Ultra HD Forum Guidelines

Red Book – Introduction to Ultra HD

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, Red Book - *Introduction to Ultra HD*, 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 September 2024 version of the UHDF Guidelines Red Book, *Introduction to Ultra HD* (v3.3), edited by Jim DeFilippis.

The *Introduction to Ultra HD* is the first in the series of Rainbow Books on the Guidelines for Ultra HD. The scope and purpose of this book is to provide a basic overview of Ultra HD and the concepts such as HDR (PQ and HLG), Wide Color (Rec 2020), High Frame Rate (HFR) and Next Generation Audio (NGA). The Foundation technologies of Ultra HD are described briefly. Throughout this book are links to both external references as well as references between the Guideline books.

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, Sept 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 Rainbow 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

7. Overview of the Guidelines

The first Guidelines document was published in 2016, at the dawn of the UHD era. Many changes in both technology and techniques have evolved that the Guidelines have captured and documented since the first publication. Historically the Guidelines were updated twice a year, at the NAB Convention (spring) and the IBC Conference (fall). Originally there were two documents, Part A, which described the foundational elements of UltraHD and Part B, which described new additional technologies that could be used to augment the foundational elements. Eventually these two parts were merged into a combined document we refer to as V2. Note that the concept of foundational technologies and beyond foundational have been incorporated into the Guidelines. Some beyond foundational technologies have become more or less mainstream; however we have kept to this dichotomy so as to provide a pathway to UHD with a minimum of features as well as the enhancements that are available. Any technology which has not reached a level of adoption by industry but shows promise of adoption will be listed in an annex to the Guidelines.

In 2022 we published the last V2 (v2.6) and embarked on a process to produce what we now refer to as the 'Rainbow Books'. We released the first version of these books at the 2023 NAB Convention. The books are color coded by general topic (see list in the Forward) to provide a convenient way to read and absorb the over 300 pages of technical information regarding Ultra HD production, post production, distribution and devices.

The Red Book, which you are now reading, provides an overview of Ultra HD technologies. It should be read prior to reading the other books to provide a baseline understanding of UHD. To aid with the organization of the books, there is the White Book of references as well as the table of contents of each book. There are links in each book which tie back to the references as well as other books that may have additional information on a particular topic. At the end of each book is a list of references used in that Book, both external references and internal references (those that point to another Rainbow Guidelines book).

Currently the Guidelines books are available in Acrobat PDF format, from our website (www.ultrahdforum.org/guidelines/). In the near future, they will also be published as interactive HTML documents on the UltraHD Forum's website.

We hope these books are informative and useful in your efforts to understand and deploy Ultra HD.

8. Foundation Ultra HD Technologies

These guidelines describe a number of Ultra HD-related technologies. The Ultra HD Forum notes that some of these technologies can be considered "Foundation" Ultra HD technologies, which are core aspects of UltraHD content and services, such as High Dynamic Range (HDR) and Wide Color Gamut (WCG). The Ultra HD Forum also describes additional technologies, which service providers may find to be valuable enhancements or enablers to a core Ultra HD service, such as dynamic HDR metadata.

Foundation Ultra HD content and services are often those that were commercially available as early as 2016, and thus have reached a level of market adoption and maturity. For the purposes of this document, Foundation Ultra HD comprises the following characteristics:

- Resolution – greater than or equal to 1080p and lower than or equal to 2160p, (progressive format; BT.2100 [5]) (BT. 2100 does not include interlaced formats)
- Wide Color Gamut – color gamut wider than BT.709 [2]
- High Dynamic Range – Image dynamic range that provides a dynamic range larger than SDR, capturing and displaying increased highlight and shadow details. Use of tone curves as referenced in BT. 2100 (HLG and PQ).
- 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 [17])
- Audio – 5.1 or higher channel surround sound or channel-based Immersive Audio
- Closed Captions/Subtitles – CTA 708 [19]/608 [18], ETSI 300 743 [20], ETSI 300 472 [21], SCTE-27 [22], IMSC1 [23]

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

- HLG: Hybrid Log-Gamma (HLG) OETF, EOTF, and OOTF transfer functions specified in BT.2100 [5]
- HLG10: HDR systems or content employing Hybrid Log-Gamma (HLG), the wide color gamut specified in BT.2100 and 10-bit depth
- PQ: HDR systems or content employing Perceptual Quantization (PQ), the wide color gamut specified in BT.2100, and 10-bit depth
- PQ10: HDR systems or content employing Perceptual Quantization (PQ), the wide color gamut specified in BT.2100, and 10-bit depth
- HDR10: HDR systems or content employing PQ10 and further including or capable of providing SMPTE ST 2086 [10], MaxFALL, and MaxCLL metadata

Table 1. Foundation Ultra HD 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 2. Foundation Ultra HD Content Parameters

Spatial Resolution	1080p[1] or 2160p
System Colorimetry	BT.709 [2], BT.2020 [3]
Bit Depth	10-bit
Dynamic Range	SDR, PQ, HLG

Frame Rate[2]	24(23.976), 25, 30(29.97), 50, 60(59.94)
Video Codec	HEVC, Main 10 Profile, Level 5 or 5.1 (single layer)[3]
Audio Channels	5.1 or higher channel surround sound or channel-based Immersive Audio
Audio Codec	AC-3, E-AC-3, DTS-HD, E-AC-3 + JOC, HE-ACC, AAC-LC
Captions/Subtitles Coding	CTA 608/708, ETSI 300 743, ETSI 300 472, SCTE-27, IMSC1

Table 2 Notes:

1. 1080p together with WCG and HDR fulfills certain use cases for Foundation Ultra HD 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.

