static metadata would change on a frame-by-frame basis.

---

<sup>14</sup> HLG10 does not specify any display metadata as it is based on normalized scene-light, rather than the absolute luminance of the signal seen on the mastering display, as described in [Section 7.2.2 of the Orange Book](#). As such, the headroom (measured in f-stops) for HLG highlights above HDR Reference White, is approximately constant regardless of the display’s nominal peak luminance. Moreover the HLG10 display EOTF, which is fully specified by the ITU-R BT.2100 [5], includes a variable display gamma to provide adjustment for a specific display’s peak brightness capabilities, along with eye adaptation; thereby allowing HLG to function in brighter viewing environments. Thus static or dynamic metadata is not required for HDR productions using HLG10.



Finally, there is no standardized way of utilizing these values in the final consumer display, so displays differ significantly in reproduction of the image. In practice some displays may ignore the values altogether. This is not consistent with the goal of displaying the image as close to the creative intent as possible on the target display.

A number of Dynamic Metadata methodologies have been developed to address the limitations of PQ10 and HDR10. Dynamic Metadata refers to metadata that describes the image at a much finer temporal granularity, scene-by-scene or even frame-by-frame and produces significantly more information about the mastering and creative intent of the scenes. In addition, most of these methodologies provide detailed information about tone mapping in the consumer display with the goal of consistent images across different manufacturers' displays. The methodologies also are designed to preserve creative intent, with the final displayed image being as close to the mastered image as the consumer display has the ability to reproduce.

Some of these methodologies go further by capturing the metadata during the color grading session and passing that metadata to consumer displays to better reproduce the creative intent. Most metadata schemes also provide for automatic metadata creation, which is useful in workflows for live content.

In general, several of these dynamic metadata schemes are additive, in that they provide additional information about the carried PQ10 image, and the HDR10 static metadata remains intact alongside the dynamic metadata. In some cases, this can provide a simple backwards compatibility to an HDR10-only display - the dynamic metadata is simply ignored.

Finally, many of these methodologies have considered how the signal can be backwards compatible with SDR displays and have built-in methods for conversion. See Dolby Vision™ described in [Section 7.1](#) and SL-HDR2 described in [Section 7.4](#) of the Indigo Book.

SL-HDR1 is another HDR dynamic metadata technology, which serves a different purpose than Dolby Vision or SL-HDR2. SL-HDR1 is intended to enable the service provider to emit an HDR/2020 service in an SDR/709 format that can be “reconstructed” to HDR/2020 by the receiver. HDR/2020 receivers that can interpret the SL-HDR1 metadata can present the HDR/2020 format to the viewer. The SDR/709 content can be displayed by receivers that cannot display HDR/2020. In this way SL-HDR1 provides a measure of backward compatibility for both HLG and PQ-based HDR content. It should be noted that SL-HDR1 requires 10-bit encoding, and so may not help address legacy SDR/709 receivers that are only capable of 8-bit decoding. See [Section 7.3 of the Indigo Book](#).



## 12.2. Ultra HD-Next Generation Audio (NGA)

Next Generation Audio (NGA) offers new tools for content creators and new experiences for consumers. Immersive audio offers a sense of being completely surrounded by the aural experience: above, below, and all around the listener. NGAs can not only make these experiences possible in a home theater environment with optimally placed speakers, but also via headphones, soundbars, or even sub-optimally placed speakers to an extent. NGA can also allow the consumer to personalize the experience, such as increasing the dialog level relative to the music and effects or choosing a particular dialog track such as alternate language or commentary. See [Section 7 of the Indigo Book](#) for details about Dolby AC-4, DTS-UHD and MPEG-H NGA systems.

### 12.2.1. NGA Post Production

Maybe the most important production aspect is that NGA systems bring new paradigms in authoring and mixing the audio content requiring additional training for sound engineers. For delivery of NGA immersive and personalized experiences, authoring of metadata plays an essential role in production. Using plugins for DAW (Digital Audio Workstations) in post-production or metadata authoring and monitoring units for live production, the metadata can be authored together with the audio data delivered to the final audio encoders. In post-production the metadata can be stored as an XML file according to the [ITU-R Audio Definition Model \(ADM\) \[72\]](#). It is anticipated that ADM profiles with tailored feature sets will be necessary at each stage of the NGA broadcast and content production chain to ensure full interoperability and quality of service.

### 12.2.2. NGA Live Production

A key consideration of implementing NGA for live production is the new requirements for dynamic metadata authoring and mixing the audio content requires additional training for sound engineers. For delivery of NGA immersive and personalized experiences, authoring and carriage of dynamic metadata plays an essential role in production.

Live productions using 3G or 12G SDI interfaces with a minimum of 16 PCM audio tracks are sufficient for delivery of immersive audio programs, with 4 height channels and several commentary tracks as objects. Dynamic audio objects can be added to the program mix, when the production tools have features such as dynamic spatial panning. Additional mechanisms for efficient and secured metadata delivered over an SDI interface are provided for NGA systems. According to SMPTE, this system will be described in [ST 2109 \[127\]](#) Audio Metadata over AES3



(pending publication by the SMPTE), and there is work to extend this to SMPTE ST 2110-40 [47] (metadata over IP).

From the many trials conducted in France, Germany and U.S. and the commercial broadcasts in South Korea using NGA, it has been found that integration of authoring and monitoring units in common SDI infrastructures (including remote facilities) is straightforward and sound engineers become familiar with mixing immersive content and authoring metadata relatively quickly. For example, Korean broadcaster SBS produced the 2018 Football World Cup using MPEG-H with immersive and interactive audio services (Korean commentary, English commentary, and stadium atmosphere alone).

Additional trials were conducted in 2018 using object-based audio and interactive audio services (including separate language tracks in some cases): by NBC during Winter Olympics using [E-AC-3 plus JOC \[35\]](#), by France Television during Roland Garros Tennis Open using both [MPEG-H \[70\]](#) and [AC-4 \[65\]](#), as well as by the EBU during the European Athletics Championships using both MPEG-H and AC-4. Since 2019, trial broadcasts of MPEG-H included: “Rock and Rio” (Brazil 2019), Eurovision Song Contest (Israel 2019), and the French Tennis Open (France 2019).

For the Tokyo 2020 Olympic broadcast, NBC/Comcast (US), Astro (Malaysia), beIN Sports (MENA) will broadcast audio using Dolby Atmos over AC-4. NHK (Japan) will deliver audio in 22.2 channels with AAC, and MBC (Korea) will broadcast the Olympics using MPEG-H audio.

### 12.2.3. NGA Production tools and workflows

#### 12.2.3.1. Introduction

To achieve a healthy and growing ecosystem with NGA content flowing from production to the end consumer, a few essential needs must be addressed in the content creation and interchange parts of the workflow. Top level needs are:

- Independence from the delivery codec solutions - ensuring open solutions from multiple vendors secures scaling of NGA content interchange and archival.
- Workflow flexibility - enabling several alternatives when transitioning from today’s workflow towards NGA enabled workflows.
- Interoperability - between traditional SDI based systems and new IP transport and cloud-based infrastructures.
- Graceful coexistence - with existing 2.0 and 5.1 content.
- Extensibility - can grow as use cases of NGA emission systems evolve.



To address these needs, several pieces must necessarily come together. A fundamental part of the solution is for the industry to collaborate around open standards that define the metadata along with the definition of how these are carried in different container formats. This is discussed further in the section on relevant standards.

Furthermore, several open-source projects are ensuring the availability of necessary NGA components to the different solution providers in the workflow, which combined with trials and interoperability testing events can facilitate both the broad adoption of NGA and interoperability between these different vendor realizations of the technology.

This section describes several “real world” use cases of NGA tools and workflows implemented and available for “real world” UHD productions.

The following information is arranged in two main sections:

- Dolby Workflows
- MPEG-H Workflows (Real-time Workflows & File-based Workflows)

### 12.2.3.2 Dolby Technology Workflows

#### **Case Study 1:** Static metadata over SDI infrastructure

Polish public broadcaster Telewizja Polska (TVP) decided to deliver Dolby Atmos via Dolby AC-4 next-generation audio for a temporary UHD channel for all Euro 2020 matches. NGA has been used in all the live transmissions as well as rebroadcasts of recorded material.

In this workflow, the existing 16-channel SDI embedded audio infrastructure was used in combination with static metadata to enable NGA content to be carried prior to final transmission encoding.

After conducting internal NGA tests using AC-4, including a successful DD+ Atmos transmissions on TVP1 HD in 2018, the Polish national broadcaster, TVP, decided to make NGA a part of their UHD transmissions on the TVP 4K channel. Their goal is to deliver special events with NGA experiences by using their standard contribution and 16-channel SDI embedded audio in-house infrastructures.



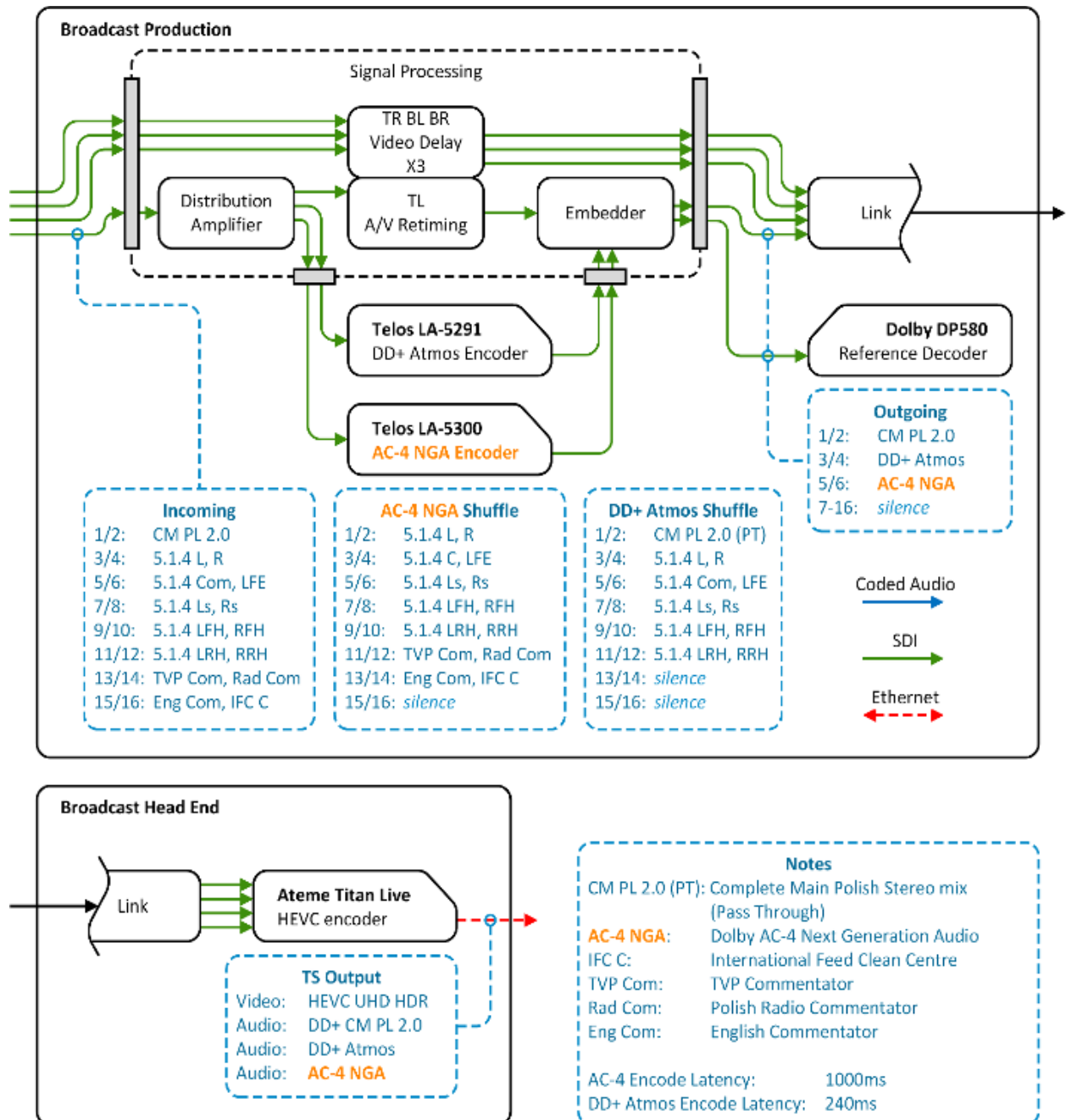


Figure 23. TVP 4k, Live Euro transmission with Atmos



Within the 16-channel SDI Infrastructure, TVP is able to carry a 2.0 Complete Main for legacy purposes as well as a 5.1.4 clean bed and three additional commentaries. The emission codec used takes care of downmixing to 5.1 upon playback if needed.

