

Report ITU-R BT.2246-8

(03/2023)

BT Series: Broadcasting service (television)

The present state of ultra-high definition television



Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from <http://www.itu.int/ITU-R/go/patents/en> where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Reports

(Also available online at <https://www.itu.int/publ/R-REP/en>)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	Broadcasting service (television)
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
TF	Time signals and frequency standards emissions

Note: This ITU-R Report was approved in English by the Study Group under the procedure detailed in Resolution ITU-R 1.

*Electronic Publication
Geneva, 2023*

© ITU 2023

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without written permission of ITU.

REPORT ITU-R BT.2246-8

The present state of ultra-high definition television

(2011-08/2012-11/2012-2014-02/2015-07/2015-2017-2020-2023)

Summary

This Report contains the results of the study on ultra-high definition television (UHDTV) and describes the present state of UHDTV. It addresses the baseband image format and the derivation of system parameter values in particular. Also included is a description of the operational experience of providing UHDTV coverage of the London Olympics.

TABLE OF CONTENTS

	<i>Page</i>
Policy on Intellectual Property Right (IPR)	ii
REPORT ITU-R BT.2246-8.....	1
1 Introduction	4
2 General considerations.....	4
2.1 Fundamental idea of UHDTV	4
2.2 Classification of image systems	4
2.3 Application	5
2.4 System model	7
2.5 Viewing conditions	8
2.6 Feasibility of technologies required for development of UHDTV	10
3 Baseband image format	11
3.1 General considerations on the baseband image format in programme signal chain.....	11
3.2 Parameters	12
4 Operational experience	45
4.1 London Olympics SUPER Hi-VISION public viewing operations.....	45
4.2 Implementation of UHDTV broadcasting	45
4.3 Initiatives for 8K/120 Hz broadcasting	45
5 Applicable specifications for UHDTV broadcasting.....	45
Annex 1 – Input documents to WP 6C or SG 6 related to UHDTV	49
Annex 2 – Analysis of sub-sampled versions of the UHDTV signals.....	54

1	Summary	54
2	Fundamental	54
2.1	Why YC transmission is needed	54
2.2	Fundamental of YC transmission.....	54
2.3	The constant luminance issue.....	55
2.4	Block diagrams for the encoder and decoder in the non-constant luminance (NCL) and CL	56
2.5	Crosstalk characteristics	57
2.6	Quality degradation by clipping.....	59
2.7	Evaluation index	60
2.8	Factors other than picture quality.....	60
3	The reference models of the NCL and CL.....	61
4	Objective and subjective evaluation results.....	62
4.1	Objective evaluation of YC encoding-decoding.....	62
4.2	The evaluation results in a production process	63
4.3	Real implementation vs. the hypothetical loss in G for the CL	65
4.4	The evaluation results in a delivery chain.....	67
5	Summary of the evaluation results.....	77
	Annex 3 – London Olympics SUPER Hi-VISION Public Viewing Operations Report.....	79
1	Introduction	79
2	Overview.....	79
3	Venue production	80
4	Transmission and editing	84
5	Distribution and screening.....	88
6	Conclusion.....	91
	Annex 4 – Implementation of UHD TV broadcasting.....	91
1	Japan.....	91
1.1	UHD TV satellite broadcasting services	91
1.2	Specifications for UHD TV satellite broadcasting	91
1.3	List of ARIB Standards	94

Annex 5 – Initiatives for 8K/120 Hz broadcasting	95
1 Development of high frame frequency 8K equipment	97
1.1 High speed camera	97
1.2 Studio Monitor	98
1.3 Video/audio encoder and decoder.....	99
1.4 Sheet-type flexible OLED display	99
2 Live production and transmission	100
Annex 6 – Practice of 8K UHD programmes broadcasting on large screens in China	101
1 Overview.....	101
2 8K UHD video and 3D audio broadcasting & distribution system workflow	101
3 8K programme production and broadcasting in CMG	103
3.1 Basic technical specifications for 8K UHD programme video and audio.....	103
3.2 8K UHD programme video and audio production requirements and specifications.....	103
3.3 8K UHD signal routing system.....	106
3.4 8K UHD broadcast system.....	106
4 8K distribution and transmission.....	107
4.1 Distribution and transmission of video and audio over private network.....	107
4.2 Distribution and transmission of audio over Internet	107
4.3 Video and audio synchronization in heterogeneous network (HetNet).....	107
5 8K signal reception and display	108
5.1 Signal reception and display of "BaiChengQianPing"	108
5.2 System architecture	108
5.3 Reception and display technical specifications.....	109
6 Main technology.....	113
6.1 AVS3	113
6.2 HDR VIVID	113
6.3 Audio Vivid.....	115

1 Introduction

Television has built its history on the fundamental desire of human beings to extend their audiovisual senses spatially and temporally. HDTV is one of the great achievements of television. People in many countries are now enjoying the benefits of HDTV, and people in the rest of the world will soon benefit in the near future.

The definition of HDTV, described in Report ITU-R BT.801 – The present state of high-definition television, is as follows:

A high-definition system is a system designed to allow viewing at about three times the picture height, such that the system is virtually, or nearly, transparent to the quality of portrayal that would have been perceived in the original scene or performance by a discerning viewer with normal visual acuity. Such factors include improved motion portrayal and improved perception of depth.

The attainment of this goal is limited in some aspects; e.g. the field of view of HDTV is only 30 arc-degrees. Our natural desire to overcome such limitations has led us to the concept of UHDTV:

UHDTV is a television application that is intended to provide viewers with an enhanced visual experience primarily by offering a wide field of view that virtually covers all of the human visual field with appropriate sizes of screens relevant to usage at home and in public places.

2 General considerations

2.1 Fundamental idea of UHDTV

UHDTV is a television application that will provide viewers with a better visual experience primarily by offering a wide field of view (FOV) which virtually covers all of the human visual field, while maintaining other features of HDTV or improving them. UHDTV could therefore be characterized as a TV system having a wide field of view supported by enhanced spatial resolution.

2.2 Classification of image systems

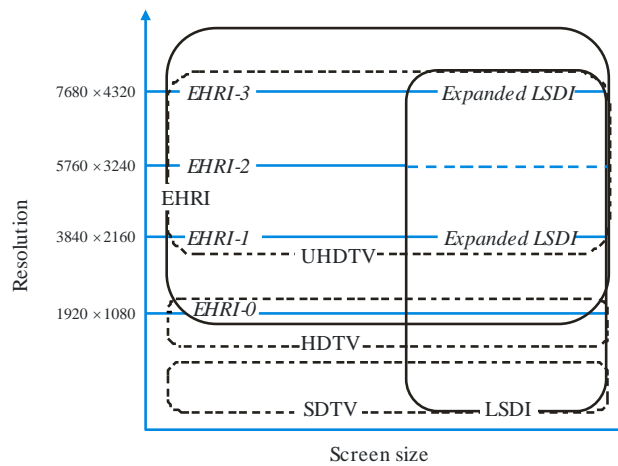
Figure 1 classifies the image systems that ITU-R deals with in terms of resolution and screen size. Resolution is one of the most important parameters in ITU-R Recommendations for television systems (image formats). Screen size has not been regarded as a parameter intrinsic to image formats.

The resolution of HDTV is described in Recommendation ITU-R BT.709 as $1\,920 \times 1\,080$. The resolution of LSDI is described in Recommendation ITU-R BT.1680, and the resolution for an expanded hierarchy of LSDI, namely $3\,840 \times 2\,160$ and $7\,680 \times 4\,320$, is described in Recommendation ITU-R BT.1769. The screen size of LSDI is described in Question ITU-R 15-2/6 as ... *large screen presentation in ... theatres, halls, and other venues.*

The resolution of EHRI is described in Recommendation ITU-R BT.1201 as a simple integer multiple of $1\,920 \times 1\,080$. The specific values given in Report ITU-R BT.2042 are 1, 2, 3 and 4. The screen size of EHRI is not described anywhere in ITU-R documents since its basic idea is not related to the screen size.

From the conceptual viewpoint, it is natural to suppose that the resolution would exceed that of HDTV. It would also be natural to suppose that the screen size would be suitable for home UHDTV receivers. Nevertheless, as in the case of HDTV, the applications of UHDTV would probably not be restricted to home use. Therefore, it is reasonable to conclude that the screen size of UHDTV would not be limited to the range applicable to home use.

FIGURE 1
Classification of image systems



Report BT.2246-01

2.3 Application

2.3.1 Overview

Compared with current HDTV, the UHDTV application should bring considerably improved benefits to its viewers. Those benefits may include:

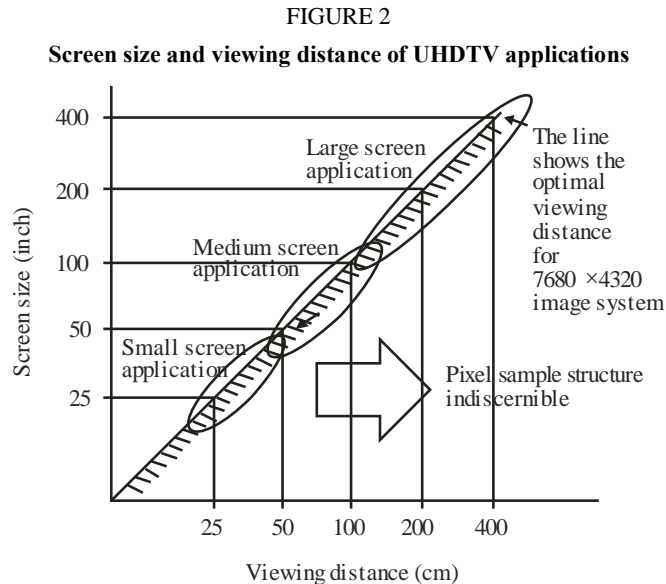
- stronger sensation of reality or presence;
- higher transparency to the real world;
- more information.

It may be presented in:

- living rooms;
- personal spaces in mobile and non-mobile environments;
- collective viewing locations such as theatres.

Each screen should be well suited to the particular form of usage.

Recommendation ITU-R BT.1845 – Guidelines on metrics to be used when tailoring television programmes to broadcasting applications at various image quality levels, display sizes and aspect ratios, shows the relationship between the screen size and viewing distance given that the optimal viewing distance is one at which one pixel corresponds to the visual angle of one arc-minute. The optimal viewing distance is, for example, 100 cm when viewing an UHDTV image on a 100-inch screen. One pixel per one arc-minute is generally regarded to correspond to a visual acuity of 1.0. This means that the pixel sampling structure is not detectable when viewing under the conditions on the right and below the line in Fig. 2.



A viewer typically watches a television while remaining stationary at around two to three metres from the display. Such a viewing style for UHDTV corresponds to a large screen application (Fig. 2) that may include theatrical environments as well as home environments. It will provide stronger sensation of reality and a stronger immersive feeling to viewers by offering a far wider field of view than current systems can offer.

The second category in Fig. 2 is a medium-size screen application with screen sizes from around 50 to 150 inches. The respective optimal viewing distances are 50 to 150 cm. The viewing style for this application may include not only the conventional one but also a new one in which viewers changes their viewing distance according to their preference or the content creator's intention. For example, fine art can be appreciated by standing as close as the painters painted them at. So too, with a UHDTV image of fine art.

The third category is a small-screen application. Similar applications for text data or still images have recently emerged and are known as electronic paper. The optimal viewing distance for a 20-inch screen is 20 cm, at which the HVS reaches the accommodation limit. The size of a 20-inch screen is close to A3, and a 7 680 × 4 320 pixel screen would have approximately 350 pixels per inch.

The use-cases described above are in accordance with our natural desire to extend our sensory limits. Some of them exceed the capabilities of conventional television. They should be studied as part of the development of future television applications.

The transmission paths to the presentation venue may vary depending on technical and economic reasons. They may include:

- satellite radio transmission;
- terrestrial radio transmission;
- cable and network;
- new wide-band networks, radio or cable, that can accommodate virtually lossless UHDTV signals;
- package delivery.

In [Input, 9], a broadcaster expressed its view on the introduction of UHDTV as follows:

- If we imagine a giant screen and a multichannel audio system with as many as 22.2 channels, the image and sound presentation arrangements required by the UHDTV are not likely to be broadly implemented or implementable in consumer homes. They might perhaps be implementable to some extent in a dedicated “TV room”, which may more properly be called a “home-cinema room”.
- Both UHDTV systems, and particularly the 7 680 × 4 320 system will likely find their most successful application in LSDI, namely in the large-screen presentation of digital programmes for collective viewing under controlled conditions, in suitably equipped venues such as auditoria, cinema-like halls and similar places, on condition, of course, that programmes in UHDTV will be available in adequate number and quality to support such a broadcasting service.

2.3.2 Considerations of audience viewing

Worldwide television audiences have acquired certain viewing habits, which they might be reluctant to change unless the viewing experience is enhanced greatly.

A typical viewing habit is to watch television while sitting on the sofa in the living room. The TV screen is watched from a position perhaps 7 feet (2.1 m) away. This viewing distance did not change when television transitioned from black and white to colour, or from SDTV to HDTV. Indeed, in the latter transition, the viewing distance remained roughly the same, while the TV screen became larger.

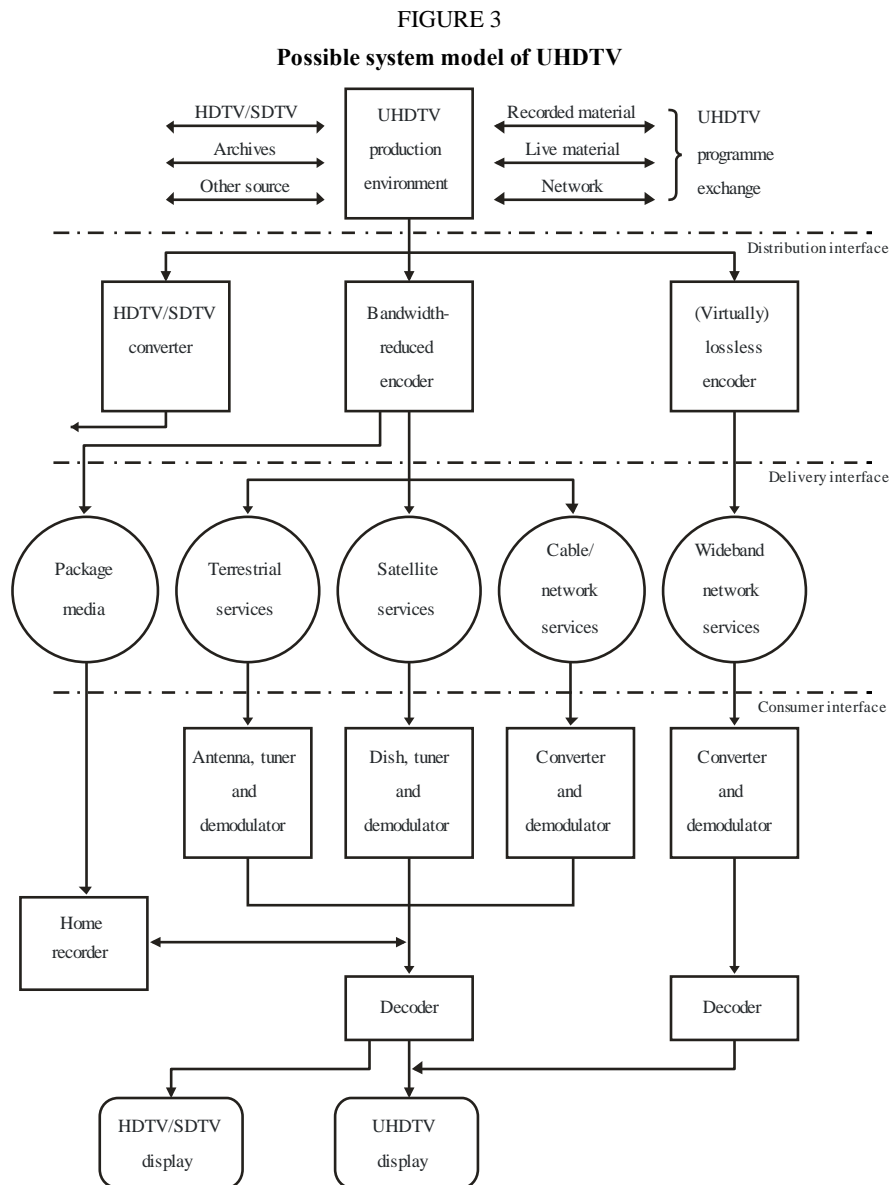
A 52” (132 cm) HDTV 1080” receiver today is viewed at a distance of about 7 feet (2.1 m). Assuming that the home viewing habits will not change dramatically when transitioning from HDTV to 3 840 × 2 160 UHDTV, it can be expected that 100” (2.5 m) screens may well exist, in some instances such a large screen would take up a good part of the living room wall. This could be considered an enhanced viewing condition.

It cannot be excluded that consumers would be prepared to install such a large screen in their living room, but it seems unlikely that they will be prepared to install the 200” (5 m) screen that could be recommended for 7 680 × 4 320 UHDTV viewing.

It could be concluded that, from the point of view of image presentation, 3 840 × 2 160 UHDTV television broadcasting to the home may find acceptance with in-home television audiences, while it may be unlikely that 7 680 × 4 320 UHDTV will do so. The 7 680 × 4 320 UHDTV image system may instead find applications for television presentations to the public in theatres, in home theatres, auditoria, theme parks, and other public venues.

2.4 System model

Figure 3 depicts a possible system model for the applications described above.



2.5 Viewing conditions

2.5.1 Field of view

Among the properties of current television pictures, field of view may be the one that is most inferior to the capabilities of the human visual system. Nevertheless, UHDTV does not intend to cover the whole surrounding field of view like that of virtual reality. The required value for a two-dimensional display to realize the applications described above should be explored.

2.5.2 Angular resolution

The “design viewing distance” gives the angular resolution (pixel/unit visual angle). It has tacitly been taken to be one pixel/arc-minute. Recommendation ITU-R BT.1127 – Relative quality requirements of television broadcast systems describes the design viewing distance as the relative distance to the picture height at which the picture quality of each system always falls into the “excellent” evaluation range. The relative viewing distance to the picture height is an alternative expression of FOV for the same television system. These expressions are related as follows.

$$\tan(a/2) = r/(2n)$$

Here, a is FOV, r is aspect ratio, and n is design viewing distance to picture height.

Recommendation ITU-R BT.1845 – Guidelines on metrics to be used when tailoring television programmes to broadcasting applications at various image quality levels and sizings defines the optimal viewing distance as the distance at which the pixel count per visual angle of one minute is one. It lists the optimal viewing distances to the picture height and the optimal field of view for image systems with various pixel counts (Table 1).

These two Recommendations suggest that the picture quality of an image system having viewing conditions in which the angular resolution is one pixel per one arc-minute falls into the “excellent” range.

With above idea in mind, the issue of pixel count should be addressed from the following viewpoints.

- What should the (minimum and maximum) FOV of UHD TV be and for what reason?
- Should the criterion for determining the picture quality of UHD TV by the pixel count be the same as that in Recommendations ITU-R BT.1127 and ITU-R BT.1845?

TABLE 1

**Optimal horizontal viewing angle and optimal viewing distance in image heights (H)
for various digital image systems**

Image system	Reference	Aspect ratio	Pixel aspect ratio	Optimal horiz. viewing angle	Optimal viewing distance
720 × 483	Rec. ITU-R BT.601	4:3	0.88	11°	7 H
640 × 480	VGA	4:3	1	11°	7 H
720 × 576	Rec. ITU-R BT.601	4:3	1.07	13°	6 H
1 024 × 768	XGA	4:3	1	17°	4.4 H
1 280 × 720	Rec. ITU-R BT.1543	16:9	1	21°	4.8 H
1 400 × 1 050	SXGA+	4:3	1	23°	3.1 H
1 920 × 1 080	Rec. ITU-R BT.709	16:9	1	32°	3.1 H
3 840 × 2 160	Rec. ITU-R BT.1769	16:9	1	58°	1.5 H
7 680 × 4 320	Rec. ITU-R BT.1769	16:9	1	96°	0.75 H

2.5.3 Viewing distance

Absolute viewing distance differs according to the kind of application. The range may be wider than that of current television because of the variety of applications. Relative distance varies with the field of view and resolution. It also determines the display size and vice versa.

2.5.4 Surrounding luminance

Surrounding luminance affects human perception of displayed images. Therefore, certain assumptions may be needed to determine the system parameters. These assumptions would depend on the kind of application.

2.5.5 Black and white luminance and electro-optical transfer function (EOTF)

The perceptual limit of the black level depends on the accommodation situation of the human visual system (HVS) and is mainly affected by the surrounding luminance. Modern (non-CRT) displays have various native EOTFs, and they can easily be converted into almost arbitrary values by using

modern digital techniques. Therefore, the major criterion for determining EOTF may be the efficiency of digital code usage.

2.5.6 Technical constraints imposed by the assumed display device

In the CRT era, system designers tacitly agreed upon the temporal sampling aperture, chromaticity coordinates of tri-stimulus (also their spectrum distribution), and EOTF. The UHDTV display will likely be some other kind of device; hence, those assumptions should be reviewed.

2.6 Feasibility of technologies required for development of UHDTV

In [Input, 9] a broadcaster gave their view on the feasibility of technologies required for UHDTV delivery as follows.

- It will probably never be possible to implement a UHDTV broadcasting service in the frequency bands currently assigned to terrestrial broadcasting, since it will not be possible to achieve the amount of bit-rate reduction necessary to fit the data rate required by UHDTV into the capacity of current terrestrial broadcasting channels.
- It may be possible to implement a $3\,840 \times 2\,160$ UHDTV broadcasting service in the 12 GHz satellite-broadcasting band in the medium future, when more efficient source-coding and modulation methods will have been developed and implemented in a reliable and viable way. Some experts think that this goal may be attained in about ten years' time, based on studies currently under way on improved source coding and modulation methods. It remains to be seen whether, at that time, the 12 GHz band will be too densely populated to accommodate the new UHDTV services.
- It will probably be possible to implement a $7\,680 \times 4\,320$ UHDTV broadcasting service in the 22 GHz broadcasting service in the far future, when another quantum leap in the efficiency of source coding and modulation methods will have been developed and implemented. Some experts think that this goal may be attained in perhaps 10 to 20 years. It remains to be seen how the 22 GHz satellite broadcasting band will be structured at that time, and which applications will be earmarked for it.

A performance evaluation of compression coding that may be used in the secondary distribution [Input, 17] shows that:

- For 75% of the sequences chosen: $DSCQS \leq 12\%$
- For the rest: $DSCQS \leq 30\%$.

In [Input, 39] the increase in bit rate needed for the picture quality improvement including motion portrayal and colour/tone rendition is discussed from the feasibility point of view as follows.

Since $3\,840 \times 2\,160$ UHDTV broadcasting will require a new television receiver, it seems logical to also ask whether the opportunity should also be seized to improve some other aspects of the television image.

Beyond the improvement in resolution, at least two other possible areas of picture improvement could be considered: improvement in colour rendition and improvement in movement rendition.

Since any improvement in service quality has an associated cost, the issue is to decide whether the improvements are worth its cost.

In the case of image resolution, the transition from HDTV to $3\,840 \times 2\,160$ UHDTV broadcasting would imply the cost of carrying four times more pixels to the user, which would result in a four-fold increase in uncompressed bit rate. Today, each HDTV broadcast service uses all or a major part of a broadcast emission channel. A four-fold increase in uncompressed bit rate would thus be well beyond the capabilities of the source coding and channel coding methods currently used for television

broadcasting. In other words, the emission of a $3\,840 \times 2\,160$ UHD TV broadcasting service would require a bit rate that exceeds the capacity of a current terrestrial television channel.

In the case of movement rendition, the first possibility would be to double the picture rate by delivering progressive signals rather than interlaced signals. The second possibility would be to further double the picture rate by delivering progressive signals at a picture rate of 100 Hz or 120 Hz, rather than 50 Hz or 60 Hz. Taking into account that the correlation among spatially co-located pixels increases as their temporal distance decreases, each one of those increases in pixel count would probably result in about 50% increase in compressed bit rate assuming no technology advances.

In the case of colour rendition, Working Party 6C has extensively discussed this topic, concluding that an improvement in colour rendition to encompass the complete gamut of visible colours would require at least 10 bits/sample, rather than the 8 bits/sample now required for SDTV and HDTV emission: a 25% increase in bit rate.

It is clear that the possible image improvements above, as well as any increase in the number of audio channels, will all compete for the available emission bit rate, with the four-fold increase in pixel count of $3\,840 \times 2\,160$ UHD TV. It will be necessary to make a choice among the improvements that can be fitted in the available emission bit rate, and those that cannot be, unless an improvement is made in source coding, and/or in channel coding, or unless multiple terrestrial TV channels are devoted to a single UHD TV program stream.

3 Baseband image format

3.1 General considerations on the baseband image format in programme signal chain

The baseband image format is the basis of every part of the programme signal chain. Therefore, it should be established first so that every part of the UHD TV broadcast system can be based on it. The formats for different parts of the programme chain may differ from each other because of the requirements for each part. In particular, the acquisition and production format may need some quality headroom for post-processing management.

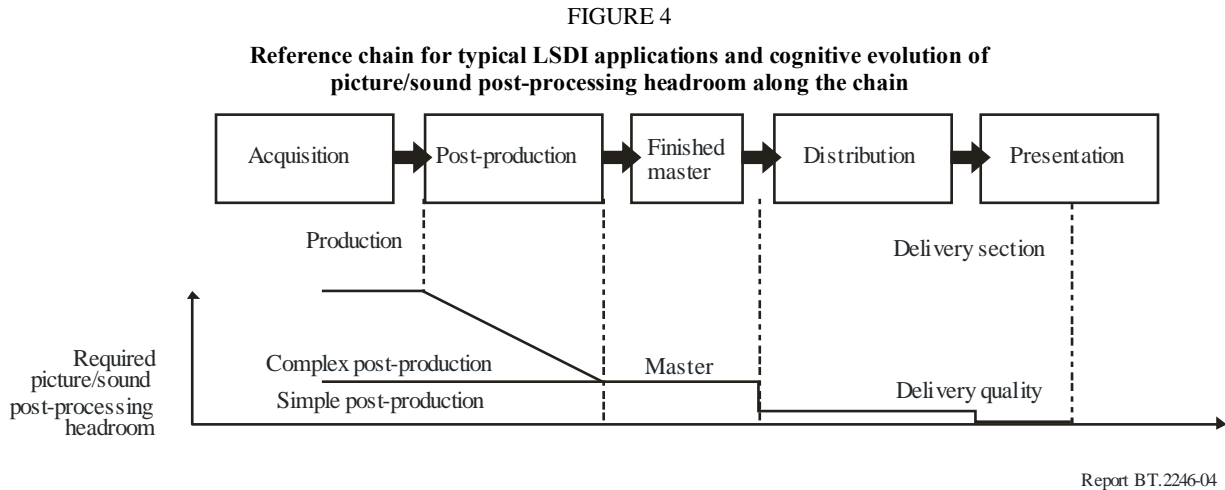
Recommendation ITU-R BT.1662 “General reference chain and management of post-processing headroom for programme essence in large screen digital imagery applications” describes the general reference chain from acquisition to presentation and recommends as follows (see also Fig. 4, taken from Recommendation ITU-R BT.1662):

- a single native format depending upon the quality level should be used for programme essence throughout the chain, if possible;
- if that is not possible due to complex post-processing requirements, then a minimum number of native formats should be used along the chain, and their parameter values should stand in a simple relationship to each other.

The acquisition block should normally use the same native format that is planned to be used for the finished master (simple post-production). However, depending on the nature of required post-production processes, the required amount of headroom on the required parameters should be provided at the input of the post-production block (complex post-production). This can be done by using different native signal formats. To fully benefit from this, the native formats should be related in a direct way to the format intended for the finished master.

This signal chain would likely be applicable to UHD TV since it will deal with programmes having high picture quality. This observation leads to the following suggestions.

- A baseband image format for high-picture-quality programme production should be developed.
- It should be based on the image format for finished masters.



3.2 Parameters

A set of parameter values for the baseband image format was proposed in 2009 [Input, 18], this proposal was further enhanced during the May 2011 series of meetings when proposals from two proponents were merged into a single proposal.

3.2.1 Aspect ratio

Aspect ratio is one of the fundamental properties of the picture because it is difficult to convert one ratio into another without perceptible distortion. A paper in 1990s [1] concludes based on a thorough study that there is no clear evidence of an aesthetic or physiological reason to choose any one aspect ratio over another.

The resolution of EHRI is described in Recommendation ITU-R BT.1201 as a simple integer multiple of $1\,920 \times 1\,080$. This leads to the choice of an aspect ratio of 16:9.

3.2.2 Spatial and temporal sampling structure

A non-orthogonal structure would fit the HVS characteristics. Such a sample structure would result in a reduction in the data rate. Still, continuing the practice of using orthogonal sampling may be desirable. For temporal sampling, interlace would not be needed to reduce the bandwidth because alternative techniques such as high-efficiency video coding are now available.

3.2.3 Pixel count

The pixel count can be derived from the required field of view and required resolution (pixels per unit visual angle).

3.2.3.1 Subjective assessment on “sense of being there” and “sense of realness”

A subjective assessment of psychophysical effects of UHDTV was conducted in Japan. The test conditions are summarized in Table 2. A flat panel display with 85-inch diagonal and $7\,680 \times 4\,320$ pixels¹ (Fig. 5) was used. Assessment results are shown in Fig. 6. To summarize the results, when viewed at a distance of less than three times the screen height, $3\,840 \times 2\,160$ and $7\,680 \times 4\,320$ systems provide viewers with a greater sense of “being there” and greater sense of realness than with the $1\,920 \times 1\,080$ system and the $7\,680 \times 4\,320$ system performs the best.

¹ The display accompanied with 22.2 channel sound system was demonstrated in September 2011 at the occasion of WP 6C meeting.

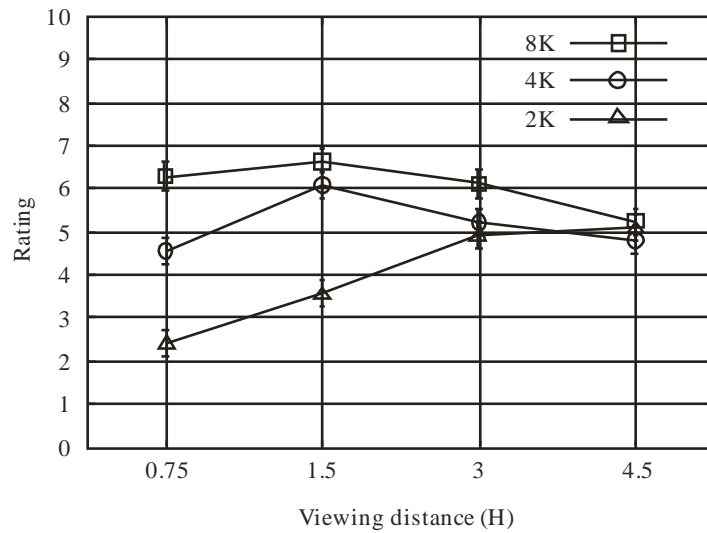
TABLE 2
Test conditions

Display	LCD, 85-inch diagonal, 7 680 × 4 320 pixels
Image resolution	8K(7 680 × 4 320), 4K(3 840 × 2 160), 2K(1 920 × 1 080) 4K and 2K images were produced from 8K by low-pass filtering and down/up-sampling.
Viewing distance (H)	0.75, 1.5, 3.0, 4.5 (H: screen height)
Viewers	53 non-experts
Assessment method	0-10 continuous quality scale

FIGURE 5
Test image displayed on 85-inch LCD

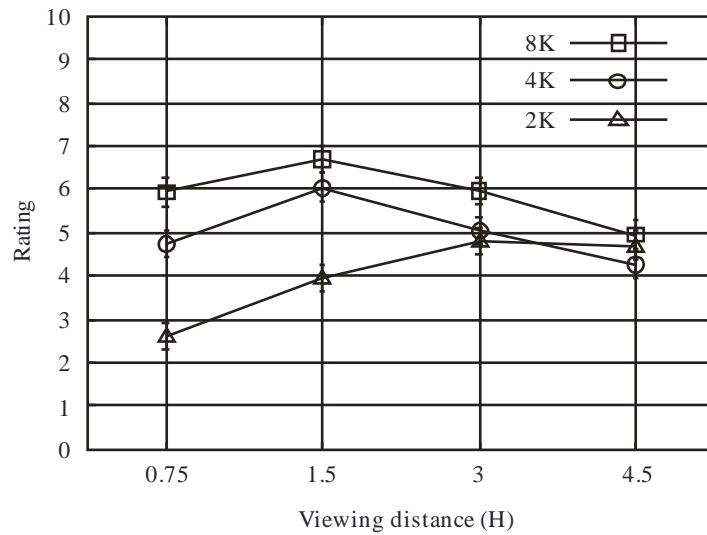


FIGURE 6
(a) Sense of “being there”



Report BT.2246-06a

(b) Sense of realness



Report BT.2246-06b

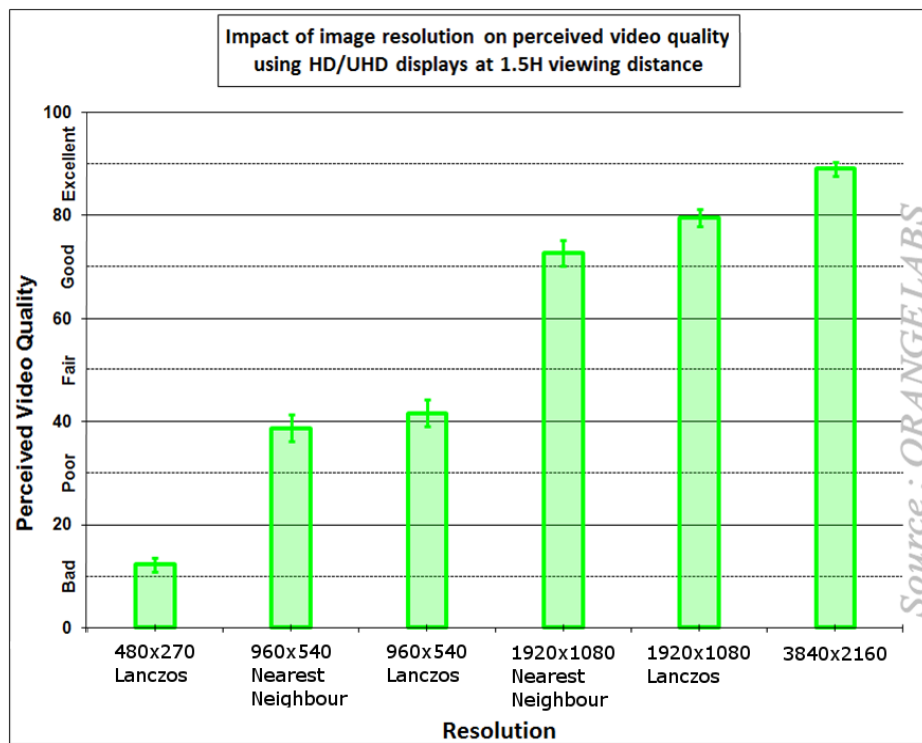
3.2.3.2 Subjective assessment on video quality [Document 6C/246]

A subjective assessment focusing on video quality of UHD TV was conducted by Orange. The test conditions are summarized in Table 3. A flat panel display with 55-inch diagonal and 3 840 × 2 160 pixels was used. Assessment results are shown in Fig. 7. The main conclusion of the study is that at 1.5H viewing distance, the video quality of 3 840 × 2 160 images is half a grade higher in quality on the five-point scale compared to 1 920 × 1 080 images.

TABLE 3
Test conditions

Display	LCD, 55-inch diagonal, 3 840 × 2 160 pixels
Image resolution	3 840 × 2 160, 1 920 × 1 080, 960 × 540 and 480 × 270. 1 920 × 1 080, 960 × 540 and 480 × 270 images were produced from 3 840 × 2 160 by low-pass filtering and down/up-sampling (Lanczos). A nearest neighbour up-sampling processing also permits to simulate a HD display
Viewing distance (H)	1.5 (H: screen height)
Viewers	26 non-experts
Assessment method	0-100 continuous quality scale according to Recommendation ITU-R BT.500-13

FIGURE 7



3.2.4 Frame frequency

The image frame frequency chosen for the capture, postproduction and international exchange of UHDTV programmes depends on various factors such as, for example:

- the programme genre;
- whether the programme is intended for use in areas that use 50 Hz or 60 Hz television;
- in the case of 60 Hz areas, whether the broadcasters use integer frame frequencies or fractional frame frequencies for programme emission;
- the creative decisions of the programme director and of the director of videography.

3.2.4.1 General

Motion reproduction is one of the most important features of television. If we take “artefact” to mean any perceptual difference between the real scene and the reproduced image, the major artefacts would be flicker, motion blur, and jerkiness (stroboscopic effect).