2. Fractional frame rates for 24, 30 and 60 fps are included for compatibility with current plant video clock reference, but ultimately not preferred. Fractional frame rates will be necessary during migration from legacy video systems. The lower frame rates may be common for cinematic content.

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

For the purpose of this document, including the above constraint on 1080p content, various combinations of these Foundation Ultra HD parameters can be combined to produce Ultra HD content. Additional, non-Foundation technologies may also be employed.

The Foundation Ultra HD codec is HEVC[1] 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.2020/2100 system colorimetry and support of PQ/HLG High Dynamic Range formats. However, the Ultra HD Forum finds 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.

In the Violet Book, Table 1 [V01] and Table 2 [V02] categorizes decoders and services in terms of Foundation Ultra HD capability. Additionally, Table 2 in the Violet Book offers some indications of which types of decoders are compatible with which service formats. Foundation decoder and service formats provide the base encoding formats for a number of enhancement features which may be implemented without causing service incompatibility with Foundation supporting devices. This will normally be described in the sections covering the enhancement features.

---

[1] For use in China, the AVS2 codec, Main10 profile, may be used instead of HEVC. [Y01]

9. SDR vs. HDR

HDR is sometimes touted as providing a significant improvement in the "immersivity" of an image, as compared to SDR, thus providing an improved sense of "being there". In some degree, this stems from a coincidental match between a dynamic range that the human visual system (HVS) can instantly perceive and the dynamic range presentable with SDR displays, vs a greater dynamic range presented by HDR displays that requires adaptation of the HVS.

In Figure 1 the total dynamic range of the human visual system is represented along the horizontal axis[2]. Photographs of scenes representative of and exposed for portions of that range are presented[3]. The left portion, from 10 cd/m2 to down around 10^{-6} cd/m2 represents levels of scene luminance where scotopic (rod) vision is active. There is only one kind of rod in humans, characterized by a specific photopigment, which leads to our perception that "night vision" is monochromatic. Accordingly, scotopic vision is somewhat less interesting for entertainment purposes, for being limited to black and white. The right portion of the axis is the range over which photopic (cone) vision is active, from 0.1 to 108 cd/m2. Human cones are present in the typical individual in three distinct kinds, corresponding to three photopigments having peak sensitivities to either long, medium, or short wavelengths of visible light.

The overlap between scotopic and photopic vision amounts to about 14% (2 decades of 14) of the total human visual range and is called mesopic vision, where both rods and cones operate simultaneously, starting from a scene luminance where cones are just barely sensitive enough yet rods are still able to register an image. At the upper end of this overlap range, the scene light is too strong for scotopic vision and saturates the rods, consuming all of the photopigment available in them, a "bleaching" effect leaving them completely desensitized. The cones, being less sensitive than the rods, allow color perception from dawn to dusk. Still, at a scene brightness around 108 cd/m2, as might occur for one staring into thin clouds backed by the sun, the cones too, become saturated and desensitized by bleaching. Recovery from bleaching is not quick. Cones can take two to five minutes to recover, while rods can take 15 to 30 minutes to "dark adapt".

---

[2] Timo Kunkel, Erik Reinhard, A Reassessment of the Simultaneous Dynamic Range of the Human Visual System, Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization, 2010 http://doi.acm.org/10.1145/1836248.1836251