As all UHD TVs in Poland support DD+ and a large proportion also support AC-4, TVP decided to broadcast two Atmos streams simultaneously: DD+ Atmos with a TVP commentator and AC-4 Atmos with the possibility of personalizing the experience for TV viewers with TVs that support AC-4.

As highlighted in the [Figure 23](#) workflow diagram, only two audio encoders are used. In this case a Telos LA5291 is used for encoding an immersive soundtrack using E-AC-3-JOC, and a Telos LA5300 for AC-4 NGA encoding. Both encoders are equipped with an audio-shuffler on the input to handle audio routing if needed. Since this is a workflow without additional dynamic metadata flow, different encoding modes on both encoders can be triggered by GPI. The E-AC-3-JOC encoder will encode a 5.1.4 immersive Complete Main, while the NGA encoder will encode a 5.1.4 immersive bed with three additional commentary substreams.

The resulting experiences are:

- Stadium Ambience Only
- Stadium Ambience + Polish Commentary
- Stadium Ambience + Radio Commentary
- Stadium Ambience + English Commentary

All experiences enable the “Dialogue Enhancement” feature to allow end-users to adjust the dialogue level to their needs.

Once the audio is encoded and embedded into the SDI, the signal will be sent to the distribution hub. At the distribution hub, an ATEME Titan Live encoder is responsible for the video encoding. As the audio is already encoded, the video encoder only needs to pass-through the already encoded audio. The emission encoder feeds DVB T2 free to air transmissions with 49 transmitters and coverage of more than 90% of the Poland population.

According to Polish regulation, NGA capable AC-4 is mandatory for DVB-T2 UHD HEVC IDTV receivers. The receivers enable personalization of the sound reception including soundtrack selection, dialogue enhancement and mixing main with additional audio broadcast as audio objects. It seems that these features are very attractive for the viewers and may be implemented in the main HD services.

Poland’s national plan foresees a switch from DVB-T to DVB-T2 in 2022.



---

**Case Study 2: Dynamic metadata live event**

French national broadcaster FranceTV is preparing to launch a UHD service over DVB T2. A key aspect of the service is Next Generation Audio and the use of metadata in open standard distribution formats to author and control the NGA experiences.

In this workflow, the metadata is authored at the original production site and is carried through the workflow to configure the emission encoder at the platform operator's site. Metadata is also used to automatically reconfigure the emission audio encoder when playout switches to pre-recorded advertisements in stereo. The emission encoder feeds DVB T2 (free to air) transmissions.

[Figure 24](#) illustrates the workflow used in this case study. From the Roland Garros tennis arena, both 2.0 and 5.1 mixes are created along with English commentary and the Umpire's audio. In total 14 audio channels are then contributed to FranceTV as multiple audio streams all using ST 302 24 bit PCM audio.

At FranceTV the audio is brought into a ST 2110 network, where the 5.1 is up mixed to 5.1.4 using a Linear Acoustic UPMAX ISC, the French commentary is added when available, and the NGA metadata is authored as [S-ADM \[124\]](#) using PMD studio. All this comes together into an ATEME Titan Live Encoder which contributes 16 audio channels to Cognac-Jay in a transport stream as multiple audio streams all using ST 302 with 24-bit PCM audio and where channel 16 is used for the S-ADM metadata.

The two or three NGA experiences that are created are:

- Stadium Ambience Only
- Stadium Ambience + English Commentary
- Stadium Ambience + French Commentary (when available)

At Cognac-Jay the transport stream feeds the ATEME Titan emission encoder, although during commercial breaks the advert-streams, which also include S-ADM, are spliced in prior to the encoder. The emission encoder feeds both the DVB-T and DVB-S transmissions.

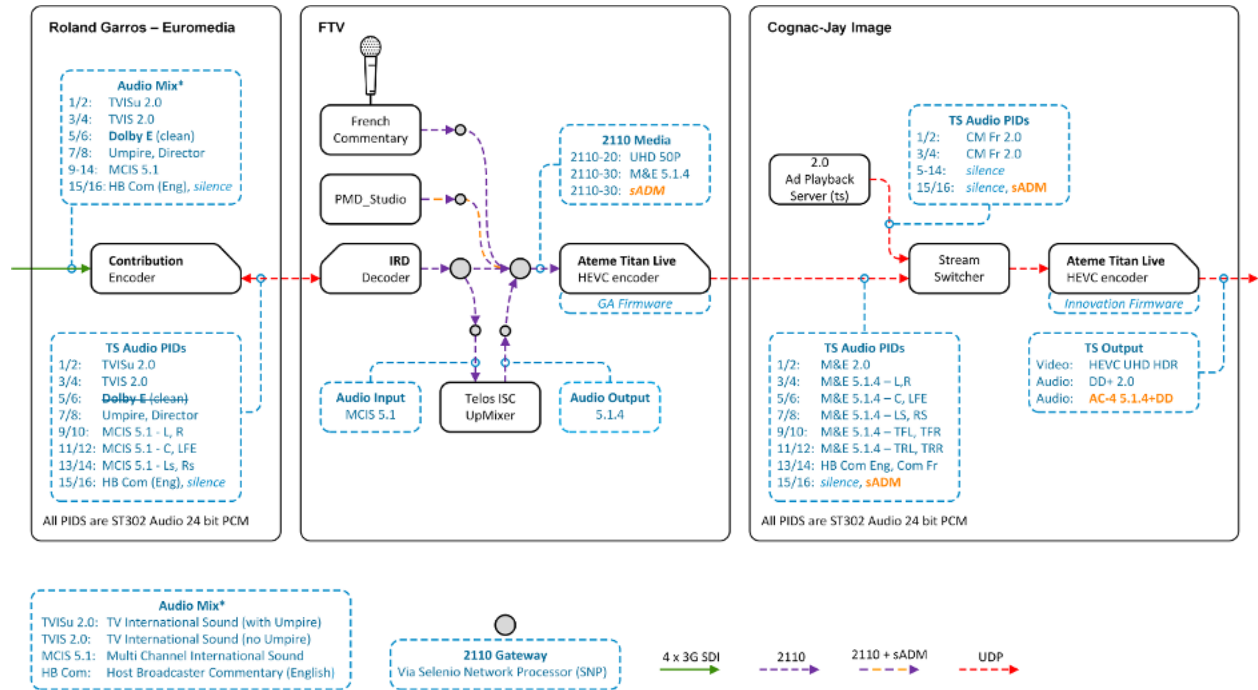
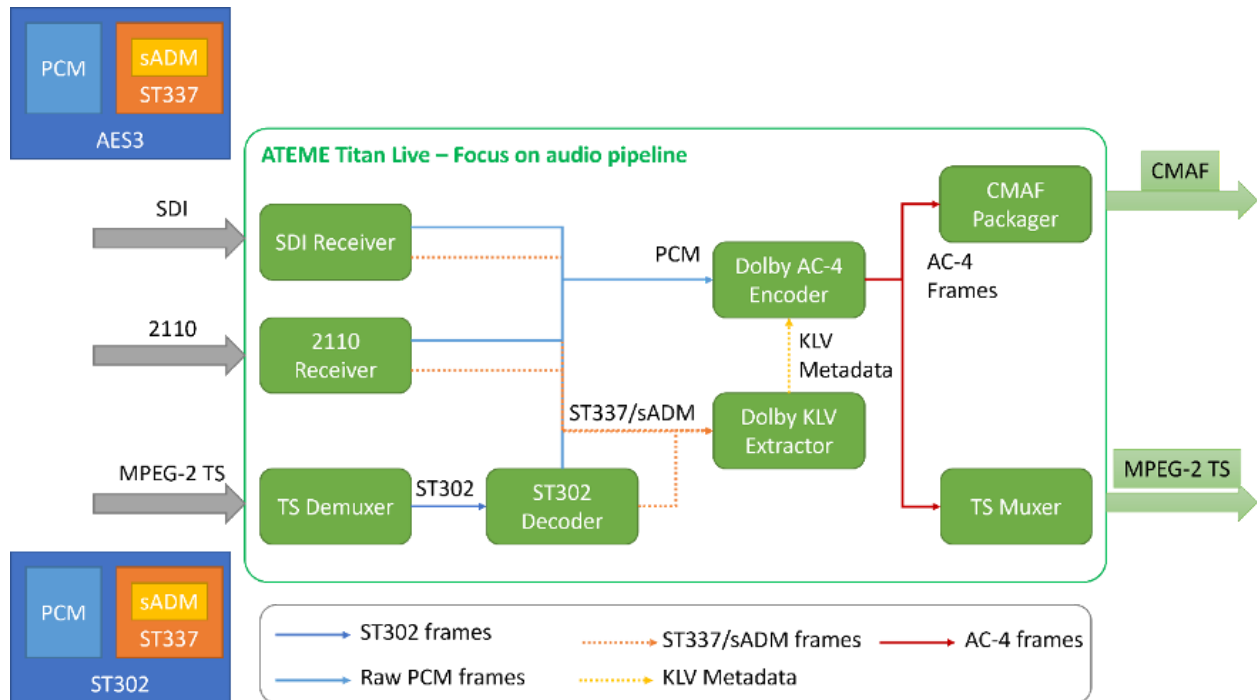


Figure 24. Dynamic Metadata Live Event example workflow



**Figure 25. Encoding pipeline using an ATEME Titan Live Encoder**

In [Figure 25](#) the audio encoding pipeline is illustrated. For this use case, the following inputs are used: MPEG 2 TS, SDI or SMPTE ST 2110. Independent of the input type, the first step is to extract the raw PCM and the S-ADM ST 337 frames. Those two types of data are then routed differently. Whilst the PCM frames are sent directly to the AC-4 encoder, the S-ADM frames need an additional processing step to extract the KLV metadata. The encoded AC-4 frames are forwarded to a CMAF packager and/or a TS muxer. These units will dynamically generate the signaling either using MPEG-DASH manifest declaration or MPEG-2 TS descriptors.

### Case Study 3: Dynamic Metadata in a SMPTE ST 2110 environment

In this workflow, dynamic metadata is used inside a SMPTE ST 2110 [\[43-47\]](#) environment to control NGA experiences for live sports.

At Swiss broadcaster Schweizer Radio und Fernsehen (SRF), the discussions concerning personalized sound experiences and the ongoing aim of optimizing the use of bandwidth found their solutions in the technologies associated with Object-Based Audio. Additionally, SRF stepped into the future of IP-based workflows in broadcast by developing Europe's first all-IP UHD OB Van (EBU Technology & Innovation Award 2019) and building a new Playout and



Switching center completely based on SMPTE ST 2110. One of the lessons learned from this innovation was the need for flexible metadata-handling in an IP environment.

Here we will highlight these two main issues and how they got married in a proof of concept in collaboration with Dolby Laboratories Inc.

The choices we made in our proof-of-concept were those available to us and with good support from the vendors and manufacturers concerned. Other solutions are evidently possible.

For our purposes, one of the biggest advantages of OBA is its ability through metadata to separate audio beds and other audio objects such as commentaries in different languages, audio description and so on. The idea of transmitting a blockbuster movie, concert or live sports event with one multichannel audio bed and a variety of three or four different languages plus audio description (AD) seems obvious. Whereas the thought of using static metadata for fixed rendering ratios for these programme elements was not an attractive option for us, the idea of delivering them with dynamic metadata came to mind.

The solution was found by taking the mixing parameters of an audio console (or any other audio processor) and using this data to control an authoring tool generating the desired metadata. Using a standard metadata protocol (in our case, S-ADM) allows the use and routing of these metadata in any standardized broadcast environment.

Open standards and protocols ensure the most interoperability between different products and as technology advances accelerate, they are becoming even more important for us as a broadcaster to embrace.