Flicker is a phenomenon in which the whole screen or a large area of the screen “flickers” due to low frame frequency. Motion blur is due to the accumulation mechanism of the capturing devices. It also happens when moving images are tracked on the screen of a hold-type display. Non-smooth reproduced motion at low frame frequencies is called jerkiness. The use of the shutter during acquisition can produce a multi-exposure-like image even at high frame frequencies. This phenomenon is sometimes called jerkiness as well. Motion blur and jerkiness depend on the object speed.

The object speed (arc-degree/second) generally tends to become faster as the field of view of the system becomes wider. For example, given the same shooting angle, an object on an HDTV screen will move faster than the one on an SDTV screen. This implies that high definition TV systems have stricter motion portrayal requirements.

3.2.4.2 Studies done in the HDTV R&D era

The study on the frame frequency was the most difficult one to reach an agreement on in the HDTV standardization. Report ITU-R BT.801 – The present state of high-definition television, describes it in detail.

The study was concerned with two major factors.

- motion portrayal;
- the relationship with film and with current and future TV systems.

After quite a lot of discussion, consequently, 60, 50, 30, 25, and 24 Hz were adopted as possible frame frequencies. The first two values are the same as in SDTV whereas the lower three frequencies are intended for special program production.

The rationale for maintaining the same frame (field) frequencies as of SDTV regardless of the desire for motion portrayal improvement was as follows.

- Dynamic resolution is improved by the use of a shutter in the camera.
Regarding the desire for flicker reduction, the rationale for maintaining the same frame (field) frequencies as in SDTV was as follows.
- It is not essential that the display refresh rate be the same as the studio or the emission signal field-rates.

The current HDTV has a 50 or 60 Hz display refresh rate, 100% aperture in acquisition, and virtually 0% (impulse-type) aperture in display.

The recent advent of FPDs that have a long hold time have solved one problem at the expense of causing another. Their hold mechanism works favourably to reduce flicker. In fact, large-screen FPDs have no flicker problems. However, the hold mechanism also creates motion blur. Many studies have been undertaken to seek a means to eliminate motion blur. So far, having the display itself create and insert frames to double, triple, or even quadruple the frame frequency is seen as a potential solution. It is also noted that newer FPD technology does not suffer from long hold times.

Our review of the studies on the frame (field) rate of HDTV suggests that we should look carefully at the UHD TV frame rate from the following aspects.

- Widening of the field of view (FOV) may affect the human visual system’s perception of the portrayed motion. The critical flicker frequency increases as FOV becomes wider, for example.

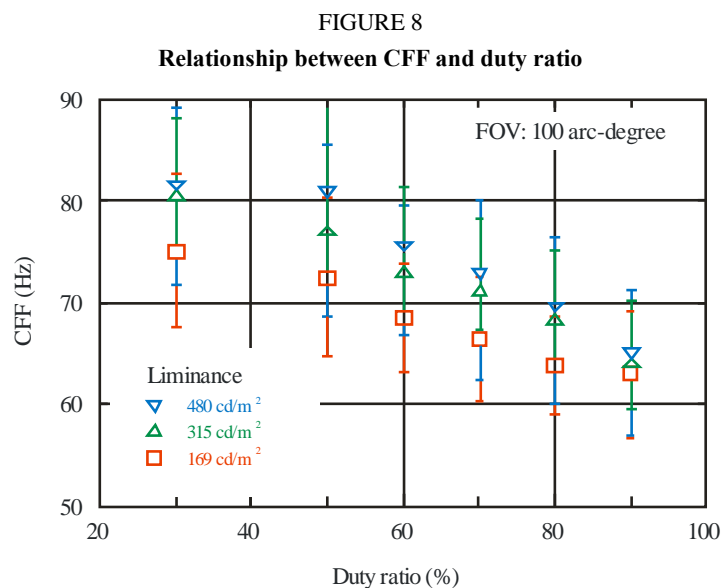
- How much improvement is worthwhile?
- The results of such investigations depend on the type of display and signal processing at the display itself.

3.2.4.3 Recent studies

Flicker [2]

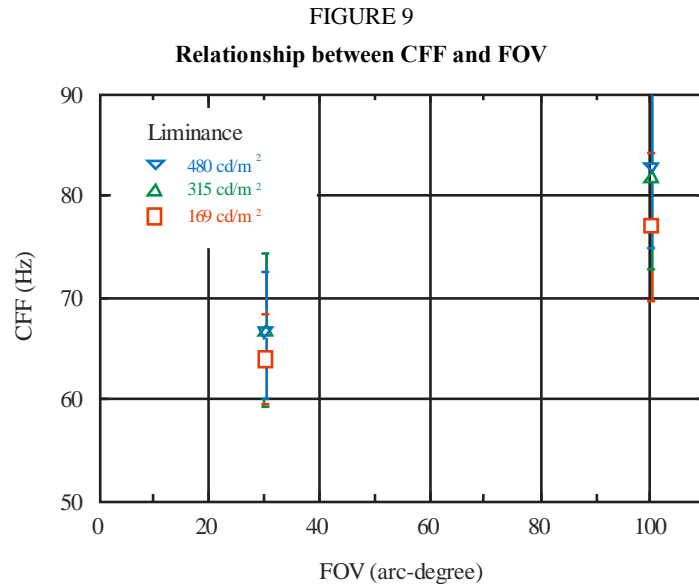
Critical flicker frequencies (CFFs) with varying duty ratios and FOVs were measured in order to characterize the parameters of the HVS while viewing a wide FOV display. Two experiments were conducted. In one experiment, CFFs were measured for various duty ratios ranging from 30% to 90% with the FOV of 100 arc-degrees. In the other experiment, CFFs were measured for 30 and 100 arc-degree FOVs under the condition of the most sensitive duty ratio for each viewer. These two experiments enabled the CFFs of wide FOV systems to be compared with those of narrow FOV systems.

Figure 8 shows the mean CFFs and standard deviations at FOV of 100 arc-degrees for duty ratios of from 30% to 90% at three levels of luminance. CFFs decreased from 80 Hz to 65 Hz as the duty ratio increased.



Report BT.2246-07

Figure 9 shows the mean CFFs and standard deviations for FOVs of 30 and 100 arc-degrees at three levels of luminance. The CFFs of the wide FOV were above 80 Hz and were higher than the CFFs of the narrow FOV (about 65 Hz).



Report BT.2246-08

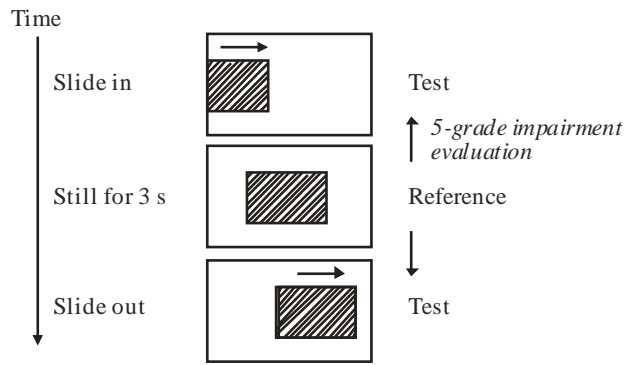
These results show that CFFs decrease as the duty ratio increases. This suggests that viewers would perceive more flicker in impulse-type displays than in hold-type displays, and that flicker perception (or the decrease in image quality caused by flicker) might be a problem specific to wide-FOV display systems. On the other hand, as mentioned above, hold-type displays have perceptible motion blur. We should optimize the trade-off between perceptions of flicker and motion blur at a certain frame frequency. Furthermore, we believe the best way to resolve this trade-off is to increase the system frame frequency. It will thus be essential to determine an adequate system frame frequency for wide FOV systems.

Motion blur and stroboscopic effect [3]

Motion blurs at various aperture times and object speeds were evaluated. One of the scenes that give viewers a very unnatural feeling is when the perceived sharpness of an object changes significantly between conditions of stillness and motion. To emulate such a scene, the test involved sliding a still picture into the field of view, stopping it at the centre for three seconds, and then sliding it out (Fig. 10). The test materials are shown in Fig. 11. The sliding speed was varied up to 32 arc-degrees/second to match the characteristics of camera panning in actual programmes. The moving part of the test materials was subject to a moving average filter so that accumulation at the camera aperture of the imaging device or at the retina in the display aperture would be simulated. The viewers evaluated the sharpness of the moving part in comparison with the still part as a reference on a 5-grade impairment scale.

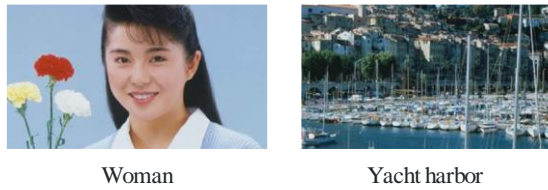
The results are shown in Fig. 12. The horizontal axis represents the number of the pixels of the filter that operated on the moving part. The three lines are for object speeds of 32, 16, and 8 arc-degrees/second, respectively. The evaluation for long filters becomes less severe as the object speed increases. 2~3 pixels/frame gives the perceptible limit, and 6~11 pixels/s gives the acceptable limit in Fig. 12a. 1~2 pixels/frame gives the perceptible limit, and 6~11 pixels/s gives the acceptable limit in Fig. 12b.

FIGURE 10
Sequence of evaluation stimuli



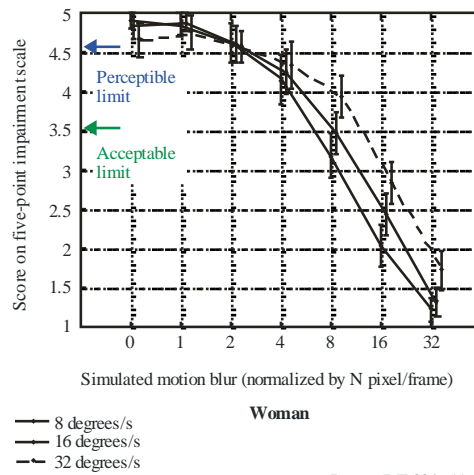
Report BT.2246-09

FIGURE 11
Test materials for blur evaluation



Report BT.2246-10

FIGURE 12a
Results of subjective evaluation of motion blur



Report BT.2246-11a

FIGURE 12b
Results of subjective evaluation of motion blur

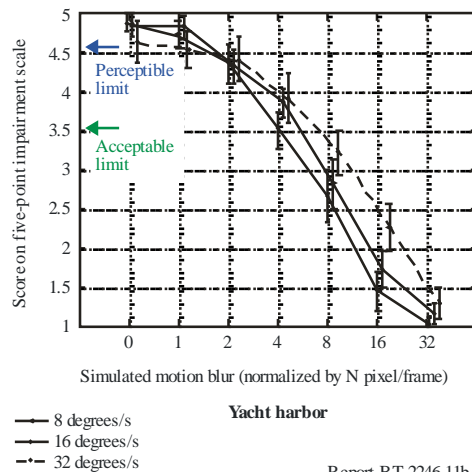
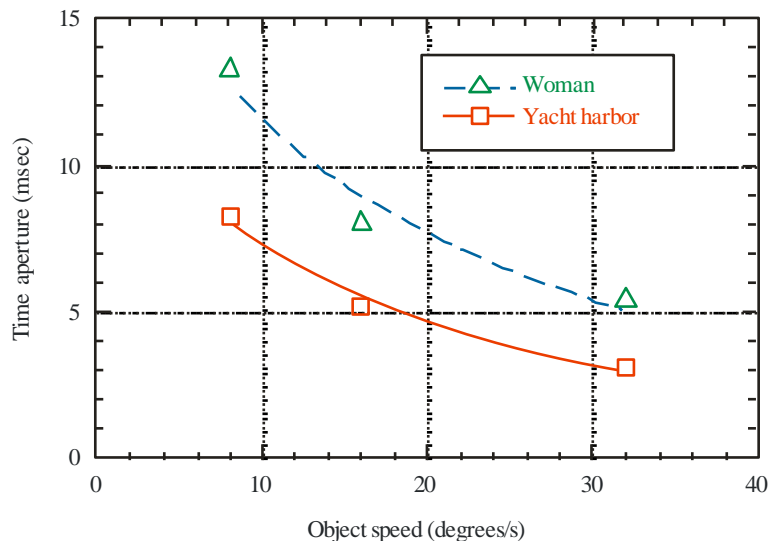


Figure 12c shows the relationship between time aperture and object speed giving an acceptable level of motion blur derived for the data in Fig. 12a and 12b [6C/456].

FIGURE 12c
Relationship between time aperture and object speed giving an acceptable level of motion blur



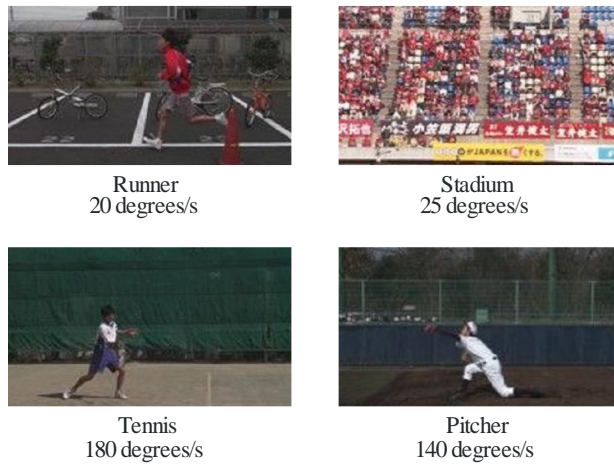
Stroboscopic effects at various frame frequencies and aperture times were evaluated. It is known that a camera shutter causes a stroboscopic effect. The relationship between the effect and temporal sampling parameters has been studied. This stroboscopic effect is sometimes called jerkiness. It may be referred to as “multiple images” that appear when the frame frequency is high relative to the reciprocal of the “time constant” of vision. In other words, the stroboscopic effect arises when the image of the object that is not or cannot be tracked by the eye moves discretely across the retina. The stroboscopic effect frequently appears in sports programmes. Therefore, the test materials were chosen from such a genre and were taken by a high-speed HDTV camera with various temporal parameter values (Fig. 13). The object speed was varied from 20 to 170 arc-degrees/second.

Figure 14a shows the results for sequences with a constant aperture time of 1/240 second and different frame frequencies. There was a degradation of 1 rank at 120 Hz/50%, 1.5 ~ 2 ranks at 80 Hz/33%, and 2 ~ 2.5 ranks at 60 Hz/25% compared with the reference at 240 Hz/100%. Figure 14b shows the relationship between image quality score and frame frequency for stroboscopic effect derived for the data in Fig. 13 (see Doc. 6C/456).

Figure 15 shows the results for a constant frame frequency of 240 Hz and different aperture times. It shows that the stroboscopic effect appears even at a frame frequency of 240 Hz for high-speed objects, whereas it is not apparent for low-speed objects.

FIGURE 13

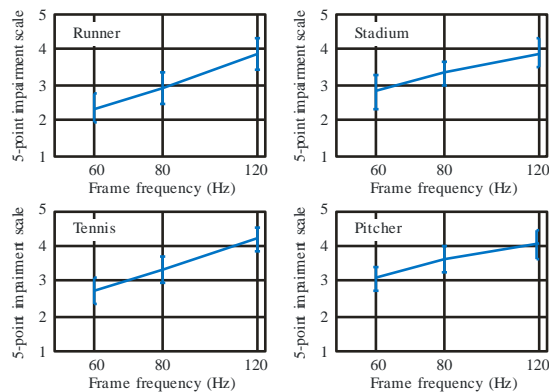
Test materials for stroboscopic effect evaluation²



Report BT.2246-12

FIGURE 14a

Results of subjective evaluation of stroboscopic effect, constant aperture time: 1/240 s

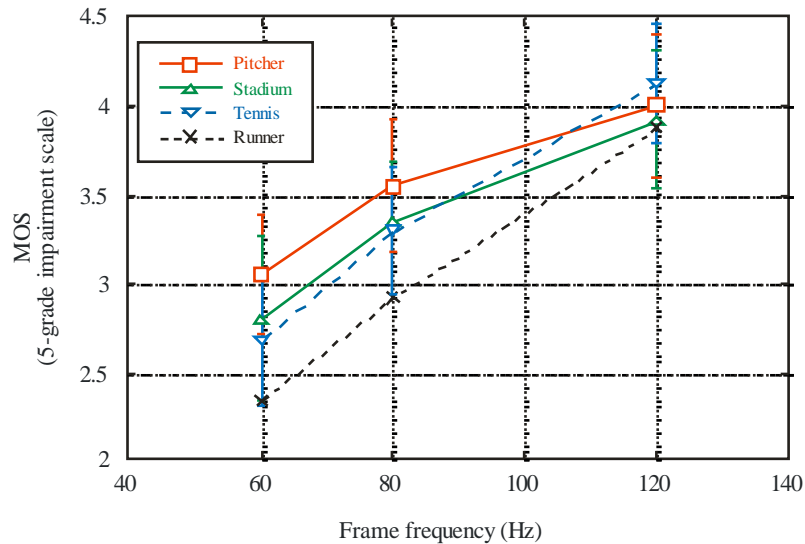


Report BT.2246-13a

² They were captured using an HDTV high-speed camera with a frame frequency of up to 240 Hz and various shutter speeds. The maximum object speed is noted.

FIGURE 14b

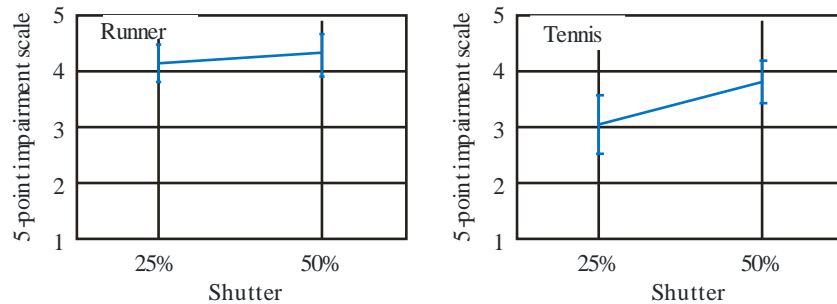
Relationship between image quality score and frame frequency for stroboscopic effect



Report BT.2246-13b

FIGURE 15

Results of subjective evaluation of stroboscopic effect, constant frame frequency: 240 Hz



Report BT.2246-14

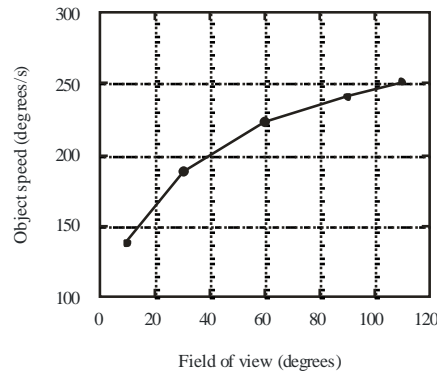
Moving pixels per aperture time can be used as an index for motion blur, and moving pixels per frame time can be used as an index for the stroboscopic effect. The indices of 1.2 pixels/frame and 6 pixels/frame give the perceptible limit and acceptable limit for motion blur, respectively. They correspond to the accumulation times of 1/1600 s and 1/320 s, respectively. These values seem slightly more demanding than the previously reported results do. The reason can be found in the comparison procedure used. Our procedure may make it easier than the previous ones to perceive differences. Our results suggest that the stroboscopic effect might not be perceived when the object motion is below 5 to 10 pixels/frame. The difference in the perceptible limits between motion blur and the stroboscopic effect may be attributed to the difference between fovea and peripheral viewing.

The index of pixel/frame is determined from the frame frequency, accumulation time, and object speed. Therefore, how a fast-moving object is dealt with is an important consideration in the system design. Data obtained from analyses of broadcasting programmes would be useful for studies aimed at improvement of current systems. On the other hand, a study aimed at development of a new system should make clear what kinds of objects are dealt with and how are they expressed on the screen.

In addition, it should be taken into account that another picture parameter affects the human visual system, and hence the requirement for motion portrayal changes. It has been pointed out that dynamic visual acuity (DVA) is affected by the field of view (FOV) through which the moving target is

presented [4]. Evidence for this effect is shown in Fig. 16. The vertical axis in the Figure is the object speed above which viewers could not recognize the object's orientation. DVA, in this case the viewer's ability to resolve an orientation of a moving target (Landolt C), increases by a factor of 1.8 as the FOV increases from 10 to 110 arc degrees.

FIGURE 16

Measurements of DVA

Report BT.2246-15

The motion portrayal of television depends on its scanning scheme and temporal sampling parameters. The parameters comprise the frame frequency and aperture time. Aperture time has not been explicitly included in television signal standards, but general program productions have been done with an aperture ratio of 100% at a camera and virtually 0% at a display because of technical constraints. The aperture ratio is thus one of the most important aspects in envisaging the limits of motion portrayal and determining the temporal sampling parameters of new high-quality television systems beyond the level of HDTV. A CRT will not be the only display device for such a future system. Inevitably, a hold-type display should be considered.

The above subjective tests were conducted with test materials similar to those used in daily broadcasting with the aim of establishing guiding principles for the design of future high-quality television systems. The results presented here can be taken as guiding principles when designing new television image formats and developing cameras and displays.

3.2.4.4 Proposed higher frame frequency

The UHD TV image format proposal suggests a frame frequency of 120/1.001 or 120 Hz as a higher frame frequency for UHD TV for the following reasons:

- The frame frequencies of 120/1.001 and 120 Hz eliminate the flicker artefact completely. This means that the shorter the light emitting time of displays during the frame period, the more the motion blur caused by the hold-mechanism is reduced without introducing flicker. The frame frequency at the camera end keeps the stroboscopic effect within the acceptable limit. The use of an electronic shutter of approximately a half to one third keeps motion blur within the acceptable limit. These results show that the frame frequencies of 120/1.001 or 120 Hz is required for a worthwhile improvement in motion image quality, although higher frame frequencies can yield further improvements for more demanding criteria.

- Some UHDTV applications may require the same frame frequencies as those of HDTV, which are up to 60/1.001 and 60 Hz. In defining frame frequencies higher than 60 Hz for UHDTV, it would be reasonable to propose a unique value. From purely a technical perspective it is desirable to adopt the least common multiple (LCM) frequency of the existing television systems frame frequencies for easy conversion. But the value of 300 or 600 Hz does not seem reasonable from the standpoint of balancing achievable performance with technical difficulty and the required bandwidth in the foreseeable future. Consequently, a conversion from/to 120/1.001 or 120 Hz to/from 50 Hz or 59.94 Hz would have to be made. These frame conversion problems could be solved by using modern sophisticated digital technologies.
- The feasibility of the camera, display, and transmission technologies needs to be considered before setting a higher frame frequency. Recent progress would indicate that a UHDTV system with a frame frequency of 120/1.001 and 120 Hz will be feasible sometime in the near future.

It is proposed that the inclusion of a frame frequency of 100 Hz in Recommendation ITU-R BT.2020 for following reasons:

- Legacy 50 Hz lighting systems at locations where the broadcaster does not have control of the lighting type used can interact adversely with cameras set to acquire at 120/1.001 or 120 fps. This can result in visible artefacts effects (see § 3.2.4.6).
- Even if the artefacts are not perceptible, it may nevertheless result in a significant reduction in coding efficiency for delivery (see § 3.2.4.6).
- Simple temporal conversion from 120 fps to both 25 fps and to 50 fps HDTV does not produce acceptable motion or image quality (see § 3.2.4.7).

Further study is needed on the issues below:

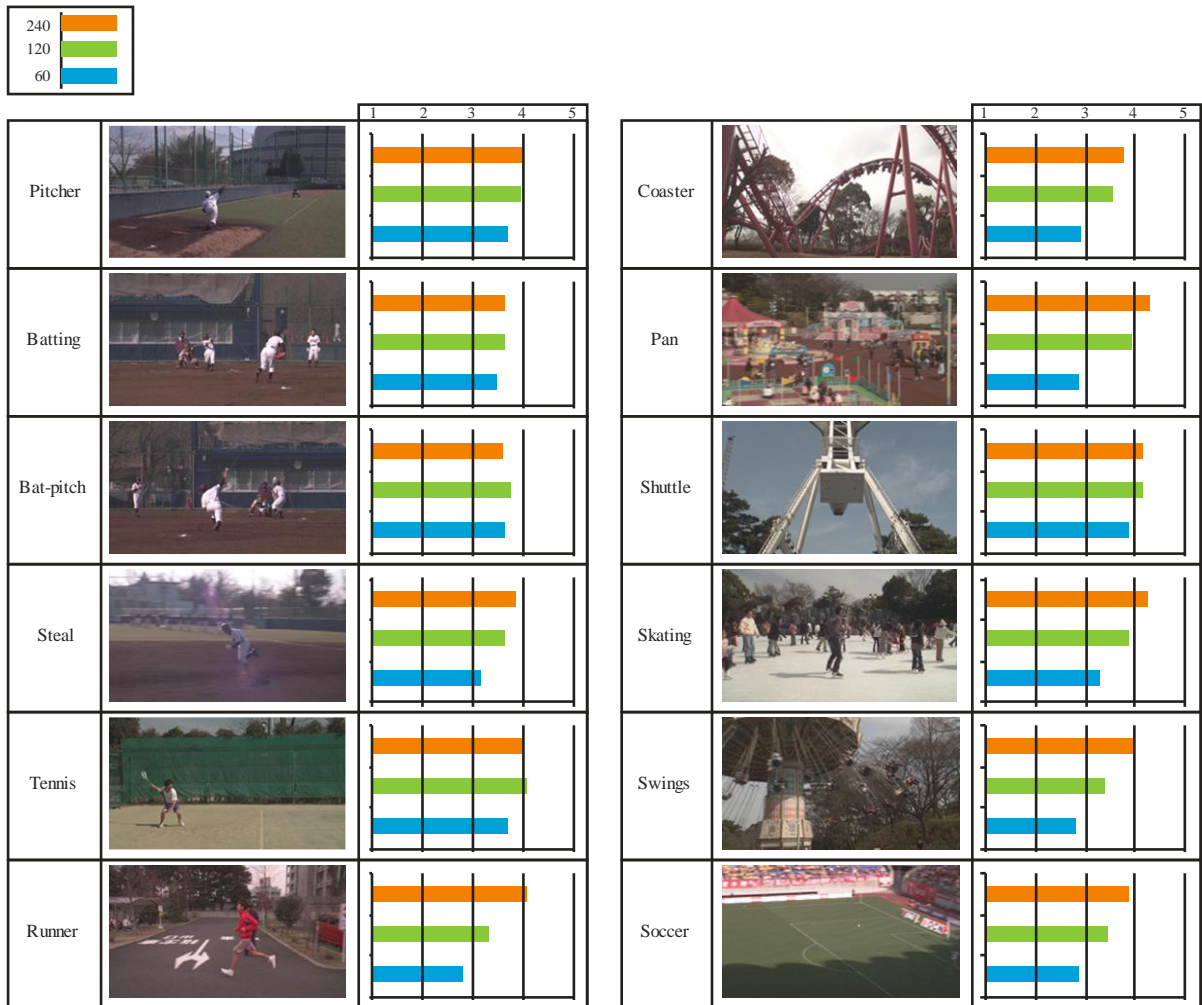
- Technical solutions for the lighting problem (see § 3.2.4.6).
- Review on the signal conversion issue considering the existing high-performance conversion method (see § 3.2.4.7).

3.2.4.5 Subjective evaluation of moving picture quality [Input, 47]

A recent paper has reported the subjective evaluation results of moving picture quality on a large screen [2]. An HDTV high-speed camera and projector that operate up to 240 Hz were used. Common images for television programmes were used as the test materials. The set of test materials included sports scenes, such as football, baseball, tennis, and running, and several scenes taken at an amusement park. Test conditions are listed in Table 3. Adjectival categorical judgement with a five-grade quality scale was used. Figure 17 shows the evaluation scores converted to an interval scale for each test material. The results show that the scores increase as the frame frequency becomes higher while how much the picture quality improves differs from material to material. The range of score improvement between 60 and 120 Hz extends from 0.14 to 1.04. The average scores are shown in Fig. 18. The improvement in scores between 60 and 120 Hz, and 120 and 240 Hz are 0.46 and 0.23 respectively. Both differences are statistically significant. It is anticipated that the increase of frame frequency will be more effective for UHDTV that has much more resolution than HDTV. These results lead to the conclusion that a frame frequency of at least 120 Hz is desirable for UHDTV to improve its moving picture quality.

FIGURE 17

Results of subjective evaluations of moving picture quality at high frame frequencies

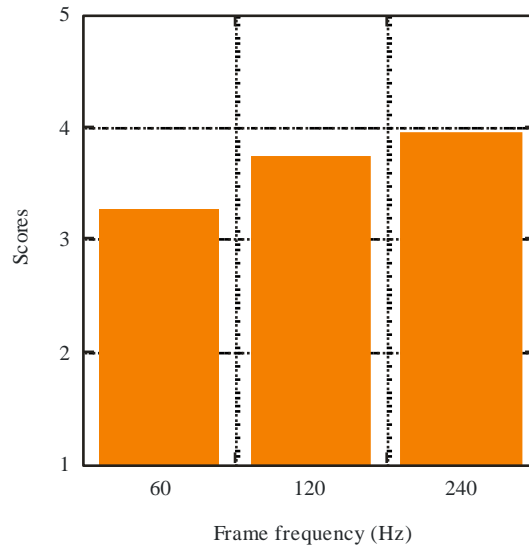


Report BT.2246-16

TABLE 4
Test conditions

Resolution	1 920 × 1 080
Scanning	60, 120, 240 Hz, progressive
Screen size	100 inches
Viewing condition	Recommendations ITU-R BT.500, ITU-R BT.710
Viewing distance	3.7 m
Participants	69

FIGURE 18
Average score



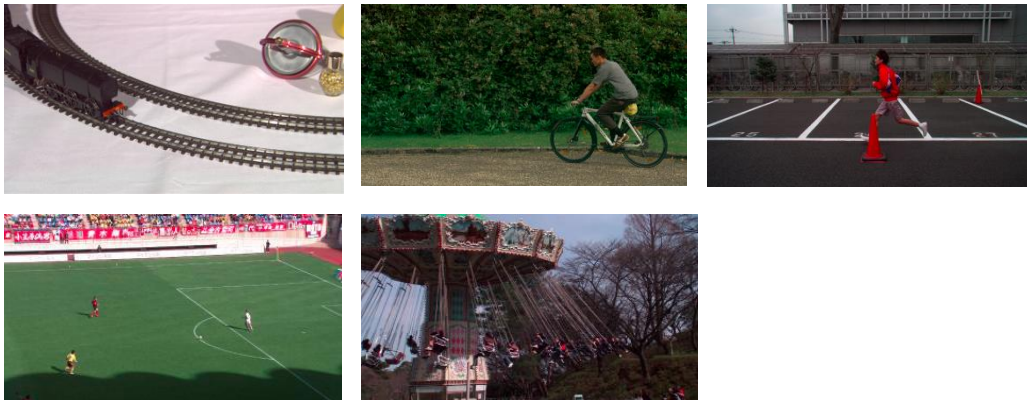
Report BT.2246-17

BTF³ assessed the quality gain provided by higher frame frequencies compared to conventional ones. Test conditions, materials and results are shown below:

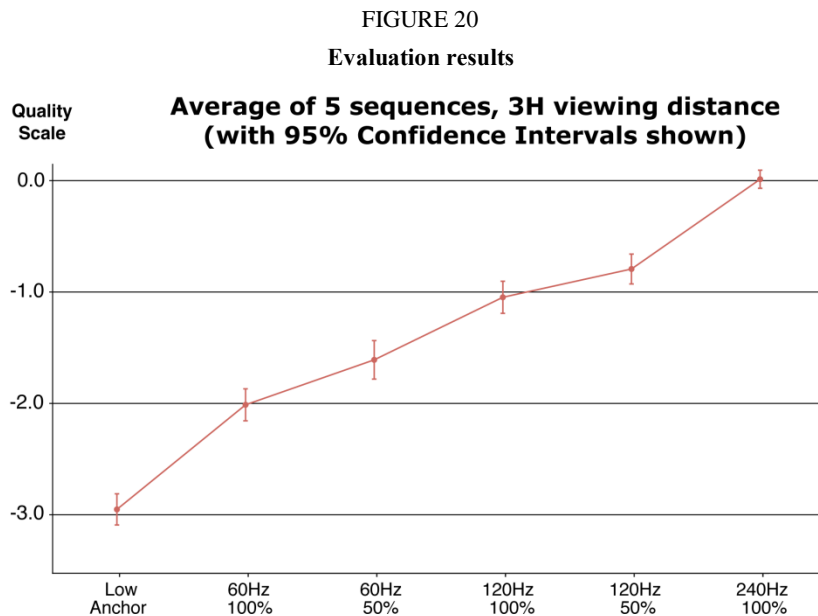
TABLE 5
Test conditions

Resolution	1 920 × 1 080
Scanning	60, 120, 240 Hz, progressive
Screen size	55 inches
Viewing distance	3 H
Participants	26

FIGURE 19
Test materials



³ Broadcast technology future group consisted of BBC, EBU, IRT, NHK, RAI, etc.



The results are summarized as follows:

- The sequences (all containing motion) gained benefit from increased frame rate (increase from 60 to 120 as well as 120 to 240).
- Between 60 Hz and 120 Hz at least one grade of difference (which is clearly perceptible) is observed.
- The interpolated results suggest that 100 Hz might be a better cost/value trade off than 150 Hz.
- The impact of a shutter is perceptible, however it is not conclusive.

3.2.4.6 Image acquisition of the scene illuminated by flickering lights at high frame frequency

Document 6C/152 which was contributed from the United Kingdom to the April 2013 meeting of Working Party 6C raised the issues of flicker when cameras capturing at 120 frames per second are used under 50 Hz lighting.

Document 6C/201 reported the study by BBC R&D on the phenomenon, namely the impact of 50 Hz based lighting on 120 fps acquisition. It carried out controlled tests using a variety of commonly used luminaries and replicate situations where the lighting sources may appear in vision or are featured in vision (practical lighting). The results show that the test sequences shot at 120 fps and various shutter speeds contained visible flicker (an example is shown in Fig. 21) while the ones shot at 100 fps contained no visible flicker. It also suggests that a compensation technique for the brightness variations will cause variations in noise performance between frames which is likely to cause both visually disturbing and impaired down-stream bit rate reduction processing. Document 6C/216 provides data for the impact of this issue on emission encoding.

Document 6C/211 reported the study results on the same subject done by NHK on the image acquisition on the scene illuminated by flickering lights. It measures the flickering characteristics of lighting equipment for programme production (an example is shown in Fig. 22) as well as the waveform of the captured image at 120 fps. It defines the required condition as *the flickering frequency of the light must be equal to or higher than the frame frequency in order not to generate the flicker artefact by the flickering lights with any temporal characteristics*. It also mentions the trend of lighting equipment which is shifting to higher flickering frequency than power frequency (Fig. 23) as well as the flicker compensation techniques.