[3] Photo credits to (top to bottom): Artur Tomaszewski, Caleb Mullins, Sterling Davis, Logan Weaver, Mike Erskine, Van Mendoza, and Thomas Ciszewski, all from Unsplash.com
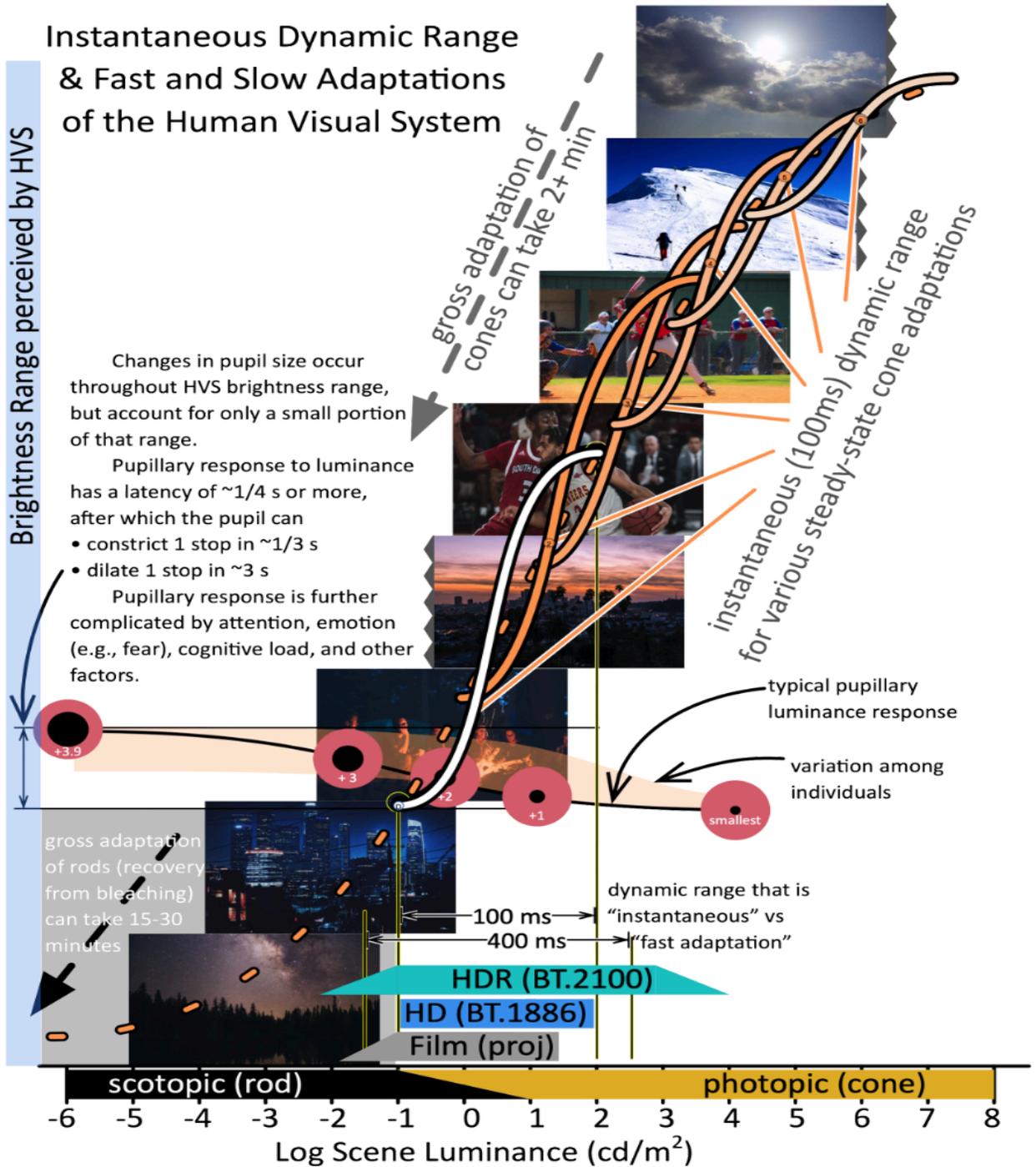
Figure 1. Comparing Dynamic Range of SDR, HDR, and the Human Visual System

A more subtle adaptation takes place, helping to normalize the significant dynamic range of human vision for higher level visual functions, such as contrast detection, texture recognition, object identification, etc. At any given moment, the human visual system is adapted to a point along the entire visual range, and while adapted to that point, is able to "instantly" (within 100 ms) see details in a small subset of the range. (See the solid white and orange sigmoid curves in Figure 1).

The instantaneous adaptation curve in white represents the simultaneously detectable range of 0.1-100 cd/m2. This corresponds to the SDR range recommended for HD television, as indicated in BT.1886 and other standards.

In parallel to adaptation that occurs at a particular light level, the pupillary luminance response is able to adjust quickly at nearly all levels of adaptation. By constricting and dilating, the pupil admits more or less scene light into the eye, with a range of almost four stops (i.e., 24:1), though this luminance response is limited once adapted to very dark or very bright scenes[4]. Overall, the pupillary adjustment represents only a small portion of the overall dynamic range of the human visual system, but is able to quickly extend the what can be seen given the current adaptation.

Pupillary response to luminance has a latency of about 1/4 second, or more. After that, the pupil can contract 1 stop (cutting light entering the eye by one-half) in about 1/3 s. Dilation is slower, taking about 3 s to open 1 stop.

Pupillary response is further driven by attention. When looking attentively into a shadow, the pupil quickly expands to see more detail, even though the overall amount of light in the scene hasn't changed. When looking at highlights, the pupil quickly contracts to better see texture. The overall adaptation needn't change, which would take many seconds, but is quickly extended at sub-second speeds, by the pupil's attention-based operation. This expanded range of dynamic adaptation is exploitable in HDR video, specified in BT.2100 to extend the luminance range both higher and lower. When properly presented, a greater sense of immersion is induced when HDR video presents details in different parts of an image that are not instantaneously visible, placing those details 400 ms or more away due to the need for dynamic attentive pupillary response and triggering involuntary responses more similar to those presented by the real world.

---

[4] C.J.K. Ellis, The pupillary light reflex in normal subjects, British Journal of Ophthalmology, 1981, 65, 754-759

10. Additional Ultra HD Technologies

In addition to the Foundation Ultra HD technologies, the Ultra HD Forum provides guidance on a number of additional technologies that can enable or enhance Foundation Ultra HD content and services.