**Test Scenario:**

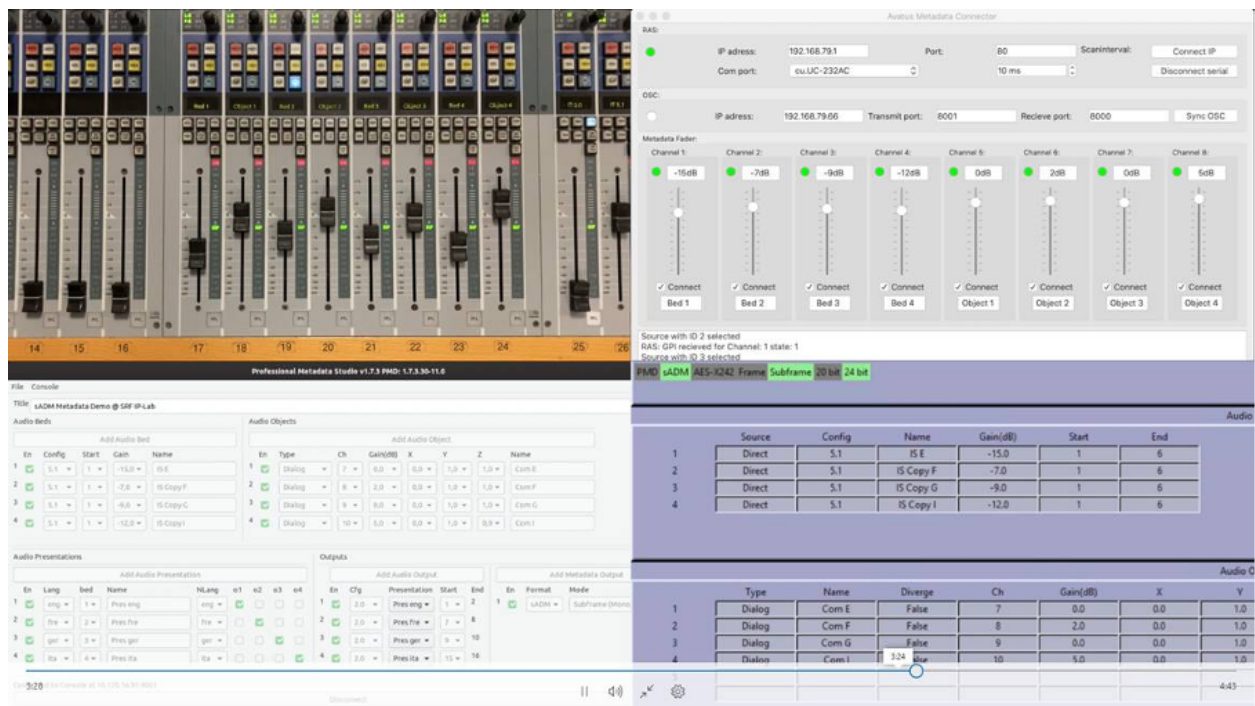
Our test was designed to prove how dynamic metadata could be used for a multilingual sports transmission. Typically, in such a programme we have an international ambience sound object (IS) and several commentary objects. Each object is transported once only, with accompanying metadata, but at the receiver side we have a different representation for each commentary. The appropriate mix for the IS and the individual commentary is controlled through the metadata.

This means that the IS object needs three different gains in the metadata, one for each representation, whilst each Commentary also needs a gain.



**Table 10. Representations Available**

Representation 1:	IS + Commentary 1
Representation 2:	IS + Commentary 2
Representation 3:	IS + Commentary 3



**Figure 26. Setup Screen Captures**

**Notes:**

- Top Left: Faders on the Desk controlling metadata only
- Top Right: Protocol converter between desk and PMD Studio
- Bottom left: PMD Studio with different gains for the same IS input
- Bottom right: AM Viewer on the receiver side



### Transport of dynamic metadata in a SMPTE ST 2110 environment:

S-ADM is an appropriate choice for such object-based audio metadata. It is clear to us that even in a modern facility there is always the need for backward compatibility as legacy SDI infrastructure and links are often encountered. A full IP-based workflow is of course the goal.

To simplify operational use, it is important to be able to generate metadata directly from a mixing desk. As there are nowadays so many controllers and web interfaces in a control room, having yet another UI for the metadata control is not an option to contemplate.

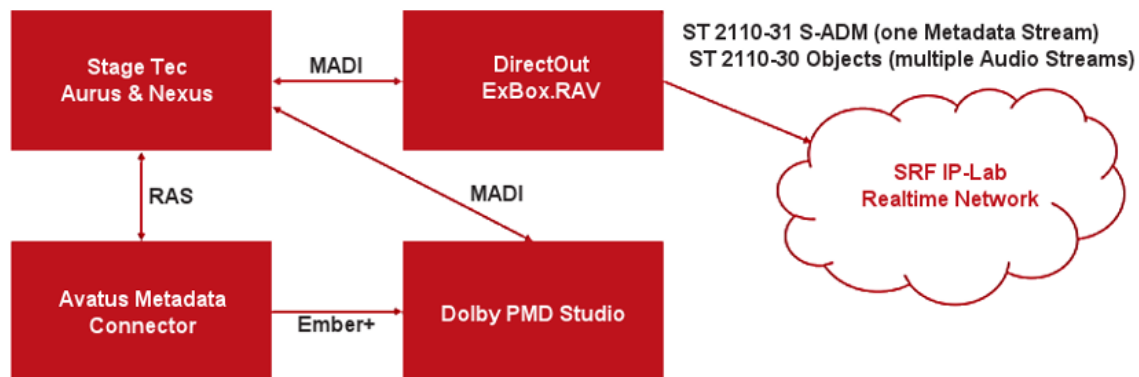


Figure 27. S-ADM setup in the audio control room

As is shown in [Figure 26](#), things rapidly get complicated, even in a very simple setup. The first challenge was to get the fader positions from our console to the metadata authoring tool (Dolby PMD Studio, in our case). There are different protocols possible, but no common standard.

Even with the chosen Ember+ protocol, parameters vary substantially between different products and vendors, as was the case here. As a result, no direct connection was available, and we had to develop our own protocol converter to interface our console with the PMD Studio.

The generated S-ADM is carried from the PMD Studio as an AES-subframe via a transparent MADI link to an AES67 Gateway (DirectOut ExBox.RAV) and is transported as a ST 2110-31 stream through our IP test environment.





Some research is still needed in this field to standardize parameters so as to reduce complexity. A potential way forward is an S-ADM profile in OSC<sup>15</sup>. As OSC is widely used in audio gear and with a defined S-ADM profile, the vocabulary of each device would be compatible. Perhaps in local setups, a pure OSC workflow could be a viable solution as many renderers for multichannel PA setups already use OSC.

**Signal path overview:**

Referring to [Figure 28](#) the delivery from origin to reproduction works as expected, even if there is conversion to and from an embedded SDI link through gateways. The open-source Dolby AM-Viewer decoded the metadata stream without any problems.

The downsides of this method may be summed up as scalability and bandwidth consumption. In a modern IP infrastructure this is not an issue though, but we think it is not a smart solution to transport dynamic metadata of a few kbit/s in a ST 2110-31 stream of approximately 3 Mbit/s. Also, as S-ADM data rates grow, the limitations of a constant bandwidth tunnel can incur even more inefficiencies. A solution with a SMPTE ST 2110-41 metadata stream will mitigate this problem.

A possible pitfall is also the fact that within the S-ADM, audio objects are only referenced by index. Whilst this is not an issue if you have a single stream containing both the audio and the metadata, in practice there will often be separate streams for metadata and the different audio elements. Furthermore, an audio element might have different metadata for different representations as per the scenario described above.

---

<sup>15</sup> Open Sound Control, <http://opensoundcontrol.org/>.

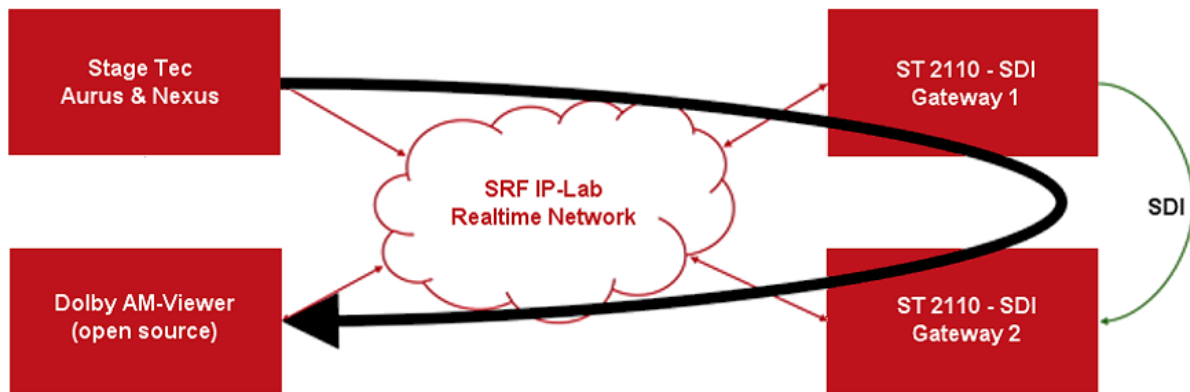


Figure 28. Signal Path overview

### 12.2.3.3 MPEG-H Audio Workflows

#### MPEG-H Audio for real-time workflow:

The MPEG-H Audio system is designed to work with today's streaming and broadcast equipment using SDI-based workflows as well as with future IP-based infrastructure. In real-time scenarios, the authoring of MPEG-H Audio scenes and the metadata export is handled by a device class called "Authoring and Monitoring Unit" (AMAU).

AMAU systems can be integrated into the existing SDI, MADI or future IP based signal flows which are used for broadcast. In content production, audio signals from the mixing console are fed into the AMAU using I/O converters. Within the AMAU, metadata creation and rendering take place. The AMAU can be controlled using a web interface or hardware controller; typically, the output of the AMAU is an SDI signal including 15 channels of audio and the Control Track on channel 16. In an IP based production facility, the audio and metadata would be transmitted in a container over IP. A video signal can be passed through for visual reference and for synchronization purposes. Finally, the audio and video signals are provided to the emission encoder.

At the time of this writing, there are three AMAUs available and used for MPEG-H Audio production in live broadcast: the Multichannel Monitoring and Authoring system from Jünger Audio (MMA), the Authoring and Monitoring System from Linear Acoustic (AMS) and the Spatial



Audio Designer - Processor from New Audio Technology (SAD-P)<sup>16</sup>. The use of AMAU systems allows productions to make use of all MPEG-H features without changing the entire workflow. Most of today's existing broadcast equipment can still be used.

For the purpose of monitoring MPEG-H Audio content, many different speaker layouts from stereo to 7.1+4H can be connected to an AMAU system. Additionally, all interactivity options and the audio quality can be monitored during production using an emulation of end-user receivers and with different reproduction configurations.

AMAU systems measure the loudness and true peak values of all channels, objects, output buses and formats, as well as every created Preset in real-time. With the resulting data, correction values are added to the metadata stream compliant with the applicable loudness regulation. The measurement of all generated DRC profiles and real-time loudness correction are also included. Additionally, AMAU production tools support the user with visualizations of all crucial measurement values.

### **MPEG-H Audio for real-time Contribution using SDI-based workflow:**

A typical MPEG-H Audio real-time workflow for live productions based on SDI infrastructure appears as in Figure 144. It consists of a chain containing an MPEG-H Authoring Unit and a Video Broadcast Encoder. In real-time scenarios, the authoring of MPEG-H Audio scenes and the metadata are handled by the AMAU as described in Section 25.3.1.