FIGURE 21

Example of acquired image waveform: Practical Fluorescent (Light box) 120 fps with 1/150 s shutter

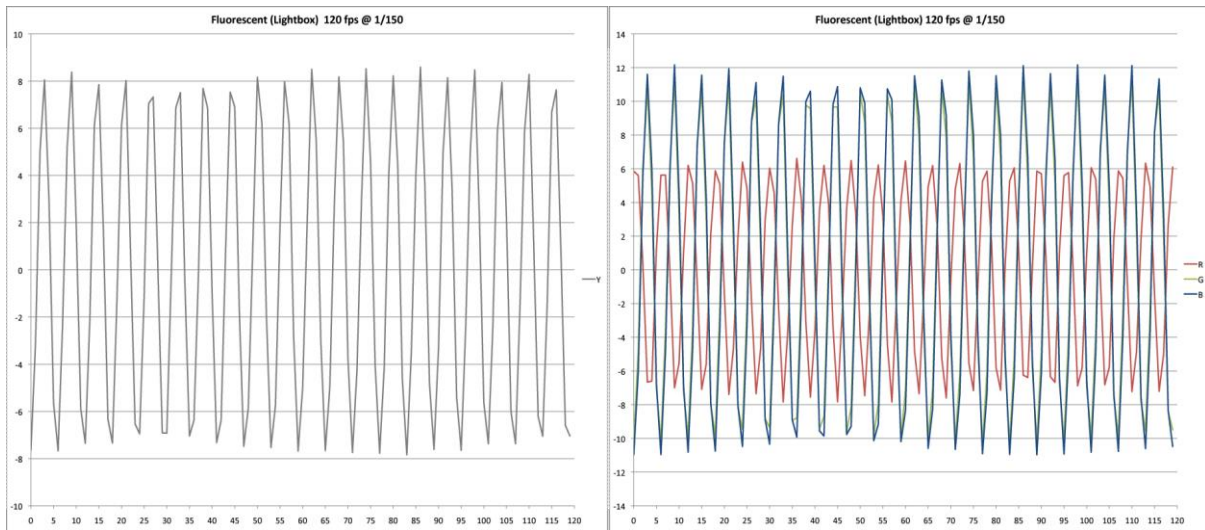
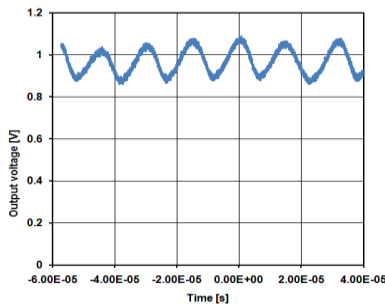
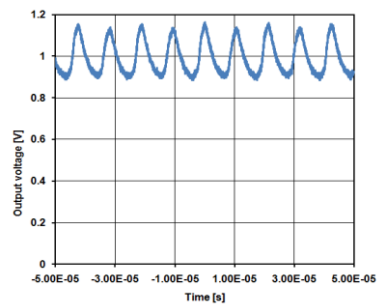


FIGURE 22

Example of flickering light: fluorescent lights

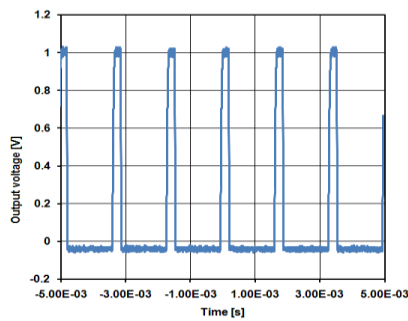
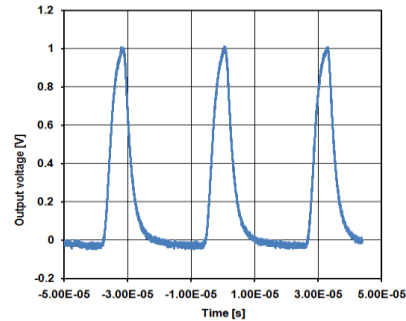
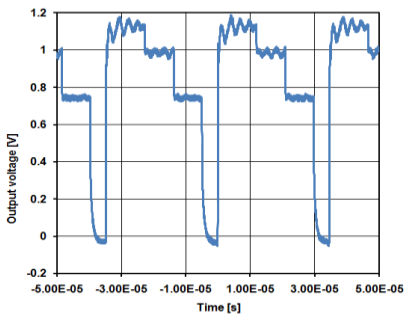
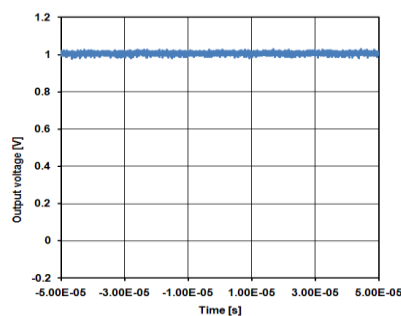


Fluorescent 55W : 60kHz



Fluorescent 36W : 100kHz

FIGURE 23

Example of flickering light: LED lights**LED (FLF-554-D-T1 :50%) : 600Hz****LED (AL-LED-FS-10 :50%) : 30kHz****LED (LC-7) : 30kHz****LED (KLA-1 x 2) : DC****3.2.4.7 Temporal signal conversion**

Temporal signal conversion was discussed as a factor in the selection of frame frequency.

Document 6C/201 provided the experimental data on the temporal conversion techniques which employ nearest frame, linear, and 11-tap FIR filter approaches. It concludes that all these approaches are not sufficient in terms of motion picture quality as well as the delay in the FIR filter.

Document 6C/211 points out that existing professional signal converters for exchanging HDTV programmes between 50 Hz and 60 Hz regions use motion-compensated frame interpolation technology which provides much better picture quality. The delay time of those devices are only 2 to 3 frames.

3.2.5 Non-linear coding function

The non-linear coding function, or gamma pre-correction, of video signals has two roles. One is to raise the code-usage efficiency by making the perceptual quantization error among the video codes more uniform (see chapter 3 for a description). This is similar to the one in analogue video to reduce the perceptual effects of noise on the transmission path.

The other role is to pre-compensate for the non-linear electro-optical transfer function (EOTF) of displays. The combination of the gamma pre-correction and the gamma characteristics of the display device yields the total transfer function, or end-to-end gamma, from the scene being shot to the reproduced image on the screen. It is well known that the end-to-end gamma is not necessarily unity; i.e. the gamma value is not necessarily 1. For example, some people believe that values of 1, 1.25, and 1.5 are appropriate for bright, dim, and dark surrounding environments, respectively [6]. Furthermore, the end-to-end characteristics are more complex than a simple power function because the optical-electro transfer function (OETF) at the acquisition end consists of a power-function segment and a linear segment. The current tone curve is analysed below with this point in mind.

[Input, 52], Proposal of required electro-optical transfer function for a FPD as a master monitor, describes the measured EOTFs of current HDTV reference monitors. According to that document, the EOTF of a CRT can be represented by equation (1).

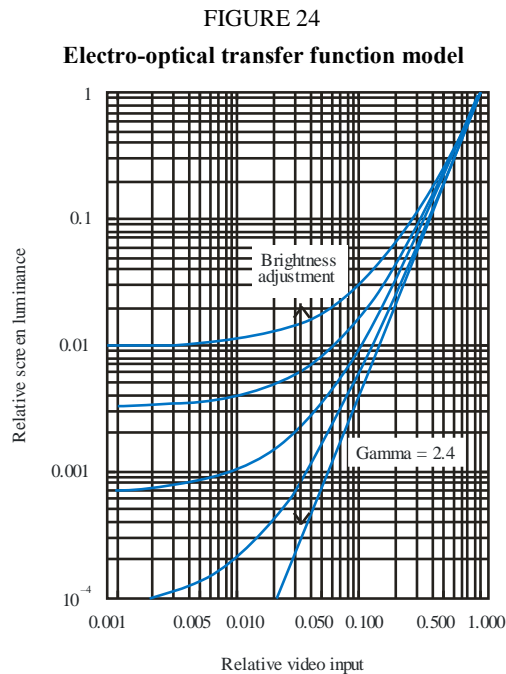
$$L = k[V + b]^\alpha + \beta \quad (1)$$

where:

- L : Screen luminance (normalized)
- V : Input video signal level (normalized)
- α : Inherent gamma of the device
- β : Luminance with no signal input
- b : Variable for brightness control
- k : Coefficient for normalization

$$[x] = x \text{ when } x \geq 0, [x] = 0 \text{ when } x < 0$$

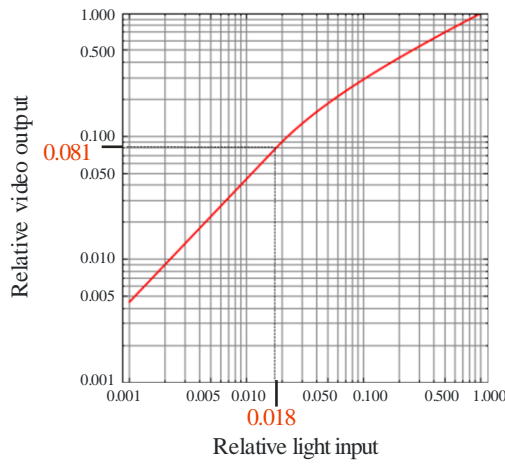
The variable “ b ” represents the brightness control, and it is adjusted to conform to the viewing environment by using the procedure described in Recommendation ITU-R BT.814. Figure 24 plots five EOTFs with different brightness control settings.



Report BT.2246-18

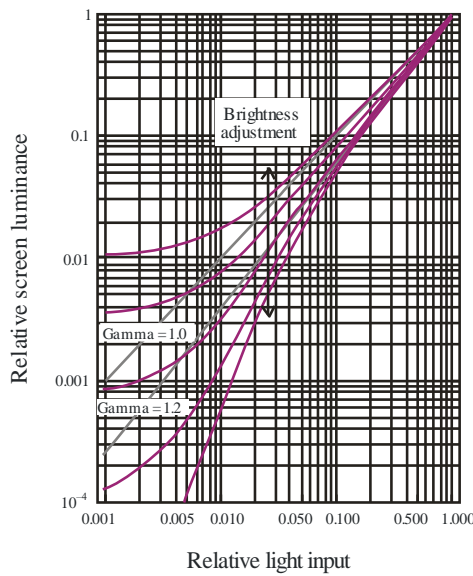
The gamma pre-correction described in Recommendation ITU-R BT.709 is applied to HDTV signals (Fig. 18). Modifications for artistic reasons will not be considered here. Calculated end-to-end characteristics are described in Fig. 26. Figure 26 illustrates that the brightness adjustment according to the viewing environment makes the end-to-end gamma vary so as to match with the visual perception predicted by the colour appearance model. Namely, increasing the brightness in bright-surroundings makes the end-to-end gamma close to unity, whereas decreasing the brightness in dark surroundings makes it greater than unity. The end-to-end gamma is an intermediate between these two values in dimly lit surroundings.

FIGURE 25
Optical-electro transfer function of Recommendation ITU-R BT.709



Report BT.2246-19

FIGURE 26
End-to-end transfer function: OETF with linear segment



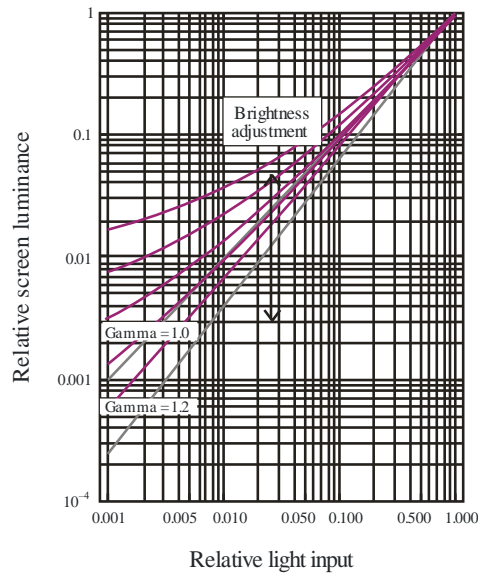
Report BT.2246-20

Figure 27 plots cases in which the OETF is a simple power function without a linear segment. A comparison between Figs 26 and 27 shows that the linear segment of the BT.709 OETF has a significant effect at low luminance.

The above discussion has illustrated an ingenious mechanism of the combination of OETF of Recommendation ITU-R BT.709 and the EOTF of CRTs. It unintentionally makes the end-to-end tone curve quite optimal.

FIGURE 27

End-to-end transfer function: OETF without linear segment



Report BT.2246-21

3.2.6 Colorimetry

The first steps to determine the UHD TV colorimetry should address the following questions.

- What are the colour rendition requirements?
- Are there other colorimetry requirements?
- What method and parameter values should be chosen in light of those requirements?

One of the points to be discussed is whether the colorimetry should be based on the display tri-colours or not. Expansion of the colour gamut will enable UHD TV to provide a better visual experience to viewers. Table 7 classifies the methods to expand the colour gamut.

One proposal [Input, 10] is that UHD TV should cover the entire real surface colours with good efficiency.

The chromaticity coordinates of primaries and reference white are proposed in [Input, 16] and [Input, 18]. They are listed in Table 6.

TABLE 6

Proposed sets of primaries and reference white

Parameters	Values			Wavelength (nm)
		x	y	
Chromaticity coordinates (CIE 1931 xy) of reference primaries and reference white	Redprimary	0.7006(a)	0.2993(a)	625
		0.7140(b)	0.2859(b)	635
	Greenprimary	0.1625(a)	0.8012(a)	531
		0.1702(b)	0.7965(b)	532
	Blueprimary	0.1314(a)(b)	0.0459(a)(b)	467
Reference white (D65)	0.3127(a)(b)	0.3290(a)(b)	–	

Sets (a) and (b) are proposed in [Input, 18] (Korea) and [Input, 16] (Japan), respectively.

Both proposals are based on the following concept.

- The colour gamut for UHDTV should be expanded by using wider colour primaries than conventional ones so that it covers real surface colours as much as possible.
- Those primaries should be real colours.

As a result, both proposals select monochromatic colours as RGB primaries. The arguments are summarized as follows, in light of both proposals on the derivation of the primaries.

- 531 nm for green and 625 nm for red are on the equi-hue lines of the current LED-backlit LCD and AM-OLED. However, 635 nm is far located from the red primaries of the current LCD and AM-OLED technology.
- 532 nm for green and 635 nm for red are easy to realize with the current laser technology. On the other hand, at present it is very difficult to manufacture 625 nm and 531 nm LDs.
- 635 nm for red has less energy efficiency because human visual sensitivity is lower at 635 nm than at 625 nm.

3.2.6.1 Examination of two sets of RGB primaries [Input, 24]

Introduction

Two RGB-primary sets were introduced in WD-PDNR ITU-R BT.[Image-UHDTV] [Input, 19]. Table 6 describes their chromaticity coordinates in the CIE xy domain and their corresponding wavelength (nm) on the spectral locus. Figure 28 shows the two RGB-primary sets and their constant hue lines which were computed using the CIELAB hue. The largest difference can be found in the red primary position. The 625 nm position was chosen in order to encompass efficiently most of the real-world surface colours and also by taking into account the characteristics of flat panel displays such as AMOLED and LCD on which UHDTV programmes will be presented. (see Attachment 1 in [Input, 19]). The 635 nm position was selected by considering laser display technology (see Attachment 2 in [Input, 19]).

FIGURE 28

The two sets of RGB primaries introduced in Table 6 and their constant hue lines in the CIE xy coordinates

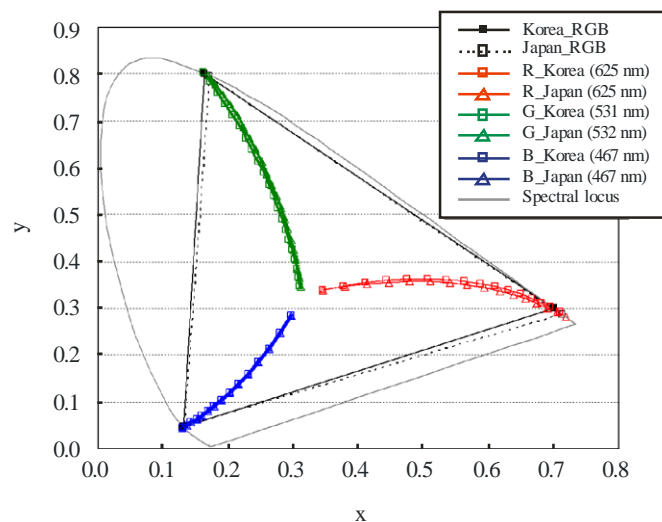


TABLE 7

Classification of methods to expand the colour gamut of image systems

Approach	Name	Description	Performance	Implementation and compatibility
Display referred approach	Negative RGB	Use negative or more than unity values for RGB while using conventional RGB primaries e.g. Recommendation ITU-R BT.1361, xvYCC	Considerably wide, but not all in some case, part of the area within colour locus will be covered Optimal code usage efficiency may be difficult	Easy implementation of relatively small circuitry in mixed environment with current systems Some standard or guideline that describes a real colour gamut expanded display may still be needed* Incompatibility with positive-value-based signal processing
	Wider RGB (real colour)	Use widened RGB primaries comprising real colours e.g. Adobe RGB	Considerably wide, but not all, part of the area within colour locus will be covered Code usage efficiency is expected to be high	No conversion circuitry is required after display compliant with the standard is introduced
	Wider RGB (imaginary colour)	Use widened RGB primaries including imaginary colour(s) e.g. [Input, 10] (Korea)	Flexible choice as to cover the target gamut (e.g. real surface colours) Small efficiency drop due to the coverage of outside part of colour locus	Some conversion circuitry is needed at display end * See above
	XYZ based	Use XYZ colorimetry and converted primaries for transmission if needed e.g. [Input, 53] (USA), SMPTE 428-1-2006	Any signal falling onto colour locus is covered Code usage efficiency is not expected to be high	Some conversion circuitry is needed at display end * See above
Scene referred approach	Multiple (more than three) primaries	Use more than three primaries, spectral distribution reproduction ultimately	Environment-independent reproduction	Completely different system

The influence of changes in red primary

To determine the best red-primary position, three unnoticeable areas around three red constant-hue lines of 625, 630 and 635 nm were compared. Figure 29 illustrates these three unnoticeable areas that can be formed by three pairs of red, blue and green lines corresponding to 625, 630 and 635 nm respectively. The colours located in each of three unnoticeable areas fall into the colour difference range of $\Delta E = \pm 2$ (in CIELAB) against each of three constant hue lines. It is generally known that if the colour difference between two stimuli is less than about 2, the two stimuli might not appear to have different colour. As mentioned in the previous section, 625 nm was chosen by considering the characteristics of flat panel displays such as AMOLED and LCD whereas 635 nm was selected by taking into account laser display technology. The unnoticeable area for the red primary of 630 nm (the area generated by two blue lines in Fig. 29) is seen to overlap largely with the two unnoticeable areas for two red primaries of 625 and 635 nm. This suggests that the red primary of 630 nm can reflect the colour characteristics of both types of display – flat panel displays and laser displays. Thus, the red primary of 630 nm can be the best choice amongst 625, 630 and 635 nm.

It can be presumed based on the above investigation that the extent and efficiency in the inclusion of real surface colours are insignificantly affected by changing red-primary position within 625-635 nm. The four candidates of RGB primaries were therefore selected and were evaluated in terms of gamut coverage and gamut efficiency. Table 8 explains the evaluation results for the four chosen candidates. In the computation of the gamut-coverage and -efficiency, the reference gamut was firstly established from the real surface colours (Pointer and SOCS database) and the reproducible colours by the RGB primaries of AMOLED [7], LCD [7], Digital-Cinema Reference Projector [7], and three standard colour spaces (Recommendation ITU-R BT.709 [9], Adobe RGB [10], and NTSC [11]). Details for computing the gamut-coverage and -efficiency were described in Attachment 1 of WD-PDNR ITU-R BT.[Image-UHDTV] [Input, 19]. The calculated results of gamut coverage/efficiency in Table 8 suggest that there is little difference for all four candidates in the aspect of coverage of the colours falling onto the reference gamut.

FIGURE 29

Three unnoticeable areas around three red constant-hue lines of 625, 630 and 635 nm in the CIE xy chromaticity diagram

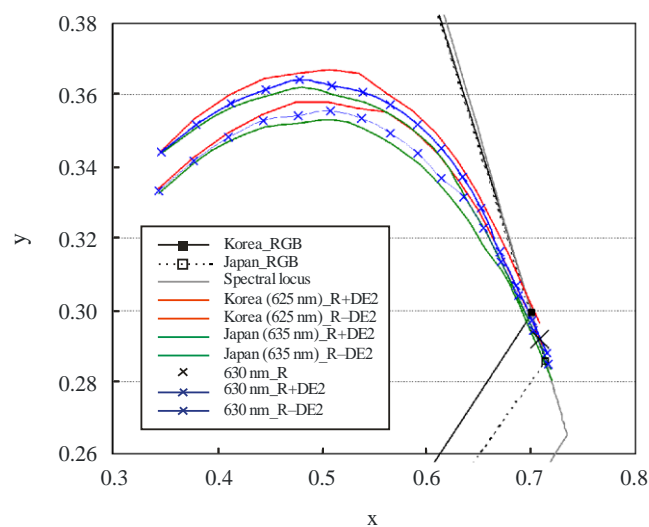


TABLE 8
Four candidates of RGB primaries for UHDTV systems

Primary	Korean Proposal (Attachment 1 in WD-PDNR ITU-R BT.[Image-UHDTV], [Input, 19])	Japanese Proposal (Attachment 2 in WD-PDNR ITU-R BT.[Image-UHDTV], [Input, 19])	Modification 1	Modification 2
Red	0.7006, 0.2993 (625 nm)	0.7140, 0.2859 (635 nm)	0.7079, 0.2920 (630 nm)	
Green	0.1625, 0.8012 (531 nm)	0.1702, 0.7965 (532 nm)	531 nm from Korean proposal	532 nm from Japanese proposal
Blue	0.1314, 0.0459 (467 nm)			
Reference White (D65)	0.3127, 0.3290			
Gamut Coverage	96%	95%	95%	95%
Gamut Efficiency	91%	90%	90%	90%

Conclusions

Two RGB-primary sets for UHDTV systems were proposed by Korean and Japanese administrations during the previous ITU-R WP 6C meeting in November, 2009 and were described in WD-PDNR ITU-R BT.[Image-UHDTV] [Input, 19]. A single suggestion was made for blue primary which was placed at 467 nm on the spectral locus. The difference in the two suggested green-primaries was insignificant, i.e. 1 nm difference occurring between 531 nm by Korea and 532 nm by Japan. The main difference in the two proposals was the location of red primary: 625 nm by Korea and 635 nm by Japan. The influence of changes in red-primary position was thus examined within 625-635 nm. The red primary of 630 nm was found to be the best choice, because this red primary can reflect colorimetric characteristics of both flat panel displays and laser displays. Two new sets of RGB primaries were then formed and evaluated in terms of gamut-coverage and -efficiency. These were constituted of the same red (630 nm) and blue (467 nm) primaries, but different green primaries (531 and 532 nm). Both two sets performed equally in covering the reference data set that were composed of real surface colours, the reproducible colours of flat panel displays such as AMOLED and LCD and Digital Cinema Reference Projector, and standard colour spaces of Adobe RGB, NTSC and Recommendation ITU-R BT.709. In conclusion, either new RGB-primary set (R 630 nm, G 531 nm, B 467 nm or R 630 nm, G 532 nm, B 467 nm) can be recommended for UHDTV systems.

3.2.6.2 Capture and display experiments of wide colour gamut images [Input, 28]

Objects of shooting

Red and yellow flowers were selected as highly-saturated natural objects commonly used in TV programmes. For blue and cyan natural objects, butterflies were selected. A red model car and stained glass tableau containing green as artificial colours are also selected.

Capture

The objects were illuminated by xenon lamps that have the spectral energy distribution close to that of the sunlight in the daytime. The spectral energy distribution was adjusted by spectral correlation filters (SOLAX, SERIC Ltd., Japan). The stained glass tableau was backlit with a fluorescent light. The tristimulus values XYZ of the objects were measured with a CCD-based imaging colorimeter (PM-1400; Radiant Imaging, USA).

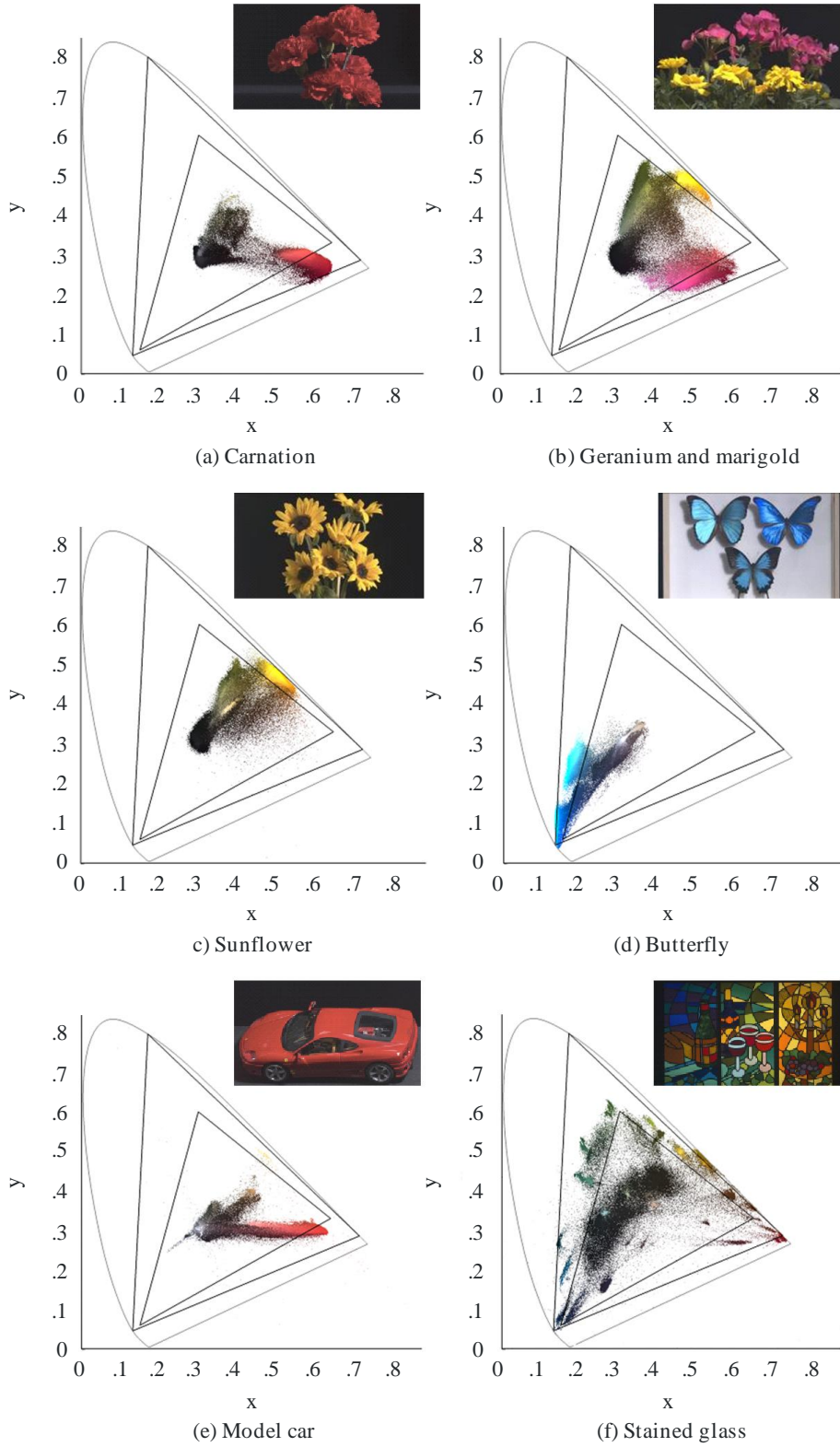
Display

The images data are converted to RGB values on the basis of a laser display (LaserVue, Mitsubishi, Japan) [12] with monochromatic RGB primaries ($\lambda_B = 447 \text{ nm}$, $\lambda_G = 532 \text{ nm}$, $\lambda_R = 640 \text{ nm}$, white point: D65). For comparison of the reproduced colours between UHDTV and HDTV, the RGB images were converted to ones the colour gamut of which were limited to that of the HDTV (Recommendation ITU-R BT.709).

FIGURE 30

Colour distribution of objects on the x-y chromaticity coordinates

(Inner triangle: HDTV primaries, Outer triangle: UHD TV primaries)



Results

Figure 30 shows the colour distribution of each object on the x-y chromaticity coordinates. It was found that some objects commonly used in TV programmes contain highly saturated colours that are not reproducible with the HDTV colorimetry, and that most colours are reproducible with the proposed colorimetry. Viewers perceived conspicuous visual differences between the wide-gamut images and ones the gamut of which is limited to that of HDTV and felt that wide-gamut images appeared closer to the real objects.

3.2.7 Colour encoding

3.2.7.1 General

The use of a colour coding in transmission that is different from the one in acquisition and display, e.g. $Y' C_B' C_R'$, can reduce the transmission bandwidth while maintaining the luminance resolution that the human visual system (HVS) is sensitive to. The suitability of this method for UHD TV should be examined as follows.

- What kind of primaries are suitable?
- What are the ratios for spatial sampling between those primaries, e.g. the ratio between luminance and colour-difference?
- What is the order of the transform matrix and non-linear conversion (Gamma correction)?

Conversion from RGB to luminance and colour difference signals was originally introduced when the colour television was developed. This was done while maintaining compatibility with black and white television. Later, it was used in the analogue component interface in a studio environment since it is less affected by error and noise. Then, when the interface was digitalized, the digital component of the signal format (Recommendation ITU-R BT.601) was standardized. The chroma sub-sample was introduced to reduce the total bit rate of the interface while taking advantage of human visual system characteristics. The major issue in the study was maintaining picture quality, when applied to chroma key processing in particular [13]. Those studies are described in detail in Reports ITU-R BT.629 – Digital coding colour television signals and ITU-R BT.962 – The filtering, sampling and multiplexing for digital encoding of colour television signals. Based on broad studies including those ones, “4:2:2” was standardized as the chroma sub-sample ratio for the digital component signal in Recommendation ITU-R BT.601. It was also used for HDTV and described in Recommendation ITU-R BT.709.

The chroma sub-sample ratio affects the required bit rate of digital interface the most. The serial digital interface (SDI) for HDTV started at 1.5 gigabits per second that was available at that time to transmit 4:2:2 signals [14][15]. Later, in response to the requirement for higher picture quality, 1.5 Gigabits-per-second dual-link system was standardized to transmit 4:4:4 signals [14][16]. Further, advances in technology have enabled 3 Gigabits-per-second SDI and a single link system to be standardized to transmit 4:4:4 signals [14][17][18].

Picture quality is a trade-off for the “cost” of the interface when the chroma sub-sample ratio varies. It is again required from the “definition” of UHD TV described in [Input, 20] that UHD TV should have picture quality no worse than that of HDTV. Higher picture quality is desirable even in a simple post-production⁴.

The interface seems slightly different from that used in HDTV where 1.5 Gigabits per second was the limit although it is true that lower bit rate is always preferable. This is because optical transmission technologies are going forward in the field of network technology, and those technologies are

⁴ The concept of simple/complex post-production is described in Recommendation ITU-R BT.1662 for LSDI application.

supposed to be applicable to the UHDTV interface. An optical transmission system was recently developed for the interface of UHDTV studio equipment [19]. It carries 72 Gigabits-per-second data that correspond to a $7\,680 \times 4\,320$ pixel-, 60 Hz-, 12-bit-, 4:4:4-signal.

The discussions related to this section during study period 2008-2011 are as follows.

It was agreed that the answer to the second question (chroma sub-sample ratio) is that 2:1 on one-dimension is acceptable as the past experimental results show. The results are reflected to the parameter values of 4:2:2 and 4:2:0.

The third question means solving the so-called “constant luminance issue”. This is discussed in § 3.2.7.2.

3.2.7.2 Analysis of sub-sampled versions of the UHDTV signal [Input, 58]

The discussion on the so-called “constant luminance issue” is summarized as follows. The detailed information is attached as an Attachment.

- 1) Sub-sampled versions of the UHDTV signals will be valuable for the production environment when reduced bandwidths are needed, and for the transmission, exchange, and for delivery environments where reduced bandwidths are essential taking into account very large amount of image data of UHDTV programmes.
- 2) Since the UHDTV formats use exclusively progressive scanning, it is only strictly necessary to specify a 4:2:0 version of the sub-sampled signal rather than a 4:2:2 version in the digital noise reduction (DNR). This provides balanced vertical and horizontal information. The work in the RG was carried out assuming a 4:2:0 sub-sampling structure. Nevertheless, a 4:2:2 version could also be included in the DNR if there is a demand, though it would not contribute to significantly higher quality results.
- 3) After a lengthy analysis, it is clear that there are potential advantages, technical and operational, for both non-constant luminance encoding and constant-luminance encoding.
- 4) Making formal subjective evaluations of the different approaches proved unlikely to be valuable, as the qualitative difference between the approaches for normal video content (natural pictures) would be too small to provide results for which the difference in mean scores with non-expert assessors falls outside the confidence intervals of the results. Essentially computer generated test pictures or test cards are needed to see differences of a half a grade or more. Nevertheless, for the quality-critical environment of UHDTV the differences are important, and they are distinguishable using objective methods, which have been largely used in the RG study reported in this document.
- 5) The current approach of “non constant luminance” encoding for sub-sampled signals may have some advantages in programme production, which include the following:
 - 5.1 Similar results would be obtained for colour mixing between R'G'B' and Y'C_B'C_R'.
 - 5.2 Production staff may find the degree of familiarity with HDTV practice comforting.
- 6) The new approach for “constant luminance” encoding for sub-sampled signals may have some advantages, which include the following:
 - 6.1 The original luminance information is less affected by sub-sampling and so the original detail and edge information is more accurately maintained.
 - 6.2 There will be higher compression efficiency for transmission and delivery that is beneficial to accommodate huge data size of UHDTV programmes. This is due in part to the higher PSNR of the input signal to the compressor. Additionally, there is also an apparent gain in the reduction of bit rates for the compression stage according to the objective analysis results.

- 7) The group has discussed the use in practice of the two approaches with a number of broadcasters in Korea, Japan, Europe, and the United States of America. The comments received have been balanced and supported both approaches. This reinforces the need to include both options in the DNR.
- 8) In a later section of the Report there is a discussion of the use of a conventional and an alternative implementation of constant luminance encoding.
- 9) The RG is convinced that the optimum DNR will include both colour encoding approaches as are included in Annex 14 to Document 6C/564 [Input, 54]. It is possible that after some years of experience, one or the other may become the dominantly used form.
 - 9.1 In summary, we can note that it is particularly valuable to include the constant luminance approach option for the sake of compression efficiency, and which may later even allow the use of 4:1:0 in some circumstances.
 - 9.2 In summary, we can note that it is particularly valuable to include the non-constant luminance approach option for the sake of the introduction of UHDTV into the production community that is familiar with HDTV production practice.

3.2.8 Bit depth

Bit depths of 8 and 10 have been empirically used for digital video signals of television applications. There is a related description in an early version of Recommendation ITU-R BT.709 as follows.

Studies are continuing in this area. There is agreement that at least 8 bits are required for R, G, B, Y, C1, C2 and that 10 bits will be required for some applications. Therefore, both 8-bit and 10-bit representations are required.

The later version of Recommendation ITU-R BT.709 recommends 8 and 10 as the bit depths, as is well known.

Recommendation ITU-R BT.1769 – Parameter values for an expanded hierarchy of LSDI image formats for production and international programme exchange, recommends bit depths of 10 and 12, in consideration of emerging requirements for higher picture quality.

Let us briefly consider what the bit depth should be for UHDTV. The quality requirement is that the discontinuity in tone reproduction should be below a certain criterion. The most demanding criterion would be the perceptible limit, and it depends on the image content and viewing conditions.

The contrast sensitivity function (CSF) of the human visual system (HVS) relates to the ability to detect discontinuities in tone reproduction. The Weber-Fechner law (equation 2) is a classical representation of contrast sensitivity.

$$\text{Minimum detectable contrast: } \Delta L/L = \text{const.} \quad (2)$$

where:

L : luminance

ΔL : minimum detectable difference in luminance.

According to this law, the minimum detectable contrast, i.e. the reciprocal of contrast sensitivity, is constant regardless luminance. It is believed that the ratio is between 1/50 and 1/100. However, it increases below and above certain luminances. Figure 31 shows measured data taken from Schreiber's book [20].

The actual CSF varies with the screen luminance, field of view, spatial frequency of the image, etc. Barten's model [21] takes such conditions into account. Figure 31 also shows the minimum detectable

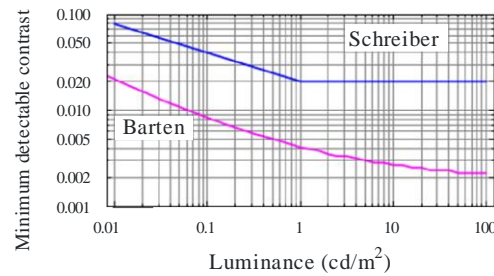
contrast calculated according to Barten's model with the assumed conditions⁵ for UHDTV applications. Note that the minimum detectable contrast is calculated by using the following equation.

$$\text{Minimum detectable contrast} = 1/\text{CSF}_{\text{Barten}} \times 2 \times 1/1.27 \quad (3)$$

Here, 2 is used for the conversion from modulation to contrast and 1/1.27 is used for the conversion from sinusoidal to rectangular waves [22].

FIGURE 31

Examples of HVS minimum detectable contrast characteristics



Report BT.2246-25

The bit depth and coding function should be determined by taking HVS characteristics into account so that the discontinuity in tone reproduction will be below a certain criterion. More specifically, the step (reproducible contrast) in the displayed luminance between consecutive codes of the coding function should be lower than the criterion, e.g. the minimum detectable contrast. Figure 32 shows the reproducible contrast of log and power coding functions. These functions are in fact the display EOTFs, i.e. the inverses of the coding functions, rather the coding functions themselves. The reproducible contrast is calculated with equation (2). The black level for the log function is set to $10E-6$ of the white level.