The Ultra HD Forum notes that these additional Ultra HD technologies can be "layered" onto Foundation Ultra HD technologies in order to upgrade the consumer experience. For example, 7.1+4 immersive audio can be transmitted instead of 5.1 surround sound and/or dynamic HDR metadata can be included with HDR10 content and/or dynamic HDR metadata can be included with HDR10-based. Content-aware encoding can be employed to increase encoding efficiency. Content producers and service providers may elect to implement one or more additional Ultra HD technologies according to market drivers, the capabilities of the end-to-end ecosystem and consumer preferences. See Beyond Foundational Technologies [Y].

10.1. Wide Color

As part of the Ultra HD basket of new features is the adoption of ITU Rec BT.2020 [3] system colorimetry. This color system extends the gamut range of video imagery beyond HDTV color, as defined in ITU Rec BT.709 [2], and covers most of the spectrum of reflected colors, also known as Pointer's Colors. The primaries are located close to the limits of the CIE locus of monochromatic colors and thus require narrow band emitters to reproduce the full range of Rec 2020 color (see Figure 2).

Table 3. System Colorimetry for ITU-R Rec BT.2020

| System Colorimetry Parameter	Values		
Opto-electronic transfer characteristics before pre-correction	Assumed Linear[1]		
Primary colours and reference white[2]	Chromaticity Coordinates (CIE, 1931)	x	y
	Red Primary [R]	0.708	0.292
	Green Primary [G]	0.170	0.797
	Blue Primary [B]	0.131	0.046
	Reference White [D65]	0.3127	0.3290

Table 3 Notes:

1. PIcture information can be linearly indicated by the tristimulus values of RGB in the range of 0-1.
2. The colorimetric values of the picture information can be determined based on the reference RGB primaries and the reference white.
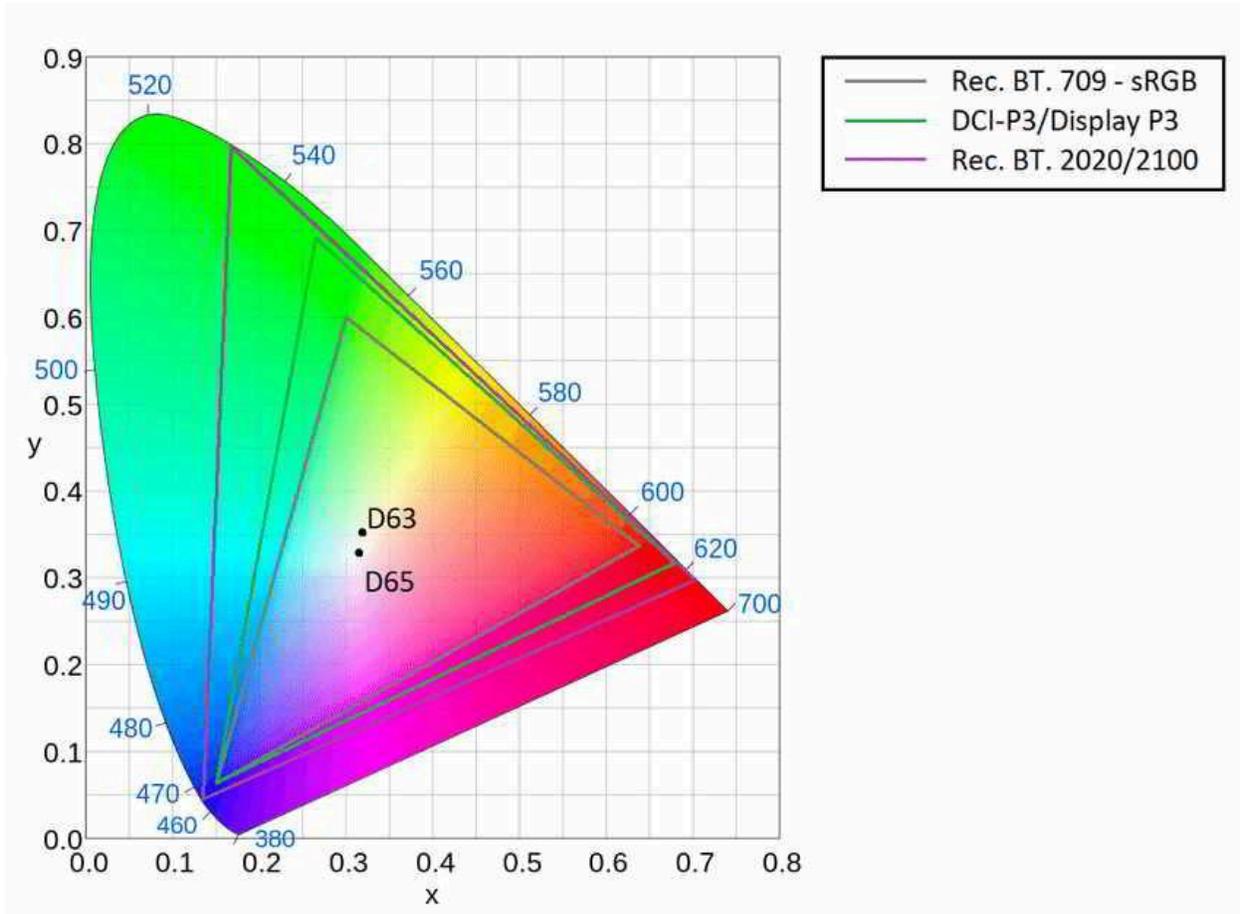
Figure 2. CIE Diagram showing Color System Gamuts, Rec2020, Rec709, P3

