International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Report ITU-R BT.2390-8

(02/2020)

High dynamic range television for production and international programme exchange

BT Series

Broadcasting service (television)

International Telecommunication Union

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 http://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

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, 2020

REPORT ITU-R BT.2390-8

High dynamic range television for production and international programme exchange

(02/2016-10/2016-03/2017-10/2017-04/2018-10/2018-04/2019-07/2019-2020)

Summary

Recommendation ITU-R BT.2100 – Image parameter values for high dynamic range television for use in production and international programme exchange, specifies parameters for High Dynamic Range television (HDR-TV) signals to be used for programme production and international programme exchange. This Report provides background information on HDR in general, and for the perceptual quantization (PQ) and hybrid log-gamma (HLG) HDR signal parameters specified in the Recommendation.

As HDR-TV is at a formative stage of research and development as presented in this Report, a call for further studies is made, in particular on the characteristics and performance of the recommended HDR-TV image parameter values, for use in broadcasting.

TABLE OF CONTENTS

Page

1 Introduction and design goals for HDR television

HDR-TV enables more natural images that contain wider variations in brightness. While HDR-TV does allow the picture average brightness to increase, the expectation is that indoor scenes produced in HDR will generally be at a similar brightness as with legacy TV systems. The brightness range available with HDR enables outdoor sunlit scenes to appear noticeably brighter than indoor scenes, thus providing a more natural look. All scenes, especially outdoor, will be able to produce small area highlights such as specular reflections or emissive light sources at much higher brightness. There is also an improvement in the ability to show details in dark areas; this feature is dependent on the black level of the display and the viewing environment.

1.1 Common misconceptions on HDR

HDR for video and display is an entire ecosystem that encompasses much more than the words underlying the acronym. Before discussing system issues, there are number of frequent misconceptions about HDR video, such as: 'It is all about brighter pictures', 'It is all about dynamic range' , 'It is all about bit-depth', 'It is primarily an image capture issue', 'It is primarily a display capability issue', 'It makes images look like paintings'.

Of these, only the first one here will be addressed. The misconception about HDR being simply brighter[1] pictures arises from the fact that the maximum luminance *capability* is indeed much higher than standard dynamic range (SDR) television. However, this higher maximum is primarily used by the highlight regions of images. While the highlights will indeed appear brighter [1], they are nearly always small in region, and the overall image may not necessarily appear brighter. This is because the overall appearance of an image's brightness is dominated by the average brightness, not the small regions usually occupied by highlights. One type of highlight is the specular reflection. The advantages of having more accurate specular reflections enabled by HDR include better surface material identification [2] as well as in depth perception, even with 2D imagery [3] [4].

By comparison, in the process of making the SDR content (whether colour grading in post-production or selection of the camera settings in live broadcast), human decisions are invariably made to fit the higher dynamic range of the scenes into the standard range.

In typical practice, highlights are processed through a shoulder operation or simply clipped. This loses not only the amplitudes of the highlights, but also the details within and around the highlights. Similarly, shadow detail is lost. Colour emissive highlights result in the colour component going through different portions of the shoulders such that the colour shifts towards white. These different aspects resulted in the realization that a new HDR signal format needed to be developed to allow for the HDR display to truly deliver an HDR experience.

There *is* another way to utilize the new range capabilities than to utilize it solely for highlights. This is to allow for more realistic scene-to-scene luminance variations. In current SDR, with a range of less than three log10 luminance, it was always difficult to render evening scenes, and nearly impossible to render the luminance differences of indoor and outdoor scenes. Acknowledging this limitation with SDR, some creatives like to use the increased dynamic range of HDR to have larger scene-to-scene variations in mean luminance. So for this particular approach, HDR may result in brighter images for some scenes.

However, despite these variations in intent for invoking increased brightness, HDR also allows for lower black levels than traditional SDR, which was typically in the range between 0.1 and 1.0 cd/m^2

[1] Brightness is technically a perceptual measure, and not linear to luminance. However, in the majority of consumer TV literature, 'brightness' is used to convey either overall luminance, or the maximum luminance. The term will be used in that sense here.

for cathode ray tubes (CRTs), and is now in the range of 0.1 cd/m^2 for most standard SDR liquid crystal displays (LCDs). So a key design question is how low should the black level be?

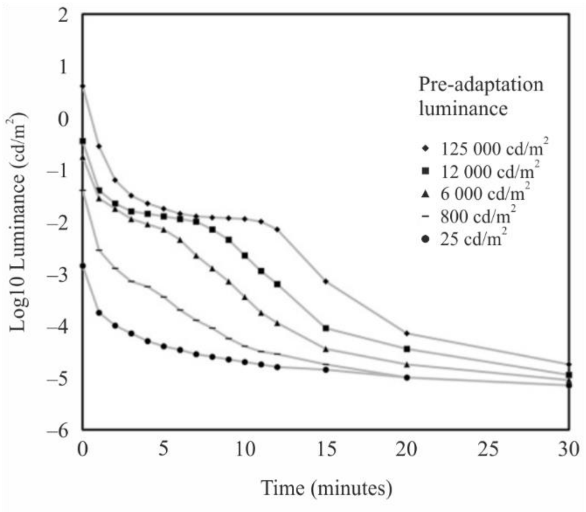
1.2 System black level determination

In order to determine the system black level, the state of light adaptation[2] is central. The classic psychophysical study on dark adaptation was by Hecht *et al* [5], which corresponds to the top data line of the plot in Fig. 1, which is a compilation of more recent studies [6]. The left branch of the curve corresponds to the cones, while the right branch of the curve corresponds to rod vision. While threshold values of less than 0.00001 cd/m^2 can be obtained, they can take significant durations of dark adaptation, which are not likely in entertainment media. If one restricts consideration to cone vision's left branch of the uppermost curve, visibility does not go as low, but it still can be below ~0.02 cd/m^2.

However, detectability as low as 0.02 cd/m^2 seems to require minutes of dark adaptation time, which in traditional entertainment media is considered unrealistic[3]. Often, the early part of the curve (< 1 minute) is used to conclude that black levels of between 0.3 and 1.0 cd/m^2 are sufficient, and in previous years display capability has been limited to be greater than 0.1 cd/m^2 (e.g. for fixed backlight LCD). Using data such as those presented in Fig. 1 to conclude that the human eye cannot see black level differences below 0.1 cd/m^2 overlooks that the curves depend on the initial adaptation condition. The other curves shown in the figure show that as the initial adaptation level is lowered, the ability to see lower luminance levels improves. While the plotted time scale does not allow for determination of adaptation ranges on the order of video scene cuts (3-5 s), the leftmost data points are enough to show that visual detectability of black level can be close to 0.001 cd/m^2 for the 25 cd/m^2 initial level, close to SDR average luminance levels (i.e. average picture level (APL)). Thus from Fig. 1, one would easily conclude that the black level of video should allow levels as low as 0.001 cd/m^2.

FIGURE 1

Black level detectability as a function of duration for different initial adaptation levels. From Stokkerman [6]



Report BT.2390-01

However, system design by the use of data as in Fig. 1 leans toward the most demanding cases, where the entire image may be dark. Other approaches consider that images generally do not consist of

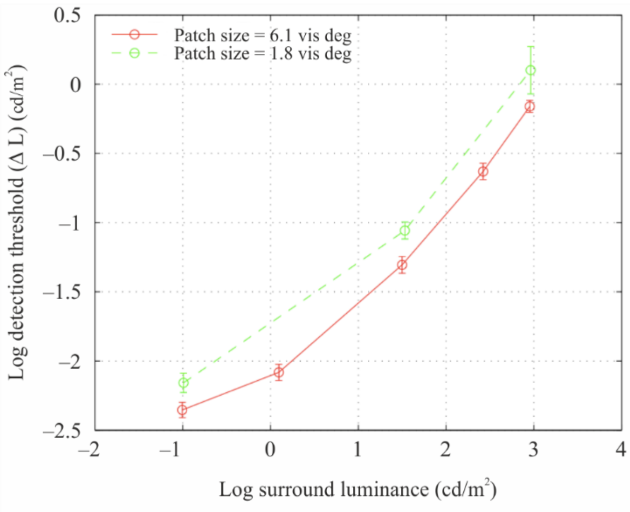
2 Sometimes called dark adaptation when adapting toward dark.

3 Creatives in production and post have desired to allow for longer periods of dark adaptation in their content.

all-dark regions; there is a mixture of different luminance levels. The general approach is to treat the image as a surround around a possible black area. Using rectangular patches with a white surround, Mantiuk *et al* [7] studied black level threshold as a function of the size of the black region. The area outside of the patch was termed the surround, and the surround serves as a surrogate for an actual image with average image luminance level. The results in Fig. 2 show the lowest black level that can be discriminated from zero luminance is ~–2.4 log10 cd/m^2 (0.0039 cd/m^2), at least for the darkest surround that they studied, which was 0.1 cd/m^2. Lower thresholds would be expected from darker surrounds, such as might occur in home theatre, or some evening viewing situations.

FIGURE 2

Detectability of black level differences for a rectangular patch of either 6.1 or 1.8 visual degrees, both as a function of surround luminance level



Report BT.2390-02

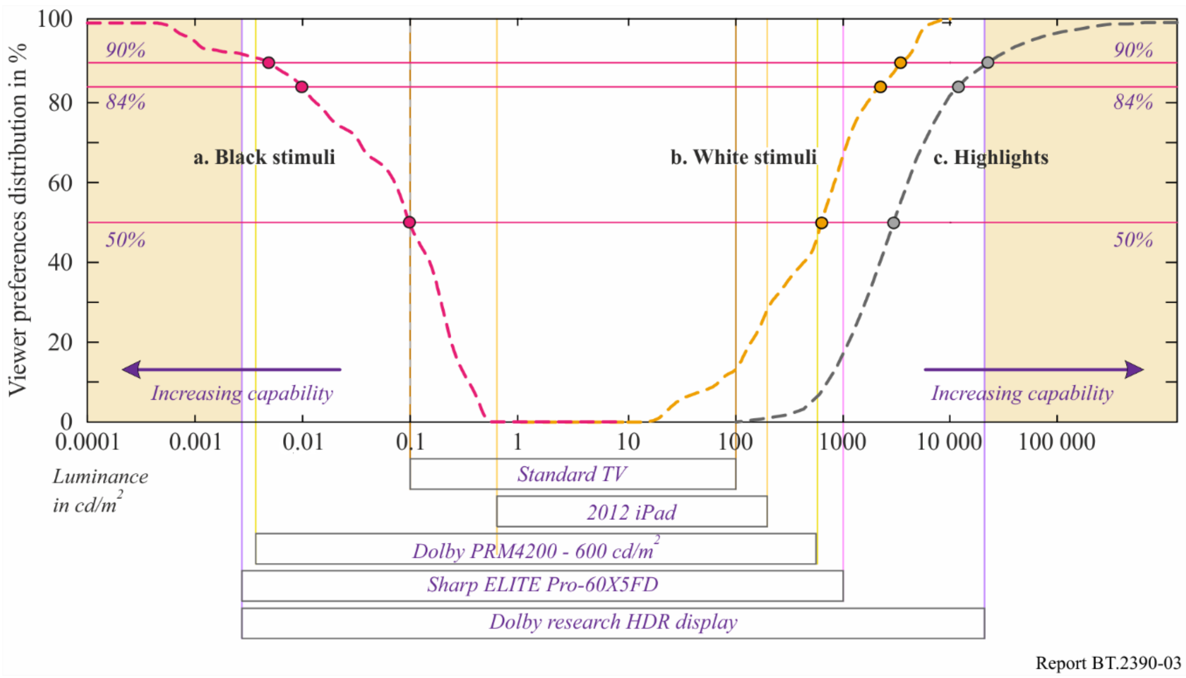
Two things are clear. As the surround luminance decreases, the detectable black level decreases. That is, the expected surround luminance that results from practical imagery can determine the necessary black level to achieve a pure black perception, as well as finding the level where dark detail is no longer distinguishable. The other effect is that thresholds for the larger black region are lower than for the smaller. Thus in designing a system black level, the expected size of the black region is a key factor. Note that the largest region studied in this work was 6 degrees, whereas the image size for HDTV viewed at 3H is approx. 35 degrees (UHDTV @ 1.5 H is ~70 degrees).

Another approach for determining system black level is to not base it on psychophysical detection tasks with abstract geometric stimuli, but rather use preferences while viewing more natural imagery. Rempel at al. [8] measured preference for display black level and brightness in short video clips (a sitcom) and found all participants consistently set the black level to the lowest possible setting, which was about 0.3 cd/m^2 for their display. So the only conclusion from this was that 0.3 is not low enough. A more recent study using an experimental HDR display with very low black level capability [9] [10] [11] found levels near *its* minimum capability, which was 0.004 cd/m^2. In order to meet the preferences of 90% of the viewers, a level of 0.005 cd/m^2 was needed. The typical current black level LCD TVs of 0.1 cd/m^2 would meet the preferences of only half of the viewers. Results are shown in Fig. 3.

The plot in Fig. 3 demonstrates the results of psychophysical experiments designed to understand the preferred dynamic range [9] [10] [11]. The experiment was based on a two-alternative forced choice paradigm using static images shown sequentially for average shot durations (2-5 s) and trial durations of around 20 s to include response times, for an experiment lasting a total of 40 minutes per participant. The stimuli were drawn from three classes of images, containing shadow detail, reflective white stimuli, and highlight stimuli. A dual modulation display was used using an LCD panel backlit by a digital cinema projector, allowing a luminance range between 0.004 and 20 000 cd/m^2. Separate experimental sessions were conducted for the black level scenes vs. the white and highlight level scenes; the results of all the experiments are plotted on the same Figure but this should not be interpreted as indication that both extremes can be perceived simultaneously.

FIGURE 3

Cumulative distribution functions for a. black stimuli, b. reflective white stimuli and c. emissive and highlights. For comparison, the dynamic ranges of common displays are given



Regarding the black level, there are a number of studies that found detectability as well as preferences well below the level of 0.1 cd/ m^2, which was common for SDR displays. Values in the range of 0.001 to 0.005 cd/ m^2 could be deduced from the studies described here, and regarding preferences there may be upward biases due to the smaller field of view used in [9] than occurs with UHDTV.

1.3 System white and highlight level determination

In video, the system white is often referred to as reference white, and is neither the maximum white level of the signal nor that of the display. When calibration cards are used to set the reference white, it is a diffuse white (also called matte) that is placed on the card, and measured. The ideal diffuse white has a Lambertian reflection. The luminances that are higher than reference white are referred to as highlights. While there are several key quality dimensions and creative opportunities opened up by HDR (e.g. shadow detail, handling indoor and outdoor scenes simultaneously, and colour volume aspects), one of the key differentiators from SDR is the ability for more accurate rendering of highlights. These can be categorized as two major scene components: specular reflections[4] and emissives (also referred to as self-luminous). They are best considered relative to the maximum diffuse white luminance in the typical image. Most scenes can be broken down into two key ranges:

[4] In traditional photography, the term ‘highlights’ is sometimes used to refer to any detail near white, such as bridal lace, which may entirely consist of diffuse reflective surfaces. In HDR literature, the use of ‘highlights’ is intended for the specular or emissive regions in an image since that is a key feature opened up by HDR.

object's diffuse reflectances and the highlights. (Some scenes would defy such categorization, e.g. fireworks at night.) The object's reflectance is important to convey its shape due to shading and other features, and the visual system has strong ability to discount the illuminant to be able to estimate the reflectance [12].

However, the human ability to perceive both types of highlights is much less accurate and less computationally sophisticated as the ability perceive reflectances [12]. Illustrations of emissives and specular highlights are shown in Fig. 4.

FIGURE 4

Emissive light sources, specular reflections, and diffuse white

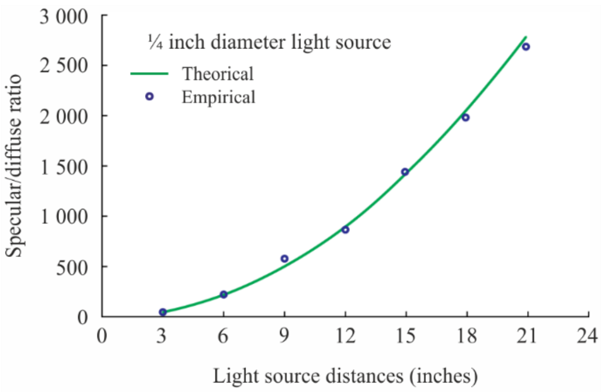
Report BT.2390-04

In traditional imaging, the range allocated to these highlights was fairly low and the majority of the image range was allocated to the diffuse reflective regions of objects. For example, in hardcopy print the highlights would be 1.1x higher luminance than the diffuse white maximum [13]. In traditional video, the highlights were generally set to be no higher than 1.25x the diffuse white. Of the various display applications, cinema allocated the highest range to the highlights, up to 2.7x the diffuse white.

Actual measurements show the specular regions can be over 1 000x higher than the underlying diffuse surface [2], which is presented in Fig. 5. This means the physical dynamic range of the specular reflections vastly exceed the range occupied by diffuse reflection. If a visual system did not have specialized processing as previously described, and saw in proportion to luminance, most objects would look very dark and the visible range would be dominated by the specular reflections. Likewise, emissive objects and their resulting luminance levels can have magnitudes much higher than the diffuse range in a scene or image. The most common emissive object, the disk of the sun, has a luminance so high (~1.6 billion cd/m^2), it is damaging to the eye to look at more than briefly, and exceeding even the speculars. A more unique aspect of the emissives is that they can also be of very saturated colour (sunsets, magma, neon, lasers, etc.).

FIGURE 5

Measurements showing that the specular regions can be over 1 000x higher in comparison to the underlying diffuse surface. After Wolff (1994)

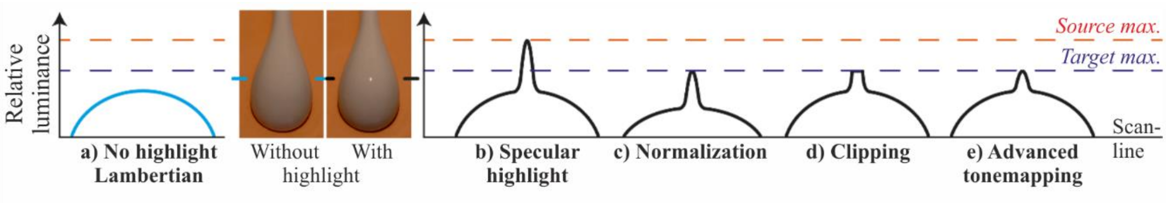


Report BT.2390-05

With traditional imaging's under-representation of highlight ranges, the question arises: what happens to the luminances of highlights? Figure 6 shows example scanlines of common distortions from a specular highlight from a glossy object, (b). It exceeds the maximum luminance of the display (or the signal), indicated as the dashed line titled 'Target Max.'. Illustration (c) shows a distortion that is seldom selected, that is, to renormalize the entire range. Another approach, (d) preserves diffuse luminances, and the highlight is simply truncated (hard-clipping). Details within the highlight region are replaced with constant values, giving rise to flat regions in the image, looking quite artificial. Typical best practices (e), have been referred to as soft-clipping, or a knee. Here the shape and internal details of the highlight are somewhat preserved, without flattened regions. HDR allows for a result closer to scanline (b). The more accurate presentation of specular highlights, (assuming the entire video pathway is also HDR), is one of the key distinctions of HDR. A number of perceptual papers have looked closely at specular reflection, as mentioned in the beginning of this section. Preferences of luminances for diffuse white and highlights are shown in Fig. 3.

FIGURE 6

Effects of highlight rendering, clipping and (tonescale) compression



Per the results shown in Fig. 3, 16% of the viewers preferred highlights ≥10 000 cd/m^2. Also shown is that 50% of the viewers preferred diffuse white levels ≥ 600 cd/m^2. This suggests that if display luminances increase in the future, some PQ content (e.g. outdoor scene in bright sun) may be produced with diffuse white levels higher than the levels indicated in Report ITU-R BT.2408. Consideration would, however, need to be given to the appearance on lower peak luminance PQ displays.

2 Television system architecture

2.1 The relationship between the OETF, the EOTF and the OOTF

This Report makes extensive use of the following terms:

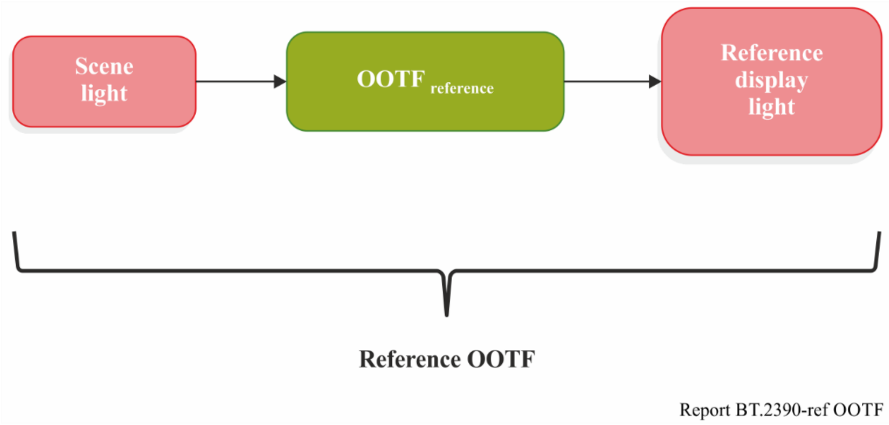
OETF: the opto-electronic transfer function, which converts linear scene light into the video signal, typically within a camera.

EOTF: electro-optical transfer function, which converts the video signal into the linear light output of the display.

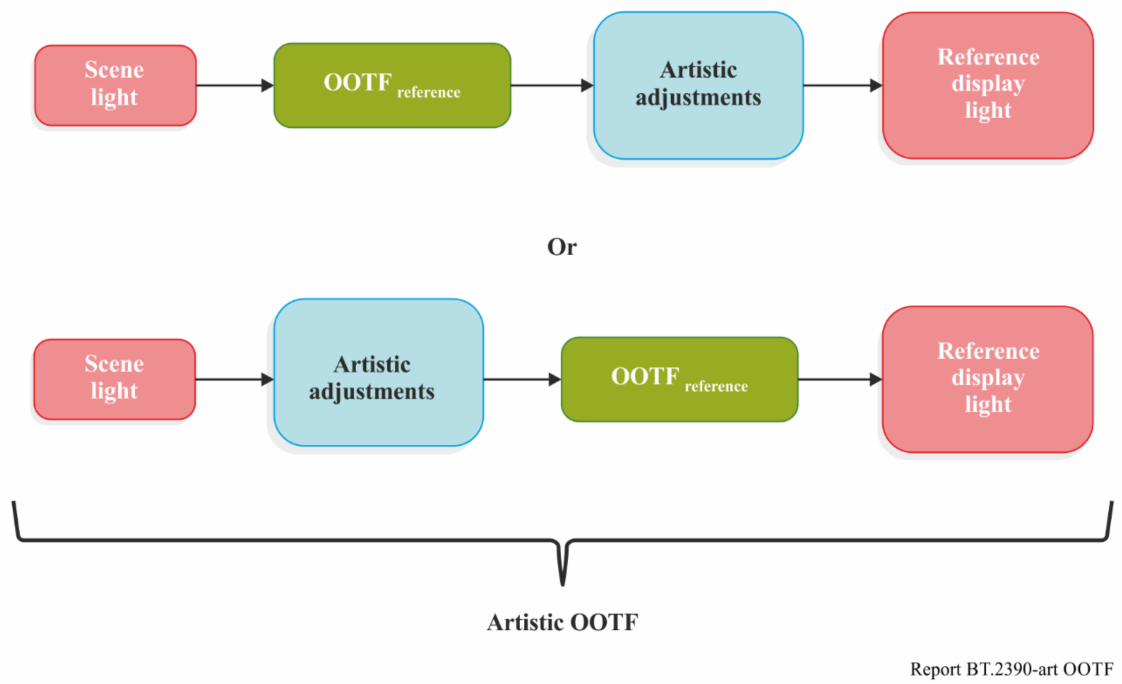
OOTF: opto-optical transfer function, which has the role of applying the 'rendering intent'.

These functions are related, so only two of the three are independent. Given any two of them the third one may be calculated. This section explains how they arise in television systems and how they are related.

