



Ultra HD Forum: Phase A Guidelines

September 14, 2018

Revision: 1.5

Ultra HD Forum

8377 Fremont Blvd., Suite 117,

Fremont, CA 94538

UNITED STATES



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 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. 2016, 2017

Revision History

Version	Date
Revision 1 released for select external review	April 18, 2016
Revision 1.1 released	July 15, 2016
Revision 1.2 released	December 16, 2016
Revision 1.3 released	April 24, 2017
Revision 1.4 released	August 25, 2017
Revision 1.5 released	September 14, 2018



Table of Contents

1. PURPOSE AND SCOPE.....	7
2. REFERENCES	9
2.1 Reference List	9
2.2 Brief Summary of ITU-R BT.709, BT.2020, and BT.2100	11
3. TERMS AND ACRONYMS	13
3.1 Terms	13
3.2 Acronyms and Abbreviations	15
4. PHASES AND TIMEFRAMES	17
4.1 Phase A	17
4.2 Future Phases	20
5. USE CASES.....	22
5.1 MVPD Platform Delivery	22
5.2 IP Network Delivery	22
6. PRODUCTION AND POST PRODUCTION.....	24
6.1 HDR/WCG Technologies	25
6.1.1 Perceptual Quantization (PQ) and PQ10	25
6.1.2 Hybrid Log-Gamma (HLG) and HLG10	26
6.1.3 Recommendation ITU-R BT.2100	26
6.1.4 Static Metadata – SMPTE ST 2086, MaxFALL, MaxCLL	26
6.1.5 HDR10	27
6.1.6 UHD Phase A HDR Technologies	27
6.1.7 HDR10 Metadata Generation	28
6.1.8 HDR10 Metadata Carriage	28
6.1.9 Signaling Transfer Function, Color Space and Matrix Coefficients	29
6.1.10 Peak Brightness – Consideration of Production, Reference Monitors, Consumer Displays and Archive Content	31
6.1.11 Studio Video over IP	32
6.2 Production for Pre-recorded Content	33
6.2.1 Camera Requirements	34
6.2.2 Reference Monitor	35
6.2.3 On-Set / Near-Set Monitoring	36
6.2.4 Color Grading	36
6.2.5 Channel-based Immersive Audio Post Production	37
6.3 Production for Live Content	38
6.3.1 Live Production in Trucks or Studio Galleries	39
6.3.2 Production with encodings other than PQ and HLG	39
6.3.3 Channel-based Immersive Audio Production	40
7. SECURITY	43
7.1 Content Encryption	43
7.2 Forensic Watermarking	44
7.2.1 Introduction	44
7.2.2 Use Cases	44
7.2.3 Distribution	45
7.2.4 One-Step Watermarking	46



7.2.5	Two-Step Watermarking Integration	46
7.2.6	Use Case: ABR VOD	51
8.	REAL-TIME PROGRAM SERVICE ASSEMBLY	53
8.1	Maintaining Dynamic Range and Color Space Parameters	53
8.2	Conversion from SDR/BT.709 to PQ10/HLG10	53
8.3	Conversion between Transfer Functions	56
8.4	Conversion from PQ10/HLG10 to SDR/BT.709	56
9.	DISTRIBUTION	58
9.1	Production Processing and Contribution	59
9.1.1	Video	60
9.1.2	Audio	62
9.1.3	Closed Captions and Subtitles	62
9.2	Broadcast Center Processing and Primary Distribution	62
9.3	Final Distribution from MVPD/OTT Provider Processing	65
9.3.1	Bit Depths	65
9.3.2	Video	65
9.3.3	Adaptive Bitrate (ABR) Streaming	67
9.3.4	Audio	68
9.3.5	Closed Captions and Subtitles	68
9.4	Transport	68
10.	DECODING AND RENDERING	70
10.1	Decoding	70
10.2	Rendering	70
10.3	Overlays Inserted at the Consumer Device	71
11.	FORMAT INTEROPERABILITY	72
11.1	Legacy Display Devices	73
11.2	Down-conversion at the Service Provider	73
11.3	Down-conversion at the STB	74
11.4	Spatial Resolution Up-conversion of Legacy Services	76
11.5	Interoperability of Atmos Channel-based Immersive Audio	78
12.	ANNEX A: REAL WORLD PHASE A UHD DEPLOYMENTS.....	79
12.1	CBS and DirecTV Major Golf Tournament	79
12.2	Amazon Major Parade	80
12.3	NBC Universal Olympics and 2018 World Cup	80
12.4	Sky/BBC Royal Wedding and Major Tennis Tournament	81
13.	ANNEX B: OVER-THE-AIR BROADCAST USE CASE.....	84
14.	ANNEX C: IC_TC_P COLOR REPRESENTATION.....	85
15.	ANNEX D: ACES WORKFLOW FOR COLOR AND DYNAMIC RANGE	86
16.	ANNEX E: ISO 23001-12, SAMPLE VARIANTS.....	88



Index of Tables and Figures

Table 1 Summary Comparison of ITU-R BT.709, BT.2020, and BT.2100	12
Table 2 UHD Phase A Workflow Parameters	18
Table 3 UHD Phase A Content Parameters	18
Table 4 Decoder Capabilities	19
Table 5 Service Formats	20
Table 6 Phase A UHD over IP Networks	23
Table 7 Signaling Transfer Function, Color Space and Matrix Coefficients	30
Table 8 SMPTE Registry UL's for Transfer Function, Color Primaries and Coding Equations ..	31
Table 9 Pre-recorded Content Interface Descriptions	34
Table 10 Compression and Distribution Nodes and Interfaces	58
Table 11 SDI Input Standards for 2160p Content	59
Table 12 Contribution Bitrates and Key Parameters	61
Table 13 Primary Distribution Bitrates and Key Parameters	64
Table 14 Existing Practices for Real-Time Program Service Distribution Formats	65
Table 15 Final Distribution Bitrates and Key Parameters	66
Table 16 Example "Real World" Bitrates as early as 2016	66
Table 17 Example Bitrates for Video Streams	68
Table 18 ACES Workflow Model	86
Figure 1 Content Production and Distribution Workflow	25
Figure 2 Pre-recorded Content Production Workflow and Interfaces	33
Figure 3 Channel-based Immersive Audio Post-Production	37
Figure 4 Channel-based Immersive Audio Live Production	41
Figure 5 Illustration of Watermark Identifier	44
Figure 6 One-Step Watermark Performed on the Client Side	46
Figure 7 Two Examples of Two-step Watermarking Systems	47
Figure 8 Watermark Embedding Using a Unique Variant Sequence	48
Figure 9 Transport at the Media Layer Using MPEG SEI NALUs	49
Figure 10 Transport at the Container Layer Using a Track in an ISO/BMFF File	50
Figure 11 Pre-processing in the Baseband Domain for Two-step Watermarking	51
Figure 12 ABR Playlist Serialization	52
Figure 13 Sample Live Workflow with Mixed Format Source Content	55
Figure 14 Distribution Nodes and Interfaces	58
Figure 15 Contribution Workflow	60
Figure 16 Primary Distribution to an MVPD or Station	63
Figure 17 Primary Distribution to an OTT Provider	64
Figure 18 Down-conversion at the Headend	74
Figure 19 Down-conversion at the STB	75
Figure 20 Spatial Resolution Up-conversion of Legacy Services	76
Figure 21 CBS and DirecTV Major Golf Tournament Workflow	79
Figure 22 Amazon Major Parade Workflow	80
Figure 23 NBCU Olympics and 2018 World Cup UHD Workflow	81
Figure 24 Sky/BBC Royal Wedding and Major Tennis Tournament Workflow	82
Figure 25 ACES Workflow Model	86



Figure 26 Transport at the Container Layer Using a Track in an ISOBMFF File.....	88
--	----



1. Purpose and scope

The purpose of this document is to describe consistent methods for the creation and delivery of Ultra HD content for consumer distribution along with a uniform set of characteristics that may be combined to produce content that can be considered “Ultra HD,” referred to as “UHD Phase A” in this document. (See Table 3 for characteristics of UHD Phase A services.) The scope includes delivery via the Internet, satellite, terrestrial broadcast and cable as transmission methods. It does not include encode and delivery of content via storage media, such as Blu-ray® disc, HDD, SCSA devices, or similar, nor does it include encode and delivery of Digital Cinema content.

The goal is to create consistency across the industry for ensuring interoperability and a high quality experience for consumers. While this document provides context with respect to content creation, the primary purpose of this document is to define guidelines for proper delivery of UHD Phase A content from the studio producer or the live event to the consumer via a linear (real-time) service.

This document recommends profiles and practices to be used across each of the elements in a distribution chain to maximize end-to-end interoperability. References supporting the recommendations are provided to the extent possible. However, in many cases, industry practices are advancing more quickly than existing documentation. In the cases where technologies are in the process of being developed and/or standardized, these guidelines also provide associated time line expectations where possible. All the recommendations represent the consensus view of the Ultra HD Forum based on these references, its members’ expertise and experience, and/or results from Ultra HD Forum Interop events.

The Ultra HD Forum intends the Phase A Guidelines to be a "living document" and plans to release new revisions as more data becomes available, more learning is accumulated across deployments, and more interops and trials are conducted, while keeping the same scope for the document.

For the purpose of this document, the Ultra HD Forum is considering the following UHD Phase A content and service types, which have different workflow characteristics:

- Content Types:
 - Live content – content that is distributed to consumers in real-time as it is produced, such as sports, news, awards shows, reality programs, talent shows, debates, etc. Live production workflows do not include a post-production step and creative intent is set in the control room or truck. Note that content produced in this manner may also be captured for subsequent re-broadcast.
 - Pre-recorded content – content that is fully produced prior to distribution to consumers, such as sitcoms, dramas, advertisements, documentaries, etc. Pre-recorded production workflows include post-production steps, and creative intent is set during post-production.
- Service Types:
 - Real-time Program Services – services consisting of a linear, pre-scheduled stream of content that is assembled in real-time for distribution to consumers such as a broadcast television channel, cable network, etc. Real-time Program Services are comprised of Live and/or Pre-recorded content and may also include graphic overlays, such as station logos, emergency text crawls, etc.



- On-Demand Services – While services such as Hulu, Netflix and MVPD VOD are largely out of scope, if the content offered on these platforms was originally Live and recorded for later distribution then these Guidelines may be relevant. See below for further explanation..

The initial and primary focus of the Ultra HD Forum is on Real-time Program Services because they may be the most challenging for an Ultra HD end-to-end workflow. On-Demand Services are largely out of scope. However, guidelines contained herein related to producing Live content may be valuable to On-Demand service providers who are offering content that was originally produced for Live distribution using a linear workflow, and has been repurposed as a VOD asset, e.g., via caching at the point of final distribution, for start-over, catch-up, or trick play. With this in mind, the scope of this document is defined as follows:

In scope:

- Pre-recorded and Live content production
 - Cameras
 - Monitoring
 - Color grading for Pre-recorded content
 - HDR/WCG technologies
 - Channel-based Immersive Audio
- Metadata
- Security
- Distribution and Compression
 - Content master format
 - Content mezzanine format
 - Encoding codecs, methods and recommendations
 - Transcode codecs, methods and recommendations
 - Approximate ranges of bitrates through the entire processing chain
 - Distribution and transport methods
- Real-time program stream assembly
- Conversion between SDR and HDR formats and among different HDR formats
- Interface guidelines for connecting systems and functions throughout the production and delivery of the content
- Backward compatibility for legacy systems

Out of scope:

- Filming techniques (e.g., lighting, camera settings, etc.)
- TV settings
- Encoder settings
- Subjective analysis of overall content quality
- TV technology guidelines (e.g., OLED vs. Quantum Dots)
- Color grading guidelines (e.g., luma, saturation and contrast preferences)
- Fixed media delivery and digital cinema



2. References

This section contains references used in this text, which are an essential component of these guidelines. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties are encouraged to investigate the applicability of the most recent editions of the materials listed in this section.

2.1 Reference List

- [1] Ultra HD Forum: “UHD Phase B Guidelines – revision 1.0,” April 7, 2018, <https://ultrahdforum.org/resources/phaseb-guidelines-description/>
- [2] Recommendation ITU-T H.222.0 | ISO/IEC 13818-1:2000, “Information Technology—Generic coding of moving pictures and associated audio information - Part 1: Systems”
- [3] Recommendation ITU-R BT.709, “Parameter values for the HDTV standards for production and international programme exchange”
- [4] Recommendation ITU-R BT.2020, “Parameter values for ultra-high definition television systems for production and international programme exchange”
- [5] Recommendation ITU-R BT.1886, “Reference electro-optical transfer function for flat panel displays used in HDTV studio production”
- [6] Recommendation ITU-R BT.2100-1, “Image parameter values for high dynamic range television for use in production and international programme exchange”, <http://www.itu.int/rec/R-REC-BT.2100>
- [7] Report ITU-R BT.2390-2, “High dynamic range television for production and international programme exchange” , <https://www.itu.int/pub/R-REP-BT.2390-2016> (companion report to ITU-R Recommendation BT.2100)
- [8] Recommendation ITU-R BT.2087-0, “Colour conversion from ITU-R BT.709 [3] to ITU-R BT.2020 [4]”
- [9] Recommendation ITU-R BT.2408, “Operational Practices in HDR Television Production”
- [10] SMPTE ST 2084, “High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays”
- [11] SMPTE ST 2086, “Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images”
- [12] SMPTE HDR Study Group report, <https://www.smpte.org/sites/default/files/Study%20Group%20On%20High-Dynamic-Range-HDR-Ecosystem.pdf>
- [13] DVB UHD-1 Phase 1 specification, http://www.etsi.org/deliver/etsi_ts/101100_101199/101154/02.01.01_60/ts_101154v020101p.pdf
- [14] DVB DASH specification, https://www.dvb.org/resources/public/standards/a168_dvb-dash.pdf
- [15] Recommendation ITU-R BT.814, “Specifications and alignment procedures for setting of brightness and contrast on displays”



- [16] HDMI 2.0a specification,
http://www.hdmi.org/manufacturer/hdmi_2_0/hdmi_2_0a_faq.aspx
- [17] DASH-IF Interoperability Points: Guidelines for Implementation, version 4.2
- [18] DCI Specification, Version 1.2 with Errata as of 30 August 2012 Incorporated
- [19] CTA-608-E R-2014, “Line 21 Data Services”, <http://www.cta.tech/Standards/Standard-Listings/R4-3-Television-Data-Systems-Subcommittee.aspx>
- [20] CTA-708-E (ANSI), “Digital Television (DTV) Closed Captioning” ,
<http://www.cta.tech/Standards/Standard-Listings/R4-3-Television-Data-Systems-Subcommittee.aspx>
- [21] ETSI 300 743, “Digital Video Broadcasting (DVB); Subtitling systems”,
http://www.etsi.org/deliver/etsi_en/300700_300799/300743/01.03.01_60/en_300743v010301p.pdf
- [22] ETSI 300 472, ”Digital Video Broadcasting (DVB); Specification for conveying ITU-R System B Teletext in DVB bitstreams”,
http://www.etsi.org/deliver/etsi_en/300400_300499/300472/01.03.01_60/en_300472v010301p.pdf
- [23] SCTE-27, “Subtitling Methods for Broadcast Cable”
http://www.scte.org/documents/pdf/standards/SCTE_27_2011.pdf
- [24] W3C: “TTML Text and Image Profiles for Internet Media Subtitles and Captions (IMSC1)”, [Candidate] Recommendation, W3C, www.w3.org.
- [25] ATSC: “Techniques for Establishing and Maintaining Audio Loudness for Digital Television,” Doc. A/85, Advanced Television Systems Committee, Washington, D.C., 12 March 2013, <http://atsc.org/recommended-practice/a85-techniques-for-establishing-and-maintaining-audio-loudness-for-digital-television/>
- [26] ISO/IEC: “MPEG-4 -- Part 10: Advanced Video Coding,” Doc. ISO/IEC 14496-10¹
- [27] ISO/IEC: “Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 2: High efficiency video coding, ” Doc. ISO/IEC 23008-2:2015²
- [28] ISO/IEC 14496-3:2009, “Informational Technology, Coding of Audio-Visual objects, Part 3: Audio”
- [29] ISO/IEC 14496-12:2015, "Information technology—Coding of audio-visual objects—Part 12: ISO base media file format”
- [30] ETSI 102 366 v1.3.1 (2014-08), “Digital Audio Compression (AC-3, Enhanced AC-3) Standard”
- [31] SMPTE RP 431-2, “D-Cinema Quality — Reference Projector and Environment”
- [32] CTA 861-G, “A DTV Profile for Uncompressed High Speed Digital Interfaces”
http://www.techstreet.com/standards/cta-861-g?product_id=1934129
- [33] W3C: “HTML5, A vocabulary and associated APIs for HTML and XHTML”
<https://www.w3.org/TR/html5/>
- [34] ETSI TS 103 433-1 v1.2.1 (2016-08) "High-Performance Single Layer Directly Standard Dynamic Range (SDR) Compatible High Dynamic Range (HDR) System for use in Consumer

¹ Also published by ITU as ITU-T Recommendation H.264.

² Also published by ITU as ITU-T Recommendation H.265: 2015.



- Electronics devices (SL-HDR1)", http://www.etsi.org/deliver/etsi_ts/103400_103499/10343301/01.02.01_60/ts_10343301v010201p.pdf
- [35] ETSI TS 103 420, "Object-based audio coding for Enhanced AC-3 (E-AC-3)".
 - [36] SMPTE ST 337, "Format for Non-PCM Audio and Data in AES 3 Serial Digital Audio Interface"
 - [37] Recommendation ITU-R BS.1770-4, "Algorithms to measure audio programme loudness and true-peak audio level"
 - [38] SMPTE ST 2067-21, "Interoperable Master Format — Application #2E," 2016
 - [39] "Guidelines for Implementation: DASH-IF Interoperability Points v3.1," <http://dashif.org/wp-content/uploads/2015/10/DASH-IF-IOP-v3.1.pdf>
 - [40] MovieLabs, "MovieLabs Specification for Enhanced Content Protection, v1.1"
 - [41] ISO/IEC 23001-12:2015, "Information technology -- MPEG systems technologies - Part 12: Sample Variants in the ISO base media file format"
 - [42] ETSI TS 102 034 v1.4.1 (2009-08), "Transport of MPEG-2 TS Based DVB Services over IP Based Networks"
 - [43] Internet Engineering Task Force (IETF) RFC 3550, "RTP: A Transport Protocol for Real-Time Applications", available at <https://www.ietf.org/rfc/rfc3550.txt>
 - [44] SMPTE ST 2110-10:2017, "Professional Media over IP Networks: System Timing and Definitions"
 - [45] SMPTE ST 2110-20:2017, "Professional Media over IP Networks: Uncompressed Active Video"
 - [46] SMPTE ST 2110-21:2017, "Professional Media over IP Networks: Traffic Shaping and Delivery Timing for Video"
 - [47] SMPTE ST 2110-30:2017, "Professional Media over IP Networks: PCM Digital Audio"
 - [48] SMPTE ST 2110-40:2018, "Professional Media over IP Networks: SMPTE ST 291-1 Ancillary Data"
 - [49] SMPTE ST 2108-1:2018, "HDR/WCG Metadata Packing and Signaling in the Vertical Ancillary Data Space"
 - [50] OpenCable Specifications, OC-SP-EP-I01-130118: "Encoder Boundary Point Specification" (01/2013). <https://apps.cablelabs.com/specification/encoder-boundary-point-specification/>
 - [51] SMPTE ST 2065-1:2012, "Academy Color Encoding Specification (ACES)"

2.2 Brief Summary of ITU-R BT.709, BT.2020, and BT.2100

The Ultra HD Forum Guidelines document refers to ITU-R BT.709 [3], BT.2020 [4], and BT.2100 [6] that address transfer function, color space, matrix coefficients, and more. The following table is a summary comparison of those three documents. Please note that this is merely a summary, and the documents contain considerably more information.



Table 1 Summary Comparison of ITU-R BT.709, BT.2020, and BT.2100

	ITU-R BT.709	ITU-R BT.2020	ITU-R BT.2100
Spatial Resolution	HD	UHD, 8K	HD, UHD, 8K
Framerates*	24, 25, 30, 50, 60	24, 25, 30, 50, 60, 100, 120	24, 25, 30, 50, 60, 100, 120
Interlace/Progressive	Interlace, Progressive	Progressive	Progressive
Color Space**	BT.709	BT.2020	BT.2020
Dynamic Range	SDR (BT.1886 [5])	SDR (BT.1886 [5])	HDR (PQ and HLG)
Bit Depth	8, 10	10, 12	10, 12
Color Representation	RGB, YC _B C _R	RGB, YC _B C _R	RGB, YC _B C _R , IC _T C _P

Table 1 notes:

*Framerates include both integer and fractional values (including 120/1.001 for BT.2020 and BT.2100).

**Refer to the ITU-R documents for actual color primary values.



3. Terms and Acronyms

3.1 Terms

This guideline contains the following terms and definitions:

Adaptive Bit Rate	A technique used in streaming multimedia over computer networks, in which multiple versions of a single content source are provided, each encoded at different bitrates; the client device monitors the available bandwidth and CPU capacity in real time, and switches between streaming the different encodings, choosing the highest bitrate (i.e., highest quality) according to available resources.
Bit Depth	The number of bits used per component. It describes the number of increments for both brightness and color.
Color Gamut	The subset of colors that can be accurately represented within a given color space, or by a certain output device.
Color Space	A specific organization of colors, allowing for their reproducible representation. BT.2020 [4] and BT.709 [3] are examples of color spaces.
Color Volume	Combined color gamut and luminance characteristics.
Color Volume Transform	A technique used to map a coordinate in one color volume to a coordinate in another color volume.
DCI-P3	Color gamut defined in SMPTE RP 431-2 [31].
Electro-Optical Transfer Function	The transfer function that maps digital pixel values to values of display light.
Forensic Watermarking	Forensic Watermarking is a technology that modifies multimedia content (e.g., a video, a song, a piece of text) to encode a Watermark Identifier without introducing artifacts that would be perceptible by a human being. The Watermark Identifier encoded by a Forensic Watermark can be recovered even if the content is altered after the watermarking operation.
HLG10	The Hybrid Log-Gamma OETF described in BT.2100 [6] together with BT.2020 [4] color space and 10-bit depth. ³ [see also Section 6.1.2]
HDR10	A PQ10-based format further capable of providing SMPTE ST 2086, MaxFALL, and MaxCLL metadata (see also Section 6.1.5).
High Dynamic Range	Greater than or equal to the contrast ratio that could be derived from 13 f-stops.
High Frame Rate	Content with a relative rate greater than 24 frames per second for motion pictures and greater than 60 fps for television content.

³ Note: HLG10, as used in DVB [13] is further limited to the Non-Constant Luminance Y'C_BC_R signal format and narrow range quantization as defined in [6].