In comparison to HDTV color gamut as well as Digital Cinema (P3) color gamut Rec 2020 gamut increases the number of colors possible to represent in Ultra HD. Further with the addition of HDR, there is an increase in the volume of colors that can be represented (see Figure 3 below).
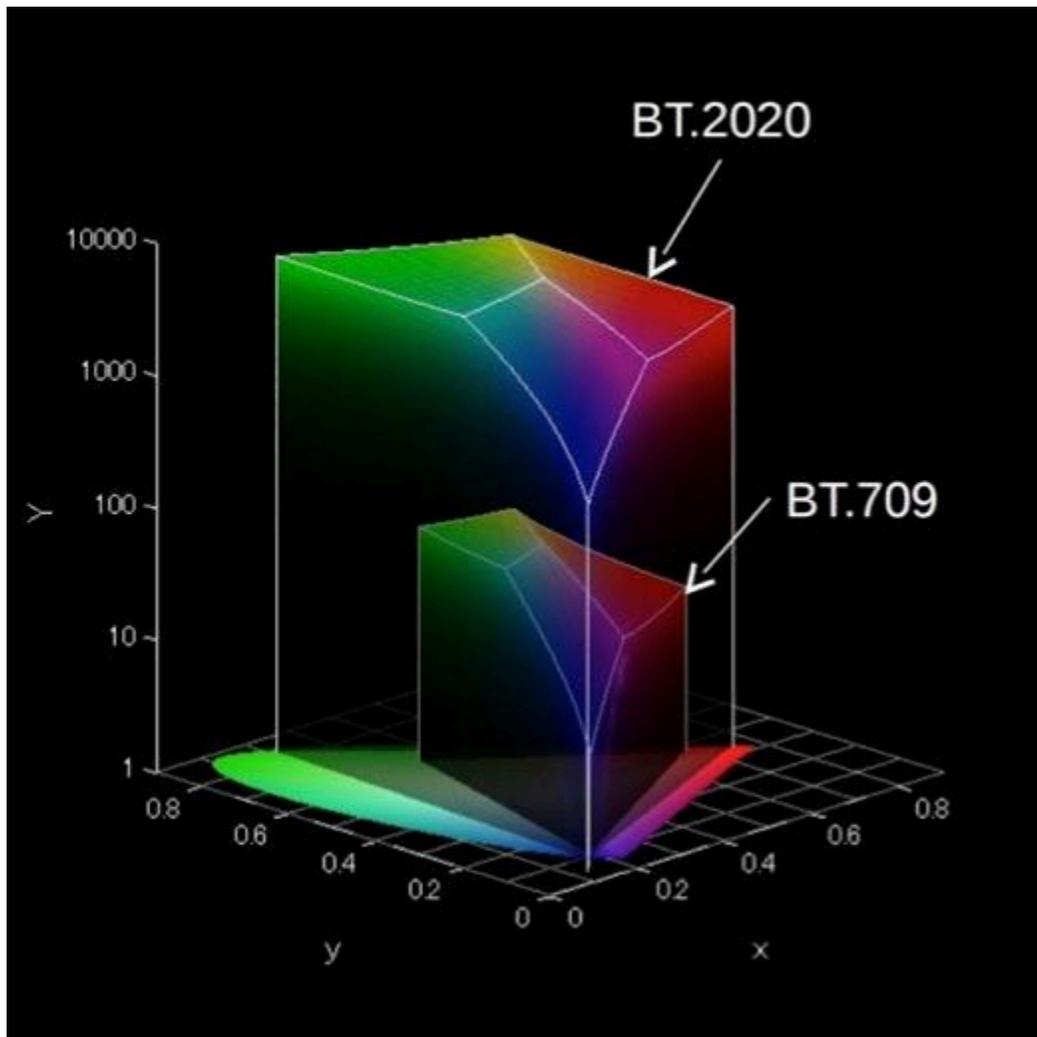
Figure 3. Color Volumes, Rec 2020, and Rec 709

Note that Rec.2020 uses a SDR luminance curve with a gamma of 0.45. ITU-R Rec BT.2100 [5] defines the use of both a PQ and HLG luminance curves with BT.2020 color gamuts.

The encoding of BT.2020 uses either constant luminance or non-constant luminance formats. The non-constant luminance is similar to the format used in Rec 709, with $Y'C'_BC'_R$ signals conveying the color and luminance information. The constant luminance format in Rec 2020 is not in wide use today. However another encoding format, that is a constant luminance approach as well, is IC_TC_p which is defined in ITU-R Rec BT.2100. IC_TC_p is derived from the RGB color

values, converted to LMS space and then have either the EOTF^{-1} of PQ or the OETF of HLG applied. Final conversion from L'M'S' to IC_TC_p values via a 3x3 matrix transform.