To ensure the integrity of metadata in an SDI environment in any production step, the metadata is delivered over SDI using the so-called "Control Track"<sup>17</sup>. The Control Track is a "time-code like" audio signal and can be treated as a regular audio channel. This ensures the synchronization of metadata with its corresponding audio and video signal. The Control Track is designed to survive A/D and D/A conversions, level changes, sample rate conversions or frame-wise editing. The Control Track does not require audio equipment to be put into data

---

<sup>16</sup> More information and the AMAU's available here:

1. Jünger Audio, Multichannel Monitoring and Authoring (MMA),  
<https://www.junger-audio.com/en/products/flexai/multichannel-monitoring-and-authoring-mma>
2. Linear Acoustic, Authoring & Monitoring System (AMS),  
<https://www.telosalliance.com/Linear/Linear-Acoustic-AMS-Authoring-Monitoring>
3. New Audio Technology, Spatial Audio Designer – Processor (SAD-P),  
<https://newaudiotechnology.com/products/spatial-audio-designer-processor/>

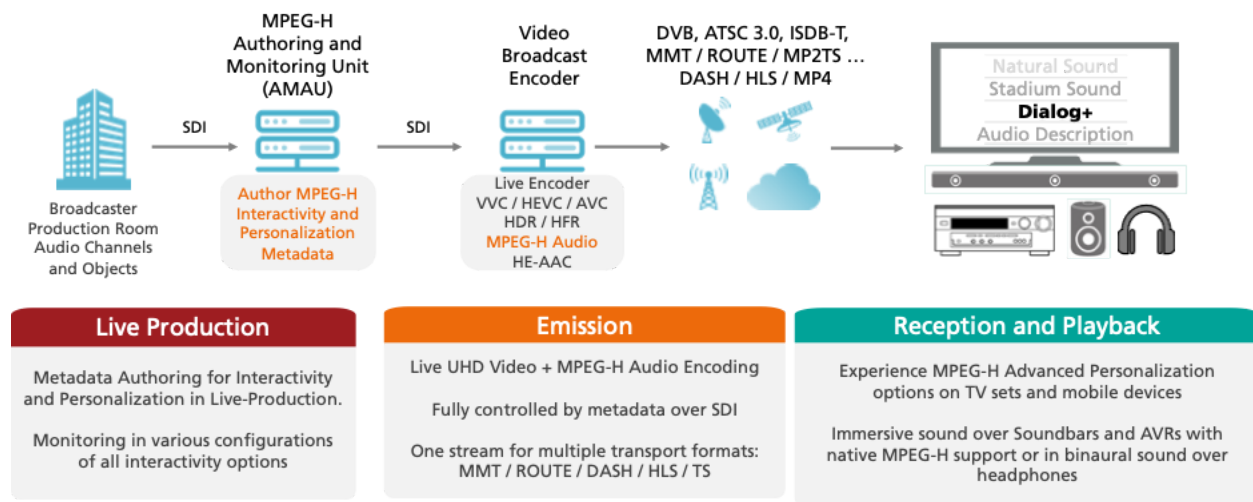
<sup>17</sup> Reference: R. Bleidt et al. "Development of the MPEG-H TV Audio System for ATSC 3.0," in IEEE Transactions on Broadcasting, vol. 63, no. 1, March 2017,  
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7874294>



mode or non-audio mode in order to pass through as it was introduced for efficient and robust usage of MPEG H Audio in SDI-based workflows from the beginning.

The metadata for the audio signal is collected into packets synchronized with the video signal and is modulated with analog channel modem techniques into a Control Track signal that fits in the audio channel bandwidth. This signal is unaffected by typical filtering, resampling, or scaling operations in the audio sections of broadcast equipment.

A detailed tutorial on how to handle MPEG-H Audio in live production can be found online in the MPEG H Audio Live Production Tutorial series<sup>18</sup>.



**Figure 29. Typical MPEG-H Workflow for Live Productions**

MPEG-H Audio authoring and encoding is also supported in SMPTE ST2110 compliant workflows. According to SMPTE ST 2110 [43-47], in IP-based workflows the media essences such as video, audio and metadata are transmitted over separate RTP connections and PTP (IEEE 1588, SMPTE ST 2059) is used to synchronize the different essence streams. The transmission of the PCM audio essence (SMPTE ST 2110-30 [46]) is already established whereas the standardization for the transport of metadata including serialized ADM (S-ADM) metadata (SMPTE ST 2110-41<sup>19</sup>) is still ongoing at the time of writing.

The serial representation of the Audio Definition Model, according to ITU-R BS.2125 [123] defines a segmentation of the original ADM for use in linear workflows such as live production

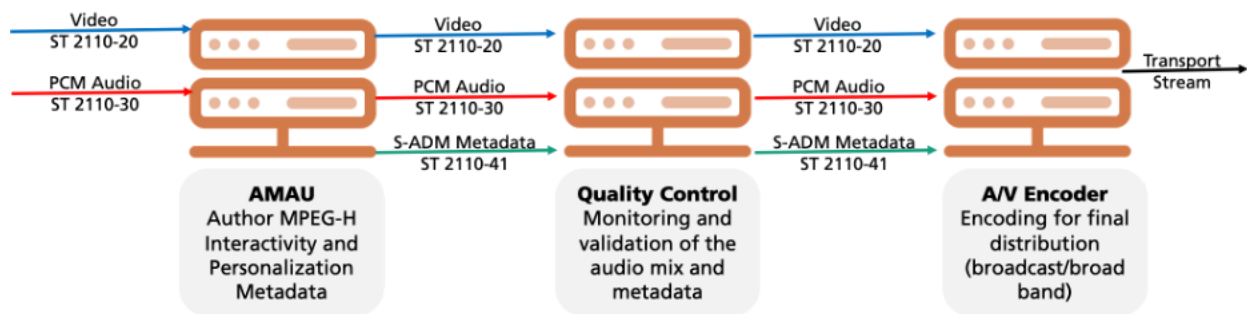
<sup>18</sup> MPEG-H Audio: Live and Broadcast Tutorial series, <https://vimeo.com/showcase/7538846>

<sup>19</sup> SMPTE ST 2110-41, Fast Metadata Framework



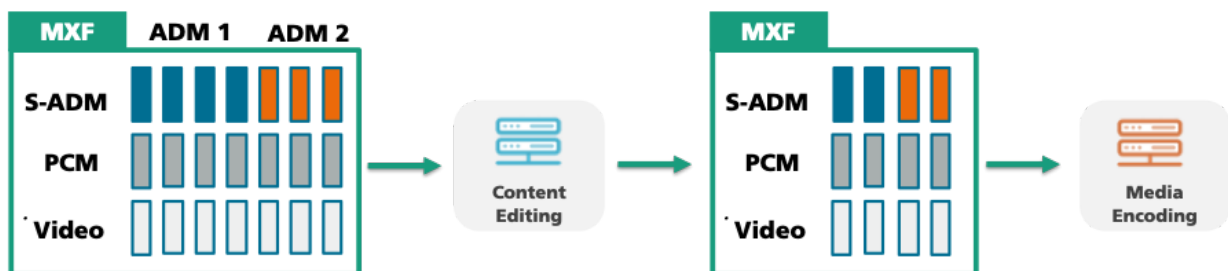
for broadcasting and streaming applications. Similar to the MPEG-H Control Track, one S-ADM frame contains a set of metadata describing at least the audio frame over the time period associated with that frame. S-ADM has the same structure, attributes and elements as those of ADM, as well as additional attributes to specify the frame format. The S ADM frames are non-overlapping and contiguous with a specified duration and start time.

For illustration purposes, a simplified ST 2110 Audio and Metadata over IP workflow for MPEG H Audio is depicted in [Figure 30](#).



**Figure 30. MPEG-H Audio production workflow based on SMPTE ST 2110 (simplified)**

For storage and playout of S-ADM based content, the standardization of the transport of PCM audio essence and S-ADM metadata inside the Material Exchange Format (MXF, [SMPTE ST 377-1 \[36\]](#)) is underway at the time of writing. The MXF Format is optimized for content interchange or archiving by creators and/or distributors and provides a complete framework for the transport of NGA. An example application of MXF in practice is shown in [Figure 31](#).



**Figure 31. Example of MXF based workflow using ADM metadata**

**MPEG-H Audio for file-based workflows:**



In file-based workflows, MPEG-H content can be created via means of file-based processing using either stand-alone software or integrated in a variety of tools such as DAW, server and platforms. In those tools several processes are executed including measuring the loudness levels of the audio components, setting of the user interactivity options and authoring of the NGA MPEG-H scene. After completion of the authoring, the metadata and the audio data are exported together as an "MPEG-H Master" file. An example of such implementation is the MPEG-H Authoring Suite, a software suite designed for authoring MPEG-H Audio scenes<sup>20</sup>.

Using the tools of the MPEG-H Authoring Suite or other post-production tools, an "MPEG-H Master" can be exported as a bundle of metadata and audio content. MPEG-H Audio metadata contains all control information for user interactivity and also all necessary information that the playback device needs for reproduction and rendering to ensure the expected experience on any platform.

MPEG-H Master content can be created from multiple source formats including object-based productions, ADM and legacy channel-based material, either by authoring new projects or by ingesting existing material. The ingest step can enhance stereo content with Dialog+ as described above or convert an existing BWD/ADM file (e.g., created with Dolby Atmos Production Suite) to MPEG-H ADM Profile and enhance the content with advanced personalization and interactivity options.

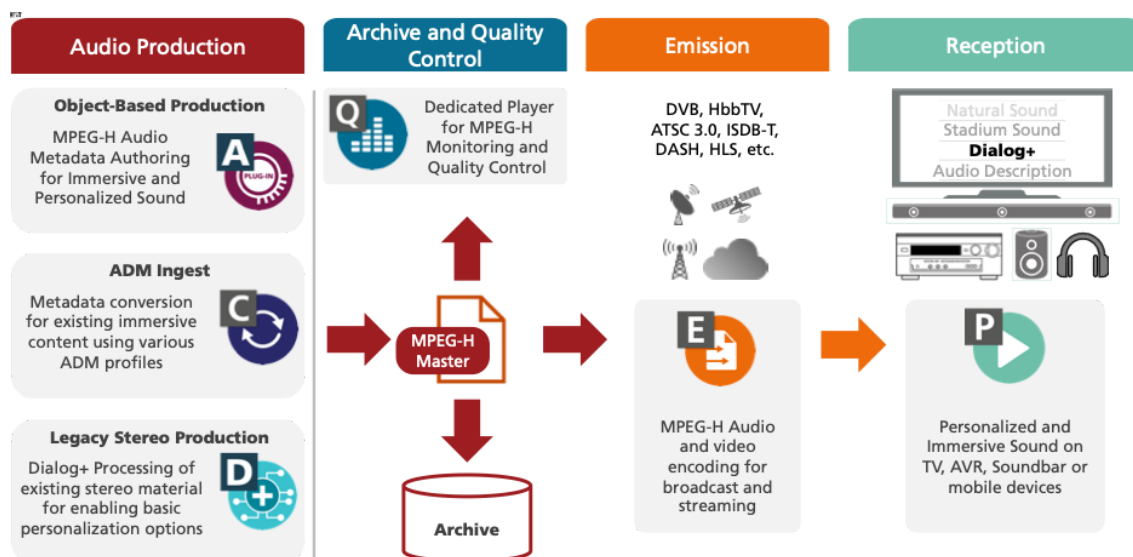


Figure 32. MPEG-H Audio Post-Production workflow (simplified)

<sup>20</sup> The MPEG-H Authoring Suite, <https://www.iis.fraunhofer.de/en/ff/amm/dl/software/mas.html>



The MPEG-H Master can be exported in the following formats:

- **MPEG-H BWF/ADM:** An MPEG-H BWF/ADM file (short for Broadcast Wave Format with embedded Audio Definition Model metadata) is a multi-channel wave-file which contains all the audio and metadata of the MPEG-H Audio scene. The exported BWF/ADM file is compliant to the MPEG-H ADM Profile<sup>21</sup>.
- **MPEG-H Production Format (MPF):** An MPF file is a multi-channel wave-file which contains all the audio and metadata of the MPEG-H Audio scene. The metadata is stored in the "Control Track", which is one of the audio tracks in the multichannel wave-file and contains a "time-code like" signal that is robust against sample rate conversions or level changes. The Control Track is used for live productions using SDI-based workflows (see [MPEG-H Audio for real-time Contribution using SDI-based workflows](#)).