Hybrid Log-Gamma	Hybrid Log-Gamma OETF, EOTF, and OOTF as defined in BT.2100 [6].
Immersive Audio	An audio system that enables high spatial resolution in sound source localization in azimuth, elevation and distance, and provides an increased sense of sound envelopment.
MaxCLL	Maximum Content Light Level – greatest pixel light value of any video frame in the program.
MaxFALL	Maximum Frame-Average Light Level – greatest average pixel light value of any video frame in the program.
Modulation Transfer Function	The contrast performance of an optical system such as a lens as a function of spatial frequency.
Multichannel Video Programming Distributor	A service provider that delivers video programming services, usually for a subscription fee (pay television).
Nit	Unit of luminance measurement, weighted by the human visual system, formally specified in “candela per meter squared” (cd/m^2); the term “nits” is used in this document for convenience.
Opto-Electronic Transfer Function	The transfer function that maps scene light captured by the camera into digital pixel values.
Opto-optical Transfer Function	The overall transfer function that maps scene light captured by the camera to light values produced by the display.
Perceptual Quantization	A high dynamic range EOTF used for HDR. PQ is specified in BT.2100 [6].
PQ10	The Perceptual Quantization EOTF described in BT.2100 [6] together with BT.2020 [4] color space and 10-bit depth (see also Section 6.1.1). ⁴
Resolution	The number of vertical and horizontal pixels available on a display device.
Set of Variants	A Set of Variants is a collection of Variants for a given segment of a multimedia asset. Variants contain the same perceptual content but different marks and can be used interchangeably. Sets of Variants for a given asset are typically generated during the first step in a two-step watermarking system.
Standard Dynamic Range	Content graded as per BT.1886 [5] and BT. 709 [3] for HD television.
Variant	A Variant is an alternative representation of a given segment of a multimedia asset. Typically, a Variant is a pre-watermarked version of the segment using a Forensic Watermarking technology. The size of the segment varies for different Forensic Watermarking technologies: a few bytes, a frame, a group of pictures, a video fragment.

⁴ Note: PQ10, as used in DVB [13] is further limited to the Non-Constant Luminance $Y'C'_B C'_R$ signal format and narrow range quantization as defined in [6].



Variant Sequence Generator	A Variant Sequence Generator (VSG) selects a single Variant in each Set of Variants to produce a Variant Sequence. The VSG is part of the second step in a two-step watermarking system.
Variant Sequence	A Variant Sequence is a sequence of Variants that encodes a desired Watermark Identifier.
UHD-1	UHD at resolution of 3840 H by 2160 V (this is a 4K resolution).
UHD-2	UHD at resolution of 7680 H by 4320 V (this is an 8K resolution).
UHD Phase A	Term used in this document to for content that conforms to the parameters shown in Table 3.
Wide Color Gamut	Color gamut wider than the gamut of BT.709 [3].
Watermark Identifier	A serialization number that is embedded in a multimedia asset using a Forensic Watermarking technology to make the asset unique. Examples of data used as a Watermark Identifier are session IDs, client IDs, device IDs, firmware versions, timestamps, etc. The Watermark Identifier is also routinely referred to as the <i>payload</i> or the <i>message</i> in the watermarking literature. See also Figure 5.

3.2 Acronyms and Abbreviations

ABR	Adaptive Bit Rate
ACES	Academy Color Encoding System
AVC	Advanced Video Coding
CA	Conditional Access
CDN	Content Delivery Network
CG	Character Generator
CGI	Computer Generated Imagery
DASH	Dynamic Adaptive Streaming over HTTP
DOCSIS	Data Over Cable Service Interface Specification
DVE	Digital Video Effects
EMB	Watermark EMBedder
ENC	Video ENCoder
EOTF	Electro-Optical Transfer Function
EPB	Encoder Boundary Point
HD	High Definition
HDR	High Dynamic Range
HEVC	High Efficiency Video Coding
HFR	High Frame Rate



HLG	Hybrid Log-Gamma
HLS	HTTP Live Streaming
HTTP	Hyper Text Transfer Protocol
IP	Internet Protocol
IPTV	Internet Protocol Television
ISOBMFF	ISO Base Media File Format [29]
LUT	Look Up Table
MPD	Media Presentation Description
MTF	Modulation Transfer Function
MVPD	Multichannel Video Programming Distributor
NALU	Network Abstraction Layer Unit
NGA	Next Generation Audio
OETF	Opto-Electronic Transfer Function
OOTF	Opto-Optical Transfer Function
OTA	Over-the-Air (i.e., terrestrial transmission of content)
OTT	Over-the-Top (i.e., Internet-based transmission of content)
PES	Packetized Elementary Stream
PQ	Perceptual Quantization
PVR	Personal Video Recorder
RTP	Real-Time Transport Protocol
SD	Standard Definition
SEI	Supplemental Enhancement Information
SDR	Standard Dynamic Range
STB	Set Top Box
TSD	Transport Stream Decoder
UDP	User Datagram Protocol
UHD	Ultra High Definition (see “UHD Phase A” in Section 2.2 above for use of this term within the scope of this document)
URI	Uniform Resource Identifier
VSG	Variant Sequence Generator
VOD	Video-on-Demand
WCG	Wide Color Gamut
WM	WaterMark
WM ID	Watermark Identifier
xDSL	Digital Subscriber Line (x indicates any variety, e.g., ADSL, HDSL, SDSL, etc.)



4. Phases and Timeframes

This document addresses UHD Phase A methods and technologies, i.e., those that were deployable as early as 2016 timeframe, which is defined as “Phase A.”

The high-level media attributes for new services included in the Phase A timeframe are listed in Section 4.1. Future phases are out of scope of this document; however, some high level guidance is suggested in Section 4.2 below and in Annex B-E. Detailed recommendations are grouped into Production / Post-Production, Distribution (includes compression and distribution for contribution, primary and final delivery stages), and Decoding / Rendering.

The Ultra HD Forum applies the following criteria to select recommended technologies for a given Phase:

1. Viability is near term based on sufficiently low technical complexity
2. There is broad interest in the technology within the industry, including from service providers
3. The technology enables Real-time Program Services consisting of Live and/or Pre-recorded content elements
4. The technology is expected to be commercially available *or* described in a reasonably accessible specification within the given Phase timeframe

4.1 Phase A

UHD Phase A content and services are those that were distributed as early as 2016 via OTT or MVPD. (OTA broadcast services were not commercially offered in 2016.)

For the purposes of this document, UHD Phase A includes the following characteristics:

- Resolution – greater than or equal to 1080p and lower than or equal to 2160p, (progressive format; BT.2100 [6] does not include interlaced formats)
- Wide Color Gamut – color gamut wider than BT.709 [3]
- High Dynamic Range – greater than or equal to the contrast ratio that could be derived from 13 f-stops of dynamic range
- Bit depth – 10-bit
- Frame rates – up to 60fps (integer frame rates are preferred; note that cinematic content may opt to use lower frame rates, e.g., see DCI specification)
- Audio – 5.1 channel surround sound or channel-based Immersive Audio⁵ (2.0 stereo is possible; however, 5.1 or channel-based Immersive Audio are preferred for a premium experience; see also Ultra HD Forum Phase B Guidelines [1] for details on next-generation audio)
- Closed Captions/Subtitles – CTA 708/608, ETSI 300 743, ETSI 300 472, SCTE-27, IMSC1

⁵ Note that Immersive Audio can also include object- and scene-based audio as well as channel-based audio; see Ultra HD Forum Phase B Guidelines [1] for details.



The following terms are used in this document for HDR and HDR plus WCG:

- HLG: The Hybrid Log-Gamma OETF defined in BT.2100 [6]
- HLG10: The Hybrid Log-Gamma OETF described in BT.2100 [6] together with BT.2020 [4] color space and 10-bit depth.
- PQ: The Perceptual Quantization EOTF defined in BT.2100 [6]
- PQ10: The Perceptual Quantization EOTF described in BT.2100 [6] together with BT.2020 [4] color space and 10-bit depth
- HDR10: The Perceptual Quantization EOTF with BT.2020 [4] color space, 10-bit depth, and ST 2086 [11] static metadata, and the MaxCLL and MaxFALL static metadata [27]; an HDR10 system or format is capable of handling these static metadata when present

Table 2 UHD Phase A Workflow Parameters

Content Creation & Mastering	Defined and documented standard workflows for Live and Pre-recorded content
Service Type	Real-time Program Services; On-Demand content that was originally offered as Live content
Network Type	Unicast (including Adaptive Bit Rate) Broadcast, Multicast
Transport/Container/Format	MPEG TS, Multicast IP, DASH ISO BMFF
Interface to TVs (source format)	IP connected (for OTT content delivered via managed or unmanaged network) HDMI (for services delivered via a STB, e.g., OTT, MVPD)
Backward Compatibility	Native (HLG), simulcast (HDR10/PQ10), decoder based (optional)

Table 3 UHD Phase A Content Parameters

Spatial Resolution	1080p* or 2160p
Color Space	BT.709 [3], BT.2020 [4]
Bit Depth	10-bit
Dynamic Range	SDR, PQ, HLG
Frame Rate**	24, 25, 30, 50, 60
Video Codec	HEVC, Main 10 Profile, Level 5.1 (single layer)
Audio Channels	Stereo or 5.1 or channel-based Immersive Audio
Audio Codec	AC-3, E-AC-3, HE-ACC, AAC-LC
Captions/Subtitles Coding	CTA 608/708, ETSI 300 743, ETSI 300 472, SCTE-27, IMSC1

Table 3 Notes:

*1080p together with WCG and HDR fulfills certain use cases for UHD Phase A services, and is therefore considered to be an Ultra HD format for the purpose of this document. 1080p without WCG or HDR is considered to be an HD format. The possibility of 1080i or 720p plus HDR and WCG is not considered here. HDR and WCG for multiscreen resolutions may be considered in the future.



**Fractional frame rates for 24, 30 and 60 fps are included, but not preferred. Lower frame rates may be best applied to cinematic content.

For the purpose of Phase A services described in this document, including the above constraint on 1080p content, various combinations of these parameters can be combined to produce “UHD Phase A” content.

Phase A is defined in terms of the HEVC codec for distribution to consumers due to its support for HDR/WCG (10-bit) as well as coding efficiency, which makes 4K content more feasible. AVC specifications have been updated to include support for 10-bit depth with BT.2020 color spaces and support of PQ/HLG High Dynamic Range formats. However, the Ultra HD Forum members find that deployed consumer decoders are mostly hardware-based with no feasible mechanism to upgrade to match these latest updates, and thus these new AVC capabilities are not generally usable for distribution purposes. For AVC, consumer decoders generally support 8-bit 4:2:0 formatted content and normally with limitations with regard to the maximum bit rates that would normally preclude decoding 4K content at the compression rates that AVC can achieve.

Using AVC for production, contribution and mezzanine workflows is more viable because changing to encoders with support of the HDR/10-bit capabilities is something that is more easily accomplished.

Table 4 and Table 5 illustrate types of decoders and services that are (or are not) considered “UHD Phase A.” Additionally, Table 5 offers some indications of which types of decoders are compatible with which service formats.

Table 4 Decoder Capabilities

Type No.	Color Container	Resolution	Frame rate	Bit Depth	HDR	SDR BT2020	HDMI	HDCP	UHDF Ph. A
1	BT.709	1080	P25/30	8	No	No	1.4	1.x	No
2	BT.709	1080	P50/60	8	No	No	1.4	1.x	No
3	BT.709	2160	P25/30	8	No	No	1.4	1.x	No
4	BT.709	2160	P50/60	8	No	No	2.0	2.2	No
5	BT.2020	1080	P50/60	10	No	Yes	2.0	2.2	No
6	BT.2020	2160	P50/60	10	No	Yes	2.0	2.2	Yes
7	BT.2020	1080	P50/60	10	PQ10	Yes	2.0a	2.2	Yes
8	BT.2020*	1080	P50/60	10, 8	PQ10	No	2.0a	2.2	Yes
9	BT.2020	2160	P50/60	10	PQ10	Yes	2.0a	2.2	Yes
10	BT.2020*	2160	P50/60	10, 8	PQ10	No	2.0a	2.2	Yes
11	BT.2020	1080	P50/60	10	HLG10/PQ10	Yes	2.0b	2.2	Yes
12	BT.2020	2160	P50/60	10	HLG10/PQ10	Yes	2.0b	2.2	Yes
13	BT.2020*	1080	P50/60	10, 8	HLG10/PQ10	No	2.0b	2.2	Yes
14	BT.2020*	2160	P50/60	10, 8	HLG10/PQ10	No	2.0b	2.2	Yes

Table 4 notes:

- The Ultra HD Forum finds that decoder types 1, 6, and 12 (in **bold** text) comprise the vast majority of decoders, with type 6 being more common in Europe than in other regions.
- Decoders capable of HDR are assumed to also be capable of SDR.
- Decoders capable of BT.2020 are assumed to also be capable of BT.709 at 10 and 8 bit depths.



- *The Ultra HD Forum InterOp Work Group has found that some decoders capable of HDR/BT.2020 are capable of SDR/BT.709, but not also capable of SDR/BT.2020.
- Decoders capable of 2160p resolution include support for 1080p resolution through up-conversion.
- Frame rates indicated are maximum supported for the decoder and include the support for lower frame rates, i.e., 24, 25, and 30, including fractional frame rates for 24, 30 and 60.

Table 5 Service Formats

Service Format Description	Attributes					Table 4 Decoders	UHDF Ph. A
	Color Container	Resolution	Frame rate	Bit Depth	HDR		
HD SDR	BT.709	1080	P30	8	No	All	No
HDp60 SDR	BT.709	1080	P50/60	8	No	2 and above	No
UHD SDR	BT.709	2160	P30	8	No	3, 4, 6, 9, 10, 12 and 14	No
UHD SDR	BT.709	2160	P50/60	8	No	4, 6, 9, 10, 12 and 14	No
HDp60 SDR2020	BT.2020	1080	P50/60	10	No	5, 6, 7, 9, 11, and 12	No
UHD SDR2020	BT.2020	2160	P50/60	10	No	6, 9, and 12	Yes
HDp60 PQ10	BT.2020	1080	P50/60	10	Yes	7 and above	Yes
HDp60 HLG10	BT.2020	1080	P50/60	10	Yes	11 and above	Yes
HDp60 HLG10*	BT.2020	1080	P50/60	10	Yes	11 and above (HDR), 5 through 10 (SDR)	Yes
UHD PQ10	BT.2020	2160	P50/60	10	Yes	9, 10, 12, and 14	Yes
UHD HLG10	BT.2020	2160	P50/60	10	Yes	12 and 14 (HDR)	Yes
UHD HLG10*	BT.2020	2160	P50/60	10	Yes	12 and 14 (HDR), 6, 9, and 10 (SDR)	Yes

Table 5 notes:

*Indicates the Service Format signals HLG10 using the SDR/BT.2020 backward compatible method. See Section 6.1.9.

4.2 Future Phases

The Ultra HD Forum continues to provide guidance beyond the Phase A. The Ultra HD Forum UHD Phase B Guidelines [1] has been released and includes technologies commercially deployable in 2017 and beyond. Although details of future phases are not in the scope of this text,



examples of additional technologies and methods are listed below with further details offered in Annexes B-E.

- Next Generation Audio (NGA) codecs such as Dolby AC-4, DTS:X, MPEG-H – immersive and/or personalized audio systems using audio objects, channels or higher-order ambisonics
- Scalable coding (Spatial, Temporal, Color gamut, Dynamic range)
- Greater than 10-bit encoding depths
- High Frame Rate – Greater than 50 frames per second (fps) for 50Hz regions and greater than 60 fps for 60Hz regions
- 8k spatial resolution
- Dynamic metadata for HDR being standardized in the SMPTE ST 2094 Dynamic Metadata for Color Volume Transform suite of standards
- Advances in captioning and subtitling, e.g., to accommodate HDR/WCG
- Single-layer HDR technologies (such as SL-HDR1 [34]; see Annex E)
- Dual layer HDR technologies (such as Dolby Vision)
- IC_{TCp} color space (see Section 14)
- Color Remapping Information (CRI)

Not all the above technologies are at the same level of maturity in terms of specification/standards development and product availability, or commercial readiness for deployment in an end-to-end ecosystem.



5. Use Cases

The following use cases are intended to provide context for the guidelines defined within this document. They are not intended to be exhaustive, yet they cover those associated with the content creation and distribution ecosystem in Phase A. Note that OTA delivery of UHD Phase A content occurred post-2016, and thus is out of scope of Phase A; see Annex B for OTA use cases.

5.1 MVPD Platform Delivery

Programming distributors may wish to provide a consumer with a high quality experience via cable, satellite or IPTV or, as a secondary delivery system, OTT. The programming distributor may deliver content to the following devices:

- Set-Top Box (STB)
- Other media device (e.g., TV, tablet, smart phone)

The content will need to be transcoded into multiple formats and bit-rates to provide the best experience to the consumer. The following will need to be considered:

- 1080p with HDR/WCG for low bit-rate delivery

Consumers may be at home or mobile, and will access content across multiple devices. Some devices may provide the consumer a superior experience based on decoding and display capability, and content should be created with these variables in mind.

5.2 IP Network Delivery

Adaptive Bit Rate (ABR) is an HTTP delivery solution suitable for multiple streaming formats, such as HLS (HTTP Live Streaming, implemented by Apple®) or DASH. ABR can be used over a managed network (e.g., DOCSIS 3.x) or an unmanaged network (i.e., the public Internet).

In a managed network, ABR content can be delivered by an MVPD that is also an Internet service provider (ISP), e.g. a cable operator or telco operator. The content is sent over a managed network in IP Services. This is also referred to as IPTV. There are a number of technologies that can be used, e.g., DOCSIS 3.x. MVPDs who are ISPs may also send their IP Services over networks managed by other ISPs in what is known as “TV Everywhere.”

In an unmanaged network, a type of service provider referred to as an edge provider, e.g., Netflix®, Amazon® and others, sends IP services over multiple ISP networks on the public Internet. This is known as over the top (OTT).

Today Live event content producers such as sports league owners and others may choose to provide consumers with a high-quality Ultra HD audio/video experience of a real-time event over an IP network. When content is captured and mastered using parameters described in this document and delivered to compatible displays, IP networks can deliver a high quality consumer experience.



So that a high quality experience can be delivered over varying network conditions, the content is transcoded into multiple versions suitable for various bitrates to form sets of encoded content at different bitrate levels. The sets allow seamless switching between the higher and lower bitrate versions of the real-time content, i.e., “adapting” as network conditions vary. In order to make the most efficient use of available bandwidth to the consumer, a content producer will get the best results using advanced picture encoding technologies (e.g., HEVC), which have been engineered specifically for such applications (see also Section 9.3.2).

The following represents the minimum, under normal network conditions, that should be supported from camera to the consumer’s television:

- 1080p resolution with HDR, WCG, and 10-bit depth

Phase A UHD ABR content may be delivered via an IP network to the following devices.

MVPD/ISPs:

- STBs managed by MVPDs that connect to a television.

OTT or TV Everywhere:

- OTT Streaming Boxes (STB) that connect to a television (e.g., game consoles, Roku® boxes and similar devices).
- OTT-capable media devices (e.g., smart TVs, tablets and smart phones) that include a display panel.

Content producers will continue to deliver programming to customers that are using ‘legacy’ devices (i.e., devices that only support SDR, BT.709 [3] color space and stereo audio). Content distribution network (CDN) partners may need to host both UHD Phase A and legacy files.

It should be noted that if the content is delivered to the consumer device via a home Wi-Fi connection, the quality of service may be impacted by the available bandwidth on the home LAN. A wired connection to the device may be preferred.

Table 6 describes various options for delivering UHD Phase A content via IP networks. A list of commercial services currently employing these (and other) methods can be found at the Ultra HD Forum website, <https://ultrahdforum.org/resources-categories/information-on-uhd-deployments/>.

Table 6 Phase A UHD over IP Networks

Operator	Protocol	Network	Format	Video
IPTV	IP multicast	Fiber / xDSL	UDP	Single bitrate
IPTV	ABR unicast (managed network)	Fiber / xDSL	HTTP (HLS/DASH)	ABR
IP cable	ABR unicast (managed network)	DOCSIS 3.x	HTTP (HLS/DASH)	ABR
OTT TV (live)	ABR unicast (un-managed network)	Fiber / xDSL	HTTP (HLS/DASH)	ABR



6. Production and Post Production

The UHD 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.

In Phase A, production practices for audio are similar to those used in current HD content creation. Audio follows multi-channel workflows established for multi-channel 5.1 surround, and Dolby Atmos delivery using (as appropriate) AC-3 [30], E-AC-3 [30], HE-AAC [28], or AAC-LC emission [28]. Although 2.0 stereo sound is possible, UHD Phase A content is considered premium content, and it is therefore recommended that productions provide at least 5.1 channels.

Phase A, 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/CTA-708, ETSI 300 743, ETSI 300 472, SCTE-27, or IMSC1 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 UHD Phase A 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.

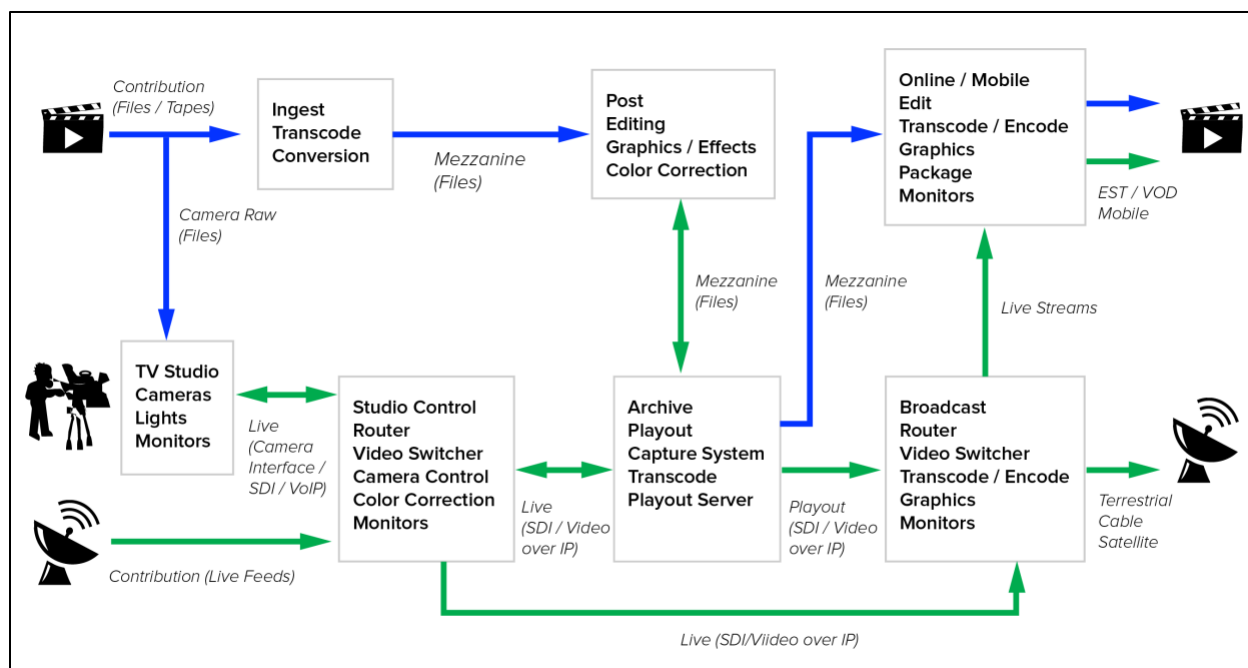


Figure 1 Content Production and Distribution Workflow

6.1 HDR/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 [4] 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 [4] color space. 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 [4]-compatible.