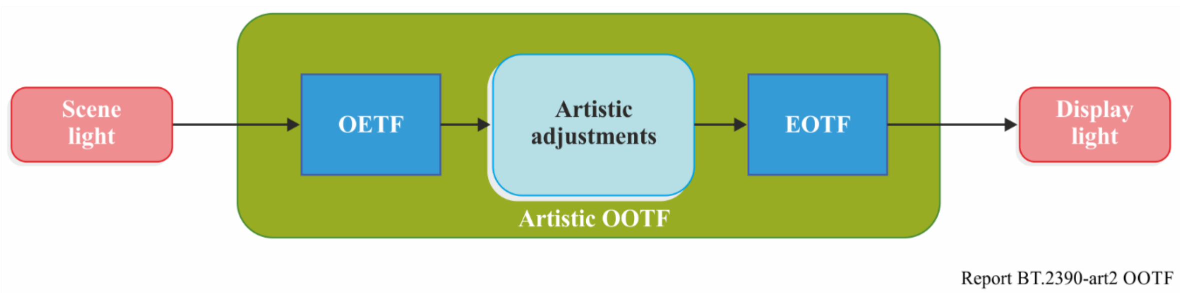
In television systems the displayed light is not linearly related to the light captured by the camera. Instead an overall non-linearity is applied, the OOTF. The "reference" OOTF compensates for difference in tonal perception between the environment of the camera and that of the display. Specification and use of a 'reference OOTF' allows consistent end-to-end image reproduction, which is important in TV production.



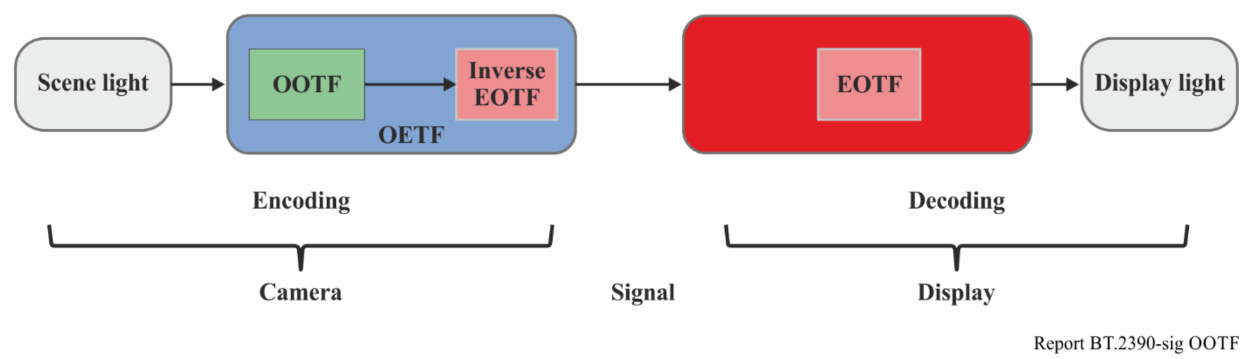
Artistic adjustment may be made to enhance the picture. These alter the OOTF, which may then be called the ‘artistic OOTF’. Artistic adjustment may be applied either before or after the reference OOTF.



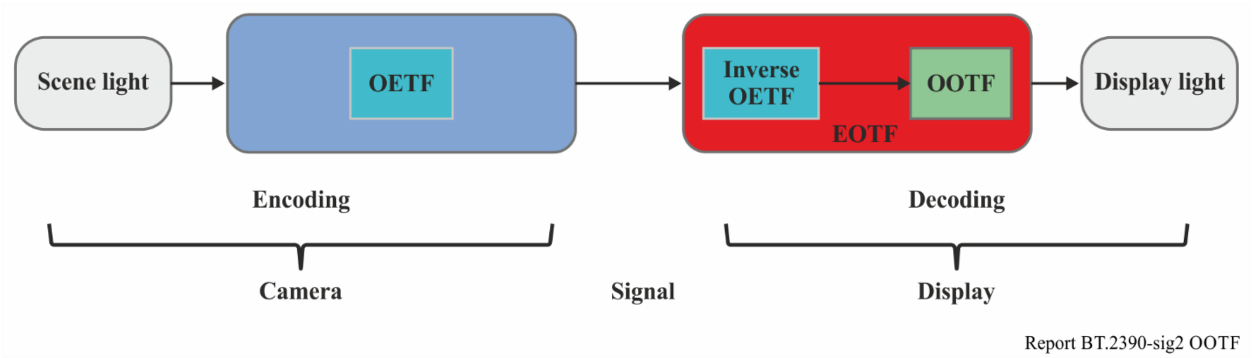
In general the OOTF is a concatenation of the OETF, artistic adjustments, and the EOTF.



The PQ system was designed with the model shown below, where the OOTF is considered to be in the camera (or imposed in the production process):



The HLG system the system was designed with the model shown below, where the OOTF is considered to be in the display:



Only two of three non-linearities, the OETF, the EOTF, and the OOTF, are independent. In functional notation (where subscripts indicate the colour component):

$$\mathrm{OOTF}_R(R,G,B) = \mathrm{EOTF}_R(\mathrm{OETF}_R(R,G,B))$$
$$\mathrm{OOTF}_G(R,G,B) = \mathrm{EOTF}_G(\mathrm{OETF}_G(R,G,B))$$
$$\mathrm{OOTF}_B(R,G,B) = \mathrm{EOTF}_B(\mathrm{OETF}_B(R,G,B))$$

This is clearer if the symbol ⊗ is used to represent concatenation. With this notation, the following three relationships between these three non-linearities can be obtained:

$$\mathrm{OOTF} = \mathrm{OETF} \otimes \mathrm{EOTF}$$
$$\mathrm{EOTF} = \mathrm{OETF}^{-1} \otimes \mathrm{OOTF}$$
$$\mathrm{OETF} = \mathrm{OOTF} \otimes \mathrm{EOTF}^{-1}$$
$$\mathrm{OOTF}^{-1} = \mathrm{EOTF}^{-1} \otimes \mathrm{OETF}^{-1}$$
$$\mathrm{EOTF}^{-1} = \mathrm{OOTF}^{-1} \otimes \mathrm{OETF}$$
$$\mathrm{OETF}^{-1} = \mathrm{EOTF} \otimes \mathrm{OOTF}^{-1}$$

The PQ approach is defined by its EOTF. For PQ the OETF may be derived from the OOTF using the third line of the equations above. In a complementary fashion the HLG approach is defined by its OETF. For HLG the EOTF may be derived from the OOTF using the second line of the equations above.

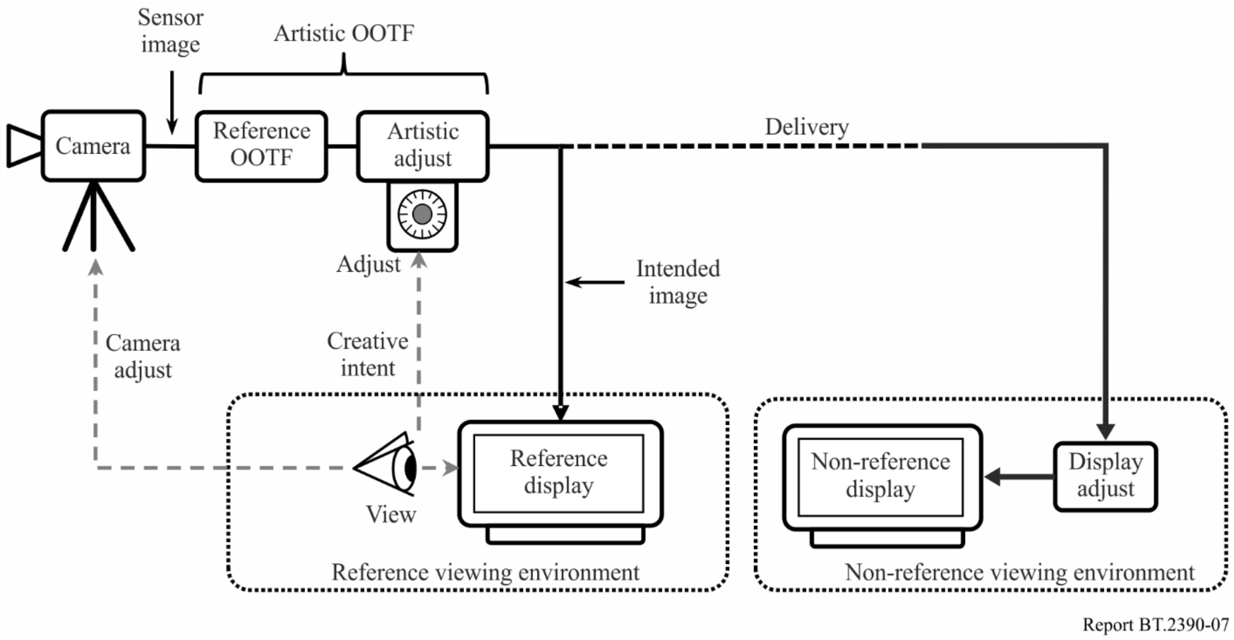
2.2 Conceptual TV system showing basic concepts

Figure 7 is a high level conceptual flow of a simplified television system that does not employ a non-linearity (such as gamma) in order to reduce the bit depth needed to represent the baseband signal;

such a non-linearity is needed in signal pipelines that have limited bit depths (e.g. limitations to 8-12 bit values), but these pipelines will be considered later and the conceptual system described here is considered to have no such restrictions. In Fig. 7, the camera outputs a linear light signal, which is representative of the scene in front of the lens. Exposure controls (camera iris and filters) perform a global scaling so the camera output is proportional to absolute scene light. The signal can be represented by high bit-depth integers, or for more efficiency, as 16-bit floating point. Non-reference viewing includes consumer viewing, as well as much TV production which often takes place in non-reference environments.

FIGURE 7

The conceptual TV system



A linear display of the scene light would produce a low contrast washed out image as illustrated in Fig. 8. Therefore, the signal is altered to impose rendering intent, i.e. a Reference OOTF (opto-optical transfer function) roughly like that shown in Fig. 9. The sigmoid curve shown increases contrast over the important mid-brightness range, and softly clips both highlights and lowlights, thus mapping the possibly extremely high dynamic range present in many real world scenes to the dynamic range capability of the TV system.

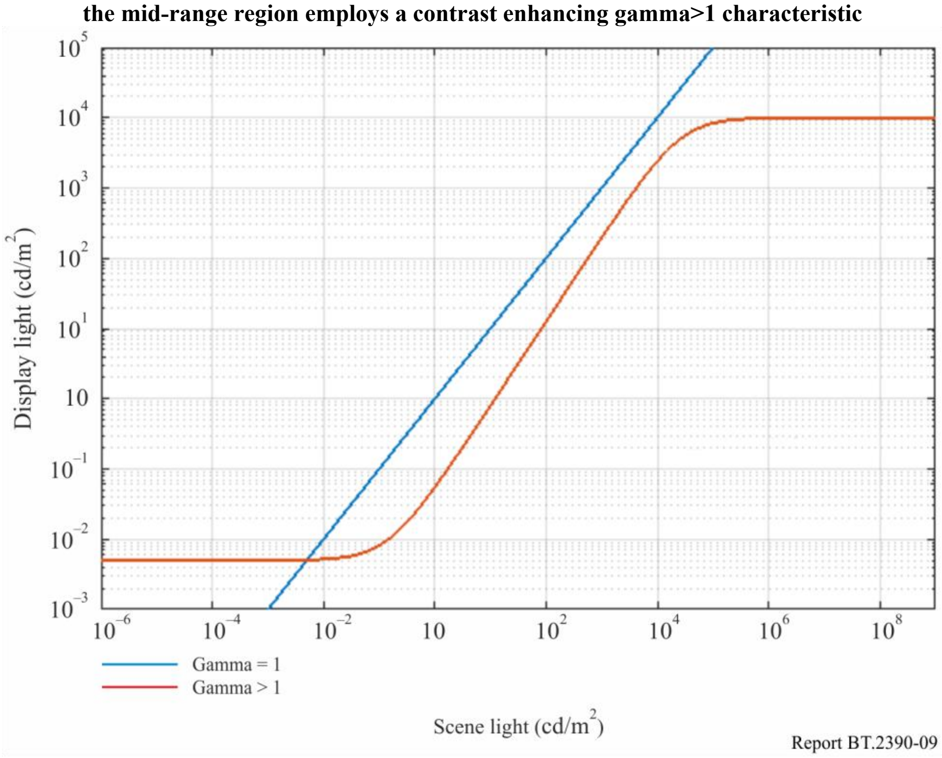
FIGURE 8

The left image has a system transfer function (or greyscale) of unity slope.
The right image has a system transfer function consistent with ITU broadcast practices

A reference display in a reference viewing environment would, ideally, be used for viewing in production, and adjustments (e.g. iris) are made to the camera to optimize the image. Use of the *Reference OOTF* to produce images, with viewing done in the reference viewing environment, allows consistency of produced images across productions. If an artistic image "look" different from that produced by the reference OOTF is desired for a specific programme, "Artistic adjust" may be used to further alter the image in order to create the image "look" that is desired for that programme. Artistic adjustments may be made through the use of camera settings or after image capture during editing or in post-production. The combination of the reference OOTF plus artistic adjustments may be referred to as the "Artistic OOTF".

FIGURE 9

Typical sigmoid used to map scene light to display light; extreme highlights and dark areas are compressed/clipped, the mid-range region employs a contrast enhancing gamma>1 characteristic



On the receive side where the consumer will view the image, if the consumer display is capable, and the consumer viewing environment is close to that of the reference viewing environment (dim room), then the consumer can view the image as intended. There may be limitations on both the viewing environment and the display itself. The viewing environment may be brighter than the reference environment, and the display may be limited in brightness, blackness, and/or colour gamut. Figure 7 shows ‘display adjust’ as an alteration made to accommodate these differences from the reference condition. To compensate for a brighter environment, display adjust may lift the black level of the signal. To accommodate limited brightness capability of the display, system gamma may be changed or a ‘knee’ may be imposed to roll off the highlights. To accommodate a limited colour gamut, gamut mapping would be performed to bring the wide gamut of colours in the delivered signal into the gamut that the display can actually show.

In practice television programmes are produced in a range of viewing environments using displays of varying capabilities. Thus similar adjustments are often necessary in production displays to achieve consistency.

3 The legacy television architecture

Since its beginning, television has employed restricted signal pipelines. Limited signal-to-noise ratios in the analogue days have transitioned to limited bit depths in the digital age. A non-linearity in the basic video signal was required in order to improve the visible signal-to-noise ratio in analogue systems, and the same non-linearity helps to prevent quantization artefacts in digital systems. This is the typical ‘gamma’ curve that is the natural characteristic of the CRT, and that is documented in Recommendations ITU-R BT.709, BT.1886 and BT.2020.

Until recently all displays were based on the CRT which, based on the common physics, all had a similar characteristic function converting the electrical signal to light, the so-called ‘electro-optical transfer function’ or EOTF. The camera characteristic of converting light into the electrical signal, the ‘opto-electronic transfer function’ or OETF, was adjusted to produce the desired image on the reference CRT display device. The combination of this traditional OETF and the CRT EOTF yielded the traditional OOTF. The non-linearity employed in legacy television systems (Recommendations ITU-R BT.601, BT.709 and BT.2020) is satisfactory in that 10-bit values are usable in production and 8-bit values are usable for delivery to consumers; this is for pictures with approximately 1 000:1 dynamic range[5], i.e. 0.1 to 100 cd/m^2.

3.1 HDTV as specified in Recommendations ITU-R BT.709 and BT.1886

Recommendation ITU-R BT.709 explicitly specifies a reference OETF function that in combination with a CRT display produces a good image. Creative intent to alter this default image may be imposed in either the camera, by altering the OETF, or in post-production, thus altering the OOTF to achieve an ‘artistic’ OOTF. As the CRT is no longer manufactured, it became impractical to rely on the inherent CRT characteristic in order to achieve uniformity in reference displays. In the year 2011 Recommendation ITU-R BT.1886 was approved; this new Recommendation specified the EOTF of the reference display to be used for HDTV production; the EOTF specification is based on the CRT characteristics so that future monitors can mimic the legacy CRT in order to maintain the same image appearance in future displays. A reference OOTF is not explicitly specified for HDTV. Nevertheless,

[5] This definition of dynamic range refers to the luminance ratio between the dimmest and brightest possible pixels presented on the display. However quantization artefacts, known as banding, may be visible, particularly in low lights, at luminance levels substantially brighter than the dimmest pixel. Quantization artefacts may, therefore, limit the “effective” dynamic range that is free from banding.

as shown in Fig. 10, in practice it exists as the cascade of the specified OETF (BT.709) and EOTF (BT.1886).

FIGURE 10

The BT.709 HDTV television system architecture

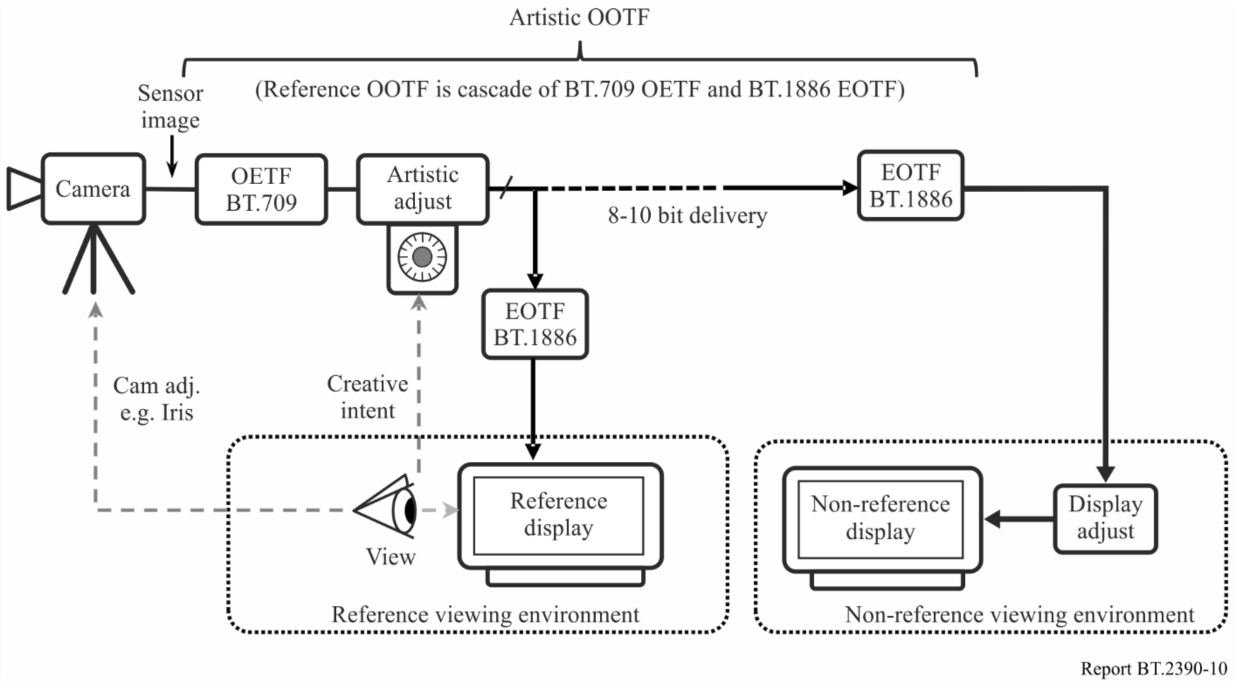


Figure 10 shows the HDTV system. The linear light is encoded into a non-linear signal using the OETF specified in Recommendation ITU-R BT.709. Creative intent may be imposed by altering this encoding or in post-production by adjusting the signal itself; this can be considered as an alteration outside of the BT.709 OETF (e.g. as 'artistic adjust' in the diagram). Recommendation ITU-R BT.1886 specifies the conversion of the non-linear signal into display light. This drives the reference display in the reference viewing environment. The image on the reference display drives adjustment of the camera iris/exposure, and if desired, artistic adjust can alter the image to produce a different artistic look. At the receiver (ideally a reference display in a reference viewing environment) the non-linear signal is converted to display light using the Recommendation ITU-R BT.1886 specified function. There is typically further adjustment (display adjust) to compensate for viewing environment, display limitations, and viewer preference; this alteration may lift black level, effect a change in system gamma, or impose a 'knee' function to soft clip highlights. (In practice the EOTF gamma and display adjust functions may be combined in to a single function.)

In a typical TV system the soft clipping of the highlights (sometimes known as the 'shoulder'), described earlier and illustrated in Fig. 3, is implemented in the camera as a camera 'knee'. This is part of the artistic adjustment of the image. Part of the low light portion of the characteristic (sometimes known as the 'toe') is implemented in the display as a black level adjustment. This adjustment takes place in the display as part of the BT.1886 EOTF and implements soft clipping of the lowlights.

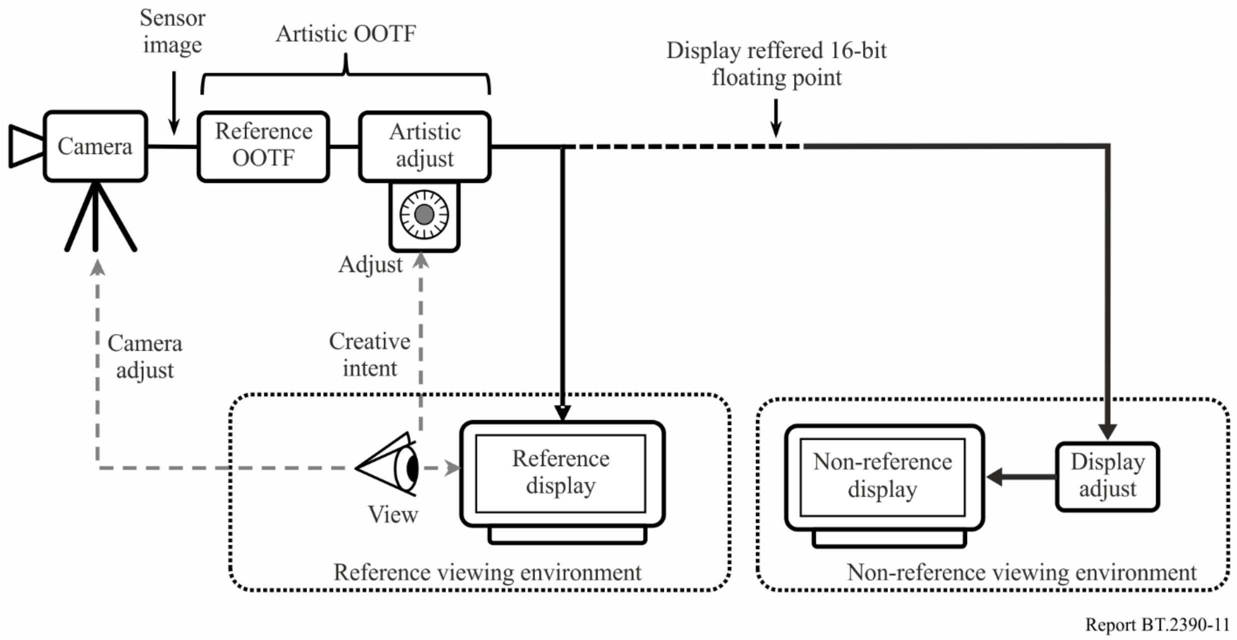
There is no clearly defined location of the reference OOTF in this system. The reference OOTF is the cascade of the OETF and the EOTF, and the actual OOTF is the cascade of those plus the artistic and display adjustments. Any deviation from the reference OOTF for reasons of creative intent must occur upstream of delivery. Alterations to compensate for the display environment or display characteristics must occur at the display by means of display adjust (or a modification of the EOTF away from the reference EOTF).

4 RGB floating point HDR-TV system

A 16-bit RGB HDR system is defined for use when 48-bit/pixel pipelines are available. This architecture is shown in Fig. 11.

FIGURE 11

HDR floating point system



The raw output of the camera is a relative scene referred floating point signal. These floating point values may be scaled such that maximum diffuse white results in R=G=B=1.0. The reference OOTF is implemented directly after camera capture of the scene, and an artistic adjustment may be used to make additional changes as desired for creative intent. Alternatively, the raw camera output can be used as input to a post-production process. The display referred output of the OOTF block (or post-production) is in the 16 bit floating point format which allows for adequate precision even for large colour volumes. Display referred floating point values directly represent light values on the display, i.e. R = G = B = 1.0 means 1.0 cd/m^2 of white for a pixel. As before, display adjust is used to compensate as much as possible for limitations of displays, and for environments that may differ from the reference viewing environment that was (ideally) used during programme production.

