Ultra HD Forum Guidelines

Green Book – Ultra HD Distribution

Ultra HD Forum

8377 Fremont Blvd., Suite 117

Fremont, CA 94538

UNITED STATES

VERSION 3.3.0

Sep 10, 2024

1. Foreword

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

This work represents over eight years of collaborative effort by the membership of the Ultra HD Forum. The Ultra HD Forum’s Guideline books would not have been possible without the leadership of Jim DeFilippis, who represents Fraunhofer, and chair of our Guidelines Work Group with invaluable support from the co-chair, Pete Sellar of Xperi as well as technical assistance from Ian Nock of Fairmile West Consulting, chair of the Interop Working Group.

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

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

The Rainbow Books are, in their entirety:

White Book	Guidelines Index and References
Red Book	Introduction to Ultra HD
Orange Book	Foundational Technologies for Ultra HD
Yellow Book	Beyond Foundational Technologies
Green Book	**Ultra HD Distribution**
Blue Book	Ultra HD Production and Post Production
Indigo Book	Ultra HD Technology Implementations
Violet Book	Real World Ultra HD
Black Book	Terms and Acronyms

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

I hope you will enjoy reading today.

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

Dr. Yasser Syed, President, Ultra HD Forum
Sept 2024

1.1 Changes from version 3.2 to 3.3

What's new in the Fall 2024 version of the UHDF Guidelines Green Book, *Ultra HD Distribution* (v3.3), edited by Jim DeFilippis.

The *Ultra HD Distribution* is the fourth of the series of Rainbow Books on the Guidelines for Ultra HD. The scope and purpose of this book is to describe the assembly of content and final processing prior to distribution via the multitude of distribution methods (OTA, Satellite/Cable, IPTV, and OTT). Detailed information on the use of Content Aware Encoding (CAE) techniques, and content security technology (watermarking) is included.

This edition has updated references.

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

Jim DeFilippis and Pete Sellar,

Guidelines Working Group Co-Chairs, Ultra HD Forum, April 2024

2. Acknowledgements

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

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

3. Notice

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

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

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

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

4. Contents

5. List of Figures

6. List of Tables

List Of Tables

7. Introduction

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

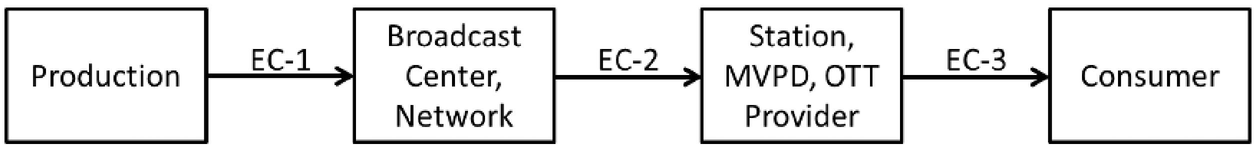

Figure 1. Distribution Nodes and Interfaces

Table 1. Compression and Distribution Nodes and Interface

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 available since 2016. The workflows described are 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 [B01]), 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/subtitles 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. Section 8 of the Indigo Book [I01] describes the Next-Generation Audio (NGA) workflow. For the purpose of Foundation Ultra HD, 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 [26] includes provisions for carrying captions and subtitles in the VUI and SEI in a similar manner to legacy video codecs.

Existing Foundation Ultra HD the production systems are likely SDI-based (1x12G or 4x3G) and therefore deployment of an SDI workflow is likely. Newer production systems have adopted an IP-based workflow, using techniques such as Media over IP: SMPTE ST 2022-6 [82] and near-lossless compression technologies such as VC-2 (Dirac), TICO, JPEG-XS, JPEG 2000, or other vendor proprietary solutions.

Methods of carrying 2160p over SDI defined by SMPTE are shown in Table 2 below.

Table 2. SDI Input Standards for 2160p Content Production

Interface	Standard	Details	Notes
4x 3G-SDI*	SMPTE ST 424 [79]	4 quadrants of 3G-SDI	
	SMPTE ST 425-1 [80]	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 [84]	12Gbps SDI native	

Table 2. notes:

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

Metadata for HDR10 can be carried over SDI in VANC, per SMPTE ST 2108-1 [47]. If HDR10 metadata is present, it can be applied at the video encoder as follows:

- In compressed format, HEVC, Main 10 Profile, Level 5/5.1 may be used for metadata signaling. The metadata is carried via VUI and SEI messages (see Section 7.2.8 HDR10 Metadata Carriage in the Orange Book [O01])
- 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 Foundation Ultra HD 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 Foundation Ultra HD workflows.

8. Production Processing and Contribution

This section describes processes for transmitting Foundation Ultra HD 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.

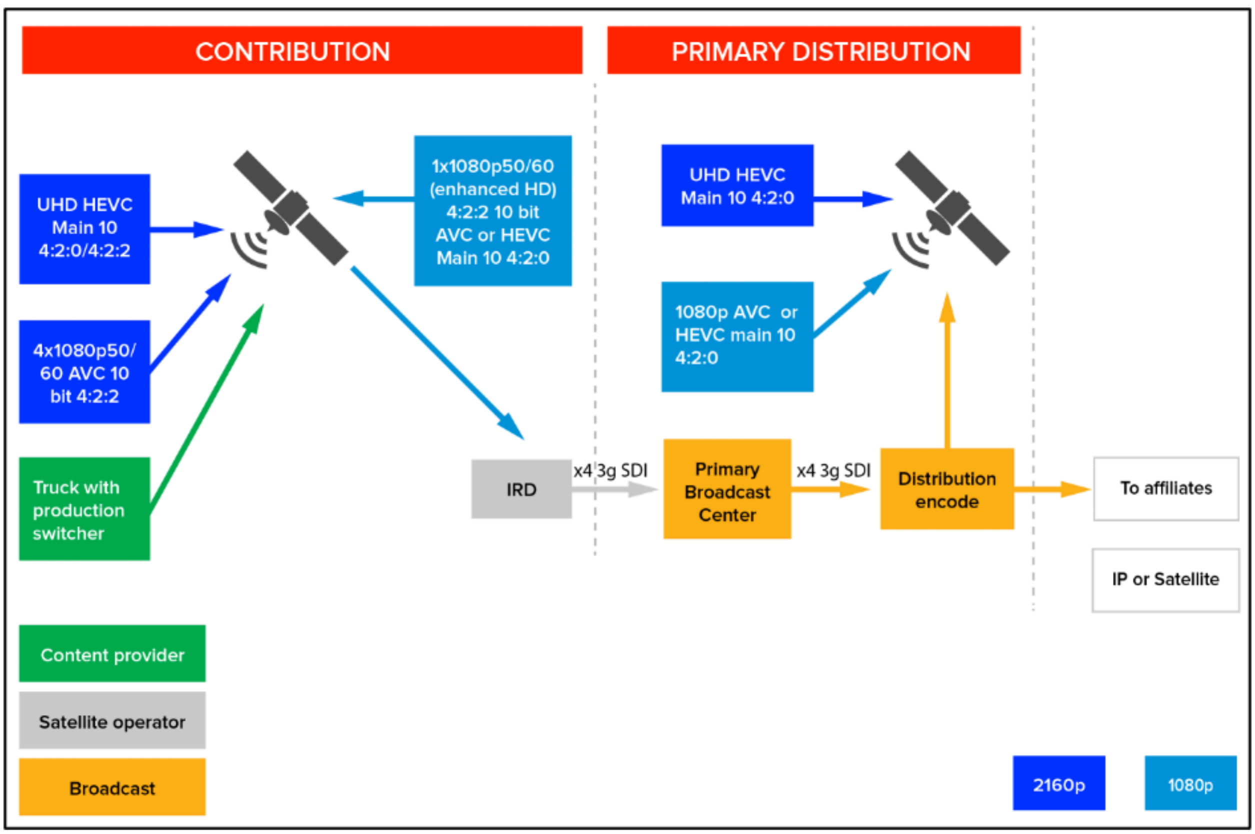

Figure 2. Contribution Workflow

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 8 of the Blue Book [B02] 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 10 of the Blue Book [B03] for details on Pre-recorded content production.

8.1. 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 since 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 system colorimetry 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 can be 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[1] contribution method is recommended for use in Foundation Ultra HD.