5 The following constants were used:

$$k = 3.0, T = 0.1 \text{ s}, \eta = 0.03, \sigma_0 = 0.5 \text{ arc min}, X_{\text{max}} = 12^\circ, \Phi_0 = 3 \times 10^{-8} \text{ s deg}^2, C_{ab} = 0.08 \text{ arcmin/mm}, N_{\text{max}} = 15 \text{ cycles}, u_0 = 7 \text{ cycles/deg}$$

where:

k : signal-to-noise ratio at 50% detection probability

T : integration time of the eye

η : quantum efficiency of the eye

σ_0 : standard deviation of the optical line-spread function

X_{max} : maximum integration area in the x direction

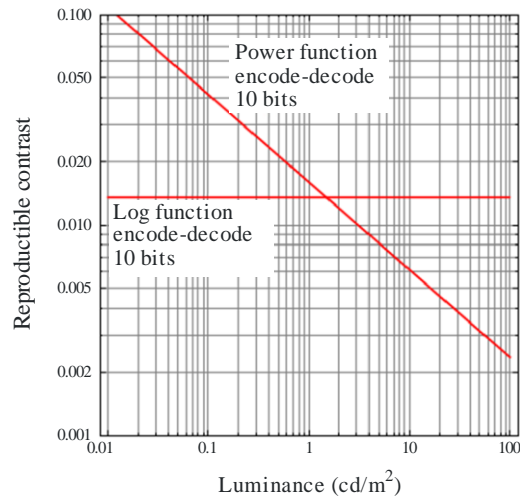
Φ_0 : spectral density of the neural noise

C_{ab} : aberration constant of the eye lens

N_{max} : maximum number of integration cycles, and u_0 is the spatial frequency limit of the lateral inhibition process.

The angular size of the object X_0 is 60 deg., and the photon conversion factor p is $1.2274 \cdot 10^6$ photons/s/deg²/Td.

FIGURE 32
Reproducible contrast comparison of log and power functions



Report BT.2246-26

The Figure shows that the log function allots relatively more codes to the low luminance region, whereas the power function allots more to the high luminance region. The minimum detectable contrast increases in the low luminance region, and the luminance of reproduced video signals generally distributes more in the high luminance region. Hence, the power function is preferable to a log function.

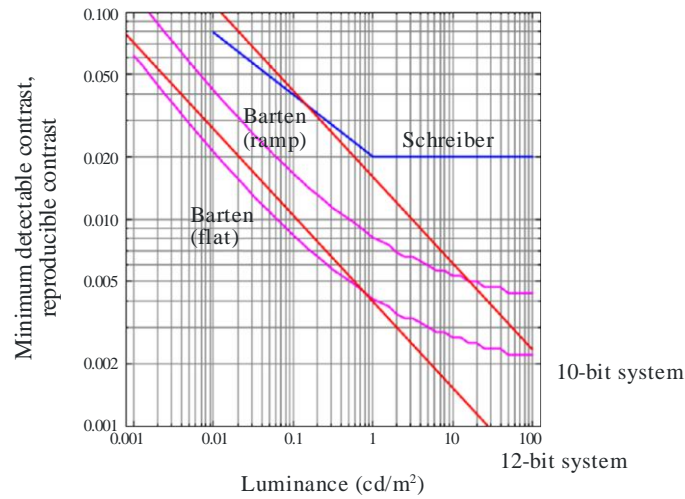
Reference [22] has derived the optimal coefficient and bit depth for digital cinema applications. The criterion is the CSF calculated with Barten's model. A coding function with a derived coefficient of 2.6 and a bit depth of 12 ensures that the reproducible contrast of the function is less than the minimum detectable contrast calculated from the CSF of the HVS. The calculation was done for the luminance range of cinema applications, i.e. approximately 50 cd/m².

A coefficient of 2.4 may be preferable for UHDTV because the current CRT reference monitor has characteristics that would be close to those with a power function with a coefficient of 2.4, and consequently the current operational practice would be preserved. It should be noted that the coding function for the signal format is determined from the assumed EOTF of reference display.

Now let us turn to the bit depth. Figure 33 shows the reproducible contrast characteristics of power functions with a coefficient of 2.4 and 10- and 12-bit precisions. Three minimum detectable contrast characteristics mentioned above are also plotted in the Figure for the sake of comparison. They are Schreiber's data, Barten's model for a flat signal, and the one for a ramp signal [22]. If we take Schreiber's data as a criterion, 10 bits is sufficient. If we take the Barten's model for rectangular signal even 12 bits is not sufficient for some ranges of luminance. However, when a natural scene is reproduced, a typical tone jump typically happens in parts where the luminance gradually changes. Therefore, a ramp signal is more appropriate than a flat signal. Taking those factors into account, Fig. 33 shows that 12 bits is sufficient to cause no tone jump artefact.

FIGURE 33

Comparison of reproducible contrasts of non-linear coding functions and HVS minimum detectable contrast characteristics

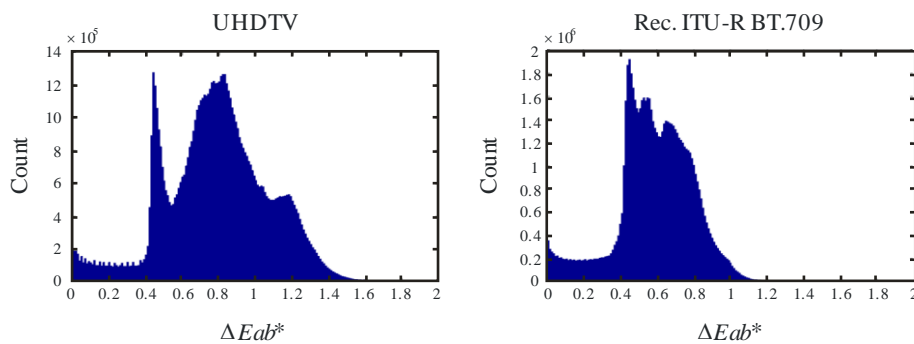


Report BT.2246-27

A wider colour gamut also expands the distance between consecutive codes given the same bit depth. Figure 34 shows the histogram of colour differences ΔE_{ab^*} between all consecutive 8-bit codes for Recommendation ITU-R BT.709 primaries and the new primaries proposed for UHDTV (Set b in Table 4 of [Input, 19]). The mean and maximum values are listed in Table 9. The mean of ΔE_{ab^*} for UHDTV is 35% larger than that of HDTV. This means that 1 bit more of precision is sufficient to make the ΔE_{ab^*} for UHDTV equal to or less than that for HDTV. The maximum value for UHDTV will be below 1 if 10-bit coding is employed.

FIGURE 34

Histogram of colour errors in 8-bit coding



Report BT.2246-28

TABLE 9

Mean and Max of colour errors in 8-bit coding

Colorimetry	Mean	Max
Recommendation ITU-R BT.709	0.58	1.45
Proposed for UHDTV	0.78	2.05

4 Operational experience

This section details the experience of those involved in producing, distributing and presenting UHDTV programmes.

4.1 London Olympics SUPER Hi-VISION⁶ public viewing operations

NHK and the BBC in cooperation with the Olympic Broadcasting Service (OBS), held public viewing of SUPER Hi-VISION at venues in Japan, the UK, and the USA for the 2012 London Olympics. This operation consists of programme production at Olympic Games venues, post-production at a BBC studio, IP distribution of live and non-live programmes through global and domestic IP networks, and regular presentations during the Olympic Games at nine presentation venues in three countries. A summarized report is shown in Annex 3.

4.2 Implementation of UHDTV broadcasting

4K and 8K UHDTV satellite broadcasting has been in operation in Japan. The detailed information and specifications are described in Annex 4.

4.3 Initiatives for 8K/120 Hz broadcasting

8K/120 Hz broadcasting is yet to be implemented and effort has been made to develop various pieces of equipment for the 8K/120 Hz ecosystem including program production, transmission and display. Live production and transmission of 8K/120 Hz programmes were also conducted to demonstrate the feasibility of 8K/120 Hz broadcasting. More details are provided in Annex 5.

5 Applicable specifications for UHDTV broadcasting

This section shows the applicable specifications for UHDTV broadcasting developed by the ITU.

⁶ SUPER Hi-VISION complies with 7 680 × 4 320 UHDTV system described in Recommendation ITU-R BT.2020 – Parameter values for ultra-high definition television systems for production and international programme exchange, and a 22.2 multichannel sound system described in Report ITU-R BS.2159-4 – Multichannel sound technology in home and broadcasting applications.

TABLE 10
Specifications for UHD TV broadcasting

Items		ITU Texts	Specifications	
Video signal	Spatial format	Rec. ITU-R BT.2020 (1) Rec. ITU-R BT.2100 (2)	7680×4320, 3840×2160	
	Frame frequency		24/1.001, 24, 25, 30/1.001, 30, 50, 60/1.001, 60, 100, 120/1.001, 120 Hz	
	Colorimetry		Wide colour gamut	
	Transfer function		Rec. ITU-R BT.2100	PQ, HLG
			Rec. ITU-R BT.2020	SDR
Coding	Rec. ITU-T H.265 (3), Rec. ITU-R BT.1870 (4), Rec. ITU-R BT.2073 (5)	MPEG-H HEVC		
Audio signal	Speaker configuration	Rec. ITU-R BS.2051 (6)	Channel-based (0+2+0 – 9+10+3), object-based	
	Coding	Rec. ITU-R BS.1196 (7)	MPEG-2 AAC, MPEG-4 AAC, MPEG-4 ALS	
Multimedia		Rec. ITU-R BT.2075 (8)	IBB, HTML5	
		Rep. ITU-R BT.2342 (9)	Closed captions	
Transport		Rec. ITU-T H.222.0 (10)	MPEG-2 TS	
		Rec. ITU-R BT.2074 (11), Rec. ITU-R BT.1869 (12)	MMT/TLV	
CAS		Rec. ITU-R BT.1852 (13)		
Satellite transmission		Rec. ITU-R BO.1784 (14)	DVB-S2	
		Rec. ITU-R BO.2098 (15)	ISDB-S3	

ITU Texts applicable for UHD TV broadcasting

- (1) Recommendation ITU-R BT.2020 – Parameter values for ultra-high definition television systems for production and international programme exchange
- (2) Recommendation ITU-R BT.2100 – Image parameter values for high dynamic range television for use in production and international programme exchange
- (3) Recommendation ITU-T H.265 – High efficiency video coding
- (4) Recommendation ITU-R BT.1870 – Video coding for digital television broadcasting emission
- (5) Recommendation ITU-R BT.2073 – Use of the high efficiency video coding (HEVC) standard for UHD TV and HDTV broadcasting
- (6) Recommendation ITU-R BS.2051 – Advanced sound system for programme production

- (7) Recommendation ITU-R BS.1196 – Audio coding for digital broadcasting
- (8) Recommendation ITU-R BT.2075 – Integrated broadcast-broadband system
- (9) Report ITU-R BT.2342 – Production, emission and exchange of closed captions for all worldwide language character sets (latin and non-latin)
- (10) Recommendation ITU-T H.222.0 – Information technology – Generic coding of moving pictures and associated audio information: Systems
- (11) Recommendation ITU-R BT.2074 – Service configuration, media transport protocol, and signalling information for MMT-based broadcasting systems
- (12) Recommendation ITU-R BT.1869 – Multiplexing scheme for variable-length packets in digital multimedia broadcasting systems
- (13) Recommendation ITU-R BT.1852 – Conditional-access systems for digital broadcasting
- (14) Recommendation ITU-R BO.1784 – Digital satellite broadcasting system with flexible configuration (television, sound and data)
- (15) Recommendation ITU-R BO.2098 – Transmission system for UHDTV satellite broadcasting

References

- [1] Mark Schubin, “Searching for the Perfect Aspect Ratio”, *SMPTE J.*, Vol. 105, No. 8, pp. 460-479, 1996.
- [2] M. Emoto and M. Sugawara, “Flicker perceptions for wide-field-of-view and hold-type image presentations”, *IDW'09*, pp. 1233-1234, 2009.
- [3] M. Sugawara, K. Omura, M. Emoto, and Y. Nojiri, “Temporal Sampling Parameters and Motion Portrayal of Television”, *SID 09 DIGEST*, pp. 1200-1203, 2009.
- [4] M. Emoto, and Y. Matsuo, “Frame Frequency for Wide Field of View Video System – Dependency of Dynamic Visual Acuity on Field of View”, *Annual Conv. of ITE*, 2008.
- [5] M. Emoto *et al.*, “Improvement in Moving Picture Quality by High Frame Rate”, *The Journal of the ITE*, Vol. 65, No. 8, pp. 1208-1214, 2011.
- [6] R. W. G. Hunt, *The Reproduction of Colour, Sixth Edition: John Wiley & Sons*, 2004.
- [7] Samsung Electronics Products produced in 2007-2009.
- [8] Digital Cinema System Specification, ver. 1.0, Digital Cinema Initiatives, LLC Member Representatives Committee, 2005.
- [9] Recommendation ITU-R BT.709-5 – Parameter values for the HDTV standards for production and international programme exchange, ITU, Geneva, Switzerland, 2002.
- [10] Adobe RGB® Color Image Encoding, *Adobe Systems Incorporated*. 2005.
- [11] SMPTE 170M-2004, “Composite analog video signal – NTSC for studio applications”, 2004.
- [12] M. Kuwata, *et al.*, “A 65-in. slim (255-mm depth) laser TV with wide-angle projection optical system”, *JSID 17(11)*, pp. 875-882, 2009.
- [13] S. Baron and D. Wood, “The Foundation of Digital Television: The Origins of the 4:2:2 Component Digital Television Standard”, *SMPTE Motion Imaging Journal*, pp. 327-334, September 2005.
- [14] Recommendation ITU-R BT.1120-7 – Digital interfaces for HDTV studio signals, December 2007.
- [15] SMPTE 292M-2008 1.5 Gb/s Signal/Data Serial Interface.

- [16] SMPTE 372M-2009 Dual Link 1.5 Gb/s Digital Interface for 1920×1080 and 2048×1080 Picture Formats.
- [17] SMPTE 424M-2006 Television – 3 Gb/s Signal/Data Serial Interface.
- [18] SMPTE 425-2008 3 Gb/s Signal/Data Serial Interface – Source Image Format Mapping.
- [19] M. Nakamura, T. Nakatogawa, and K. Oyamada, “Development of Optical Interface for Full-Resolution Super Hi-Vision Production Systems”, *ITE Tech, Rep.*, Vol. 33, No. 32, pp. 5-8, July 2009.
- [20] W. F. Schreiber, *Fundamentals of Electronic Imaging Systems, Third Edition*: Springer-Verlag, 1992.
- [21] P. G. J. Barten, *Contrast Sensitivity of the Human Eye and its Effects on Image Quality*: SPIE PRESS, 1999.
- [22] M. Cowan, G. Kennel, T. Maier, and B. Walker, “Constant Sensitivity Experiment to Determine the Bit Depth for Digital Cinema”, *SMPTE Motion Imaging Journal*, Vol. 2004, pp. 281-292, 2004.

Annex 1

Input documents to WP 6C or SG 6 related to UHDTV⁷

Input Ref.	Year	Document No.	Title	Submitter
Study Period 2008-2011				
1	2008	6C/39	Proposed draft new Questions ultra-high definition television (UHDTV)	Japan
2		6C/40	Report on recent progress of technologies that are applicable to expanded hierarchy of LSDI, EHRI and UHDTV	Japan
3		6/62	Draft revision of Question ITU-R 40/6 – Extremely high-resolution imagery	WP 6C
4		6C/115	Proposed framework of work plan for the study on ultra-high definition television	Japan
5		6/92	Research and development of Ultra-high Definition Television (UHDTV)	EBU, BBC, RAI, NHK
6		6C/133, Annex 16	Framework of work plan for the study on ultra-high definition television	Chairman, WP 6C
7		6C/133, Annex 18	Decision – Appointment of a Rapporteur Group – Commencement of study on UHDTV	Chairman, WP 6C
8	2009	6C/174	Report on the work plan and study of UHDTV	Rapporteur Group / UHDTV
9		6C/176	Tentative forecast for the applications and the time frame of UHDTV broadcasting services	CBS
10		6C/195	Proposed preliminary draft new Recommendation – Parameter values for extremely high-resolution imagery for production and international exchange of programmes	Korea
11		6C/210	Proposed definition of UHDTV (Doc. 6C/TEMP/130)	WP 6C
12		6C/210, Annex 9	Working document towards a preliminary draft new Recommendation – Parameter values for extremely high-resolution imagery for production and international programme exchange	Chairman, WP 6C
13		6C/210, Annex 14	Working document for the study on UHDTV	Chairman, WP 6C
14		6C/210, Annex 20	Continuation of the Rapporteur Group – Study on UHDTV	Chairman, WP 6C
15		6C/246	Report on the study of UHDTV	Rapporteur Group / UHDTV
16		6C/255	Proposed colorimetry for UHDTV	Japan
17		6C/257	Performance evaluation of compression coding for expanded hierarchy of LSDI and EHRI signals	Japan
18		6C/277	Proposed preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] – Parameter values for UHDTV systems for production and international programme exchange	Korea
19		6C/287, Annex 9	Working document toward preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV]	Chairman, WP 6C

⁷ Documents related to audio are not included.

Input Ref.	Year	Document No.	Title	Submitter
20		6C/287, Annex 13	Studies relating to UHDTV	Chairman WP 6C
21		6C/287, Annex 25	List of Rapporteurs and Rapporteur Groups	Chairman, WP 6C
22	2010	6C/319	Report on the study of UHDTV	Rapporteur Group / UHDTV
23		6C/329 Rev.1	Technical modification to WD-PDNR ITU-R BT.[IMAGE-UHDTV]	Korea
24		6C/330	Examination of two sets of new RGB primaries proposed in WD-PDNR ITU-R BT.[IMAGE-UHDTV]	Korea
25		6C/333	Proposed modification to working document toward preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV]	Japan
26		6C/353	Liaison statement on draft Recommendation J.LSSYS: Real-time transmission system of exLSDI signals under spatial image segmentation for parallel processing	ITU-T SG 9
27		6C/388	Work plan towards a new Recommendation ITU-R BT.[IMAGE-UHDTV]	Korea
28		6C/401	Capture and display experiments of wide colour gamut images	Japan
29		6C/404	Update of compression coding performance report for expanded hierarchy of LSDI and EHRI signals	Japan
30		6C/405	Proposed preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] – UHDTV system parameters for production and international programme exchange	Japan
31		6C/407	Proposed draft new Recommendation ITU-R BT.[COLOUR] – Approaches of extended colour representation for television image systems	Japan
32		6C/415, Annex 7	Working document toward preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] – Parameter values for UHDTV systems for production and international programme exchange	Chairman, WP 6C
33		6C415, Annex 18	Working document toward draft new Report ITU-R BT.[UHDTV] – The present state of ultra-high definition television	Chairman, WP 6C
34		6C415, Annex 19	Work plan for the study of the baseband image format for UHDTV systems	Chairman, WP 6C
35	2011	6C/425	Liaison statement – Real-time transmission system of exLSDI signals under spatial image segmentation for parallel processing	ITU-T SG 9
36		6C/437	Proposed preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] – Parameter values for UHDTV systems for production and international programme exchange	Korea
37		6C/455	Consideration of luminance and colour-difference signal formats for UDTV image parameters	Japan
38		6C/456	Proposed modification to working document toward preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] – Inclusion of a higher frame frequency – UDTV system parameters for production and international programme exchange	Japan

Input Ref.	Year	Document No.	Title	Submitter
39		6C/461	Some considerations on broadcasting applications for UHDTV	CBS, Inc.
40		6C/466	Current status of discussions concerning colour equations for UHDTV systems	Chairman, WP 6C
41		6C/490, Annex 5	Preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] – Parameter values for UHDTV systems for production and international programme exchange	Chairman, WP 6C
42		6C/490, Annex 14	Preliminary draft new Report ITU-R BT.[UHDTV] – The present state of ultra-high definition television	Chairman, WP 6C
43		6C/490, Annex 17	Work plan for the study of the baseband image format for UHDTV systems	Chairman, WP 6C
44		6C/496	International Organization for Standardization - ISO/IEC JTC1/SC29/WG11 – Coding of moving pictures and audio – Liaison Statement on High Efficiency Video Coding (HEVC)	ISO
45		6C/501	Comment on PDN Recommendation ITU-R BT.[IMAGE-UHDTV] – Parameter values for UHDTV systems for production and international programme exchange	Italy
46		6C/504	Draft new Recommendation ITU-R BT.[IMAGE-UHDTV] – Parameter values for UHDTV systems for production and international programme exchange	United States of America
47		6C/519	Recent study on high frame frequency television to support parameter values described in PDNR ITU-R BT.[IMAGE-UHDTV] – UHDTV system parameters for production and international programme exchange	Japan
48		6C/520	Proposed modifications to preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] for a draft new Recommendation – Parameter values for UHDTV systems for production and international programme exchange	Japan
49		6C/528	Performance evaluation of compression coding for expanded hierarchy of LSDI and EHRI signals	Japan
50		6C/535	Report on the study of UHDTV	Rapporteur Group / UHDTV
51		6C/553	Proposed draft new Recommendation ITU-R BT.[IMAGE-UHDTV] – Parameter values for UHDTV systems for production and international programme exchange	Korea
			Other input documents to WP 6C	
52	2008	6C/114	Proposal of required electro-optical transfer function for a FPD as a master monitor	Japan
53	2009	6C/170	A perception based image representation for inclusion into the draft modification of Recommendation ITU-R BT.1361	United States of America
			Study Period 2012-2015	
54	2012	6C/564 Annex 4	Preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] – Parameter values for UHDTV systems for production and international programme exchange	Chairman, WP 6C
55		6C/564 Annex 14	Establishment of a Rapporteur Group on colour-encoding scheme for UHDTV	Chairman, WP 6C
56		6C/570	Liaison statement on High Efficiency Video Coding (HEVC)	ISO

Input Ref.	Year	Document No.	Title	Submitter
57		6C/572	Reply liaison statement – 3D broadcasting service and EHRI	ITU-T SG 9
58		6C/11	Draft new Recommendation ITU-R BT.[IMAGE-UHDTV] – Parameter values for UHDTV systems for production and international programme exchange	RG on colour encoding methods
59		6C/24	Activity Report – Analysis of sub-sampled versions of the UHDTV signals	RG on colour encoding methods
60		6C/29	A study on the presence effect during UHDTV viewing	Russian Federation
61		6C/33	Proposed modifications to preliminary draft new Recommendation ITU-R BT.[IMAGE-UHDTV] – Parameter values for UHDTV systems for production and international programme exchange	Japan
62		6C/56	Reply liaison statement on High Efficiency Video Coding (HEVC) and Ultra-high Definition Television (UHDTV)	ITU-T SG 16
63		6C/57	Liaison statement on new work item for 4K and 3D scalable video service over HD video service	ITU-T SG 9
64		6C/59	Liaison statement on Interactive Applications and HEVC	ISO
65		6C/60	Liaison statement on UHDTV, HEVC, and new Questions in ITU-R	ISO
66		6C/77	Proposed preliminary draft new Report – Image dynamic range in television systems	USA
67		6C/78	Proposal to include 50 Hz based frame rates in Recommendation ITU-R BT.2020	EBU
68		6C/81	Progress Report on UHDTV – London Olympics SUPER Hi-VISION public viewing operations	NHK, BBC

Input Ref.	Year	Document No.	Title	Submitter
			Study Period 2012-2015	
	2013	6C/141	Improved picture quality provided by capturing, editing, finishing, and archiving HDTV programmes in 3 840 × 2 160 (“4K”) UHDTV	BBC, CBS, Inc.
		6C/152	120FPS acquisition in 50 Hz territories	United Kingdom of Great Britain and Northern Ireland
		6C/176 Annex 2	Preliminary draft new Recommendation ITU-R BT.[HQPROD] – Use of UHDTV image systems for capturing, editing, finishing and archiving high-quality HDTV programmes	Chairman, WP 6C
		6C/181	Liaison statement on equations in Recommendation ITU-R BT.2020 and related specifications from ITU-T Question 6/16	ITU-T SG 16
		6C/201	Impact of 120 fps acquisition in 50 Hz territories	United Kingdom of Great Britain and Northern Ireland

Input Ref.	Year	Document No.	Title	Submitter
		6C/202	Proposed revision of Recommendation ITU-R BT.2020 – Parameter values for ultra-high definition television systems for production and international programme exchange	United Kingdom of Great Britain and Northern Ireland
		6C/211	Image acquisition of the scene illuminated by flickering lights at high frame rate	NHK
		6C/212	Proposal to include 100 frames per second in Recommendation ITU-R BT.2020	EBU
		6C/214	Proposed new Opinion of Study Group 6 – New television programme and post-production method for improved picture quality in programme distribution and broadcasting	BBC, CBS, Inc.
		6C/215	Soliciting adoption of preliminary draft new Recommendation ITU-R BT.[HQPROD] (Annex 2 to Document 6C/176)	BBC, CBS, Inc.
		6C/216	Impact of 120 fps acquisition in 50 Hz territories on emission encoding	BBC
		6C/219	Transfer characteristics and related specifications in Recommendation ITU-R BT.2020	Japan
		6C/220	Comments on Documents 6C/201 and 6C/202 – Frame frequencies above 60 Hz for UHD TV	Japan
		6C/228	Real-time production interfaces for UHD TV	SMPTE
		6C/233	Comments on Contribution 6C/202 “Proposed revision of Recommendation ITU-R BT.2020”	Italy
		6C/246	Impact of image resolution on perceived video quality using HD/UHD displays at 1.5H viewing distance	Orange (F)
		6C/247	Viewing distance issue for subjective quality assessment of UHD TV	Orange (F)
		6B/145	Proposed working document towards a draft new Recommendation on digital interfaces for UHD TV studio signals	Japan
		6B/155	Proposed new Recommendation ITU-R BT [UHD TV-TAILOR] – Tailoring of UHD TV programmes for their distribution and broadcasting in HDTV	BBC, CBS, Inc.
		6B/156	Proposed preliminary draft new Recommendation ITU-R BT.[UHD TV-IF] – Transport of ultra-high definition digital television programmes for international programme exchange in broadcasting	Australia
		6C/257 Annex 5	Preliminary draft revision of Recommendation ITU-R BT.2020 – Parameter values for ultra-high definition television systems for production and international programme exchange	Chairman, WP 6C
		6C/257 Annex 9	Working document towards a draft revision of Report ITU-R BT.2246 – The present state of ultra-high definition television	Chairman, WP 6C
		6C/257 Annex 11	Appointment of a Rapporteur to complete a draft revision of Report ITU-R BT.2246 – The present state of ultra-high definition television	Chairman, WP 6C
		6B/180 Annex 2	Working document towards a new Recommendation ITU-R BT.[UHD TV-TAILOR] – Tailoring of UHD TV programmes for their distribution and broadcasting in HDTV	Chairman, WP 6B

Input Ref.	Year	Document No.	Title	Submitter
		6B/180 Annex 3	Working document towards a draft new Recommendation ITU-R BT.[UHDTV-IF] – Real-time digital interfaces for UHDTV studio signals	Chairman, WP 6B

Annex 2

Analysis of sub-sampled versions of the UHDTV signals

1 Summary

See § 3.2.7.2.

2 Fundamental

2.1 Why YC transmission is needed

The YC format, in which image data is handled as one luminance signal and two colour signals, was first used in the early days of colour television. Its primary advantages include compatibility with conventional (i.e. black and white) television and its ability to reduce bandwidth when applied to a composite signal. It is also less susceptible to DC errors and external noise in transmission than RGB format. It has therefore been widely used in the analogue age.

Studies on the digital component format in the 1980s led to the adoption of a 2:1 ratio for the luminance channel and colour channels. The YC format has been continuously used in serial digital transmission and continues to be necessary in the digital age because it can reduce bandwidth or bit rate and is compatible with conventional operational practice.

2.2 Fundamental of YC transmission

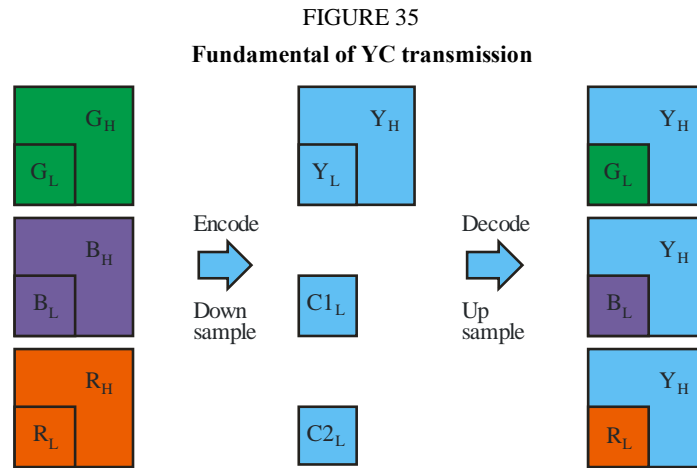
The principle of bandwidth reduction in YC transmission is based on the human visual system (HVS) characteristic of being less sensitive to high frequency colour signals than to the luminance signal. Here, high frequency means any frequency above the Nyquist frequency of colour sub-sampling. RGB signals are converted to YC signals, which reduces the bandwidth of C signals (Fig. 35) with little degradation to perceptual picture quality. The spatial information of luminance to which HVS is sensitive is maintained over the entire bandwidth. The high frequency spatial information of colour, to which HVS is less sensitive, is not transmitted. This reduces the total signal bandwidth. Overall, the YC format can be described as a system in which the low frequency components of input RGB signals are maintained and the high components become achromatic.

What are the appropriate signals for Y and the two C channels? Obviously, luminance Y in the XYZ colorimetry was selected for the Y channel. It might be desirable to select independent luminance signals for the C channels. Early researchers selected colour differences B-Y and R-Y for the C channels, possibly due to the ease of implementation.

Another bandwidth reduction technique in signal format is use of the non-linear transfer function, also called gamma correction. The original purpose of this function was the pre-compensation of

CRT's electro-optical transfer function, but coincidentally, it matches the HVS tone sensitivity as well. Therefore it can be used even in the post-CRT era.

Both YC colour encoding with bandwidth reduction and gamma correction with non-linear conversion include a non-linear signal process. This means that differently ordered processes yield different outputs. Early researchers selected processes in which gamma correction was performed first, followed by YC colour encoding. They probably did this due to the simplicity of the hardware. It is particularly advantageous that colour-decoded signals could be directly fed to a CRT display at the receiver end. In a previous study on HDTV, the Y signal was calculated before gamma correction [1], but this method was ultimately not adopted as the HDTV standard.



Report BT.2246-29

2.3 The constant luminance issue

Calculating the Y' signal after the gamma correction of the RGB signals causes a so-called constant luminance (CL) issue. More specifically, the problems caused by the Y' signal being calculated from gamma-corrected RGB signals differ from those when the Y signal is calculated from linear RGB signals then gamma corrected. Hereafter, the first case is called "Luma" and the second is called "Luminance".

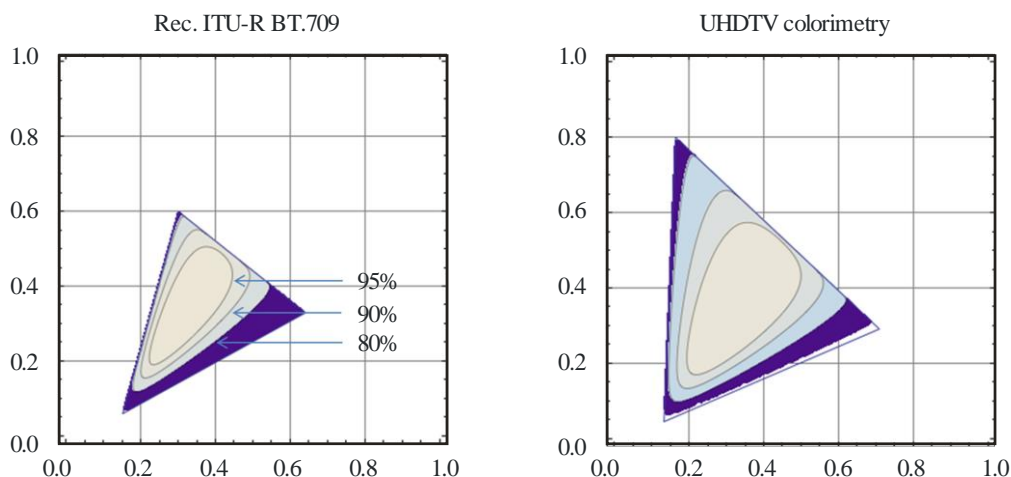
Luminance matches human brightness sensitivity more closely than Luma. The ratio between Luma and Luminance in the linear domain is called the constant luminance index [2].

$$k = \frac{(rR^{1/\gamma} + gG^{1/\gamma} + bB^{1/\gamma})^\gamma}{(rR + gG + bB)}$$

The constant luminance index is 100% for achromatic signals. Then it decreases as the colour saturates (Fig. 36). In most cases, it does not matter whether the Y' signal carries the true luminance or not because the luminance of a reproduced image after colour decoding is a result of the matrix calculation from Y'C_B'C_R' to RGB, and the original luminance is reproduced. The problem occurs when the reproduced luminance is affected by the bandwidth reduction of the C_B' and C_R' signals. This phenomenon is described as the decreased luminance level at highly saturated and high-frequency parts. This leads to the loss of luminance detail and/or a drop in the average luminance.

FIGURE 36

Constant luminance index



Report BT.2246-30

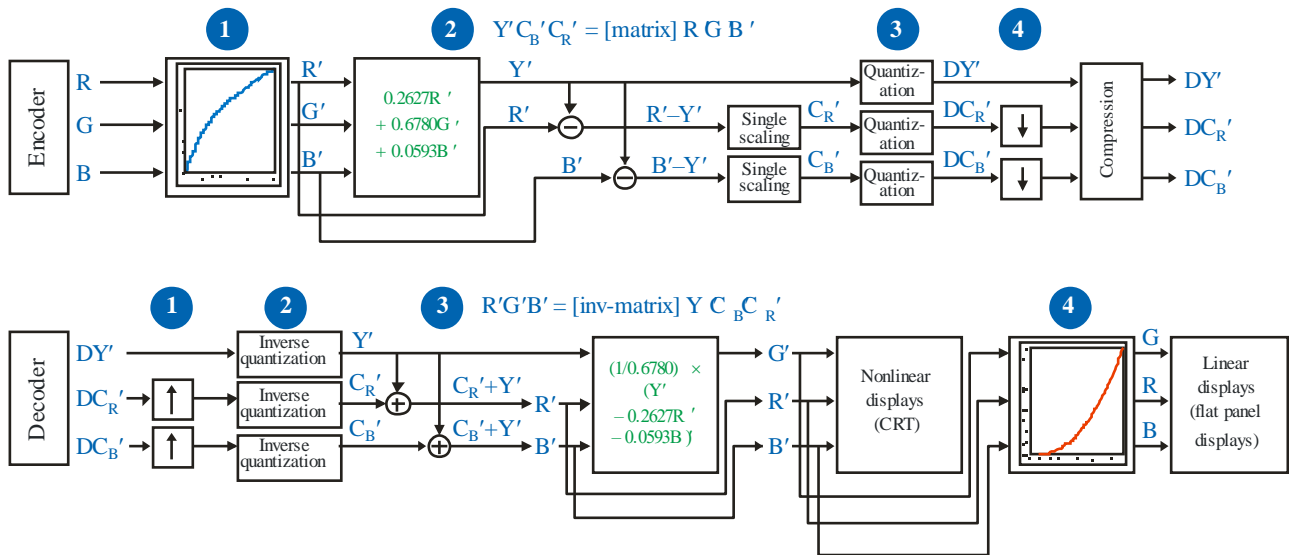
2.4 Block diagrams for the encoder and decoder in the non-constant luminance (NCL) and CL

Figures 37 and 38 illustrate the computational procedures in the encoder and decoder for each of the NCL and CL signal formats proposed in the PDNR ITU-R BT.[IMAGE-UHDTV] (Annex 4 to Doc. 6C/564) [Input, 54]. The computational step in each of the NCL and CL is described using the number inside a circle where linear RGB signals are assumed to be input to the encoder.

The colour encoding process defined in Recommendation ITU-R BT.709 (same as in the NCL) was actually proposed to comprise the least computational steps, and so the non-linear R'G'B' signals can be used without further transformation as an input to compensate the intrinsic non-linear property of a CRT receiver. On the contrary, a linear display such as LCD and AMOLED mainly will be used for UHDTV systems. Therefore, it is reasonable to have linear RGB signals at the end of the decoder. The computational steps in Figs 37 and 38 are described by taking into account real implementations. The quantized full-resolution DR'DG'DB' or down-sampled DY'DC_B'DC_R' signals which are produced from video cameras are further manipulated in a studio post production. The calculations in the encoder are generally applied for video cameras and those in the decoder are for a TV receiver. If needed during the post-production in a studio, both encoding and decoding processes are used.

