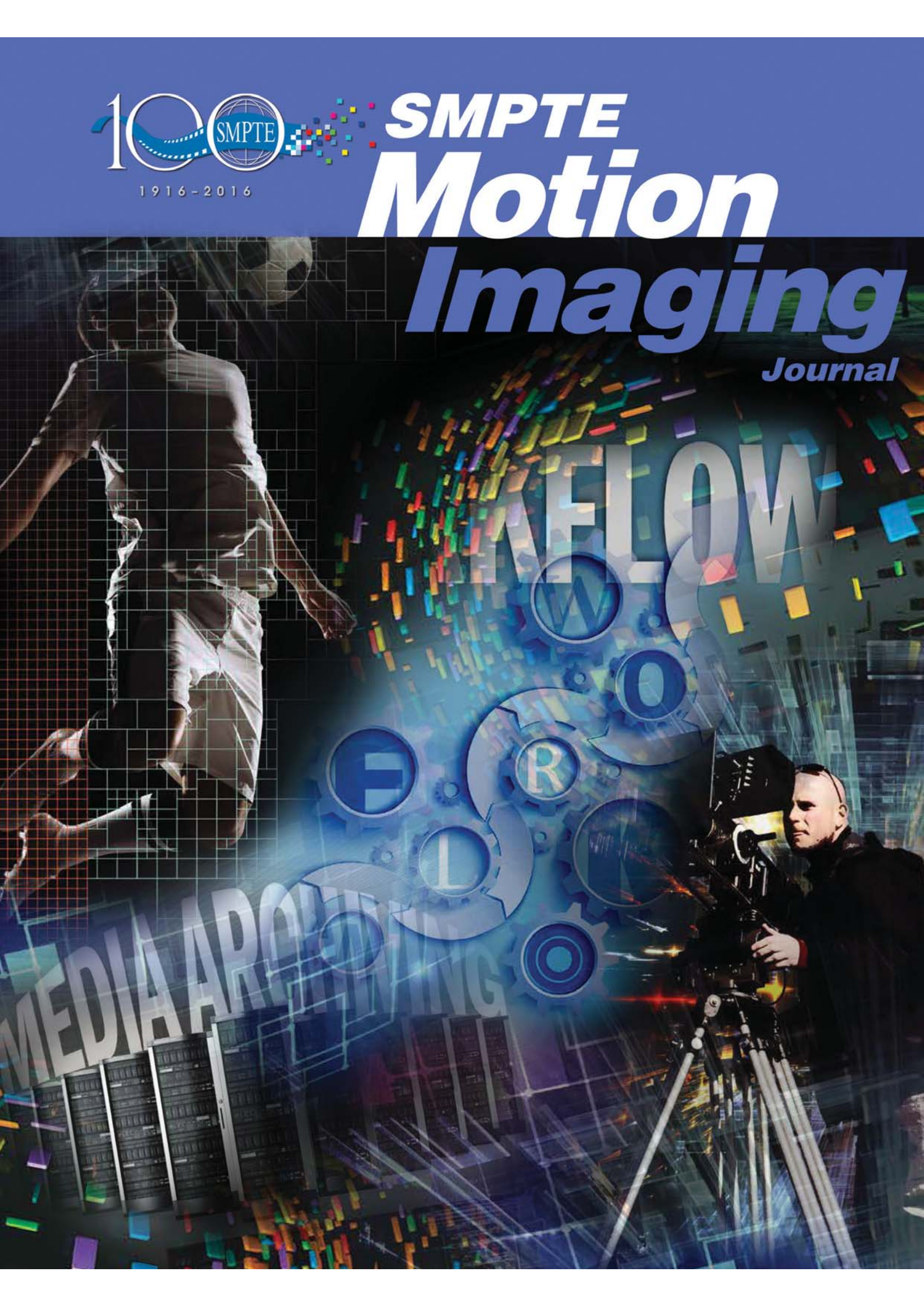

100
SMPTE
1916-2016
SMPTE
Motion
Imaging
Journal
WORKFLOW
MEDIA ARCHIVING

O artigo apresentado nesta edição aborda uma importante questão: Como a ampla gama de cores e as imagens de vídeo de alto alcance dinâmico devem ser codificadas? A principal conclusão a que os autores chegam é que a técnica Y'CBCR e as suas variantes são perfeitamente adequadas para uma gama dinâmica moderada, mas um menor desempenho quando combinadas com HDR. E, por isso, são necessárias novas técnicas de codificação.