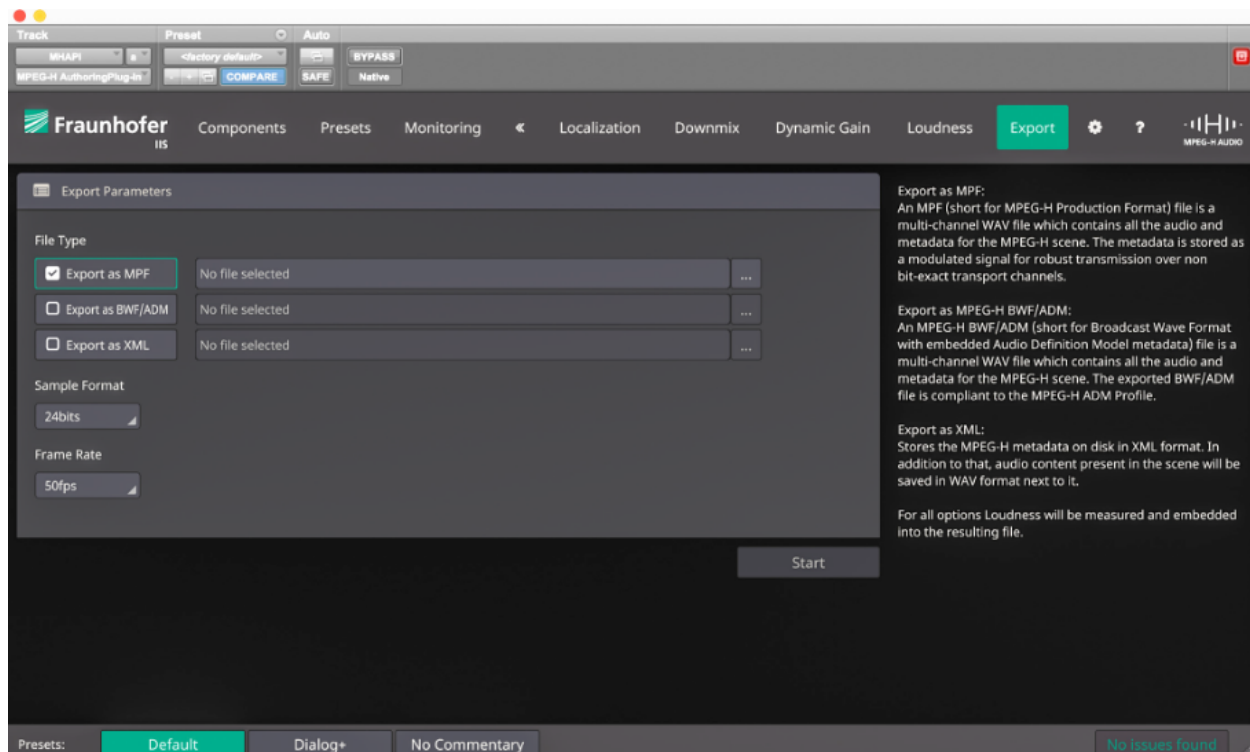


Figure 33. Export window of the MPEG-H Authoring Plug-in

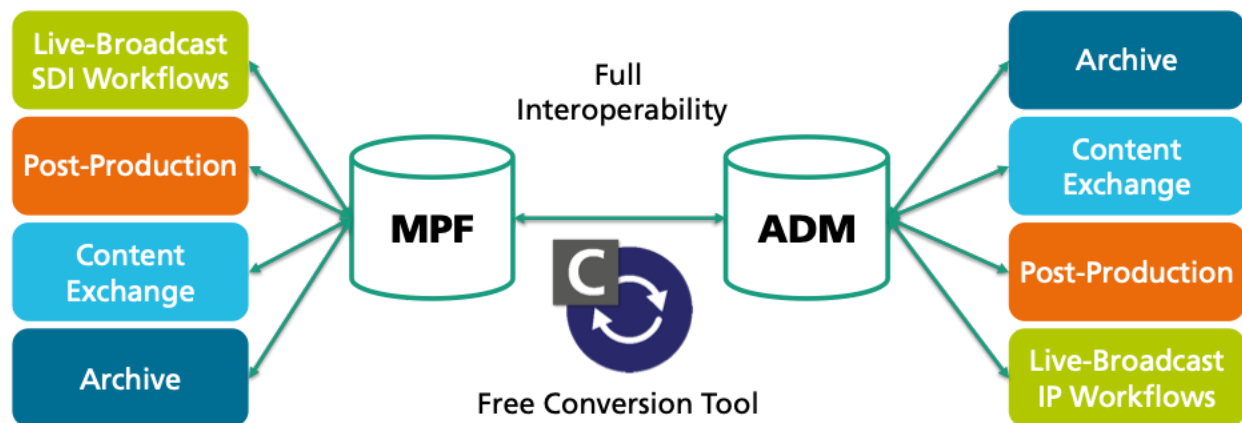
<sup>21</sup> The MPEG-H ADM Profile, <https://www.iis.fraunhofer.de/en/ff/amm/dl/whitepapers/adm-profile.html>





The MPEG-H Masters can be used for both live and offline encoding or stored for archive purposes. Interoperability is made possible thanks to transcoding tools which allow conversion from and to MPF (MPEG-H master file) and ADM formats.

As shown in [Figure 33](#) the MPF format is suitable for existing SDI-based live workflows and it may be used for all other purposes such as post-production, content exchange and archiving. On the other hand, the ADM format is ideal for content storage, content exchange using MXF based workflows, and with its serialized option for IP-based workflows according to SMPTE ST 2110 [\[43-47\]](#) suite of standards.



**Figure 34. Full Interoperability between MPEG-H Masters and ADM-based workflows**

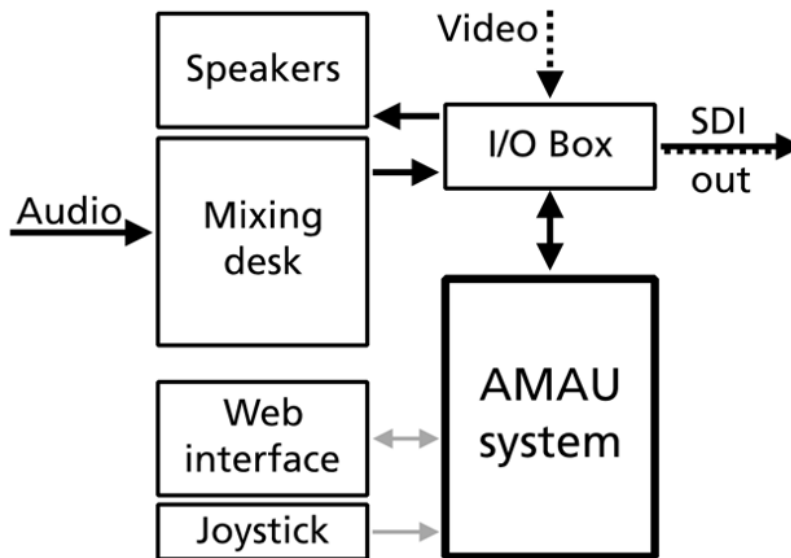
### **MPEG-H Audio Production Tools:**

MPEG-H operations, both real-time or file-based, are supported by a comprehensive variety of hardware and software tools which allow flexible and efficient authoring and monitoring of NGA content. This section provides short descriptions of the most common tools in use at the time of this writing.

### **Tools for MPEG-H Audio real-time Workflows:**

In a real-time production environment, as well as during post-production, a 3D Audio scene needs to be created and monitored. At the same time metadata needs to be generated to describe the audio scene and to enable the features provided by MPEG-H. For the purpose of generating or monitoring MPEG-H Audio content, a device class called Audio Monitoring and Authoring Unit (AMAU) has been developed.





**Figure 35. Signal flow of an MPEG-H AMAU System**

The AMAU can be controlled using a web interface or hardware controller; typically, the output of the AMAU is an SDI signal including 15 channels of audio and the Control Track on channel 16 (see [Figure 35](#)). In an IP based production facility, the audio and metadata would be transmitted in a container over IP. A video signal is passed through for visual reference. After the audio and video signal leaves the AMAU it can be passed on to the master switch.

An AMAU device can also be used after the master switch for monitoring purposes of the MPEG H scene and its related metadata, right before the emission encoder in order to check the quality of the signal. At this point, usability of interactivity features or loudness parameters can be checked.

Currently, there are three AMAU products available to the market (see [footnote 16](#)):

- Authoring and Monitoring System by Linear Acoustic (AMS)
- Multichannel Monitoring and Authoring system by Jünger Audio (MMA)
- New Audio Technology Spatial Audio Designer Processor (NAT SAD-P)

The use of AMAU devices allows productions that support all MPEG- H features without changing the entire workflow, as most of existing equipment still can be used and does not need to be replaced.



---

A video tutorial series is available to guide you step-by-step through the live authoring, monitoring, and encoding (see [footnote 18](#)).

### **MPEG-H Encoders and Decoders:**

Several manufacturers support MPEG-H Audio encoding/decoding in their product families including ATEME, KaiMedia, Sumavision, KuVision, DS Broadcast, MediaKind, and Pixtree.

### **Tools for MPEG-H Audio File-based Operations:**

Typically, DAWs are used in audio post-production operations. Most common DAWs support the integration of AAX, VST or AU based plugins, which extend the capabilities of the host. Plugins such as the MPEG-H Authoring Plugin – part of the freely available MPEG-H Authoring Suite – or the Spatial Audio Designer (SAD) by New Audio Technology can be integrated into the existing post-production workflows for MPEG-H audio content creation.

When the production workflow allows, file-based operations provide several benefits. Compared to real time panning, post-production operations allow for more advanced object movements because automation can be edited and retrieved. The number of objects that move at the same time can be higher for the same reasons. Another benefit is that they can overcome the DAWs bus width limitation, e.g., enabling a high number of objects to loudspeaker layouts not natively supported in the DAW, such as 5.1+2H, 5.1+4H or 7.1+4H.

Furthermore, to monitor the different presets or switch groups that have been defined, a full MPEG H renderer is included in the authoring tools. Finally, when the mix is finished and all metadata entries are authored, the session can be exported to an MPEG-H Master. During this stage, the loudness of all components and presets is measured offline and embedded into the corresponding metadata fields.

MPEG-H software tools also allow conforming to legacy channel-based audio content. For example, in case a stereo or 5.1 audio mix should be prepared for MPEG-H broadcast, the mix does not need to be modified as only metadata needs to be generated. For this, stand-alone tools such as the MPEG-H Authoring Tool (MHAT), part of the MPEG-H Authoring Suite<sup>22</sup>, are available. It is also possible to monitor the created scene and presets and export the final MPEG-H Master.

---

<sup>22</sup> The MPEG-H Authoring Suite, <https://www.iis.fraunhofer.de/en/ff/amm/dl/software/mas.html>



---

To control created MPEG-H Masters with or without an accompanying picture, the MPEG-H Production Format Player, part of the MPEG-H Authoring Suite, can be used to ensure that the quality of the mix and the authoring is matching the expectations.

**MPEG-H Authoring Suite for file-based Operations:**

The MPEG-H Authoring Suite (MAS) is a comprehensive set of tools that make the production of MPEG H Masters easy, fast and intuitive. These tools support the MPEG-H ADM Profile, as well as binaural monitoring for immersive audio reproduction over headphones. The MPEG H Authoring Suite can be freely downloaded. An exhaustive tutorial on how to use the MPEG-H Authoring Suite is also provided<sup>23</sup>. The main components of the MPEG-H Authoring Suite are:

- MPEG-H Authoring Plug-in (MHAPi)
- MPEG-H Authoring Tool (MHAT)
- MPEG-H Conversion Tool (MCO)
- MPEG-H Production Format Player (MPF Player)

While dialog enhancement (DE) can be directly enabled for new MPEG-H Audio productions using object-based audio, it is also important to enable DE for existing stereo content (e.g., archive material). This would offer a consistent user experience and the option to improve intelligibility for hearing impaired audience. To achieve this, different solutions for Dialog separation can be used with the MPEG-H Audio system. Interoperability is ensured through the BWF/ADM open format.

---

<sup>23</sup> Video Tutorial MPEG-H Audio Authoring Suite,  
<https://www.vimeo.com/user/41127038/folder/1871919MPEG-H>



---

At the time of this writing, additional authoring tools available are Minnetonka AudioTools Server, Da Vinci Resolve 16 Studio and New Audio Technology Audio Designer (SAD)<sup>24</sup>.

### **MPEG-H Utility Tools:**

#### **MPEG-H Dialog+ demonstrator:**

Dialog+ is based on artificial intelligence and reduces the volume of music, sound effects or background noise to put spoken dialogue in the foreground. Developed by Fraunhofer IIS, the technology is easy to integrate into existing production processes. It does not replace the original audio mix, but instead offers an alternative that viewers can switch to if they so choose. Complementing MPEG-H Audio object-based productions, the main application of Dialog+ is increasing the speech intelligibility of existing material transmitted in both legacy and NGA services at the same time, allowing for cost-effective end-to-end workflows.