6.1.1 Perceptual Quantization (PQ) and PQ10

One HDR transfer function defined for use in television is the “Perceptual Quantization” (PQ) proposed by Dolby. This is defined as a display transfer curve, or EOTF. The curve is designed to minimize visibility of banding on a display over the brightness range of 0 to 10,000 cd/m². SMPTE standardized the PQ curve in ST 2084. 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 BT.2100 [6] describes the code mapping for both narrow range and full range code values.⁶

The Ultra HD Forum has defined the term PQ10 to refer to the “base package” of parameters that PQ content typically has; i.e., the Perceptual Quantization EOTF described in BT.2100 [6] (which references ST 2084 [10]) together with BT.2020 [4] color space and 10-bit depth. As such,

⁶ CTA 861-G [32] offers additional information about code ranges.



PQ10 content or systems may or may not be metadata-capable. For example, HDR10 is a Phase A PQ10-based format that has the capability of including certain static metadata (see Section 6.1.5).

6.1.2 Hybrid Log-Gamma (HLG) and HLG10

Another HDR transfer function defined is “Hybrid Log-Gamma” (HLG) proposed 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 compatibility with SDR/BT.2020 displays. The HLG curve has been specified in BT.2100 [6].

The Ultra HD Forum has defined the term HLG10 as HLG with 10-bit representation, black at code 64, and nominal peak at code 940⁷.

6.1.3 Recommendation 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 [6]. 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
- Color space: Same as BT.2020 [4]
- Reference Viewing Environment: 5 cd/m² background and less than or equal to 5 cd/m² surround
- Reference non-linear transfer functions EOTF, OOTF, OETF: PQ and HLG
- Color representation: Y'C'B'C'R and IC_TC_P
- Color sub-sampling: same alignment as specified in BT.2020 [4]
- Bit depth value range: 10 and 12 bits, Narrow (64-940) and Full (0-1023) ranges
- Floating point signal representation: Linear RGB, 16-bit floating point

BT.2100 [6] 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 difference format, integer (10 or 12 bits, narrow or full range) or floating point.

Not all of the parameters listed above were deployed as early as 2016, but are included as informative details.

6.1.4 Static Metadata – SMPTE ST 2086, MaxFALL, MaxCLL

SMPTE has specified a set of static metadata in the ST 2086 [11] Mastering Display Color Volume Metadata 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

⁷ CTA 861-G [32] offers additional information about code ranges.



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 color gamut of the mastering display is the DCI-P3 gamut in the BT.2020 [4] color space, and the luminance range is a subset of the 0 to 10,000 cd/m² range provided by 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 transform of the received program to better reproduce the creative intent as shown on the mastering display, given the capabilities of the display device. However, the use of MaxFALL and MaxCLL static metadata have limitations for use with Live broadcasts since it is difficult to determine a program's maximum pixel light values during a Live production. 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.

6.1.5 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 curve: BT.2100 [6] (PQ)
- Color space: BT.2020 [4]
- Bit depth: 10 bits
- Metadata: ST 2086, 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 curve 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, MaxFALL, and MaxCLL); however, HDR10 distribution systems must deliver the static metadata, when present.

6.1.6 UHD Phase A HDR Technologies

Two HDR/WCG “packages” are recommended for Phase A due to their conformance with the criteria listed in Section 4 above. The two technologies are:



- HLG10 – inclusive of HLG OETF per BT.2100 [6], BT.2020 [4] color space, and 10-bit sample depth. (See also Section 10.1 for information about carriage of HLG transfer function signaling over HDMI interfaces.)
- PQ10 – inclusive of PQ EOTF per BT.2100 [6], Rec. ITU-R BT. 2020 [4] color space, and 10-bit sample depth

HLG10 and PQ10 as defined herein may be used without metadata. Provided that the reference display peak luminance specifications do not exceed the consumer display, PQ10 can deliver a greater level of certainty in what the viewer will see based on the display-referred nature of its design, particularly when highlights or average picture level are very bright. This is a result of pixel-encoded values, which are prescribed to represent specific picture luminance level. Another distinction drawn between the two schemes is backward compatibility. Content produced using HLG can be displayed on SDR/WCG devices with a degree of compatibility that may be judged acceptable for programs and services according to Report ITU-R BT.2390 [7]. PQ10 content requires a down-conversion step in order to provide acceptable SDR quality. See Section 11 below for a deeper discussion of backward compatibility, including the pros, cons and open questions that apply to various possible methods.

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). In the sections below, both HDR technologies (PQ10, HLG10) will be discussed.

6.1.7 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. The special values of '0' (meaning, "unknown") are 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 [11], 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.

6.1.8 HDR10 Metadata Carriage

When the Phase A Guidelines were originally published, standards did not exist to carry HDR10 static metadata over an SDI interface. The metadata had to be either embedded within the content



intended for encoding (for Pre-recorded content) or supplied to the encoder via a file or UI. That situation changes with the publication of SMPTE ST 2108-1 [49], which allows a variety of static and dynamic color transform metadata to be carried in the Vertical Ancillary Data Space (VANC) as Ancillary Data (ANC) Messages over an SDI interface or over an SMPTE ST 2110 IP connection (per SMPTE ST 2110-40 [48], and see Section 6.1.11). In HEVC [27] and AVC [26] bitstreams, MaxFALL and MaxCLL may be carried in the Content Light Level (CLL) static SEI message (sections D.2.35 (Syntax) and D.3.35 (Semantics), [27]). SMPTE ST 2086 metadata is carried in the Mastering Display Color Volume (MDCV) SEI, (section D.2.28 (Syntax) and D.3.28 (Semantics), [27]). Both the CLL and MDCV SEI messages can be represented by their corresponding ANC messages under SMPTE ST 2108-1 and 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 Phase B dynamic metadata technologies (see Ultra HD Forum UHD Phase B Guidelines [1]).

Note that 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/display, it is expected that the decoder/display will employ “best effort” to render the content accurately. Note that such an end-to-end chain is operating in a manner indistinguishable from PQ10 content.

Given that the insertion, carriage and preservation of static metadata presents new challenges in the broadcast environment, the use of PQ10 or HLG10 may be a practical alternative to the use of HDR10 in some Live workflows. See also Section 6.2.5 on Live content production and Section 9 on content distribution for additional details.

6.1.9 Signaling Transfer Function, Color Space and Matrix Coefficients

The color space, 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 and QuickTime are often used.

In linear workflows, these values are typically signaled in the VUI of an H.265/HEVC or H.264/MPEG-4 AVC bitstream. As shown in the table below, there are two methods of signaling the HLG transfer function.

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 losslessly convert 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 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.

The table below summarizes HEVC Main10 Profile bitstream PQ and HLG indicators. (In HEVC and AVC specifications, the bitstream elements are bolded and italicized to distinguish them from temporary variables and labels.)

Table 7 Signaling Transfer Function, Color Space and Matrix Coefficients

Color Volume	Color Primary (Logical)	Transfer Function (Logical)	Color Representation	Minimum Bit Depth (bpc)	MPEG Color Primary Index Value	MPEG Transfer Function Index Value	MPEG Conversion Matrix Index Value	AVC/HEVC Codec Tiers
SDR/BT.709	BT.709-5	BT. 1886	$Y'C_B'C_R'$ NCL	8	1	1	1	MPEG2 MAIN 8-bit
		BT. 1886	$R'G'B'$ 4:4:4	8	1	1	0	AVC High 4:4:4 8-14 bits OR HEVC Main 8-bit
SDR/WCG		BT. 1886	$Y'C_B'C_R'$ NCL	10	9	1 or 14*	9	HEVC Main10 and Main12
		BT. 1886	$R'G'B'$ 4:4:4			1 or 14*	0	HEVC Main 4:4:4 16 and Main 4:4:4 Intra
HDR/WCG	BT.2020-2	PQ (ST.2084 / BT. 2100)	$Y'C_B'C_R'$ NCL			16	9	HEVC Main10 and Main12
		HLG / BT.2100	$Y'C_B'C_R'$			18	9	
		PQ (ST.2084 / BT. 2100)	$IC_T C_P$			16	14	
		HLG / BT.2100	$IC_T C_P$			18	14	
		PQ (ST.2084 / BT. 2100)	$R'G'B'$ 4:4:4			16	0	HEVC Main 4:4:4 16 and Main 4:4:4 Intra
		HLG / BT.2100	$R'G'B'$ 4:4:4			18	0	

*Note that both “1” and “14” are valid values to signal SDR transfer function; for example, DVB documents require “1” when the color container is BT.709 and “14” when the color container is BT.2020.

Details on SEI and VUI messaging are available in the HEVC specification [27], in particular, Appendix D (SEI) and Appendix E (VUI).



Table 8 SMPTE Registry UL's for Transfer Function, Color Primaries and Coding Equations

Color Volume	Color Primary (Logical)	Transfer Function (Logical)	Color Representation	Min. Bit Depth (bpc)	Color Primaries	Transfer Functions	Coding Equations (Matrix Conversions)
SDR/BT.709 [Color 3]	BT.709-5	BT. 1886	Y'C _B 'C _R ' NCL	8	06.0E.2B.34.04.01.01.06.04.01.01.01.03.03.00.00	06.0E.2B.34.04.01.01.01.04.01.01.01.01.02.00.00	06.0E.2B.34.04.01.01.01.04.01.01.01.02.02.00.00
HLG HDR-SDR [Color 3]		HLG / BT.2100	Y'C _B 'C _R ' NCL	10		06.0E.2B.34.04.01.01.0D.04.01.01.01.01.0B.00.00	
SDR/WCG [Color 5]	BT.2020-2	BT. 1886	Y'C _B 'C _R ' NCL	10	06.0E.2B.34.04.01.01.0D.04.01.01.01.03.04.00.00	06.0E.2B.34.04.01.01.0E.04.01.01.01.01.09.00.00	06.0E.2B.34.04.01.01.0D.04.01.01.01.02.06.00.00
HDR/WCG [Color 7]		PQ (ST.2084 / BT. 2100)	Y'C _B 'C _R ' NCL			06.0E.2B.34.04.01.01.0D.04.01.01.01.01.0A.00.00	
		HLG / BT.2100	Y'C _B 'C _R ' NCL			06.0E.2B.34.04.01.01.0D.04.01.01.01.01.0B.00.00	

Additional use cases for production-level signaling wrappers include Apple QuickTime™ Wrapper NCLC Color Tags using MPEG Color Tags and MXF Image Descriptors from ST.2067:21:2016 [38].

As described in Section 8.1, service providers should convert or remap all content into a single, consistent color space and transfer function. Setting the initial values in the encoder should be adequate assuming the encoder is dedicated to a single format.

As of 2016, there was no defined means of carrying signaling for transfer function, color space, and HDR-related metadata through the end-to-end supply chain. Gaps existed in the SDI interface, the HDMI interface, mezzanine file wrappers, and other areas. The Ultra HD Forum observes that standards bodies have addressed these issues and provided documentation, much of which is referenced herein.

6.1.10 Peak Brightness – Consideration of Production, Reference Monitors, Consumer Displays and Archive Content

Luminance values can be coded as “absolute values” on a scale of 0-10,000 nits (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 nits, 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 nits, it is helpful to the display to indicate the actual maximum pixel value to be expected that are creatively pertinent as a basis for color volume transform. HDR10 systems are capable of providing static metadata to assist mapping the transmitted content to 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 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 black levels.

HLG is not specified for use with metadata, and instead has a specified relationship between overall system gamma (implemented as part of the display EOTF) and peak display brightness. An overall system gamma of 1.2 is specified for HLG when displayed on a 1,000 nit monitor. BT.2390 [7] states that the artistic intent of HLG content can be preserved when displaying that content on screens of different brightness (even when that display is brighter than the mastering display), through changing the system gamma. BT.2100 [6] provides a formula to determine overall system gamma based on the desired display peak brightness. To preserve creative intent, this formula is recommended.

In the case of HLG, ITU-R BT.2408 [9] recommends the reference level for HDR graphics white (aka “reference white”) be set to 75 IRE (equivalent to 203 cd/m² on a 1,000 cd/m² reference display, or 343 cd/m² on a 2000 cd/m² 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. Some experts have found that setting graphics white at this level as an anchor can produce a darker SDR image and note that compensating for this during conversion to SDR by remapping levels higher can conversely cause clipping, and a suggestion is that dynamic conversion can potentially improve this.

Note that 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” (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.

6.1.11 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 Realtime Transport Protocol (RTP) as described in RFC 3550 [43], 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 [44] specifies the network interface requirements, overall system timing model, and session description protocols. SMPTE ST 2110-20 [45] specifies the format for uncompressed video essence, while SMPTE ST 2110-21 [46] identifies timing characteristics for these video streams and requirements for compliant senders and receivers. SMPTE ST 2110-30 [47] specifies the format for uncompressed digital audio. Lastly, SMPTE ST 2110-40 [48], 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.

6.2 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) and OTT (and OTA in Phase B). 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.

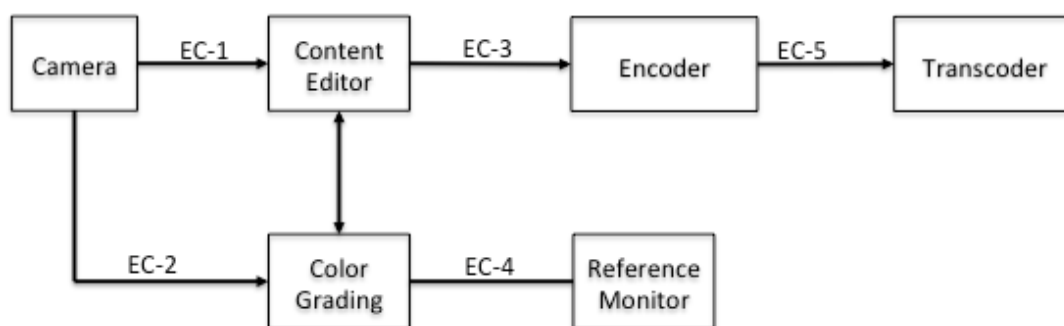


Figure 2 Pre-recorded Content Production Workflow and Interfaces

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.



Table 9 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.

6.2.1 Camera Requirements

6.2.1.1 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 Phase A, capturing in the native camera format allows for future content grades being used for next generation Ultra HD distribution.

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.

6.2.1.2 Imaging Devices: Resolution, Dynamic Range and Spectral Characteristics

In the creation of UHD Phase A 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 color space and spectral characteristics, the more advanced sensing devices will exhibit characteristics approaching the color gamut of BT.2020 [4], while more typical devices will produce color performance approximating the DCI-P3 gamut or just beyond the gamut of BT.709 [3].

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 UHD Phase A film content, the 16:9 aspect ratio of UHD Phase A 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 UHD Phase A differently than digital sources, e.g. film grain, achievable resolution.

6.2.2 Reference Monitor

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 Film employs slightly different techniques, using calibrated monitors and LUTs to replicate the viewing characteristics of different consumer viewing conditions.

The recommendation for UHD Phase A 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 UHD Phase A, a reference monitor can ideally render at least the following: resolutions up to 3840x2160, frame rates up to 60p, BT.2020 [4] color space (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 UHD Phase A delivery.

6.2.3 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 or the latest version of an equivalent standard. BT.2100 [6] 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 nits of peak brightness. Some 2015 and later RGB OLED mastering monitors are capable of 1,000 nits of peak brightness. Note that future HDR content delivered to the consumer may be intended for substantially greater than 1,000 nits peak display brightness.

6.2.4 Color Grading

6.2.4.1 Grading Work Flow

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 [11] Mastering Display Color Volume Metadata.

Current industry workflows such as ACES [51] 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 Annex in Section 15.)

6.2.4.2 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.

6.2.4.3 Grading Monitor Configuration

A professional mastering monitor should be specified and configured according to the required deliverable: Color Space, Transfer Matrix, EOTF, White Point and RGB level range.

6.2.4.4 Grading System Configuration



As of 2016, no *de facto* standards 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. Annex A may provide some insight into this topic.

6.2.5 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. A high-level diagram of a post-production mixing environment for this format is shown below:

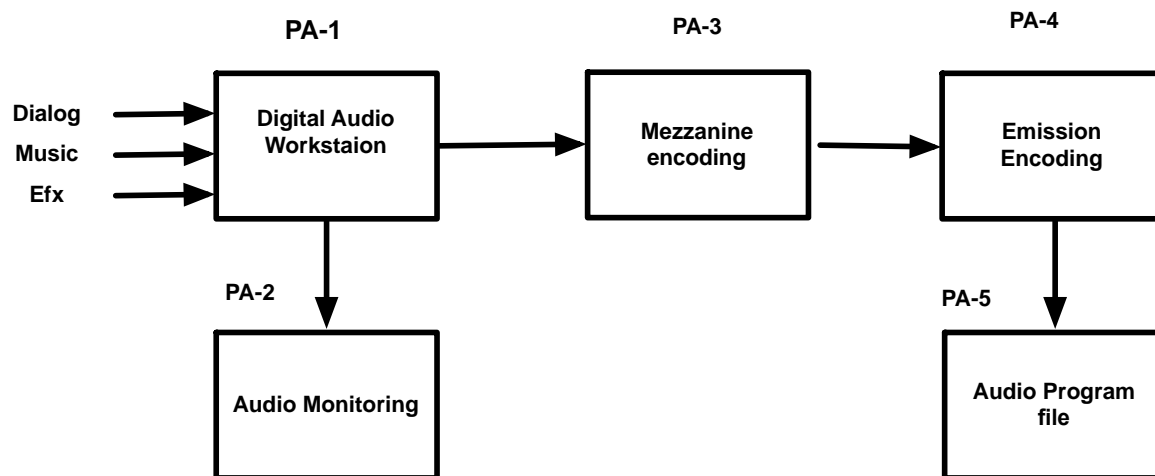


Figure 3 Channel-based Immersive Audio Post-Production



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 ⁸ for final emission delivery
PA-5	Audio Program File	SMPTE ST 337 [36] embedded in a delivery file.

6.2.5.1 Monitor 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.2051⁹ (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).

6.2.5.2 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.

6.3 Production for Live Content

The Ultra HD Forum’s objectives include understanding the implications of creating and delivering UHD Phase A 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, awards shows, etc. Pre-recorded content is captured and produced in advance (see Section 0). Examples include soap operas, sitcoms, drama series, etc. Real-time Program Services may include Live programs and Pre-recorded programs,

⁸ See Section 11.5 regarding backward compatibility.

⁹ 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.



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) and OTT (and OTA in Phase B) 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.

6.3.1 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 color space. See section 7.2 for details on converting assets from SDR to HDR or vice versa.

Performing the above functions on UHD Phase A content may be different than for HD/SDR content. UHD Phase A 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 color space preset in the display. For instance, for HLG10, a displayed picture on a legacy display that uses BT.1886 [5] / BT.709 [3] color space will be incorrect. See Section 11 for details and caveats.

6.3.2 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 color space 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. When both PQ and HLG masters target the same peak brightness, they are typically expected to exactly match each other when viewed on the same mastering display in the respective corresponding mode. In some practices, if the PQ and HLG masters are checked in this way and they do not appear identical to each other, the overall grade is rejected.

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.

6.3.3 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 downmixes, 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 is shown in Figure 4 below.

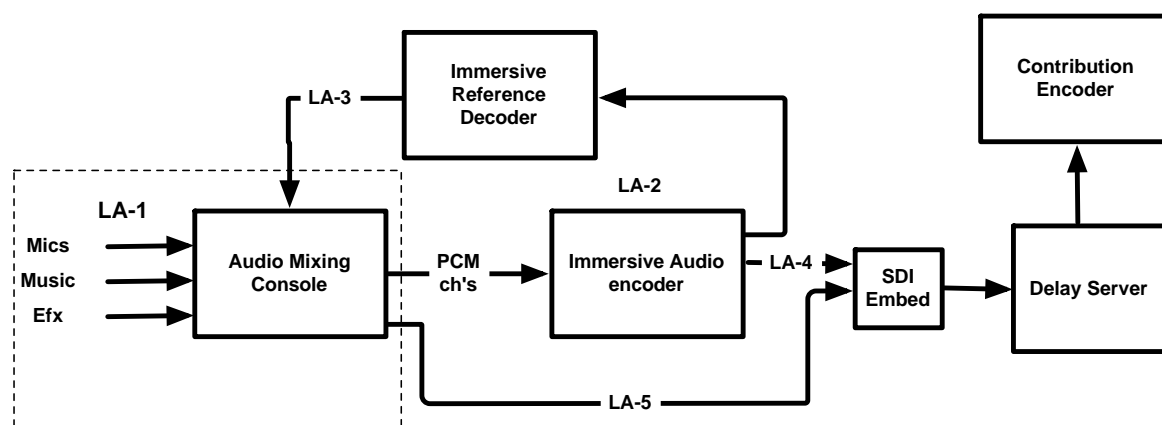


Figure 4 Channel-based Immersive Audio Live Production

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 ¹⁰ .
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 EAC-3 is delivered as a 5.1 channel bitstream + parametric side-data (steering data) to the contribution encoder ¹¹ 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.

In this channel-based immersive audio example, a Dolby Atmos enabled E-AC-3 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 (specified in ETSI TS 103 420 [33]) that are carried in the bitstream, along with the 5.1 audio, for use by a Dolby Atmos E-AC-3 decoder. The full Atmos decode process reconstructs the original

¹⁰ The downstream Atmos Channel-Based Immersive emissions encoder, using E-AC-3 with ETSI TS 103 420 will render a legacy 5.1 audio program. It is recommended to verify the rendered 5.1 audio program using a suitable Atmos E-AC-3 decoder in the monitor chain.

¹¹ See Section 11.5 regarding backward compatibility.



channel-based immersive audio source from the 5.1 backwards compatible representation and the additional parameters. A typical immersive program can be encoded, for example, at 384-640 kbps total, thus fitting into existing emission scenarios.



7. Security

7.1 Content Encryption

Digital content has always been exposed to illegal reproduction and illegal distribution. Various content protection technologies have been developed, involving different scrambling algorithms. Some of these algorithms, still in use, were designed more than 20 years ago and are no longer resistant to sophisticated attacks.

As a general rule, it should be considered that scrambling algorithms with key size less than 64 bits are not resistant to sophisticated attacks; in fact, the time needed to succeed in such attacks is measured in seconds not hours.

A well-known algorithm is Data Encryption Standard (DES), designed in the 1970's, and referred as FIPS 46-3. It was withdrawn from the list of algorithms recommended by the US National Institute of Standards and Technologies (NIST) in May 2005.

Another well-known algorithm is DVB CSA, approved by DVB in 1994, and described in ETSI ETR 289. Its design was initially secret but was subsequently disclosed in 2002. Since then, many publications describe various attacks. In particular, the publication "Breaking DVB-CSA", by Tews, Wälde and Weiner, Technische Universität Darmstadt, 2011 describes an implementation of such attack. This paper reveals that with very reasonable investment, DVB-CSA with a 48-bit key size (also known as DVB-CSA1) could be reversed in real-time.