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 3. 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	As per MPEG VUI/SEI signaling in MPEG 4 [25]	20 – 50 Mbps
	HEVC, Main 10, Level 5.1 4:2:2/4:2:0	As per MPEG VUI/SEI signaling HEVC [26]	10 – 40 Mbps
	JPEG 2000	Not standardized	100 – 150 Mbps
2160p @ 50/60 fps	AVC 4:2:2 10-bit (4 quadrant)	As per MPEG VUI/SEI signaling in MPEG 4	90 – 140 Mbps total

[1] For use in China, the AVS2 codec, Main10 profile, is used in lieu of HEVC. See AVS2 in Sec 7.4.1 of the Yellow Book [Y01].

	HEVC, Main 10, Level 5.1 4:2:2/4:2:0	As per MPEG VUI/SEI signaling in HEVC	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 Foundation Ultra HD, 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 to bitrate. 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.

8.2. Audio

In Foundation Ultra HD, production practices for audio are similar to those used in current HD content creation. In Foundation Ultra HD audio follows multi-channel workflows established for 5.1 surround for delivery using one of the following emission codecs as appropriate for contribution applications: AC-3, E-AC-3+JOC (an instance of Dolby Atmos channel-based immersive audio), HE-AAC, or AAC-LC.

Beyond Ultra HD Foundational audio includes a set of Next Generation (NGA) audio codecs: AC-4, DTS-UHD and MPEG-H. NGA provides additional capabilities such as personalization, object audio, and immersivity as well as providing enhanced accessibility features such as AD (audio descriptive) and dialog enhancement. More information can be found in the Red Book Sec 9.2 [R01].

8.3. Closed Captions and Subtitles

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

8.4. 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 a DTT, MVPD or OTT provider.

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.

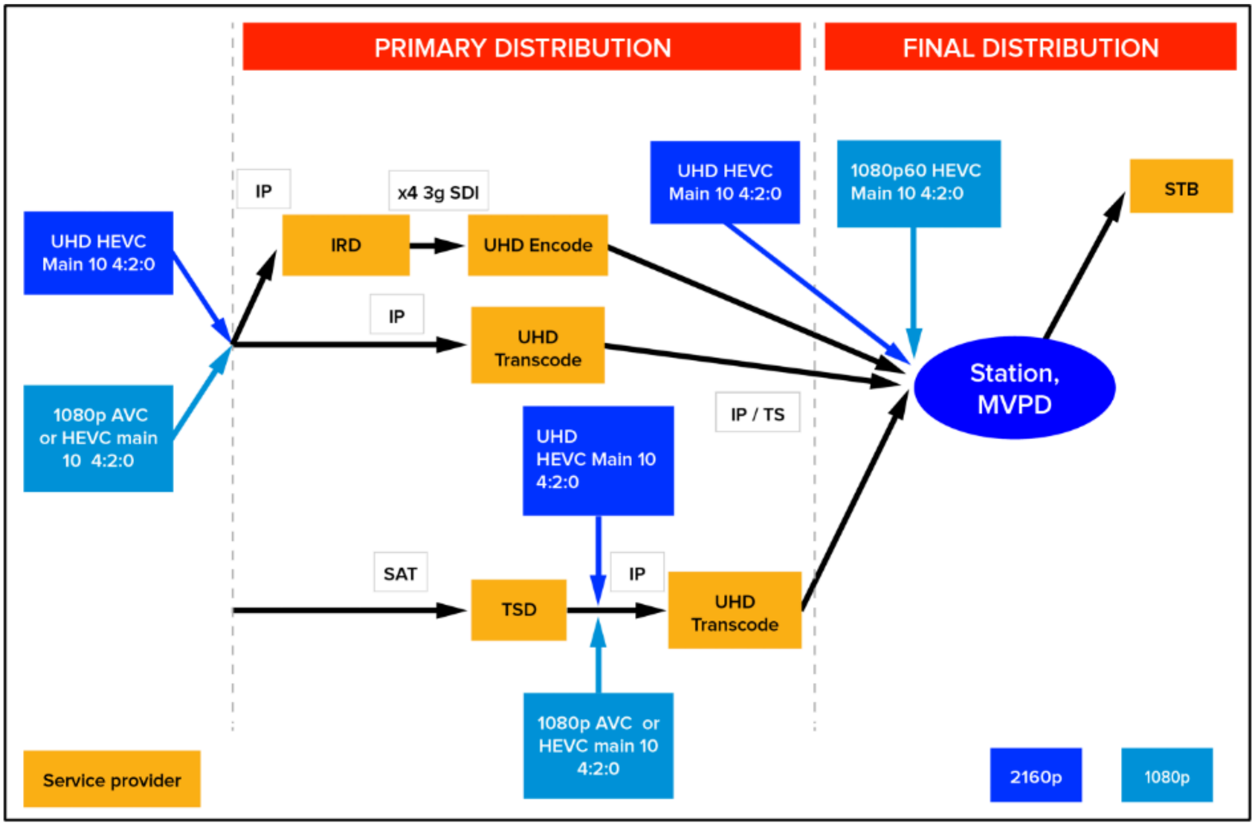

Figure 3. Primary Distribution to an MVPD or TV Station

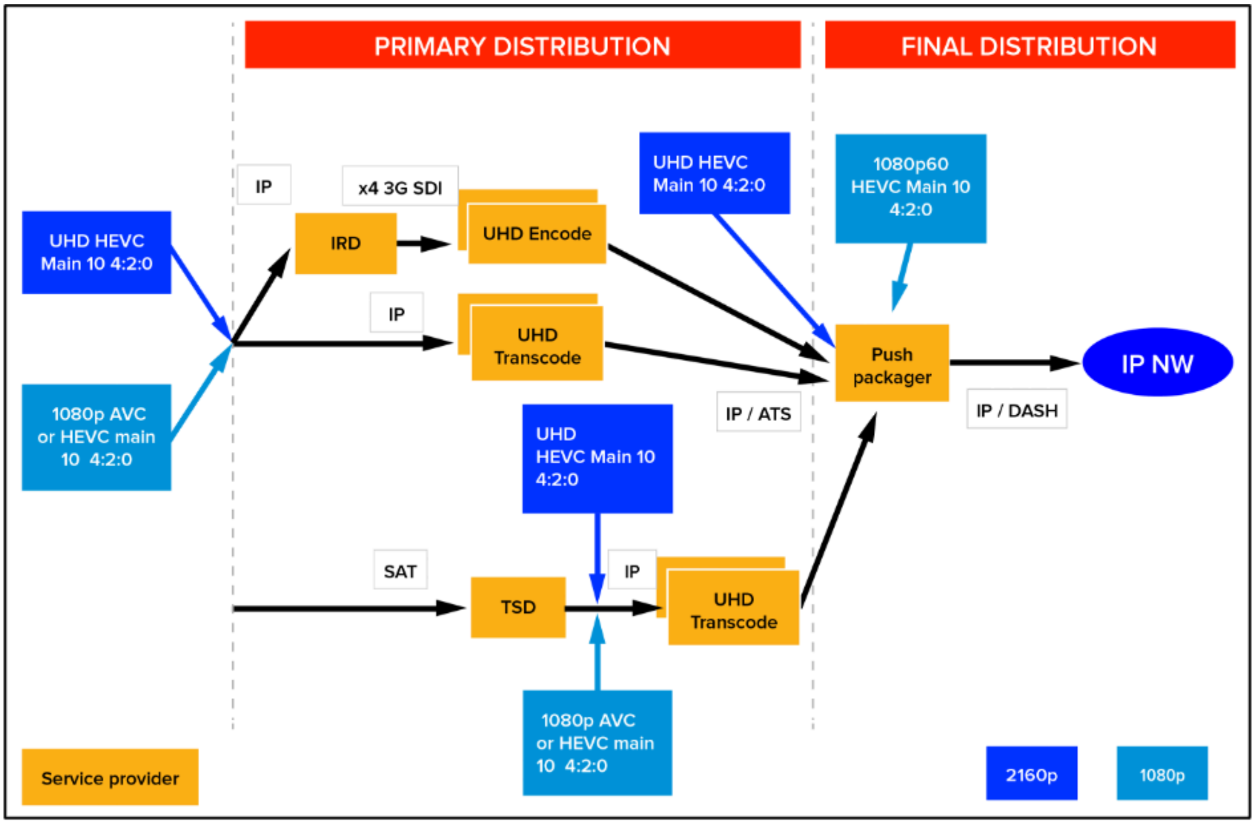

Figure 4. Primary Distribution to an OTT Provider

Figure 3 depicts different mechanisms of Primary Distribution delivery in TS formats to an MVPD. Figure 4 is an example of Primary Distribution delivery in an IP format to OTT distributors.