FIGURE 37

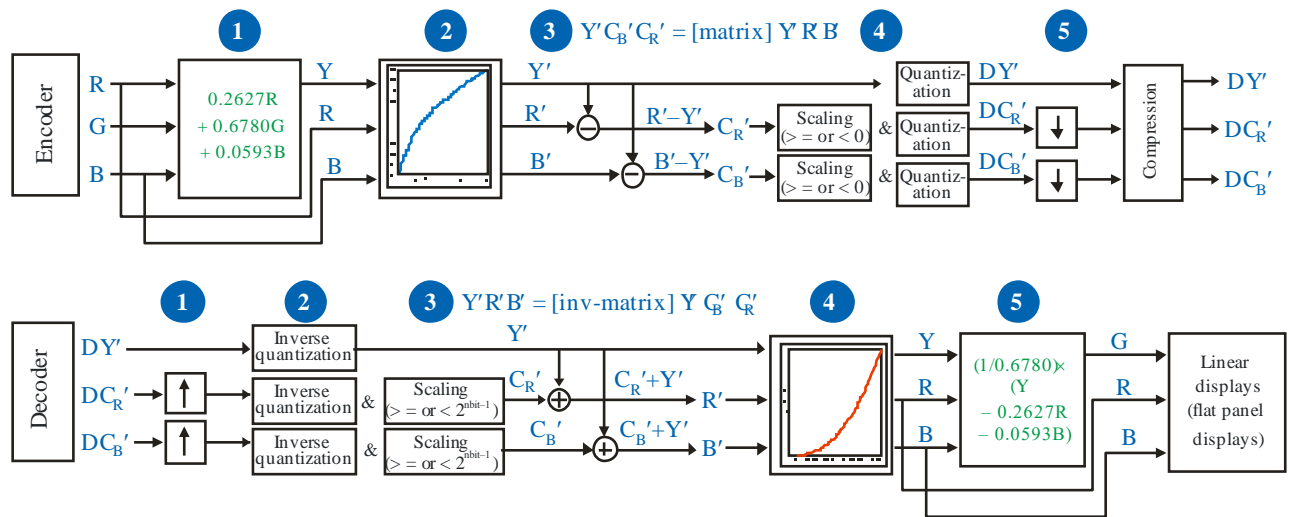
The NCL colour encoding method



Report BT.2246-31

FIGURE 38

The CL colour encoding method



Report BT.2246-32

NOTE – Scaling can be included into the quantization equation and so both scaling and quantization can be carried out simultaneously in a real implementation of the CL.

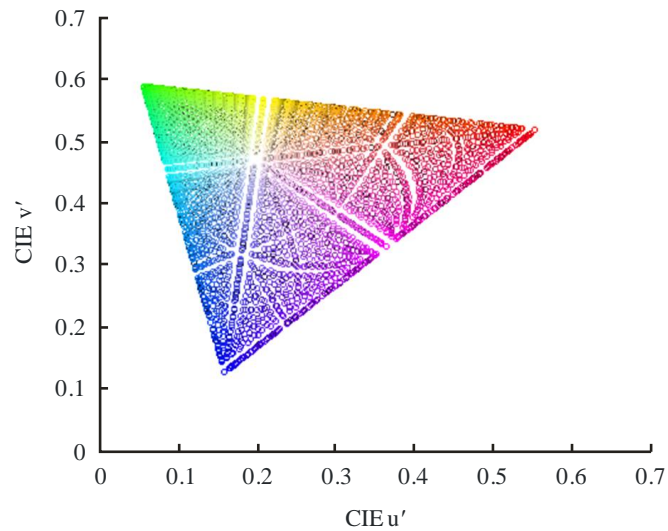
2.5 Crosstalk characteristics

The crosstalk characteristics in each of three pairs ($Y'-C_B'$, $Y'-C_R'$ and $C_B'-C_R'$) were evaluated in terms of the correlation coefficient using 4,096 colours sampled to encompass the UHDTV colour gamut (see Fig. 39). The coefficients were computed in the two sets of NCL and CL video signals. Figure 40 shows the computed correlation coefficients in the 10 sub-sample groups divided according to 10 equal luminance segments of 10 cd/m² each (0-100). The blue and red bars represent the coefficient values of the NCL and CL video signals respectively. Overall, higher coefficient values are found in all $Y'-C_B'$, $Y'-C_R'$ and $C_B'-C_R'$ signal pairs of NCL. Significant crosstalk in the $Y'-C_R'$ pair of NCL is forecasted from much higher coefficient values compared with those in the CL, i.e. much taller blue bars than red bars in Fig. 40(b). This fact suggests that the luminance and red-green

chrominance information is not accurately separated in the Y' - C_R' pair of NCL. For very bright colours, the traverse of luminance information to the yellow-blue chrominance signal C_B' of NCL is tended to be occurred (taller blue bars in Y80 – Y100 in Fig. 40(a)). The red-green and yellow-blue chrominance components, C_B' - C_R' , include not only its own information but also the others' for bright colours in NCL and for very dark colours in CL (taller blue bars in Y80 – Y100 and taller red bars in Y10-Y20 in Fig. 40(c)).

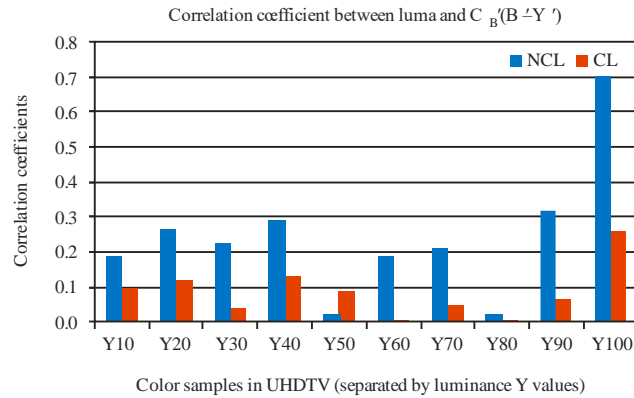
FIGURE 39

The 4096 colours chosen to cover the UHDTV gamut in the CIE u' - v' diagram

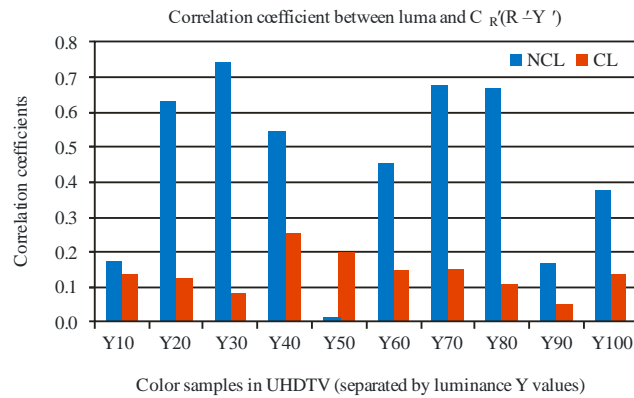


Report BT.2246-33

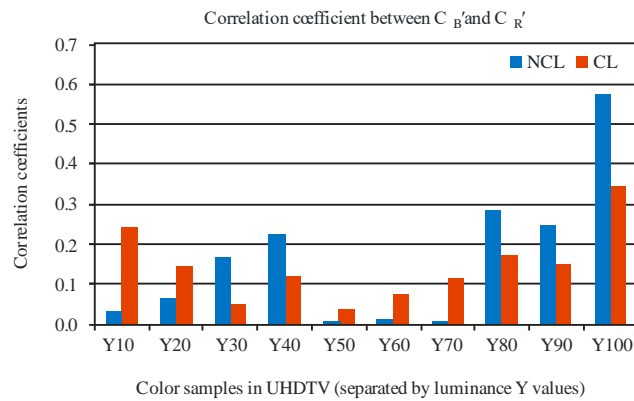
FIGURE 40
 The correlation coefficients in (a) $Y'-C_B'$, (b) $Y'-C_R'$ and (c) $C_B'-C_R'$ pairs
 for the NCL and CL video signals



(a)



(b)



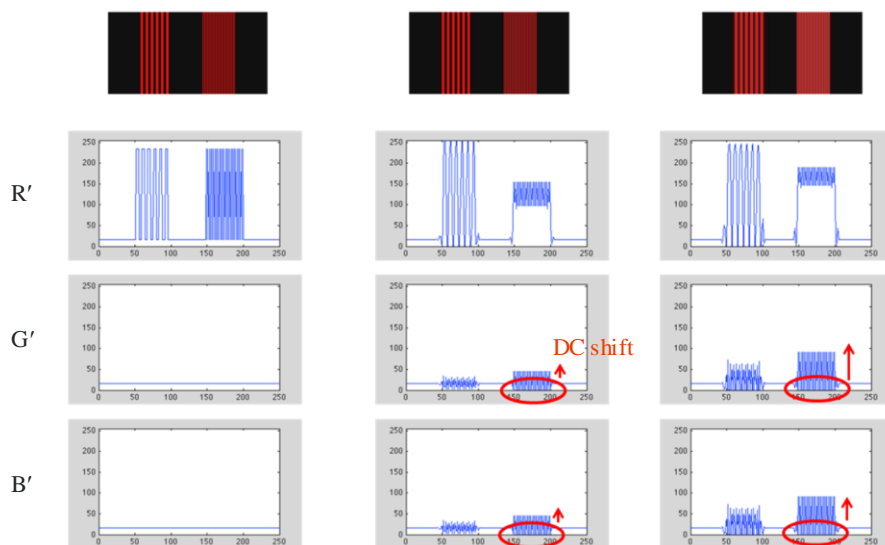
(c)

2.6 Quality degradation by clipping

Clipping (Fig. 41) is another factor that contributes to quality degradation in YC transmission, although it is not related to the constant luminance issue. In the clipping process, the luminance signals are added to the RGB signals as YC transmission makes the high frequency component achromatic. In some cases, the values after adding the luminance may exceed the “legal value” range between black and white, and consequently they are clipped at the highest or lowest values. This leads

to a highly perceptible change in the average level of the portion and occurs in both the constant and non-constant approaches, although the degree is different depending on the case.

FIGURE 41
Mechanism of clipping in YC transmission



Report BT.2246-35

2.7 Evaluation index

With television applications, the ideal evaluation method is subjective evaluation because television is primarily focused on visual presentation. One objective index to use is the lightness difference between the original RGB image and the images that appear after each encoding process. We should point out here that the index is significant when it corresponds to the subjective evaluation; also, the lightness difference between the original RGB image and the resulting images are distributed in the high frequency component. This means that the degradation is less visually perceptible compared with one distributed all over the frequency range.

An additional evaluation should be performed if the low frequency component of the original RGB signals are maintained. There is no consensus for this in the index at present, but R'G'B' signals or lightness, saturation, hue, or delta E calculated from the R'G'B' signals may be used. These indexes should be calculated from the low frequency component because the YC system makes the high component achromatic.

2.8 Factors other than picture quality

When considering the colour equations, not only picture quality and compression efficiency but also various other factors should be taken into account. These issues were discussed with broadcasters in Attachment 1. Some examples are as follows; hardware implementation by manufacturers, and consideration of compatibility between HDTV and UHDTV.

3 The reference models of the NCL and CL

Two colour encoding equations were proposed in the PDNR ITU-R BT.[IMAGE-UHDTV] (Annex 4 to Document 6C/564) [Input, 54]: Non-Constant Luminance (NCL) and Constant Luminance (CL) colour encoding methods. The Rapporteur Group prepared reference models for the NCL approach and the CL approach. The software was implemented using MATLAB.

A block diagram of the NCL model is shown in Fig. 42. The low-pass filter characteristics for sub-sampling and up-sampling are based on MPEG documents [3, 4]. Model CL (Fig. 43) changes the calculation of the Y' signal from the NCL model. It is noted that the model CL incorporates separate gains for the negative and positive parts of the C_{B'} and C_{R'} signals though they are not explicitly described in the Figure.

Further examinations were carried out to find out how the approaches illustrated in Fig. 43 would perform in a practice. The examination results are given in § 4.3.

FIGURE 42

Reference model for non-constant luminance approach

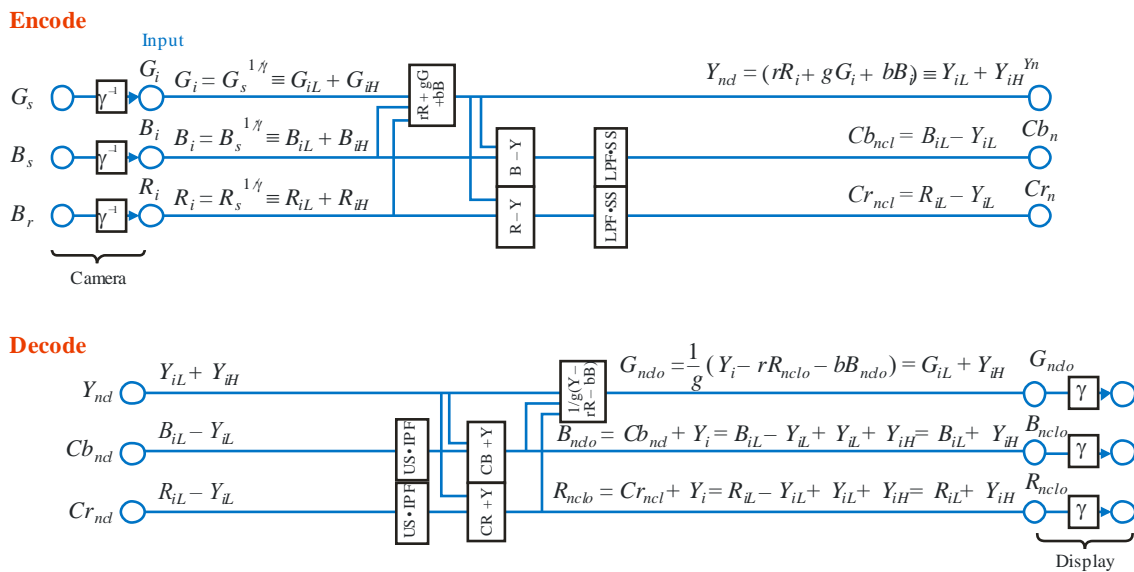
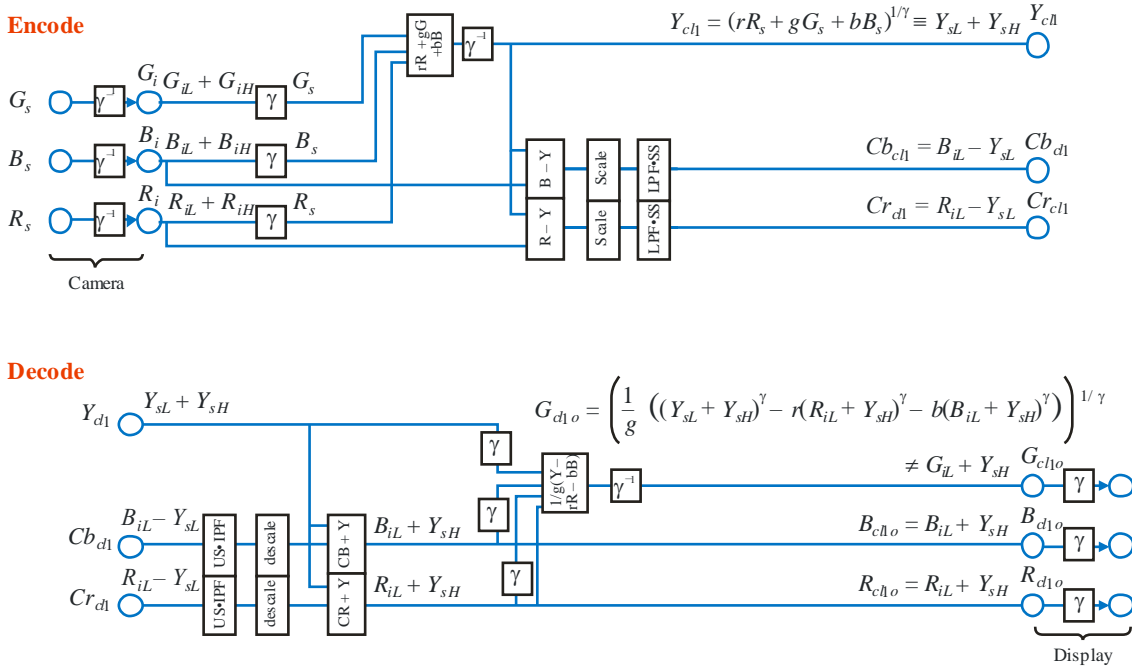


FIGURE 43

Reference model for constant luminance approach



Report BT.2246-37










4 Objective and subjective evaluation results

4.1 Objective evaluation of YC encoding-decoding

The NCL and CL were objectively evaluated by using natural images [5]. The difference of lightness and low frequency components of R'G'B' between the source image and encoded-decoded images were calculated as peak signal noise ratio (PSNR). The results were listed in Table 11.

It is obvious that CL shows higher PSNR lightness compared with NCL. This is because the CL signal carries the luminance while NCL carries so-called luma. Regarding the low frequency components of R'G'B', CL shows slightly low levels for G'. This may be because there is some interference between high frequency component of Y and low frequency component of G signal as described in Fig. 43. Note that the luminance (lightness) information is composed of RGB three components, not individual R or G or B.

TABLE 11
Objective evaluation results

Picture	Index	NCL	CL	Picture	Index	NCL	CL	Picture	Index	NCL	CL
Chroma	L*	56.2	83.6	Hat	L*	53.6	89.1	Tulip	L*	44.8	57.0
	R'L	48.1	48.0		R'L	44.2	44.6		R'L	38.6	37.9
	G'L	56.2	54.4		G'L	54.1	50.7		G'L	48.2	45.2
	B'L	40.0	40.3		B'L	41.1	40.8		B'L	36.0	35.9
Couple	L*	65.8	105.3	Sweat	L*	57.9	96.3	Woman	L*	63.5	104.4
	R'L	51.7	51.5		R'L	49.5	49.6		R'L	52.7	52.0
	G'L	60.0	60.4		G'L	59.8	57.9		G'L	61.8	61.0
	B'L	47.9	48.0		B'L	44.9	44.7		B'L	48.0	48.1
Eiffel	L*	54.0	66.6	Tour	L*	50.6	75.7	Yacht	L*	51.5	65.6
	R'L	46.1	45.3		R'L	42.5	42.2		R'L	43.1	42.4
	G'L	55.4	55.7		G'L	52.0	50.4		G'L	53.8	53.6
	B'L	43.8	43.7		B'L	39.5	39.5		B'L	39.9	39.8

4.2 The evaluation results in a production process

Current programme productions in HDTV are mostly done in gamma-corrected domain with R'G'B' and YC format. Mixing of two images is performed with R'G'B', NCL-Y'CB'CR', and CL-Y'CB'CR' formats and the results are shown in Fig. 44.

Figure 44 shows that there is a slight difference between CL and R'G'B' while no subjective difference is seen between NCL and R'G'B'. The colour appearance of the cork-board in the CL domain looks different compared to that in each of the NCL and R'G'B'. This is due to the fact that the luma equations of CL are not the linear combinations of gamma-corrected R'G'B' while those of NCL are. The dual normalizing gain for CB' and CR' in CL is also supposed to be a part to the cause.

Table 12 introduces another mixing example using two different images in the four different domains of linear RGB, non-linear R'G'B', NCL-Y'CB'CR' and CL-Y'CB'CR'. The mixed image in the linear RGB domain looks brighter than the others. There is not a notable visual difference in the three output images in the non-linear R'G'B', NCL and CL.

Comparing the output image appearance between the NCL and the CL domains after the mixing operation, the results can be summarized as follows.

- In some cases of mixing, output appearance differences are visible in parts of the image, but not in the entire image.
- In other cases, no significant difference is observed.
- Relative to the output colour image appearance in the linear RGB domain, overall a slightly darker appearance is seen in the other three domains of R'G'B', NCL and CL.

The mixing operation is done to avoid users noticing a sudden transition between two dissimilar images in a video stream. Thus, it is preferable that the brightness of the mixed scene does not drop or increase sharply from the two original scenes.

Note that producers can adjust the colour appearance of the output image after the mixing operation according to their preferences. There is no one 'correct' result from a mix, and there can be vast numbers of colour combinations mixed.

FIGURE 44
Mixed results of two images
“Eiffel” and “Sweat”, upper left portion

(a) NCL



(b) RGB









(c) CL



TABLE 12

Mixing two different natural images (in comparison with Fig. 44)

Original two different images to be mixed		
Domain	Linear RGB	Non-linear R'G'B'
Mixed colour image appearance	 Brighter than the other three mixed versions	
Domain	NCL-Y'C_B'C_R'	CL-Y'C_B'C_R'
Mixed colour image appearance		

4.3 Real implementation vs. the hypothetical loss in G for the CL

In accordance with the conceptual mathematical analysis in Fig. 43 in the previous section, the low-frequency component in the input G' signal at the encoder is not maintained at the decoder in the CL. This is described subsequently in terms of the mathematical expression in Fig. 43.

$$G_{cl10} = \left(\frac{1}{g} \left((Y_{sL} + Y_{sH})^\gamma - r(R_{iL} + Y_{sH})^\gamma - b(B_{iL} + Y_{sH})^\gamma \right) \right)^{1/\gamma} \neq G_{iL} + Y_{sH}$$

Further examination was carried out to find out how this mathematically explained analysis result appears in a practical implementation. The preservation of the low-frequency G' component is important to keep to the original luminance information as much as possible after manipulations such as down-sampling and compression. Note that the green component among RGB contributes most to luminance information in an image.

One stimulus composed of a) grey and red alternately and b) yellow and magenta alternately, though fairly rarely seen in natural programmes, was chosen to investigate the difference between the mathematically derived approach and a practical implementation, to emphasize the visual difference more clearly. Figures 45 and 46 illustrate the red-grey stimulus and the yellow-magenta stimulus respectively.

The original red-grey stimulus and the reconstructed stimuli after down-sampling the colour-difference components of NCL-C_B'C_R' and CL-C_B'C_R' are seen in Figs 45(a), (b) and (c) respectively. Table 13 shows the non-linear R'G'B', tristimulus XYZ and CIELAB L*a*b* values for the original and its reconstructed images after down-sampling. The Y in the tristimulus values indicates luminance information.

Comparing the output R'G'B' values in the NCL and CL cases with the R'G'B' values of the original red area, the differences in the NCL are -60 ΔR', +21 ΔG', +23 ΔB' and those in the CL are -42 ΔR', +59 ΔG', +48 ΔB'. If only ΔG' is taken into account, the CL colour encoding looks to cause larger deviation from the original G' after the down-sampling process. However, all the R'G'B' values in the reconstructed images after down-sampling are varied from the original values. Then, let us compare the luminance (Y) value between the original and its counterpart images in the NCL and CL. Larger ΔY is found between the original and its manipulated image in the NCL domain than in the CL domain: -6.99 ΔY in the NCL vs. -0.02 ΔY in the CL from the original red area. Owing to this difference, it is expected that human observers can perceive darker appearance in the reconstructed image in the NCL domain in the difference of 11.8ΔL* against the original red area.

FIGURE 45
A stimulus composed of red and grey colours

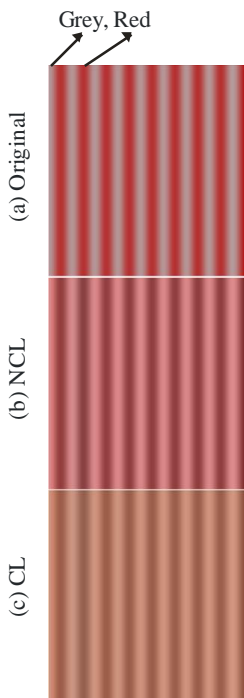


TABLE 13

The non-linear R'G'B', tristimulus XYZ and CIELAB L*a*b* values for the original and its reconstructed images after down-sampling NCL-C_B'C_R' and CL-C_B'C_R' signals seen in Figs 45(a), (b) and (c)

Nonlinear R'G'B'		Grey			Red		
		R'	G'	B'	R'	G'	B'
Org		180	180	180	180	16	16
NCL	-17	235	158	160	120	37	39
CL		222	163	148	138	75	64

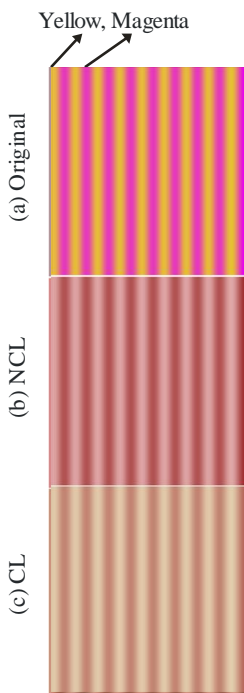
→ All R'G'B' values are varied.

Tristimulus XYZ		Grey			Red		
		X	Y	Z	X	Y	Z
Org		53.41	56.19	61.18	35.79	14.76	0
NCL		77.2	57.76	47.33	15.63	7.77	2.55
CL	-0.01	69.06	56.18	40.16	22.59	14.74	6.9

CIELAB L*a*b*		Grey			Red		
		L*	a*	b*	L*	a*	b*
Org		79.72	0	0	45.31	96.81	78.12
NCL		80.6	50.12	15.06	33.51	60.57	28.11
CL	0	79.72	36.94	21.61	45.27	45.59	25.91

Figure 46 shows the other stimulus consisting of yellow and magenta alternately. Table 14 describes the non-linear R'G'B', tristimulus XYZ and CIELAB L*a*b* values for the original and its reconstructed images after down-sampling. Again, the same tendency is observed as in the red-grey stimulus in Fig. 44: larger ΔY between the original and its reconstructed image in the NCL domain than in the CL domain, i.e. $-17.6 \Delta Y$ in the NCL vs. $-0.09 \Delta Y$ in the CL from the original yellow area, and $-14.9 \Delta Y$ in the NCL vs. $-0.05 \Delta Y$ in the CL from the original magenta area.

FIGURE 46
A stimulus composed of yellow and magenta colours



Report BT.2246-40

TABLE 14
The non-linear R'G'B', tristimulus XYZ and CIELAB L*a*b* values for the original and its reconstructed images after down-sampling NCL-C_B'C_R' and CL-C_B'C_R' signals seen in Figs 46(a), (b) and (c)

	Nonlinear R'G'B'			Yellow			Magenta		
	R'	G'	B'	R'	G'	B'	R'	G'	B'
Org	230	230	16	230	16	230	230	16	230
NCL	235	189	189	164	57	57	164	57	57
CL	235	222	186	186	112	98	186	112	98

→ All R'G'B' values are varied.

	Tristimulus XYZ			Yellow			Magenta		
	X	Y	Z	X	Y	Z	X	Y	Z
Org	74.59	89.78	2.68	76.91	30.73	101.24	76.91	30.73	101.24
NCL	83.28	72.16	68.27	30.99	15.81	5.56	30.99	15.81	5.56
CL	86.64	89.69	66.45	44	30.68	16.99	44	30.68	16.99

	CIELAB L*a*b*			Yellow			Magenta		
	L*	a*	b*	L*	a*	b*	L*	a*	b*
Org	95.96	-21.15	134.75	62.28	128.52	-60.24	62.28	128.52	-60.24
NCL	88.05	29.98	8.21	46.73	73.78	33.95	46.73	73.78	33.95
CL	95.87	2.62	23.24	62.24	49.55	27.22	62.24	49.55	27.22

The fact that the low-frequency component in the input G' signal is not maintained in the CL colour encoding as claimed by the mathematical analysis (see Fig. 43) turns out contrarily in the real implementation. The CL colour encoding was more effective in the preservation of the original luminance information than the NCL. The disagreement occurring between the mathematical analysis and the real implementation is because the luminance information is determined by all the RGB values, not by G alone.

4.4 The evaluation results in a delivery chain

4.4.1 The evaluation schemes and test sequences

The influence of different video signals of NCL and CL on compression efficiency was examined using HEVC (Ver. 4). The random access condition (GOP = IBBP, Intra Period = 16) in the HEVC with four quantization parameters of 22, 27, 32 and 37 was applied for the evaluation. The evaluation scheme was divided into two categories: the same video signal from down-sampling to compression

(see the evaluation scheme I in Table 15) vs. different video signals applied for different processes (see the evaluation scheme II in Table 16). The motivation for the formation of the evaluation scheme II was to figure out the corresponding stage (down-sampling or compression) contributing to gaining higher PSNR-lightness and colour difference values, i.e. less difference between the original and its compressed versions. The down-sampling was carried out inevitably twice for the evaluation scheme II. The evaluation scheme (ES II-Down NCL, HEVC NCL) in the second row in Table 16 was used as an anchor to find out the impact of applying CL for either down-sampling or compression in replacement of NCL.

TABLE 15

The evaluation scheme I where the same video signal from down-sampling to compression

	Down-sampling (RGB 444 to Y'Cb'Cr' 420)	Compression using HEVC Ver. 4 (Y'Cb'Cr' 420)
ES I-NCL(Anchor)	NCL	NCL
ES I-CL	CL	CL

TABLE 16









The evaluation scheme II where different video signals applied for different processes

	(1) Down-sampling (RGB 444 to Y'Cb'Cr' 420) (2) Up-sampling (Y'Cb'Cr' 420 to RGB 444)	Down-sampling (RGB 444 to Y'Cb'Cr' 420)	Compression using HEVC Ver. 4 (Y'Cb'Cr' 420)
ES II- Down/HEVC NCL (Anchor)	NCL	NCL	NCL
ES II- Down CL & HEVC NCL	CL	NCL	NCL
ES II- Down NCL & HEVC CL	NCL	CL	CL
ES II- Down/HEVC CL	CL	CL	CL

Eight test sequences were chosen in order to cover a wide range of temporal correlation and colourfulness (see Table 17). These were purchased from The Institute of Image Information and Television Engineers in Japan (ITE) and were provided in TIFF format composed of non-linear R'G'B' values of Recommendation ITU-R BT.709 for 59.94 Hz/interlace scanning [6]. These test sequences were also recommended for HDTV test materials in Report ITU-R BT.2245 [5]. Only 60 frames in the original test sequences were used to analyse the impact of different video signals on compression efficiency. The original vertical resolution was adjusted to convert interlace to progressive type and to be a multiple of 16. Only odd lines in the original sequences were firstly taken (1920H × 1080V to 1920H × 540V). A total of 28 lines from top and bottom was secondly cropped (540 V to 512 V) so that the vertical resolution could be a multiple of 16. Finally, the vertical resolution was increased from 512 to 1024 using the up-conversion method provided by MPEG [4] to mimic the original aspect ratio of 16:9.

TABLE 17

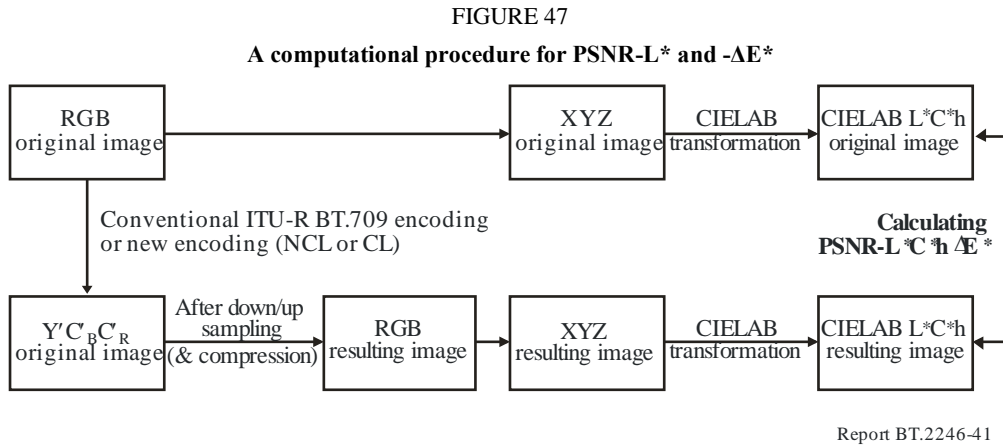
The eight test sequences chosen for a delivery chain evaluation

		Temporal correlation		
		High	Middle	Low
Colourfulness	High (colourful)	ITE-159 	ITE-160 	
	Middle	ITE-155  ITE-110 	ITE-106 	ITE-126 
	Low (neutral)		ITE-151 	ITE-165 

4.4.2 The evaluation measure

Luma (Y') and colour-difference ($C_B'C_R'$) components are all manipulated on the compression process whereas only colour-difference signals are altered on the chroma down-sampling process. This indicates that both luminance and chrominance information in an original image will be changed after the compression process. Hence, variations in the chroma and hue domains as well as those in the lightness domain were calculated between the original and its compressed images in terms of PSNR-lightness (CIELAB L^*) and colour difference (CIELAB ΔE^*). The PSNR- ΔE^* represents the image difference in all the aspects of lightness, chroma and hue. Figure 47 illustrates the computational procedure for the PSNR-lightness and colour difference between an original image and its counterpart generated after down-sampling and compression processes.

Bjontegaard Delta-PSNR and -Rate (BD-PSNR and BD-Rate) were used as a measure to evaluate the influence of two different video signals of NCL and CL on compression efficiency [7]. For the evaluation scheme I in Table 15, the BD-PSNR and -Rate values were computed to assess bit-rate saving in the unit of percent and objective image-quality improvement in the unit of dB, respectively when the CL was applied for both down-sampling and compression in the comparison with the employment of NCL. For the evaluation scheme II in Table 16, the impact of the application of CL (instead of NCL) for either down-sampling or compression stage on compression efficiency was analysed using the computed BD-PSNR and -Rate values. The BD-PSNR value represents the measure for changes in objective image quality for a given bit rate. The BD-Rate indicates the measure in the aspect of bit-rate saving for a given objective quality.



4.4.3 The evaluation results in terms of BD-PSNR and BD-Rate

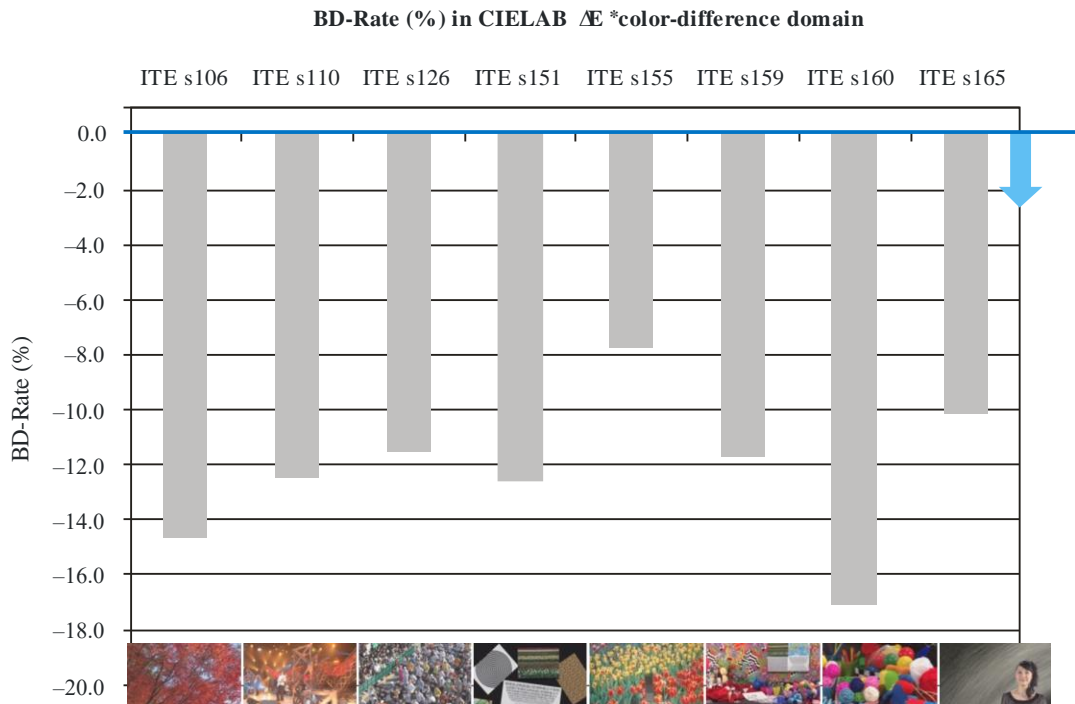
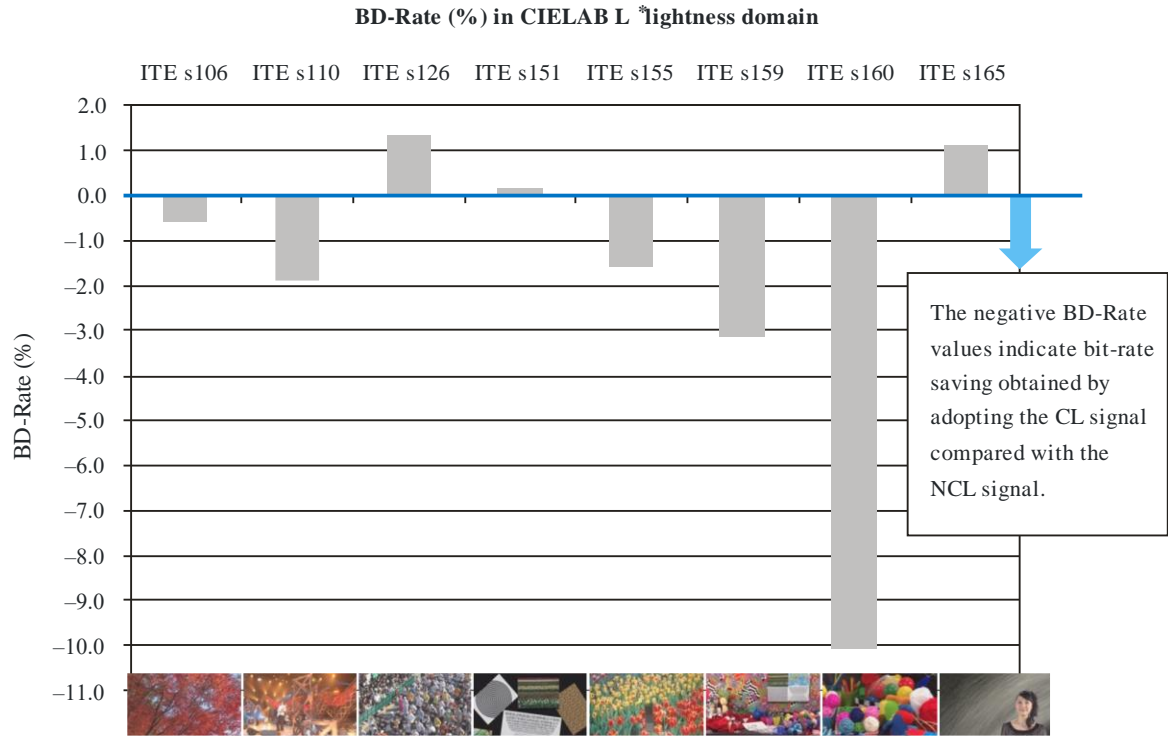
4.4.3.1 The evaluation scheme I where the same video signal from down-sampling to compression

The BD-Rate (%) and BD-PSNR (dB) were calculated to assess the bit-rate saving and the improvement in the objective quality of a compressed image that are expected to gain by applying the CL signal for all the production and delivery processes against the NCL signal. The BD-PSNR values in terms of lightness (L^*) and colour-difference (ΔE^*) were calculated between the original and its compressed versions. The BD-PSNR ΔE^* can indicate the objective deviation between the original and its compressed image in all the aspects of lightness, chroma and hue. Figures 48(a) and (b) show the computed BD-Rate and BD-PSNR values for the eight test sequences seen in Table 17. The estimated results in bit-rate saving and in the improvement of the objective quality of the compressed image are summarized hereinafter in the application of the CL signal for production and delivery compared with the NCL signal.