Table 4. Color Encoding for ITU-R Rec BT.2020

	Constant Luminance	Non-Constant Luminance
Derivation of Y'$_C$ and Y'	Y'$_C$=(0.2627R + 0.6780G + 0.0593B)'	Y'=0.2627R' + 0.6780G' + 0.0593B'
Derivation of Color Difference Signals	C'_{BC}= $\dfrac{B'-Y'_C}{-2N_B}$, $N_B \leq B'-Y'_C \leq 0$ $\dfrac{B'-Y'_C}{2P_B}$, $0 < B'-Y'_C \leq P_B$ C'_{RC} = $\dfrac{R'-Y'_C}{-2N_R}$, $N_R \leq R'-Y'_C \leq 0$ $\dfrac{R'-Y'_C}{2P_R}$, $0 < R'-Y'_C \leq P_R$ Where $P_B = \alpha(1-0.0593^{0.45}) = 0.7909854\ldots$ $N_B = \alpha(1-0.9407^{0.45})-1 = -0.9701716\ldots$ $P_R = \alpha(1-0.2627^{0.45}) = 0.4969147\ldots$ $N_R = \alpha(1-0.7373^{0.45})-1 = -0.8591209\ldots$ **For practical purpose, the following values can be used:** P_B= 0.7910, N_B = -0.9702 P_R = 0.4969, N_R = -0.8591	$C'_B = \dfrac{B'-Y'}{1.8814}$ $C'_R = \dfrac{R'-Y'}{1.4746}$

<div align="center">**Table 5. Color Encoding for ICTCP**</div>

Parameter	Values PQ	Values HLG
L,M,S Color Space	**L = (1688R + 2146G + 262B)/4096** **M = (683R+2951G + 462B)/4096** **S = (99R + 309G + 3688B)/4096**	
Derivation of L', M', S' **(see note 1)**	**$\{L',M',S'\}$ = EOTF^{-1} (F_D)** **where F_D = $\{L_D,M_D,S_D\}$**	**$\{L',M',S'\}$ = OETF (E)** **where E = $\{L_S, M_S, S_S\}$**
Derivation of I	**I = 0.5 L' + 0.5 M'**	
Derivation of Color Difference Signals	**C_T = (6610L'-1361M'+7003S')/4096** **C_P = (17933L'-17390M'-543S')/4096**	**C_T = (3625L'-7465M'+3840S')/4096** **C_P = (9500L'-9212M'-288S')/4096**

Table 5 Notes:

> 1- The subscripts D and S refer to display light and scene light, respectively.

10.2. Next Generation Audio (NGA)

Complementing the visual enhancements that Ultra HD will bring to consumers, Next Generation Audio (NGA) provides compelling new audio experiences:

- Immersive – 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
- Personalized – Enabling consumers to tailor and interact with their listening experience, e.g. selecting alternative audio experiences, switching between languages, enhancing dialogue intelligibility.
- Consistent – Playback experience automatically optimized for each consumer device, e.g. home and mobile
- Object-based Audio – Audio elements are programmed to provide sound from specific locations in space, irrespective of speaker location. By delivering audio as individual elements, or objects, content creators can simplify operations, reduce bandwidth, and provide a premium experience for every audience

- Scene-Based Audio – An arbitrarily large number of directional audio elements composing a 3D sound field are mixed in a fixed number of PCM signals according to the Higher-Order Ambisonics format. Once in the HOA format, the Audio Scene can be efficiently transmitted, manipulated, and rendered on loudspeaker layouts/headphones/soundbars.
- Flexible Delivery - NGA can be delivered to consumers over a number of different distribution platforms including terrestrial, cable, and satellite broadcast, IPTV, OTT, and mobile. It could also be delivered over a hybrid of broadcast and OTT
- Flexible Rendering - NGA can be experienced by consumers through headphones or speakers (e.g., TV speakers, home theater systems including ceiling speakers, sound bars) as shown in Figure 4.
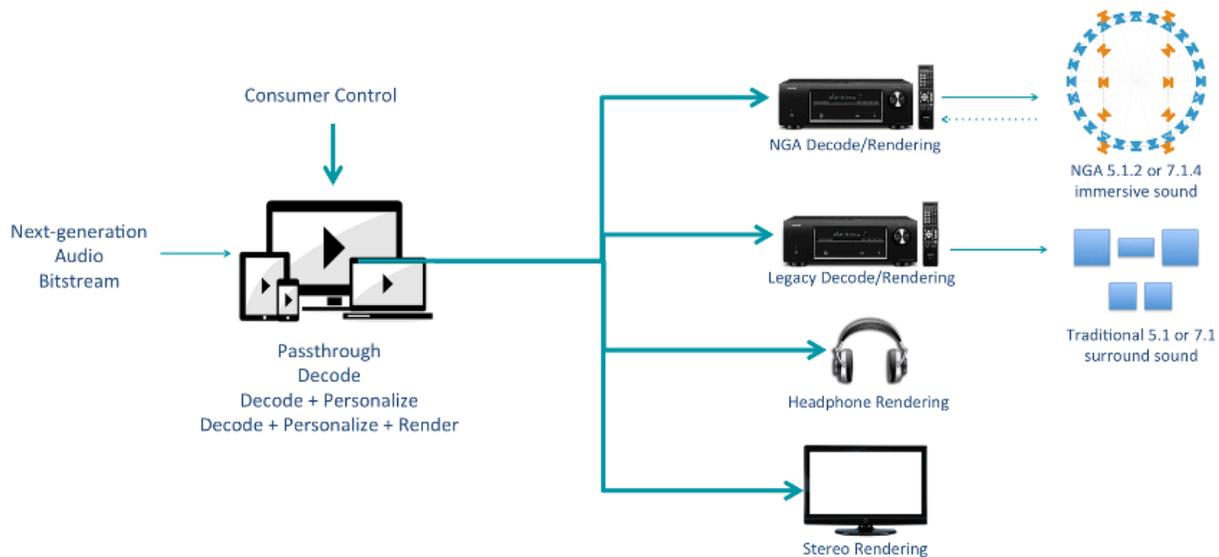

Figure 4. NGA in the consumer domain

10.2.1 NGA Audio Codecs

There are three Next Generation Codecs described in the Guidelines, AC-4, DTS-UHD and MPEG-H. A quick overview of each system is described below.