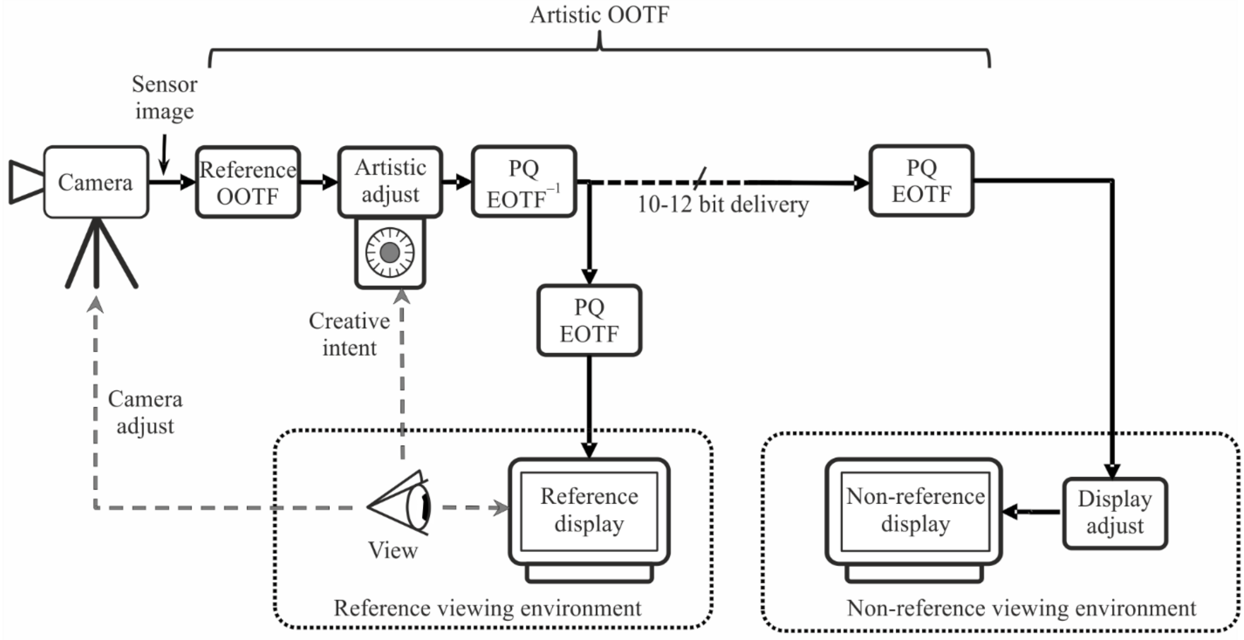
5 PQ HDR-TV

5.1 PQ system architecture

When bit-constrained pipelines are required for television production systems, then an HDR implementation very similar to the current HDTV system of Fig. 10 can be constructed. This implementation is shown in Fig. 12.

FIGURE 12

PQ HDR-TV system with 10 or 12 bit integer values



Report BT.2390-12

An optimized non-linear signal representation is used so that 10-12 bit depth values can accommodate the larger colour volume of HDR; otherwise this system is very similar to the HDTV system in use today. The PQ EOTF replaces the BT.1886 function of SDR HDTV, and the corresponding PQ OETF replaces the BT.709 OETF as the default camera capture curve. Once again an artistic adjustment may be used to further modify the creative intent of the image, and a display adjustment is used to adapt the signal for different display characteristics and display environments. No use of metadata is shown or required.

5.2 Design of the PQ non-linearity

As described in [14] the traditional gamma nonlinearities of Recommendations ITU-R BT.709 and BT.1886 are unsatisfactory when stretched to the much larger dynamic ranges desired for future television productions.

FIGURE 13

Contrast step size vs. display luminance for 12 bit signals

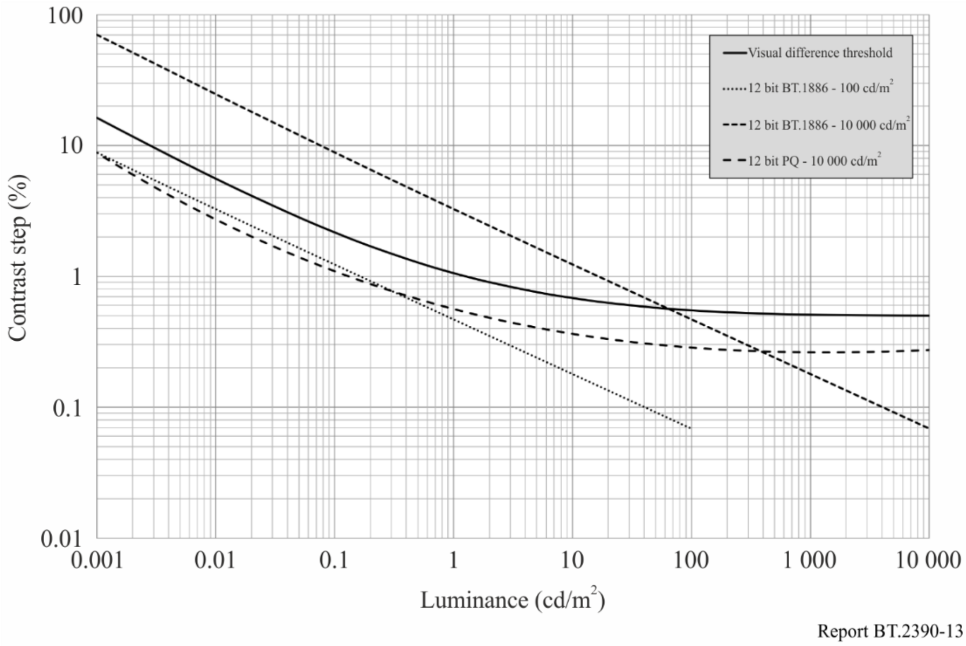
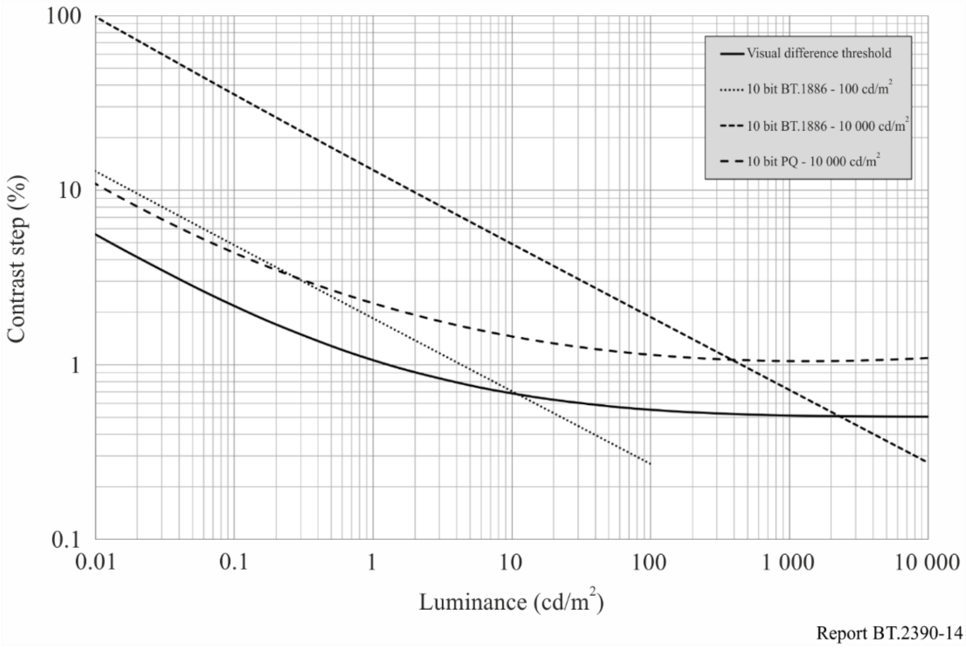


Figure 13 shows the approximate visual difference threshold as a solid black curve on a log-log plot with luminance on the x-axis and contrast step size (due to bit depth limitation) in % on the vertical axis. This threshold is based on the detailed Barten model of the human visual system. Lines which fall below this threshold curve will not exhibit any visible quantization artefacts such as image banding, while lines above the threshold curve may exhibit visual artefacts. While the legacy Recommendation ITU-R BT.1886 operating with a peak level of 100 cd/m^2 is comfortably below the threshold curve when using 12 bits, it rises substantially above the visual threshold when operating with a 10 000 cd/m^2 peak. A traditional “gamma” power function is not a good approximation for human vision over an extended range of luminance values (too many code words allocated to very bright regions and not enough allocated to dark regions). This inefficiency was not a serious problem with SDR systems due to their limited dynamic range, but when trying to represent HDR luminance ranges, an improved curve is required. By using the same Barten model as the visual threshold calculation itself, an optimized nonlinear function was developed for the PQ signal, which can operate over the entire range from 10 000 cd/m^2 down to less than 0.001 cd/m^2 without any visible quantization artefacts using 12 bit coding precision.

FIGURE 14

Contrast step size vs. display luminance for 10 bit signals



Report BT.2390-14

Figure 14 shows the same plots as Fig. 13 but with all three systems using 10 bit quantization. Though the signal lines all come above the threshold curve to some extent, experience has shown that with realistic camera noise levels, the slight quantization artefacts predicted for 100 cd/m^2 Recommendation ITU-R BT.1886 or 10 000 cd/m^2 PQ are masked and thus do not present real problems in television production.

5.3 OOTF and OETF

This subsection describes the PQ opto-optical transfer function (OOTF) and the resulting opto-electronic transfer function (OETF). The PQ opto-optical transfer function is normatively specified in Recommendation ITU-R BT.2100, which is intended to be compatible with existing SDR BT.709 signal sources and BT.1886 compliant displays. This maximizes compatibility for mixed source applications wherein some sources are HDR and some are SDR. It is desired that the image from an SDR source and that from an HDR source match everywhere the HDR image brightness overlaps the range of the SDR source (the HDR OOTF extends up to the maximum PQ displayed light level of 10 000 cd/m^2).

5.3.1 Generalized OOTF from Recommendation ITU-R BT.1886 in combination with Recommendation ITU-R BT.709

In order to maximize compatibility with existing SDR signals, it is desired an OOTF consistent with the effective OOTF of existing practice which is:

$$OOTF_{SDR} = EOTF_{1886}\left[OETF_{709}\right] \tag{1}$$

It is only needed to extend the range of $OETF_{709}$ and $EOTF_{1886}$ for HDR.

The extension factor for displayed light is 10 000 / 100 = 100.

As the SDR OOTF has a roughly gamma = 1.2 characteristic at the high end, the extension relative to scene light (the input to OOTF) is approximately 100 $^{1/1.2}$ = 46.42. When the exact equations for Recommendations ITU-R BT.709 and BT.1886 are used, the extension for HDR is 59.5208.

To expand the range of $OETF_{709}$ to G_{709} for HDR the equation is therefore (HDR E normalized to range of 0 to 1):

$$E' = G_{709}\left[E\right] = \begin{array}{ll} 1.099\left(59.5208E\right)^{0.45} - 0.099 & \text{for } 1 > E > 0.018/59.5208 \\ 4.5\left(59.5208E\right) & \text{for } 0.018/59.5208 \geq E > 0 \end{array} \qquad (2)$$

Consequently, the range of E' is [0, 6.813] for HDR while it remains [0,1] for SDR. To expand the range of $EOTF_{1886}$ to G_{1886} for HDR no change to the equation is necessary, the argument is simply allowed to extend to 6.813 (from 1) and hence the range increases from 100 to 10 000:

$$G_{1886}\left[E'\right] = 100(E')^{2.4}$$

These extensions satisfy the boundary conditions:

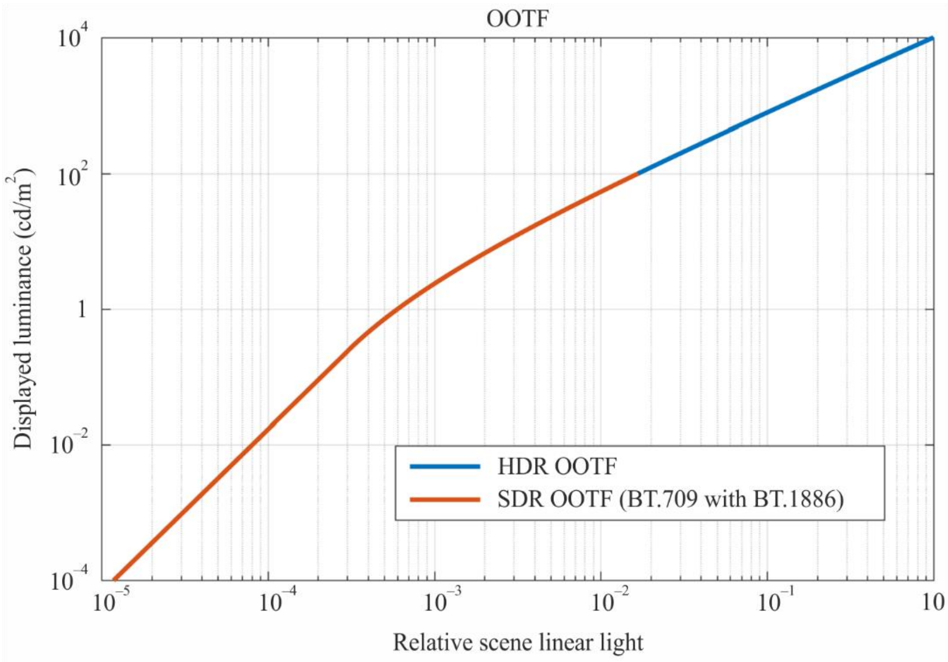
a) E = 1 produces a displayed luminance of 10 000 cd/m^2

b) E = 1/(59.5208) produces a displayed luminance of 100 cd/m^2

The resulting OOTF is shown in Fig. 15. The x-axis, relative scene light is the same as E for SDR while for HDR it is 59.5208*E since the domain of E is [0,1]:

FIGURE 15

PQ and SDR OOTF



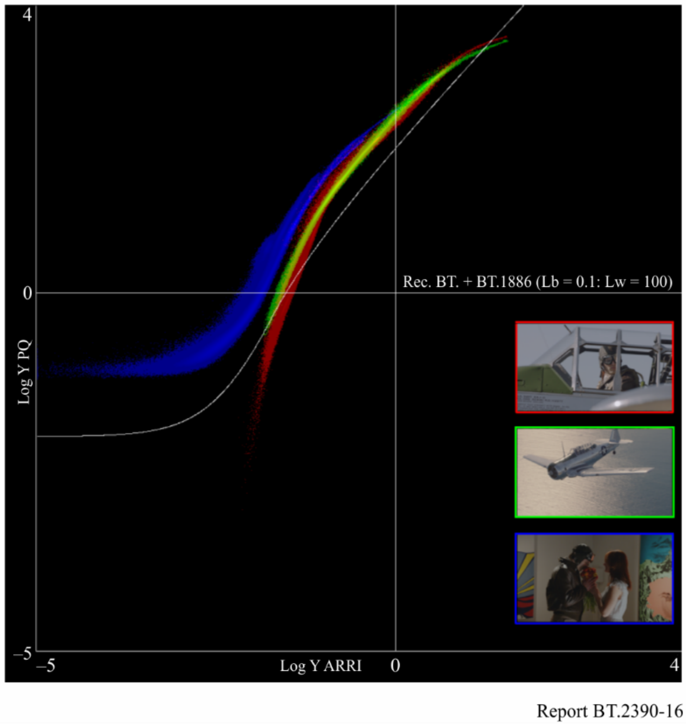
Report BT.2390-15

5.3.2 Actual OOTFs from manually graded content

It is instructive to compare this proposal with the actual OOTFs that are imposed when manually grading camera RAW output. The OOTF is the ratio of the graded linear output to the RAW linear input. Figure 16 shows several examples from the HDR sequence “Fantasy Flights”:

FIGURE 16

Extracted OOTFs from Fantasy Flights (3 of 3)



Report BT.2390-16

These Figures show scatter plots of the log of the output luminance derived from the PQ grade versus the log of the relative input luminance derived from the ARRI RAW camera output. These scatter plots are colour-coded (RGB) to match the images shown in the lower right corner of each figure. For comparison, the OOTF from the combination of Recommendations ITU-R BT.1886 and BT.709 are plotted in white. This shows that the extracted OOTFs are, as one would expect, a bit brighter than SDR. Some preliminary conclusions can be drawn from this experimental data:

1 For this manually graded content, the OOTF is not a straight line, and thus the actual OOTF does not correspond to an overall "system gamma".

2 Darker indoor scenes tend to be noise limited at the bottom end and the OOTF exhibits a very clear toe.

3 The extracted OOTFs appear to have roughly the same curvature in the mid-tones as the proposed model.

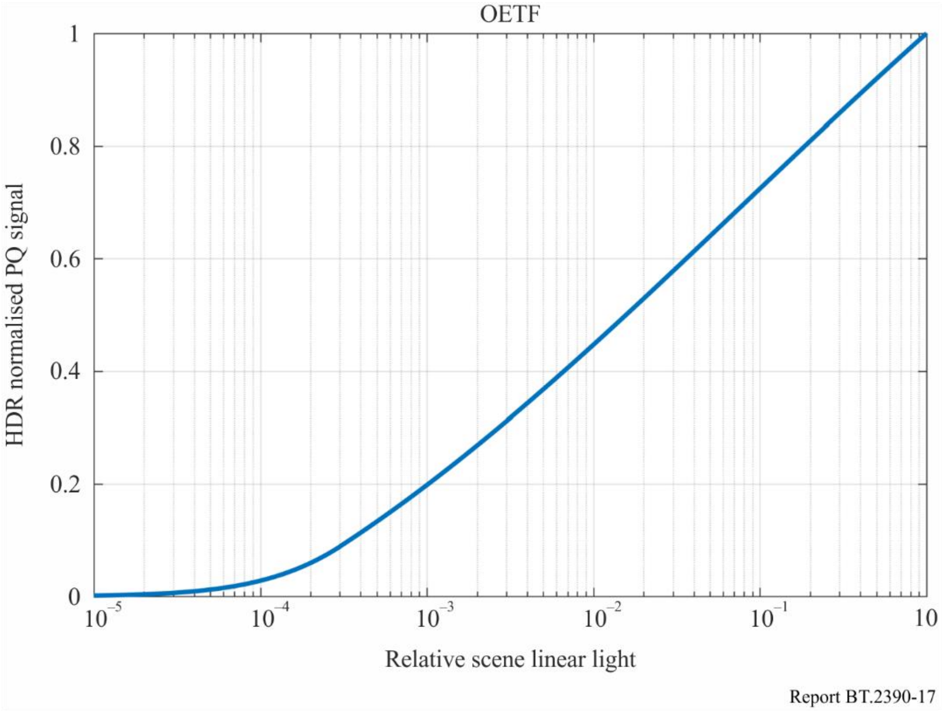
5.3.3 Resultant OETF

This OOTF can be combined with the inverse of the EOTF to produce an OETF. That OETF is shown in Fig. 17.

In actual cameras there is noticeable noise at low signal levels, and in practice the OETF slope at low levels is limited so as to 'crush' the noise in black, thereby putting a 'toe' into the response. The reference OETF does not have such a 'toe', but one is apparent in the OOTF plot for the indoor scene of "Fantasy Flights" shown above.

FIGURE 17

HDR OETF



This OETF:

– emulates the ‘look’ of Recommendation ITU-R BT.709 plus Recommendation ITU-R BT.1886 for display light up to the limit of SDR;

– facilitates mixing of legacy BT.709 signals and PQ HDR signals;

– offers reasonable behaviour for levels above those of SDR.

5.4 Display mapping

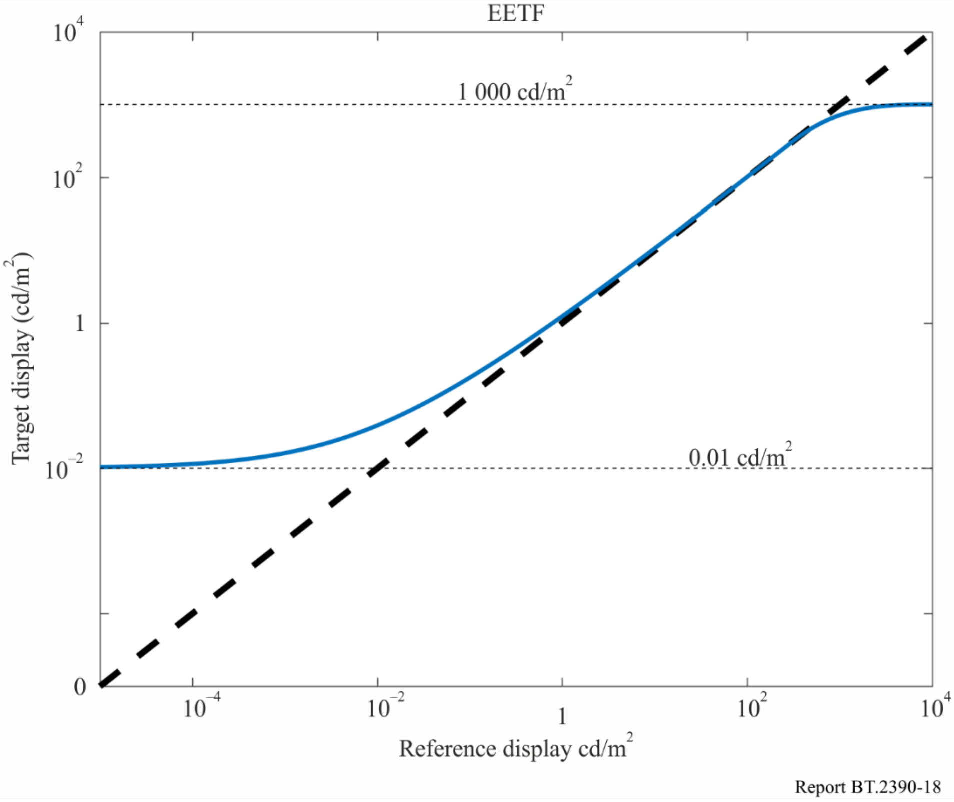
The PQ HDR system generates content that is optimum for viewing on a reference monitor in a reference viewing environment. The reference monitor would ideally be capable of accurately rendering black levels down to or below 0.005 cd/m^2, and highlights up to 10 000 cd/m^2. Also, the ideal monitor would be capable of showing the entire colour gamut within the BT.2020 triangle. The viewing environment would ideally be dimly lit, with the area surrounding the monitor being a neutral grey (6 500 degree Kelvin) at a brightness of 5 cd/m^2. However, content often must be viewed or produced in environments brighter than the reference condition, and on monitors that cannot display the deepest blacks or brightest highlights that the PQ signal can convey. In these cases the display characteristic needs to be changed in a process often referred to as display mapping (DM).

5.4.1 Mapping to display with limited brightness range

High dynamic range content may be viewed on displays that have less dynamic range than the reference display used to master the content. In order to view HDR content on displays with a lower dynamic range, display mapping should be performed. This can take the form of an EETF (electrical-electrical transfer function) in the display. This function provides a toe and knee to gracefully roll off the highlights and shadows providing a balance between preserving the artistic intent and maintaining details. Figure 18 is an example EETF mapping from the full 0 – 10 000 cd/m^2 dynamic range to a target display capable of 0.01 – 1 000 cd/m^2. The EETF may be introduced into the PQ signal; the plots show the effect of the mapping, i.e. how the intended light is changed into actual displayed light. In practice the mapping is done on the PQ signal.

FIGURE 18

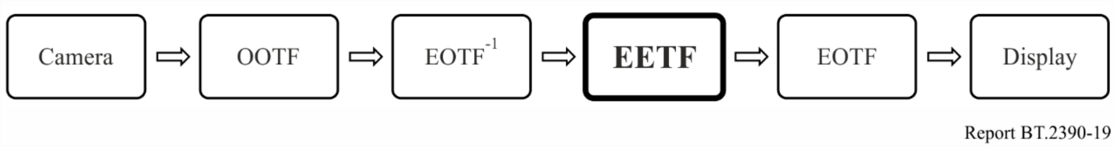
Example EETF From 0 – 10 000 cd/m² to 0.01 – 1 000 cd/m²



Below are the mathematical steps that implement this tone mapping function for displays of various black and white luminance levels. Figure 19 shows the block diagram of where the EETF should be applied.