Fortunately, DVB-CSA with a 64-bit key size (also known as DVB-CSA2) is still resistant against these attacks and is likely to remain so for another few years.

Content protection technologies are now using more recent scrambling algorithms with a larger key size, usually a minimum of 128 bits.

The algorithms that have been standardized for protection of digital content are:

- B7 AES, Advanced Encryption Standard, published in FIPS-197, NIST, 2001,
- B7 DVB-CSA3 published in 2011 by DVB in ETSI TS 100 289 V 1.1.1.,
- B7 DVB-CISSA published in 2013 by DVB and described in ETSI TS 103 127 V1.1.1

The Ultra HD Forum recommends the following regarding content security:

- B7 UHD Phase A content should not be scrambled with DVB-CSA1, nor with DES scrambling algorithms.
- B7 UHD Phase A content should be scrambled with AES or DVB-CSA3, using a minimum of 128 bits key size.
- B7 DVB-compliant service providers should use DVB-CSA3 or DVB-CISSA when transmitting Live or linear UHD Phase A content.
- B7 In the case where DVB-CSA3 is still not deployed, it is acceptable to use DVB-CSA2 with a crypto-period between 10 and 30 seconds, during the time needed to upgrade the equipment to DVB-CSA3.

7.2 Forensic Watermarking

7.2.1 Introduction

Forensic Watermarking complements content protection technologies such as Digital Rights Management (DRM) and Conditional Access Systems (CAS) by providing a means to deter piracy. In this case, Forensic Watermarking is used to generate individualized copies of a video asset thereby allowing the recovery of forensic information to address security breaches. For instance, such forensic information could identify the user receiving the content or the device receiving or displaying the content, its model ID, or other information that can help identify a piracy source.

A watermarking technology used for Forensic Watermarking is characterized by a number of properties, including:

- **Fidelity:** the modifications made to the content to introduce the Watermark Identifier shall remain imperceptible for a human being;
- **Robustness:** the Watermark Identifier shall be detectable after post-processing operations that may be performed by pirates, e.g., re-compression, camcording, screencasting, etc.
- **Payload length:** the size of the Watermark Identifier expressed in number of bits that can be inserted.
- **Granularity:** the duration of multimedia content that is needed for detecting a Watermark Identifier, usually expressed in seconds and dependent on the payload length.
- **Security:** the Watermark Identifier shall withstand targeted attacks from adversaries that know the Forensic Watermarking technology in use as well as its technical details.

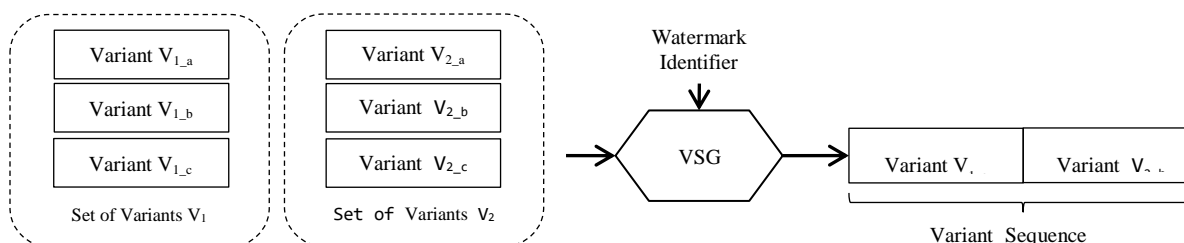


Figure 5 Illustration of Watermark Identifier

The following sections provide:

- Information about watermarking terminology, use cases and applications.
- Consideration for performance, feasibility, privacy, and other aspects.
- High level information flow for some use cases.
- Communication specification for integration of Forensic Watermarking.

7.2.2 Use Cases

Forensic Watermarking is routinely used nowadays in professional environments, e.g., for distributing movies prior to theatrical releases and in Digital Cinemas. In 2014, MovieLabs



released a specification [40] that mandates the use of Forensic Watermarking on UHD content. In this context, it is currently being considered at different stages of the content distribution pipeline. This section provides a comprehensive overview of these different stages and how different watermarks contribute to the forensic analysis. However, the remaining sections will only focus on the use case “Forensic Watermarking Identifying Consumers Devices.”

7.2.2.1 Forensic Watermark Identifying Operators

It is common practice for content owners to incorporate a master Watermark Identifier in the copy of the movies that they are shipping to operators. It provides them with means to track which resellers have been pirated in order to ask them to adopt relevant security countermeasures. Such master Watermark Identifiers are embedded in very high quality content and fidelity is therefore of paramount importance. In this case, fast turnaround processing time is not as critical as in other application use cases. The watermark embedding process can be performed offline and watermark detection results can be returned after a few days.

7.2.2.2 Forensic Watermark Identifying Redistribution Networks

An operator may distribute content through several channels: satellite, terrestrial broadcast, over-the-top, etc. These redistribution channels can be part of the operator’s own redistribution network. They may also be operated by some affiliate networks, and in such a situation, when piracy is detected to originate from an operator, there is some uncertainty about the source of the leak. To disambiguate the situation, alternate Watermark Identifiers can be embedded for the different redistribution channels. In this context, fast detection is usually not required. On the other hand, watermark embedding may need to be performed on-the-fly to accommodate redistribution of some video content, e.g., live broadcast.

7.2.2.3 Forensic Watermark Identifying Consumers Devices

The MovieLabs specification [40] requires identification to stop leakage on the level of device or device type. This can provide means to analyze piracy at the finest granularity in order to deploy targeted anti-piracy measures accordingly. Based on this requirement for premium content, it is likely that Forensic Watermarking could be useful to help deter piracy of live events such as sports and music concerts.

The live service application use case has stronger requirements, as the watermark embedding operation must not introduce significant delay in the content transmission chain. In addition, it may be even more critical to quickly identify the source of an illegal retransmission in order to stop it, ideally before the end of the pirated live program such as a sporting event. On the other hand, the watermark fidelity constraint might be relaxed on some occasions to speed up the detection process. Finally, while it is important to secure the Watermark Identifiers against malicious attacks, the attacker will have more time to execute attacks on VOD content than live content, which quickly expires in value even in pirated form.

7.2.3 Distribution

For Forensic Watermarking that is identifying consumers’ devices, different methods of media distribution may result in different workflow and process optimizations.

- **Physical and Broadcast Distribution:** The content is distributed either via physical media such as a Blu-ray disc or via broadcast.
- **File Distribution:** A single file is made available for playback, used in environments with guaranteed bandwidth streaming, or (progressive) download.



- ABR Streaming: Streaming without guaranteed bandwidth but adaptation to the available bandwidth is done based on several available bandwidth options made available in a manifest or playlist.

The distribution mechanism has impacts on the integration of the watermarking system. Sections 7.2.4 and 7.2.5 present in detail two major approaches, namely those that operate in a single step (one-step) and those that require a preprocessing step (two-step). In addition to the distribution mechanism, the selection of one-step versus two-step watermarking is guided by other aspects such as the integration complexity, the availability of client-side marking capabilities, the ability to modify components on the head-end side, etc.

7.2.4 One-Step Watermarking

This approach to create forensically watermarked content is to mark decompressed baseband video content in a single step:

- During encode on the server, or
- During playback in the secure video pipeline.

One-step watermarking on the distribution server side requires delivering individualized watermarked content to every device requesting it. As such, it may not be suitable for serving a large number of recipients due to scalability constraints when encoding a stream for each recipient.

One-step watermarking is therefore usually applied on the client side, where the watermark is enabled. It typically involves communication between the Conditional Access (CA) and the watermarking (WM) modules as shown in Figure 5.

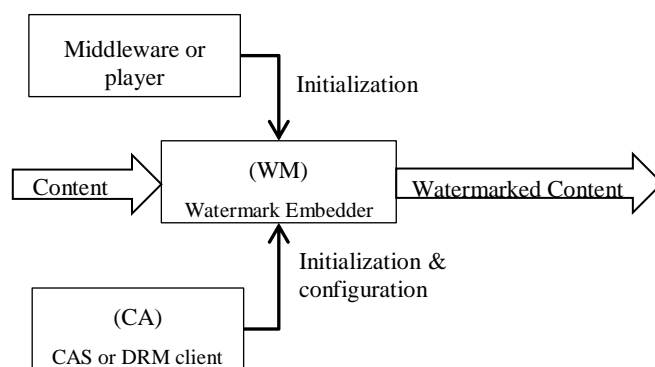


Figure 6 One-Step Watermark Performed on the Client Side

7.2.5 Two-Step Watermarking Integration

Two-step watermark integration requires (i) integration of the preprocessing step at the head-end, (ii) integration of the Individualization step somewhere in the distribution pipeline, and (iii) defining a mechanism to transport the Variant metadata along the video.

7.2.5.1 Step One: Content Variants Preparation

The generation of Variants is performed at the head-end so that distributed video can be forensically watermarked easily further along the distribution pipeline.

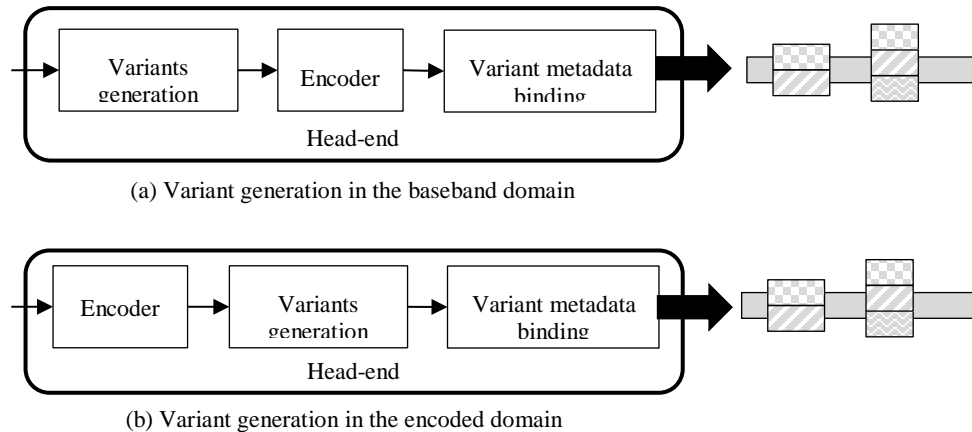


Figure 7 Two Examples of Two-step Watermarking Systems

The resulting integration with encoders will depend on whether the generation of Variants is performed in the baseband domain or in the encoded domain.

- **Variant generation in the baseband domain:** In this case, the preprocessing module is fed with baseband video content in input and generates uncompressed Variants. The resulting Sets of Variants are then forwarded to the encoder to produce encoded Variants.
- **Variant generation in the encoded domain:** In this case, the preprocessing module is fed with encoded video content produced by the encoder and generates Variants of encoded content. Such a preprocessing module can operate independently of the encoder, however, for latency-constrained applications (e.g., live broadcast). it is desirable to have a more intimate integration to minimize processing delay.

7.2.5.2 Step Two: Individualization

Further along the video distribution pipeline, an individualization agent produces the forensically watermarked videos. This individualization agent has typically access to a WM ID and comprises:

- A Variant Sequence Generator (VSG) that receives Sets of Variants in a relevant container, selects a single Variant from each Set of Variants, and thus produces a Variant Sequence that encodes the desired WM ID, and
- A Content Assembler that merges the Variant Sequence into the compressed video bitstream in order to obtain the forensically watermarked video encoding the desired WM ID. If Variants are associated to independent encryption blocks, the Assembler may operate in the encrypted domain without knowledge of the decryption key.

The resulting forensically watermarked video can be decoded for rendering, stored for later use (PVR) or forwarded further along the transmission pipeline.

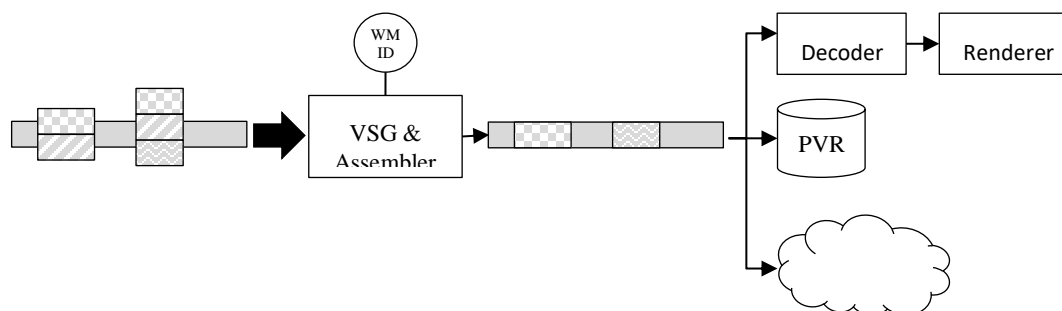


Figure 8 Watermark Embedding Using a Unique Variant Sequence

Deployment scenarios differ with respect to the time and location of the individualization agent, which themselves are dependent on integration preferences and network types. Individualization can be performed on the server side. It creates requirements on the server such as VOD server or CDN edge, but it is transparent to the client. Alternatively, individualization can be done on the client side before decoding:

- **Just-in-Time Server Individualization:** The delivery server hosts the Sets of Variants computed in the preprocessing step and possibly (parts of) the original content. When a client requests the content, the server identifies the session or client using known authentication methods such as HTTP cookies, tokens or authenticated sessions and associates a WM ID to the request. The individualization agent can then operate and deliver forensically watermarked video to the requesting client.

The storage overhead on the server amounts to the Variants data. The smaller is the Variants data compared to the original content, the less overhead is induced by the Forensic Watermarking technology.

- **Prepared Server Individualization:** In this case, the delivery server hosts pre-individualized video segments to lower the complexity of delivering forensically watermarked content.

In the most extreme case, the whole content is individualized using a pool of WM IDs resulting in a collection of forensically watermarked contents that are placed in a stack. When a client requests the content, the server delivers the next pre-watermarked content in the stack and records in a database the link between the session and the WM ID. This approach does not scale well though as it induces significant storage and caching overhead.

In practice, it is usually more efficient to pre-individualize video content at a finer granularity. For instance, a watermark vendor may use several Sets of Variants to encode a single WM ID symbol. The baseline idea is to then pre-compute alternate versions of a video segment, each version encoding one of the possible symbol. The delivery server then only has to forward one of these pre-computed versions for each video segment based on the client's WM ID.

When using a binary alphabet, this solution amounts to having two versions of the content being stored and transmitted downstream. The boundaries of the pre-computed video segments can be made to align with the boundaries of the transmitted video packets, e.g. a MPEG TS packet or an ABR chunk. This further lower the complexity of the delivery server. A typical example is A/B watermarking routinely used in ABR and described in detail in Section 3.1.5.



- **Client Device Individualization:** In some application uses cases, individual stream delivery is not feasible, e.g. one-way distribution such as broadcast or physical media. In this case, the individualization process needs to be performed on the client side. The client device therefore receives the Variant metadata along with the encoded video content. Using the Watermark Identifier of the device, the VSG produces a Variant Sequence that is then used by the Assembler to produce a watermarked video bitstream to be forwarded to the video decoder. To secure the watermarking operation, the individualization process shall be performed in a secure environment e.g. along the secure video path under the TEE control. Another security mechanism includes controlling access to Variants with decryption keys. This approach is useful for scenarios where individual stream delivery is not feasible, such as one way distribution, including physical media.

7.2.5.3 Variants Transmission Mechanisms

The steps of pre-processing and watermark embedding operations are unlikely to be co-located. It is therefore necessary to define a mechanism to transport the Sets of Variants along the content to be watermarked. A first challenge to address is the temporal synchronization between the video content and the Variants metadata e.g. to support skip modes. Depending on the implementations, the transport mechanism may also define which portion of the Watermark Identifier applies to a given Set of Variants:

Transport at the Media Layer

Variants metadata can be incorporated at the media layer. For instance, MPEG standards indicate that any proprietary information can be placed in the video bitstream as dedicated NALUs referred to as Supplemental Enhancement Information (SEI). Variants SEI NALUs can be signaled using a dedicated identifier [26], [27].

The advantage of such a low-level transport mechanism is twofold. First, the Variants metadata is finely interleaved with media essence, thereby providing the necessary temporal synchronization between the content and the Variants metadata. Second, Variants metadata can inherit the protection provided by CAS or DRM systems. On the other hand, such low level signaling may induce integration along the secure video path and/or the trusted execution environment, which requires collaboration from chip vendors for adoption.

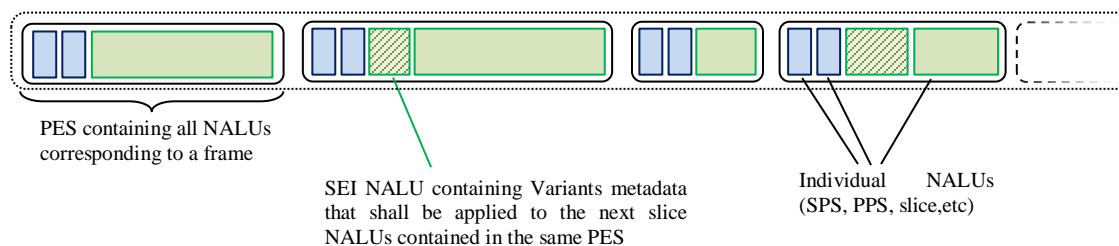


Figure 9 Transport at the Media Layer Using MPEG SEI NALUs

Transport at the Container Layer

Variants metadata can be transmitted in an alternate transmission channel next to the video content at the container layer. For instance, Variants metadata could be placed within a MPEG2-TS bitstream using a dedicated PID and using the PCR to synchronize the video and metadata channels. Alternately, Variants metadata could be incorporated as an extra track in an ISO-BMFF file. In that case, synchronization can be achieved by aligning samples across different tracks.

When Variants metadata is handled as a component separate from the video, proper care shall be taken to guarantee its protection if needed with relevant content protection techniques.

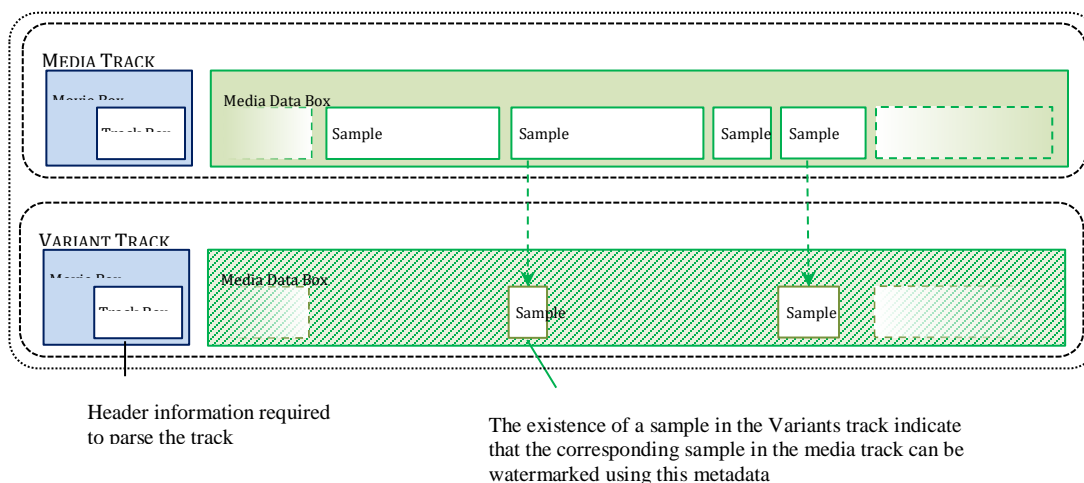


Figure 10 Transport at the Container Layer Using a Track in an ISO BMFF File

An example of how to transmit Variants metadata as an extra track in an ISO BMFF file is described in [41]¹². This standard applies to file-based delivery, e.g. on physical media with embedding on the client side. The baseline principle is to define a dedicated track in an ISO BMFF file that describes how to construct watermarked video samples. For instance, a constructor for a sample indicates which portion of the video track shall be kept and which portions shall be replaced by a Variant available in the variant track. Access to the MPEG variant constructors is subject to cryptographic keys. Different users/devices will have a different set of keys and thereby been able only to produce different watermarked video samples using different constructors. Moreover, the Variants are double-encrypted to serve as an anti-bypass mechanism. A player that would not perform the watermark embedding operation would not be able to display a good-quality video since some segments of the video would still be encrypted. The strong link between encryption and watermarking requires collaboration between CAS/DRM and watermarking systems e.g. for key management and provisioning. The virtue of this design is that it enables a secure integration of the watermark embedding module on open devices outside of the secure video path or trusted execution environment.

Note: The embed location may be communicated out of band, in a common format or can be embedding as extra data during the encoder and is then removed during use by the packetizer, e.g. using a format like EBP [50].

Out-of-Band Transport

Transporting Variants metadata at the media as well as container layer requires some level of video parsing. However, in some integrations scenarios, such a requirement may not be acceptable, e.g. for scalability reasons or because parsing is impossible at the integration point due to encryption. In such a situation, it is desirable to have a direct access to the Variants metadata either

¹² It shall be noted that the terminology “variants” is slightly different in the MPEG standard and these UHD guidelines. In the MPEG standard, a variant is a full MPEG sample composed of parts of the original bitstream and parts of the Variants, as defined in this document i.e. segments of bitstream that can be used interchangeably at a given location in the bitstream.



with a separate file containing the metadata itself or with a separate file containing relevant pre-computed parsing information to access the Variants metadata directly in the video file without performing the actual parsing. This is typically the case on a CDN Edge server.

A CDN Edge server is serving video chunks on request to clients, among other things. To guarantee the scalability of the service, it is of the utmost importance to reduce parsing to a minimum. A possible solution is to store Sets of Variants in the container in such a way that one of the Variants can be selected simply by skipping the data of the other Variants. The indices of where the Variants are located in the stream can then be stored in a separate file which can be loaded in memory by the CDN edge server. As a result, the server can perform the watermark embedding operation directly without any video parsing.

Another use case of the out-of-band transport is when the video stream is being broadcasted but the CDN Edge server is not modifiable for the watermark embedding operation. In such scenario, broadcaster can supply two streams, the video stream and the Variants metadata stream, to the client devices using the conventional CDN. The VSG at the client side can directly access to the broadcasted Variants and can perform the client device individualization. This approach demands additional bandwidth but is applicable when the service provider is not able to control network configurations between streaming servers and end-user clients.

7.2.6 Use Case: ABR VOD

The characteristics of this use case requires that segment boundaries are known before step #1.