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 [26]. Table 4 below describes the formats that can be used for Primary Distribution transmission.

Like contribution bitrates, the typical bit rates 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.

Table 4. 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 [26]	VUI/SEI signaling In HEVC	10-20 Mbps
2160p	HEVC Main 10 Profile, 10-bit	VUI/SEI signaling In HEVC	40-50Mbps

8.5. Final Distribution Processing from MVPD/OTT/DTT Providers

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

The MVPD, OTT, or DTT 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. Table 5 below illustrates the bit depths currently in use.

Table 5. Existing Practices for Real-Time Program Service Distribution Formats

Case	Bit Depth*	Color Gamut	Peak White at D65	Color Volume	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	Ultra HD Forum Guidelines

Table 5 notes:

* To ensure proper interchange of video, SMPTE ST.274 [131] section 6.4 specifies that signal levels shall be contained between reference black and reference white (code values 64 and 940 in 10bit) with the exception of overshoots and undershoots.

BT.709-6 [2] specifies a nominal peak of 940 (10-bit). While it includes video data above that, this is typically understood as being reserved for overshoots which are artifacts due to processing.

EBU r103 [106]] defines an Expected Video Range (10-bit code values 64 to 940); a Preferred Range (10-bit code values 20 to 984) and, in accordance with ITU-R Recommendation BT.709, the Total Signal Range (10-bit code vales 4 to 1019). The EBU recommends that the RGB components and the corresponding Luminance (Y) signal should not normally exceed the Preferred Minimum/Maximum range of digital sample levels. The Preferred Minimum/Maximum range is equivalent to -5%/+105%.

The goal of these documents is to provide consistent video levels that pass through an entire broadcast infrastructure, including its encoders and all the way home to a consumer display.

In the U.S. market specifically, EBU r103 doesn't exist and many infrastructure devices do not support signals beyond SDR reference white. It is therefore recommended that infrastructures that may contain devices that only support video between reference black and reference white (legal or narrow range) signals should be pre-legalized to prevent unwanted artifacts caused by low quality signal processing. It may be worth investigating your infrastructure to discover where unwanted clipping may occur prior to final distribution encoding so this is not necessary, but even if all of this is addressed be advised that delivery may occur to some end-point devices that don't support overshoots and may cause clipping.

Finally, in single-stream HDR workflows it is recommended that a "predictive LUT" be provided to the video shader so they can see what will occur to a legalized SDR video signal (otherwise SDR quality may suffer).

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 Foundation Ultra HD services, and Case 4 is preferred. Cases and 1 and 2 are included for context.

In Cases 3 and 4, SMPTE ST 2086 [10] 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 [3] color primaries; however, in current practice the color gamut rarely exceeds DCI P3 [17] primaries.

8.5.1. 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 system colorimetry.

HEVC[2] Main 10 Profile, 10-bit [26] is recommended as the only final distribution or emission codec as shown in Table 6 below, as all Ultra HD decoders support HEVC.

[2] For use in China, the AVS2 codec, Main10 profile, is used in lieu of HEVC. See Sec 7.4.1 of the Yellow book on AVS 2 [Y01].

Table 6. 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 [26]	VUI/SEI signaling In HEVC	5-18 Mbps
2160p	HEVC Main 10 Profile, 10-bit	VUI/SEI signaling In HEVC	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 Table 7 below showing 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 Foundation Ultra HD services. This information was provided voluntarily, and not all volunteers were able to provide all the metrics.

Table 7. Example "Real World" Bitrates

Delivered via	Transfer function	Frame rate	Bit depth	System Colorimetry	Audio codec(s)	Bitrate	Notes
Satellite	SDR	30fps				27Mb/s	
Satellite	HLG or SDR	59.94fps	10-bit	BT.2020	AC3(DD5.1) 384Kb/s DD+JOC Atmos 640Kb/s Total: 1 Mb/s	31 Mb/s	Bitrate works for sports and HDR content
Satellite or IPTV	SDR	50fps	8 or 10-bit		AC-3	24-30 Mb/s	
	PQ	59.94fps	10-bit		AAC, AC-3	32Mb/s	
IPTV	PQ	50fps	10-bit	BT.2020	MPEG2, AC-3, DD Total Bit rate: 1Mb/s	25Mb/s	Sports content
Satellite	SDR	50fps	10-bit	BT.709		30-38 Mb/s	Drama, movie content
VOD	PQ	25fps	10bit	BT.2020	AC3 ,DD + Atmos AC 1Mb/s	19 Mb/s	Based on VBR with avg BR of 19Mb/s

Table 7 notes:

- all the examples in Table 7 are 2160p spatial resolution and use HEVC encoding with 4:2:0 color subsampling.

8.5.2. Audio

In Foundation Ultra HD, 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+JOC Codec [35].

Many broadcasters and operators are governed by regional legislation regarding managing loudness of broadcast audio. In the United States, for example, a DTT or MVPD provider 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. Note that while E-AC-3+JOC Codec delivery of Atmos soundtracks using channel-based production and delivery is described as a Foundation Ultra HD technology, E-AC-3+JOC is also capable of delivering the dynamic spatial objects that are described as an NGA feature in Section 8 of the Indigo Book [I01].

8.5.3. Closed Captions and Subtitles

Production practices for closed captions and subtitles are similar to those of HD content creation in Foundation Ultra HD. Closed captions and subtitles follow workflows established for CTA 608 [18] /708 [19], ETSI 300 743 [20], ETSI 300 472 [21], SCTE-27 [22], or IMSC1 [23]. HEVC carries captions and subtitles in User data registered by Rec. ITU-T T.35 SEI defined in HEVC specification [26], section D.2.36 (syntax) and D.3.36 (semantics).

8.5.4. Transport

Operators deploying 2160p HDR/WCG content over MPEG-2 TS can use the DVB UHD-1 Phase 2 specification. Operators can carry Foundation Ultra HD content over RTP/UDP/IP per DVB IPTV [41] (single bitrate only, not ABR), i.e., MPEG-2 TS over RTP/UDP/IP.

In addition to DVB UHD-1 Phase 2 specification, DTT operators can refer to ATSC A/331 [53] which specifies MPEG Media Transport (MMT) and ROUTE/DASH for carriage of UHD content.

For OTT services, MPEG DASH is used to transport Foundation Ultra HD content as follows:

- DASH per DVB DASH specification [13] for live applications
- DASH 265 for live from DASH-IF Guidelines [59]

9. Distribution 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.

9.1. Digital Terrestrial Transmission

In February 2017, Korea launched terrestrial UHD TV commercial services using ATSC 3.0 with 4K spatial resolution and Next Gen Audio encoding. In the US, ATSC 3.0 roll-out has reached more than 66 markets and that number is expected to reach 75% of US television households in 2023[3] . For reasons described below, today's US broadcast services are almost exclusively HD, with some using HDR. While Japan launched UHD TV services in 2018 via satellite and cable, no defined date for terrestrial transmissions has been announced. In Europe, the Digital Video Broadcasting members are developing standards for UHD Digital Terrestrial Transmission (DTT) services using DVB-T2.

DTT 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 DTT. It is possible that in some countries, broadcasters will use satellite to deliver an Ultra HD experience until spectrum is allocated on terrestrial networks.

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.

[3] Reference: Press release from ATSC, January 03, 2023. https://www.atsc.org/news/atsc-marks-milestones-for-deployment-of-next-generation-broadcasting-3/

9.2. 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)

Collectively we refer to these devices as 'Edge Devices' or just 'Devices'. More information regarding these can be found in the Violet Book, Section 8 [V01].

The content will need to be transcoded into multiple formats and bit-rates to provide the best experience to the consumer. Content providers should consider using 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.