FIGURE 19

Block diagram of signal chain showing location of EETF application



Calculating the EETF

The central region of the tone mapping curve is defined as a 1:1 mapping. A 'knee' roll off may be calculated using a hermite spline to create a mapping that will reduce the luminance range to the capability of the display. The black level lift is controlled by an offset, b, which would be determined by a PLUGE adjustment. The difference between this proposal and the black level adjustment per Recommendation ITU-R BT.1886 is the addition of a tapering factor $(1 - E_2)^4$. Without such a tapering factor, a constant offset throughout the entire signal range has the effect of increasing the brightness at the high end. With Recommendation ITU-R BT.1886 this effect was limited and not problematic due to the large number of code values at the high end of the gamma curve. The perceptual uniformity of the PQ EOTF causes this effect to be unacceptable. The tapering function allows fine-tuning the lift without a significant impact on mid-tones or highlights.

In the case where the mastering display minimum black and peak white luminances are known or reasonably can be assumed, the first step in applying the EETF is to normalize the PQ values based on the mastering display black and white luminances, L_B and L_W:

$$E_1 = (E' - \mathrm{PQEOTF}^{-1}[L_B])/(\mathrm{PQEOTF}^{-1}[L_W] - \mathrm{PQEOTF}^{-1}[L_B])$$

where E' is the I, Y′or R′, G′, or B′ PQ component and E_1 is the corresponding mastering display black and white normalized PQ component.

In the case where the mastering display minimum black and peak white luminances are not known and reasonably cannot be assumed, a value of 0 can be used for L_B and a value of 10 000 can be used for L_W, corresponding to the entire PQ encoding luminance range.

The next step is to calculate the mastering display black and white normalized PQ values, *minLum* and *maxLum*, corresponding to the target display minimum (L_{min}) and maximum (L_{max}) luminances, including ambient, as follows:

$$minLum = (\mathrm{PQEOTF}^{-1}[L_{min}] - \mathrm{PQEOTF}^{-1}[L_B])/(\mathrm{PQEOTF}^{-1}[L_W] - \mathrm{PQEOTF}^{-1}[L_B])$$

$$maxLum = (\mathrm{PQEOTF}^{-1}[L_{max}] - \mathrm{PQEOTF}^{-1}[L_B])/(\mathrm{PQEOTF}^{-1}[L_W] - \mathrm{PQEOTF}^{-1}[L_B])$$

The next step is to calculate the 1:1 mapping and knee (E_2). The turning point (KneeStart or KS) for the spline is the point where the roll off will begin [15], as follows:

$$KS = 1.5\ maxLum - 0.5$$

$$\mathrm{b} = minLum$$

The next step is to solve for the EETF (E_3) with given end points.

Step 3.1:

$$E_2 = E_1 \qquad \text{for } E_1 < KS$$

$$E_2 = P[E_1] \qquad \text{for } KS \le E_1 \le 1$$

Step 3.2:

$$E_3 = E_2 + b(1 - E_2)^4 \qquad \text{for } 0 \le E_2 \le 1$$

Hermite spline equations:

$$P[B]=(2T[B]^3-3T[B]^2+1)KS+(T[B]^3-2T[B]^2+T[B])(1-KS)+(-2T[B]^3+3T[B]^2)maxLum$$

$$\mathrm{T[A] = (A - KS)/(1 - KS)}$$

The last step is to invert the normalization of the PQ values based on the mastering display black and white luminances, L_B and L_W, to obtain the target display PQ values.

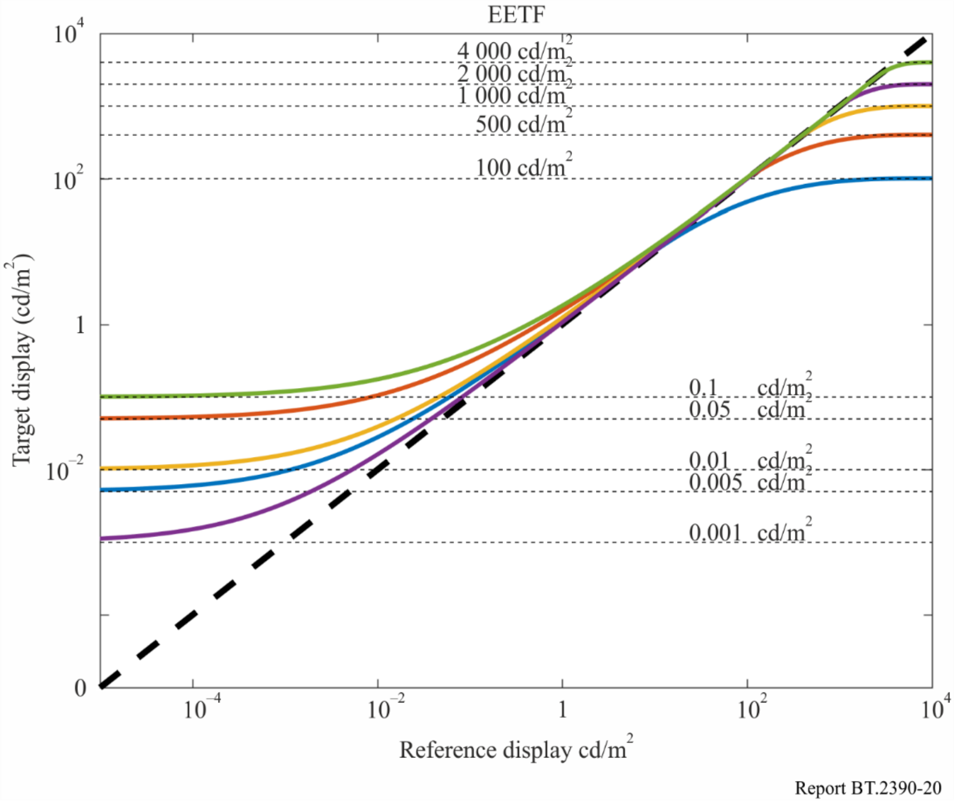
$$E_4 = E_3\,(\mathrm{PQEOTF}^{-1}[L_W] - \mathrm{PQEOTF}^{-1}[L_B]) + \mathrm{PQEOTF}^{-1}[L_B]$$

Practical application

The sample curves shown in Fig. 20 are designed for tone mapping to display black level up to 0.1 cd/m^2 and display white level as low as 100 cd/m^2.

FIGURE 20

Example EETFs of various target displays



Here are the notable options:

1) IC_TC_P

$$I_2 = EETF(I_1)$$

$$C_{T2}, C_{P2} = min\left(\frac{I_1}{I_2}, \frac{I_2}{I_1}\right) \times (C_{T1}, C_{P1})$$

2) $Y'C'_BC'_R$

$$Y'_2 = EETF(Y'_1)$$

$$C'_{B2}, C'_{R2} = min\left(\frac{Y'_1}{Y'_2}, \frac{Y'_2}{Y'_1}\right) \times (C'_{B1}, C'_{R1})$$

3) $YRGB$

$$Y_1 = 0.2627R_1 + 0.6780G_1 + 0.0593B_1$$

$$Y_2 = EOTF_{PQ}(EETF(EOTF_{PQ}^{-1}(Y_1)))$$

$$(R_2, G_2, B_2) = \frac{Y_2}{Y_1} \times (R_1, G_1, B_1)$$

4) $R'G'B'$

$$(R'_2, G'_2, B'_2) = EETF(R'_1, G'_1, B'_1)$$

6 HLG HDR-TV

The hybrid log-gamma (HLG) HDR-TV signal parameters were designed from the outset to offer broadcasters and programme producers an evolutionary approach to HDR production and distribution. The signal characteristic is similar to that of a traditional standard dynamic range camera with a ‘knee’ and requires no production metadata. It is therefore compatible with conventional standard dynamic range production equipment, tools and infrastructure. Furthermore, the HLG HDR-TV signal parameters were designed to provide a significant degree of compatibility on BT.2020 colour SDR displays (see § 6.4). Thus HDR monitors are only necessary in critical monitoring areas. The design of the HLG HDR signal parameters is intended to allow distribution networks to provide a single HEVC Main 10 bitstream that can target both SDR and HDR receivers, where those SDR receivers support the BT.2020 colour container (e.g. DVB and ARIB HEVC UHD receivers).

6.1 The hybrid log-gamma opto-electronic transfer function (OETF)

In the brighter parts and highlights of an image the threshold for perceiving quantization is approximately constant (known as Weber’s law). This implies a logarithmic OETF would provide the maximum dynamic range for a given bit depth. Proprietary logarithmic OETFs are in widespread use. But in the low lights it becomes increasingly difficult to perceive banding. That is, the threshold of visibility for banding becomes higher as the image gets darker. This is known as the De Vries-Rose law. The conventional gamma OETF used for SDR comes close to matching the De Vries-Rose law, which is perhaps not coincidental since gamma curves were designed for dim CRT displays. So an ideal OETF might be logarithmic in the high tones and a gamma law in the low lights, which is essentially the form of the hybrid log-gamma OETF.

The dynamic range of modern video cameras is considerably greater than can be conveyed by a video signal using a conventional OETF gamma curve (e.g. Recommendation ITU-R BT.709 or Recommendation ITU-R BT.2020). In order to exploit their full dynamic range conventional video cameras sometimes use a “knee” characteristic to extend the dynamic range of the signal. The knee characteristic compresses the image highlights to prevent the signal from clipping or being “blown out” (overexposed). Knee characteristics are discussed, for example, in “Circles of Confusion”, by Alan Roberts, published by the EBU. The “shoulder” characteristic of conventional photochemical film used in movie cameras provides a similar effect. When a hybrid log gamma HDR video signal is displayed on a conventional SDR display the effect is similar to the use of a digital camera with a knee or using film. It is not surprising therefore, that the HLG video signal is highly compatible with conventional SDR displays, because what you see is very similar to the signal from an SDR camera. Indeed the “knee” characteristic of the HLG OETF, defined in Table 5 of Recommendation ITU-R BT.2100 (and shown in Fig. 21A below), provides an extended highlight range that is comparable to some “knees” used for SDR. Note that the “knee” curve in the figure is diagrammatic for illustrative purposes only. Whilst knees are sometimes described in the literature as linear, as in this figure, in practice they are “smooth” and avoid the discontinuous gradient shown here, which can result in objectionable colour shifts.

An HLG signal is defined as:

OETF:

With E is normalized to the range [0:1] then the equation for the OETF is:

$$E' = \mathrm{OETF}\left[E\right] = \begin{cases} \sqrt{3E} & 0 \le E \le 1/12 \\ a \cdot \ln\left(12E - b\right) + c & 1/12 < E \le 1 \end{cases}$$

where:

E: signal for each colour component $\{Rs, Gs, Bs\}$ proportional to scene linear light and scaled by camera exposure, normalized to the range [0:1]

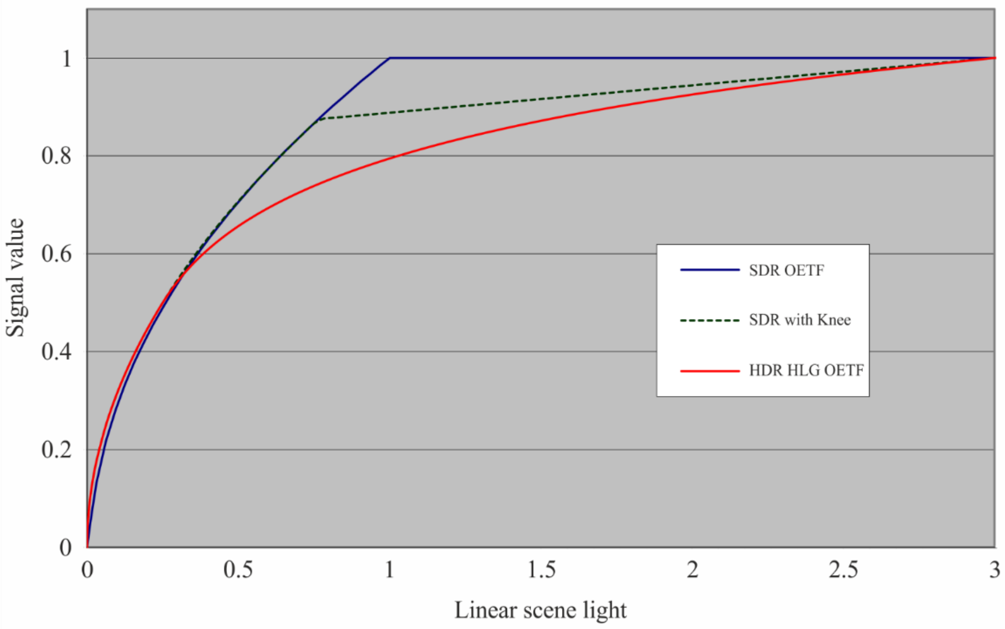
E': resulting non-linear signal $\{R', G', B'\}$ in the range [0:1].

$$a = 0.17883277,\ b = 1 - 4a,\ c = 0.5 - a \cdot \ln(4a)$$

The HLG OETF is shown in Fig. 21 alongside the conventional SDR OETF and an (illustrative) knee characteristic. These plots assume that two cameras, one Recommendation ITU-R BT.2020 and the other BT.2100 (that is, one SDR and one HDR), are set up with the same sensitivity. For example, if both cameras were looking at the same 18% grey chart, then their sensitivities (gain, iris, and shutter time) could be adjusted so that the signal level was 42.5% of nominal full signal level for both cameras. A notional SDR "knee" is shown on the same plot, with a breakpoint of 87.5% signal level, which extends the SDR dynamic capture range substantially.

FIGURE 21

Comparison of SDR and HLG HDR OETFs



Report BT.2390-21

When the two cameras' (SDR and HDR) sensitivities are equalized then both the SDR (BT.2020) and HDR responses to light amplitude would be almost the same for signal levels at or below 50%. Above 50% signal level the HDR OETF is logarithmic, which means it can capture higher light levels (such as specular reflections and highlights) without clipping. There are small differences between the two plots below 50% of nominal signal range. This is because SDR OETFs include a linear portion near black to avoid excessive noise amplification. HLG, by contrast, uses a pure square root OETF at low levels. This allows HLG to achieve higher dynamic range "in the blacks", but it does mean that camera manufacturers must use an alternative to the linear part of the SDR OETF to avoid excessive noise amplification in the black.

Note that the conventional 'narrow range' digital signal can actually support signal levels of up to 109% of nominal full scale. This is to accommodate overshoots and highlights. If this additional signal range is used (though not all equipment supports it) then even higher light levels may be captured without clipping.

Considering a nominal full scale signal (i.e. 100% signal level), and with the cameras set up as above, then the SDR camera can capture objects no brighter than 100% reflective (i.e. no highlights). The HLG camera increases the luminance that can be captured by a factor of 3. If the signal is allowed to excurse to the maximum 109% range (super-whites) then SDR can capture luminance equivalent to 120% reflectivity, whereas HLG can capture nearly a factor of 5 more luminance than 100%

reflectivity. It is the limitations in the ability of SDR displays to accurately render highlights that prompts the use of camera knees.

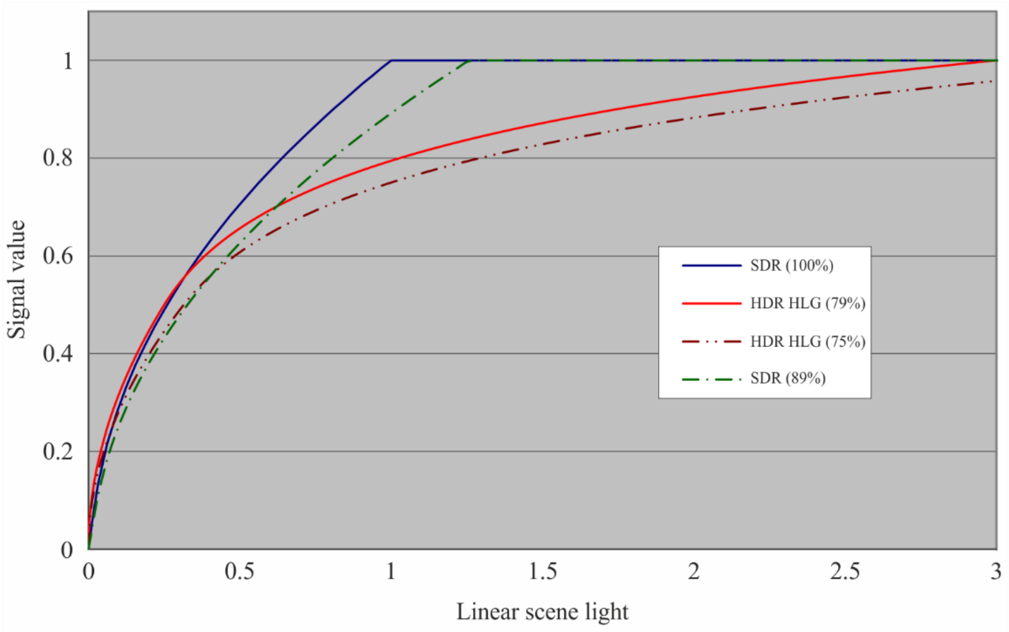
A naïve interpretation of these plots might suggest that the dynamic range of HLG is only 3 times greater than SDR, but this is not the case. HDR is about more than just increasing the brightness of highlights. Creating the detail in lowlights and "in the black" is also very important and HLG adds much dynamic range here. Secondly, the OETF describes the capture dynamic range. The dynamic range on the display is greater because of overall system gamma, discussed below. With a typical system gamma of 1.2, and the camera sensitivity adjusted as described, HLG supports display highlights which are a factor of 3.7 (or 6.9 with super-whites) higher than diffuse white.

However, the foregoing discussion assumes that "diffuse white" produces 100% signal output for SDR cameras. Whilst this may be true for some programmes, the signal level for diffuse white is not defined for SDR signals. In practice it varies between about 90% and 115% depending on genre, geographical region, and artistic preference. Drama, in particular, tends to set diffuse white at a lower signal level. This supports more artistically pleasing pictures that can contain some highlight detail. HLG supports a much greater dynamic range than SDR, and can take advantage of this by setting diffuse white at a lower signal level to support more highlight dynamic range.

Report ITU-R BT.2408 indicates that, for HLG HDR, diffuse white should be set at a signal level of 75%. This can be configured by making the output from an 18% grey card correspond to a signal level of 38%, rather than the 42.5% stated above. The OETFs for this camera setup are illustrated in Fig. 22 below, which also include the plots above for comparison. Setting 18% grey to 42.5% and 38% results in the diffuse white signal level being 100% and 89% respectively for SDR, and 79% and 75% respectively for HLG. The traces on the plots are labelled accordingly.

FIGURE 22

Comparison of HLG OETFs with BT.2408 signal levels for diffuse white



Report BT.2390-22

With cameras configured to produce this slightly lower signal level for diffuse white, the dynamic range available for highlights is increased. SDR can now support scene luminance equivalent to 125% of diffuse white, and HDR can support scene luminance of 375% diffuse white. These figures increase to 150% and about 620% if super-whites are used. So the use of super-whites is much more advantageous for HLG than it is for SDR. Note that these figures increase further to 163% and 890% at the display when a typical system gamma of 1.2 is used.

6.2 System gamma and the opto-optical transfer function (OOTF)

As is well known, and explained in § 2.2, the light out of a television display is not proportional to the light detected by the camera. The overall system non-linearity, or "rendering intent" is defined by the opto-optical transfer function, or OOTF. The OOTF maps relative scene linear light to display linear light. Rendering intent is needed to compensate for the psychovisual effects of watching an emissive screen in a dark or dim environment, which affects the adaptation state (and hence the sensitivity) of the eye. Traditionally movies were, and often still are, shot on negative film with a gamma of about 0.6. They were then displayed from a print with a gamma of between 2.6 and 3.0. This gives movies a system gamma of between 1.6 and 1.8, which is needed because of the dark viewing environment. Conventional SDR television has an OOTF which is also a gamma curve with a system gamma of 1.2. But, for HDR, the brightness of displays and backgrounds/surround will vary widely, and the system gamma will need to vary accordingly.

Colour images consist of red, green and blue components and this affects how the OOTF should be applied. Simply applying a gamma curve to each component separately as is done for SDR television distorts the colour; in particular it distorts saturation but also to a lesser extent the hue. As an illustration, suppose the red, green and blue components of a pixel have (normalized) values of (0.25, 0.75, 0.25). Applying a display gamma of 2, (i.e. squaring the value of the components) is obtained (0.0625, 0.5625, 0.0625). In this example, the pixel has got slightly darker and the ratio of green to blue and red has increased (from 3:1 to 9:1). This means, a green pixel would have appeared as a discernibly different shade of green. This approach is far from ideal if it is wished to avoid distorting colours when they are displayed.

Instead of the current SDR practice of applying a gamma curve independently to each colour component, for HDR it should be applied to the luminance alone. The luminance of a pixel is given by a weighted sum of the colour components; the weights depend on the colour primaries and the white point. According to Recommendation ITU-R BT.2100, luminance is given by:

$$Y_s = 0.2627R_S + 0.6780G_S + 0.0593B_S$$

where Y_S represents normalized linear scene luminance and R_S, G_S and B_S represent the normalized, linear scene light (i.e. before applying OETF) colour components. By applying rendering intent (OOTF) to the luminance component only it is possible to avoid colour changes in the display.

The HLG reference OOTF is therefore given by:

$$F_D = OOTF[E] = \alpha Y_S^{\gamma-1} E$$

$$R_D = \alpha Y_S^{\gamma-1} R_S$$

$$G_D = \alpha Y_S^{\gamma-1} G_S$$

$$B_D = \alpha Y_S^{\gamma-1} B_S$$

where:

F_D: luminance of a displayed linear component {R_D, G_D, or B_D}, in cd/m^2

E: signal for each colour component {Rs, Gs, Bs} proportional to scene linear light and scaled by camera exposure, normalized to the range [0:1].

α: user adjustment for the luminance of the display, commonly known in the past as a "contrast control". It represents L_W, the nominal peak luminance of a display for achromatic pixels in cd/m^2

γ: is an exponent, which varies depending on L_W as described below, and which is equal to 1.2 at the nominal display peak luminance of 1 000 cd/m^2

In order to determine the appropriate system gamma for a 1 000 cd/m^2 reference display, NHK conducted a series of experiments with an indoor test scene. Lighting was adjusted so that the luminance level of the diffuse white was 1 200 cd/m^2. The subjects were requested to adjust the system gamma and camera iris with reference to the real scene so that a tone reproduction similar to the scene could be obtained on the display. It was found that personal preference has an impact in determining the optimum system gamma for a given brightness display. But for a 1 000 cd/m^2 OLED display (Sony BVM-X300) the average optimum system gamma was found to be 1.18. Similar tests were repeated using a 2 000 cd/m^2 peak luminance LCD display (Canon DP-V3010), where it was found that the average preferred system gamma was 1.29.

Similarly, the BBC conducted subjective tests to determine the value of system gamma that delivers the best compatible SDR image. For those tests two Sony BVM-X300 OLED displays were used, one in its SDR mode (Recommendation ITU-R BT.1886, 100 cd/m^2 peak luminance) and the other a running prototype HLG HDR firmware (1 000 cd/m^2 peak luminance). In those tests the BBC found that the value of system gamma that delivers the best SDR compatible picture with a ~1 000 cd/m^2 display was 1.29. A value of 1.18 was found to be the best value when the peak brightness of the display was reduced to 500 cd/m^2.

Notably both NHK and the BBC reported values of 1.29 and 1.18 independently, albeit at different peak brightness values.

When designing the HLG HDR system, it was considered more important to weigh the choice of gamma value in favour of HDR production, rather than backwards compatibility with SDR displays. So a value of 1.20 was adopted for the reference 1 000 cd/m^2 display.