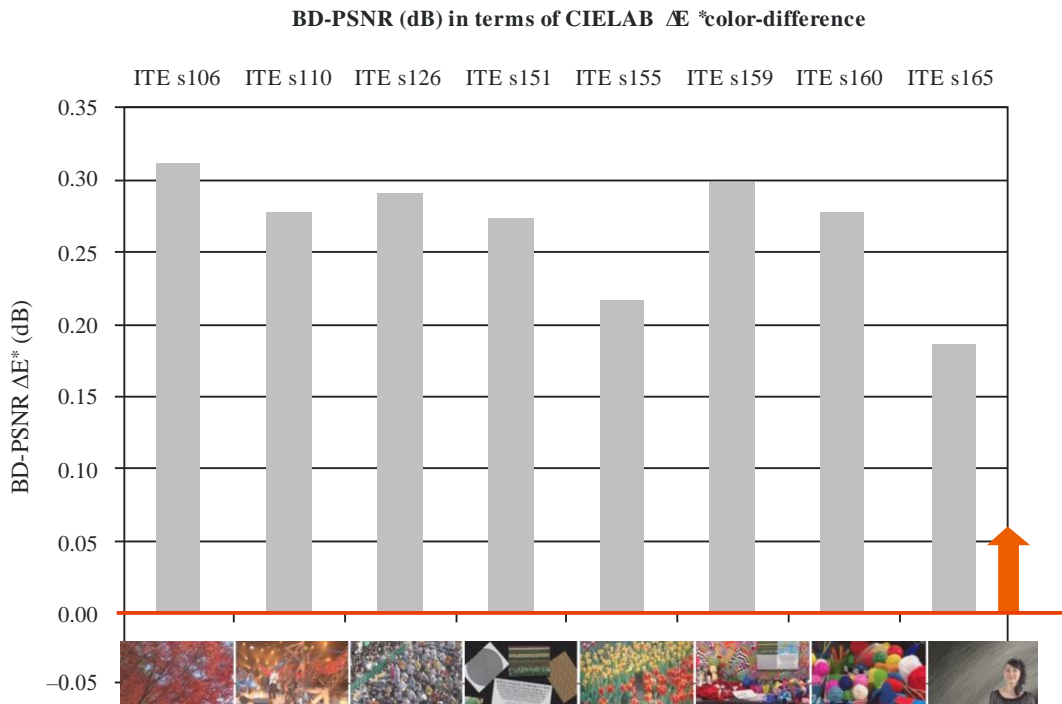
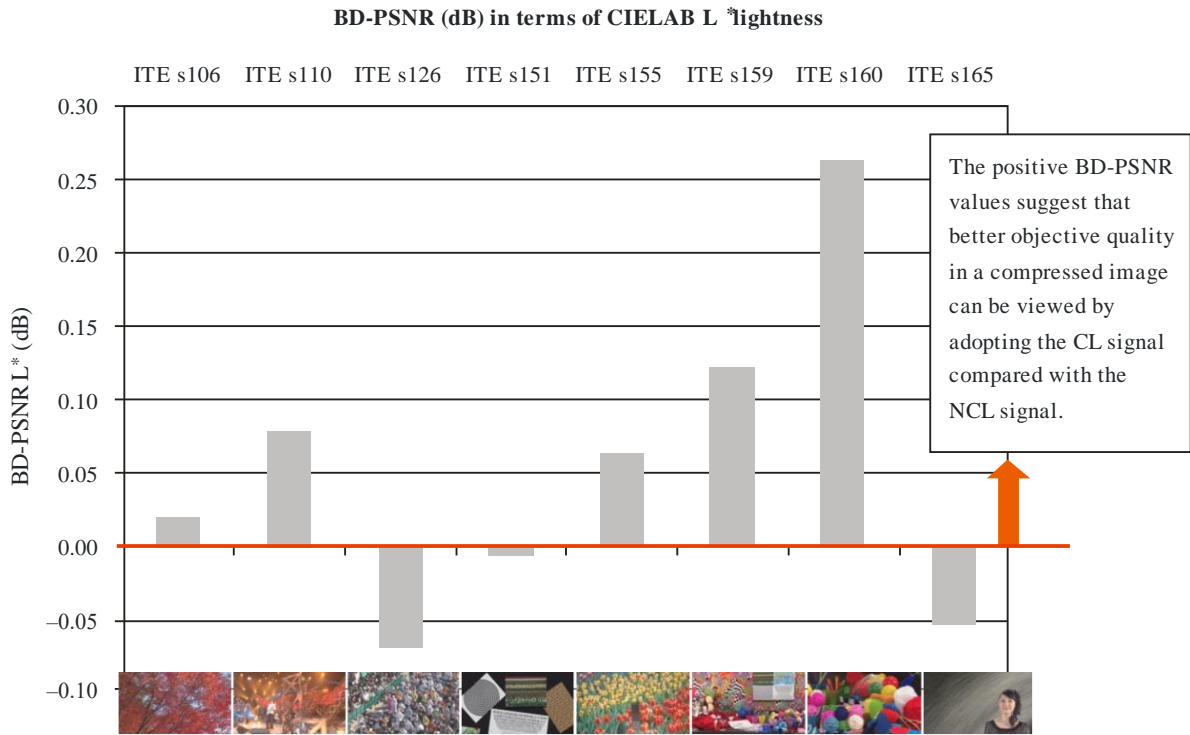
- The bit-rate reduction can be obtained by adopting the CL approximately -1.9% in the lightness domain and -12.2% in the colour-difference domain on average.
 - The maximum bit-rate saving in the lightness domain is -10.1% for ITE 160, but the slight increased bit rate of about 1% is also found for the two sequences of ITE 126 and ITE 165.
 - The maximum and minimum bit-rate saving in the colour-difference domain is -17% for ITE 160 and -7.7% for ITE 155.
- There are the average gains of 0.05 dB and 0.27 dB in the lightness and colour-difference domains for the CL, suggesting better preservation in the quality of the original image after down-sampling and compression processes.
 - The largest positive gain in the lightness domain is 0.26 dB for ITE 160, but negative gains are also observed for the two sequences of ITE 126 (-0.07 dB) and ITE 165 (-0.05 dB).
 - The largest and smallest positive gains in the colour-difference domain are 0.31 dB for ITE 106 and 0.19 dB for ITE 165.

FIGURE 48

(a) The BD-Rate (%) in the CIELAB lightness and colour-difference domains
 (The estimated bit-rate saving by applying the CL for down-sampling and compression in replacement of NCL)



(b) The BD-PSNR (dB) in terms of lightness and colour-difference
 (The estimated objective image-quality improvement by applying the CL for down-sampling and compression in replacement of NCL)



4.4.3.2 The evaluation scheme II where different video signals applied for different processes

One type of the video signal – CL or NCL – was employed throughout all the processes from down-sampling to compression in the evaluation scheme I (see Table 15). The evaluation results showed that the application of the CL signal can enhance compression efficiency in comparison with the NCL. For more detail analysis, the evaluation scheme II (see Table 16) was also designed with the following aim.

- To figure out the corresponding stage (down-sampling or compression) contributing to gaining higher PSNR-lightness and -colour difference values, i.e. positive BD-PSNR (dB) that is presumed to obtain by using the CL signal in replacement of the NCL.

The evaluation scheme (ES II-Down/HEVC NCL) in the second row in Table 16 was used as an anchor in the computation of the BD-Rate (%) and BD-PSNR (dB). In turn, the bit-rate saving and the improvement in the objective quality of a compressed image, which are expected to gain by applying the CL signal for either down-sampling or compression, could be assessed against the adoption of the NCL signal for all the processes. The evaluation scheme II is composed of four sub-schemes: (1) Down/HEVC NCL (anchor), (2) Down CL & HEVC NCL, (3) Down NCL & HEVC CL and (4) Down/HEVC CL. The down-sampling was carried out inevitably twice for the evaluation scheme II.

Figures 49(a) and 49(b) show the computed BD-Rate and BD-PSNR values that indicate the estimated ‘bit-rate saving’ and ‘gain in the PSNR-lightness and -colour difference’ when the sub-schemes (2) – (4) were applied against the anchor scheme (1). The average of bit-rate reduction in the unit of BD-Rate (%) and the average gain in the unit of the BD-PSNR (dB) are given in Table 18. The assessed results in the bit-rate saving and in the improvement of the objective quality of a compressed image are summarized subsequently.

TABLE 18

The average of bit-rate saving (%) and the average gain (dB)

Anchor scheme: Down/HEVC NCL		Down CL & HEVC NCL	Down NCL & HEVC CL	Down/HEVC CL	
BD-Rate (%)	lightness	–1.01%	–1.75%	–1.97%	Increase in the bit-rate saving (%)
	colour-difference	–0.93%	–10.80%	–10.63%	
BD-PSNR (dB)	lightness	0.04	0.06	0.07	Increase in the gain (dB)
	colour-difference	0.02	0.25	0.24	

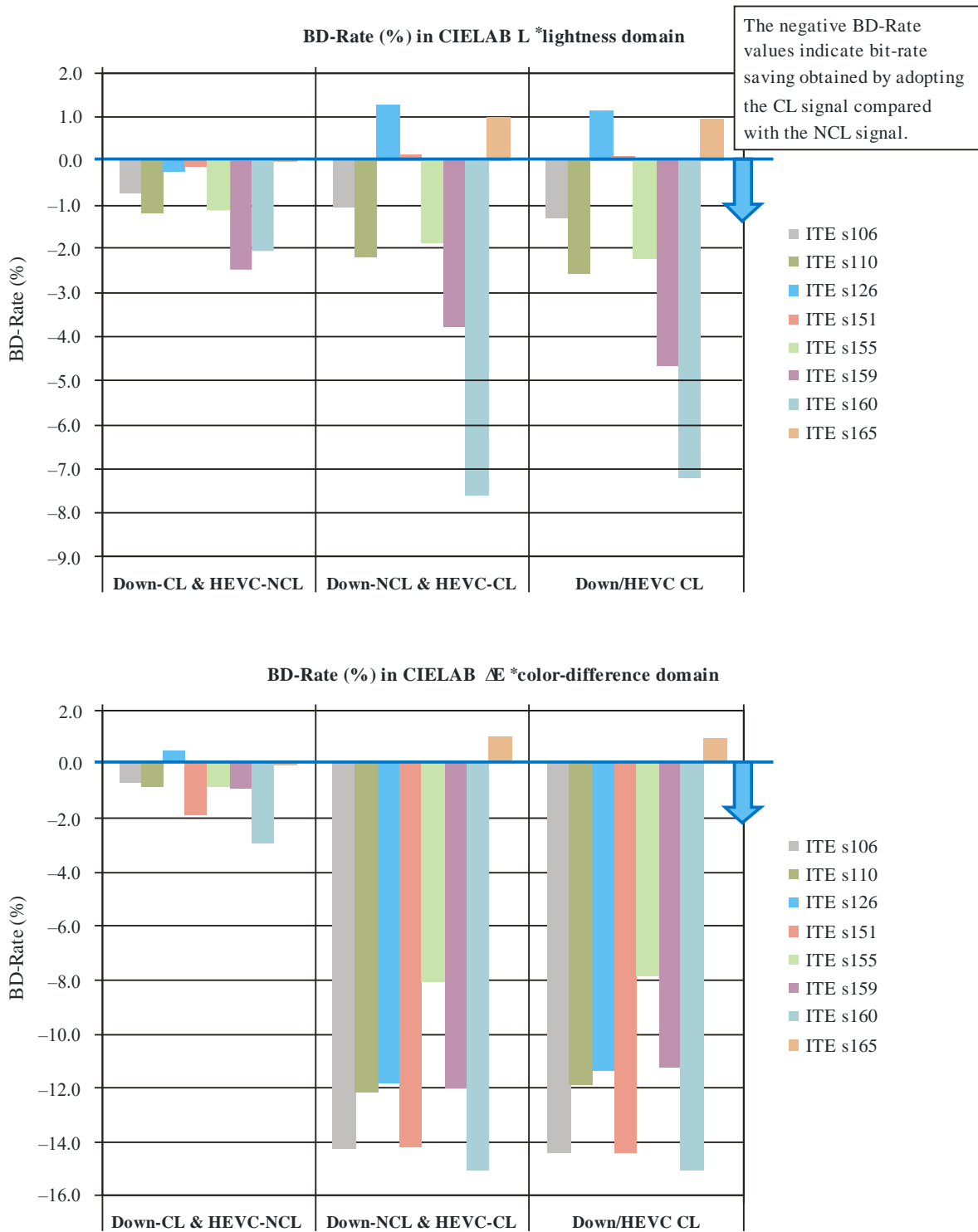
- Comparing with the application of the NCL for all the processes from down-sampling to compression, a bit-rate reduction and a gain in the PSNR are obtained by using the CL for even one or both processes, that is, negative % in the BD-Rate and positive dB in the BD-PSNR on average in all the three sub-schemes against the anchor scheme in Table 18.
- More bit-rate reductions (–1.01% to –1.75% and –0.93% to –10.8% in the lightness and colour-difference domains) and increased gains (0.04 to 0.06 and 0.02 to 0.25 in the lightness and colour-difference domains) are achieved in the application of the CL signal for the compression stage than for the down-sampling stage (‘Down CL & HEVC NCL’ vs. ‘Down NCL & HEVC CL’). This indicates that the usage of the CL for the compression stage contributes more significantly to decreasing bit rates and to improving the objective quality of the compressed image than for the down-sampling stage.

- There is an enhancement tendency both for bit-rates saving and for increasing PSNR values in the lightness domain when the CL signal is employed in all the processes than just in one of the down-sampling or compression process: $-1.01\% \rightarrow -1.75\% \rightarrow -1.97\%$ and $0.04 \text{ dB} \rightarrow 0.06 \text{ dB} \rightarrow 0.07 \text{ dB}$ in ‘Down CL & HEVC NCL’ vs. ‘Down NCL & HEVC CL’ vs. ‘Down/HEVC CL’.
- Greater reduction in the bit rates ($\Delta\text{BD-Rate}\%$) and larger gain in the PSNR values ($\Delta\text{BD-PSNR dB}$) are seen in the colour-difference domain than in the lightness domain between ‘Down CL & HEVC NCL’ vs. ‘Down NCL & HEVC CL’: -0.65% vs. -9.87% , and 0.02 dB vs. 0.23 dB in lightness- vs. colour difference-domain. This trend suggests that the usage of the CL for the compression is more effective in the maintenance of the chrominance property (chroma and hue) of the original in comparison with the usage of the NCL. Its reasons are presumed as follows.
 - As shown in § 2.5, more chrominance information is incorporated in the luma component and vice versa for the NCL, especially larger crosstalk between the luma and the colour-difference component of C_R for the NCL.
 - The luma rather than the colour-difference components is mainly manipulated in the compression process because reliable temporal information, which is a key factor in the data redundancy, is obtained from the luma. Therefore, larger variation in the chrominance information can be generated in the application of the NCL for the compression because the chrominance information included in both luma and colour-difference components is manipulated.

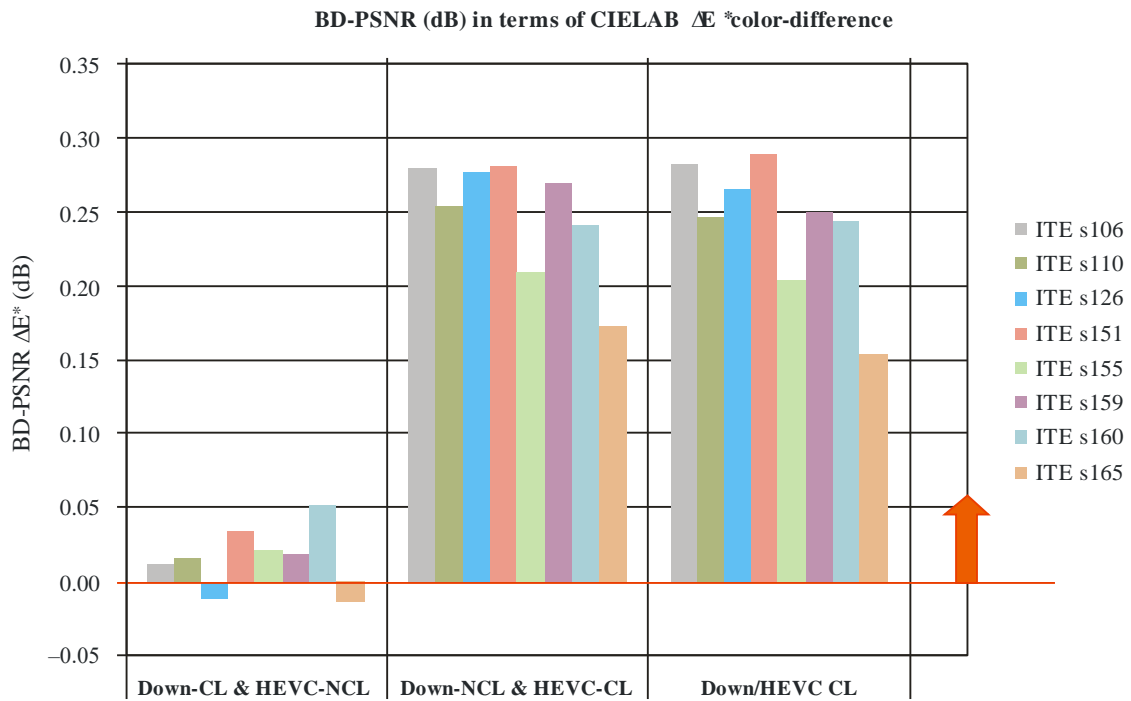
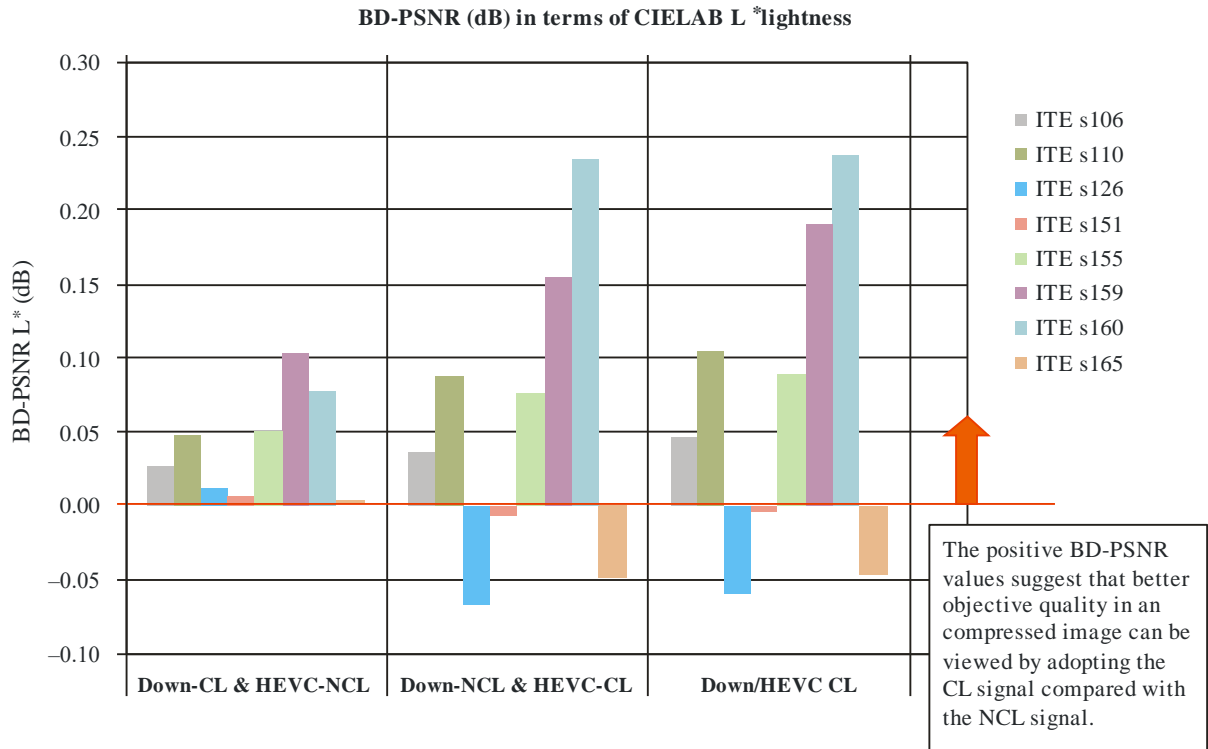
In conclusion, the enhancement results in the compression efficiency resulted from applying the CL seems to be drawn from more dominant contribution in the usage of the CL for the compression stage than for the down-sampling stage. However, the replacement of the NCL with the CL only in the down-sampling stage also provides an apparent improvement in the bit-rate saving and in the increase of the objective quality of the compressed image.

FIGURE 49

(a) The BD-Rate (%) in the CIELAB lightness and colour-difference domains
 (The estimated bit-rate saving by applying the CL
 for either down-sampling or compression in replacement of the NCL)



(b) The BD-PSNR (dB) in terms of lightness and colour-difference
 (The estimated objective image-quality improvement by applying the CL for either down-sampling or compression in replacement of the NCL)



5 Summary of the evaluation results

The deviations of YC encoded-decoded signals from original RGB signals are summarized in Table 19. They are categorized into effects associated solely with the constant luminance issue (No. 1), those associated with the way CL could be implemented internally in equipment (No. 2), those associated with clipping (No. 3) and those associated with signal processing (No. 4). This classification may help the analysis of picture quality degradation of YC signals.

RG found that the qualitative difference due to Nos 1, 2, and 3 between the approaches for normal video content (natural pictures) would be too small to provide results for which the difference in mean scores with non-expert assessors falls outside the confidence intervals of the results.

The difference in the PSNR lightness between NCL and CL observed in § 4.1 is supposed mainly due to No. 1. The difference in PSNR G'_L between NCL and CL is supposed mainly due to No. 2.

TABLE 19
Deviations of YC signals from RGB signals

No.	Model	Mechanism	Signal deviation from RGB model	Possible subjective recognition
1	NCL	Combination of Y signal calculation after gamma-correction and chroma sub-sample	Luminance level in high frequency component with high colour saturation becomes lower.	High frequency component modulation mostly drops. Average level becomes darker
2	CL	Interaction between high frequency component and low frequency component in decoded G signal which the mathematical analysis shows is possible (see Fig. 43 in § 3)	Low component of G signal may deviate from the source signal.	The colour of low frequency component may change hypothetically. ⁸
3	NCL CL	Luminance signal is equally added to R, G, B signals and the results exceed the legal value.	Clipping at black or white.	The colour of low frequency component changes.
4	CL	The luma signal Y' is formed by applying the nonlinear function to the sum of linear RGB signals. Then, the $C_B'C_R'$ signals are produced by subtracting the Y' from the gamma-corrected B'R'.	Mixed results of two images in the CL domain may not be the same as those in the R'G'B' domain.	The output colour image appearances after the mixing operation is similar in some cases and is partially different in other cases (see § 4.2).

The impact of different signal formats of NCL and CL on a studio post production and a delivery process was examined since the WP 6C meeting in September, 2011. The evaluation results are very briefly described in Table 20.

⁸ However, in practical implementations this effect can be overshadowed by the benefits of the preservation of the original luminance information (see § 4.3).

TABLE 20

Summary of the evaluation results for the NCL and CL video signal formats

		NCL-Y'C _B 'C _R '	CL-Y'C _B 'C _R '
Studio post-production	(1) Mixing operations – Natural images (see § 4.2)	<ul style="list-style-type: none"> – Similar output appearance compared with that in non-linear R'G'B' domain – Overall darker output appearance compared with that in linear RGB domain 	<ul style="list-style-type: none"> – Comparing the output appearances between the NCL and the CL domains, the results can be summarized as follows: <ul style="list-style-type: none"> • In some cases, output appearance differences are visible in parts of the image, but not in the entire image. • In other cases, no significant difference is observed. – Overall darker output appearance compared with that in linear RGB domain
	(2) Down-sampling operations (see § 4.1)	<ul style="list-style-type: none"> – Larger loss (i.e. smaller PSNR-lightness) in the original luminance information after down-sampling process than in the CL 	<ul style="list-style-type: none"> – Compared with the NCL, less loss (i.e. larger PSNR-lightness) in the original luminance information after down-sampling colour difference signals
Delivery using HEVC Codec (see § 4.4)			<ul style="list-style-type: none"> – Beneficial to improve compression efficiency due to more orthogonal CL video signals (see § 2.5) compared with the NCL signals

References

- [1] Report ITU-R BT.801-4 (1990) – The present state of high-definition television.
- [2] Livingston, D. C. (1954), “Colorimetric aspects of the NTSC colour television system”. *Proc. IRE42*: 138-150.
- [3] ISO/IEC JTC1/SC29/WG11 MPEG3/N6295 (2003), “Colour format down-conversion for test sequence generation”.
- [4] ISO/IEC JTC1/SC29/WG11 MPEG3/N6296 (2003), “Colour format up-conversion for test sequence generation”.
- [5] Report ITU-R BT.2245 (2011) – HDTV test materials for assessment of picture quality.
- [6] The Institute of Image Information and Television Engineers in Japan.
- [7] An excel add-in for computing Bjøntegaard metric and its evolution, *Document VCEG-AE-07*, ITU-T Study Group 16 Question 6, VCEG, ITU-T, Jan. 2007.

Annex 3

London Olympics SUPER Hi-VISION Public Viewing Operations Report

1 Introduction

For the London Olympics, the British Broadcasting Corporation (BBC) and Nippon Hoso Kyokai (NHK) held SUPER Hi-VISION (SHV) public viewings (hereinafter PV) at venues in Japan, the UK (Fig. 50), and the USA, in cooperation with the Olympic Broadcasting Service (OBS).

SHV is a next-generation broadcasting system being developed by NHK, which reproduces strong sensations of presence, as though the viewer was actually there, through the use of ultra-high-resolution, 33-megapixel video (7 680 pixels horizontally by 4 320 pixels vertically) and 22.2 multichannel audio (22.2ch audio).

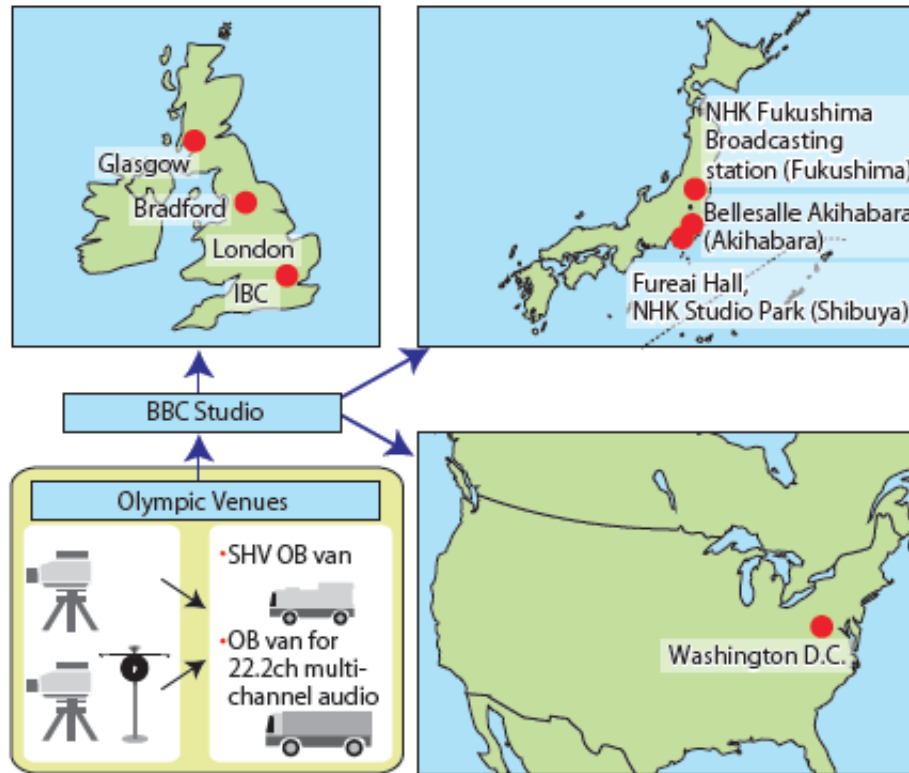
FIGURE 50
PV venue (Bradford, UK)



2 Overview

Content from seven events at the Olympic venues was produced. These were the opening ceremony, swimming, basketball, athletics, cycling, synchronized swimming and the closing ceremony. Content created at the venues was sent by optical fibre to a temporary production and transmission base established in TC0, a studio at the BBC Television Centre London. Live content was output as is from Television Centre while other content was edited and packaged. Both live and non-live programmes were compressed into transmission streams (TS) with a rate of approximately 300 Mbit/s, then distributed by IP lines to PV venues within the UK except for the International Broadcasting Centre (IBC), in the USA, and in Japan for presentation (Fig. 51).

FIGURE 51
System overview



An uncompressed signal was sent from the BBC Television Centre to a PV theatre in the IBC within the Olympic Park to hold PVs for broadcasting-related attendees gathered from around the world at the Olympic venues.

The opening and closing ceremonies and swimming competitions were presented live in the UK and the USA. Due to time differences, only the morning (local time) part of the July 30 session of the swimming competition was shown live in Japan.

3 Venue production

1) Outside broadcasting-van system (OB-van)

Two trucks were used at each location: one for video (Fig. 52), and the other for audio (Fig. 53). They were staffed with a single crew and after each event they had to de-rig the equipment before moving to the next venue.

FIGURE 52
OB-van for Video



FIGURE 53
OB-van for Audio



The video OB-van was a rental truck that was furnished with empty equipment racks before being shipped to London from Japan. Once there, it was rigged with the SHV production equipment to create the video OB-van. Production equipment included an eight-input switch system with resources including two SHV cameras (three at the opening and closing ceremonies), two SSD live-slow-motion devices, a host HD signal up-converter, and graphics equipment.

2) *Shooting*

Three cameras equipped with 1.25-inch four CMOS image sensors were used at the opening and closing ceremonies, this was reduced to two cover the competitions. Two sets of lenses were used, each included a 5×12 mm to 60 mm lens (Fig. 54) and a 10×18 mm to 180 mm lens. These were selected by venue.

FIGURE 54
1.25-inch 4-CMOS camera and 5x lens



One feature of SHV is the different viewing distance. The optimal viewing distance for ordinary Hi-Vision is three-times the screen height, but for SHV it is 0.75-times the screen height. This gives it a horizontal field of view of 100 degrees. SHV covers a wide viewing angle with high resolution images and produces an immersive sensation when viewed at close distances. With large screens greater than 400 inches, viewers get a sensation of being wrapped in the field of view, producing a sense of presence, even though they cannot recognize any information in the peripheral areas without consciously looking there. This meant shots had to be composed to place the most important action near the centre of the screen. The camera viewfinder had “action area” markings to help the camera operators maintain a good shot framing (Fig. 55).

FIGURE 55

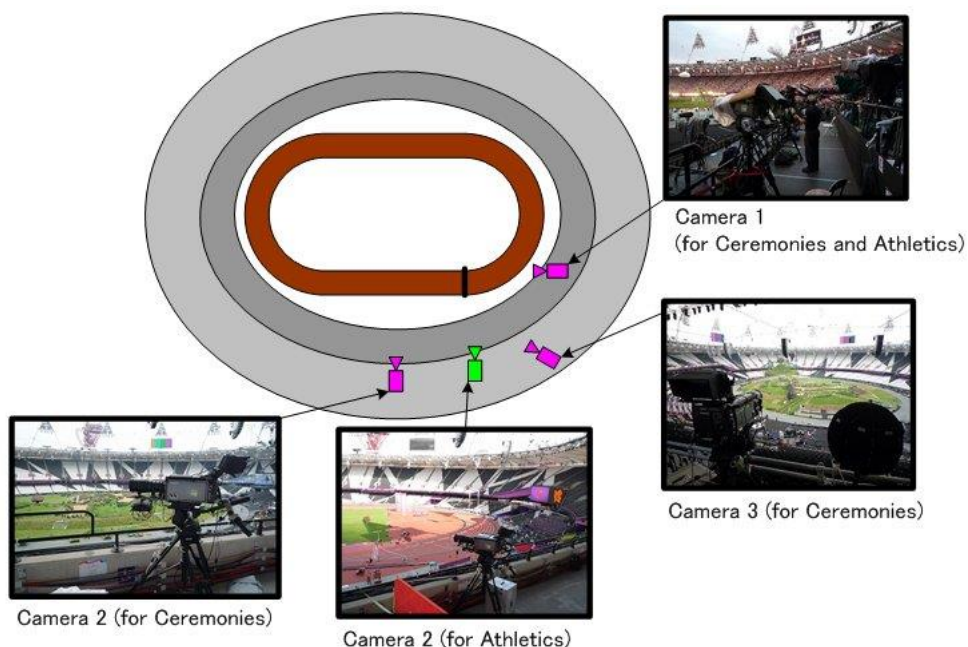
Viewfinder markers showing large screen



It was decided that each venue would have a master or base camera in a high position while the other camera would give a different angle from a low position and was primarily used for close-ups. Depending on the venue, the low-position camera gave more depth and was able to convey a greater sense of presence, so it was switched in and out with the base camera shot during each event. The camera positions for athletics and opening/closing ceremonies are described in Fig. 56.

FIGURE 56

Camera positions for athletics and opening/closing ceremonies



Care was also taken to make sure there was a link between the sense of presence given by the images and the sense of presence provided by the audio. An audio monitoring environment with 5.1-channel audio down-mixed from the 22.2 ch audio was provided in the OB-van production room, so that video switching could complement the sounds from the venue and from the cheering crowds.

Due to space limitations in the video monitoring environment, work was done with 28-inch 4K monitors, so technicians had to struggle with a perceptual gap between the monitoring in the OB-van and the actual viewing environments at the large PV screens. In order to check the content in a real

viewing environment, the OB crew had to view the 145-inch PDP display in the IBC theatre, or go to the PV venue in London to view the 300-inch screen there.

3) *Audio*

A temporary, locally rented, audio OB-van (Fig. 53) was used. It was installed with a 22.2 ch live mixing board and 22.2 ch speakers (Fig. 57). The mixing desk was developed at NHK Science & Technology Research Laboratories (NHK STRL) with operations optimized for live production and incorporating 3D sound-image positioning (3D panning) functions.