9.3. 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 3.0 [77], [136], [137], [138], [139] & DOCSIS 3.1 [97], [140], [141], [142] [143]) or an unmanaged network (i.e., the public Internet). Content-aware Encoding can be an enabling technology capable of added efficiency.

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).

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 8.5.1.).

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. Ultra HD 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 [2] system colorimetry and stereo audio). Content distribution network (CDN) partners may need to host both Ultra HD 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 8 describes various options for delivering Ultra HD 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/uhd-service-tracker/

Table 8. Ultra HD 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

9.4. Ultra HD Live Distribution over OTT

The live streaming Ultra HD workflow can be divided into several major components (see Figure 5 below):

- Live source or content acquisition/production
- Transmitting the live content to the encoding farm
- Encoding the content to the appropriate distribution profiles
- Packaging the content into its appropriate distribution formats
- End Device detection and determination of the appropriate formats
- Dynamic delivery to the end user devices, depending on available bandwidth
- End user device feedback (dropped packets, stall, overflow
- Optionally DRM may be required.

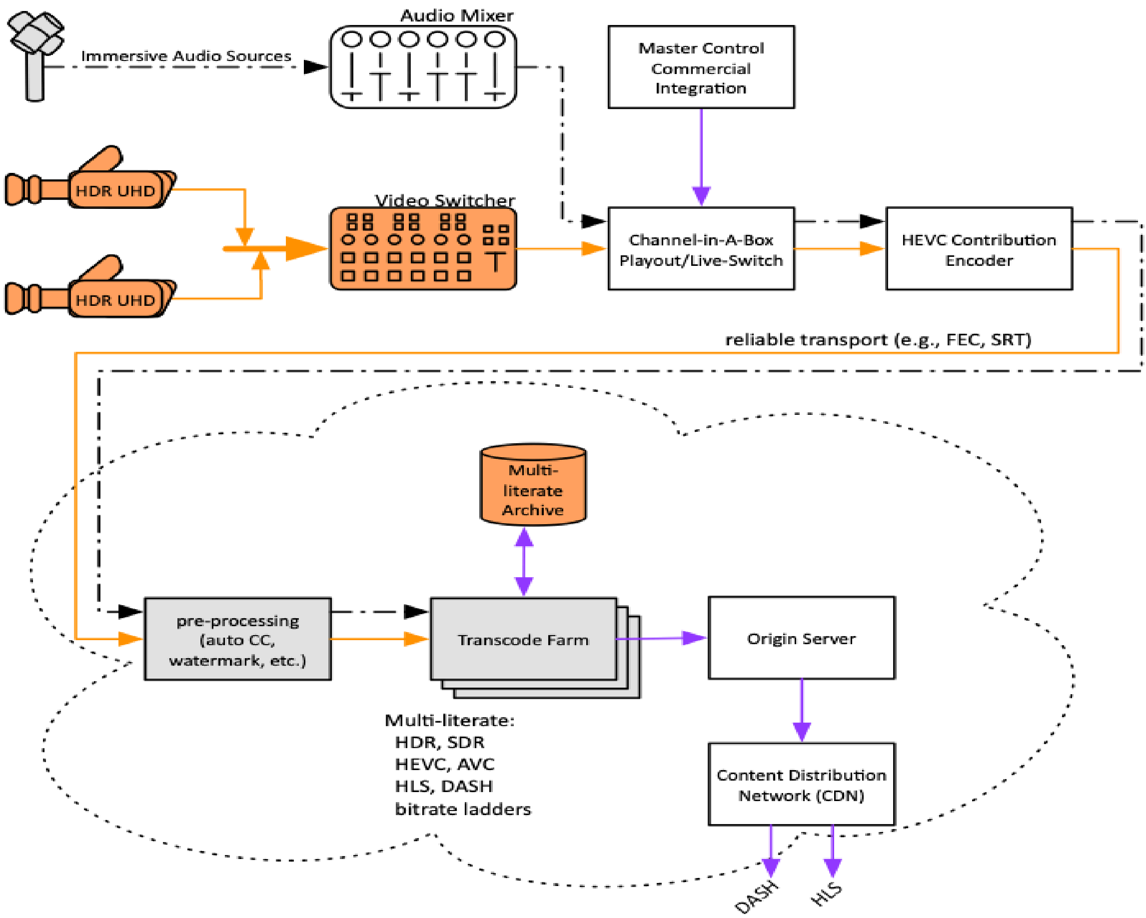

Figure 5. Live Event Streaming over OTT Workflow

9.5. Typical Live Streaming Workflow

The requirements for Ultra HD live production consist typically of cameras, graphics, video replay units, and vision mixer. The Ultra HD production may be at 4K resolution with or without HDR. In the case of the use of SDR cameras, it is recommended that the source be upconverted to HDR (see Section 8.2 of the Blue Book [B04]). Note: 1080p production along with HDR is considered an Ultra HD format.

The output of the live production is connected to a contribution encoder. The contribution encoder is configured to minimize bandwidth requirements while maintaining sufficient image quality using HEVC encoding in a MPEG-TS format. Typical contribution bit rates are from 20 to

30Mb/s for HEVC. For the purpose of security and guaranteed delivery, MPEG-TS is usually encapsulated in SRT[4] or other forward error correction protocols (FECs). Once the contribution stream is delivered to the cloud based transcode farm, assuming that there are sufficient resources, the content is either passed thru as well as encoded in lower bitrates with a mix of compression levels. Typical HEVC distribution bitrates are presented in Table 9. To offer the best Quality of Experience (QoE), it is desirable to provide finer granularity of Ultra HD profiles, with a corresponding increase in encoding resources. This could be a challenge for Live Streaming, especially for thin viewership events.

Assuming that the incoming contribution bitrate is sufficient, it is desirable to encode the Live stream into the bit rates listed in Table 9 (these may vary based on the internet contribution quality as well as access to encoding resources)[5]. In addition, it is also required to provide an SDR AVC version that will play on all devices: TV, but also Mobile and all web browsers. (See Table 10).

Table 9. 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

[4] See https://www.srtalliance.org

[5] See also https://www.bbc.co.uk/rd/blog/2018-05-uhd_hdr_world_cup_2018 and https://www.bbc.co.uk/rd/blog/2017-12-uhd-hdr-trial-blue-planet-bbc-iplayer

Table 10. Typical SDR Representations

Resolution	frame rate	Codec	HDR	AVC bitrate (max CAE rate)
3840x2160	p60/50	AVC	No	25-30 Mbps
3840x2160	p60/50	AVC	No	15-20 Mbps
1920x1080	p60/50	AVC	No	10-15 Mbps
1920x1080	p60/50	AVC	No	8-12 Mbps

9.6. Packagers

One of the most important components in the live streaming workflow is the packager. The packager should be capable of packaging the live Ultra HD stream into multiple formats. For compatibility with Apple devices, a fragmented MP4/HEVC format has to be used(see DASH-IF IoP [16]). To support Roku devices either a fragmented MP4/HEVC or an HLS/HEVC Apple HLS [67] format can be used. For smart TVs, HLS/HEVC is the most desirable format.

For the compatibility of the stream with non-HDR or older devices, the use of AVC-H264 based codecs using suitable HLS-bitrates is recommended for the output format of the encoders/packagers. When DRM is required, Android devices will require a DASH formatted output from the packager.

Common Media Application Format (CMAF) unifies the streaming payload format, so that one package can accommodate both DASH and HLS manifests. The CMAF format has been selected by CTA WAVE to address both HLS and DASH internet services (see CTA 5001 [108]), and a DASH/HLS interoperability guideline was published as CTA-5005 [115]. Apple guidelines [109] describe the use of CMAF with iOS devices. From 2021 some prominent services are using CMAF exclusively and it is expected that most major streaming services expand their use, as legacy devices fall out of service.

9.6.1. Devices

OTT services can be delivered either to an Edge Device provided by the operator or by a Edge Device or STB / TV provided by the user. To highlight the difference between the two scenarios:

1. An operator service delivers to a limited set of devices. The difficulties in delivery are minimized as the operator controls the delivery formats and device, usually via an STB, though the TV appliance is not under their control.
2. Consumer distribution, where the service operator distributes to consumer devices via a specific OTT application. The decoding device can either be a BYOD STB (i.e., Fire TV, Roku TV, Apple TV, Chromecast) or a connected SmartTV running the service operator's application in the TV. More information on Edge devices can be found in the Violet Book, Section 8.1 [V02].