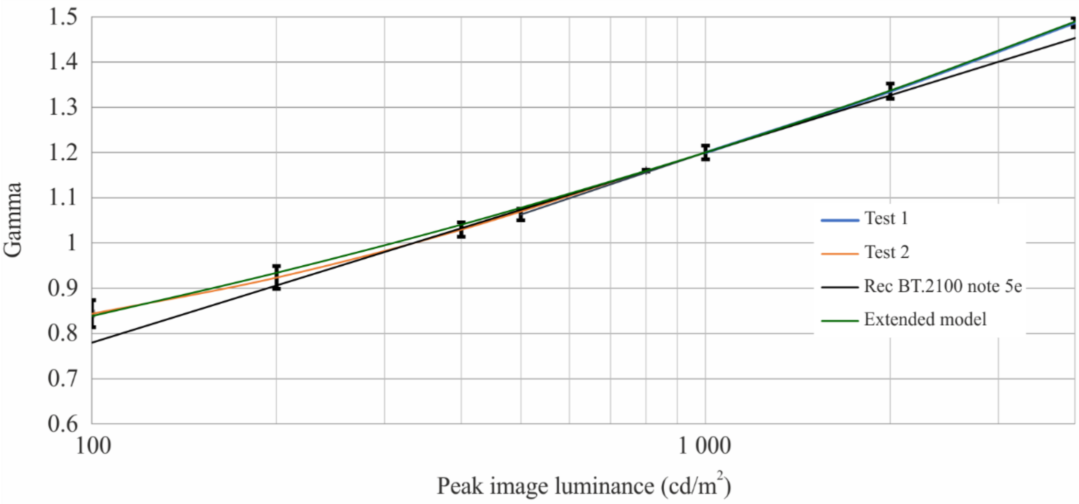
The clear indication from both of these studies is that system gamma needs to vary according to display peak brightness. In order to establish a more precise relationship between the gamma and display brightness, the BBC conducted further subjective tests where images were viewed with different gammas at different luminances (and with a fixed background luminance of 5 cd/m^2). The pictures were derived from HDR linear light images selected from Mark Fairchild's HDR Photographic Survey. Test subjects were asked to perceptually match as closely as possible an image displayed with a reference peak brightness to the same image with a non-reference peak brightness by adjusting the system gamma applied to the non-reference brightness image. The images were displayed on a calibrated SIM2 HDR47E display using its LogLuv input. The minimum black level viewable in the test environment was determined using an HDR PLUGE test signal, and an appropriate "brightness" offset added to the test images.

The initial tests varied peak brightness between 500 and 4 000 cd/m^2. The results were confirmed in subsequent BBC tests for a 1 000 cd/m^2 to 500 cd/m^2 change using a prototype Sony BVM-X300 OLED display. These results are also consistent with the ratio of gamma values found by NHK for a 2 000 cd/m^2 LCD display and a 1 000 cd/m^2 OLED display, and with the ratio of values determined by the BBC for optimum SDR compatibility at 1 000 cd/m^2 and 500 cd/m^2. The BBC then extended these tests to lower peak luminances [17].

The results of the BBC tests are illustrated in Fig. 23. Here, test 1 corresponds to peak luminances from 1 000 to 4 000 cd/m^2, and test 2 from 100 to 1 000 cd/m^2. Both tests are normalised so that gamma=1.2 at 1 000 cd/m^2.

FIGURE 23

Gamma value to match images for different screen peak brightness



Report BT.2390-23

Bringing together the results of all studies, it is found that the appropriate system gamma (γ) for different brightness displays, in the reference environment, can be determined using the following equation:

$$\gamma = 1.2 + 0.42\,\mathrm{Log}_{10}\left(L_W/1000\right)$$

where L_W is nominal peak luminance of the display in cd/m^2.

According to the subjective tests conducted by the BBC, displays for a range of different values of nominal peak luminance, specifically the range from 400 cd/m^2 to 2 000 cd/m^2, can be shown to provide a consistent look by varying the value of gamma in the HLG OOTF in accordance with the equation above. This allows programmes to be made using displays with different peak luminance. Outside this range of peak luminance the match of this simple model to the experimental detail starts to deteriorate. An extended model, described in [18] and also illustrated in Fig. 23, is given by:

$$\gamma = 1.2 * \kappa^{\mathrm{Log}_2(L_W/1000)}$$

where:

$$\kappa = 1.111$$

This may be used for displays with peak luminance outside the range above. Within that range the two models are virtually identical and will provide equally good performance.

It should be noted that using a gamma adjustment to adapt to different peak luminances has its limitations. Television receivers typically apply different and more sophisticated methods. The acceptability of displays with different peak luminance values is a decision for individual producers, and might differ between productions.

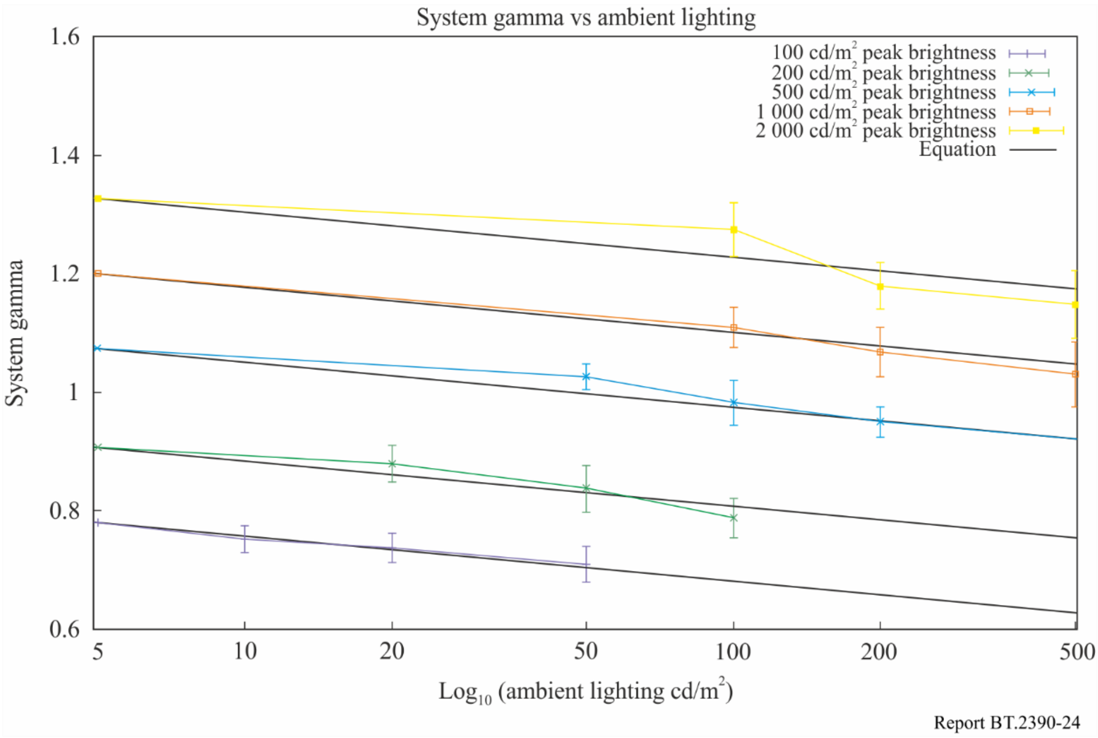
Many television programmes are produced in environments that differ considerably from the reference viewing environment. The luminance of the surround may be considerably higher than the recommended 5 cd/m^2.

Recommendation ITU-R BT.2100 recognises that the HLG display gamma may need to be reduced in brighter viewing environments, to compensate for the differences in the adaptation state of the eye.

The BBC conducted subjective tests to measure the change in gamma necessary to perceptually match images displayed across a range of peak luminances in the reference and in non-reference environments. Twenty-one viewers participated in the tests. The results, from 21 viewers, that show the reduction in gamma as the surround brightness increases are presented below in Fig. 24.

FIGURE 24

Graph of system gamma vs. ambient lighting for a number of different screen luminances, with lines of best fit



The line of best fit, which provides an indication of how gamma should be adjusted in non-reference environments, is given by the equation below:

$$\gamma_{bright} = \gamma_{ref} - 0.076 \log_{10}(\frac{L_{amb}}{5})$$

where:

γ_{bright} : system gamma for display surrounds greater than 5 cd/m^2

γ_{ref} : system gamma for reference environment, calculated according to Recommendation ITU-R BT.2100 Note 5e (and above)

L_{amb} : ambient luminance level in cd/m^2.

By adjusting the display gamma to compensate for non-reference viewing environments in this way more consistent results may be achieved in a wide range of production environments.

An alternative model is described in reference 18 which matches the form of the extended model for the variation of gamma with peak display luminance and which also includes the variation of gamma with surround luminance:

$$\gamma = \gamma_{\text{ref}} \cdot \kappa^{\text{Log}_2(L_W/L_{ref})} \cdot \mu^{\text{Log}_2(L_{surround}/L_{surround\,ref})}$$

where γ_{ref} is 1.2, $\mu = 0.98$ and the reference surround luminance $L_{reference}$ is 5 cd/m^2.

6.3 The hybrid log-gamma electro-optical transfer function (EOTF)

In order to specify the complete television system an EOTF is needed, as well as the OETF defined in § 6.1. The HLG EOTF maps the HLG signal representing the scene to the light emitted from the display.

The EOTF mapping should:

1) preserve the artistic intent of the programme maker (and provide a suitable rendering intent),
2) allow for the dynamic range of the display from black level to peak white, and
3) minimize quantization artefacts.

The EOTF defined in Table 5 of Recommendation ITU-R BT.2100 and described below is similar to the conventional display gamma curve, thereby maximizing backward compatibility, whilst also meeting the three preceding requirements.

$$\begin{aligned} F_D &= \mathrm{EOTF}\left[\max\left(0,(1-\beta)E'+\beta\right)\right] \\ &= \mathrm{OOTF}\left[\mathrm{OETF}^{-1}\left[\max\left(0,(1-\beta)E'+\beta\right)\right]\right] \end{aligned}$$

where:

F_D: luminance of a displayed linear component $\{R_D, G_D, \text{or } B_D\}$, in cd/m^2

E': non-linear signal $\{R', G', B'\}$ as defined for the OETF.

The inverse OETF, OETF^{-1}, is given by:

$$E = \mathrm{OETF}^{-1}[E'] = \begin{cases} E'^2/3 & 0 \le E' \le \tfrac{1}{2} \\ \{\exp\left((E'-c)/a\right)+b\}/12 & \tfrac{1}{2} < E' \le 1 \end{cases}$$

and β, the black level lift, is given by:

$$\beta = \sqrt{3(L_B/L_W)^{1/\gamma}}$$

L_W is nominal peak luminance of the display in cd/m^2 for achromatic pixels.

L_B is the display luminance for black in cd/m^2.

The black level lift, conventionally known as the "brightness" adjustment in CRT displays, adapts the EOTF to the minimum luminance that can be seen in the actual, not necessarily reference, viewing conditions. The appropriate value for β may be determined in any particular circumstance by using the PLUGE test signal specified in Recommendation ITU-R BT.814.

6.4 Compatibility with SDR displays

Both PQ and HLG provide limited compatibility when directly connected to legacy SDR displays with BT.709 colorimetry. In the absence of additional processing HLG has a degree of compatibility when shown on SDR UHDTV displays that have been designed to accept signals in the BT.2020 colour space.

Concerning the degree of compatibility achieved by HLG, hue changes can be perceptible on the SDR display should images contain bright areas of highly saturated colour or very high code values. Generally such high code values would be used for specular highlights and thus constitute a small proportion of the picture. The acceptability of the degree of compatibility of HLG might be a commercial decision by specific broadcasters or for a specific application.

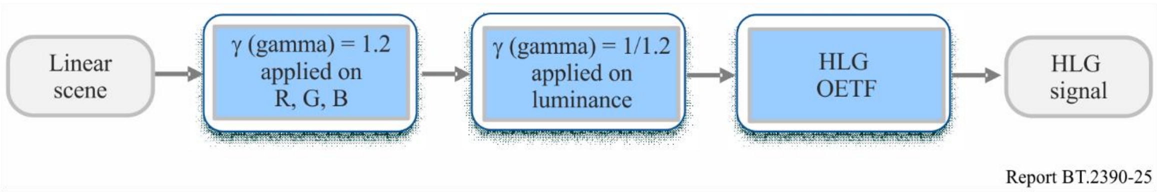
When PQ or HLG HDR signals are converted for use in SDR ITU-R BT.709 facilities, the conversion process is expected to perform the colour space, HDR to SDR and any video format conversion in such a way as to minimise perceptible changes in colour for all types of HDR content, regardless of the code value ranges in use.

6.5 Traditional colour reproduction for camera signals

The HLG OOTF (system gamma applied on luminance) uses scene-referred camera signals that result in a display that closely preserves the chromaticity of the scene as imaged by the camera. This differs from the traditional colour reproduction provided by the HDTV and UHDTV OOTFs, which produce more saturated colours which viewers of existing SDR content have become familiar with. Should such a traditional colour reproduction be desired, a gamma of 1.2 could be applied on the RGB components of a camera signal to produce more saturated colours. This approach is illustrated in Fig. 25.

FIGURE 25

Block diagram of signal chain to produce more saturated colours



In this Figure (linear) light from the camera is first processed by applying a gamma curve ($\gamma = 1.2$) independently to the red, green and blue colour components. Applying gamma separately to red, green and blue components does two things. Firstly, it adjusts the overall tone curve. Secondly, because it is applied separately to the colour components, the colour saturation is increased. The second processing block undoes the modification of the tone curve by applying an inverse gamma ($\gamma = 1/1.2$) to the luminance component of the signal. Applying gamma to the luminance component only (as in the HLG OOTF) leaves the ratio of the red to green to blue components unchanged and, hence, does not change the saturation.

Overall, the effect of applying such processing is to increase colour saturation whilst leaving the overall tone curve unchanged. Conversely, it would be possible to use similar processing to modify a signal representing the traditional look to instead more closely represent the chromaticity of the scene as imaged by the camera.

7 Conversion between PQ and HLG

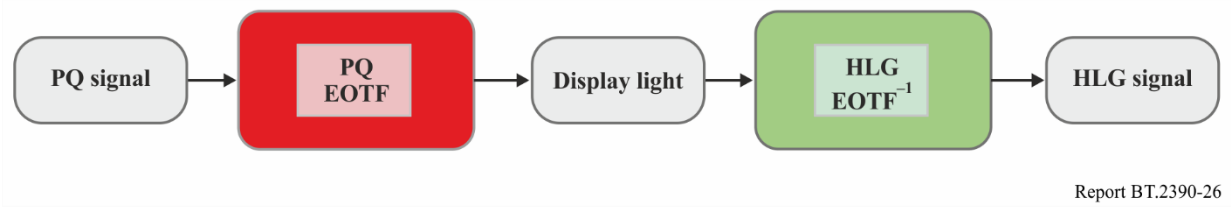
7.1 Transcoding Concepts

Transcoding aims to produce identical display light when the transcoded signal is reproduced on a display of the same peak luminance as the original signal. This section describes how a PQ signal may be transcoded to an HLG signal and vice versa, although cascaded conversions are to be discouraged to avoid risking loss of quality.

The following diagram illustrates the concept behind transcoding from the PQ signal to the HLG signal. The PQ signal is decoded by the PQ EOTF to yield a signal that represents linear display light. This signal is then encoded by the HLG inverse EOTF to produce an equivalent HLG signal. When this HLG signal is subsequently decoded by the HLG EOTF in the display, the result will be the same display light that would be produced by decoding the original PQ signal with the PQ EOTF. The HLG inverse EOTF is the HLG inverse OOTF followed by the HLG OETF.

FIGURE 26

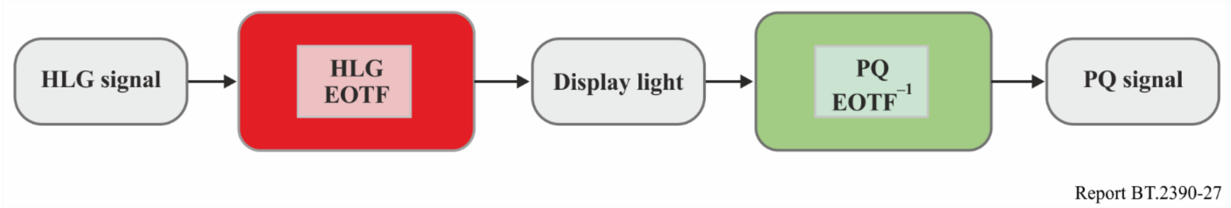
Concept of transcoding from PQ to HLG



The following diagram illustrates the concept behind the transcoding from the HLG signal to the PQ signal. The HLG signal is decoded by the HLG EOTF to yield a signal that represents linear display light. This signal is then encoded by the PQ inverse EOTF to produce an equivalent PQ signal. When this PQ signal is subsequently decoded by the PQ EOTF in the display, the result will be the same display light that would be produced by decoding the original HLG signal with the HLG EOTF.

FIGURE 27

Concept of transcoding from HLG to PQ



7.2 Conversion concepts using a reference condition at 1 000 cd/m^2

The transcoding concepts in the previous section produce the same displayed light for both PQ and HLG signals only when they are viewed on displays with the same peak luminance.

However, the difference in the way that PQ and HLG signals are rendered on displays of different peak luminance complicates the conversions between PQ and HLG signals. If, for example, PQ signals, representing different peak luminances, are simply transcoded to HLG, the signal level for diffuse white will vary. Similarly, when HLG content is transcoded to PQ the brightness of diffuse white will vary depending on the assumed peak luminance of the HLG display.

To avoid such brightness changes, it is needed to convert, rather than simply transcode, the signals. Consistent brightness in the converted signals may be achieved by choosing a reference peak displayed luminance (L_W) for the HLG signal, and requiring that PQ signal be limited to the same peak luminance. With these constraints consistent brightness is achieved in the converted signals.

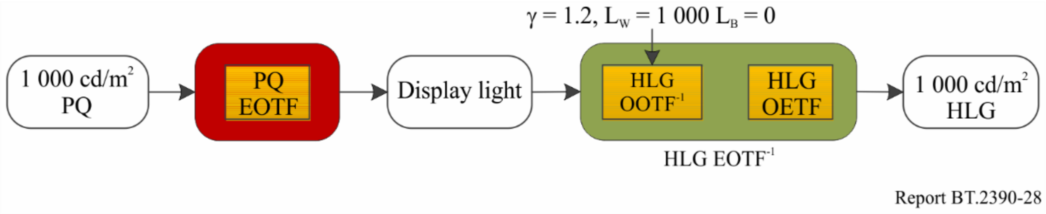
Therefore it is desirable that conversion between PQ and HLG should take place using the same reference peak displayed luminance for the signals used in the conversion. There is currently an industry consensus that this common peak luminance should be 1 000 cd/m^2.

For both transcoding and conversion a black level for the HLG EOTF also needs to be specified. The HLG black level, L_B, should be set to zero for transcoding and conversion.

With the choice of 1 000 cd/m^2 as the common peak luminance, the conversion outlined above is completely specified for any HLG signal to PQ and, for PQ signals not exceeding 1 000 cd/m^2, from PQ to HLG. Figure 28 illustrates the conversion from PQ to HLG.

FIGURE 28

Conversion from PQ to HLG at a common peak luminance of 1 000 cd/m²



The following is an elaboration of the corresponding Figure above in terms of the three most fundamental transformations:

(1) The PQ EOTF and its inverse

(2) The HLG OETF and its inverse

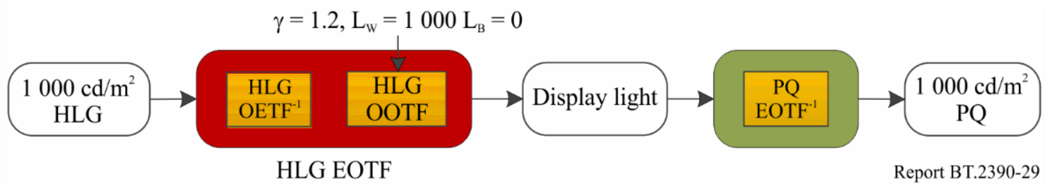
(3) The HLG OOTF and its inverse.

The HLG EOTF is derived from (2) and (3). The Figure also includes the parameters for HLG $OOTF^{-1}$. The resulting HLG signal will produce images identical to the original PQ images for all content that is within the colour volume of the 1 000 cd/m^2 HLG reference display.

Analogously, the conversion from HLG to PQ at 1 000 cd/m^2 is the inverse of the above as illustrated in Fig. 29.

FIGURE 29

Conversion from HLG to PQ at a common peak luminance of 1 000 cd/m²



This conversion always produces a PQ image identical to HLG.

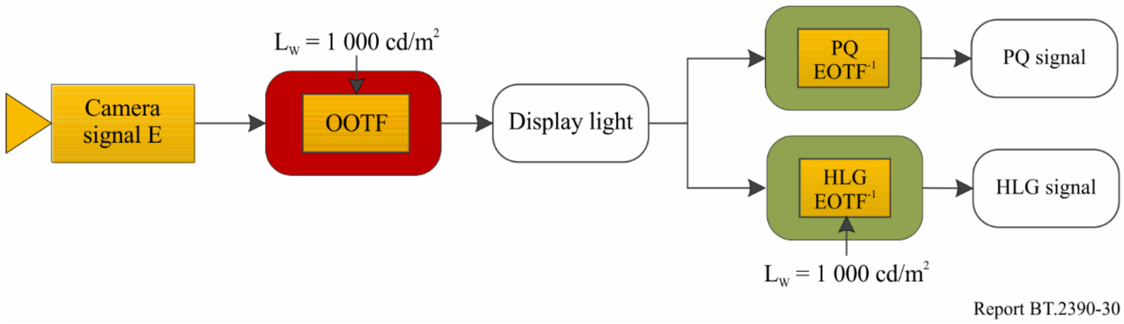
7.3 Cameras using a common OOTF at a reference peak luminance of 1 000 cd/m^2

Cameras could apply a common OOTF to produce PQ and HLG signals with identical displayed images at a reference peak luminance of Lw = 1 000 cd/m^2.

This OOTF could be the PQ OOTF, or the HLG OOTF, and might include additional modifications applied in the camera, as illustrated in Fig. 30. PQ and HLG signals are obtained using their respective inverse EOTFs.

FIGURE 30

Use of a common OOTF to provide both PQ and HLG at a common peak luminance of 1 000 cd/m²



The appearance of the displayed images will be the same on displays with a peak luminance capability of 1 000 cd/m^2, for both the PQ and HLG signals. The appearance of the image is determined by the OOTF.

7.4 Handling PQ signals with greater than 1 000 cd/m^2 peak luminance

PQ signals can represent a peak luminance of up to 10 000 cd/m^2. In order to enable the reference conversion described above, PQ content must be limited to have a peak luminance that does not exceed 1 000 cd/m^2. There are, in general, three approaches to achieving this:

(1) Clip to 1 000 cd/m^2

(2) Static mapping to 1 000 cd/m^2 (e.g. using an EETF curve like those described in § 5)

(3) Dynamic mapping to 1 000 cd/m^2

The first method, clipping to 1 000 cd/m^2, is simple to implement. While multiple round trip conversions between PQ and HLG are to be discouraged, with this method content undergoes no additional limiting/clipping in the event of multiple round-trip conversions (i.e. PQ->HLG->PQ->HLG) beyond the initial clipping.

The second method, static mapping to 1 000 cd/m^2 can be implemented by a LUT containing an EETF such as that described in § 5.4.1. While this avoids hard clipping of detail in the highlights, it is not invariant under blind multiple round-trip conversions.

The third method, dynamic mapping to 1 000 cd/m^2, utilizes adaptive processing, for example on a frame-by-frame, or scene-by-scene basis. An adaptive algorithm could vary the EETF described in § 5.4.1 based on statistics of the image content (scene maximum for example). For non-live content, dynamic mappings could be generated offline by the content producer (either manually or using algorithmic processing). Except for the initial stage of limiting the PQ signal to 1 000 cd/m^2, this approach could survive multiple round-trip conversions, because subsequent dynamic processing should be inactive given that the signal would already have been limited to 1 000 cd/m^2.

7.5 Possible colour differences when converting from PQ to HLG

In principle, the conversion of PQ images to HLG could give rise to hue shifts or desaturation on bright highly saturated areas of the picture, although such effects are believed to be rare in practice.