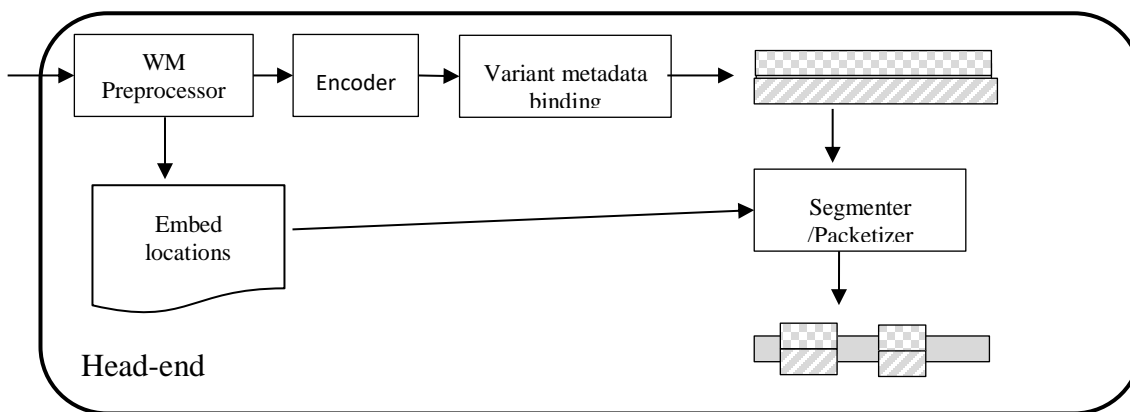


Figure 11 Pre-processing in the Baseband Domain for Two-step Watermarking

In one scenario, video segments are pre-serialized to lower the complexity of delivering forensically watermarked content. The baseline idea is to prepare serialized versions of video segments comprising Sets of Variants for a single WM ID symbol. A typical example is for instance to associate all the Sets of Variants for an ABR fragment to a single WM ID bit. The server can then host two versions of the ABR fragments, corresponding to the WM ID bit equal to 0 or 1. When a client requests an ABR fragment, the server only has to deliver the version of the fragment corresponding to the WM ID of the client.

In the most extreme case, the whole content is pre-serialized using a pool of WM IDs resulting in a collection of forensically watermarked contents that are placed in a stack. When a client requests the content, the server delivers the next pre-watermarked content in the stack and records in a database the link between the session and the WM ID.



This strategy implies duplication of content at a bigger granularity than in the previous case. The resulting storage is therefore usually more significant and since files are individual per user, caching is not applicable.

In another more scalable approach, server hosts pre-serialized ABR fragments. However, the server has no notion of session and is therefore unable to individualize content delivery based on some WM ID.

Forensic Watermarking is achieved by providing individualized manifests/playlists to requesting clients. In ABR, the manifest essentially declares the list of addresses where video fragments can be acquired. The playlist server therefore personalizes the manifest so that each client only ‘sees’ the video fragments that encode its WM ID. When consuming the playlist, the client requests pre-serialized ABR fragments (encoding its WM ID) that are served by the streaming server.

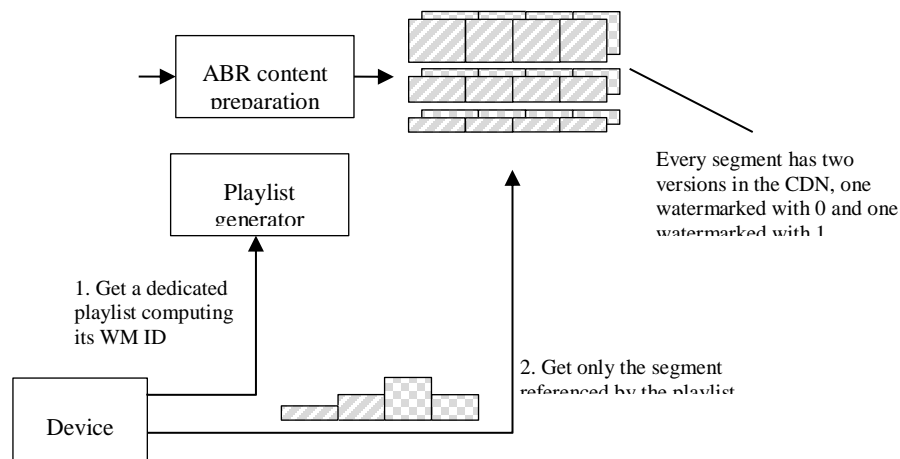


Figure 12 ABR Playlist Serialization



8. 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 are comprised 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 Table 10) 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.

8.1 Maintaining Dynamic Range and Color Space Parameters

Different dynamic range and color spaces 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 color space¹³. 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 color space, 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 remap SDR/BT.709 content into HLG10 or PQ10, to up-convert SDR/BT.709 content to HLG10 or PQ10 and vice versa, and to convert content from PQ10 to HLG10 or vice versa. The following subsections offer guidelines for normalizing content in the headend for this purpose. See also Section 11 for conversion possibilities in consumer equipment for backward compatibility.

8.2 Conversion from SDR/BT.709 to PQ10/HLG10

In multiple areas of the production chain, it is anticipated that a requirement exists 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 (practicing BT.1886 [5] gamma and BT.709 [3] color space). To make it possible to combine such

¹³ See “Guidelines for Implementation: DASH-IF Interoperability Points v3.3”, Section 6.2.5, which is a work in progress as of 2016, but may be found at <http://dashif.org/wp-content/uploads/2016/06/DASH-IF-IOP-v3.3.pdf> for reference.



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 [6]) and color space (i.e., BT.709 [3] primaries “mapped” into the BT.2020 [4] container). Such conversions can utilize:

- Remapping: SDR/BT.709 content is decoded and repackaged as PQ10 or HLG10 containers, but while changing the color space, 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 90 IRE, 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 IRE (equivalent to 203 cd/m² on a 1,000 cd/m² reference display, or 343 cd/m² on a 2000 cd/m² 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-conversion: SDR/BT.709 is decoded and then enhanced/modified to emulate PQ10/HLG10 and repackaged as above. Blind (not supervised) up-conversion can lead to undesirable results depending on the technology so care should be used and any algorithms or equipment should be carefully evaluated.

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 UHD Phase A 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-conversion may be appropriate. For example, up-conversion 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 abruptly cuts 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 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

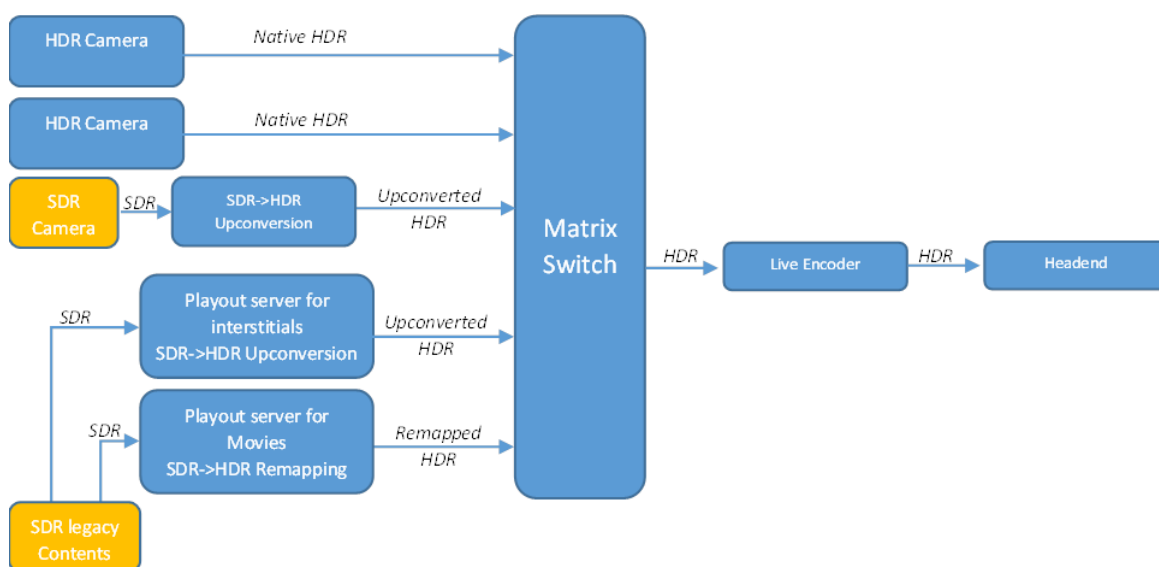


Figure 13 Sample Live Workflow with Mixed Format Source Content

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 judgement 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 11). 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 color space. Real-time Program Service providers should not expect consumer equipment to render seamless transitions between segments of content that have different transfer function or color spaces.

8.3 Conversion 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 Phase A. 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.

8.3.1.1 PQ10 to HLG10

It is possible that Pre-recorded programs are produced in PQ10 in Phase A; however, some service providers may favor HLG10 delivery for its degree of backward compatibility. It is possible to perform the required conversion with a LUT. As the PQ program may have pixel values as high as 10,000 nits, some color volume transform of highlights may occur in content converted from PQ10 to HLG10 and then rendered on a lower peak luminance display.

Note that in an HEVC program stream, it is possible to deliver PQ-related metadata (ST 2086, CLL) in the SEI even when HLG is signaled as the transfer function. At this time, that practice is not referenced in any current standard, thus it should be avoided or employed with caution.

8.3.1.2 HLG10 to PQ10

It is possible that Live programs are produced in HLG10 in Phase A; 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 LUT prior to encoding. It must be considered that this conversion involves translation from a scene-referred relative signal to a display-referred absolute signal. As such, a target peak luminance needs to be preset as part of the LUT conversion function. At the time of this writing, grading monitors typically have a peak brightness of 1,000 nits so that value may be used; i.e., the maximum value of the HLG signal would map to the PQ value of 1,000 nits.

8.4 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



TVs¹⁴ and legacy HD networks. (See also Section 11 on Format Interoperability including STB conversions.)

It is possible to do this conversion using a 3D LUT mechanism as described in Section 8.2. Another method is an invertible down-conversion process described in ETSI TS 103 433 [34], in which HDR/WCG content is down-converted in real time to SDR/BT.709 at or prior to the emission encoder. As a commercially deployed system in 2016, this down-conversion process is considered to be a Phase A technology.

¹⁴ Although HLG offers a degree of backward compatibility with SDR/WCG displays, there is no evidence that HLG offers backwards compatibility for BT.709 [3] displays without additional processing as described in this section. However, both DVB and ARIB require support of BT.2020 color space (and HDMI 2.0) 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.



9. Distribution

This section describes the various stages of distribution for a UHD Phase A workflow. It shows the primary nodes and distribution paths in a typical workflow and describes each node and interface in the chain.

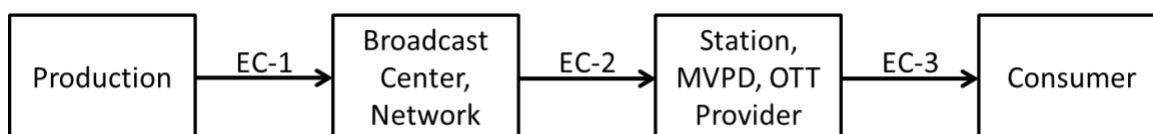


Figure 14 Distribution Nodes and Interfaces

Table 10 Compression and Distribution Nodes and Interfaces

Production	Content production location, e.g., studio, live venue truck, etc.
EC-1	Compression/distribution from Production to a central location
Broadcast Center, Network	A central location such as a broadcast network where content is aggregated for further distribution to station affiliates or service providers; this node is often used in distribution of Live content
EC-2	Compression/distribution to the final point before distribution to consumers
Station, MVPD, OTT Provider	A service provider that distributes content to consumers, e.g., a local television station affiliate, an MVPD or an OTT service provider; this node also often produces Live content, such as news programming
EC-3	Compression/distribution to the consumer device
Consumer	The viewer device, such as a television set, tablet, PC, STB connected to a display, etc.

Some workflows may be simpler, with only three nodes (production, service provider and consumer) while others may be more complex with more than 4 nodes in the chain. The concepts that encompass a four-node workflow can largely be applied to both simpler and more complex scenarios.

The workflows described include those in use in 2016. The workflows described will be able to carry metadata except where noted otherwise. The workflows apply to Real-time Program Services and to On Demand content that was originally offered as Live content.

Typical distribution practices involve decoding, modification and re-encoding the content as it flows from production to consumer. Carriage of transfer, color container, color matrix, and optional HDR10 static metadata is possible in production over both SDI and IP (see Section 6.1.11), in contribution feeds using AVC or HEVC, and in distribution using HEVC. When content is decoded at a node, modified or otherwise, and then re-encoded, attention must be given



to preserving this data at each step. Audio and caption/subtitle are similar to those used in HD content distribution, and thus do not require the same attention. For pre-recorded content, embedded test patterns at the head or tail of the program can be useful for verifying accurate signaling.

The Ultra HD Forum Phase B Guidelines [1] describes the Next-Generation Audio (NGA) workflow. For the purpose of Phase A UDH, audio follows workflows established for Dolby-E and PCM in contribution applications, and AC-3, E-AC-3, HE-AAC, and AAC-LC as the 5.1 emission codecs.

Captions and subtitles follow workflows established for CTA 708/608, ETSI 300 743, ETSI 300 472, SCTE-27, and IMSC1 formats. HEVC [27] includes provisions for carrying captions and subtitles in the VUI and SEI in a similar manner to legacy video codecs.

In Phase A the production system is likely SDI-based (1x12G or 4x3G) and therefore deployment of an SDI workflow is likely. In the future, an IP-based workflow should be implemented, using techniques such as Media over IP: ST 2022-6 and near-lossless compression technologies such as VC-2 (Dirac), JPEG 2000, or other vendor proprietary solution.

Methods of carrying 2160p over SDI defined by SMPTE are shown in Table 11 below.

Table 11 SDI Input Standards for 2160p Content

Interface	Standard	Details	Notes
4x 3G-SDI*	SMPTE ST 424	4 quadrants of 3G-SDI	
	SMPTE ST 425-1	3G-SDI source input format mapping (Level A and Level B)	2 options: quad split or 2 sample interleave
1x 12G-SDI	SMPTE ST 2082	12Gbps SDI native	

* For 1080p, only 1x 3G-SDI is needed.

Metadata for HDR10 can be carried over SDI in VANC, per SMPTE ST 2108-1 [48]. If HDR10 metadata is present, it can be applied at the video encoder as follows:

- In compressed format, HEVC, Main 10 Profile, Level 5.1 may be used for metadata signaling. The metadata is carried via VUI and SEI messages (see Section 6.1.8).
- In a “light compression” format, such as a 12G-SDI signal mapped into a 10GbE, there are multiple options including VC-2 (Dirac), JPEG 2000, and other vendor proprietary solutions.

For UHD Phase A content, only uncompressed video over SDI (4x 3G-SDI or 1x 12G-SDI) or compressed video using HEVC, Main 10 Profile Level 5.1 is recommended for Phase A workflows.

9.1 Production Processing and Contribution

This section describes processes for transmitting UHD Phase A content from an event venue or production studio to a central facility such as a Broadcast Center or Network. Note that in some cases the content may be distributed to more than one Broadcast Center, and different facilities may have different standards or requirements for content acquisition. For example, an international



sports event program may be transmitted to a Broadcast Center, which in turn distributes the program to broadcast stations in different countries, which may have different format requirements, e.g., frame rate.

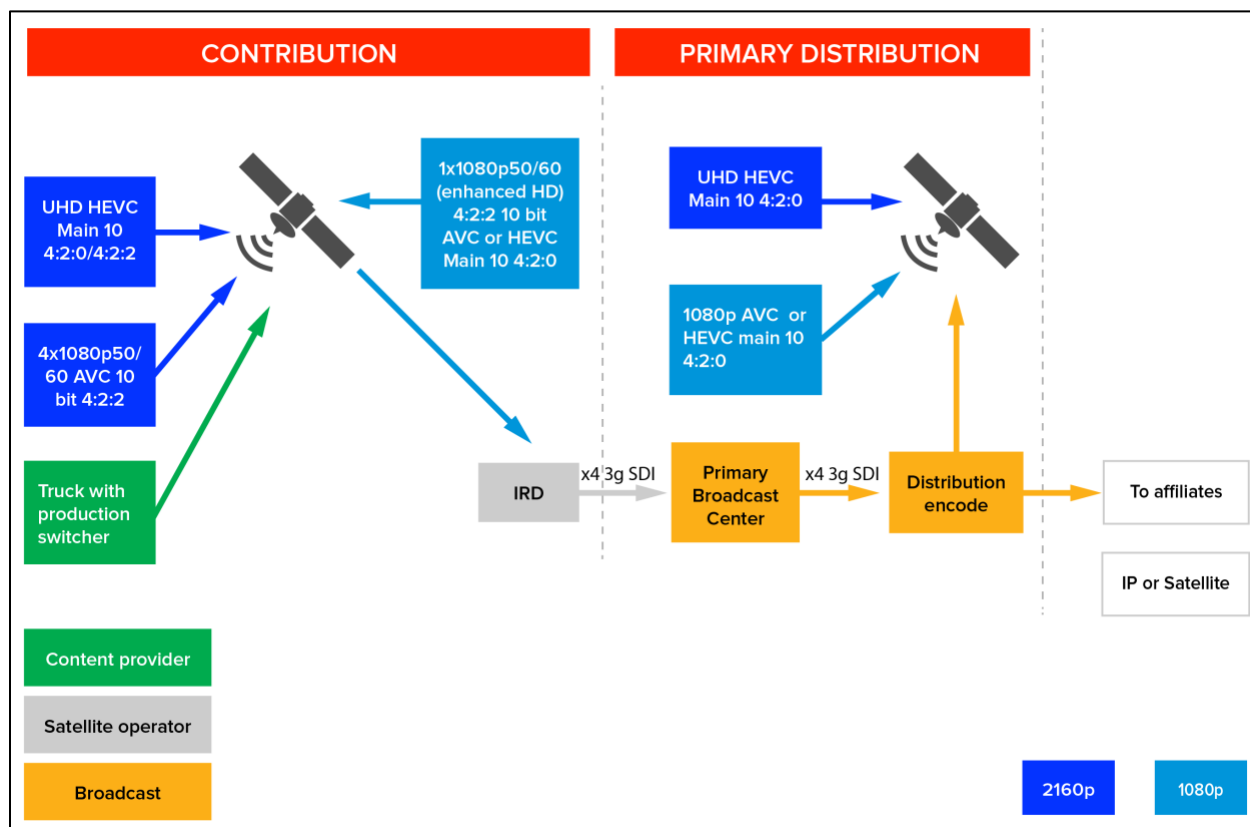


Figure 15 Contribution Workflow

In the case of Live content, the first stage is either a truck or remote production gallery, equipped with a production switcher, camera racking equipment, graphics and audio mixing. The production switcher typically will be receiving baseband signals from cameras and other sources such as digital disc recorders (DDRS) or other playback devices. The Technical Director (vision mixer) in the truck then applies transitions, lower thirds, etc., and may splice in Pre-recorded content (slow motion, interviews, match analysis, etc.). See Section 6.3 for details on Live content production and Section 8.2 for details on mixing source content of different formats in a Live program production.

In the case of Pre-recorded content, the studio has the opportunity to apply post-production edits. See Section 0 for details on Pre-recorded content production.

9.1.1 Video

It has been tested and confirmed by several Ultra HD Forum member companies that in the content production part of the chain, HLG10 and PQ10 can be used with 10-bit 4:2:2 workflow methods and equipment available as early as 2016.



Examples of image processing functions can include mixes, wipes, fades, chroma-keying, linear or luma-keying, motion tracking, DVE moves, slow motion, freeze frames, addition of graphics tickertapes or logos, use of virtual sets, etc. Since image processing may occur many times before final delivery to the consumer, this can be particularly important for Live programming, in which much of the workflow must be fully automated.

Once the content is ready, it is sent to a contribution encoder. Live production workflows typically feed a modulated uplink to a satellite that delivers the content to a Broadcast Center or Network. In some cases, fiber will be used as well as or instead of a satellite. For cost and bandwidth reasons, the satellite feeds will be compressed.

A 2160p feed may be compressed in HEVC, Main 10 Profile, 4:2:0 or 4:2:2 or quadrant-based in 4x1080p in AVC 10-bit 4:2:2. A 1080p HDR/WCG feed may be compressed in either HEVC, Main 10 Profile, 4:2:0 or AVC 10-bit 4:2:2. Note that when fiber links are used, intra-frame encoding and decoding, such as JPEG 2000, may be used. The HDR transfer function and color space must be predetermined and used uniformly for the production.

The satellite operator decodes the signal back to baseband using an integrated receiver decoder (IRD).

Quadrant based (4x1080p 4:2:0 or 4:2:2 10-bit) encoding/decoding is commonly used to create a 2160p image. AVC, AVC-I and JPEG 2000 all could be used for the quadrant codec. In the case that quadrant streams are sent, 4 encoders and 4 IRDs are synced. Single frame 2160p solutions using HEVC are likely to replace quadrants over time as they offer better compression efficiency (HEVC vs. AVC and single frame encoding) and are simpler to operate. This 2160p HEVC contribution method is used in Phase A.

AVC, HEVC and JPEG 2000 differ in the expected bitrate of the contribution file and in the mechanism for carrying HDR/WCG signaling.

Approximate examples of the contribution bandwidth required and HDR carriage signaling are shown below:

Table 12 Contribution Bitrates and Key Parameters

Source	Contribution Format	HDR/WCG Carriage Signaling	Approximate Typical Bitrate Range
1080p 50/60 fps	AVC 4:2:2 10-bit	Under definition in MPEG	20 – 50 Mbps
	HEVC, Main 10, Level 5.1 4:2:2/4:2:0	As per MPEG VUI/SEI signaling [27]	10 – 40 Mbps
	JPEG 2000	Not standardized	100 – 150 Mbps
2160p 50/60 fps	AVC 4:2:2 10-bit (4 quadrant)	Under definition in MPEG	90 – 140 Mbps total
	HEVC, Main 10, Level 5.1 4:2:2/4:2:0	As per MPEG VUI/SEI signaling [27]	50 – 80 Mbps
	JPEG 2000	Not standardized	450 – 550 Mbps

The Ultra HD Forum offers these bitrates based on the general experience of its members. It is important to note that the actual contribution bitrates can vary substantially from the figures shown



depending on many factors, such as latency, quality of source content, type of source content, type of network, multi-hop contribution, etc.

In Phase A, the main factor affecting contribution bitrates is the step between 1080p and 2160p spatial resolution; the addition of HDR or WCG has a much smaller impact. HDR and WCG do, however, require 10-bit encoding precision and modifications to the signal that, if not maintained, will ‘break’ the overall performance of the system resulting in an unacceptable image.

9.1.2 Audio

In 2017, the Ultra HD Forum Phase B Guidelines [1] document was released and included description of Next-Generation Audio (NGA) systems. In Phase A, production practices for audio do not differ from those used in current HD content creation. In Phase A audio follows multi-channel workflows established for 5.1 surround and Dolby Atmos for delivery using one of the following emission codecs as appropriate for contribution applications: AC-3, E-AC-3, HE-AAC, or AAC-LC.

9.1.3 Closed Captions and Subtitles

Production practices for closed captions and subtitles are similar to those of HD content creation in Phase A. Closed captions and subtitles follow workflows established for CTA 608/708, ETSI 300 743, ETSI 300 472, SCTE-27, or IMSC1.

9.2 Broadcast Center Processing and Primary Distribution

This section describes the processes and functions involved in Primary Distribution, i.e., transmitting content from a central facility such as a Broadcast Center or Network to a service provider such as an MVPD or OTT provider. (UHD Phase A terrestrial broadcasts did not commence until after 2016 and thus are out of scope of this document.)