In order to address the problem of OTT compatibility, the BBC developed a list of 700 tested devices during the FIFA'18 World Cup that was transmitted over OTT, that properly supported BBC iPlayer during this event.

For more in depth description of device capabilities, see Real World Ultra HD (Violet Book [V]).

9.6.2. TV Sets

TV Set behavior may vary based on the chosen input: USB, network (cable, DTH, terrestrial), IPTV and HTTP (streaming).

TV capabilities also differ by region, as the services in each market are different, so one needs to understand the target market.

Some TVs can be software (SW) upgraded. Note that TV manufacturers do not in general provide upgrades after 12 months, unless a major flaw has been found, although at the IBC 2019 Android TV announced three years of SW upgrade for its devices, including TVs.

The TV set capabilities must be separated from the app capabilities. For example, a viewer with a Dolby Vision enabled TV set, may be using an app that is limited to HDR10 only.

For more in depth description of TV set capabilities, see Real World Ultra HD (Violet Book [V]).

9.6.3. BYOD STB

For the best experience, the capabilities of the BYOD STB and the TV must be aligned in terms of HDR capability. Table 11 below shows what happens when HDR is sent to the STB.

Table 11. STB HDR/SDR Modes

Mode	STB out	TV out	Note
Aligned	HDRx	HDRx	Best experience
STB mismatch	SDR	SDR>HDRy	Quality depends on TV tone expander
TV mismatch	HDRx	HDRy	STB needs to convert HDR format. In some cases like Dolby Vision, this is well specified, in other cases STB has the liberty to do its own conversion.

As with a TV, on a BYOD STB, the STB and the app capabilities are separate, for example, with an STB HW capable of Dolby Vision running an app that might be limited to HDR10 only.

For a more in depth description of TV set capabilities, see Real World Ultra HD (Violet Book [V]).

9.7. Content Distribution Network (CDN)

For the live streaming of Ultra HD content, it is recommended to use a multi CDN solution as bit rates of 18 Mb/s or higher present a challenge to deliver to a mass audience of viewers. Best practice is to use diverse CDN services to facilitate delivery to a large, global audience.

9.8. User Interface (UI)

The UI of a device or application should present to the viewer which Ultra HD formats (with or without HDR) are available for a live streaming program. In some cases, due to incompatibilities between the device and display capabilities or the lack of support in the device, it is useful to provide a manual choice in the UI so that the viewer can override automatic selection of the image format.

10. Content Aware Encoding

CAE is an Ultra HD technology beyond Foundation Ultra HD that works with OTT devices equipped with HLS or DASH- compliant video players. The encoders implementing CAE should also follow guidelines defined for content preparation for HLS [67] or DASH [16] formats respectively.

Content Aware Encoding or Content-Adaptive Encoding (CAE) is a class of techniques for improving efficiency of encodings by exploiting properties of the content. By using such techniques, “simple” content, such as scenes with little motion, static images, etc. will be encoded using fewer bits than “complex” content, such as high-motion scenes, waterfalls, etc. By so doing, content-aware techniques aim to spend only a minimum number of bits necessary to ensure quality level needed for delivery. Since “simple” content is prevalent, the use of CAE techniques results in significant bandwidth savings and other benefits to operators (e.g., some systems may also reduce the number of encodings, deliver higher resolution in the same bits as the previous systems required for lower resolution, better overall quality, etc.). The CAE process is the “secret sauce” of an encoder company as described in several references[6].

Content Aware Encoding, also referred to as Content-Adaptive Encoding, or CAE, is a technique applied during the encoding process to improve the efficiency of encoding schemes. It can be used with any codec, but in the context of this document we will solely focus on HEVC.

CAE is not a standard, but a technique applied at the encoder side that is expected to be decoded by an HEVC Main 10 decoder.

Regarding adaptive streaming, the original specification is iOS 11, now Apple HLS Authoring Specification [67]. However MPEG DASH IF supports CAE and ABR.

As opposed to other techniques such as HDR, WCG, NGA or HFR, where new devices or network equipment are required, CAE just requires an upgrade of the encoder and should work with any decoder.

For OTT, ABR is already the most common way to deliver content. CAE is applied on top of ABR in the encoding process.

[6] http://info.harmonicinc.com/Tech-Guide-Harmonic-EyeQ
http://media2.beamrvideo.com/pdf/Beamr_Content_Adaptive_Tech_Guide.pdf
https://www.brightcove.com/en/blog/2017/05/context-aware-encoding-improves-video-quality-while-cutting-costs

Cable operators can broadcast Live over either QAM or ABR or over IP (DOCSIS® 3.0 [77]). The IP delivery may be performed in Unicast as the traffic is not expected to be high, and may later be scaled using ABR Multicast as described in a CableLabs Technical Report [58].

For a more complete description of CAE and its uses, please see Yellow Book-Section 7.4.2 [Y02].

11. Security - Content Protection

Digital content has always been exposed to illegal reproduction and illegal distribution. Various content protection technologies have been developed, involving different scrambling algorithms and protection mechanisms. These are covered under the terms Conditional Access and Digital Rights Management. Ultra HD based content generally falls under the description premium content, and requires the deployment of enhanced content protection mechanisms. Broadly the Ultra HD Forum recommends that the MovieLabs Specification for Enhanced Content Protection [39] be followed, with the following additional context and guidance.

11.1. Conditional Access

Over the years, various scrambling algorithms have been developed for protecting linear delivered content, generally delivered over broadcast networks. 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 toas 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[7] 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.

[7] https://link.springer.com/chapter/10.1007/978-3-642-34159-5_4

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

- AES, Advanced Encryption Standard, published in FIPS-197, NIST, 2001,
- DVB-CSA3 published in 2011 by DVB in ETSI TS 100 289 V 1.1.1,
- 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:

- Ultra HD content should not be scrambled with DVB-CSA1, nor with DES scrambling algorithms.
- Ultra HD content should be scrambled with AES or DVB-CSA3, using a minimum of 128 bits key size.
- DVB-compliant service providers should use DVB-CSA3 or DVB-CISSA when transmitting Live or linear Ultra HD content.
- 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.

11.2. Digital Rights Management

Digital Rights Management is the evolution of content protection beyond just controlling access to content based on a variety of business rules. Digital Rights Management enhances beyond scrambling of the content through to evolved management of the distribution of keys for the playback and usage of the content which can include onward distribution, and is eminently used for the delivery of content, both linear and on demand, in IP based networks.

DRM comes in several forms but there are two primary solutions that are suited for Ultra HD content when implemented in accordance with the Movielabs Specification for Enhanced Content Protection [39], PlayReady SL3000 and Widevine L1, although we do not restrict to these or recommend the use of any specific DRM as long as it is compliant with the Movielabs Specification.

12. Security - Forensic Watermarking

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 the 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.

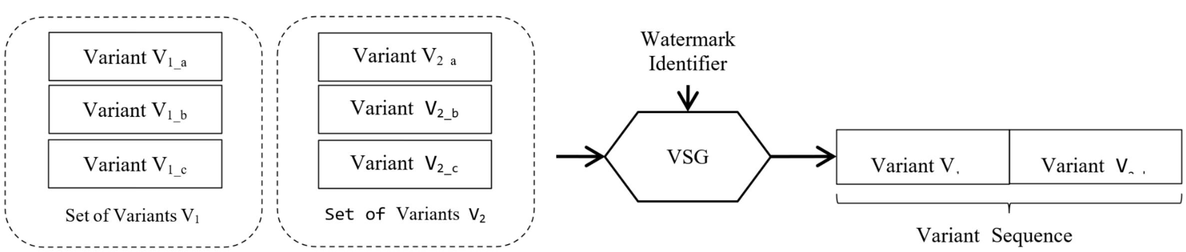

Figure 6. Illustration of Watermark Identifier

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.

12.1. 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 [39] that mandates the use of Forensic Watermarking on Ultra HD 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.”

12.1.1. 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.

12.1.2. 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.