Mathematically, this arises because the OOTF applied in the display for HLG is a function of overall luminance rather than identical functions of R, G, and B. Consider the equations for luminance in both the display and scene domains along with the EOTF for HLG:

$$Y_D = 0.2627R_D + 0.6780G_D + 0.0593B_D$$

$$Y_S = 0.2627R_S + 0.6780G_S + 0.0593B_S$$

$$R_D = \alpha Y_S^{\gamma-1} R_S$$
$$G_D = \alpha Y_S^{\gamma-1} G_S$$
$$B_D = \alpha Y_S^{\gamma-1} B_S$$

The Table below summarizes the peak values that can be displayed for pure white, and for the red, green and blue primaries, for a 1 000 cd/m^2 PQ monitor, and for a 1 000 cd/m^2 HLG monitor. The value 'x' is the signal value required such that when $R = G = B = x$ the resulting white is 1 000 cd/m^2. For PQ, this occurs when x is approximately 0.76; for a 1 000 cd/m^2 HLG display, this occurs when $x = 1.0$. For a 1 000 cd/m^2 PQ display, the maximum luminance of each of these colours is calculated using Y_D and is shown in the middle column of the Table. For HLG, the EOTF can be simplified by normalizing scene colours within [0,1]. Thus:

$$R_D = 1000 Y_S^{\gamma-1} R_S\text{, etc.}$$

This determines $\{R_D, G_D, B_D\}$ and the resulting luminance is calculated using Y_D . The peak luminance achievable with HLG is tabulated in the rightmost column.

Colour	BT.2100 PQ Y cd/m^2	BT.2100 HLG Y cd/m^2
{x,x,x} // Peak white	1000.0	1000.0
{x,0,0} // Peak red	262.7	201.1
{0,x,0} // Peak green	678.0	627.3
{0,0,x} // Peak blue	59.3	33.7

In summary, PQ signals that have had peak luminance limited to 1 000 cd/m^2 could potentially contain bright saturated colours that cannot be displayed identically by a 1 000 cd/m^2 HLG monitor. As only scene highlights are very bright, and highlights are generally not highly saturated colours, such signals are rare. Nevertheless they can occur and need to be considered. As described in § 7.4, such signals may be clipped (default), static mapped using a LUT (i.e. soft clipped), or dynamically limited using a dynamic colour processor.

8 Colour representation for chroma sub-sampling

The legacy $Y'C'_BC'_R$ non-constant luminance format is a colour-opponent based encoding scheme (in which signals are interpreted based on colour differences in an opposing manner) intended to separate luma from chroma information for the purposes of chroma subsampling (i.e. 4:2:2 and 4:2:0). High dynamic range and wide colour gamut content reveal the limitations of existing colour encoding methods. Errors that were previously small with standard dynamic range can become magnified. Recommendation ITU-R BT.2020 provides an alternative to $Y'C'_BC'_R$, i.e. the $Y'_CC'_{BC}C'_{RC}$ constant luminance format. This format resolves the issue of chroma leakage into the Y' luma signal, but does not solve the problem of luminance contamination of the C'_{BC} and C'_{RC} components. Recommendation ITU-R BT.2100 provides an alternative method for colour difference encoding called constant intensity, which is based on IPT colour space [17] developed by Ebner and Fairchild.

8.1 Non-constant luminance (NCL) $Y'C'_BC'_R$

$Y'C'_BC'_R$ is widely used for standard dynamic range content and requires a specific conversion based on the primaries being encoded and decoded. Recommendation ITU-R BT.2100 specifies PQ as a non-linearity to be used with the BT.2020 colour primaries. While $Y'C'_BC'_R$ performs satisfactorily in many cases, some limitations have emerged for its use in high dynamic range wide colour gamut scenarios.

Limitations of $Y'C'_BC'_R$ with wide colour gamut and high dynamic range

– Quantization distortions due to bit depth limitations with the increased colour volume.

– Chroma subsampling distortions due to a perceptually uneven distribution of code words.

– Colour volume mapping distortions due to incorrectly predicted hue and luminance.

– Error propagation from chroma to luma channels.

The constant luminance method specified in Recommendation ITU-R BT.2020 helps reduce the last of these, but this solution is not being widely adopted because the benefits are considered modest and entail some additional complexity.

8.2 Constant intensity IC_TC_P encoding

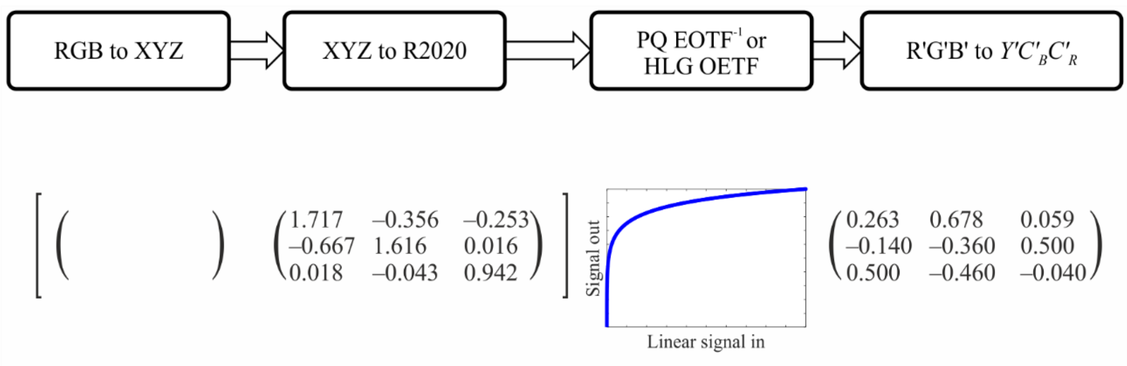
An alternative to constant luminance (CL) $Y'_CC'_{BC}C'_{RC}$ is the constant intensity (CI) IC_TC_P colour representation. Like $Y'C'_BC'_R$, IC_TC_P is a colour-opponent based encoding scheme intended to separate luma from chroma information. CI offers the same benefit as CL in that the chroma channels are lacking luminance, but IC_TC_P has the advantage that the lines of constant hue are straighter, and the MacAdam's ellipses are more circular. The CI neutral (grey) axis is encoded with the PQ or HLG non-linearity to match the human visual system, and to optimize it for high dynamic range signal encoding. The alternative 3x3 colour matrices used to generate the colour difference channels have been optimized [18] for the human visual system perception of HDR and WCG. The in-camera encoding and in-display decoding steps for IC_TC_P are identical to those for NCL $Y'C'_BC'_R$, so IC_TC_P is compatible with that hardware.

8.2.1 Constant intensity IC_TC_P encoding

Below are the conversion steps needed to get from camera linear RGB sensor signals into $Y'C'_BC'_R$ and into IC_TC_P [19]. Note that the matrix coefficients are decimal values that differ very slightly from the values shown in Recommendation ITU-R BT.2100; the values shown in the Recommendation should be used in actual implementations.

FIGURE 31

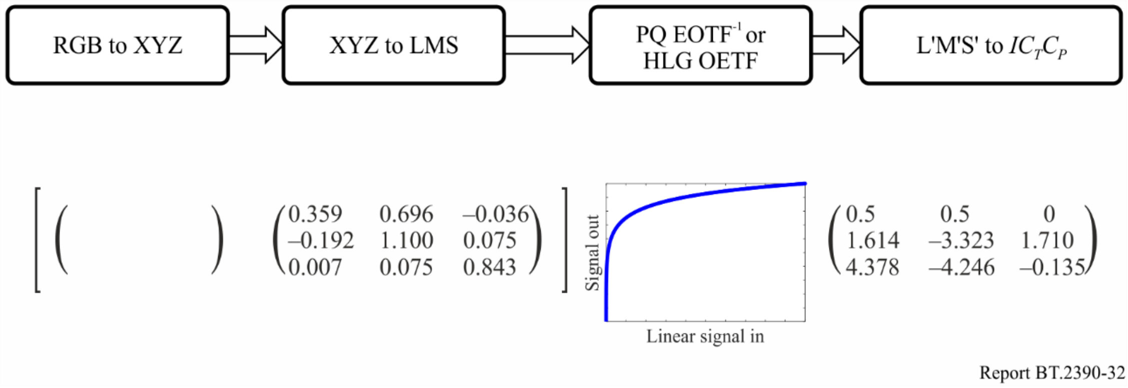
Camera RGB conversion To $Y'C'_BC'_R$



Report BT.2390-31

FIGURE 32

Camera RGB conversion to IC_TC_P



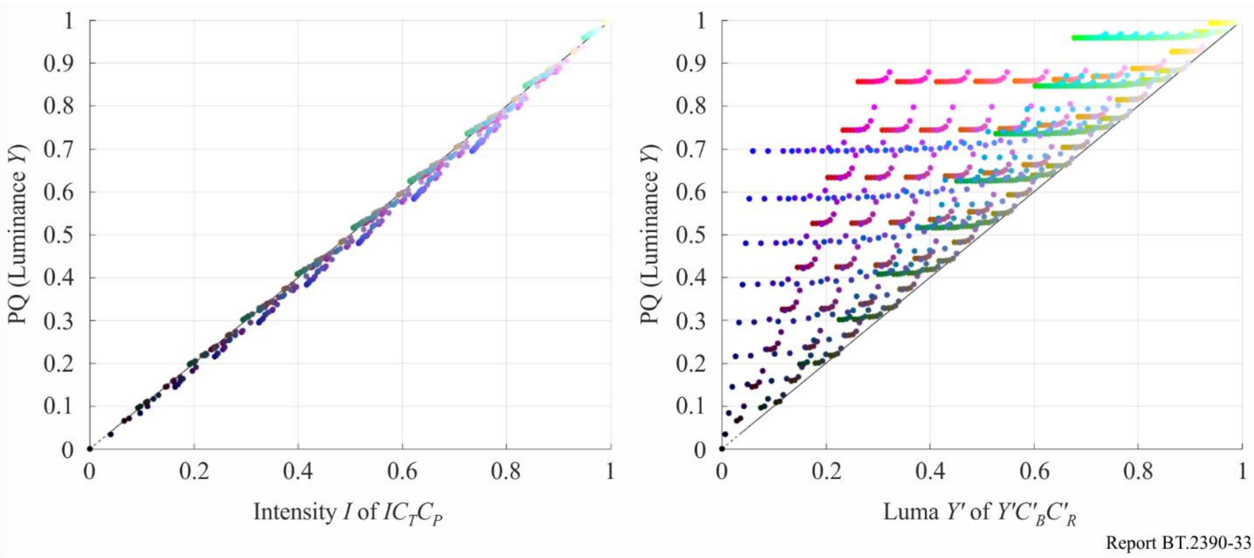
8.2.2 Advantages of constant intensity IC_TC_P

The specific design of the constant intensity colour space provides several benefits versus the Non-Constant Luminance colour space when used with the PQ or HLG non-linearity to provide HDR.

Achromatic channel: The achromatic axis of $Y'C'_BC'_R$ (Y' encoded in PQ or HLG) does not fully decorrelate luminance from colour. Therefore distortions introduced into the chroma channels can propagate to luminance where they become much more noticeable. As shown in Fig. 33, the achromatic axis of IC_TC_P (I) corresponds very closely with luminance (where luminance is a weighted sum of linear R,G,B). This is an indicator of how well IC_TC_P separates luma from chroma information. This reduces errors that can be introduced when spatially sub-sampling the chroma components compared to conventional non-constant luminance encoding. The axes in Fig. 31 are from zero to full scale in PQ space. (The luminance errors shown for $Y'C'_BC'_R$ are not as large for legacy systems using standard dynamic range with gamma encoding.)

FIGURE 33

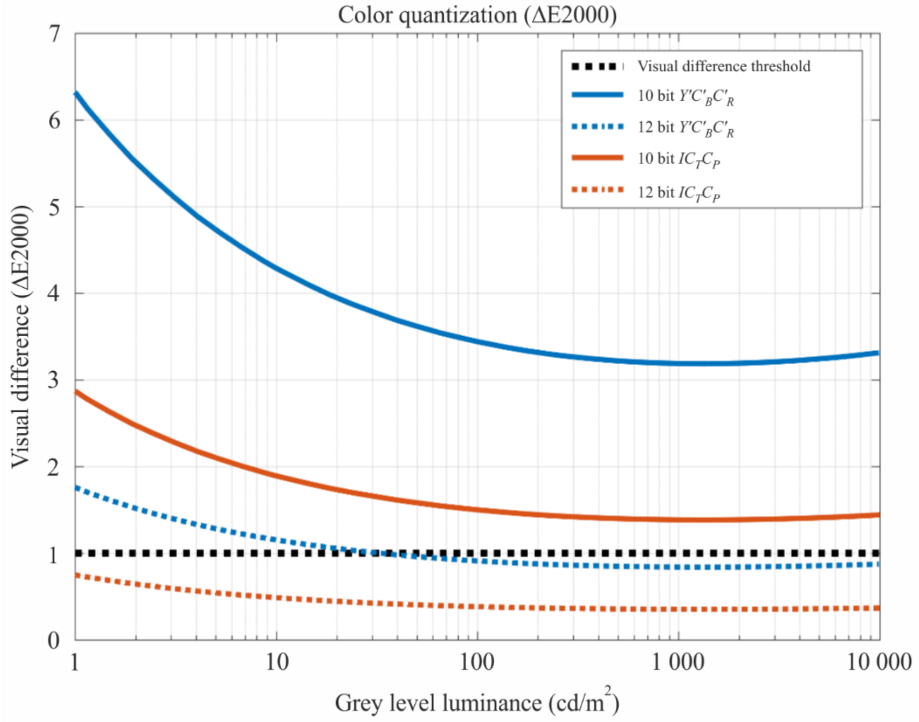
Luminance correlation



Quantization to limited bit-depth: Figure 34 shows the worst case visual colour difference between chroma channel code values (using ΔE2000) at various luminance levels. 10-bit IC_TC_P provides an approximately 1.5 bit colour difference improvement over 10-bit $Y'C'_BC'_R$. At less than an average of 1.0 ΔE above the visual difference threshold, use of IC_TC_P significantly decreases visible distortions thus enabling excellent colour performance with 10-bit encoding.

FIGURE 34

Maximum colour deviation at various bit-depths

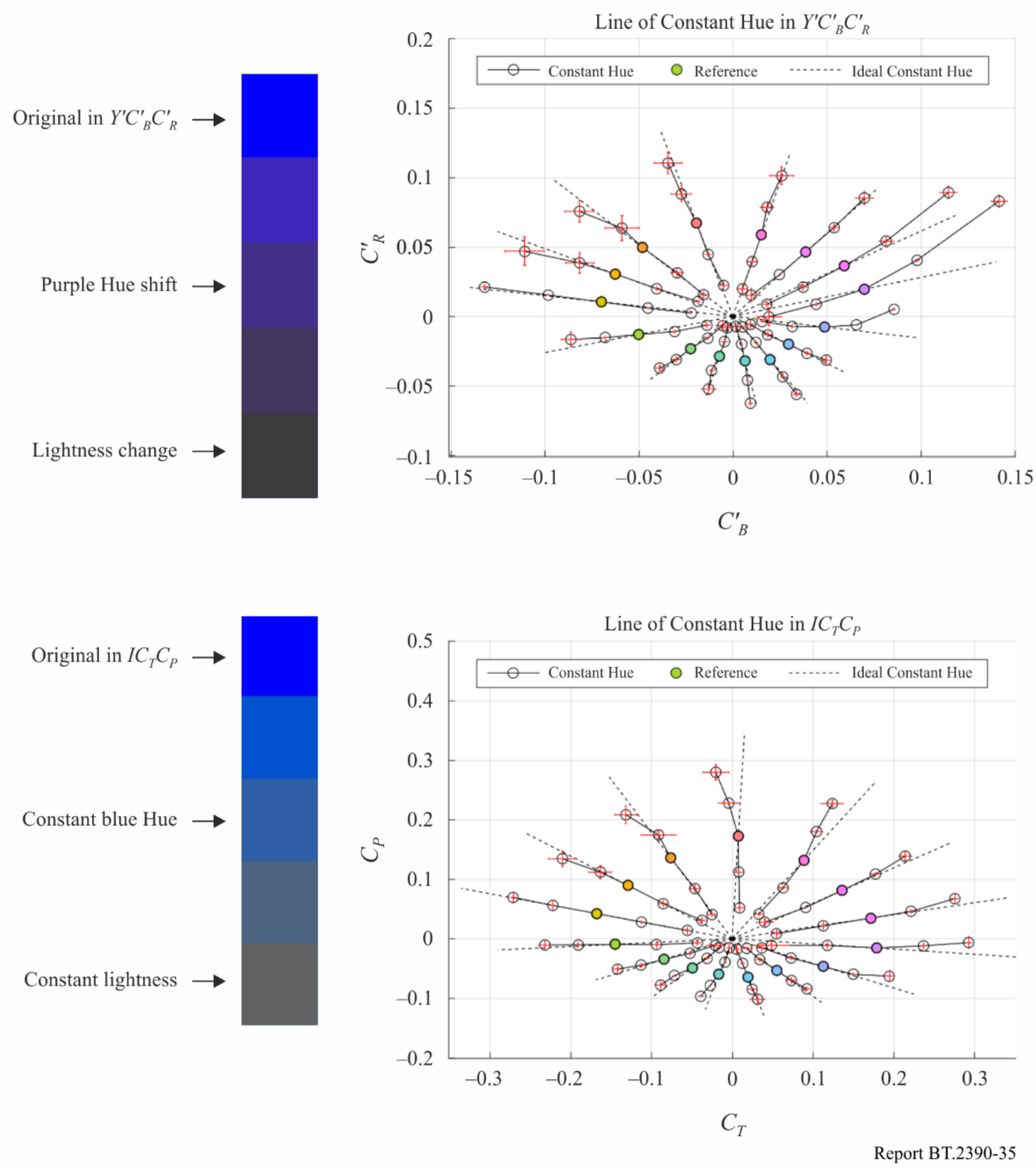


Report BT.2390-34

Uniformity and hue linearity: A colour space is hue linear when the hue remains constant as saturation or intensity are changed. Hue linearity is important during any interpolation such as colour volume mapping, chroma subsampling, and blending/fading. $Y'C'_BC'_R$ has large deviations (see Fig. 35) that cause hue shifts with highly saturated colours. IC_TC_P was designed to minimize deviation from lines of constant hue thereby reducing hue shifts. In addition, IC_TC_P has a more uniform distribution of colours. This improves efficiency, reduces worst case quantization and interpolation errors.

FIGURE 35

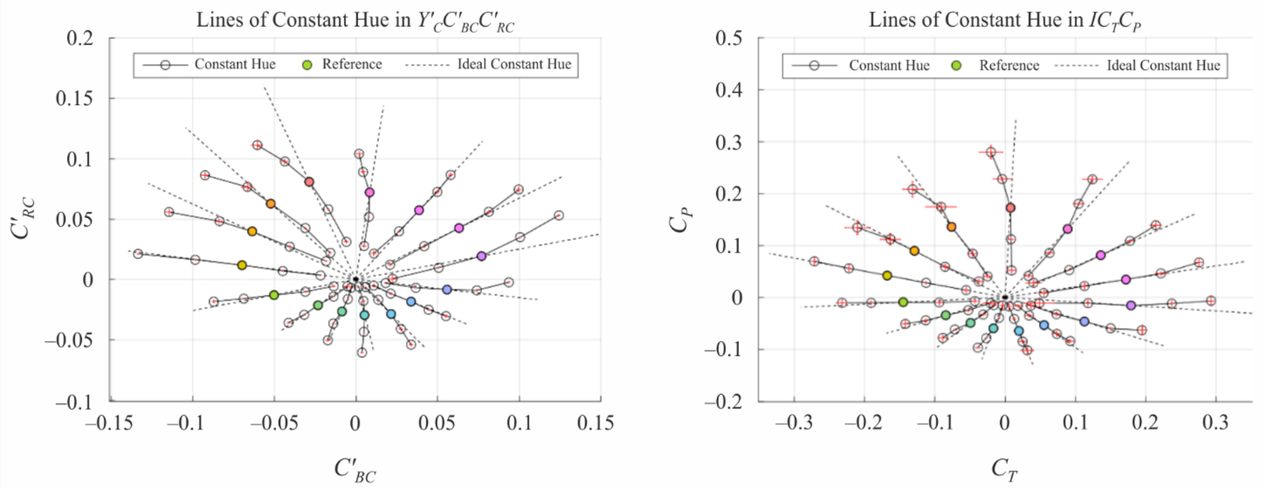
Blue Hue Linearity Comparison (using PQ)



If the CL format specified in Recommendation ITU-R BT.2020 is applied to HDR, the $Y'_C C'_{BC} C'_{RC}$ representation introduces additional (over NCL) errors in skin tones. The blue is significantly improved versus NCL (but still contains errors) and CL has significantly worse errors in the red and green regions (see Fig. 36). (The BT.2020 CL coefficients were designed for use with the SDR camera characteristic, and thus were not optimized for use in HDR.)

FIGURE 36

Constant luminance vs. constant intensity hue linearity comparison (PQ)

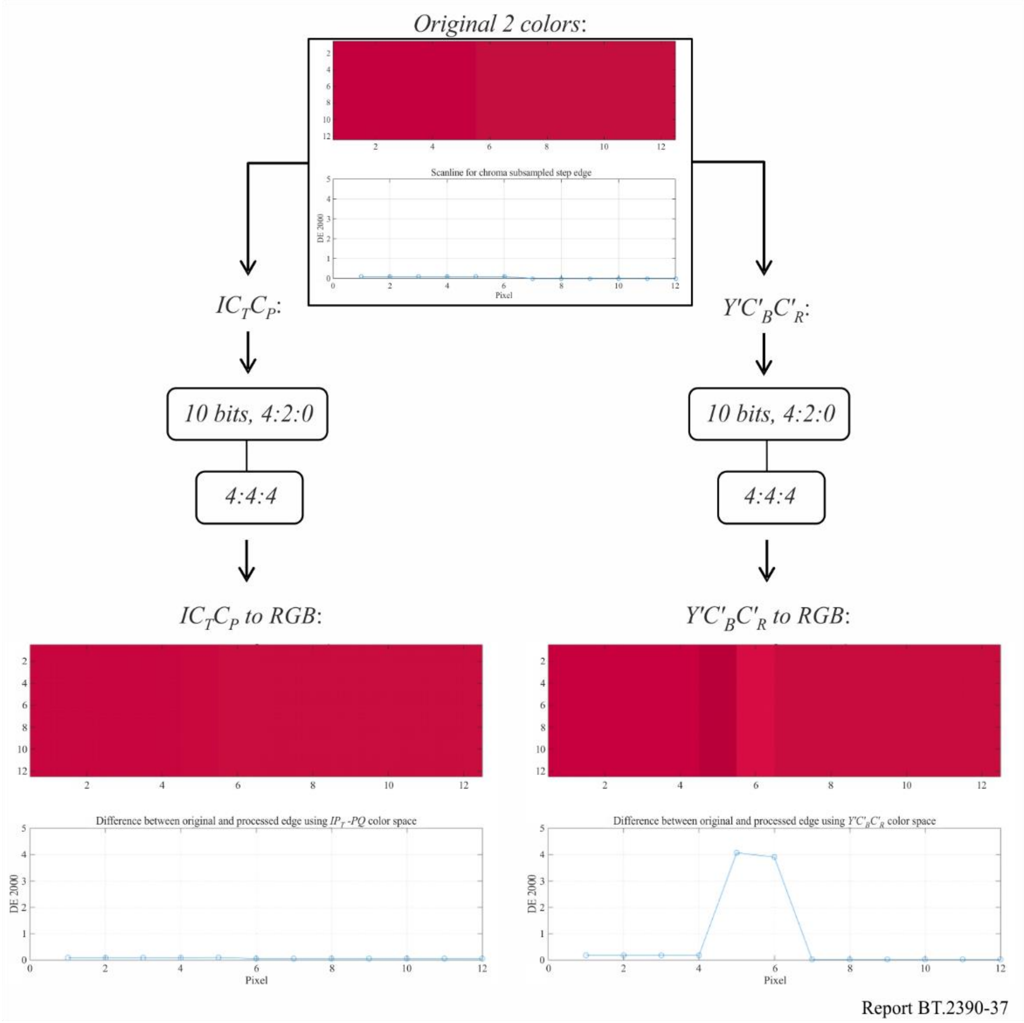


Report BT.2390-36

Colour sub-sampling: Figure 37 shows a practical example of a colour sub-sampling distortion due to NCL encoding. Two very similar colours with a Δ*E*2000 of 0.1 were sub-sampled to 10 bits 4:2:0 in $Y'C'_BC'_R$ and IC_TC_P and reconstructed. Due to the poor decorrelation between Y' of $Y'C'_BC'_R$ and luminance (Y), errors introduced in chroma during sub-sampling spread to the luminance and became more visible with a Δ*E*2000 of 4.0. Constant intensity IC_TC_P has a higher tolerance for chroma error and the colours remain indistinguishable with a Δ*E*2000 of 0.2.