In the Broadcast Center, PQ10 or HLG10 signals can follow roughly similar workflow methods as those used for HD programming during image processing operations using a presentation switcher.

The output of the Primary Distribution encoder will go to a MVPD or OTT provider that will typically decode the content, modify it, and re-encode it to the final distribution format.

Figure 16 depicts different mechanisms of Primary Distribution delivery in TS formats to an MVPD.

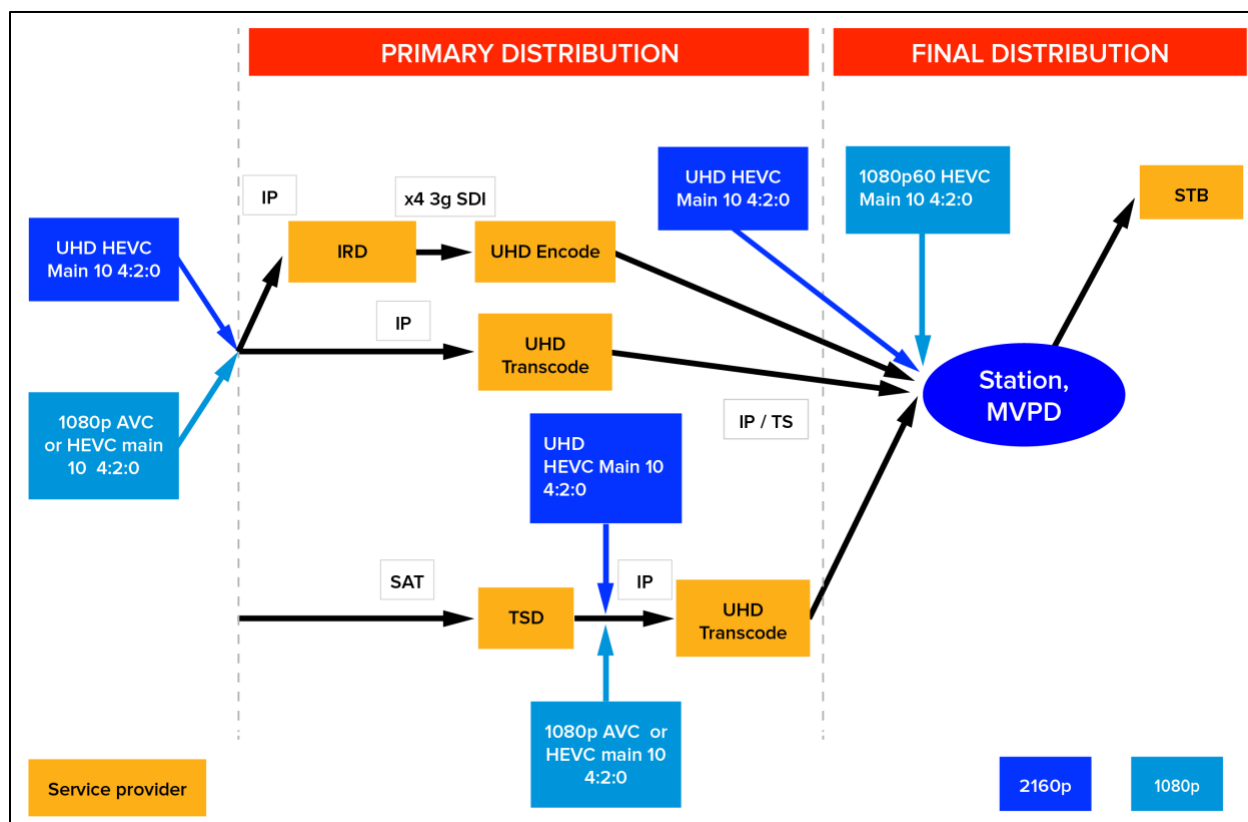


Figure 16 Primary Distribution to an MVPD or Station

Primary distribution to an OTT provider follows a similar scheme, except that the content is delivered using MPEG DASH instead of MPEG TS as shown in Figure 17 below.

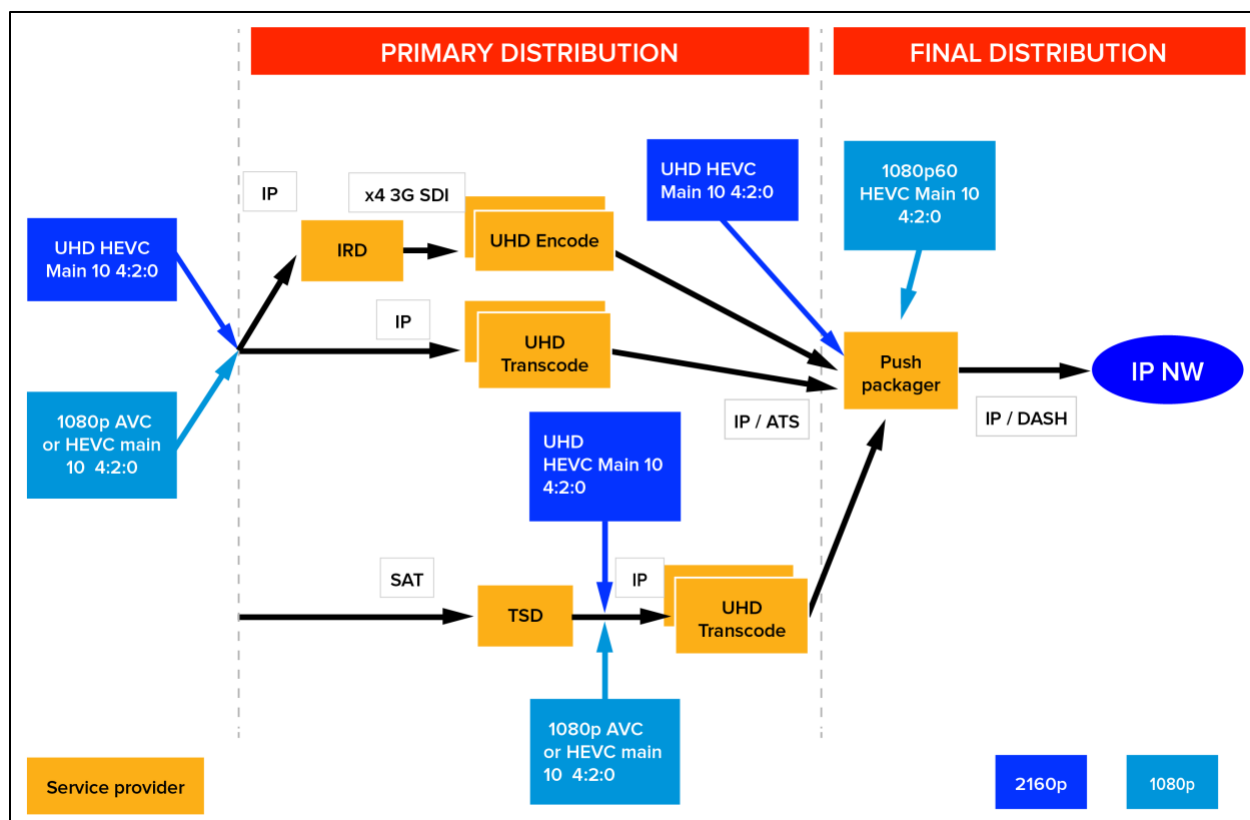


Figure 17 Primary Distribution to an OTT Provider

Primary distribution encoding to affiliates or other partner organizations is expected to be 4:2:2 or 4:2:0 10-bit. Note that frame rate (temporal) conversion may be required for international distribution.

For Live content, the production of the signal may be done via a live ingest server and playout server whose output will be SDI (1x12G or 4x3G or 1x3G in the case of 1080p content). The signal encoded by the Primary Distribution encoder should be HEVC, Main 10 Profile, 10-bit depth [27].

Table 13 describes the formats that can be used for Primary Distribution transmission.

Table 13 Primary Distribution Bitrates and Key Parameters

Spatial Resolution	Primary Distribution Format	HDR/WCG Carriage Signaling	Approximate Typical Bitrate Range
1080p	HEVC Main 10 Profile, 10-bit [27]	VUI/SEI signaling [27]	10-20 Mbps
2160p	HEVC Main 10 Profile, 10-bit [27]	VUI/SEI signaling [27]	40-50Mbps

Like contribution bitrates, the typical bitrates shown for HEVC 2160p are early approximations only based on the general experience of Ultra HD Forum members; Primary Distribution bitrates depend on many factors, such as latency, the quality of the source content, type of source content, type of network, etc.



9.3 Final Distribution from MVPD/OTT Provider Processing

This section describes image processing that may occur at the MVPD or OTT provider and the encoding, multiplexing and packaging required to prepare content for final distribution to consumers.

The MVPD or OTT service provider receives the Primary Distribution feed from the broadcast or network center. In many cases, the service provider will decode the signal and re-encode it for final distribution to the consumer.

9.3.1 Bit Depths

The below table illustrates bit depths in use in 2016.

Table 14 Existing Practices for Real-Time Program Service Distribution Formats

Case	Bit Depth	Color Gamut	Peak White at D65	Color Container	HDR	Use Case
1	8	BT.709	100	BT.709	No	Deployed
2	10	BT.709	100	BT.709	No	Deployed
3	10	Up to BT.2020	100	BT.2020	No	DVB UHD-1, Phase 1 Scenario
4	10	Up to BT.2020	Up to 10,000	BT.2100	Yes	Phase A Guidelines

The Ultra HD Forum finds that Case 2 has not been widely deployed and may be phased out quickly. Only Cases 3 and 4 are recommended for Phase A UHD services, and Case 4 is preferred. Cases 1 and 2 are included for context.

In Cases 3 and 4, SMPTE ST 2086 can be used to signal Peak White. It should also be noted that in Cases 3 and 4, the color gamut can be up to BT.2020 color primaries; however, in current practice the gamut does not exceed DCI P3 primaries.

9.3.2 Video

In this final part of the chain, image manipulation may still be needed for ad insertion, ‘squeeze & tease’, channel logo insertion, etc. PQ10 or HLG10 can again follow roughly similar workflow methods as used for HD programming as they require only signaling describing the transfer function and color space.

HEVC Main 10 Profile, 10-bit [27] is recommended as the only final distribution or emission codec as shown in Table 15 below, as all UHD decoders support HEVC.



Table 15 Final Distribution Bitrates and Key Parameters

Spatial Resolution	Final Distribution Format	HDR Carriage Signaling	Approximate Typical Bitrate Range
1080p	HEVC Main 10 Profile, 10-bit [27]	VUI/SEI signaling [27]	5-18 Mbps
2160p	HEVC Main 10 Profile, 10-bit [27]	VUI/SEI signaling [27]	10-40 Mbps

The Ultra HD Forum provides guidance for bitrates used throughout the final distribution chain only in the form of ranges due to the many parameters that influence the choice of the final distribution bitrate. History suggests that encoding rates improve significantly over time.

The bitrates used will depend on factors such as:

- Whether 2160p or 1080p is used
- Whether the source content is p60 or p50 (or lower) frame rate
- The quality criteria of the operator
- The performance of the encoder

The Ultra HD Forum members offer the below table of bitrates that were in use in services as early as 2016. These “real world” bitrates are intended to offer an additional benchmark for potential bitrates for Phase A UHD services. This information was provided voluntarily, and not all volunteers were able to provide all the metrics.

Table 16 Example “Real World” Bitrates as early as 2016

Delivered via	Transfer function	Frame rate	Bit depth	Color space	Audio codec(s)	Bitrate	Notes
Satellite	SDR	30fps				27Mbps	
Satellite	HLG or SDR	59.94fps	10-bit	BT.2020	Audio adds 5Mbps	35Mbps	Bitrate works for sports and HDR content
Satellite or IPTV	SDR	50fps	8 or 10-bit		AC-3	24-30Mbps	
	PQ	59.94fps	10-bit		AAC, AC-3	32Mbps	
IPTV	PQ	50fps	10-bit	BT.2020	MPEG2, AC-3, DD	25Mbps	Sports content
Satellite	SDR	50fps	10-bit	BT.709		30-38Mbps	Drama, movie content

Note that all the examples in Table 16 are 2160p spatial resolution and use HEVC encoding with 4:2:0 color subsampling.



9.3.3 Adaptive Bitrate (ABR) Streaming

ABR streaming works to minimize playback interruptions when content is delivered OTT through variable quality Internet connections. Multiple encodings enable switching to lower bit rates, which allow a bias towards continuous playback versus higher quality. The encodings are separated into tracks for audio, video, and subtitles. The separate tracks allow real-time multiplexing and addition of tracks as needed without changing the original tracks. Guidelines for UHD Phase A ABR for Live content and Real-time Program Service assembly are as follows:

- Use DASH-IF framework (Guidelines for Implementation: DASH-IF Interoperability Points [17]), which includes manifest format of available assets (Media Presentation Description [MPD]), container (ISO BMFF), video segmenting, security, and HEVC profile settings. DASH-IF may be extending their work to include additional Ultra HD signaling in their Guidelines for Implementation.
- It is recommended that the adaptation set include 720p, 1080p, and 2160p resolutions. However, it should be noted that DASH-IF does not have any specific recommendations for spatial resolutions of Representation in one Adaptation Set, but only specifies maximum resolution for each DASH-IF Interoperability Point (since in practice the input video may be in any arbitrary resolution).
 - Adaptation Sets should be generated such that seamless switching across all Representations in an Adaptation Set is possible. For this purpose, a Representation may be encoded in a lower resolution to provide suitable quality at lower bit rates, permitting down-switching and continuous playout. In this case it is expected that the decoder and renderer apply upscaling to the display resolution in order to ensure seamless switching.
 - Each Adaptation Set is expected to have sufficient annotation to be selected at start-up based on decoding and rendering capabilities. Typically, if one MPD targets different receiving classes, then multiple Adaptation Sets in one MPD for the media type are present. Selection is a key functionality of DASH to support backward and device-compatibility; i.e. not every player has to support all formats described in an MPD.
- Keep HDR and WCG parameters the same across the Adaptation Set resolutions¹⁵. Aspect ratio must also be constant across the adaptation set, and it may also be advisable to maintain the same framerate across the adaptation set in some circumstances.
- Since segment sizes for the same segment duration for 2160p will be larger than for lower resolutions, ensure the CDN is configured to optimize the larger segment sizes, or reduce segment duration to accommodate the CDN capability for segments at 2160p resolution.

Table 17 offers an example bitrates that could be used for OTT services. Note that the combined rows in the table do not represent a suggested “adaptation set” of streams that are

¹⁵ See “Guidelines for Implementation: DASH-IF Interoperability Points v4.0”, Section 6.2.5, which may be found at <http://dashif.org/wp-content/uploads/2016/12/DASH-IF-IOP-v4.0-clean.pdf> for reference.



intended for seamless switching in a given program. Each row represents an independent example of possible expected bitrates.

Table 17 Example Bitrates for Video Streams

Resolution	frame rate	Approximate HEVC bitrate
3840x2160	p60/50	15-20Mbps
1920x1080	p60/50	5-10Mbps
1280x720	p60/50	2-5Mbps
1280x720	p30/25	1-2Mbps
720x404	p30/25	<1Mbps

9.3.4 Audio

In 2017, the Ultra HD Forum released the Phase B Guidelines [1] which includes description of Next-Generation Audio (NGA) systems. In Phase A, Audio may be delivered using legacy or currently employed multi-channel audio codecs, e.g. AC-3, E-AC-3, HE-AAC, and AAC-LC. E-AC-3 and HE-AAC, are considered to offer similar quality at reduced bitrates. While 2-channel stereo can be delivered, it is recommended to deliver 5.1 channel soundtracks when available for an improved sonic experience. Dolby Atmos soundtracks are available for some programs and can be delivered using the E-AC-3 codec.

Many broadcasters and operators are governed by regional legislation regarding managing loudness of broadcast audio. In the United States, for example, an MVPD is obligated to manage loudness per the CALM act, and thus should ensure the audio transmission is compliant with ATSC A/85. OTT providers in the U.S. may also wish to be compliant with CALM in the event the FCC decides to consider them as MVPD providers. Other territories should consider any local specifications or regulations regarding loudness, e.g. EBU R-128.

9.3.5 Closed Captions and Subtitles

Production practices for closed captions and subtitles are similar to those of HD content creation in Phase A. Closed captions and subtitles follow workflows established for CTA 608/708, ETSI 300 743, ETSI 300 472, SCTE-27, or IMSC1. HEVC carries captions and subtitles in User data registered by Rec. ITU-T T.35 SEI defined in HEVC specification section D.2.36 (syntax) and D.3.36 (semantics) [27].

9.4 Transport

MPEG-2 Transport Stream (TS) is being worked on in several SDOs (e.g., DVB, ATSC, SCTE) to accommodate content with HDR/WCG. Operators deploying 2160p HDR/WCG content over MPEG-2 TS in 2016 can use the DVB UHD-1 Phase 2 specification. Operators can carry UHD



Phase A content over RTP/UDP/IP per DVB IPTV [42] (single bitrate only, not ABR), i.e., MPEG-2 TS over RTP/UDP/IP.

For OTT services, MPEG DASH is used to transport UHD Phase A content as follows:

- DASH per DVB DASH specification [14] for live applications
- DASH 265 for live from DASH-IF Guidelines 3.1 [39] (3.3 is a work in progress and includes HDR/WCG)



10. Decoding and Rendering

This section covers guidelines for implementation of decoding capabilities in the consumer player device, picture processing capability of the consumer display device as well as the interface between the consumer player device and the consumer display device. There are two possible architectures for decoding and rendering in the consumer home: 1) STB decoder connected to a display, and 2) integrated decoder/display.

The extent to which the consumer decoder or display is able to switch between SDR/BT.709/SDR/BT.709 and PQ10/HLG10 content or switch between PQ10 and HLG10 seamlessly is not proven, nor is it specified by any standards. It is recommended that service providers employ conversions as needed to ensure that program content, interstitial material, and graphic overlays (bugs, crawls, etc.) within in a given program are either entirely SDR/BT.709 or entirely PQ10 or entirely HLG10, to the extent possible. Section 7.2 offers details on conversions at Production and Distribution and Section 11 has details on conversions in consumer equipment for backward compatibility.

This section addresses equipment that is compatible with UHD Phase A content streams. Methods for addressing Backward Compatibility for legacy decoders and displays are discussed in Section 11. Note that decoders that support only 8 bits are not considered UHD Phase A decoders. (This was the first generation of “UHD” decoders.)

10.1 Decoding

Phase A consumer decoder device capabilities:

- Able to decode HEVC, Main 10 Profile, Level 5.1
- Able to process BT.2020 [4] color space
- Able to process PQ transfer characteristics
- Able to process HLG transfer characteristics
- Able to process HDR10 content (with or without metadata)
- For the STB-display architecture, the STB also supports the following:
 - Output Interface – HDMI 2.0a/b*
 - Optionally able to transmit ST 2086 [11] metadata, MaxCLL, and MaxFALL to the connected display device
- Able to decode multi-channel Dolby AC-3, E-AC-3, DTS-HD, HE-AAC and AAC-LC audio
- Able to decode closed captions and subtitles per CTA- 608708, ETSI 300 743, ETSI 300 472, SCTE-27, or IMSC1

*Note that HLG transfer function requires at least HDMI 2.0b interface.

10.2 Rendering

The characteristics of Phase A consumer display devices differ significantly from those of professional displays used to grade and master the content. These characteristics include luminance



range, color gamut, screen size (smartphone, tablet, TV), and more. In order to compensate for these differences, Phase A consumer display devices are capable of processing incoming UHD Phase A content so that the rendered video reproduces the creative intent as optimally as possible, for example by appropriate color volume transformation of an HDR/WCG video signal to the display panel.

Phase A consumer rendering device capabilities:

- Able to process PQ transfer characteristics
- Able to process HLG transfer characteristics
- Able to process HDR10 (with or without metadata)
- Able to process BT.2020 [4] color space
- Able to render 60p frame rates
- Able to render content having 2160p spatial resolution
- Able to process multi-channel 5.1 channel surround sound
- Optionally able to render Atmos immersive soundtracks delivered by E-AC-3
- For STB-display architecture:
 - Input Interface – HDMI 2.0a/b*
 - Transmission of Extended Display Identification Data (EDID) information including peak and minimum luminance
 - Transmission of supported EOTFs
 - (Optional) Transmission of RGB primaries

*Note that HLG transfer function requires at least HDMI 2.0b or later interface.

10.3 Overlays Inserted at the Consumer Device

Closed captions, subtitles and graphic overlays may be rendered by a STB, a display connected to a STB, or an integrated decoder/display. In the case of the STB-display architecture, it is possible that both the STB and the display are rendering overlays at different times or simultaneously (e.g., STB rendering an EPG and the display rendering a volume control indicator).

The current specifications regarding closed captioning and subtitles are based on BT.709 [3] color space and SDR. When overlaying closed captions and/or subtitles onto BT.2020 [4] color space and HDR video, Phase A decoders should remap RGB values of closed captions and/or subtitles as needed to ensure that color shifts do not occur while mixing two elements having different color spaces and/or dynamic ranges.

Similar care should be taken when displaying graphics, EPGs, user interface overlays and so on.



11. Format Interoperability

There are a number of requirements for format interoperability, when considering the needs of broadcasters or service providers working with both UHD Phase A and HD (and even SD) content. One example of this is Backward Compatibility, i.e., the means of delivering a UHD Phase A service to a legacy consumer device, in such a way that it can be viewed at some level of acceptable quality on an SDR/BT.709 display.

Backward compatibility that conveys the full creative and artistic intent of the original UHD Phase A content is not attainable. UHD Phase A gives producers, camera operators, directors, editors, production designers, etc. more creative possibilities. Since legacy screens will not be able to display the full resolution, dynamic range and color gamut of the original production, some of the original creative intent will be lost.

UHD Phase A services are distributed via OTT or MVPD. (OTA broadcast services were not commercially offered in 2016 and thus are not part of Phase A.)

This section addresses Phase A backward compatibility for the installed base of SDR 2160p TVs, both BT.709 [3] and BT.2020 [4] displays. Thus, not all facets of UHD Phase A content are considered for Backward Compatibility in Phase A. Specifically:

- Spatial resolution down-conversion is not in scope; only 2160p-capable decoder/displays for UHD Phase A content are included
- Frame rate down-conversion is not in scope; only 50/60 Hz-capable decoder/displays for UHD Phase A content are included
- HDR and WCG are the primary parameters being considered for Phase A backward compatibility

Backward compatibility for OTT and MVPD services involves either:

- For HLG10:
 - HLG10 technology is designed to produce acceptable results using the same content stream on both HDR and SDR devices, provided that the SDR device can process BT.2020 [4] color space. Validation of the quality of the SDR output is being explored as of the publication of this document. In the event that the SDR rendering of HLG10 content does not produce acceptable results, schemes similar to those proposed for HDR10/PQ10 may be used (see below).
- For HDR10 or PQ10
 - Simulcasting multiple broadcast streams, one in HDR10 or PQ10 and the other in SDR/BT.709 (see Section 11.2), and/or
 - Using a STB that can decode the UHD Phase A stream and deliver material suitable for an HDR/WCG, HDR/BT.709, or SDR/BT.709 display. In the case of HDR10, the STB may be able to take advantage of the HDR10 static metadata 6.1.5, when present, in creating the down-conversion. Ideally, the Phase A STB is capable of serving any of these displays so that when a consumer decides to take advantage of HDR services, only a new display is needed.