A 22.2 ch-audio, one-point microphone (Fig. 58) was used. It was a fixed, monolithic microphone holder consisting of a 45-cm sphere divided into upper, middle and lower layers. Each layer was partitioned into eight directions using sound baffles and a compact microphone installed in each partition. The microphone was positioned near the base camera, and by panning to arbitrary locations on the sound desk (using the outputs from each microphone) a sound field similar to that heard by the event audience could be reproduced. It was not permitted to place microphones freely for SHV production in the venues, so audio was produced using the one-point microphone as a base, and mixing it with the microphone feeds distributed from the international feed. Use of 3D reverb equipment (22.2 ch audio reverb) also helped in reproducing the expansive feeling in the venues faithfully.

The 22.2 ch output feeds were recorded using a multi-track recorder in case the signals from live PVs are used for other purposes later.

FIGURE 57

Inside audio OB-van



FIGURE 58

22.2ch one-point microphone



4) *Other issues*

Presentations at the IBC on-site theatre also gathered large audience, suggesting interest in SHV increased daily amongst overseas broadcasters. Within the venue compounds, people from the media visited in groups to see the video and audio OB-vans. As promoting SHV was one of the objectives of this SHV trial, in addition to holding the PVs, visitors were welcomed and given explanations of SHV video and 22.2 ch audio. Visiting broadcaster staff (Fig. 59) came to see the high resolution video and explanations of SHV in the OB-van were done using 4K monitors. In the audio OB-van, visitors were able to sit in the main-mixer seat and experience 22.2 ch sound.

FIGURE 59

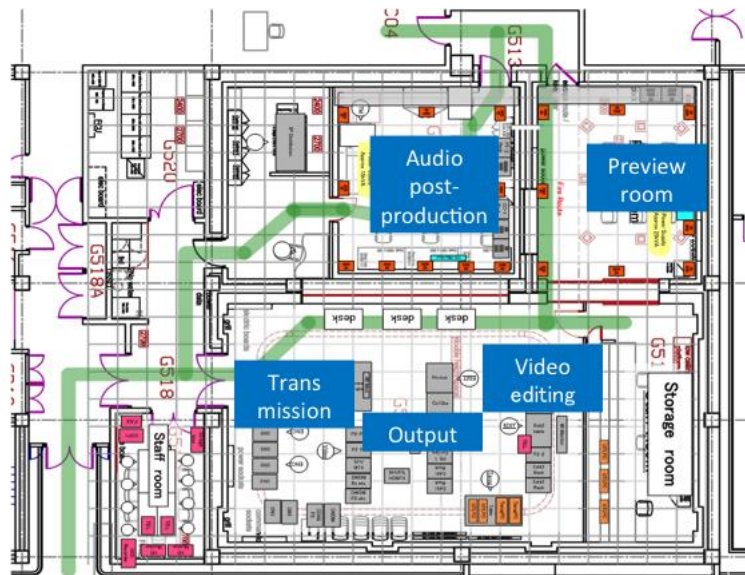
Broadcasting staff viewing the monitor beside the OB-van

4 **Transmission and editing**

For production of the SHV PVs at the London Olympics, transmission, play-out, video editing, audio post-production, and a monitor room were combined in a compact layout in the BBC TC0 studio (Fig. 60). Much consideration was given to the layout of cabling and provision of quality monitoring.

FIGURE 60

BBC TC0 studio layout



1) *Line and transmission systems (Fig. 61)*

An SHV 8×8 routing switcher system (an HD 128×128 router configured to switch 16 channels at a time) was used to access the various resources in TC0 effectively, including the live signals (main and backup) brought in from each venue by optical lines, and the P2 recorders (main and backup) for playing back content. An additional monitoring system was also built separately from the main system to allow the signals from the relay site to be checked even during a PV (Fig. 62).

FIGURE 61
Line and transmission equipment



FIGURE 62
8K-4K down-converter and 28-inch 4K screen for monitoring (left)



2) *Video post-production*

For competitions that were not covered live, the main and backup feeds were recorded using the P2 recorders in the TC0 studio. After the competition, two SHV editors were used to produce content for the next day. Two versions were edited simultaneously, a World Feed (WF) for the UK and USA, and a Japan Feed (JF) for Japan. Segments of approximately ten minutes were edited according to medal standings or popularity, and several events were combined into a package of approximately 45 minutes duration. Editing work was done overnight from midnight to 10 am, for advance distribution to Japan and presentation in the UK and USA starting at noon. The SHV recorder used 17 P2 cards for recording. One of these was used to record HD proxy data.

This recording was then used with an off-line HD proxy editor for editing (Fig. 63). The SHV video was later matched to the HD proxy using this editing data (Fig. 64). Captions were created in 8K ($7\,680 \times 4\,320$) TIFF files using graphics overlay equipment and imported to the SHV editor to create video, and a 24-channel WAV file of the completed audio master was imported and merged into the SHV editor time line to complete the content.

FIGURE 63
HD proxy editor



FIGURE 64
SHV editor



3) *Audio post-production*

For live transmissions, the on-site audio mix was transmitted to the PV venues, but for edited productions, a 22.2 ch multichannel audio post-production process had to be undertaken. Two simultaneous audio edits were produced (to match the two video edits) by using the HD proxy editing data. Two audio post-production systems were built that included a digital audio workstation (DAW) with a mixing desk that supports 22.2 ch audio and a backup DAW.

These were set up in the audio mixing room (Fig. 65). The audio editing systems also used the 22.2 ch sound speaker system (Fig. 66) in the SHV preview room (see next section). Audio post-production made it possible to manage the overall audio balance for multiple events, improving the quality of the content.

FIGURE 65
Main mixing room



FIGURE 66
Sound equipment in the preview room



4) *SHV preview room*

A preview room was created using an 85-inch SHV LCD monitor and a 22.2 multichannel audio system to enable video and audio signals for live transmissions and the edited content to be previewed reliably (Fig. 67). This was primarily to manage the quality of the signals, but it also allowed viewing on a large monitor at the ideal distance and with the ideal speaker layout. This was used to give advice to on-site teams on aspects such as the camera work and switch timing in order to increase the sense of presence and immersion given by SHV.

FIGURE 67
SHV preview room



5 Distribution and screening

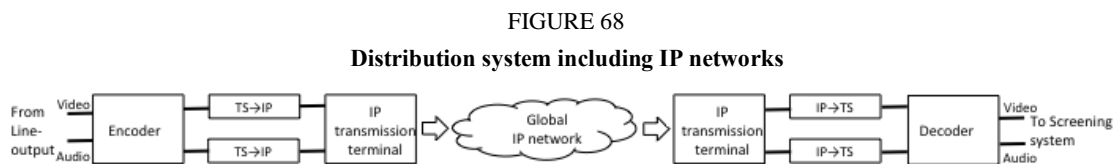
1) Distribution and screening overview

Distribution and screening refers to delivery of the packaged content and live signals from the BBC studio to each of the PV venues and the presentation at the venues. Within the overall system shown in Fig. 51, this section corresponds to the part after the BBC studio and includes encoding of the baseband signals from the BBC studio equipment, transmission over IP, the transport system that decodes and restores the baseband signals and the screening system that reproduces the video and audio signals from the baseband signals.

IP networks within the UK were used to transmit signals to the three locations (London, Bradford, and Glasgow). Dark fibre was used to transmit uncompressed signals to the IBC within the Olympic complex. Global IP networks were used to transmit to Washington in the USA, and to Tokyo in Japan. The domestic network in the USA was used to transmit from the terminus of the global network to the venue within the USA. From the Tokyo terminus, signals were transmitted to NHK STRL by domestic lines and then distributed by IP multicast over dedicated domestic lines to three locations in Japan (Shibuya, Akihabara, and Fukushima).

2) Distribution system

A system diagram of the distribution system [1][2], consisting of encoder, transport equipment and IP network, is shown in Fig. 68.



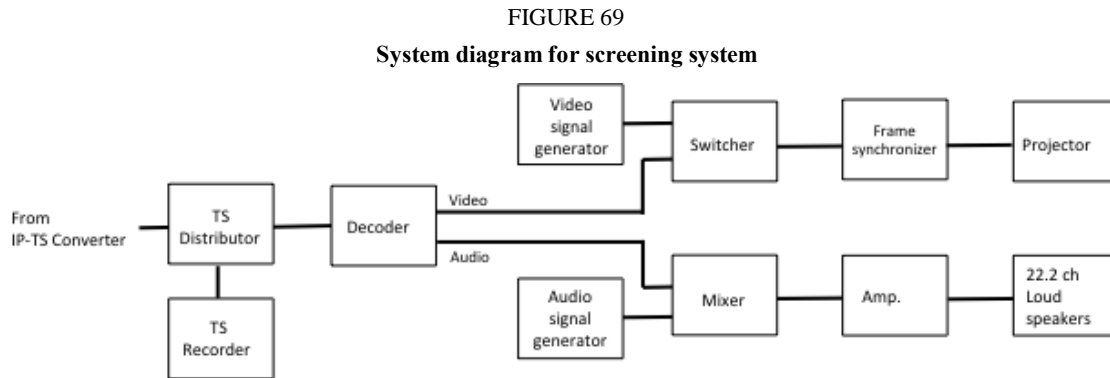
The encoder [3] used AVC/H.264 to encode the video and MPEG-2 AAC-LC to encode the audio. The interfaces between SHV equipment converted the so-called dual-green format equivalent to the RGB Bayer matrix signals to ordinary YUV signals that can be compressed using AVC/H.264. In this process, the green signal composed with diagonally-pixel-offset two-channel signals with $3\,840 \times 2\,160$ pixels is converted to the $7\,680 \times 2\,160$ format by shifting the one of the two signals by half pixel. Then Y signal with $7\,680 \times 2\,160$ pixel is calculated from this new green signal and up-converted red and blue signals. The YUV signals were partitioned into eight $1\,920 \times 1\,080$ pixel signals (1 080-60P signals) for handling. An encoder for each of these elementary signals compressed the 1 080-60P signal using AVC/H.264, and up to four audio channel signals using MPEG-2 AAC-LC. Finally, the MPEG-2 TS signals output by the encoder units were multiplexed into two MPEG2-TS signals by TS multiplexers. After the MPEG-2 TS signals output by the encoder equipment were converted to IP signals, they were transmitted over IP networks.

In order to transmit the 300 Mbit/s compressed MPEG-2 TS signal SHV video and audio over global IP networks, the required bandwidth had to be secured, fluctuations that occurs on the network had to be compensated for, and a mechanism for maintaining security was needed. Specifically, functions to control jitter for synchronous transmission, for real-time encoding and decoding, to handle packet loss, and for advanced error correction [4] were needed. These functions were built into the IP transmission terminal equipment.

The decoding equipment performed the same processing as the encoding equipment, in reverse, to produce the baseband video and audio signals.

3) Screening system

A system diagram of the screening system is shown in Fig. 69. Content was distributed from NHK STRL to the PV venues in Japan, and from the BBC Studio TC0 to the PV venues in the UK and the USA. The video and audio signals were decoded then presented using SHV projectors with 22.2 ch audio equipment at each venue. Each venue was also equipped with TS recording and playback equipment. This was to be used as a backup system if the lines between NHK STRL and each venue experienced difficulty or, if for any reason any of the venues needed to make their own playback schedule. A baseband signal generator is also built into the system to enable adjustments to the audio and video systems.



For video display two types of projector were used as well as an 85-inch LCD, a 145-inch PDP, and a 360-inch multi-screen LCD. At five of the theatre venues the SHV projectors [5] used had 8-megapixel display devices and a pixel-shifting technology called e-shift to increase the resolution, resulting in compact size and low power consumption. The projector in the NHK Minna no Hiroba Furea Hall used 33 megapixel display devices for RGB [6], making it a so-called full-resolution projector. This projector has a high output power so was used as here as the venue has a 520-inch large screen.

The 85-inch LCD [7] was used for direct viewing in the Washington venue, and also in the Akihabara and Fukushima venues. The 85-inch display was combined with a sound system to demonstrate a possible home system. The 145-inch PDP [8] was used to set up a theatre in a room at the IBC and the 360-inch multi-screen LCD was installed at the entrance to the NHK Studio Park tour for NHK visitors.

A new 22.2 multichannel audio system was developed for direct viewing displays and combined with the 85-inch LCDs and the 145-inch PDP. In theatres using a projector, the theatre sound system was used.

4) *Screening venues and reactions*

TABLE 21

List of screening venues

Country	City	Venue	Screen size	Display device
Japan	Tokyo	NHK Minna no Hiroba Fureai Hall	520	Projector
		NHK Studio Park	360	Multi-screen LCD
		Belle Salle Akihabara	300	Projector
	Fukushima	NHK Fukushima broadcasting station	350	Projector
UK	London	BBC Broadcasting House	300	Projector
		IBC	145	Plasma
	Bradford	National Media Museum	250	Projector
	Glasgow	BBC Pacific Quay	350	Projector
US	Washington DC	Comcast (NBC)	85	LCD

The display equipment and screen size used in each PV venue are shown in Table 21.

PVs were held in Japan from July 28 to August 12 at the NHK Minna no Hiroba Fureai Hall in Shibuya (Fig. 70), NHK Studio Park, in the Belle Salle Akihabara event space near the JR Akihabara station, and at the NHK Fukushima broadcasting station. At each venue several related events were held at the same time as the screening of Olympic content in order to attract many more people and to ensure that those who came really enjoyed it. The Shibuya venue catered to families during the summer holidays, the Akihabara venue to young people and those interested in technology, and the Fukushima venue appealed to supporters of local athletes. A total of over 200 000 people attended over all three events. At the venues, there were exclamations of wonder as the screening began, and many comments about the strong sense of presence and intensity of the presentation.

FIGURE 70

PV in the Fureai Hall

In the UK, the BBC took the lead with NHK cooperating, and screenings were held from July 23 to August 12 at the BBC Broadcasting House in London, at the National Media Museum in Bradford, and at the BBC Scotland building in Glasgow. At the IBC, the OBS took the lead providing demonstrations mainly for broadcasting-related attendees. Before the opening ceremony, a short item

introducing SHV was screened along with a NHK and BBC co production featuring sights and sounds around London and the Olympic venues in the run-up to the games. Many dignitaries and people related to broadcasting visited during the screenings, and there were many comments about how wonderful they were, and inquiries of when broadcasting would begin.

In the USA, NBC took the lead with cooperation from NHK, and screenings were held from July 27 to August 12 in meeting rooms in the Comcast building in Washington DC, mainly by invitation to people from government and the content, communications and electronics industries.

6 Conclusion

We have advanced development of SHV with “presence” as its strongest feature, and these events have once again shown the extremely strong sense of presence delivered by SHV video and audio, and the unprecedented levels of emotion can be imparted on viewers, giving them a sense they are actually at the Olympic venue. We also showed SHV can operate much like ordinary broadcasting, by producing and transmitting programmes continuously, every day during the Olympic Games using live coverage and recorded and edited content.

A completely different style was also used in production of the content, without using voice-overs (announcing or comments), and using mainly wide camera angles and long (slow) cut ratios. These were met with many comments of surprise, admiration, and of the new possibilities presented for broadcasting businesses.

Annex 4

Implementation of UHD TV broadcasting

1 Japan

1.1 UHD TV satellite broadcasting services

UHD TV satellite broadcasting has been operating in Japan on the basis of the domestic standards developed in 2014 and additional specifications for HDR-TV issued in 2016. There are two platforms for UHD TV satellite broadcasting in Japan, which are called BS and CS. In March 2015, 4K broadcasting via CS was commenced by SKY Perfect JSAT Corporation. 4K and 8K test broadcasting via BS was launched by NHK (Japan Broadcasting Corporation) in August 2016 and was taken over by the joint operation by NHK and A-PAB (Association for Promotion of Advanced Broadcasting Services) in December 2016. 4K and 8K regular broadcasting services via BS started on December 1, 2018. As of October 2020, nine channels of 4K and one channel of 8K broadcasting services via BS, and eight channels of 4K services via CS are in operation.

1.2 Specifications for UHD TV satellite broadcasting

Japanese domestic standards for broadcasting consist of Ministerial Ordinances and Notices and ARIB (Association of Radio Industries and Business) Standards and Technical Reports. They have been developed on the basis of international standards including ITU-R Recommendations, ITU-T Recommendations, and ISO/IEC International Standards.

Figure 71 illustrates the protocol stack of the ISDB-S3 system and Table 22 shows the specifications for UHDTV satellite broadcasting.

The UHDTV video signal and 22.2-channel audio signal are encoded by ITU-T H.265|MPEG-H HEVC and MPEG-4 AAC, respectively. As for closed captions and multimedia content, Timed Text Markup Language (TTML) and HTML 5 are used for the coding, respectively.

These coded signals are encapsulated into MPEG Media Transport (MMT), which is a media transport protocol over IP. IP packets constructed like this are multiplexed over the broadcast channel with the TLV multiplexing scheme.

FIGURE 71
Protocol stack of ISDB-S3 system

4K·8K HLG video	22.2-ch audio	Closed caption	Multimedia content
HEVC	AAC, ALS	TTML	HTML5
MMT/IP			
TLV multiplexing scheme			
ISDB-S3 channel coding & modulation			

TABLE 22
Specifications for UHDTV satellite broadcasting in Japan

		Wide-band bandwidth: 34.5 MHz	Narrow-band bandwidth: 27 MHz
Video signal	Format	8K: 7680×4320/59.94/P, 4K: 3840×2160/59.94/P, (2K: 1920×1080/59.94/P)	4K: 3840×2160/59.94/P, (2K: 1920×1080/59.94/P)
		ARIB STD-B32 ^{*1} , Rec. ITU-R BT.2020, Rec. ITU-R BT.2100	
	Colorimetry	Wide colour gamut	
		ARIB STD-B32, Rec. ITU-R BT.2020, Rec. ITU-R BT.2100	
	Transfer function	HLG, (PQ), SDR	
ARIB STD-B32, Rec. ITU-R BT.2100			
Coding	H.265 HEVC Main 10 profile		
	ARIB STD-B32, Rec. ITU-T H.265 ISO/IEC 23008-2, Rec. ITU-R BT.1870, Rec. ITU-R BT.2073		
Audio signal	Speaker configuration	22.2, 7.1, 5.1, 2.0	5.1, 2.0
		ARIB STD-B32, Rec. ITU-R BS.2051	
	Coding	MPEG-4 AAC LC, MPEG-4 ALS	MPEG-2 AAC
ARIB STD-B32, Rec. ITU-R BS.1196			
Multimedia	HTML5		None
	ARIB STD-B62 ^{*2} , Rec. ITU-R BT.2075, Rep. ITU-R BT.2342		
Transport	MMT/TLV, (MPEG-2 TS)		MPEG-2 TS
	ARIB STD-B32, ARIB STD-B60 ^{*3} , Rec. ITU-R BT.2074, Rec. ITU-R BT.1869		ARIB STD-B10 ^{*4} , ARIB STD-B32, Rec. ITU-T H.222.0 ISO/IEC 13818-1
CAS	AES, (Camellia)		Multi 2
	ARIB STD-B61 ^{*5} , Rec. ITU-R BT.1852		Private
Transmission	ISDB-S3		DVB-S2
	ARIB STD-B44 ^{*6} , Rec. ITU-R BO. 2098		Rec. ITU-R BO.1784
Receiver configuration	ARIB STD-B63 ^{*7}		ARIB STD-B1 ^{*8} , ARIB STD-B16 ^{*9}
Operational guidelines	ARIB TR-B39 ^{*10}		Private

NOTE 1: Parameter values in parentheses are neither specified in the operational guidelines nor currently used in the broadcasting.

NOTE 2: Superscripts indicate the relevant standards shown in Attachment 1.

1.3 List of ARIB Standards⁹

- 1) ARIB STD-B32, “Video coding, audio coding and multiplexing specifications for digital broadcasting”,
http://www.arib.or.jp/english/html/overview/doc/6-STD-B32v2_1-E1.pdf
- 2) ARIB STD-B62, “Multimedia coding specification for digital broadcasting (2nd generation)”,
http://www.arib.or.jp/english/html/overview/doc/6-STD-B62v1_2-1p2-E1.pdf
- 3) ARIB STD-B60, “MMT-based media transport scheme in digital broadcasting systems”
- 4) ARIB STD-B10, “Service information for digital broadcasting system”,
http://www.arib.or.jp/english/html/overview/doc/6-STD-B10v4_6-E2.pdf
- 5) ARIB STD-B61, “Conditional access system (second generation) and CAS program download system specifications for digital broadcasting”
- 6) ARIB STD-B44, “Transmission system for advanced wide band digital satellite broadcasting”,
http://www.arib.or.jp/english/html/overview/doc/6-STD-B44v2_0-E1.pdf
- 7) ARIB STD-B63, “Receiver for advanced wide band digital satellite broadcasting”
- 8) ARIB STD-B1, “Digital receiver for digital satellite broadcasting services using communication satellite”
- 9) ARIB STD-B16, “Digital receiver commonly used for digital satellite broadcasting services using communication satellites”
- 10) ARIB TR-B39, “Operational guidelines for advanced digital satellite broadcasting”

References

- [1] Nojiri, Iguchi, Noguchi, Fujii, Ogasawara: “National Super Hi-Vision Transmission Test using IP Networks for Global Research and Education”, Broadcast Technology (Hoso Gijutsu), Vol. 64, No. 6, pp. 135-141 (2011) (Japanese).
- [2] Sakaida, Iguchi, Kimura, Ogasawara, Fujii: “International Super Hi-Vision Transmission Test and Exhibition of Related Equipment at IBC2011”, Broadcast Technology (Hoso Gijutsu), Vol. 65, No. 1, pp.151-156 (2012) (Japanese).
- [3] Y. Shishikui, K. Iguchi, S. Sakaida, K. Kazui, A. Nakagawa: “Development of High Performance Video Codec for Super Hi-Vision”, 65th NAB Broadcast Engineering Conference, pp. 234-239 (2011).
- [4] Y. Tonomura, D. Shirai, T. Nakachi, T. Fujii, H. Kiya: “Layered Low-Density Generator Matrix Codecs for Super High Definition Scalable Video Coding System”, IEICE Trans. on Fundamentals, Vol. E92-A, No. 3, pp. 798-807 (2009).
- [5] F. Okano, M. Kanazawa, Y. Kusakabe, M. Furuya, Y. Uchiyama: “Complementary Field Offset Sampled-Scanning for GRB Video Elements”, IEEE Trans. Broadcasting, Vol. 58, No. 2, pp. 291-295 (2012).
- [6] T. Nagoya, T. Kozakai, T. Suzuki, M. Furuya, and K. Iwase, “The D-ILA device for the world’s highest definition (8K4K) projection system”, Proc. Int. Display Workshop (IDW) 2008, Vol. 15, pp. 203-206 (2008).

⁹ Some of the ARIB Standards are available in English and free of charge.

- [7] T. Kumakura, M. Shiomi, S. Horino, Y. Yoshida, S. Mizushima: “Development of Super Hi-Vision 8Kx4K Direct-View LCD for Next Generation TV”, SID 2012 Digest, pp. 780-783 (2012).
- [8] K. Ishii, T. Usui, Y. Murakami, Y. Motoyama, M. Seki, Y. Noguchi, T. Furutani, T. Nakakita, T. Yamashita: “Developments of a 145-inch Diagonal Super Hi-Vision Plasma Display Panel”, SID 2012 Digest, pp. 71-74 (2012).


Annex 5

Initiatives for 8K/120 Hz broadcasting


Various pieces of production equipment as well as transmission and display technologies for high frame frequency 8K programmes have been developed to achieve much higher quality with clearer and smoother subject movement (Fig. 72). A complete system ranging from live on-site production to encoding, satellite transmission, and display reproduction at a frame frequency of 120 Hz (120/1.001 Hz) was exhibited at the NHK STRL Open House 2019 in cooperation with Broadcasting Satellite System Corporation (Fig. 73).

FIGURE 72
8K/120 Hz equipment


Production




(a) High-speed camera




(b) Video mixer




(c) Mezzanine compression IP transceiver




(g) Online editor



(d) Compact camera




(e) Compression Recorder




(f) 120-Hz HDR 31.5-inch LCD

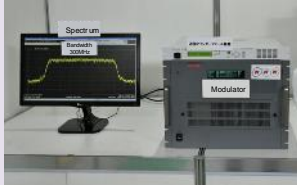
Delivery




(g) Video/audio encoder



(h) Video/audio decoder




(i) Wideband modulator




(j) Wideband demodulator

Reproduction



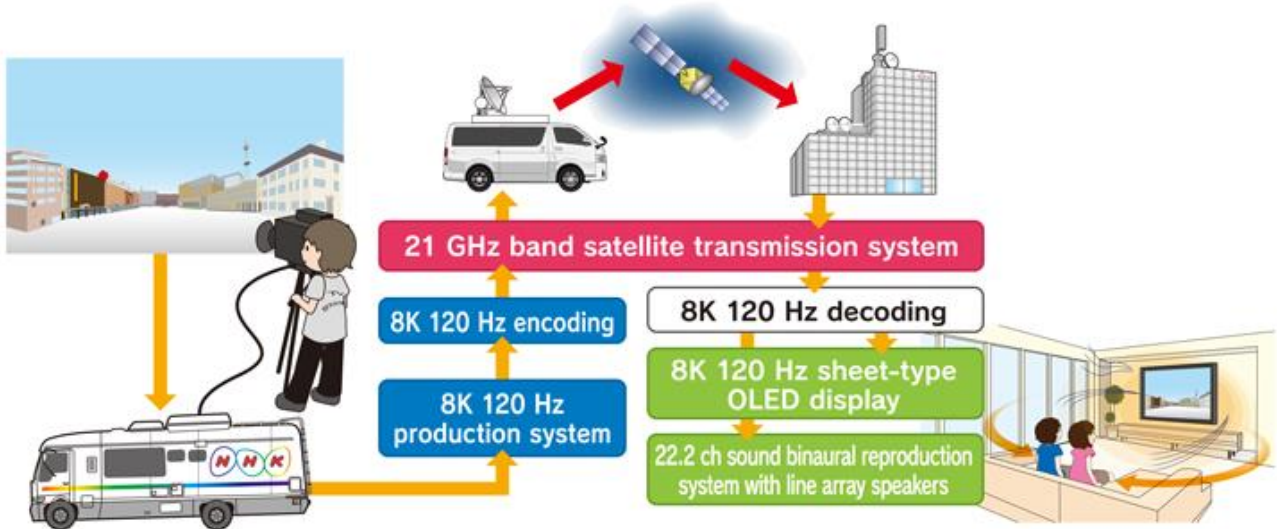
(k) Projector



(l) Sheet-type OLED display

FIGURE 73

8K/120 Hz live broadcast experimental system



1 Development of high frame frequency 8K equipment

1.1 High speed camera

An 8K/240 Hz full-resolution high-speed camera (Fig. 72(a)) and slow-motion replay server system for the creation of sports content have been developed. The high-speed camera acquires 8K/240 Hz video via three 1.25-in 33-Mpixel 240-Hz CMOS image sensors. Signals are transmitted between a camera head and a camera control unit (CCU) via a four-core optical camera cable, and the CCU then processes the sensor signal and outputs 8K/240 Hz RGB 4:4:4 video signals. The 8K/240 Hz slow-motion replay server system can record up to 240 min of 8K/240 Hz video while simultaneously playing back 8K/60 Hz slow-motion video. The input 8K/240 Hz video is coded for bit-rate reduction by a factor of 12 and stored on solid-state drives. The replay server offers loop-recording, live editing, and slow-motion replay functions with a dedicated remote controller. The specifications of the high-speed camera and slow-motion replay server are shown in Tables 23 and 24, respectively.

TABLE 23

Specification of high speed 8K camera

Optical size	1.25-in
Image sensor	33-Mpixel CMOS
Active pixel count	7680(H) × 4320(V)
Frame frequency	239.76 (240/1.001), 119.88 (120/1.001), 59.94 (60/1.001) Hz
Colour gamut	Rec. ITU-R BT.2020
HDR system	Hybrid Log-Gamma as per Rec. ITU-R BT.2100
Transmission between head and CCU	Head to CCU: 448 Gbit/s (1310 nm) CCU to Head: 40 Gbit/s (1550 nm)
Signal output	U-SDI×2ch (RGB 4:4:4) as per Rec. ITU-R BT.2077 Part 2
Sensitivity	F2.8 / 2000 lux
S/N	50 dB

TABLE 23 (*end*)

Dynamic range	450%
MTF	12% at 3200 TVL
Size	Head: 162 (W) × 206 (H) × 353 (D) mm CCU: 8RU (19-in rack size)
Weight	Head: 18 kg, CCU: 37 kg
Power consumption	Head: 300 W, CCU: 750 W

TABLE 24

Specification of slow-motion replay server

Video format	Input: 8K/240 Hz RGB 4:4:4 Output: 8K/60 Hz YC _B C _R 4:2:2
Signal input	U-SDI × 2
Signal output	12G-SDI × 4
Compression ratio	1/12 (intraframe compression)
Recording media	12 TB SSD
Recording time	240 min
Replay latency	Less than 10 frames
Size	5RU (19-in. rack size)
Functions	Instant replay, loop recording, variable replay speed, live editing
Power Consumption	800 W

1.2 Studio Monitor

A 31.5-inch, 8K/120Hz monitor was developed for the studio production. The appearance is shown in Fig. 72(f) and the specifications are given in Table 25. The monitor has a dynamic range of more than 100,000:1 and maximum brightness of more than 1000 cd/m² by controlling the backlight with local dimming.

TABLE 25

Specification of studio monitor

Screen size	31.5-in
Number of pixels	7680 × 4320
Frame frequency	119.88 (120/1.001) Hz
Bit depth	12-bit equivalent
EOTF	HLG, PQ, SDR
Maximum brightness	1000 cd/m ²
Dynamic range	100,000:1
Coverage of colour gamut	85% (Rec. ITU-R BT.2020)

1.3 Video/audio encoder and decoder

An 8K/120 Hz real-time video codec using high-efficiency video coding (HEVC) has been developed that complies with 8K broadcasting standards in Japan (see § 1.2 of Annex 4). The encoder processes 8K/120 Hz video by 12 4K/60 Hz sub-encoders (four and three partitions in the spatial and temporal directions, respectively). Video and audio decoders were implemented in a PC using highly parallelised software. The codec has achieved high image quality with a bit rate of 85 Mbit/s, which is equivalent to the bit rate of 8K/60 Hz broadcasting, and can run at higher bit rates of up to 400 Mbit/s. The appearance of the 8K/120 Hz codec is shown in Fig. 72(g) and (h), and the specifications of the encoder and the decoder are presented in Tables 26 and 27, respectively.

TABLE 26

Specification of 8K/120 Hz video/audio encoder

Video	Coding scheme	MPEG-H HEVC ITU-T H.265 Main 10 profile (4:2:0 / 10 bits), Level 6.2
	Format	7680 × 4320/119.88 Hz (120/1.001 Hz)
Audio	Coding scheme	MPEG-4 AAC Low complexity profile
	Format	22.2 ch / 24 bits / 48 kHz
Multiplexing	Multiplexing scheme	MPEG-2 TS SMPTE ST 2022-2

TABLE 27

Specification of decoder

CPU	2 × Intel Xeon E5-2669 v4 (2.20 GHz, total 44 cores)
Main memory	DDR4-2400 ECC Registered 128 GB
Graphics board	2 × NVIDIA Quadro P6000

1.4 Sheet-type flexible OLED display

An 8K/120 Hz sheet-type flexible OLED display was developed for the easy enjoyment of ultra-high-definition, highly realistic images at home. The appearance of the display is shown in Fig. 72(l) and the specifications are shown in Table 28.

TABLE 28

Specification of 8K/120 Hz sheet-type OLED display

Screen size	88-in
Number of pixels	7680 × 4280
Frame frequency	119.88 (120/1.001) Hz

2 Live production and transmission

In the live production and transmission of 8K/120 Hz programmes, low-latency, mezzanine compression IP transmission was used to transmit audio and video from the live venue to the production centre, where the transmitted video was edited in real time by online equipment and 22.2-channel audio mixing was performed. The appearance and specifications of the IP transmission are demonstrated in Fig. 72(c) and Table 29, respectively. The completed 8K/120 Hz content was coded using the HEVC codec at 250 Mbit/s for transmission via 21-GHz band satellite. The wideband modulator and demodulator (Fig. 72(i) and (j)) have been developed for 21-GHz band satellite transmission. The BSAT-4a satellite, which is equipped with a 21-GHz-broadband, high-capacity transponder was used. The specifications of the satellite transmission are shown in Table 30. The live video transmitted from the venue was presented on the 88-inch sheet-type OLED display and the 22.2-channel sound was reproduced by a binaural reproduction system using line-array speakers.

TABLE 29

Specification of IP transmission

Video	Coding scheme	TICO (Tiny Codec)
	Format	7680 × 4320/119.88 (120/1.001) Hz
	Compression ratio	1/6
	Transmission bandwidth	24 Gbit/s
	Error correction	FEC (Exclusive OR)
	Interface	4 × 10 GbE
Audio	Format	2 × 64ch / 24 bits / 48 kHz (bidirectional)
	Transmission bandwidth	250 Mbit/s
	IP protocol	AES67
	Interface	Gigabit Ethernet

TABLE 30

Specification of 21-GHz satellite transmission

Satellite	BSAT-4a
Uplink frequency	18.24375 GHz
Downlink frequency	21.84375 GHz
Diameter of uplink and downlink antennas	1.5 m
Bandwidth of satellite transponder	300 MHz (21.7-22.0 GHz)
Symbol rate	250 Mbaud
Roll-off factor	0.1
Modulation scheme	QPSK
Error correction (code rate)	LDPC (2/3)
Transmission capacity	327.726 Mbit/s
Interface	Gigabit Ethernet

Annex 6

Practice of 8K UHD programmes broadcasting on large screens in China

1 Overview

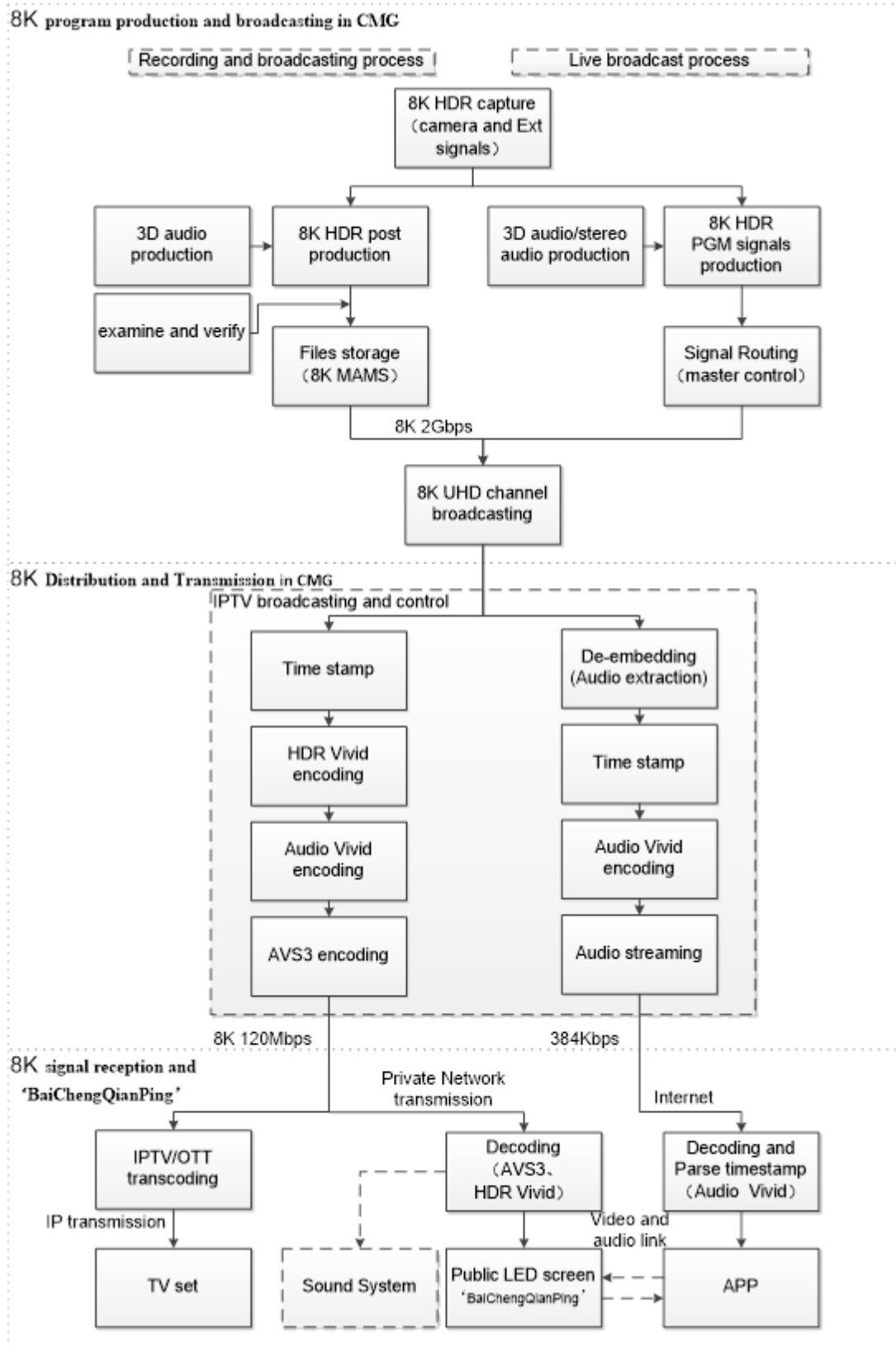
In October 2021, China Media Group (CMG) began to launch a three-year “Bai Cheng Qian Ping” (100 cities, 1,000 screens) UHD video and 3D audio promotion activities, cooperating with IT, radio and television, culture, tourism, transportation and other industries. In January 2022, CMG launched a specific 8K UHD channel to provide 8K programmes for the event. The campaign mainly showcased the UHD video and 3D audio-visual effects through the outdoor and indoor UHD large screens in public, using multicast technology to distribute CMG 8K programmes over IP to more than 70 cities, realizing the display of more than 310 UHD large screens for the Beijing Winter Olympics.