FIGURE 37

Colour sub-sampling distortions based on correlation with luminance



9 Some considerations on the use of high dynamic range in TV image capture, mastering, distribution and presentation

This section focuses on the operational issues introduced by HDR imagery, offering a number of operational considerations on the desirable amount of image dynamic range, in relation to the processing that may be required along the various stages that television images go through, from image capture to production, postproduction, mastering, versioning, distribution and presentation of television programmes to the public at large.

The introduction of HDR imagery poses a number of housekeeping challenges, associated with the increased number of picture formats that will be in use.

9.1 Television image capture, production, postproduction and mastering

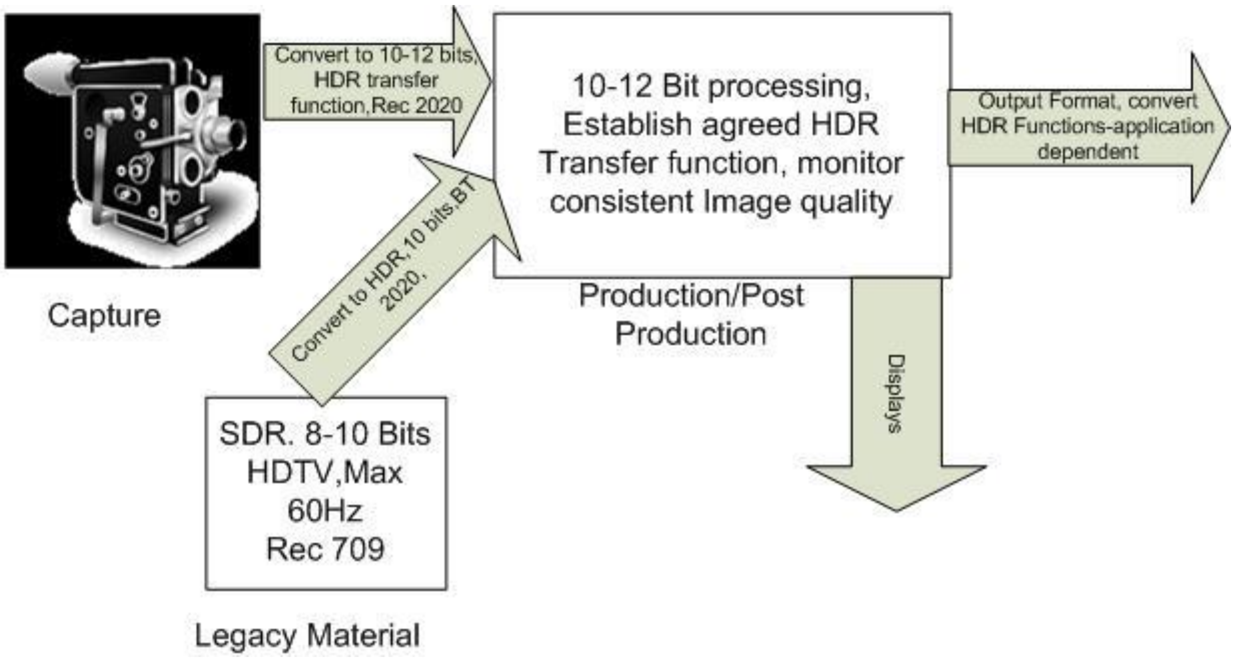
During image capture and production, depending on the envisaged subsequent image postproduction, it may be desirable to capture television programme images at HDR even if they are not intended for distribution as HDR images. Since image capture and production at HDR provides an extended image postproduction headroom, programme directors may wish to exploit that headroom in order to achieve their creative intent.

As a generalization, there are two forms of image capture; live events, and content that will or could be subject to further signal processing.

Live capture typically uses image parameters that will be retained during the entire production process.

Content captured for subsequent post-processing and for multiple distribution channels is increasingly using image parameters that are not defined by standards organizations; in particular the image pixel depth, typically 16 bits, and the image transfer function(s) are left to a manufacturer's choice which in many cases is proprietary.

FIGURE 38



The above illustration is intended to show possible conversion points in the production process. The output from a capture camera for non-live transmission would typically be about 16 bits raw data (and could also be in some proprietary HDR format). In non-real-time file processing, the 16 bits could well be edited. However, if legacy material is to be added to the production, some conversion processes will need to be performed to create either the final HDR output or some intermediate image format. Suffice to say that interface limitations will need to be addressed if serial digital interfaces (SDI) interfaces are used.

The image dynamic range of the programme at the end of the postproduction process will normally be the one required for programme mastering and distribution. However, at this time it is not clear which various options will be chosen by programme producers. Clearly, consideration will need to be given to available emission bandwidth and to the environment in which end users will watch television.

Programmes most likely will be mastered at the image dynamic range required to meet the needs of the most demanding media targeted for their distribution.

Historically, during the production process image quality is monitored and checked for consistency throughout a given production; the introduction of HDR images complicates this process. The display devices used in production will most likely be watched under different viewing conditions to those found in the home or elsewhere; the peak brightness and other display settings will be different. In the absence of some strict guidelines and recommended operating practices, the intended creative intent of HDR programmes may fail to be adequately displayed to end users.

9.2 Television programme versioning

When a programme master has been completed, it is often necessary to generate various versions of it, depending on the requirements of the targeted distribution markets and media, each of which may have its own specific needs, e.g. in terms of language, subtitles, moral codes, etc.

Programme versioning also potentially includes the generation of versions at different image standards, e.g. at different image dynamic ranges or resolutions, depending on the targeted markets.

9.3 Television programme presentation

In broad terms, the mission of broadcasting is to provide information, education and entertainment to the public at large. The entertainment mission implies that programme presentation should be pleasurable: programmes should be presented to the public, conveying the creative intent of the programme director in an attractive way, to maximize the quality of viewers' experience, e.g. without causing viewers' annoyance or discomfort.

In this respect, an extended image dynamic range may play a role in improving viewers' satisfaction on condition that the programme director properly uses it and the end user watches it under appropriate viewing conditions, yet to be defined and recommended.

The statement above begs a question: how much image dynamic range is needed to optimize viewers' satisfaction? Obviously, too little would make little difference to viewers, and too much could well cause viewer's discomfort. It also begs the question of what should be done with inserting legacy material into an HDR production?

The limit to image dynamic range in the home depends mainly on two boundaries, namely

– the peak image brightness that the consumer display can provide without compromising other aspects of its performance such as colour fidelity, display life and resolution, which depend on the design of each display model; and
– the minimum image brightness at which consumers can still discern details on the screen of their display; this depends on the design of each display model and on the consumers' prevailing viewing environment, namely their "home" environment.

It also depends on other considerations such as pixel depth, available bandwidth, and real-time or non-real-time transfers of data etc., which are not directly related to home viewing.

It is important that a reference viewing environment has been defined in Recommendation ITU-R BT.2100; this requirement has not been part of previous image Recommendations.

How bright should peak brightness be, for how long should peak brightness appear on the screen, how much of the screen area should be allowed for peak brightness? Concerns have been expressed that there might be a risk of a video version of "loudness" until such time as a Recommendation on Operational Practices for HDR-TV is in place.

9.4 The typical home viewing environment

Concerning the consumer's television viewing environment, Recommendation ITU-R BT.2022 specifies "general viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays". Section 1.2 of its Annex 1 specifies general viewing conditions in a home environment clarifying that, in its context, "the home viewing environment is intended to provide a means to evaluate quality at the consumer side of the TV chain. General viewing conditions in § 1.2 reproduce a home environment. These parameters have been selected to define an environment slightly more critical than the typical home viewing situations".

In view of the recent date of approval of Recommendation ITU-R BT.2022 (08/2012), the viewing environment specified therein can be assumed to still closely represent a typical home viewing environment, although it could be that some extensions may be required for consumers' HDR viewing in the home.

Recommendation ITU-R BT.2022 specifies the home viewing environment as being characterized by a value of 200 lux of incident light falling on the screen from the environment (e.g. from the windows during the day or the room lamps at night, as well as the light from the TV set itself, diffused back to it by the room walls and furniture). Of course, with an illumination of 200 lux on the TV screen from the environment, details in deep dark areas of the screen will appear to be somewhat washed out and the image dynamic range, as perceived by home viewers, will appear to be reduced with respect to its appearance when viewed in a dark environment.

Based on available indications, the reflectance of the screen of a modern consumer television set can be expected to be of the order of about 2% of the incident light falling on it, when the set is inactive and it uses a "black" screen with a semi-matte surface. This means that, in the consumer viewing environment specified in Recommendation ITU-R BT.2022, the luminance of the blackest image black would be of the order of 4 cd/m^2.

10 Mapping of SDR content into HLG and PQ

Standard dynamic range (SDR) content may either be directly mapped or inverse tone mapped (or 'up-converted') into an HDR format for inclusion in HDR programmes.

Methods for mapping place SDR content into an HDR container, analogously to how content specified using BT.709 colorimetry may be placed in a BT.2020 container. This approach is intended to preserve the 'look' of the SDR content when shown on an HDR display.

In contrast, inverse tone mapping ('up-conversion') is intended to expand luminance values to use the available luminance range, and thereby leverage the display capabilities to emulate the apprarance of HDR content. This section describes methods for mapping.

There are two possible approaches to SDR mapping, depending on the application:

– Display-referred mapping is used when the goal is to preserve the colours and relative tones seen on an SDR BT.709 or BT.2020 display, when the content is shown on a BT.2100 HDR display. An example of which is the inclusion of SDR graded content within an HDR programme.

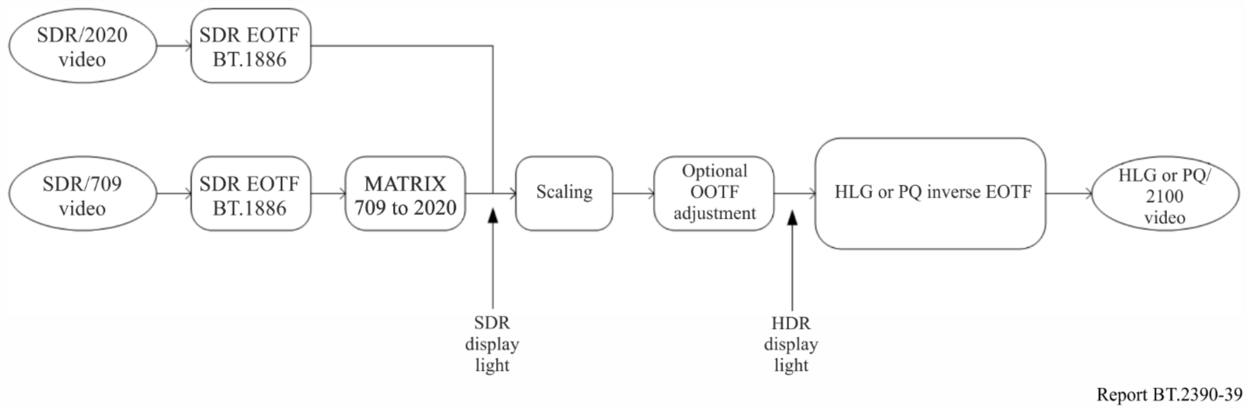
– Scene-referred mapping is used where the source is a direct SDR camera output and the goal is to match the colours and tones of a BT.2100 HDR camera. An example of which is the inter-mixing of SDR and HDR cameras within a live television production.

10.1 Display referred mapping

Figure 39 illustrates the display-referred mapping of SDR signals into either HLG or PQ.

FIGURE 39

Generic method for "display-referred" mapping of SDR into HLG or PQ



The SDR signal is first passed through the BT.1886 reference EOTF to derive SDR linear display light. An approximation of the electro-optical transfer function (EOTF) from Recommendation ITU-R BT.1886 may be used:

$$E = (E')^{2.40} \qquad , \; 0 \leq E' \quad \leq 1$$

where:

- E' is the non-linear signal (R', G', B') in the range [0:1]
- E is the normalised linear display light in the range [0:1].

A colour space conversion from BT.709 primaries to BT.2020/BT.2100 colour primaries is performed if necessary, details of which can be found in Recommendation ITU-R BT.2087.

The linear SDR display light may then be scaled to ensure that SDR and native HDR content have a similar level for HDR reference white. Where scaling is performed, a small optional adjustment to the OOTF may then be applied to compensate for the subjective change in appearance of the SDR signal arising from a simple linear scaling; thereby ensuring that the visibility of detail in the shadows is maintained and that the level of skin tones in HDR and mapped SDR content are similar.

Having scaled and adjusted the SDR display light, the resulting signal is passed through an HLG or PQ inverse EOTF to provide either an HLG or PQ signal.

10.1.1 Display referred mapping of SDR into PQ

The following procedure may be followed to achieve consistent mid-tone luminance levels when mapping standard dynamic range content into PQ.

Standard dynamic range BT.2020 content should be mapped to PQ by applying the ITU-R BT.1886 display EOTF and then applying the PQ EOTF^{-1}.

$$E' = {EOTF_{PQ}}^{-1}\left[scaling \;\times\; EOTF_{BT.1886}[V, L_W, L_B]\right]$$

- V: Input SDR video signal level (normalized, black at $V = 0$, to white at $V = 1$)
- L_W: SDR screen luminance for white = 100 cd/m^2
- L_B: Screen luminance for black = 0 cd/m^2
- E': Output PQ video signal level (normalized [0:1])
- *Scaling*: $EOTF_{PQ}\,(E'_{V=1})$ / 100 cd/m^2

Example: for scaling = 2.0, $E'_{V=1} = 0.58$ and $EOTF_{PQ}\,(E'_{V=1})$ = 200 cd/m^2

For unity mapping the peak signal of standard dynamic range content would be set to 100 cd/m^2 or 51% PQ.

Unity mapping does not change the display of the SDR content (it will display on the PQ HDR reference monitor the same as it displayed on the reference SDR monitor). Thus, no OOTF adjustment of the SDR display light signal is necessary.

If the SDR content is being inserted into HDR programming, and there is desire to more closely match the brightness of the HDR content, and that brightness is known, scaling can be done to bring up the brightness of the mapped SDR content. Scaling should be performed with care lest scaled SDR content, in particular skin tones, becomes brighter than in the HDR content.

A scaling factor of 2.0 is consistent with the HDR level guidance provided in Report ITU-R BT.2408, as that will map the 100 cd/m^2 peak white level of SDR to the 200 cd/m^2 level suggested for HDR or 58% PQ. Also noteworthy is that MovieLabs has recommended a scaling factor of 2.0 when converting for consumer displays, as MovieLabs has found this to provide a good match to the way such displays show SDR content in their “home cinema” viewing modes [19].

For standard dynamic range BT.709 content the same process may be used, with the BT.709 to BT.2020 conversion matrix applied before the scaling as shown in Fig. 39.

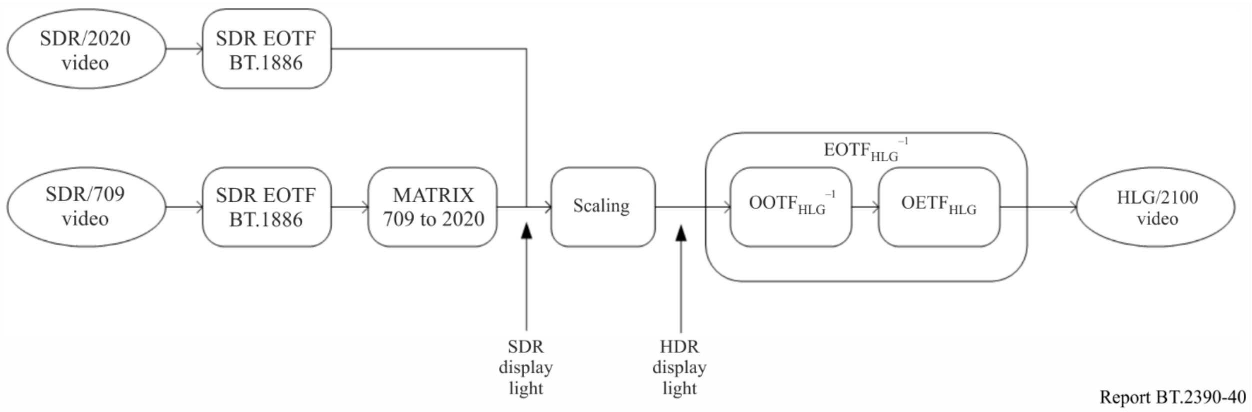
10.1.2 Display referred mapping of SDR into HLG

10.1.2.1 Mapping without gamma adjustment

The ‘display-referred’ method of mapping SDR content in to a Hybrid Log-Gamma (HLG) container is illustrated below in Fig. 40.

FIGURE 40

SDR to HLG mapping without gamma adjustment (display-referred)

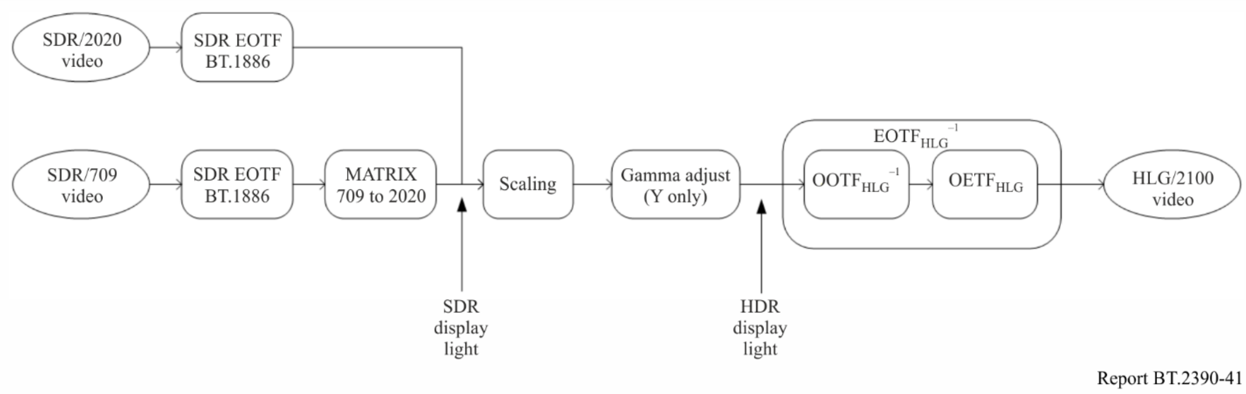


10.1.2.2 Mapping with gamma adjustment

For the case when gamma adjustment is made to the scaled SDR display light, the process is shown in Fig. 41.

FIGURE 41

Model for "display-referred" mapping with gamma adjustment of SDR into HLG



The linear SDR display light is scaled to ensure that 100% of the SDR signal is mapped to the HLG reference level 75% HLG. A small gamma adjustment may then optionally be applied to the luminance component, to compensate for the subjective change in appearance of the SDR signal arising from a simple linear scaling of the SDR display light signal.

Having scaled and adjusted the SDR display light, the resulting signal is passed through an HLG inverse EOTF to provide the HLG signal.

10.1.2.3 Scaling

When (100X)% SDR signal is mapped to (100Y)% HLG signal, a scaling gain is calculated by the following equation:

$$\text{Gain} = \frac{\text{EOTF}_{\text{HLG}}(\text{Y})}{\text{EOTF}_{\text{SDR}}(\text{X})}$$

For example, when 100% SDR signal is mapped to 75% HLG (203 cd/m^2 on a 1 000 cd/m^2 display), the scaling gain is calculated as follows:

$$\text{Gain} = \frac{\text{EOTF}_{\text{HLG}}(0.75)}{\text{EOTF}_{\text{SDR}}(1.0)} = \frac{\text{OOTF}_{\text{HLG}}\left(\text{OETF}_{\text{HLG}}^{-1}(0.75)\right)}{\text{EOTF}_{\text{SDR}}(1.0)} = \frac{0.265^{1.2}}{1.0^{2.4}} = 0.203$$

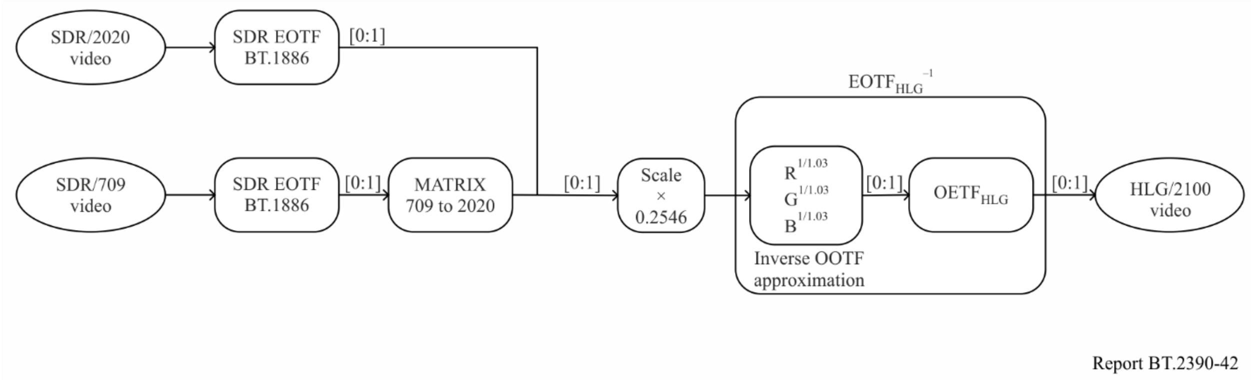
10.1.2.4 Simplification of the HLG mapping process

Through careful choice of the HLG inverse EOTF parameters, it is possible to avoid the need to scale and adjust the gamma of the SDR linear display light signal. By configuring the HLG inverse EOTF with a nominal peak luminance, L_W, of 392 cd/m^2, an input of 100 cd/m^2 from the SDR EOTF will directly deliver an HLG signal of 75%, satisfying the requirement to map 100% SDR signal to 75% HLG signal, without further scaling and gamma adjustment.

Figure 42 illustrates how, for all but the most critical applications, it is possible to simplify the conversion yet further. When applying the HLG inverse EOTF with L_W set to 392 cd/m^2, Note 5e of BT.2100 requires a gamma value of 1.03. As this is close to unity, in most applications there is no need to apply the inverse OOTF gamma to the luminance component, it can instead be applied independently to R, G and B components; greatly simplifying the mapping process. Colour distortions that usually arise through applying gamma to red, green and blue, rather than luminance, are barely visible for such low values of gamma.

FIGURE 42

Simplified (display-referred) SDR to mapping into HLG



As normalised signals are used throughout, a different scaling is required to match the signal ranges of the SDR EOTF and HDR inverse EOTF, thereby ensuring that 100% SDR signal maps to 75% of the HLG HDR signal. Note that as the normalised signals are dimensionless, the scaler is not adjusting the peak luminance of the SDR display light, so no additional gamma compensation for the signal scaling is required. Allowing for the inverse OOTF gamma of 1.03, the correct scale factor is 0.2546.

10.2 Scene referred mapping

It is particularly important that the scene-referred mapping is used for matching signals from BT.709 and BT.2020 SDR cameras with signals from HLG cameras. This is because, direct from the camera (and prior to subjective adjustment), both signals represent light from the scene captured by the camera.

If the display-referred mapping were used, which maintains the appearance of SDR images on an HLG display, the signals from SDR cameras and HLG cameras would not match. This is because the displayed 'look' of SDR and HLG images, from cameras that implement the reference OETFs, is different (see § 6.5).

Scene-referred mapping will also work for mapping SDR to PQ. However, because the 'look' of PQ and BT.2020 SDR signals is very similar, for BT.2020 SDR signals the display-referred mapping will generally work well. To best match the PQ 'look', BT.709 SDR camera signals could be converted to BT.2020 SDR camera signals (using an OETF-based conversion similar to that specified in Recommendation ITU-R BT.2087) before display-referred mapping is applied.

The schematic diagram of the scene-referred mapping is illustrated in Fig. 43 for both PQ and HLG. It includes an optional artistic OOTF adjustment, for example to match the 'traditional colour reproduction' described in § 6.5.