Creating a backward compatible version of the content that is acceptably rendered on a 2160p SDR/BT.709 display may take place at various points in the supply chain between production and the display (see also Section 7.2):

- Content producers can generate both HDR/WCG and SDR/BT.709 versions, applying creative intent adjustments to both versions. This results in the highest quality conversion but requires time and resources and both versions must be carried throughout the supply chain. This option may not be practical for Live content workflows.
- Professional equipment can down-convert HDR/WCG to SDR/BT.709 with or without the benefit of creative adjustments. This equipment may be sophisticated and thus may be the best option if automated conversion is necessary.
 - HDR10 static metadata 6.1.5, when present, may assist this process.
- Consumer equipment (i.e., STB) can down-convert HDR/WCG to SDR/BT.709 without the benefit of creative adjustments. This equipment is likely to be less sophisticated than professional equipment, but may be a viable alternative when it is impractical to offer multiple versions of the content to the consumer premises.
 - HDR10 static metadata 6.1.5, when present, may assist this process.

11.1 Legacy Display Devices

In Phase A, the Ultra HD Forum is considering legacy display devices that are connected to the MVPD STB or are receiving a suitable unicast stream from the OTT provider. In the latter case, the OTT provider offers a suitable stream, and it is up to the provider to determine which devices it can support. A STB that can ingest a Phase A 2160p stream and output a stream that a legacy display device can render is considered. The variety of legacy display devices that a STB can accommodate varies by product as does the quality of the down-conversion.

A backwards compatible distribution solution or STBs capable of down-conversion can address first-generation 2160p SDR televisions, i.e., devices that can render 2160p resolution content with BT.709 [3] or BT.2020 [4] color gamut but only in SDR. In the absence of one of these solutions, a direct IP stream can be used to address HDR TVs, e.g., using an embedded HTML5 [33] or RVU client¹⁶ that extracts the received broadcast stream and re-encapsulates it into an IP stream that can be transmitted to a TV via a Local Area IP Network. Note that currently some UHD displays are capable of accepting BT.2020 [4] content, but as of 2016, no direct view display is available that is capable of rendering the full gamut of colors in the BT.2020 [4] color space. It is assumed that in these cases, the device employs “best effort” color volume transform tailored to its particular display characteristics, and thus these devices are considered BT.2020 [4]-compatible for the purpose of this discussion.

11.2 Down-conversion at the Service Provider

This option may be employed by OTT providers or by MVPDs. With this method, providers offer both UHD Phase A and legacy versions of the service and send the appropriate stream to devices

¹⁶ See <http://rvualliance.org/what-rvu>.



(unicast) or simulcast both streams. In general, providers that use DASH as a transport method may use unicast and providers that use MPEG-2 TS may use simulcast. The variety of legacy devices served is a function of how many different streams a given service provider chooses to make available. This method may require content producers to deliver multiple versions of the content to the service provider and/or utilize professional conversion equipment at the headend (see Section 7.2).

The below diagram illustrates this method of backward compatibility.

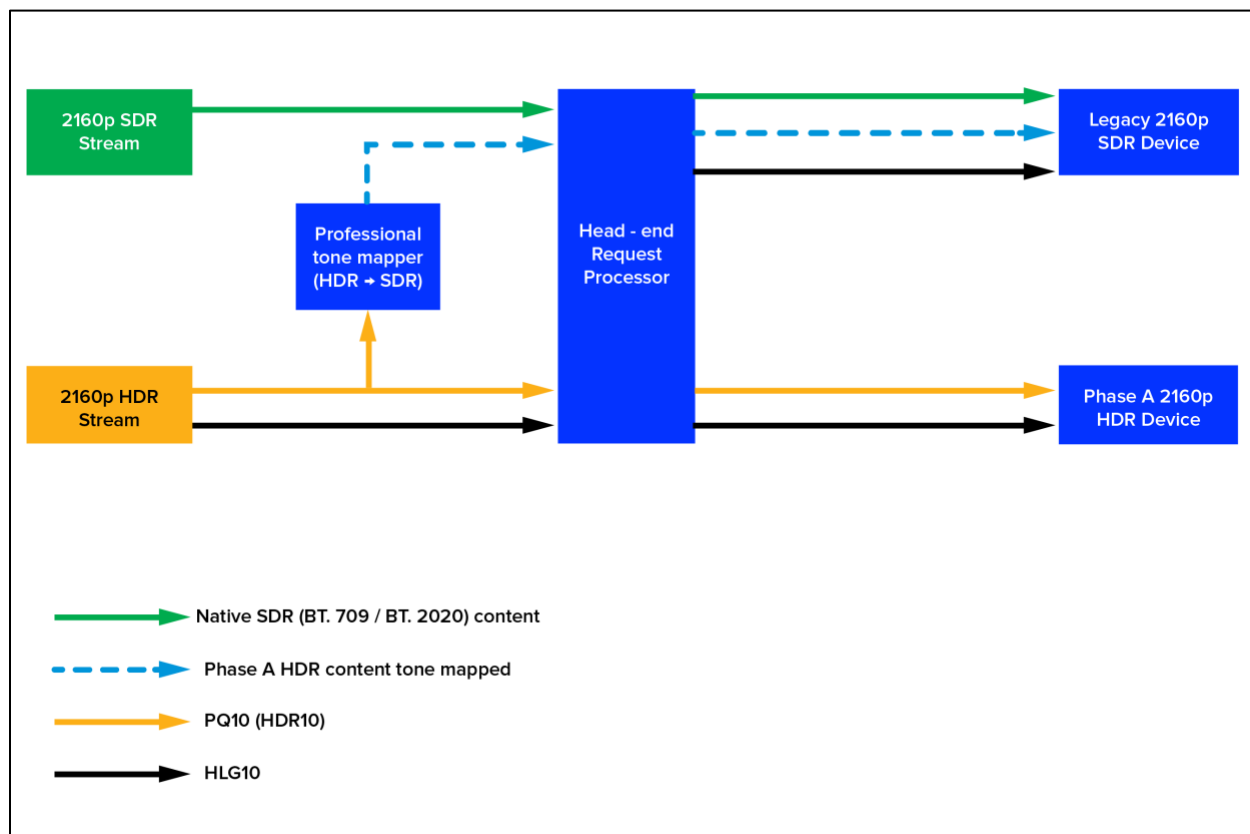


Figure 18 Down-conversion at the Headend

In the above diagram:

1. Operator receives legacy and UHD Phase A content from different content sources.
2. Operator can convert* UHD Phase A streams for legacy 2160p SDR devices.
3. Device requests content from headend based on its capabilities.
4. Headend request processor provides appropriate stream.

*Note that conversion could occur upstream of the headend; i.e., the content producer could provide the operator with both SDR and HDR versions of the content.

11.3 Down-conversion at the STB

This option may be employed in Phase A by MVPDs that prefer not to use the bandwidth required for offering multiple unicast streams, such as via switched digital video technologies, or multiple



simulcast streams. In this case, the STB is capable of decoding a Phase A stream and is also capable of down-converting the content. As stated above, there are compromises with down-conversion of UHD Phase A content and service providers should test the quality of the output for acceptability.

Although there is no standardized method of down-converting BT.2020 [4] to BT.709 [3], it is expected that some STBs may have this capability. STBs may also have the capability of down-converting PQ10 or HDR10 to SDR. The diagram below illustrates this method.

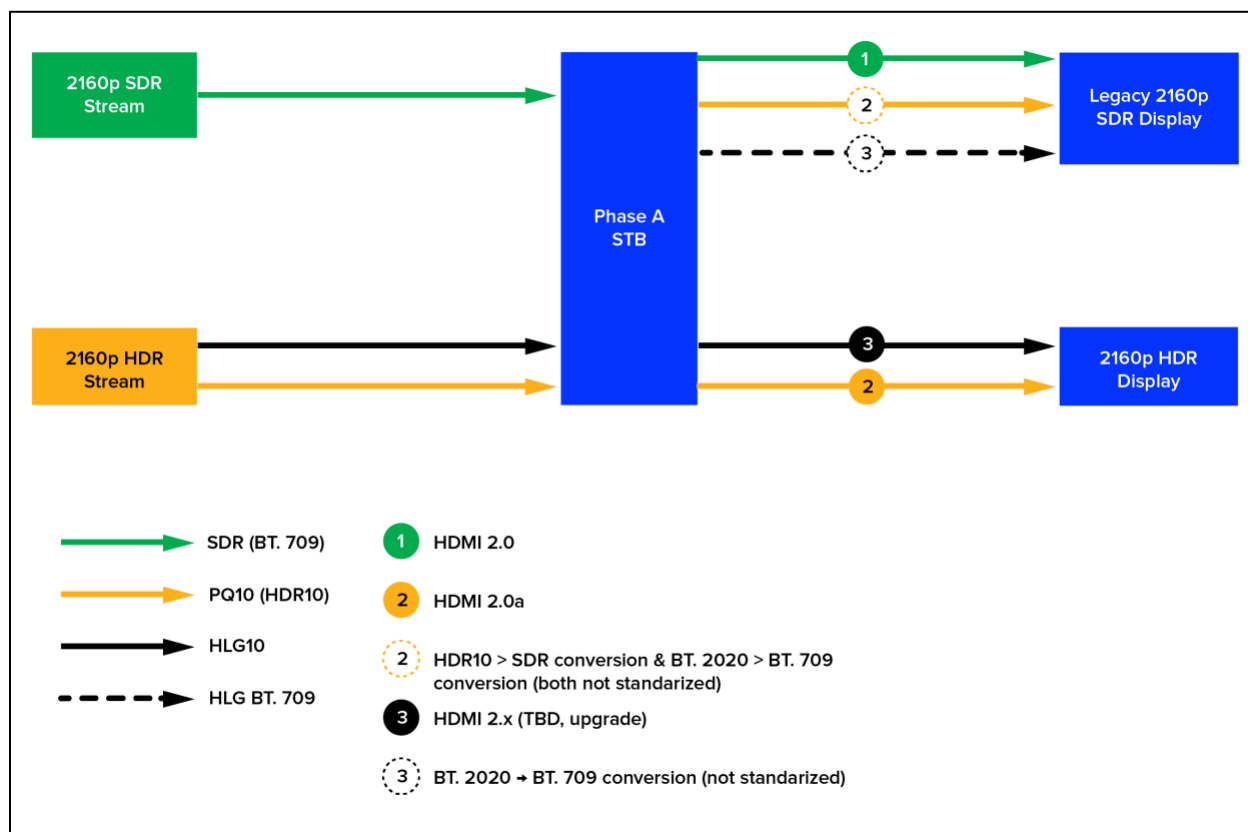


Figure 19 Down-conversion at the STB

Details:

- The Phase A STB supports HEVC, Main 10 Profile, Level 5.1, BT.2020 [4], and HDMI 2.0a and optionally IP output.
- In this example, the legacy 2160p SDR display supports BT.709 [3] but does not support BT.2020 [4].
 - Therefore, in the diagram, the Phase A STB would convert the video from BT.2020 [4] to BT.709 [3] before transmitting it to the legacy 2160p SDR display.
 - Note that some legacy 2160p SDR displays may support BT.2020 [4] and for these displays, a Phase A STB does not need to convert from BT.2020 [4] to BT.709 before transmitting to the TV.



11.4 Spatial Resolution Up-conversion of Legacy Services

This option may be employed in Phase A by MVPDs that prefer not to use the bandwidth required for offering multiple unicast streams, and when Phase A STBs are not be able to convert a UHD Phase A stream to an appropriate format and/or with sufficient quality for display on a legacy 2160p SDR display. Phase A STBs (as well as legacy 2160p SDR displays) are expected to have the capability of upscaling 720p/1080i/1080p SDR channels to 2160p resolutions. This option requires simulcasting; however, the 720p/1080i/1080p SDR stream/service often already exists, e.g., during a transition period. In this case, the legacy 2160p SDR display gets the legacy stream and up-converts the spatial resolution to 2160p. Only the 2160p HDR display gets the UHD Phase A stream. There are compromises with up-conversion of 720p/1080i/1080p content and service providers should test the quality of the output for acceptability.

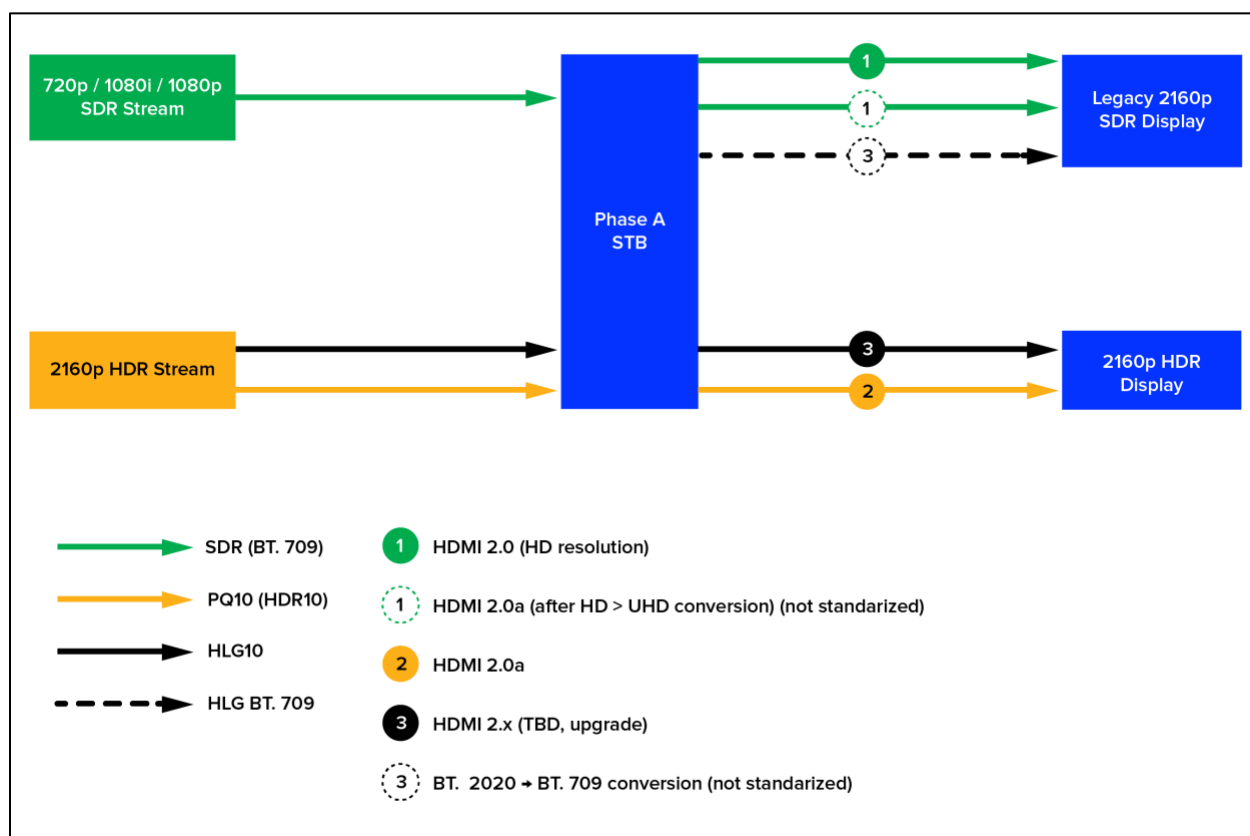


Figure 20 Spatial Resolution Up-conversion of Legacy Services

Details:

- The Phase A STB decodes the 2160p HDR stream when connected to 2160p HDR displays.
- The Phase A STB decodes the 720p/1080i/1080p SDR stream when connected to legacy 2160p SDR displays. The STB can either transmit the decoded



720p/1080i/1080p SDR video or convert the 720p/1080i/1080p SDR video to 2160p SDR video before transmitting it to the legacy 2160p SDR display.

- Note that SDR to HDR conversion, if needed, is best performed in the display device rather than in the decoder device.



11.5 Interoperability of Atmos Channel-based Immersive Audio

For emission of Atmos Channel-based Immersive Audio, an E-AC-3 encoder with ETSI 103 420 [35] functionality is required. Internally, the encoder will create a backward compatible 5.1 channel version rendered from the 7.1.4 (or 7.1.2 or 5.1.2 or 5.1.4) input. This 5.1 channel render is E-AC-3 coded as normal and information about the render, as described in ETSI 103 420 [35], is also carried. Legacy E-AC-3 decoders will reproduce the backward compatible base surround program while advanced E-AC-3 decoders, compliant with ETSI 103 420 [35] will reproduce the full 7.1.4 (or 7.1.2 or 5.1.2 or 5.1.4) immersive audio program.



12. Annex A: Real World Phase A UHD Deployments

This annex describes several “real world” use cases of Phase A UHD deployments. The Ultra HD Forum gratefully acknowledges the many organizations that contributed their experiences to this document.

12.1 CBS and DirecTV Major Golf Tournament

Thanks to CBS and DirecTV for providing this information to the Ultra HD Forum about this workflow.

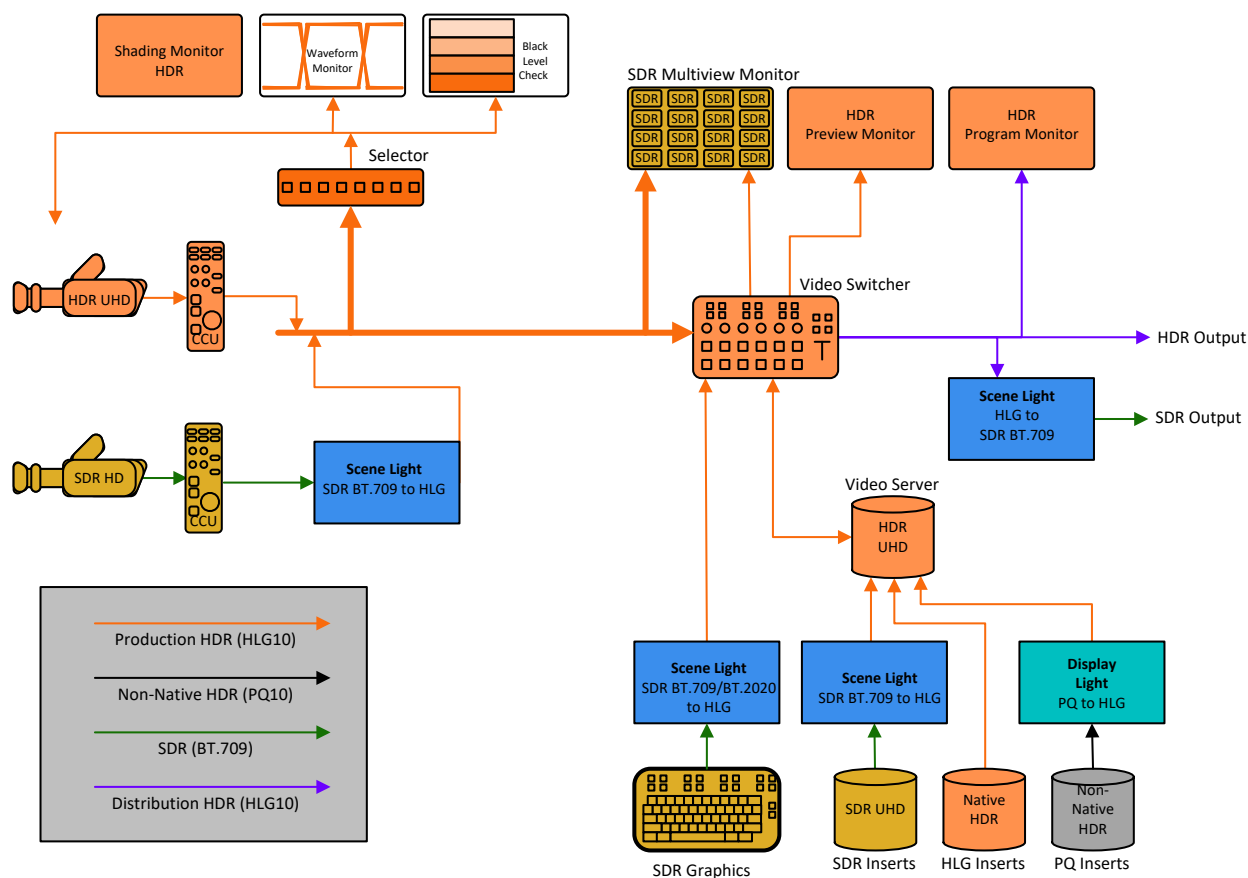


Figure 21 CBS and DirecTV Major Golf Tournament Workflow

For the production illustrated in this diagram, five holes of a golf course were captured in 4K HDR using Sony 4300 cameras operating in BT.2100 HLG mode. HDR content is mastered at 1,000 nits, with BT.2020 color.

The remaining elements of production were captured in SDR.



The video server shown in lower right of the diagram is used for replays and for inserting motion graphic elements. SDR interstitials and SDR graphics are inserted in a “Scene light” manner and switched through the Video Switcher. Graphics, legacy SDR inserts and interstitials are native SDR and are converted to HDR with 100% SDR mapped to approximately 75% HLG. There is no highlight expansion of SDR sources.

An SDR version of the program is created as is shown in the “SDR Output” source on the right side of the diagram.

12.2 Amazon Major Parade

Thanks to Amazon for providing information to the Ultra HD Forum about this workflow.