See Indigo Book: AC-4 [I01], DTS UHD [I02] and MPEG-H [I03]

26

10.2.2 Metadata in NGA

NGA codec systems have a rich set of audio metadata features and functions. Each codec has its own set of definitions; however, there is a common framework for audio metadata developed by the EBU called the Audio Definition Model (ADM) [61].

In general, there are three types of audio metadata:

1. Descriptive - Provides information regarding the available audio program features (i.e., Channel configuration, alternate languages, VDS).
2. Functional - Provides information regarding how the audio should be rendered or presented (i.e., preselections, object audio locations, loudness controls, downmix coefficients).
3. Control - Allows for personalization and user preferences (e.g. Dialog Enhancement, language preference, program preselection)

Immersive programming requires generating and delivering dynamic metadata to playback devices. For immersive programming, object position and rendering control metadata are essential for enabling the optimum set of experiences regardless of playback device or application. This section provides an overview of these important metadata parameters and how they are utilized during the creative process.

10.2.3 Audio Rendering

Audio Rendering is the process of composing an Audio Preselection and converting all the Audio Program Components to a data structure appropriate for the audio outputs of a specific receiver. Rendering may include conversion of a Channel Set to a different channel configuration, conversion of Audio Objects to Channel Sets, conversion of Scene-based sets to Channel Sets, and/or applying specialized audio processing such as room correction or spatial virtualization. In addition, the application of Dialog Enhancement as well as Loudness Normalization are parts of the audio rendering functionality.

Some uses of audio rendering include support for Audio Description services, multiple languages and personalized audio features without having to create separate Complete Mixes for each additional audio service. There are two main concepts of personalized audio:

1. Personalization selection – The bitstream may contain more than one Audio Preselection where each Audio Preselection contains pre-defined audio experiences (e.g., "home

team" audio experience, multiple languages, etc.). A listener can choose the audio experience by selecting one of the Audio Preselections.

2. Personalization control – Listeners can modify properties of the complete audio experience or parts of it (e.g., increasing the volume level of an Audio Element, changing the position of an Audio Element, etc.).

10.3 High Frame Rate Video (HFR)

For the purpose of this document, High Frame Rate (HFR) refers to frame rates of 100 fps or higher, including 100, 120/1.00[5] and 120, Standard Frame Rate (SFR) refers to frame rates of 60fps or lower, which are commonly used including 24/1.001, 24, 25, 30/1.001, 30, 50, 60/1.001 and 60.

According to a SMPTE/HPA paper authored by Mark Schubin[6], frame rates of 100, 120/1.001 and 120 fps add significant clarity to high motion video such as sports or action scenes. Schubin also notes that high dynamic range put new demands on temporal resolution. He notes that, "Viewers of HDR imagery sometimes report increased perception of motion judder... Increased frame rate, therefore, might be necessary to accompany HDR."
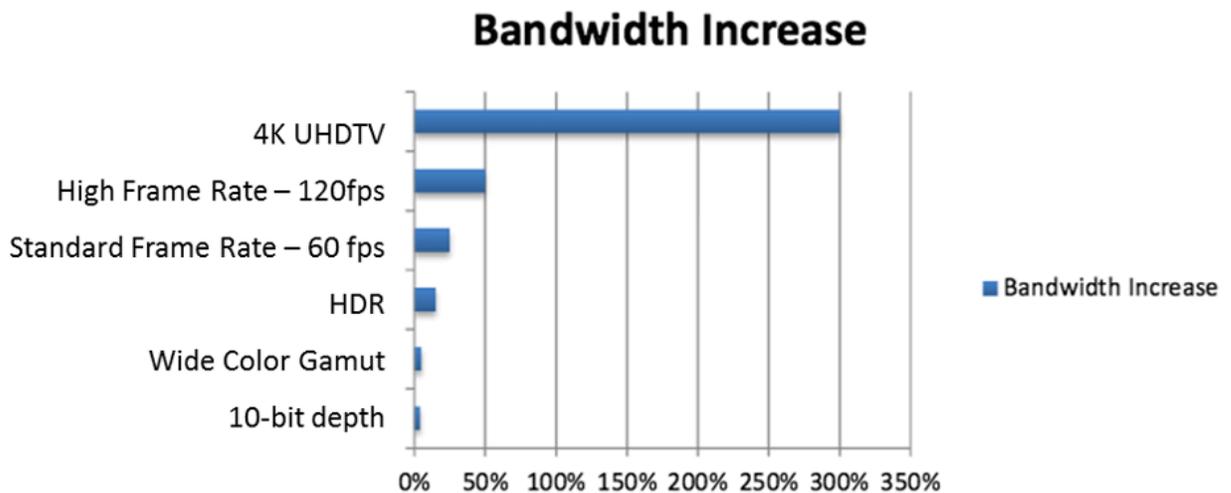
Citing Schubin again, "... the [EBU[7]] found ... that in going from 60 frames per second (fps) to 120 fps or from 120 fps to 240 fps — a doubling of the frame rate — it is possible to achieve a full grade of improvement." Further, doubling the frame rate from 50/60 fps to 100/120 fps is a very efficient means of gaining that full grade of improvement when compared to going from 2K to 4K spatial resolution, as illustrated in Figure 4.