12.1.3. Forensic Watermark Identifying Consumers Devices

The MovieLabs specification [39] 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.

12.2. 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 12.3 One Step and Two Step12.4 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.

12.3. 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 7.

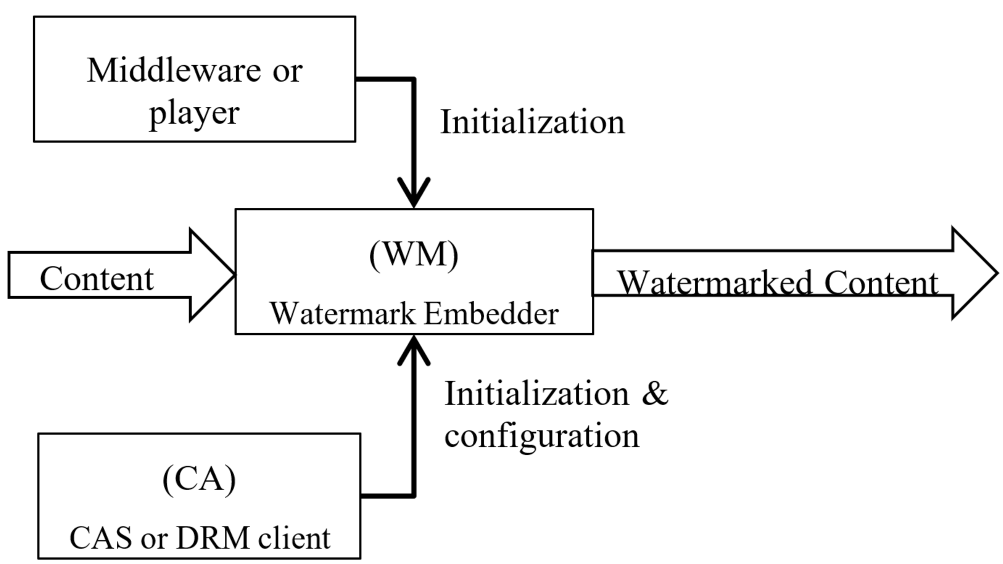

Figure 7. One-Step Watermark Performed on the Client Side

12.4. 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.

12.4.1. 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.

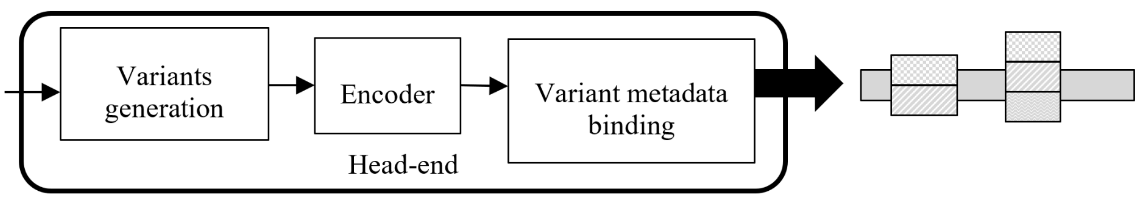

Figure 8. Example (a) Variant generation in the baseband domain

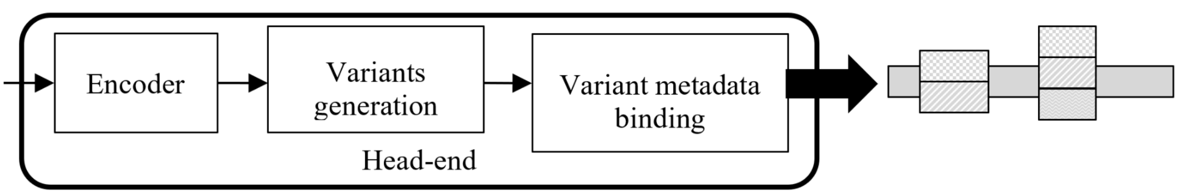

Figure 9. Example (b) Variant generation in the encoded domain

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.

12.4.2. 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.

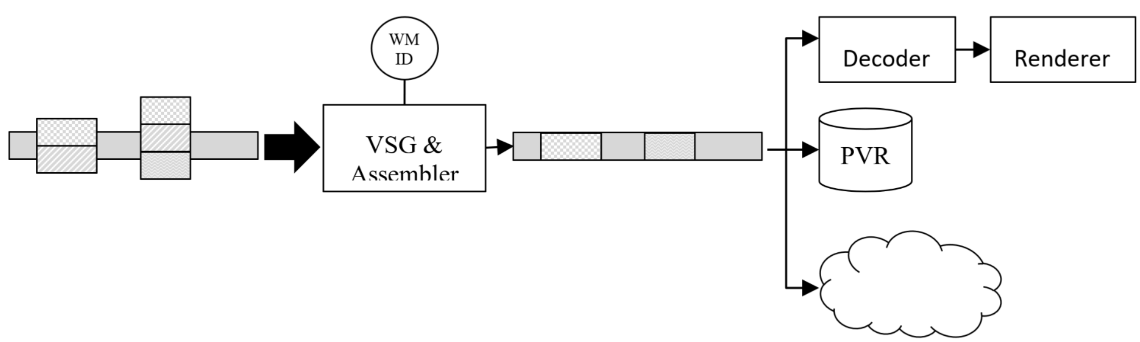

Figure 10. Watermark Embedding Using a Unique Variant Sequence

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 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 symbols. 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 12.5.

• **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.

12.4.3. 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:

12.4.3.1. Transport at the Media Layer

Variant 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 in MPEG 4 [25], or HEVC [26].

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.

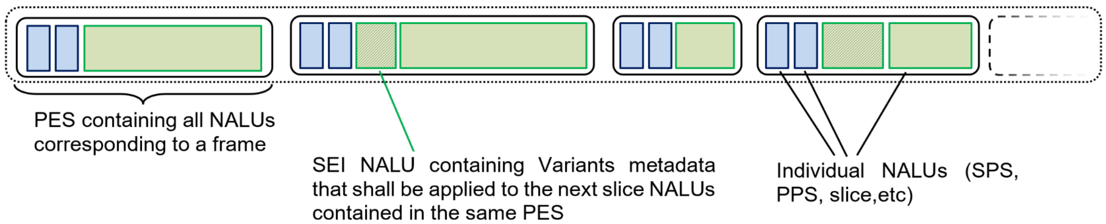

Figure 11. Transport at the Media Layer Using MPEG SEI MALUs

12.4.3.2. Transport at the Container Layer

Variant 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. Alternatively, Variants metadata could be incorporated as an extra track in an ISOBMFF 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.

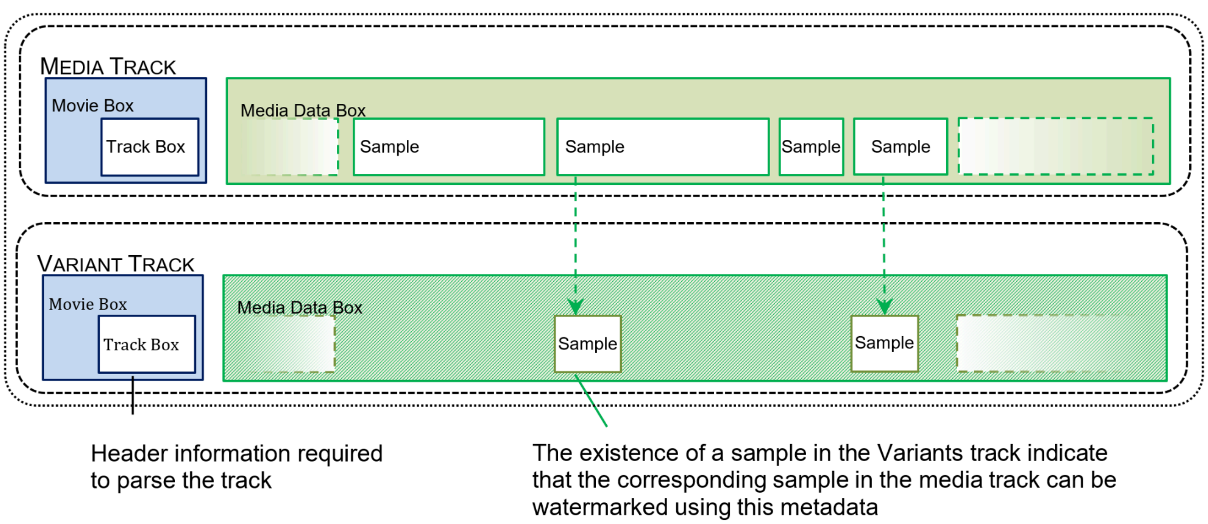

Figure 12. Transport at the Container Layer Using a Track in an ISOBMFF File

An example of how to transmit Variants metadata as an extra track in an ISOBMFF file is described in [40][8]. 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 set of keys and thereby be 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 Encoder Boundary Point in Open Cable Specification for EBP [49].