FIGURE 43

SDR to HDR mapping (scene-referred)

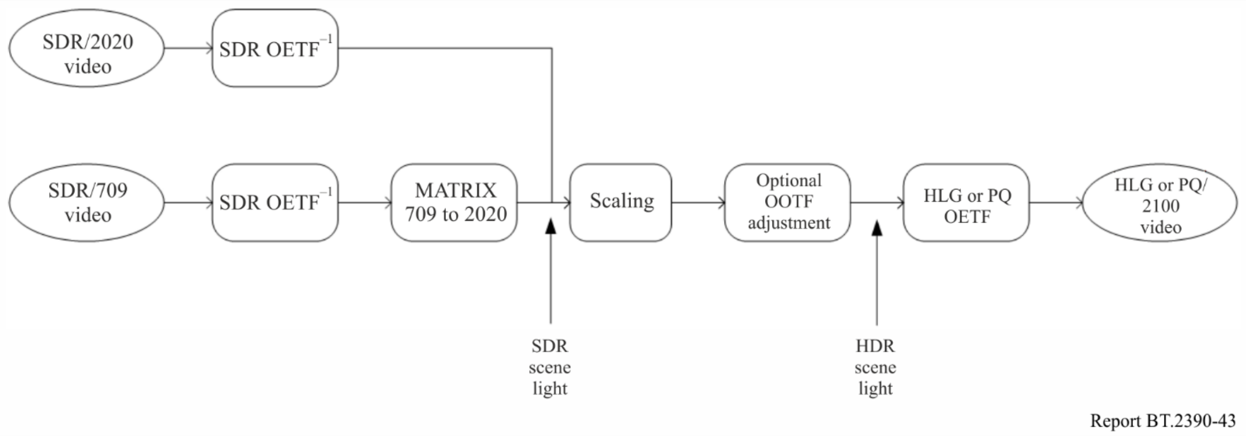


Figure 43 shows how the non-linear SDR BT.709 or BT.2020 video signal is converted to linear 'scene light' by applying the approximate inverse of SDR OETF, $E=(E')^2$, as described in BT.2087. When the SDR source is with the BT.709 colorimetry, the conversion is followed by the colour conversion matrix as described in Recommendation ITU-R BT.2087.

The scene light signal is then scaled so that the non-linear signal, after applying the reference PQ or HLG OETF, is at the appropriate signal level for HDR reference white: 58 %PQ or 75 %HLG respectively. Following any OOTF adjustment, the HLG or PQ OETFs are applied to derive the non-linear signals.

Section 10.2.1 describes how to calculate the scale factor for HLG, as well as how to adjust the OOTF to preserve a traditional SDR look.

10.2.1 Scene referred mapping of SDR into HLG

When (100X)% SDR signal is mapped to (100Y)% HLG signal, a scaling gain is calculated by the following equation:

$$\text{Gain} = \frac{\text{OETF}_{\text{HLG}}^{-1}(\text{Y})}{\text{OETF}_{\text{SDR}}^{-1}(\text{X})}$$

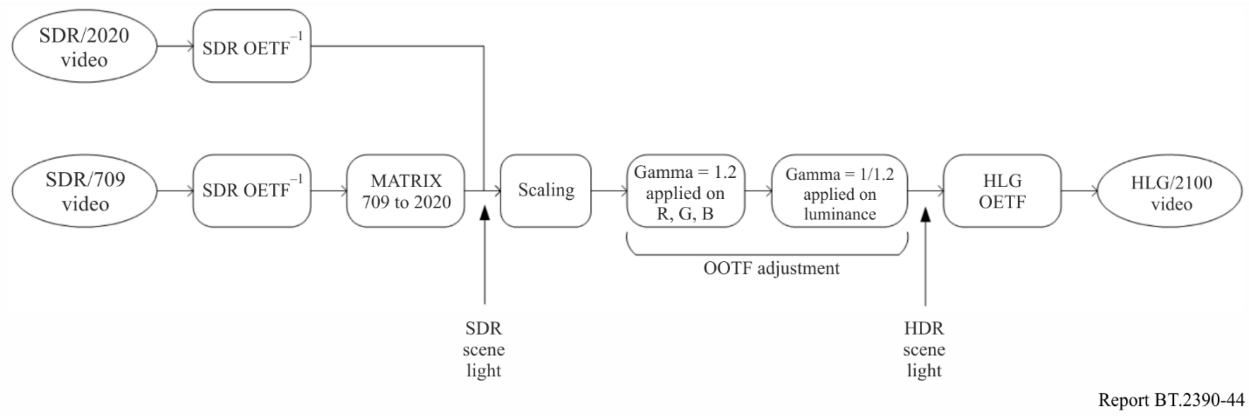
For example, when 100% SDR signal is mapped to 75% HLG signal, the scaling gain is calculated as follows:

$$\text{Gain} = \frac{\text{OETF}_{\text{HLG}}^{-1}(0.75)}{\text{OETF}_{\text{SDR}}^{-1}(1.0)} = \frac{0.265}{1.0^{2.0}} = 0.265$$

Where the SDR "look" is maintained during the conversion from SDR to HDR or the HLG camera is designed to deliver a traditional 'look' (see § 6.5), a small optional adjustment to the OOTF may then be applied to compensate for the subjective change in appearance of the SDR signal arising from a difference between HLG and SDR OOTFs. For the case when gamma adjustment is made to the scaled SDR scene light, the process is illustrated in Fig. 44.

FIGURE 44

SDR to HLG mapping with gamma adjustment (scene-referred)



11 Conversion practices for camera and display RGB colorimetry

Several camera and display systems, for both professional and consumer applications, use their own colour primaries, a practice that may give them certain advantages during capture or display respectively. However, content captured or displayed on such devices would still have to be transformed to or from a Recommendation ITU-R BT.2100 workflow, respectively. It should be noted that the transformations in this document only apply under the following conditions:

– The source and target white points are the same and should be equal to D65.

– The source and target white point brightness is the same. For scenarios where brightness is different, refer to Report ITU-R BT.2446.

Furthermore, these transformations are not applicable for camera raw signals.

Camera and display systems are commonly defined by their normalized primary matrix, NPM, which is specified as follows:

$$\mathrm{NPM} = \begin{bmatrix} X_R & X_G & X_B \\ Y_R & Y_G & Y_B \\ Z_R & Z_G & Z_B \end{bmatrix}, \quad (1)$$

where the elements of the matrix depend on the chromaticity coordinates, (x_R, y_R), (x_G, y_G), (x_B, y_B), and (x_W, y_W) for red, green, blue, and white, respectively, that characterize each system.

The NPM is needed for the conversion process to and from the CIE XYZ colour space and the BT.2100 colour space. Its elements could be computed as follows:

First, compute the z coordinates for all colour primaries as follows:

$$z_R = 1 - (x_R + y_R) \quad (2)$$

$$z_G = 1 - (x_G + y_G) \quad (3)$$

$$z_B = 1 - (x_B + y_B) \quad (4)$$

$$z_W = 1 - (x_W + y_W) \quad (5)$$

Then the matrix elements of NPM are derived as follows:

$$X_R = \frac{\left((y_G * z_B - y_B * z_G) * x_W + (x_B * z_G - x_G * z_B) * y_W + (x_G * y_B - x_B * y_G) * z_W\right) * x_R}{\left(x_R * (y_G * z_B - y_B * z_G) - x_G * (y_R * z_B - y_B * z_R) + x_B * (y_R * z_G - y_G * z_R)\right) * y_W} \quad (6)$$

$$X_G = \frac{\left((y_B * z_R - y_R * z_B) * x_W + (x_R * z_B - x_B * z_R) * y_W + (x_B * y_R - x_R * y_B) * z_W\right) * x_G}{\left(x_R * (y_G * z_B - y_B * z_G) - x_G * (y_R * z_B - y_B * z_R) + x_B * (y_R * z_G - y_G * z_R)\right) * y_W} \quad (7)$$

$$X_B = \frac{((y_R*z_G-y_G*z_R)*x_W+(x_G*z_R-x_R*z_G)*y_W+(x_R*y_G-x_G*y_R)*z_W)*x_B}{(x_R*(y_G*z_B-y_B*z_G)-x_G*(y_R*z_B-y_B*z_R)+x_B*(y_R*z_G-y_G*z_R))*y_W} \quad (8)$$

$$Y_R = \frac{((y_G*z_B-y_B*z_G)*x_W+(x_B*z_G-x_G*z_B)*y_W+(x_G*y_B-x_B*y_G)*z_W)*y_R}{(x_R*(y_G*z_B-y_B*z_G)-x_G*(y_R*z_B-y_B*z_R)+x_B*(y_R*z_G-y_G*z_R))*y_W} \quad (9)$$

$$Y_G = \frac{((y_B*z_R-y_R*z_B)*x_W+(x_R*z_B-x_B*z_R)*y_W+(x_B*y_R-x_R*y_B)*z_W)*y_G}{(x_R*(y_G*z_B-y_B*z_G)-x_G*(y_R*z_B-y_B*z_R)+x_B*(y_R*z_G-y_G*z_R))*y_W} \quad (10)$$

$$Y_B = \frac{((y_R*z_G-y_G*z_R)*x_W+(x_G*z_R-x_R*z_G)*y_W+(x_R*y_G-x_G*y_R)*z_W)*y_B}{(x_R*(y_G*z_B-y_B*z_G)-x_G*(y_R*z_B-y_B*z_R)+x_B*(y_R*z_G-y_G*z_R))*y_W} \quad (11)$$

$$Z_R = \frac{((y_G*z_B-y_B*z_G)*x_W+(x_B*z_G-x_G*z_B)*y_W+(x_G*y_B-x_B*y_G)*z_W)*z_R}{(x_R*(y_G*z_B-y_B*z_G)-x_G*(y_R*z_B-y_B*z_R)+x_B*(y_R*z_G-y_G*z_R))*y_W} \quad (12)$$

$$Z_G = \frac{((y_B*z_R-y_R*z_B)*x_W+(x_R*z_B-x_B*z_R)*y_W+(x_B*y_R-x_R*y_B)*z_W)*z_G}{(x_R*(y_G*z_B-y_B*z_G)-x_G*(y_R*z_B-y_B*z_R)+x_B*(y_R*z_G-y_G*z_R))*y_W} \quad (13)$$

$$Z_B = \frac{((y_R*z_G-y_G*z_R)*x_W+(x_G*z_R-x_R*z_G)*y_W+(x_R*y_G-x_G*y_R)*z_W)*z_B}{(x_R*(y_G*z_B-y_B*z_G)-x_G*(y_R*z_B-y_B*z_R)+x_B*(y_R*z_G-y_G*z_R))*y_W} \quad (14)$$

11.1 Conversion of normalized linear colour signals to Recommendation ITU-R BT.2100

In the case for conversion to the BT.2100 colour space, where the source colour space is linear, normalized within the [0:1] range, and defined by a particular NPM, conversion can be done as follows:

$$\begin{bmatrix} E_R \\ E_G \\ E_B \end{bmatrix}_{BT.2100} = \begin{bmatrix} 1.716651 & -0.355671 & -0.253366 \\ -0.666684 & 1.616481 & 0.015769 \\ 0.017640 & -0.042771 & 0.942103 \end{bmatrix} * \mathrm{NPM}_{Source} * \begin{bmatrix} E_R \\ E_G \\ E_B \end{bmatrix}_{Source}$$

and:

$$\begin{bmatrix} E_R \\ E_G \\ E_B \end{bmatrix}_{BT.2100} = \begin{bmatrix} 1.716651 & -0.355671 & -0.253366 \\ -0.666684 & 1.616481 & 0.015769 \\ 0.017640 & -0.042771 & 0.942103 \end{bmatrix} * \begin{bmatrix} X_R & X_G & X_B \\ Y_R & Y_G & Y_B \\ Z_R & Z_G & Z_B \end{bmatrix}_{Source} * \begin{bmatrix} E_R \\ E_G \\ E_B \end{bmatrix}_{Source} \quad (15)$$

Finally, since not all colours in the source representation may be within the BT.2100 representation, an additional clipping process may be performed. The negative values may be clipped to zero. The positive values may also be clipped to the capabilities of the interface. Although both soft or hard clipping could be performed (see Report ITU-R BT.2407, in many applications hard clipping is preferred. In the scenario that hard clipping of only the negative values is performed the process would be as follows:

$$E_R = \mathrm{Max}(0, E_R) \quad (16)$$

$$E_G = \mathrm{Max}(0, E_G) \quad (17)$$

$$E_B = \mathrm{Max}(0, E_B) \quad (18)$$

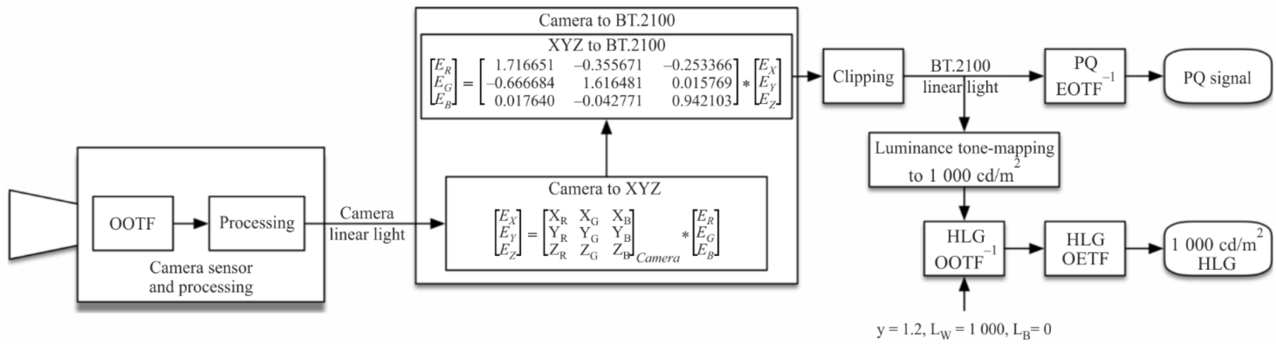
The above transformations could be applied in both display and scene referred workflows.

The conversion process, assuming a display referred camera workflow, as well as the final conversion to a BT.2100 representation, is shown in Fig. 45. For conversion to HLG, a bridge point of 1 000 cd/m^2 is assumed, and can therefore use the reference OOTF[6].

[6] Refer to § 7.4 of Report ITU-R BT.2390.

FIGURE 45

Conversion of arbitrary display referred linear light signals to BT.2100 signals using a display referred workflow

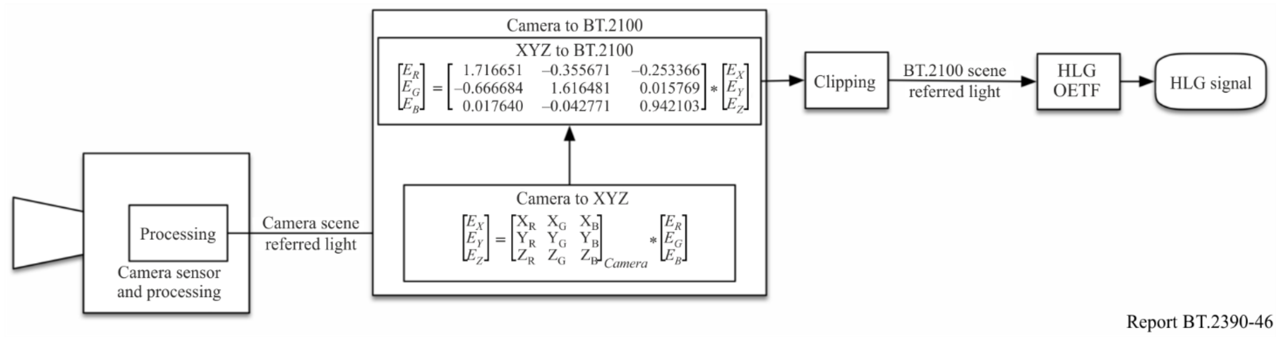


Report BT.2390-45

Figure 46 depicts the conversion process when applied on a scene referred workflow with the HLG Recommendation ITU-R BT.2100 signal as its output. Figure 47 depicts the same conversion process when applied on a scene referred workflow with the PQ Recommendation ITU-R BT.2100 signal as its output.

FIGURE 46

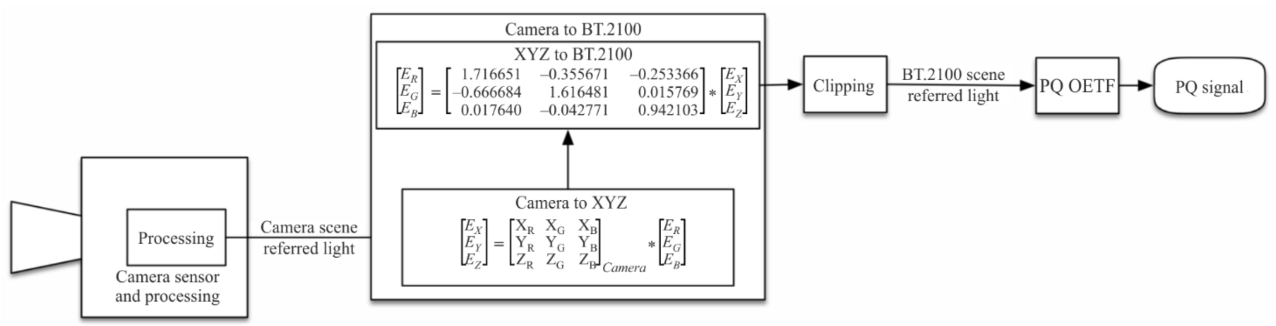
Conversion of arbitrary scene referred light signals to a BT.2100 HLG signal using a scene referred workflow



Report BT.2390-46

FIGURE 47

Conversion of arbitrary scene referred light signals to a BT.2100 PQ signal using a scene referred workflow



Report BT.2390-47

11.2 Conversion of BT.2100 to arbitrary linear colour signals for display systems

Similarly, conversion from linear and normalized Recommendation ITU-R BT.2100 RGB primaries to the RGB primaries of an arbitrary display system can be performed as follows:

$$\begin{bmatrix} E_R \\ E_G \\ E_B \end{bmatrix}_{Display} = \mathrm{NPM}_{Display}^{-1} * \begin{bmatrix} 0.636958 & 0.144617 & 0.168881 \\ 0.262700 & 0.677998 & 0.059302 \\ 0.000000 & 0.028073 & 1.060985 \end{bmatrix} * \begin{bmatrix} E_R \\ E_G \\ E_B \end{bmatrix}_{BT.2100}$$

and:

$$\begin{bmatrix} E_R \\ E_G \\ E_B \end{bmatrix}_{Display} = \begin{bmatrix} \mathrm{X_R} & \mathrm{X_G} & \mathrm{X_B} \\ \mathrm{Y_R} & \mathrm{Y_G} & \mathrm{Y_B} \\ \mathrm{Z_R} & \mathrm{Z_G} & \mathrm{Z_B} \end{bmatrix}_{Display}^{-1} * \begin{bmatrix} 0.636958 & 0.144617 & 0.168881 \\ 0.262700 & 0.677998 & 0.059302 \\ 0.000000 & 0.028073 & 1.060985 \end{bmatrix} * \begin{bmatrix} E_R \\ E_G \\ E_B \end{bmatrix}_{BT.2100} \quad (19)$$

Not all colours in the original representation may be within the target representation.

The negative values may be clipped to zero. The positive values may also be clipped to the capabilities of the display. Although both soft or hard clipping could be performed, in many applications, such as when using a reference display, hard clipping is preferred. In the scenario that hard clipping of only the negative values is performed the process would be as follows:

$$E_R = \mathrm{Max}(0, E_R) \quad (20)$$

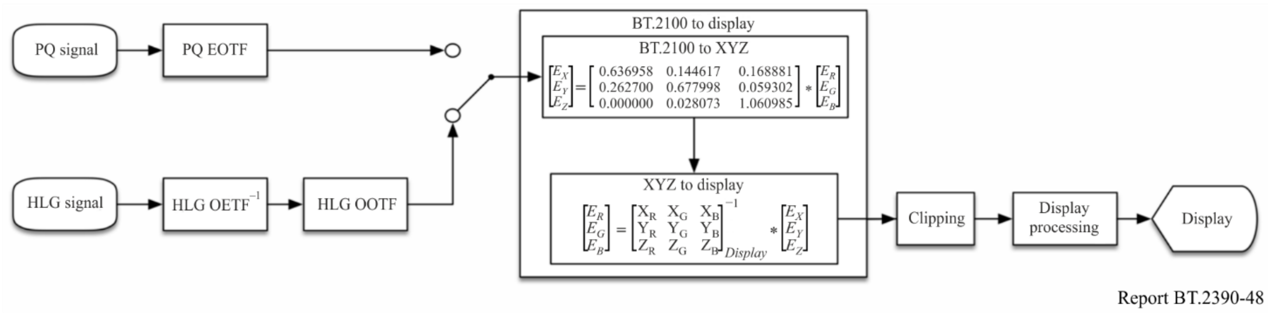
$$E_G = \mathrm{Max}(0, E_G) \quad (21)$$

$$E_B = \mathrm{Max}(0, E_B) \quad (22)$$

Figure 48 depicts this conversion process assuming a display referred workflow for both PQ and HLG. For conversion from HLG, the nominal peak luminance of the target display (and the appropriate system gamma) is used for the HLG OOTF.

FIGURE 48

Conversion of Recommendation ITU-R BT.2100 signals to an arbitrary display using a display referred workflow



References

[1] L.B. Wolf. Relative brightness of specular and diffuse reflection. Optical Engineering, 33(1), pp. 285-293, Jan 1994.

[2] V. Interrante, H. Fuchs, and S. Pizer (1997), Conveying the 3D shape of transparent surfaces via texture. ICASE report No 97-27. NASA contractor report 201705.

[3] R. Dror, A. Willsky, and E.H. Adelson (2004), Statistical characterization of real world illumination. JOV 4, pp. 821-837.

[4] A. Blake and H. Bulthoff (1991), Shape from specularities: computation and psychophysics. Phil Trans from the Royal Society of London. Series B, 33, pp. 237-252.

[5] S. Hecht, C. Haig, A.M. Chase, The influence of light adaptation on subsequent dark adaptation of the eye, Journal of General Physiology, 20 (1937), pp. 831–850.

[6] M.G. M. Stokkermans and I. Heynderickx (2014),Temporal dark adaptation to spatially complex backgrounds: effect of an additional light source. JOSA A V31 #7, pp. 1485-1494.

[7] R. Mantiuk, S. Daly, and L. Kerofsky (2010), “The luminance of pure black: exploring the effect of surround” SPIE Electronic Imaging Conference, San Jose.

[8] Rempel, A., Heidrich, W., Li, H., and Mantiuk, R., Video viewing preferences for HDR displays under varying ambient illumination," [Proc. of the 6th Symposium on Applied Perception in Graphics and Visualization], 45{52, ACM (2009).

[9] S. Daly, T. Kunkel, X. Sun, S. Farrell, and P. Crum (2013), “Viewer Preferences for Shadow, Diffuse, Specular, and Emissive Luminance Limits of High Dynamic Range Displays”, SID Display Week, paper 41.1, Vancouver, Canada.

[10] R. De Valois and K. De Valois (1990), Spatial Vision, p26, Ch.2, Oxford Science Publications., Oxford University press.

[11] N. Salvaggio (2008), Basic photographic materials and processes, 3rd edition. Focal press.

[12] S. Miller, M. Nezamabadi, S. Daly, “Perceptual Signal Coding for More Efficient Usage of Bit Codes,” SMPTE Mot. Imag. J., 122(4):52-59, May-June 2013.

[13] P.Heckbert, “Non-Overshooting Hermite Cubic Splines For Keyframe Interpolation”. New York Institute of Technology. February 1985.

[14] F. Ebner, M. Fairchild, “Derivation and Modelling of Hue Uniformity”. Rochester Institute of Technology. July 1998.

[15] “Encoding Color Difference Signals for High Dynamic Range and Wide Gamut Imagery”, Froehlich *et al*., Color and Imaging Conference, Volume 2015, Number 1, October 2015, pp. 240-247(8).

[16] J. Pytlarz, R. Atkins, T. Lu, F. Pu, P. Yin, T. Chen, W. Husak, “Overview of IPT-PQ Color Space”, Doc. m37266, October 2015, Geneva, Switzerland.

[17] T. Borer, A. Cotton, M. Pindoria and S. Thompson, "Approaches to high dynamic range video," 2016 Digital Media Industry & Academic Forum (DMIAF), Santorini, 2016, pp. 71-76. doi: 10.1109/DMIAF.2016.7574905.

[18] T. Borer, "Display of high dynamic range images under varying viewing conditions", Proc. SPIE 10396, Applications of Digital Image Processing XL, 103960H (19 September 2017); doi: 10.1117/12.2274253; http://dx.doi.org/10.1117/12.2274253.

[19] MovieLabs, “MovieLabs Best Practices for Mapping BT.709 Content to HDR10 for Consumer Distribution v1.0”, https://movielabs.com/download/938/.