The Dialog+ technology was tested in a public trial in October 2020 together with German public broadcaster WDR where the audience was able to try out and evaluate their listening experience<sup>25</sup>.

#### **MPEG-H Info Tool:**

The MPEG-H Info Tool (MHIT)<sup>26</sup> provides an overview of all information associated with content masters that are supported by the MPEG-H Audio system. These range from file-related information (e.g., date of creation, file size, etc.) to a representation of the metadata which is carried in the file (e.g.: labels, number of presets, etc.). The Scene Summary provides a list of information with several levels of detail. The unfolded tree view shows all the metadata of an MPEG-H Scene that is relevant for experts who need a deeper insight into the details. In addition to providing information, the MHIT is also equipped with automated tests for ADM-based content.

---

<sup>24</sup> For more information see:

1. Telos Alliance Minnetonka AudioTools Server MPEG-H Audio Authoring, <https://www.audioblog.iis.fraunhofer.com/minnetonka-audiotools-server>
2. DaVinci Resolve, <https://www.blackmagicdesign.com/products/davinciresolve/studio>
3. New Audio Technology, Spatial Audio Designer (SAD), <https://newaudiototechnology.com/products/spatial-audio-designer/>

<sup>25</sup> Experiencing Dialog+ on linear television, <https://www.audioblog.iis.fraunhofer.com/dialogplus-on-linear-television>

<sup>26</sup> MPEG-H Info Tool, <https://www.iis.fraunhofer.de/en/ff/amm/dl/software/mpeg-h-info-tool.html>



The MPEG-H Info Tool can help with this task by providing automated tests of ADM-based content. The free tool runs a conformance framework equipped with exhaustive sets of checks derived from [ITU-R BS.2076-2 \[72\]](#), [ITU-R BS.2088-1 \[73\]](#), and [ITU-R BS.2125-0 \[123\]](#) as well as from all ADM profiles related to the MPEG-H Audio system. After analysis, the tool displays a summary of the conformance checks' findings and indicates whether the provided content is supported by the MPEG-H Audio system.

At the time of this writing, additional MPEG-H info tools include BBright UHD-Play and DekTec StreamXpert<sup>27</sup>.

### **EAR Production Suite:**

The EAR Production Suite (EPS) is a joint open-source development of BBC R&D and IRT consisting in a set of VST plugins for digital audio workstations (DAWs) that enable sound engineers to produce immersive and personalizable content using the Audio Definition Model (ADM) format and to monitor it using the ITU ADM Renderer. The EAR Production Suite enables the import and export of ADM files, compliant to the EBU ADM Production profile. The VST plugins are currently optimized for the Reaper DAW, which features an extension interface that is used to import and export ADM files within a BW64 container.

For more information go to: <https://ear-production-suite.ebu.io/>

### **MPEG-H Use Cases:**

The MPEG-H Audio system has been successfully integrated into various HD and UHD production workflows. In the following, several case studies, where the MPEG-H Audio system has been used during live broadcast or streaming of major events, are described. This gives a brief overview of the operational requirements and processes required during an NGA production and broadcast.

#### **Globo broadcast using MPEG-H Audio during Rock in Rio:**

Rock in Rio, one of the largest music festivals in the world, with over a million people attending, was broadcast live over the ISDB-Tb terrestrial TV service, HLS streaming and 5G using MPEG-H Audio in 2019<sup>28</sup>. It was the first live end-to-end broadcast using MPEG-H Audio in Brazil.

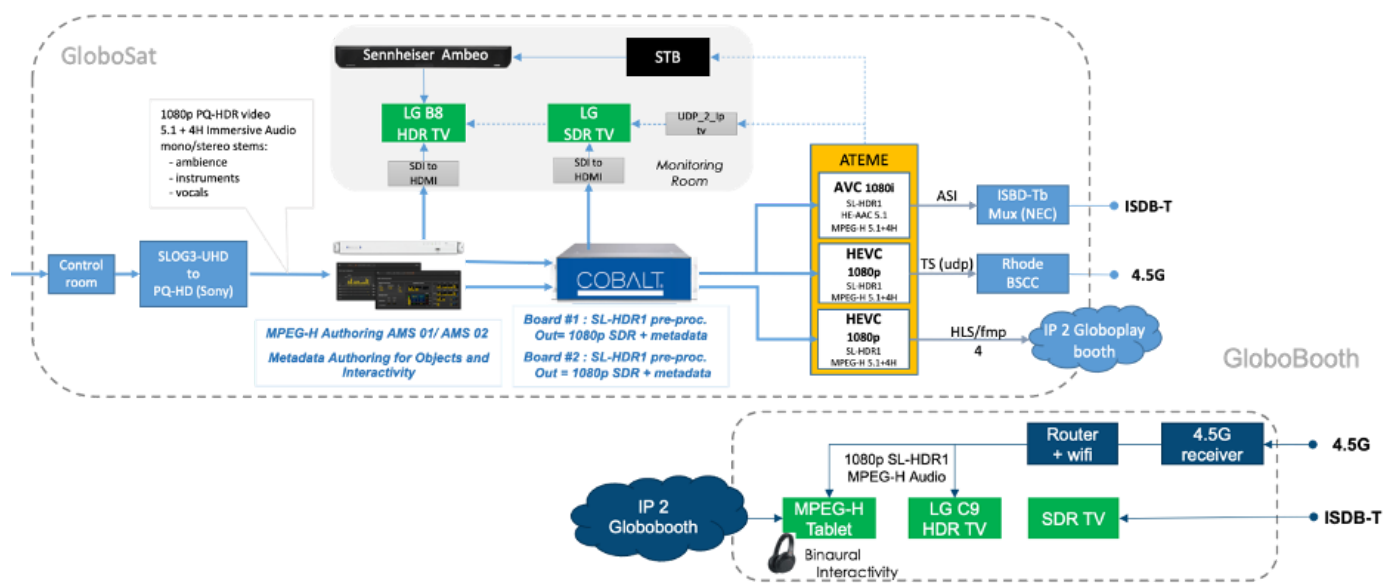
---

<sup>27</sup> BBright UHD Player, <https://www.audioblog.iis.fraunhofer.com/nab-2019-mpeg-h-bbright>  
DekTec StreamXpert, <https://dektec.com/products/applications/streamxpert/>

<sup>28</sup> A world's first: Globo successfully used MPEG-H Audio for ISDB Tb terrestrial and 5G broadcast during Rock in Rio, <https://www.audioblog.iis.fraunhofer.com/globo-rockinrio-mpeg-h-isdbtb-5g>



Globo, the largest media group in Brazil, successfully used MPEG-H Audio for a live broadcast over the enhanced ISDB-Tb system during one of the world's biggest music festivals, Rock in Rio. Using the MPEG-H Audio format, the musical performances on the two main stages, "Mundo" and "Sunset," were delivered over the air with immersive and personalized sound in the Rio de Janeiro area. In addition, an enhanced video quality was provided using Advanced HDR by Technicolor. A brief overview of the Rock in Rio setup is given in [Figure 36](#).



**Figure 36. Rock in Rio 2019 setup using MPEG-H Audio**

With MPEG-H Audio, the viewers were able to experience MPEG H immersive sound of the concerts as if they were present at the venue and moreover had the opportunity to interact with the audio content and to benefit from its advanced personalization options using tablets and binaural rendering over headphones at the Globo and Globo play booths at the festival venue. In the offices of Globosat, Fraunhofer demonstrated the ease of use of MPEG-H, bringing immersive sound into any living room with the MPEG-H-enabled Sennheiser Ambeo soundbar, which offers studio-grade sound quality.

The immersive mix using 5.1 surround plus four height channels was produced by Globosat sound engineers. Additionally, Globosat made available several mono and stereo stems for ambience, instruments or vocals. Fraunhofer sound engineers used these stems to enrich the live mix with audio objects, enabling personalization of the user experience by enhancing the level of the ambience or moving the vocals into the height loudspeaker layer. Besides immersive



audio, personalization is one of the key features of MPEG-H Audio to dramatically improve the attractiveness of a service.

The same MPEG-H Audio production was simultaneously delivered over multiple distribution platforms. Besides the ISDB-Tb terrestrial broadcast, an HLS streaming service using MPEG H Audio was offered, with commercially available LG TV sets playing back the MPEG-H Audio stream. And with the support of Rohde & Schwarz, MPEG-H Audio was embedded in an experimental broadcasting UHF channel for the first 5G broadcast transmission field test in Brazil. ATEME's TITAN Live encoder created all three services in parallel. According to the SBTVD specification, the ISDB Tb service also contained an HE-AAC audio stream, which was used for legacy receiving devices, while the MPEG H audio stream could be consumed with MPEG-H enabled receivers and soundbars.

### **EBU test with MPEG-H Audio during the Eurovision Song Contest:**

A good example of a case study where the MPEG-H Audio capabilities have been tested in a complex and very challenging real-time workflow is the production of the Eurovision Song Contest (ESC) in 2018 and 2019<sup>29</sup>. The Eurovision Song Contest is an international song competition produced by the European Broadcasting Union (EBU) and one of the most viewed non-sporting events worldwide.

### **Eurovision Song Contest 2018:**

Held in Altice Arena Lisbon, Portugal and watched live by 186 million viewers, the production was supported by the national Portuguese broadcaster Rádio e Televisão de Portugal (RTP). The EBU and Fraunhofer IIS collaborated to conduct a field test for immersive and interactive live audio production, based on the MPEG-H Audio system. The goal of the field test was to evaluate future production and reproduction scenarios. The authoring and monitoring of the immersive mixes were done live for demonstration purposes as well as executed in post-production after the event. The immersive mix was not broadcasted in this case but distributed live locally during the event.

The legacy audio production format for the international feed at ESC was a 5.1 surround sound mix and an additional 2.0 stereo mix. Both were created by the host broadcaster's OB van on location. The signals used were a mix of microphone feeds and pre-produced material, such as special sound effects (SFX) and trailers for the participating competitors. In total, 232 microphones were employed by the broadcaster to capture sound from the main stage, hosts,

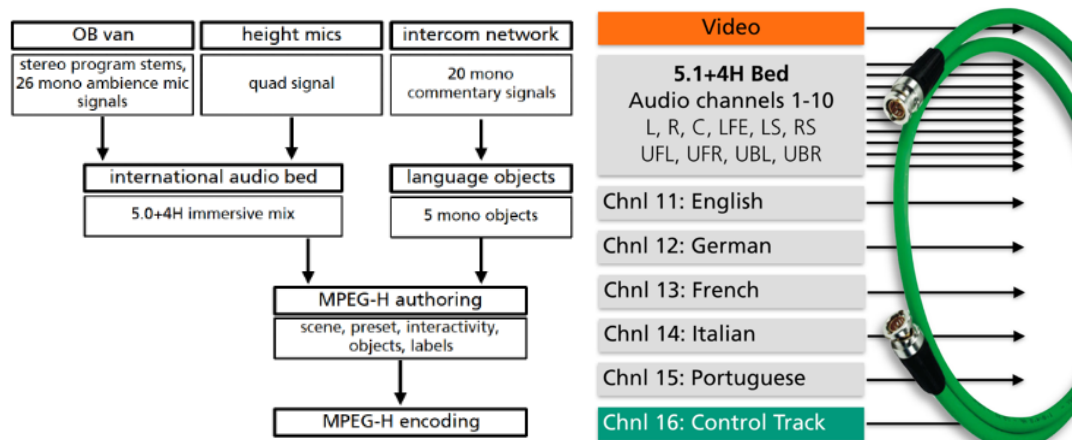
---

<sup>29</sup> Immersive and Personalized Audio at the Eurovision Song Contest,  
<https://tech.ebu.ch/news/2019/05/immersive-and-personalized-audio-at-the-eurovision-song-contest>



interview partners and arena audience. Based on these signals, international feeds were mixed. 26 ambience microphones were put into different zones throughout the audience, both in front of the stage and on the rear floor area, aiming to the upper tiers and downwards from the PA rigs.