Acompanhe nas próximas páginas, um texto claro e conciso onde os pesquisadores demonstram como chegaram a essa conclusão e desfrute da experiência única de compartilhar conhecimento fomentado pelo acordo entre a SET e o SMPTE, que nos permite traze-lhe a melhor da informação do mundo *broadcast*. Boa Leitura!

Tom Jones Moreira

Deploying Wide Color Gamut and High Dynamic Range in HD and UHD

by Charles Poynton, Jeroen Stessen, and Rutger Nijland

Twenty years ago, Poynton presented a paper at IBC 1994 entitled "Wide gamut device-independent color image interchange." The CCIR 709 standard had just been adopted (in 1990), and, by 1994, sRGB deployment in desktop computing was well under way. That paper anticipated commercial interest in exchange for wide-gamut imagery. As it turned out, wide gamut was not imminent: We've had 20 years of very stable color encoding for video in the form of BT.709 for HD (augmented recently by BT.1886, which finally standardizes gamma), and the 709-derivative sRGB that remains ubiquitous in the computer domain. Now, however, dramatic changes are under way. Wide color gamut (WCG), enabled mainly by RGB LED backlights for liquid crystal display (LCD) displays, has already seen initial deployment in consumer television. High dynamic range (HDR) cameras are commercially available; and HDR displays, mainly enabled by spatially modulated LED backlights, are on the verge of commercialization. Many industry experts agree that consumers will experience WCG and HDR as more significant than increasing spatial resolution from HD ("2K") to "4K." This paper revisits the topic of the 1994 paper, but now with some urgency, to address the question: How should wide color gamut and high dynamic range video imagery be encoded? The main conclusion is that the Y'CBCR technique and its variants are perfectly adequate for moderate dynamic range, but yield less than optimum performance when combined with HDR.

New encoding techniques are needed. We conclude that:

- A new high dynamic range opto-electronic conversion function (HDR OECF) (perceptual quantizer) should replace the conventional gamma function to enable HDR.
- HDR should be encoded with at least 10 bits per component, to suppress "banding." 10 bits $Y''C_BC_R$ 4:2:0 is at this moment the accepted standard for encoding HDR, and Philips will support the developments deriving from that choice.
- Going to 12 bits $Y''C_BC_R$ 4:2:0 will bring too little perceived improvement on natural content; further improvement must come from other changes.
- CBCR (chroma) subsampling performs worse in combination with the HDR OECF; we propose encoding and decoding constant luminance, with modified $u'v'$ chromaticity components instead of C_BC_R.
- Therefore, for the future, we propose going to 10 bits $Y''u''v''$ 4:2:0.

INTRODUCTION

For 20 years HD material has been mastered to a fixed set of display primaries, those standardized in ITU-R Rec. BT.709, which are best described as having moderate color gamut. The BT.709 primaries were chosen in 1990 to closely approximate the CRT phosphors that had been in use since about 1965. Liquid crystal displays (LCD) commercialized since 1990 have been designed to have primaries comparable to those of BT.709, partly because virtually all of the available content was mastered to those primaries, and partly because BT.709 primaries were easily achieved by cold cathode fluorescent lamp (CCFL) and white light-emitting diode (LED) backlight units (BLUs). Now, though, BLUs incorporating red, green, and blue LEDs are economical.