2 8K UHD video and 3D audio broadcasting & distribution system workflow

By promoting industrialization in 2022, China has now fully built the end-to-end all-elements 8K production and broadcast chain, and realized the technical support of the whole workflow of 8K content from capture, recording, production, signal routing, broadcasting, encoding and transmission to terminal displays. The 8K UHD signal broadcasting-distribution workflow is shown in Fig. 74.

FIGURE 74

8K UHD signal broadcasting-distribution workflow



3 8K programme production and broadcasting in CMG

3.1 Basic technical specifications for 8K UHD programme video and audio

Basic technical specifications for 8K UHD programme video and audio refers to Table 31.

TABLE 31

Basic technical specifications for 8K UHD programme video & audio

No.	Item	Technical specification
1	Picture aspect ratio	16:9
2	Pixel count Horizontal × vertical	7 680×4 320
3	Sampling lattice	Orthogonal
4	Pixel aspect ratio	1:1 (square pixels)
5	Pixel addressing	Pixel ordering in each row is from left to right, and rows are ordered from top to bottom.
6	Frame frequency	50 Hz
7	Scan mode	Progressive
8	Quantization levels	10bit
9	Color gamut	BT.2020
10	HDR	HLG 1 000 cd/m ²
11	Sampling	4:2:2
12	Audio encoding	PCM 24 bit
13	Audio sampling frequency	48 kHz
14	Channel	Supports 16-channel PCM audio
15	File format	MXF OP-1a
18	8K video coding format	JPEG XS High 4:4:4 12 Profile as defined by ISO/IEC 21122-2
19	8K video encoding bitrate	2 073.6 Mbit/s

3.2 8K UHD programme video and audio production requirements and specifications

3.2.1 8K UHD programme video and audio production requirements

The 8K UHD studio/ outside broadcasting van generally uses IP system to produce 8K UHD signals, realizing the whole workflow of 8K UHD production functions from signal source, graphics, subtitle, real-time monitoring, to program recording and time-delayed broadcasting. The external signal in the studio comes from the 8K HDR signal provided by the master control, or the 4K signal provided by the master control and other 4K signal sources are used after up-conversion in the studio.

Post-production includes functions such as 8K signal recording and storage, editing and production, colour grading and rendering. In CMG, 8K editing production integrated network system has the ability to co-product 8K HDR programme, with automated colour management, 8K multi-format mixing, high-performance 8K real-time editing, 3D audio and surround sound production, etc., realizing 8K multi-format high-code mixing and real-time output with different 8K encoding and different dynamic range. 8K colour grading and rendering has the function of colour reproduction

correction for the characteristics of different types of 8K materials. The 8K UHD programme files from the post-production system are reviewed and stored in the MAMS for programme broadcasting, or directly delivered to 8K UHD TV broadcasting system.

The audio production uses surround sound/3D audio and achieves PCM audio signal transmission of surround sound and 3D audio over IP in the live broadcast workflow. Surround sound/3D audio production picks up signals such as special effects, sound effects, commentary and hosting, and mix them together with playback sounds and other signals. With the live broadcast mixer, adjust the location, distance and other information of sound elements to create PGM signals, and transmit them to the master control with 8K video.

After surround sound/3D audio are recorded in the studio or outfield, they are pushed to the video and audio post-production platform in the form of files for post-production. The platform is based on the digital audio workstation, and is used to adjust the frequency, dynamics, location, distance and other information of sound elements, create the surround sound/3D audio files, which are pushed to the MAMS or broadcasting system with the video files.

3.2.2 8K UHD programme production basic specifications

3.2.2.1 Basic technical specifications of 8K video signals and files

Basic technical specifications of 8K UHD signal and programme files refers to Table 32.

TABLE 32

Basic technical specifications of 8K video signals and files

No.	Item	Channel
1	Video format	7 680 × 4 320/50/P
2	Sampling	4:2:2
3	Bit depth	10 bit
4	Coding format	JPEG XS High 444 12 Profile
5	Horizontal wavelet transform layers	5
6	Vertical wavelet transform layers	2
7	Dequantization method	Uniform quantizer, Qp _{ih} = 1
8	Level	8k-2
9	Sub Level	Sublev3 bpp
10	Coding method	Intra-frame encoding with a fixed frame length
11	Compression ratio	16:1
12	Bitrate	2 073.6 Mbit/s

3.2.2.2 Audio production system basic requirements

3.2.2.2.1 Surround sound production

Surround sound audio production technical specifications refer to Table 33, audio level and loudness specifications refer to Table 34, channel allocation of surround sound refer to Table 35.

TABLE 33

Technical specifications of surround sound production

No.	Item	Technical specifications
1	Format	PCM
2	Bit depth	24 bits
3	Sampling frequency	48 kHz

Audio level and loudness specifications

No.	Item	Technical specifications
1	Maximum true peak level	-2 dBTP
2	Integrated loudness	-24 LKFS
3	Variations of loudness	±2 LU

Channel distribution of surround sound

No.	Track	Channel
1	Track 1	Left (L)
2	Track 2	Right (R)
3	Track 3	Centre (C)
4	Track 4	Low frequency effects (Lfe)
5	Track 5	Left surround (Ls)
6	Track 6	Right surround (Rs)
7	Track 7	Stereo left (L)
8	Track 8	Stereo right (R)

3.2.2.2.2 3D audio production

In 3D audio production, the audio level and loudness requirements can be referred to surround sound-related specifications, the 3D audio channel allocation reference Table 34.

TABLE 34

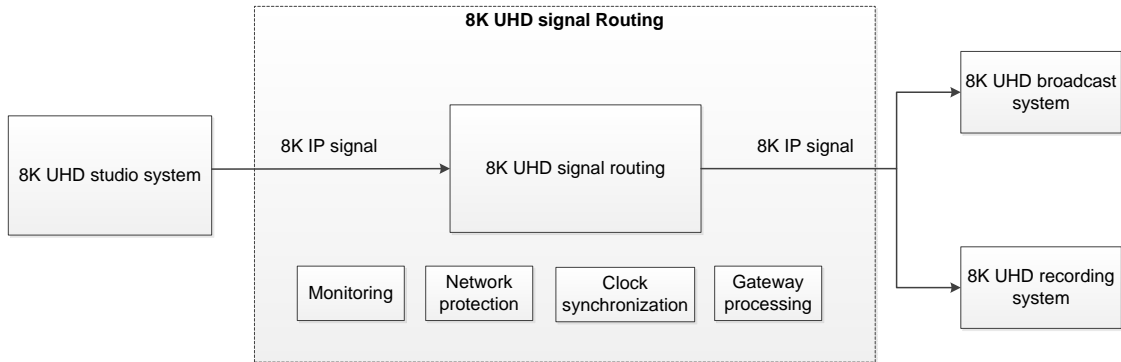
Channel distribution of 3D audio

No.	Track	Channel
1	Track 1	Left (L)
2	Track 2	Right (R)
3	Track 3	Centre (C)
4	Track 4	Low frequency effects (Lfe)
5	Track 5	Left surround (Ls)
6	Track 6	Right surround (Rs)
7	Track 7	Stereo left (L)
8	Track 8	Stereo right (R)
9	Track 9	Left top front (Ltf)
10	Track 10	Right top front (Rtf)
11	Track 11	Left top rear (Ltr)
12	Track 12	Right top rear (Rtr)

3.3 8K UHD signal routing system

The 8K UHD signal routing system is mainly responsible for the routing of 8K UHD signals exchange among the studio, the recording system and the broadcast system. The system adopts an all-IP system architecture to realize the routing, recording, broadcasting, transmission and distribution of 8K UHD signals, as shown in Fig. 75.

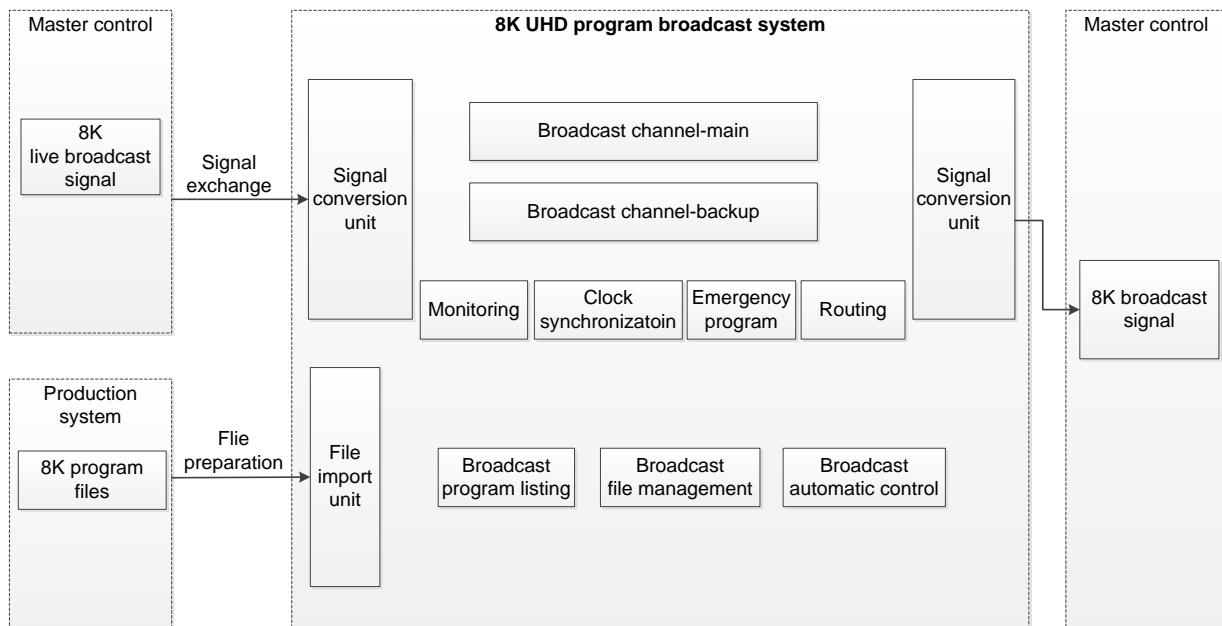
FIGURE 75
8K UHD signal routing system



3.4 8K UHD broadcast system

The 8K UHD broadcast system takes the broadcast video server as the core, realizing the functions of live broadcast signal routing, broadcast file decoding and playback, signal switching, monitoring, and emergency programme. The broadcast system is logically divided into several processing units, among which the 8K UHD video and audio signals conversion unit adopts IP technology. The file import unit migrates the 8K programme files from the production system to the broadcast video server for storage, and completes the decoding and broadcasting of the files according to the pre-programmed programme playlist, as shown in Fig. 76.

FIGURE 76
8K UHD programme broadcast system



4 8K distribution and transmission

The 8K UHD broadcast signal from the 8K UHD TV broadcast system is sent to the IPTV broadcasting control platform, which carries out HDR Vivid encoding, Audio Vivid encoding and AVS3 encoding, then distributes and transmits the encoded 8K 120 Mbit/s IP stream.

4.1 Distribution and transmission of video and audio over private network

The video and audio encoding system in the IPTV broadcasting control platform provides timestamp embedding and HDR Vivid and Audio Vivid metadata are extracted by the AVS3 encoder. After inserting the timestamp through the encoder, the 8K signal uses the PES packet of TS stream to carry and transmit the timestamp data for extraction with decoder system. Meanwhile, the encoder extracts the HDR Vivid metadata and Audio Vivid metadata from the uncompressed live streams, and encodes by AVS3 and outputs the live streams. 8K Streaming Service routes and distributes the received IP streams and pushes them to the private network for transmission.

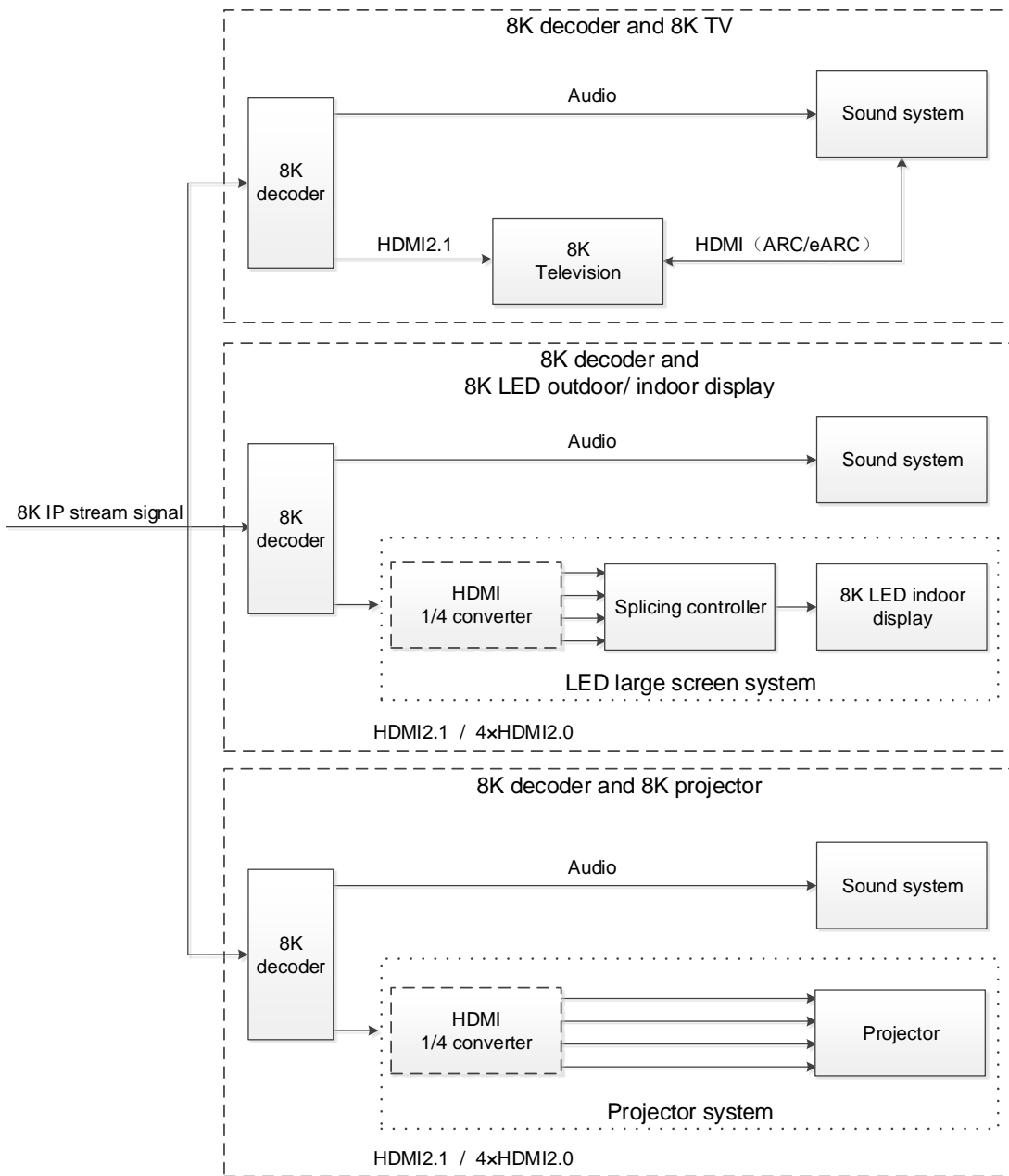
4.2 Distribution and transmission of audio over Internet

The Audio Vivid encoder system in IPTV broadcast control platform provides metadata extraction, timestamp embedding and encoding distribution functions for Audio Vivid distributed over Internet. The 8K signal is de-embedded by the Audio Vivid encoder, which then extracts the PCM audio signal for Audio Vivid encoding. Audio Vivid encoder system use the bitrate of 384 kbit/s and insert a timestamp, which is carried and transmitted by the PES packet of TS stream, and the Audio Vivid stream is packaged and pushed to the internet and transmitted to the "EtherealSound " mobile APP. The video and audio are synchronized using the timestamp extracted by the mobile receiver.

4.3 Video and audio synchronization in heterogeneous network (HetNet)

Since the 8K TV accompaniment cannot be played through high-power speakers in the public square, a HetNet AV synchronization technology was developed. The audience are able to hear the TV accompaniment sound while watching 8K video through the public large screen. Automatic end-to-end delay detection is achieved by inserting timestamp into the stream information in the chain, and Audio Vivid with a timestamp is sent to the "EtherealSound" mobile APP Through the delay time measurement between large screen and mobile terminal, audio distribution delay control, and mobile decoding and rendering delay dynamic compensation technology, the audio and video delay of different network transmission is controlled within $-90\text{ ms}\sim+120\text{ ms}$, which meets the requirements of the Chinese Standards "Relative timing of sound and vision for broadcasting" (GB/T 22150-2008). The video and audio synchronization workflow in HetNet is shown in Fig. 77.

FIGURE 78
System architecture



5.3 Reception and display technical specifications

5.3.1 Decoder specifications

8K professional decoder should have dual-link access and NTP (Network Time Protocol) client, support NTP protocol. It has the functions of remote control, operation data feedback, security authentication, remote upgrade of systems and applications, and log management. The 8K decoder specifications refer to Table 35.

TABLE 35

8K decoder specifications

No.	Item	Technical specifications	
	Type of input transport stream	Transport streams conformed by GB/T 17975.1-2010	
	Video streaming protocol	UDP and RTP transmission methods specified in T/UWA012.2-2022	
	Input transport stream packet length	Input 188-byte packet length and 204-byte packet length	
	File format	TS file	
	Single-channel video decoding	Profile and level	AVS3 basic profile 10bit (profile_id value 0x22), level 10.0.60
		Frame frequency	50 Hz (Progressive)
			100 Hz (Progressive)
		Picture aspect ratio	16:9
		Sampling	4:2:0
		Quantization levels	10 bits
			12 bits
		Pixel count Horizontal × vertical	7 680 × 4 320 maximum
Colour gamut	ITU-R BT.2020		
HDR	ITU-R BT.2100		
	HDR Vivid		
	Audio output	Stereo audio	
		5.1 Surround sound encoding signal passthrough and output	
	Audio interface	HDMI, fiber S/PDIF; L/R RCA	
	Video interface	HDMI2.1	
		HDMI2.0×4, SQD	

5.3.2 LED Screen specifications

At present, there are about 40 cities in China which have UHD large LED screens larger than 200 m². Take Shenzhen COCO Park as an example, the screen resolution is 7 680 × 4 416 pixels, the dot pitch is P2.5, the maximum brightness is 5 500 cd/m², and the display supports HDR and WCG. The present of UHD large LED screen in Shenzhen COCO Park are shown in Figs 79 and 80, the multi-city present effect is shown in Fig. 81.

FIGURE 79
Shenzhen COCO Park in the day



FIGURE 80
Shenzhen COCO Park at night



FIGURE 81
Multi-city present effect



The LED screen should follow the technical specifications in Table 36.

TABLE 36
LED display specifications

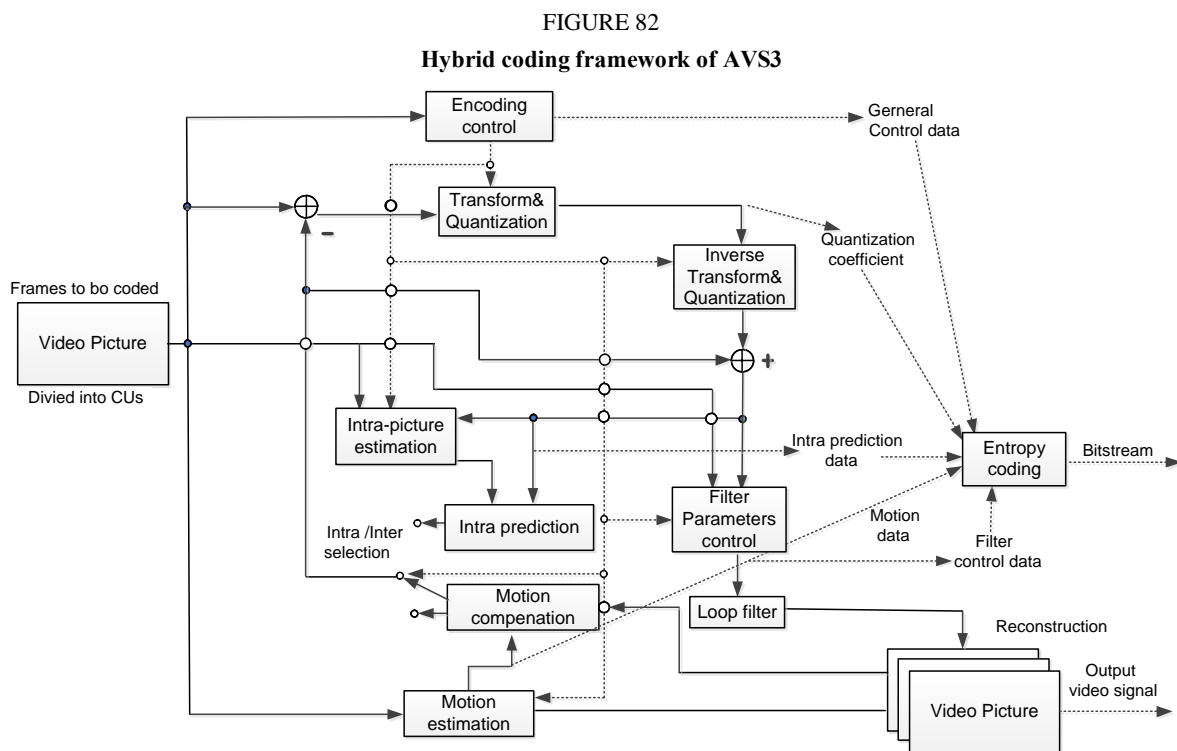
No.	Item	Technical specifications
	Physical resolution	$\geq 7\ 680 \times 4\ 320(8K)$
	Maximal luminance	white lamp $\geq 5\ 000\ \text{cd/m}^2$
		black lamp $\geq 3\ 500\ \text{cd/m}^2$
	Viewing angle of luminance	horizontal $\geq 140^\circ$
		vertical $\geq 120^\circ$
	Viewing angle of chromaticity	horizontal $\geq 150^\circ$
		vertical $\geq 150^\circ$
	Contrast	$\geq 5\ 000:1$
	Large area uniformity	$\geq 95\%$
	Area coverage of the BT.2020 gamut	$\geq 78\%$ (CIE 1931)
	Display frame rate	$\geq 50\ \text{Hz}$
	Single lamp refresh rate	$\geq 3\ 840\ \text{Hz}$
	Bit depth supported by input	10 bits
	Colour temperature	adjustable from 3 200 K~9 300 K

6 Main technology

6.1 AVS3

AVS3 (The 3rd generation of the Audio Video coding Standard, T/AI 109.2-2021: “Intelligent Media Coding – Part 2: Video”) is the third-generation video coding standard for 8K and 5G industrial applications in China. It included base profile finished in March 2019 and advanced profile finished in June 2021. In 2022, A001 (r20) published by DVB and TS 101 154 (v2.8.1) published by ETSI (European Telecommunications Standardization Association) incorporated AVS3 codec into the coding specification part.

AVS3 uses hybrid coding framework of block-based prediction coding and transform coding, which includes block division, intra prediction, inter prediction, transform quantization, entropy coding, loop filtering, etc. Figure 82 shows the hybrid coding framework of AVS3.



6.2 HDR VIVID

HDR Vivid is an HDR video technology standard recommended by UWA (UHD World Association, <http://theuwa.com/>), which released its core part: High Dynamic Range Video Technology Part 1: Metadata and Tone mapping in September 2020, and has released 7 companion standards in the past two years.

HDR Vivid technology is an end-to-end system solution. Based on human vision perception, it defines dynamic metadata in HDR video content, thus be adaptable to various applications.

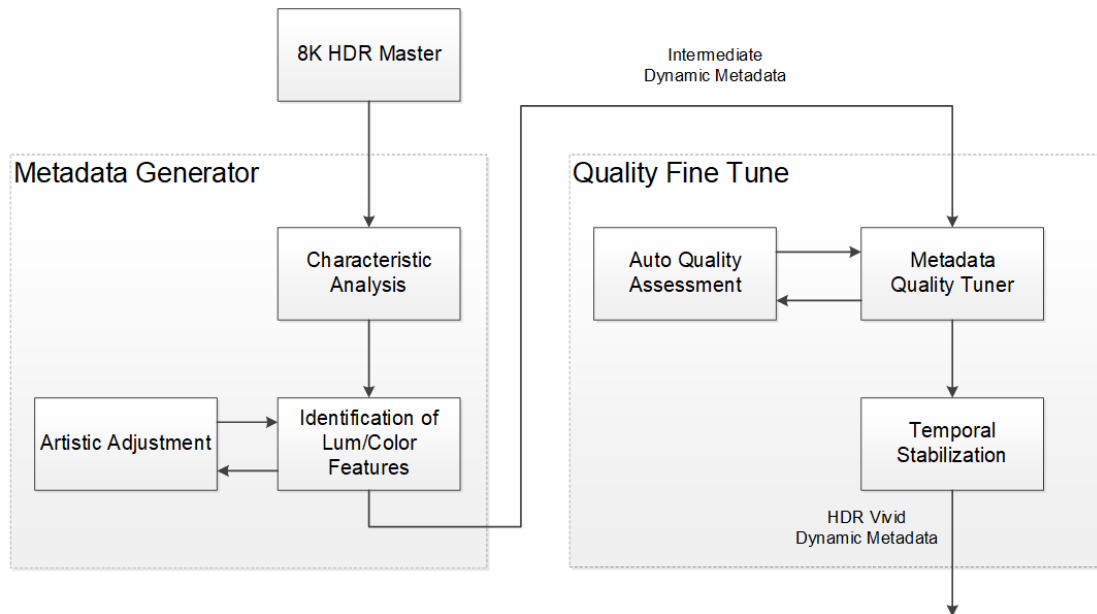
6.2.1 Introduction to HDR Vivid encoding

The dynamic metadata assures the consistency across different terminal displays, and allows creators to incorporate their preferred artistic style. Metadata generator module first analyses the characteristics of the HDR master content, including image brightness features, histogram distribution, etc. Based on the characteristics of different scenes, multiple luminance colour feature

points are identified. At this point, the content creator can adjust these feature points according to their preference. After the feature points mentioned above are determined, the metadata quality tuning module can evaluate and adjust the quality of the dynamic metadata generated by the algorithm without human intervention. Finally, the temporal stabilizer module prevents video flickering that might be introduced by dynamic metadata. The HDR Vivid encoding workflow is shown in Fig. 83.

FIGURE 83

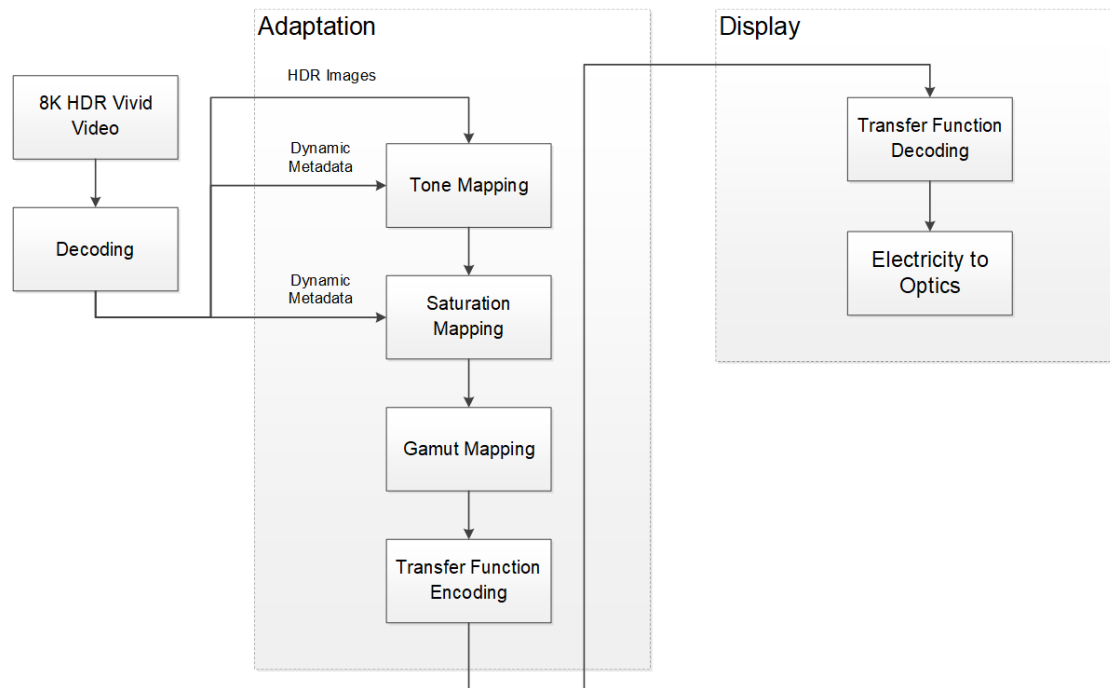
HDR Vivid encoding workflow



6.2.2 HDR Vivid decoding

The HDR Vivid terminal first decodes the AVS3 video stream. Then, the HDR Vivid image is rendered with luminance and colour gamut for display according to dynamic metadata. This adaptation process is carried out through Tone Mapping and Saturation Mapping to maximize the original artistic intention of HDR Vivid content on various terminals. The final display process is achieved through transfer function decoding and electro-optical conversion to render HDR Vivid content. The HDR Vivid decoding workflow is shown in Fig. 84.

FIGURE 84
HDR Vivid decoding workflow



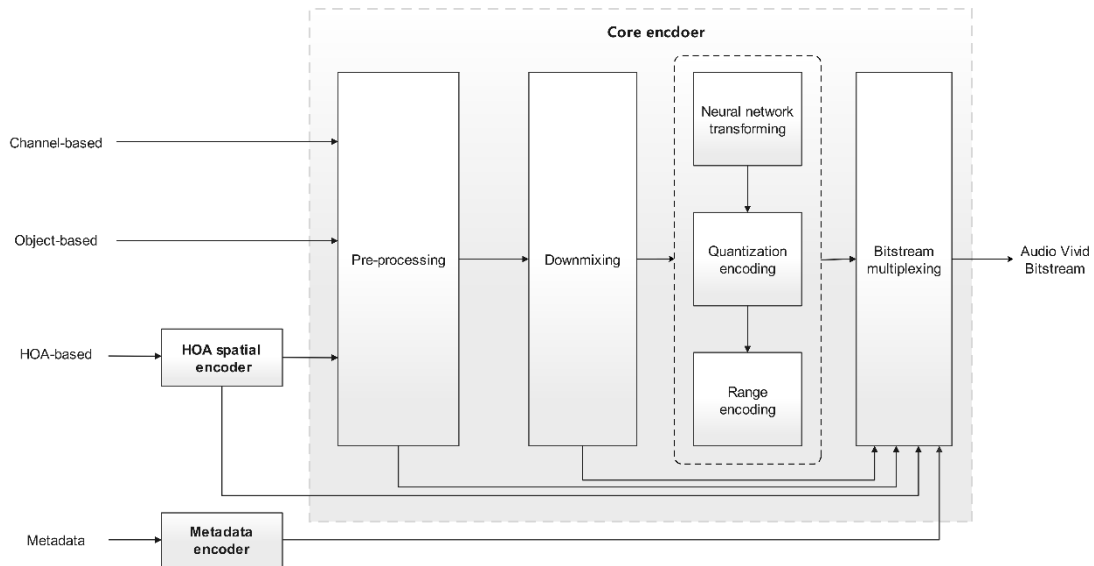
6.3 Audio Vivid

Audio Vivid is a surround/3D audio Codec and rendering technology developed by China for the “BaiChengQianPing” platform, and is a high quality, end-to-end audio production and broadcast core processing technology solution based on AI technology. It was standardized by a joint effort of UHD World Association (UWA) and Audio Video Coding Standard Workgroup of China (AVS), and was first published in April 2022. Audio Vivid adopts a novel hybrid AI Codec architecture that extracts abstract features and increases entropy coding efficiency through deep learning. It also incorporates HOA spatial coding technology, makes novel 3D audio metadata structure and definitions, supports multi-speaker rendering and binaural rendering. Audio Vivid brings enhanced sense of sound spatialization and localization, fulfils customer’s demand on high fidelity, high immersive sound experience while enabling personalization, interaction ability unique to NGA system. Audio Vivid encoder and distribution/transmission server can support package of various protocols and multiple source formats.

6.3.1 Introduction to Audio Vivid

The Audio Vivid encoder comprises a lossy audio encoding tool, a lossless audio encoding tool and a metadata encoding tool. It supports encoding of channel signal, object signal, HOA signal and metadata. Particularly, the lossy audio encoding tool consists of a HOA spatial encoder and a core encoder. The core encoder consists of a pre-processing module, a downmixing module and neural network-based modules for feature transform, Quantization and range coding, and is able to encode channel/object signals into bitstream. The core encoder together with the HOA spatial encoder can encode HOA signals. The Audio Vivid metadata consists of two parts – basic metadata and extended metadata, where the basic metadata references to the audio definition model defined in ITU-R BS.2076-2 (ADM) and the extended metadata are newly defined to serve for enhanced binaural rendering experience. Audio Vivid metadata can be either forward compatible and backward extendable which offers sufficient flexibility and extendibility while fulfilling the requirement of metadata global interworking.

FIGURE 85
Audio Vivid encoding workflow



6.3.2 Audio Vivid decoding and rendering

The decoding workflow of Audio Vivid is basically a reverse of encoding workflow. Audio Vivid comprises two default renderers – a multi-speaker renderer and a binaural renderer and also supports interface to 3rd party renderers. Both the default renderers support rendering/playback in content production and device and can receive channel signals, object signals, HOA signals and their combinations so to offer enhanced experience, more creative ideas and allow interaction and personalization. The default binaural renderer also supports 3DoF immersive experience.

References

- [1] GB/T 22150-2008 Relative timing of sound and vision for broadcasting.
 - [2] GY/T 358-2022 Technical requirements for display adaptation metadata of high dynamic range television systems.
 - [3] UWA official website: <http://theuwa.com/>.
 - [4] TUWA 012.1-2022 Technical requirements of program broadcasting of UHD video and audio broadcasting system for “Bai Cheng Qian Ping”.
 - [5] TUWA 012.2-2022 Coding of UHD video and audio broadcasting system for "Bai Cheng Qian Ping": system.
 - [6] TUWA 012.3-2022 Coding of UHD video and audio broadcasting system for "Bai Cheng Qian Ping": video.
 - [7] TUWA 012.4-2022 Technical requirements of Professional Decoder of UHD video and audio broadcasting system for “Bai Cheng Qian Ping”.
 - [8] TUWA 012.5-2022 Technical requirements of the public display system (outdoor) of UHD video and audio broadcasting system for “Bai Cheng Qian Ping”.
 - [9] TUWA 012.6-2022 Technical requirements of network transmission of UHD video and audio broadcasting system for “Bai Cheng Qian Ping”.
 - [10] TUWA 012.7-2022 Measuring method of UHD encoder for “Bai Cheng Qian Ping”.
 - [11] TUWA 012.8-2022 Measuring method of the public display system (outdoor) of UHD video and audio broadcasting system for “Bai Cheng Qian Ping”.
 - [12] TUWA 012.9-2022 Technical requirements of the public display system (indoor LCD) of UHD video and audio broadcasting system for “Bai Cheng Qian Ping”.
-