12.4.3.3. Out-of-Band Transport

Transporting Variants metadata at the media as well as container layer requires some level of video parsing. However, in some integration 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 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

[8] It shall be noted that the terminology “variants” is slightly different in the MPEG standard and these Ultra HD 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.

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 a scenario, broadcasters 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.

12.5. Use Case: ABR VOD

The characteristics of this use case requires that segment boundaries are known before the start of the watermark processing.

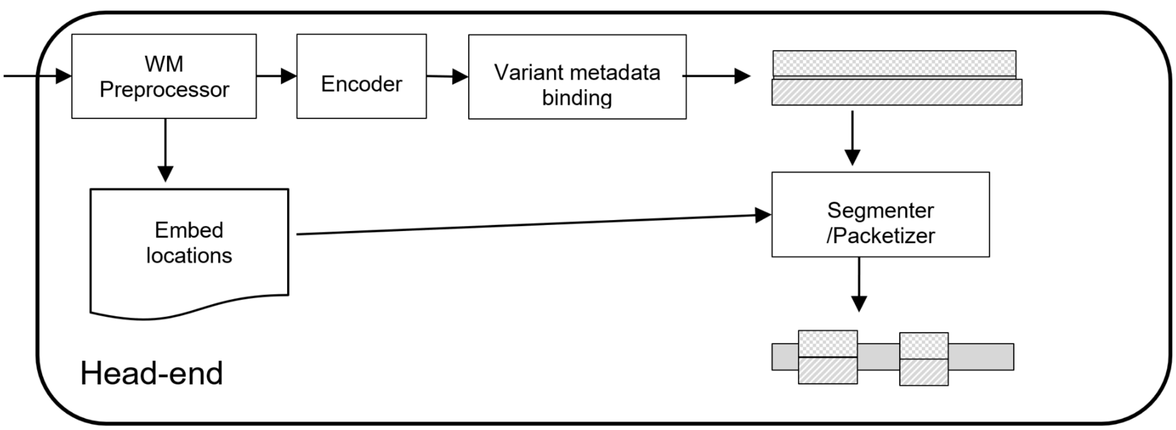

Figure 13. Pre-Processing in the Baseband Domain for Two-step Watermarking

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, the 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.

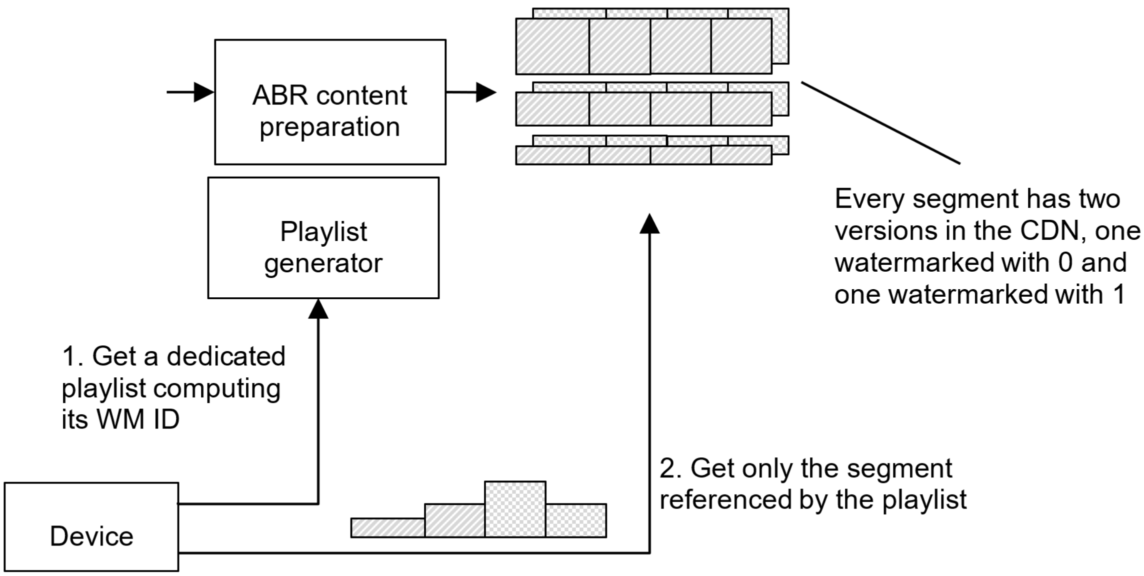

Figure 14. ABR Playlist Serialization

13. References

[2] Recommendation ITU-R BT.709-6:2015, “Parameter values for the HDTV standards for production and international programme exchange”, July 2015, https://www.itu.int/rec/R-REC-BT.709-6-201506-I/en

[3] Recommendation ITU-R BT.2020:2015, “Parameter values for ultra-high definition television systems for production and international programme exchange”, Oct 2015, https://www.itu.int/rec/R-REC-BT.2020-2-201510-I/en

[10] SMPTE ST 2086:2018, “Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images”, April 2018, https://my.smpte.org/s/product-details?id=a1BVR0000008kaI2AA

[13] DVB DASH specification, ETSI TS 103 285 v1.4.1:2023, “MPEG-DASH Profile for Transport of ISO BMFF Based DVB Services over IP Based Networks”,Sept 2023, https://www.etsi.org/deliver/etsi_ts/103200_103299/103285/01.04.01_60/ts_103285v010401p.pdf

[16] DASH-IF Interoperability Points: Guidelines for Implementation, version 4.3, November 2018, https://dashif.org/docs/DASH-IF-IOP-v4.3.pdf

[17] DCI Specification, Version 1.4.5 with Addenda as of 29 May 2024 Incorporated, https://www.dcimovies.com/specification/

[18] CTA-608-E S-2019, “Line 21 Data Services”, Dec 2019, https://shop.cta.tech/products/line-21-data-services#:~:text=ANSI%2FCTA%2D608%2DE%20S%2D2019%20is%20a,of%20the%20NTSC%20video%20signal.

[19] ANSI/CTA-708-E S-2023, “Digital Television (DTV) Closed Captioning”, Aug 2013, https://shop.cta.tech/products/digital-television-dtv-closed-captioning

[20] ETSI 300 743, “Digital Video Broadcasting (DVB); Subtitling systems”, v1.6.1, July 2018, https://www.etsi.org/deliver/etsi_en/300700_300799/300743/01.06.01_20/en_300743v010601a.pdf

[21] ETSI 300 472, "Digital Video Broadcasting (DVB); Specification for conveying ITU-R System B Teletext in DVB bitstreams", v1.4.1,Jan 2017, https://www.etsi.org/deliver/etsi_en/300400_300499/300472/01.04.01_20/en_300472v010401a.pdf

[22] ANSI/SCTE-27 2016 (R2021), "Subtitling Methods for Broadcast Cable", 2021, https://wagtail-prod-storage.s3.amazonaws.com/documents/ANSI_SCTE_27_2016_R2021.pdf

[23] W3C Recommendation: "TTML Text and Image Profiles for Internet Media Subtitles and Captions (IMSC1)", v1.1, Nov 2018 updated April 2020, https://www.w3.org/TR/ttml-imsc1.1/

[25] ISO/IEC: Doc. ISO/IEC 14496-10:2022, "MPEG-4 -- Part 10: Advanced Video Coding", Nov 2022, https://www.iso.org/standard/83529.html

[26] ISO/IEC: Doc. ISO/IEC 23008-2:2020 "Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 2: High efficiency video coding", Aug 2020, https://www.iso.org/standard/75484.html