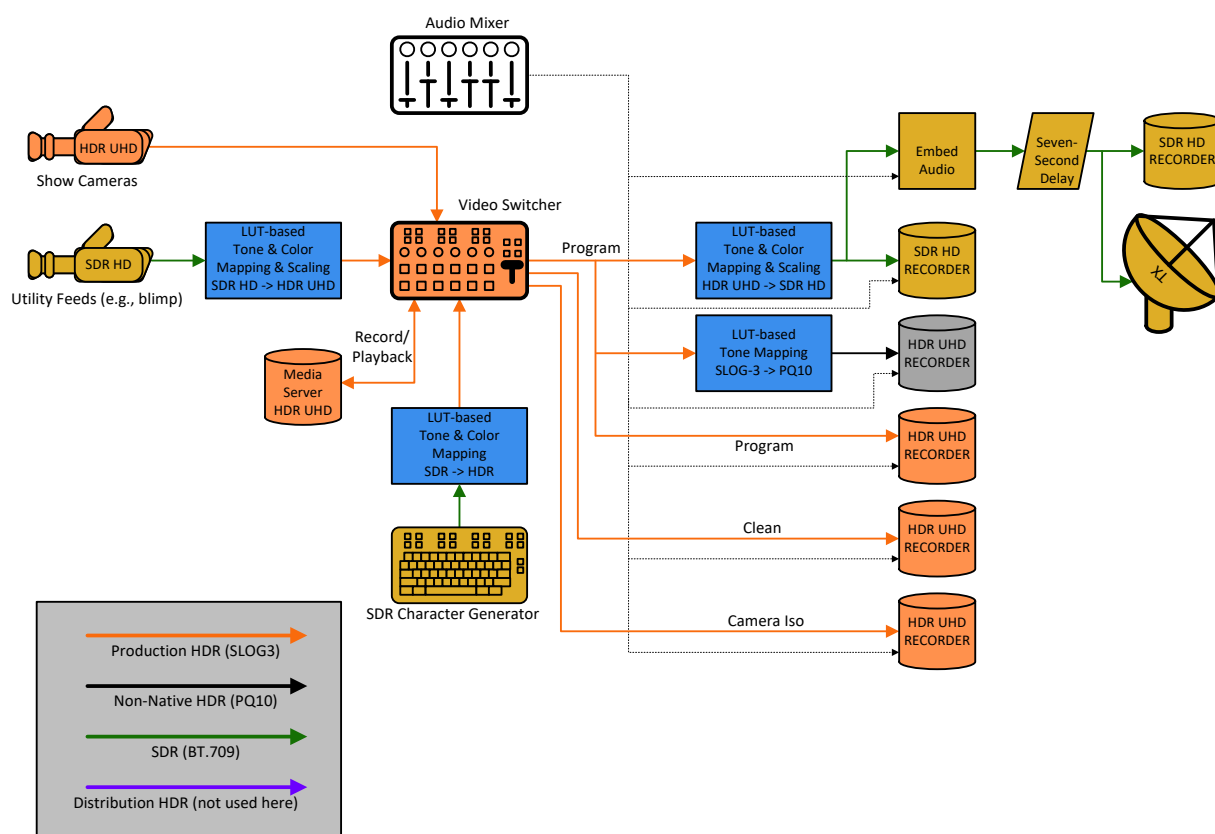


Figure 22 Amazon Major Parade Workflow

12.3 NBC Universal Olympics and 2018 World Cup

Thanks to NBC Universal (NBCU) for providing information to the Ultra HD Forum about their “production HDR” to “PQ Distribution HDR” for linear (live) and VOD workflows.

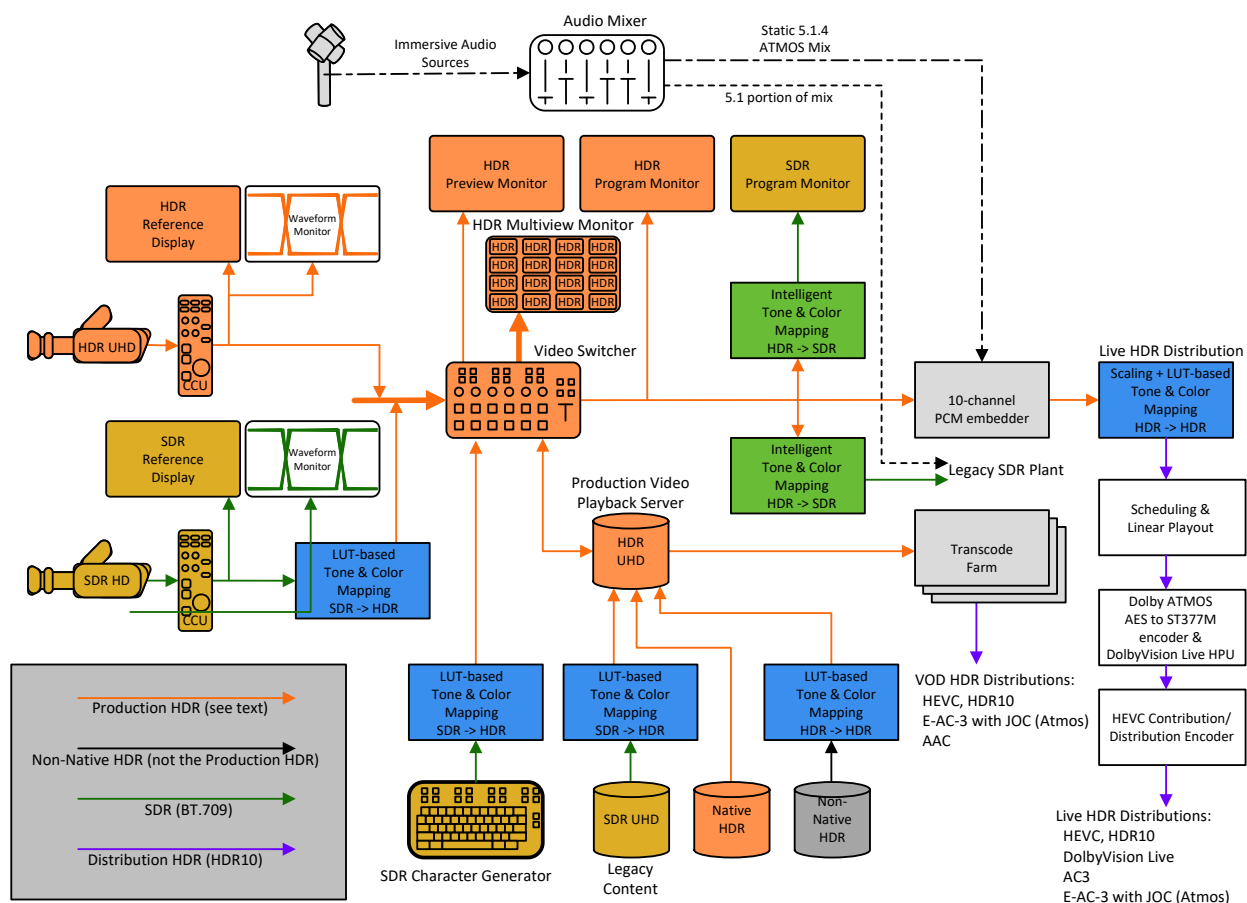


Figure 23 NBCU Olympics and 2018 World Cup UHD Workflow

This workflow includes three different production formats: S-Log3 Live, HLG, and PQ. NBCU chose PQ as a final distribution format, and converted all other formats, including SDR, to this common production format. The Pyeongchang Olympics were captured in HLG and the World Cup was captured in S-Log3 Live.

NBCU has strong preference to for PQ for the distribution format, citing that PQ ensures consistency in terms of final delivery of content quality. NBCU has found S-Log3 provides a good basis for conversion to PQ. BT.2020 color is used.

For shading, NBCU chose a starting point is of diffuse white at 200-300nits and then balances SDR and HDR for the best tonal range.

12.4 Sky/BBC Royal Wedding and Major Tennis Tournament

Thanks to Sky for providing information to the Ultra HD Forum about their distribution of the Royal Wedding in the UK in 4K resolution, a joint UHD production between Sky and the BBC. See <https://news.sky.com/story/world-first-for-sky-news-royal-wedding-coverage-11355412> for additional details.

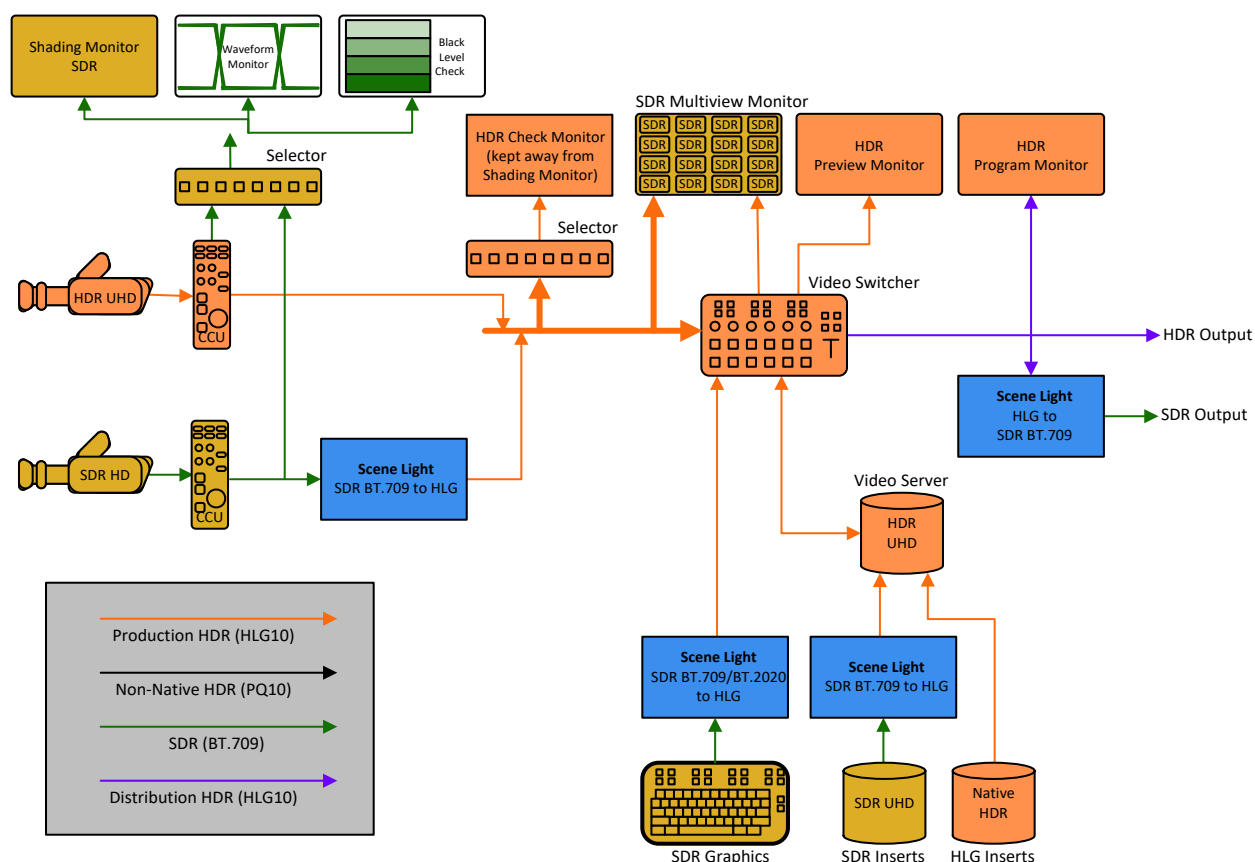


Figure 24 Sky/BBC Royal Wedding and Major Tennis Tournament Workflow

Although the BBC captured their portion of the content in wide color gamut BT.2100 HLG HDR, the main delivery for this event was HD in SDR/709 with Sky distributing UHD SDR/709 in the UK. The use of a single production workflow was necessary and SDR/709 the prime delivery, but maintaining the capability to capture the wedding chapel in HDR was a key desire. Sky worked with the BBC to convert BT.2100 HLG to SDR/709 for the main Sky 4K and HD distribution feed. Numerous tests and trials were conducted in advance to ensure the conversion LUT was working satisfactorily across all colors and luminances. The initial “technically” correct LUTs followed the Macbeth chart closely, but when the color volumes were pushed by adding shiny saturated colored surfaces and glare (using a custom-built test chart), the chroma matching to SDR-matrixed cameras strayed. This had to be addressed due to the highly saturated and nature of the fabric of the guards’ uniforms and other elements of the event that were expected to challenge the BT.709 capabilities. See also <https://www.nepgroup.co.uk/post/the-royal-wedding-in-high-dynamic-range> and <https://www.bbc.co.uk/rd/blog/2018-05-ultra-high-definition-dynamic-range-royal-wedding-uhd-hdr> for additional details.

Sky notes that while there were challenges, the event was a good experience and a big step toward having a single live production using both UHD HLG HDR camera outputs and SDR/709 cut seamlessly together delivering both HDR and SDR, UHD and HD.

A similar workflow was adopted by NEP for their Centre Court coverage of the Wimbledon 2018 tennis championships (<https://www.nepgroup.co.uk/post/live-from-wimbledon-host-broadcaster-pacing-like-an-expectant-father-as-coverage-goes-in-house-for-the-first-time>). This



time, however, a mixture of HLG HDR and SDR/BT.709 specialist cameras were used. In order to ensure a good color match between HDR and SDR cameras, “scene-light” conversions were utilized and the signal “clippers” on the SDR cameras relaxed to EBU R103 levels (-5%/+105%) to extend their effective color gamut.



13. Annex B: Over-the-Air Broadcast Use Case

Commercial deployment of Ultra HD services via OTA delivery did not commence in 2016, and is thus not a Phase A use case. In February 2017, Korea launched terrestrial UHD TV commercial services using ATSC 3.0. Japan, the U.S. and Europe may also be adopters.

OTA broadcast of 2160p content is an expensive proposition in terms of pixels/Hz. Some broadcasters, such as those in Korea, may have access to sufficient spectrum to deliver Ultra HD content in 2160p resolution. However, in other parts of the world, such as the U.S. or Europe, broadcasters' network capacity may be limited, especially if legacy HD/SD simulcasting and/or channel sharing is necessary. In this case, broadcasters may choose to offer advanced services in 1080p, 50/60fps, HDR/WCG format, which may be deployed in under 10Mbps. Where simulcasting and channel/sharing are not necessary and HEVC compression is sufficient, 2160p content can be broadcast OTA. It is possible that in some countries broadcasters will use satellite to deliver an Ultra HD experience until spectrum is allocated on terrestrial networks.

Japan has announced that 8K resolution services are on their deployment roadmap for 2018 by satellite and optical fiber. Terrestrial broadcast technologies are still in development. With technology available in 2016, 8K resolution bitrates are expected to be in the range of 100Mbps; thus, only satellite can enable the transmission of 8K resolution services. For example, in Japan BS Digital is already transmitting 8K content via satellite.

In countries where bandwidth is constrained, the expectation is that a single HDR/WCG service with direct backwards compatibility may be most desirable (i.e., not simulcast with a separate stream in SDR/BT.709). Because broadcasters target TVs directly, they must also consider backward compatibility with deployed consumer electronics or other infrastructure that is not capable of processing some aspects of Ultra HD content so simulcasting may be a necessity nonetheless.



14. Annex C: IC_{TCP} 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_BC_R (hue linearity and constant luminance) will become more prevalent and objectionable.

IC_{TCP} is a color representation designed for high dynamic range (HDR) and wide color gamut (WCG) imagery and is intended as an optional replacement for Non Constant Luminance Y'C_BC_R with HDR and WCG signals. IC_{TCP} is a more perceptually uniform color representation that is based on the human visual system. The improved decorrelation of saturation, hue, and intensity make IC_{TCP} ideal for the entire imaging chain from scene to screen. IC_{TCP} follows the same operations as NCL Y'C_BC_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_{TCP} allows for complex tasks such as color volume transform to be easily performed on HDR and WCG imagery with minimal error.

IC_{TCP} is included in BT.2100 [6] and is being deployed by OTT service providers as well as implemented by numerous consumer TV manufacturers.

15. Annex D: ACES Workflow for Color and Dynamic Range

The ACES [51] project provides a thorough workflow that can be used to model the processing of HDR/WCG video signals from source to display in a series of stages that mark the boundaries of significant transformations, each with a specific purpose. Most, if not all, Ultra HD Blu-Ray (BD-ROM 3.1) content were authored in ACES workspace.

Source code and documentation for ACES is available at: <https://github.com/ampas/aces-dev/>

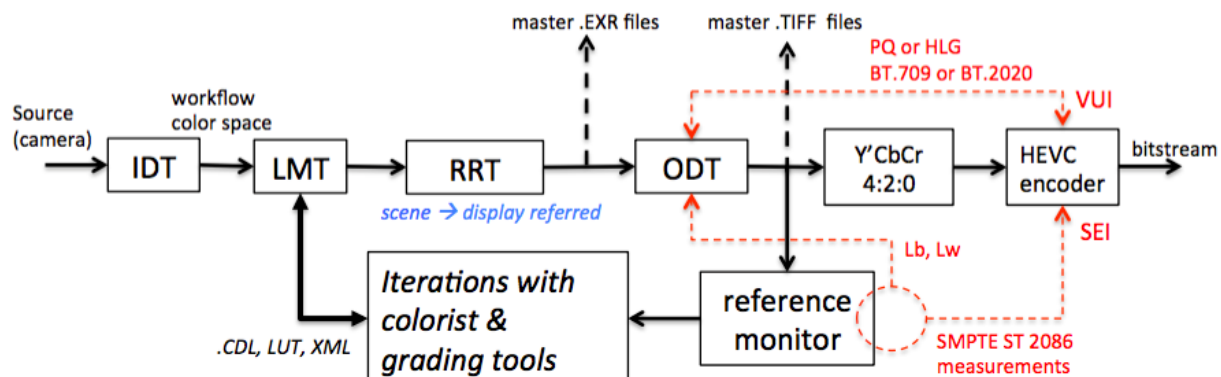


Figure 25 ACES Workflow Model

The basic ACES workflow model stages are described in the following table.

Table 18 ACES Workflow Model

ACES Stage	Purpose
IDT	Input Device Transform: camera format (Bayer RAW, Slog3, etc.) to workflow color space (ACES)
LMT	Look Modification Transform provides an appearance such as “dark night”, “indoor lighting”, etc. established by cinematographer.
RRT	Reference Rendering Transform: converts scene referred signal to display referred signal, with knowledge of reference viewing environments and limited display ranges.
ODT	Output Device Transform. Maps display referred to a specific display range (black and peak white levels), container color primaries (BT.709, BT.2020), and transfer function (gamma, PQ, HLG).

In the diagram, the blue annotates the Reference Rendering Transform (RRT, e.g., “scene linear → display referred”). The red detail indicates metadata input to the encoder for the color space and transfer function (VUI = video usability information). This could include the Master Display Color Volume (MCDV) metadata (e.g., for HDR10). VUI can optionally populate an



Alternative transfer characteristics SEI message to support backwards compatibility. (See also Section 6.1.9.)

The UHDTV color grading process will start with ingesting digital camera rushes or scanned film at whatever the useable resolution, gamut and dynamic range is available from the source material. If the source content is in a format specific to the capture device, the source signal will undergo transformation to a more universal processing space in a stage such as the IDT depicted above.

Colorists working on UHDTV projects are likely to continue the practice of first setting the overall mood of the film or program, then deciding how particular scenes fit that mood and finally how the viewer's interest is directed to characters, objects etc. on specific shots. A set of look-modification transforms, conceptualized in the LMT stage depicted above, reshape content according to the intent of directors, cinematographers, and colorists.

Source material will not necessarily be Rec. 2020 or Rec. 709, (unless it is a re - mastering or restoration project), because there are widely different capabilities in cameras, film stocks etc. Therefore the common starting point for colorists or compositors, will be to bring in all what's available, as this allows more scope in post production.

The colorist or compositor will then make decisions about what to select from that available resolution, range and gamut and how to present it as "legal" PQ10 or HLG10 based content to UHDTV consumer screens, which support Rec. 2100 containers. The display rendering stages (RRT and ODT) shape content to fit within the capabilities of a range of target displays, modeled by the mastering display monitor. The colorist will re-grade content and adjust the look based on the appearance of the rendered content on the mastering display reference monitor used to preview the final appearance. A more advanced workflow configuration could account for additional distortions added by reduced integer precision, $Y'CbCr$ color space conversion, 4:2:0 chroma resampling, and video codec quantization by feeding the output of the decoder to the reference monitor.

Additional considerations would be needed if format interoperability (back compatibility) were being attempted, for example to 4K SDR / Rec. 709 or HD SDR / Rec. 709. Producing deliverables in a Rec. 709 / HDR rendered format is not recommended and it is not clear what the long-term market use case for this would be. HDR and WCG are intrinsically linked in the HSL (hue, saturation, and lightness) or RGB color spaces and most importantly also in the way humans 'see'. They are different dimensions of the unified perceptual experience.

16. Annex E: ISO 23001-12, Sample Variants

For Forensic Watermarking as described in Section 7.2, transport of Variants can be done by different mechanisms. One alternative is transport at the container layer

Variants metadata can be transmitted in an alternate transmission channel next to the video content at the container layer. For instance, Variants metadata could be placed within a MPEG2-TS bitstream using a dedicated Packet Identifier (PID) and using the Program Clock Reference (PCR) to synchronize the video and metadata channels. Alternately, Variants metadata could be incorporated as an extra track in an ISOBMFF [29] file. In that case, synchronization can be achieved by aligning samples across different tracks. When Variants metadata is handled as a component separate from the video, proper care shall be taken to guarantee its protection if needed with relevant content protection techniques.

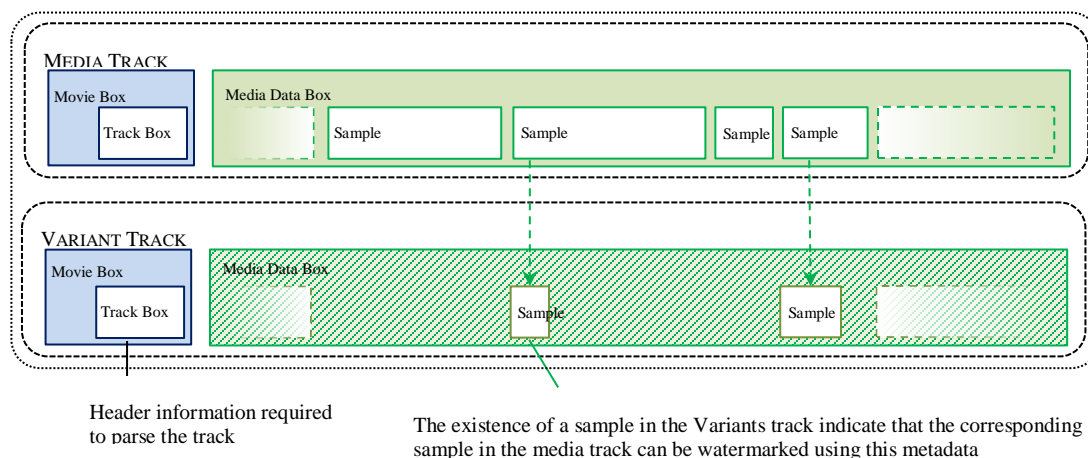


Figure 26 Transport at the Container Layer Using a Track in an ISOBMFF File

An example of how to transmit Variants metadata as an extra track in an ISOBMFF [29] file is described in [41]¹⁷. This standard applies to file-based delivery, e.g., on physical media with embedding on the client side. The baseline principle is to define a dedicated track in an ISOBMFF file that describes how to construct watermarked video samples. For instance, a constructor for a sample indicates which portion of the video track shall be kept and which portions shall be replaced by a Variant available in the variant track. Access to the MPEG variant constructors is subject to cryptographic keys. Different users/devices will have a different sets of keys and thereby would be able to only produce different watermarked video samples using different constructors. Moreover, the Variants are double encrypted to serve as an anti-bypass mechanism. A player that would not perform the watermark embedding operation would not be able to display a good quality video since some segments of the video would still be encrypted. The strong link between

¹⁷ It shall be noted that the terminology “variants” is slightly different in the MPEG standard and these UHD guidelines. In the MPEG standard, a variant is a full MPEG sample composed of parts of the original bitstream and parts of the Variants, as defined in this document i.e. segments of bitstream that can be used interchangeably at a given location in the bitstream.



encryption and watermarking requires collaboration between CAS/DRM and watermarking systems, e.g., for key management and provisioning. The virtue of this design is that it enables a secure integration of the watermark embedding module on open devices outside of the secure video path or trusted execution environment.

End of Document