Each of the red, green, and blue LED types has a rather narrow spectral spread (between about 25 nm and 35 nm); the narrow spectral coverage leads to the possibility of wide color gamut (WCG), wider than BT.709. Another opportunity is backlights based on Quantum Dot technology, from 3M, and Sony's TriLuminos backlights. Typical RGB LED BLU technology enables display gamut approximately matching the P3 gamut of the digital cine-

ma initiative (DCI). The possibility arises for studios to deliver movie-class color gamut to consumers; movies would benefit, and so would sports and live events.

Consumers seem to like colorful pictures. Consumer electronics (CE) manufacturers have found that television sets producing colorful pictures are more profitable than those delivering pictures that are less colorful. Today, however, there is no WCG program material available. So, consumer manufacturers have built signal-processing circuitry to expand the color range of BT.709 material. The colors that are displayed are not faithful to the original. One goal of our work is to allow content creation with wide gamut and to encode and decode in a way that makes it possible to display authentic wide gamut color to consumers.

The second development is high dynamic range (HDR)[2]. Conventional high definition (HD) video is approved at a contrast ratio of about 1000:1; diffuse white is portrayed at about 100 nit; and the blackest black is about 0.1 nit. (We use nit as an abbreviation of candela per meter squared, cd/m^2.) Consumers prefer brighter pictures than those displayed at program creation: today's consumer experiences diffuse white at between 300 and 500 nit; black level is typically between 0.3 and 2 nit. For this contrast range, at consumer quality level, 8-bit components coded using a 2.4-power function, as defined in BT.1886, are sufficient. 10-bit components are used in the studio, and 10-bit components would deliver standard dynamic range (SDR) video with better performance to consumers than today's 8-bit components.

Much work has been done in HDR acquisition, and capture of live action at HDR is now fairly simple using several different camera types. On the display side, a Canadian company called Brightside developed a particularly interesting type of display technology[3]. Such HDR display involves an area array backlight comprising around 1,000 LED clusters (instead of the more common linear array with a few dozen LED clusters); backlights are individually controlled, achieving spatial backlight modulation. Some CE manufacturers conceptualize the scheme starting with fully-on backlights and call the scheme "local dimming." We prefer to say, "local brightening."

Should active matrix organic light emitting diode (AMOLED) displays be commercialized for consumer television, we expect at least some of them also to be able deliver HDR-class imagery. Our goal is to take WCG/HDR material at the approval stage of production (prior to mastering), and encode into signals that can be presented to a conventional H.264/265 compressor. After decompression at the consumers' premises, we decode for WCG/ HDR display. We seek unlimited color gamut, but we expect HDR displays to have gamut approximating that of the Digital Cinema Initiatives (DCI) P3 standard. Luminance of the portrayal of diffuse white need not be higher than about 500 nit, but we seek to portray specular highlights and directly light sources using luminance levels perhaps ten times higher than diffuse white, a capability unavailable in today's systems. We also seek to enable HDR displays to present black darker than today's 0.3 nit or so.

CONCEPTS

We will speak of an *encoding standard*. Historically, we would have said *transmission standard*, but that term fails to encompass modern distribution technologies. Decoding is as close as possible to the inverse of encoding; however, encoding is somewhat lossy, so encoding is not perfectly inverted. We are not encoding the scene; we encode the material that is presented for approval at the final stage of postproduction. By encoding and decoding, we refer to representing tone (grayscale) and color. (The terms *encoding* and *decoding* are ambiguous because these terms are also used to refer to motion-compensated transform-based compression systems such as H.264/265.) Here, wide gamut and HDR image data is encoded into three components that are presented to such a compressor. After distribution, and decompression, the three color components are decoded prior to display[4].

When $R'G'B'$ signals are conveyed (or decoded from $Y''C_BC_R$), they are conveyed in display-referred form: the decoded XYZ components represent the colors intended to be displayed. Cathode ray tube (CRT) displays historically had physical primaries matching the reference primaries defined in the encoding color space, and they also had an intrinsic electro-optical conversion function (EOCF) matching the encoding standard.