[35] ETSI TS 103 420 v1.2.1 (2018-10), "Object-based audio coding for Enhanced AC-3 (E-AC-3)". https://www.etsi.org/deliver/etsi_ts/103400_103499/103420/01.02.01_60/ts_103420v010201p.pdf

[39] MovieLabs, "MovieLabs Specification for Enhanced Content Protection, v1.3", Dec 2020. https://movielabs.com/ngvideo/MovieLabs_ECP_Spec_v1.3.pdf

[40] ISO/IEC 23001-12:2018, "Information technology -- MPEG systems technologies - Part 12: Sample Variants in the ISO base media file format" https://www.iso.org/standard/74431.html

[41] ETSI TS 102 034 v2.1.1 (2016-04), "Transport of MPEG-2 TS Based DVB Services over IP Based Networks" https://www.etsi.org/deliver/etsi_ts/102000_102099/102034/02.01.01_60/ts_102034v020101p.pdf

[47] SMPTE ST 2110-40:2023, “Professional Media over IP Networks: SMPTE ST 291-1 Ancillary Data” https://pub.smpte.org/doc/st2110-40/20231231-pub/

[49] OpenCable Specifications, OC-SP-EP-I01-130118: “Encoder Boundary Point Specification” (01/2013). https://www.cablelabs.com/specifications/encoder-boundary-point-specification

[53] ATSC: A/331:2024-04, “, “Signaling, Delivery, Synchronization, and Error Protection”, April 3, 2024, https://www.atsc.org/wp-content/uploads/2024/04/A331-2024-04-Signaling-Delivery-Sync-FEC.pdf

[58] CableLabs OC-TR-IP-MULTI-ARCH-C01-161026:2016, “IP Multicast Adaptive Bit Rate Architecture Technical Report”, November 26, 2016 ,https://www.cablelabs.com/specifications/ip-multicast-adaptive-bit-rate-architecture-technical-report

[59] DASH-IF: “Guidelines for Implementation: DASH-IF Interoperability Points for ATSC 3.0, Version 1.1,” DASH Interoperability Forum”, June 2018, https://dashif.org/docs/DASH-IF-IOP-for-ATSC3-0-v1.1.pdf

[67] APPLE “ “HLS Authoring Specification for Apple Devices”, May 10, 2024, https://datatracker.ietf.org/doc/html/draft-pantos-hls-rfc8216bis

[77] SCTE: 1135-1:2024, “Data-Over-Cable Service Interface Specification version 3.0”, 2024, (also known as “DOCSIS 3.0 Part 1: Physical Layer Specification”) https://account.scte.org/standards/library/catalog/scte-135-1-docsis-3-part1-physical-layer-specification/

[79] SMPTE ST 424:2012, “3 Gb/s Signal/Data Serial Interface”, October 8, 2012,https://pub.smpte.org/doc/st424/20121008-pub/

[80] SMPTE ST 425-1:2017, “Source Image Format and Ancillary Data Mapping for the 3 Gb/s Serial Interface”, November 1, 2017, https://pub.smpte.org/doc/st425-1/20171023-pub/

ST 425-3:2019, “Image Format and Ancillary Data Mapping for the Dual Link 3Gb/s Serial Interface”, May 2019,https://pub.smpte.org/doc/st425-3/20190304-pub/ ; and

ST 425-5:2019, “Image Format and Ancillary Data Mapping for the Quad Link 3Gb/s Serial Interface”, May 2019,https://pub.smpte.org/doc/st425-5/20190304-pub/

[82] SMPTE ST 2022-6:2012, “Transport of High Bit Rate Media Signals over IP Networks (HBRMT)”, October 9, 2012, https://pub.smpte.org/doc/st2022-6/20121009-pub/

[84] SMPTE ST 2081-10:2018, “2160-line and 1080-line Source Image and Ancillary Data Mapping for 6G-SDI”, March 12, 2018, https://pub.smpte.org/doc/st2081-10/20180102-pub/

[96] Deprecated. See [77] ANSI/SCTE 135-1 2019

[97] ANSI/SCTE 220-1 2022, “DOCSIS 3.1 Part 1: Physical Layer Specification”, 2022, https://account.scte.org/standards/library/catalog/scte-220-1-docsis3-1-part1-physical-layer-specification/

[106] EBU R 103, “Video signal tolerance in Digital Television Systems”, Version 3.0, June 2020, https://tech.ebu.ch/docs/r/r103.pdf

[108] CTA 5001-A, “Web Application Video Ecosystem-Content Specification”, December 2018, https://cdn.cta.tech/cta/media/media/resources/standards/pdfs/cta-5001-a-final.pdf

[109] Apple, “About the Common Media Application Format with HTTP Live Streaming”, 2020, https://developer.apple.com/documentation/http_live_streaming/about_the_common_media_application_format_with_http_live_streaming

[115] CTA-5005, “Web Application Video –Ecosystem—DASH-HLS Interoperability Specification”, May 2021. https://cdn.cta.tech/cta/media/media/resources/standards/pdfs/cta-5005-final.pdf

[131] SMPTE ST 274:2008, SMPTE Standard-1920x1080 Image Sample Structure, Digital Representation and Digital Timing Reference Sequences for Multiple Picture Rates, Jan 2008, https://pub.smpte.org/doc/st274/20080129-pub/

[136] ANSI/SCTE 135-2 2019, DOCSIS 3.0 Part 2: MAC and Upper Layer Protocols, 2019, https://www2.scte.org/SCTEDocs/Standards/ANSI_SCTE 135-02 2019.pdf

[137] ANSI/SCTE 135-3 2019, DOCSIS 3.0 Part 3: Security Services, 2019, https://www.scte.org/standards/library/catalog/scte-135-3-docsis-3-part3-security-services/

[138] ANSI/SCTE 135-4 2019, DOCSIS 3.0 Part 4: Operations Support Systems Interface, 2019, https://www2.scte.org/SCTEDocs/Standards/ANSI_SCTE 135-4 2019.pdf

[139] ANSI/SCTE 1135-5 2022, DOCSIS 3.0 Part 5: Cable Modem to Customer Premise Equipment Interface, 2022, https://account.scte.org/standards/library/catalog/scte-135-5-docsis-3-part-5-cable-modem-to-customer-premise-equipment-interface/

[140] ANSI/SCTE 220-2 2022 DOCSIS 3.1, Part 2: MAC and Upper Layer Protocols Interface, 2022, https://account.scte.org/standards/library/catalog/scte-220-2-docsis3-1-part2-mac-and-upper-layer-protocols-interface/

[141] ANSI/SCTE 220-3 2022 DOCSIS 3.1, PART 3: Cable Modem Operations Support System Interface-Specification, 2022, https://account.scte.org/standards/library/catalog/scte-220-3-docsis3-1-part3-cable-modem-operations-support-system-interface-specification/

[142] ANSI/SCTE 220-4 2022 DOCSIS 3.1, Part 4: CCAP Operations Support System Interface Specification, 2022, https://account.scte.org/standards/library/catalog/scte-220-4-docsis3-1-part4-ccap-operations-support-system-interface-specification/

[143] ANSI/SCTE 220-5 2016 DOCSIS 3.1, Part 5: Security Specification, 2022, https://account.scte.org/standards/library/catalog/scte-220-5-docsis3-1-part-5-security-specification/

[R] **Red Book** – Introduction to Ultra HD

[R01] Section 10.2, Next Generation Audio (NGA)

[O] **Orange Book** --Foundational Technologies for Ultra HD

[O01] Section 7.2.8, HDR10 Metadata Carriage

[Y] **Yellow Book** – Beyond Foundational Technologies for Ultra HD

[Y01] Section 7.4.1, AVS2 & AVS3

[Y02] Section 7.4.2, Content Aware Encoding (CAE)

[B] **Blue Book** – Ultra HD Production and Post Production

[B01] Section 7.5.2, Studio Video over IP

[B02] Section 8, Conversions Between Transfer Functions

[B03] Section 10, Production for Pre-Recorded Content

[B04] Section 8.2, Conversion of SDR to HDR: Up-Mapping

[I] **Indigo Book --** Ultra HD Technology Implementations

[I01] Section 8, Monographs on HDR

[V] **Violet Book –** Real World Ultra HD

[V01] Section 8, Consideration of Edge Devices

[V02] Section 8.1.1, Monitor

(End of Green Book)