Positioned this way, a versatile mix of audience reactions can be created, surrounding the listener by blending more diffuse and more direct ambience signals and giving the mixer the possibility to select the most appropriate sounds. Microphone signals representing the diffuse upper part of the arena, preferably with minimum direct PA sound, were missing in the conventional production workflow to create an immersive experience. For that, a Hamasaki Square microphone was additionally placed underneath the roof in the center of the arena about 25 meters above the floor. Minimizing direct sound from PA and audience by its polar patterns and direction, it provided the upper layer ambience signal by adding just four additional microphones. A 3D audio mixing room was set up, receiving all OB van's signals including sub-mixes and additional commentator feeds from a second service provider, as shown in [Figure 37](#).



**Figure 37. Schematic audio production and authoring workflow of the ESC MPEG-H field test**

For creating an immersive, three-dimensional sound impression, the Hamasaki square signals were panned 100% to the upper layer of the loudspeakers, which created the intended "dome" effect. Microphone signals from the upper part of the arena that are panned to the middle layer for regular 5.1 audio production, have been used differently for the immersive mix. Here, they were panned slightly upwards to bring them to their natural position in the sound field, about 15 degrees upwards from the arena floor.





Once the immersive ambience was created, remaining stems like music and hosts were mixed in a legacy way, panning them in the mid-front line of the immersive loudspeaker configuration.

Additionally, a selection of the 20 available commentary feeds were added to the audio scene, whereupon personalization options were configured and monitored.

The resulting 5.1+4H mix could be monitored using an AMAU or DAW with MPEG-H enabled plugins to monitor downmix from 5.1+4H to 5.1 and stereo as well as the binauralized playback. The field test showed that the already existing infrastructure for sound capturing and processing only needs a few additional changes to be able to produce immersive and interactive sound. In the described case, four additional microphones and loudspeakers plus an MPEG-H enabled authoring tool were required. The resulting immersive and interactive audio mix received positive feedback compared to the legacy stereo or 5.1 production, as stated by involved producers and audio engineers during on-site demonstrations.

### **Eurovision Song Contest 2019:**

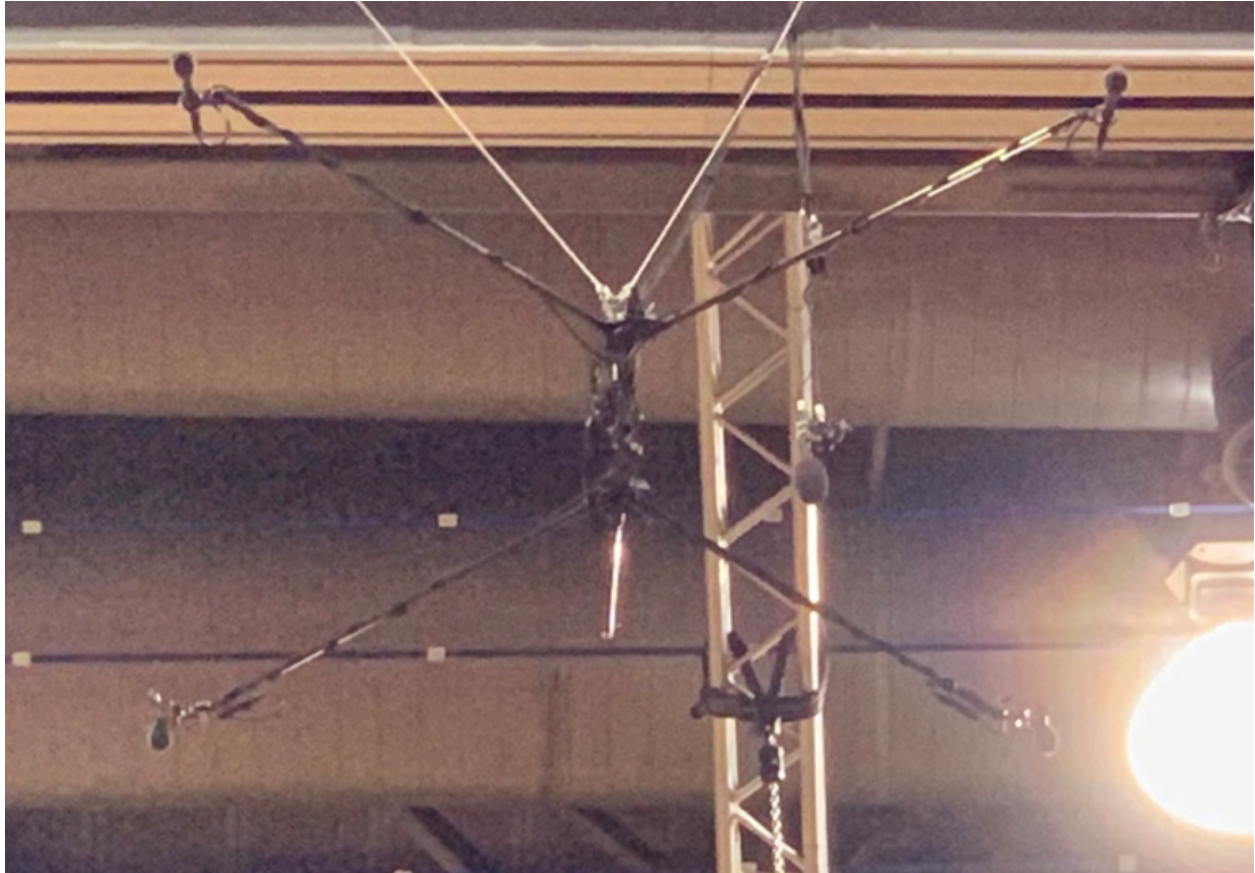
For the 2019 Eurovision Song Contest (ESC) in Tel Aviv the EBU and its technology partners did a parallel immersive live production of the contest, which was mixed and transmitted using the MPEG-H Audio NGA system. Compared with the 2018 edition, in 2019 the focus was on demonstrating a complete workflow including the transmission part. EBU members on site in Tel Aviv were able to experience the MPEG-H Audio enhanced show as a DASH livestream and as a DVB-compliant live broadcast stream. The latter was also delivered live to broadcasters abroad via the Eurovision FINE network. The "produce once, deliver anywhere" feature is one of the great advantages of the MPEG-H Audio system.

During the ESC 2019 broadcast trial Fraunhofer's sound engineers used the System T S500m broadcast console provided by Solid State Logic to produce the 5.1+4H immersive mix of the show and handle the additional host and commentary signals in four different languages as audio objects.



**Figure 39. Audio Monitoring Room for Immersive Sound Production during ESC 2019**

The MPEG-H immersive and interactive mix was done in parallel with the main broadcast program production. The MPEG-H production room (see [Figure 39](#)) was located next to the main OB Van, receiving most of the microphone feeds together with OB Van stereo mixes. Similar to the event in Lisbon, Fraunhofer IIS sound engineers have placed additional Hamasaki Square microphones on the ceiling of the arena for better ambience capturing (see [Figure 40](#)).



**Figure 40. Additional Hamasaki Square and Ambience microphones on the arena's ceiling**

The Hamasaki Square was mounted in the middle of the arena at 26-meter height, above the coverage of the PA system. This delivered a decorrelated ambience with as little music spill as possible and the final 5.1+4H immersive mix was created by pan and delay of the ambience microphone signals and up-mix of the vocals and effects via an Illusonic IAP16. The mix was monitored live using a Genelec 5.1+4H speaker setup.

The MPEG-H metadata was authored live using both Telos Alliance Linear Acoustic AMS and Jünger Audio MMA. One AMAU was used for the main production while the second one was used as fallback and for ad-insertion. The metadata was authored for offering the user multiple options to choose between different versions of the content (presets): a default preset, a dialog enhancement preset and a venue preset reproducing the experience of being present in the



arena. For each preset, various interactivity features were enabled, such as level and position interactivity.

During production, the sound engineers were able to create new presets "on the fly", change the interactivity ranges for any available audio object, disable interactivity for some presets or even completely switch to a different configuration in a seamless way. ATEME's TITAN Live UHD broadcast encoder was used to create a DVB-compliant broadcast transport stream and, at the same time, DASH and HLS streams for delivery over the internet.



**Figure 41. MPEG-H Audio Immersive and Interactive demonstration during the ESC 2019**

The MPEG-H programs were demonstrated in Tel Aviv, Geneva, and Madrid, using the Sennheiser Ambeo soundbar (see [Figure 41](#)) able to decode and playback the MPEG-H streams for creating an immersive experience without the need to place loudspeakers all around





the room. Additionally, the viewers were able to switch between different versions of the content, select their preferred language and level of dialog, as well as use the advanced menus to adjust the balance between the ambience of the arena and the dialog to their personal preference. Moreover, during the demonstration visitors had the chance to move the audio objects “away from the screen” to another part of the room.

### **MPEG-H Audio broadcast over DVB-T2 during the French Tennis Open:**

In 2018 and 2019, France Télévisions set up a dedicated Ultra HD event channel during the French Tennis Open in Paris using MPEG-H Audio for producing a sound immersive and personalized user experience<sup>30</sup>. The Ultra HD channel was offered to French viewers over DVB-T2 jointly with TDF and at the same time over DVB-S2 with the support of Fransat. The 4K resolution and high dynamic range (HDR) program was associated with MPEG-H Audio immersive and interactive features. Moreover, an HLS streaming service was setup and using an LG TV set supporting MPEG-H Audio, the streaming services was showcased in Paris.

During the tournament, France Télévisions provided a 5.1+4H immersive sound bed for the court atmosphere together with an additional object for the commentary in French, allowing the viewers to personalize the content.

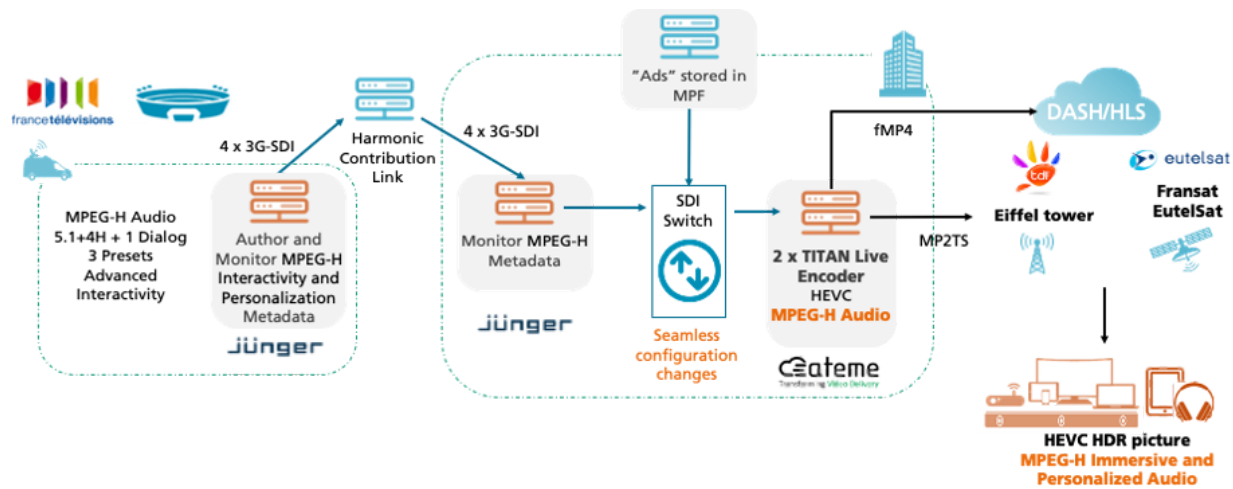
As described in the articles listed in [footnote 30](#), on site the production truck of France Télévisions (La Fabrique – Toulouse) produced in parallel the international HD signal in addition to the Ultra HD video signal. The sound from the Philippe Chatrier court was captured by means of ultra-directional microphones which cover the field, the players and the referee, associated with several microphones placed in the stands, including an ORTF-3D mic from Schoeps which serves as main sound capture for the ambience.

To generate the metadata necessary for the adaptation and interactivity of the Next Generation Audio MPEG-H signal, the MPEG-H Authoring and Monitoring unit from Jünger (MMA) was integrated in the OB van, as shown in [Figure 42](#).