---

[5] Although 120/1.001 is considered an example of HFR, the Ultra HD Forum recommends using integer frame rates for all Ultra HD content whenever possible.

[6] "Higher Resolution, Higher Frame Rate, and Better Pixels in Context The Visual Quality Improvement Each Can Offer, and at What Cost", SMPTE/HPA paper, Mark Schubin, 2014, https://www.smpte.org/publications/industry-perspectives/schubin-HPA2014

[7] Rep. ITU-R BT.2246 The present state of ultra-high definition television [130], p 26.

Bandwidth Increase

Figure 4. Bandwidth increases for various video format improvements compared to HD

HFR has been included in newer DTT television standards including ATSC 3.0 [54] and DVB [63]. As such, the Ultra HD Forum considers HFR to be viable Ultra HD technology that can be layered onto Foundation Ultra HD for DTT. Both systems include a backward compatibility mechanism that enables 50/60 fps decoders to render a 50/60 fps version of the content while 100/120 fps decoders render the full HFR experience. See Section 7.3 in the Yellow Book [Y02] for more information about backward compatibility.

Deployments of HFR, 4K HFR may exceed the capabilities of some portions of the end-to-end ecosystem. For example, while HDMI 2.1 supports 4K 120 fps or 8K 60 fps, most production environment transport systems currently support only 2K with 100/120 fps. Although this is likely to change in the future, the Ultra HD Forum describes HFR with 2K spatial resolution in order to provide an HFR guideline for a full end-to-end system. The parameters for 2K HFR content are shown in Table 6.

Table 6. 2k High Frame Rate Content Parameters

HFR Parameters	Value
Frame Rate	100 fps, 120 fps[1]
Spatial Resolution	HD (1920x1080)
Scan Type	Progressive
Dynamic Range	SDR, HDR
System Colorimetry	ITU BT. 709, ITU BT. 2020
Bit Depth	10
Distribution Codec	HEVC Main 10 Level 5.1
HDMI Interface[2]	100 fps: at least v1.4 120 fps: at least v2.0

Table 6 Notes:

1. Earliest HDMI interfaces that support 10-bit, 1920x1080p, SDR high frame rate. HDMI 1.4 also supports 120fps with 4:2:2 chroma subsampling. HDR support requires HDMI 2.0a for HDR 10 and 2.0b for HLG.

2. Note that 120/1.001 may be used for backward compatibility; however, the Ultra HD Forum recommends using integer frame rates for all Ultra HD content whenever possible.

For more complete information on HFR , please see [Y02].

11. References

[1]   Recommendation ITU-T H.222.0 | ISO/IEC 13818-1:2021, "Information Technology—Generic coding of moving pictures and associated audio information - Part 1: Systems", June 2021, https://www.itu.int/rec/T-REC-H.222.0-202106-I/en

[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

[5]   Recommendation ITU-R BT.2100, "Image parameter values for high dynamic range television for use in production and international programme exchange", July 2018, http://www.itu.int/rec/R-REC-BT.2100

[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=a1BVR0000008kal2AA

[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.pdfhttp://www.etsi.org/deliver/etsi_en/300700_300799/300743/01.03.01_60/en_300743v010301p.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/

[54]   ATSC: A/341:2024-04, "Video-HEVC", April 3, 2024, https://www.atsc.org/wp-content/uploads/2024/04/A341-2024-04-Video-HEVC.pdf

[61]   EBU Tech 3364, "Audio Definition Model Metadata Specification Ver. 2.0", June 2018, https://tech.ebu.ch/docs/tech/tech3364.pdf

[63]   ETSI TS 101 154 v2.4.1 (2018-02), "Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcasting", February 14, 2018, https://www.etsi.org/deliver/etsi_ts/101100_101199/101154/02.04.01_60/ts_101154v020401p.pdf

[130]   Report ITU-R BT.2246-8, "The present state of ultra-high definition television", Mar 2023, https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-BT.2246-8-2023-PDF-E.pdf

[Y]   **Yellow Book** – Beyond Foundational Technologies

[Y01]   Section 7.4.1, AVS2 & AVS3

[Y02]   Section 7.3, High Frame Rate Video (HFR)

[I]   **Indigo Book** – Ultra HD Technologies

[I01]   Section 8.1, Dolby Vision

[I02]   Section 8.2, Dual Layer - Scalable High-Efficiency Video Coding (SHVC)

[I03]   Section 8.3, SL-HDR1

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

[V01]   Section 9.1, Table 1, Foundation Decoder Capabilities

[V02]   Section 9.2, Table 2, Foundation Service Formats

(End of Red Book)