---

<sup>30</sup> Successful Terrestrial And Satellite Reception Of MPEG-H Audio During The Roland Garros French Open,  
<http://idfrcetv.fr/successful-terrestrial-and-satellite-reception-of-mpeg-h-audio-during-the-roland-garros-french-open/>

Roland Garros 2019 – An Ultra HD event channel with MPEG-H Audio:  
<http://idfrcetv.fr/roland-garros-2019-an-ultra-hd-event-channel-with-mpeg-h-audio/>



**Figure 42. Setup of the UHD FTV Channel using MPEG-H Audio during French Tennis Open**

The Ultra HD signal and the associated audio were then transmitted over IP by a single-mode optical fiber from the venue to France Télévisions headquarter. As shown in [Figure 43](#), the MPEG-H Audio system allowed for seamless ad-insertion during the tournament. Pre-recorded clips were played back during the break of the live feed and every transition was seamless for the viewers.

It is the final control room of France Télévisions, the CDE (Center of Diffusion and Exchange) which recovered the signal in order to realize the switch of live and replays, trailers, advertisements, insertion of logos, scrolling texts, etc. then adapted the characteristics of this chain to the different distribution networks.

As shown in [Figure 43](#) the MPEG-H Audio feed was demonstrated at the Philippe Chatrier court, where the various personalization features could be experienced:

- Switch between different mixes: default mix, dialog enhanced version as well as a venue version without a commentator;
- Enhance the dialog level or even change the position of the dialog in the 3D space.



**Figure 43. MPEG-H Audio during French Tennis Open 2019: Demonstrations at the Philippe Chatrier court**

All interactivity options were controlled by the sound engineer in the OB-van using Jünger's MMA. In the control room at the FTV broadcast center (see [Figure 44](#)), several ATEME Titan encoders have been used for encoding the various feeds. Additionally, a soundbar have been installed for reproducing the MPEG-H immersive and interactive sound.

The MPEG-H Audio stream was included in the regular live Ultra HD broadcast of the Philippe Chatrier court.



**Figure 44. MPEG-H Audio during French Tennis Open 2019: Control Room**

### **Multiple MPEG-H Audio programs at the European Athletics Championships:**

In 2018 the EBU, together with EBU members BBC, France Télévisions, IRT, RAI and ZDF, as well as several technology provider partners, conducted a complex broadcast trial in Europe at the European Athletics Championships in Berlin<sup>31</sup>. The trial demonstrated the world's first live production and distribution of Ultra High Definition (UHD) content, with High Frame Rates (HFR), High Dynamic Range (HDR), and Next Generation Audio (NGA). The Fraunhofer IIS team, together with technology partners, supported the onsite sound engineers of EBU member stations in order to familiarize them with the features and the new possibilities of MPEG-H Audio.

The various feeds were multiplexed for a live transmission via the Eurovision satellite network to RAI's experimental test bed in the Aosta Valley, Italy, and via the Eurovision fiber infrastructure to the European Championships Broadcast Operations Centre (BOC) based at BBC Glasgow.

MPEG-H Audio was used for several different feeds and the Fraunhofer IIS engineering team worked closely with technology partners ATEME, Jünger Audio, MediaKind, Kai Media, b<>com and Qualcomm to demonstrate the complete set of features and capabilities of the MPEG-H Audio system (see [Figure 38](#)).

<sup>31</sup> SMPTE Periodical – European Athletics Championship: Lessons from a Live, HDR, HFR, UHD, and Next-Generation Audio Sports Event, <https://ieeexplore.ieee.org/document/8673671>

MPEG-H Audio Trial During European Athletics Championships 2018, <https://www.audioblog.iis.fraunhofer.com/mpeg-h-trial-eac-ebu>





**Figure 45. European Athletics Championships 2018 in Berlin: Control Room**

Various devices from different manufacturers supported MPEG-H Audio work in the real-time chain and were fully interoperable, including authoring and monitoring units and post production plug-ins, contribution links as well as emission and playback devices, and demonstrating the benefit of open standards.

As illustrated in [Figure 46](#), three different MPEG-H Audio services were running in parallel, with content produced and delivered live, using available equipment from different manufacturers operated in existing SDI-based broadcast facilities.

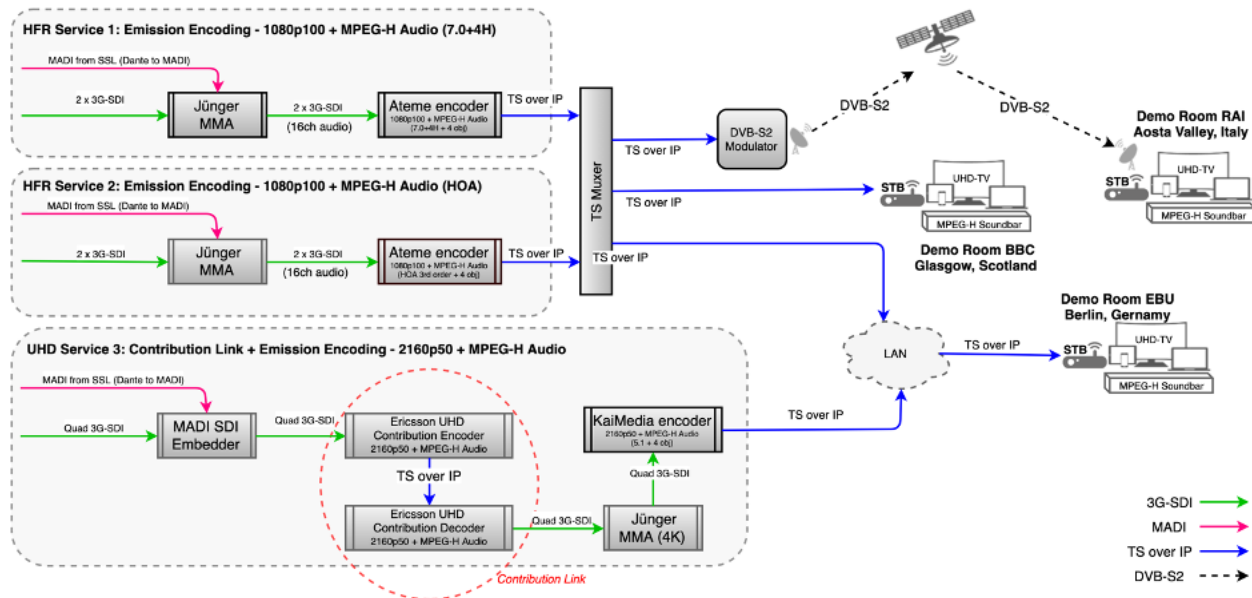


Figure 46. MPEG-H Audio Transmission setup during European Athletics Championships 2018

### Service 1: “MPEG-H Audio with 1080p100 HFR and HLG video”:

The main MPEG-H Audio service was designed to show the benefit of immersive and interactive sound together with High Frame Rate (HFR) and High Dynamic Range (HDR) video for sport events. This offered the viewer a truly immersive experience. The NGA mix (7.0+4H) was created by BBC, France Télévisions, RAI and ZDF sound engineers using SSL's new console with support for 3D Immersive Audio.

In addition to the immersive sound bed, four interactive mono object signals were used for two commentaries and two audio descriptions in English and French. The live metadata authoring of the MPEG-H immersive audio content was made possible through the use of the Multichannel Monitoring and Authoring (MMA) system provided by Jünger Audio. The personalization and interactivity options were enabled through the MPEG-H Audio metadata, which treats individual elements of an audio mix as objects that can be adjusted by the viewer afterwards.

For the live broadcast, the TITAN encoder was provided by Ateme, which supports MPEG-H Audio and enabled the use of immersive and personalization features of NGA in the program production and distribution.



---

**Service 2: “MPEG-H Audio (HOA) with 1080p100 HFR and HLG video”:**

Together with Qualcomm, a second MPEG-H service was provided, using scene-based production tools offered by b<>com. The MPEG-H Audio native support for Higher Order Ambisonics (HOA) allows the use of a scene-based bed instead of a channel-based one, so for this service the immersive mix was created using a scene-based (3rd order Ambisonics) bed together with the same additional four objects.

Similar to the main service, a second Jünger MMA system was used for metadata authoring, together with an additional Ateme encoder for the emission.

**Service 3: “MPEG-H Audio with 2160p50 HLG video, including a contribution chain”:**

The third service enabled the usage of a contribution link based on MPEG-H Audio before the UHD emission encoder. The purpose of a contribution link is to provide compression which is relatively transparent and provides robust and secure transport of the metadata, allowing downstream manipulation of the signals and of the metadata. The contribution link consisted of the AVP 2000 Contribution Encoder and MediaKind Content Processing (MKCP) Decoder, both devices manufactured by MediaKind. The contribution encoder provides certain authoring capabilities, allowing the broadcaster to manipulate the metadata to be fed into the bitstream. This is done through the GUI of the encoder.

For the emission part, the Kai Media emission encoder was used due to the Korean company encoder manufacturer having integrated MPEG-H Audio for use in the deployment of ATSC 3.0 UHD in Korea. Further, this encoder enabled DVB transmissions by incorporating MPEG2-TS for the transport layer.

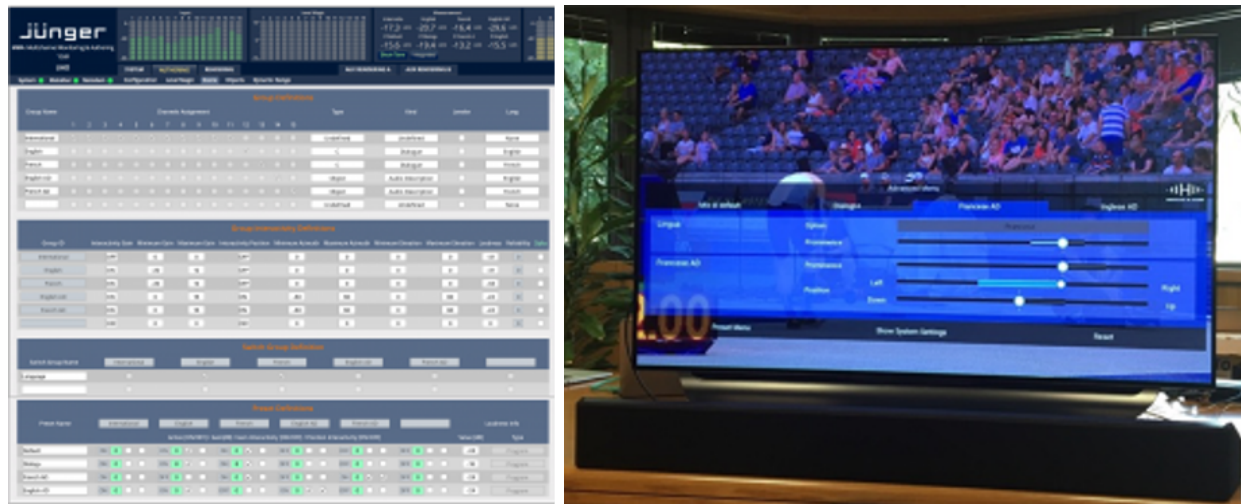
A third MMA system was provided by Jünger Audio and used between the contribution decoder and the emission encoder for monitoring of the audio scene. The service also provided a different audio setup compared to the other two MPEG-H programs, using a 5.1 downmix together with the two commentaries and audio descriptions provided in English and French.

**Metadata Authoring and Playback:**

For all three services, the MPEG-H metadata that was authored enabled users to personalize their audio experience. First of all, they had the option to choose between different versions of the content (presets): a default preset, a dialog enhancement preset and two presets for the audio description. In addition, various interactivity features such as level and position interactivity, were enabled for each preset. During production, the sound engineers had the opportunity to change or adjust live all interactivity features using the MMA systems (see [Figure](#)



47, left). They were able to create new presets on the fly, change the interactivity ranges for any available audio object, disable interactivity for some presets, or even completely switch to a different configuration in a seamless way.



**Figure 47. European Athletics Championships 2018 in Berlin: Control Room**

The MPEG-H programs were demonstrated in Berlin, Glasgow and the Aosta Valley, using an immersive soundbar that was able to decode and playback the MPEG-H streams and thus generating an immersive experience without the need for loudspeakers all around the room (see [Figure 47](#), right). Additionally, with MPEG-H Audio the viewers could switch between different versions of the content, select their preferred language and level of dialogue and use the advanced menus to adjust the balance between the stadium sound and commentary to their personal taste. Moreover, during the demonstration the users were also able to move the audio description “away from the screen” to the left/right or above, resulting in a higher separation from the main dialog localized in the center channel.

**(End of Blue Book)**