In professional imaging, and especially entertainment imaging, encoding typically has only an indirect connection to the scene and the camera. In computer animation, or other synthetically generated content, there is no physical scene and no physical camera at all! In the general case, what is important to content creators is that the image displayed to the consumer is a reasonable approximation of the image as displayed on an approval display (e.g., studio reference display) at the end of the production and post-production chain.

Upstream of approval, there may be science, but what really matters is art and craft. Downstream of approval, ideally there is just science.

Today's world offers a wide diversity of display devices; these display devices have a diversity of tone and color characteristics. We expect a transform to take place at the viewing device to adapt transmission encoding to the native device. (For example, in today's LCDs, the LCD driver circuitry incorporates compensation of the native LCD S-shaped EOCF function.) Today's BT.1886 is not capable of HDR; in order to accommodate HDR content in the transmission chain, we'll need an

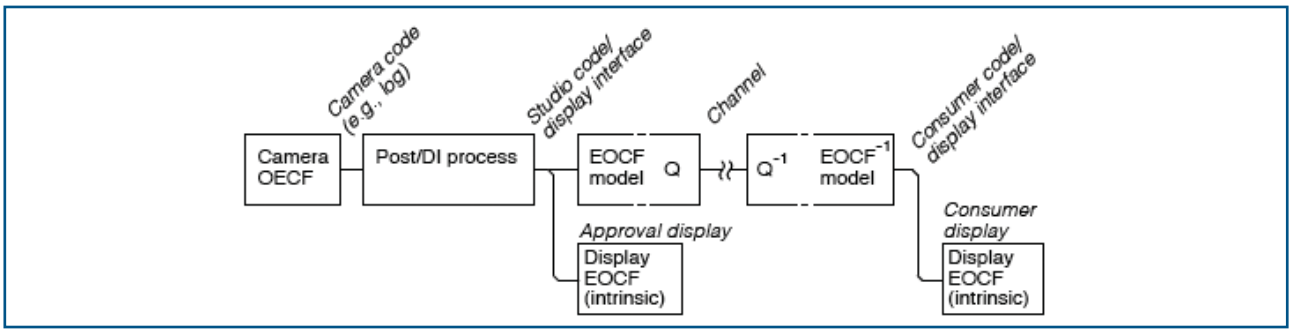


Figure 1 Image production flow

HDR-capable quantizer. A perceptual quantizer has been proposed by Miller[5].

The diversity of display devices comes with a diversity of physical display primaries. Signal processing to accomplish a color transform from the encoding ("transmission" or "interchange") primaries to the display device is expected to be implemented at the viewing device. Many television engineers are familiar with colorimetric transforms implemented as 3 x 3 "linear-light" matrix processing to transform from one primary set to another. These transforms are perfectly suitable when the source color space is completely contained within the destination space, as is the case in transforming BT.709 to wide-gamut primaries such as those of DCI P3 or BT.2020. However, a nontrivial colorimetric transform never fills the destination color space; CE manufacturers typically implement a non-colorimetric transform that stretches colors into the destination gamut in order to make the picture more colorful. When transforming to a destination space that has a smaller gamut than the source, a colorimetric transform is bound to clip some colors. We expect that colorimetric transforms will not suffice; we hope that some standards concerning gamut mapping can be established.

Historical video systems have used pure power functions at the display. $(Y''C_BC_R)$ interchange signals are converted to (R', G, B') image signals, and each of the (R', G, B') image signals is raised to approximately the 2.4-power to yield the display tristimulus (R,G,B). The scheme is better than using code values proportional to light intensity ("linear-light"); however, power-function coding places many more digital codes in the light tones than are needed, and not enough codes in the deep blacks. The way to optimize the coding for visual perception is to determine how many "just noticeable difference" (JND) steps are perceived by human vision and to quantize accordingly. The late Peter Barten, of Philips, completed a very detailed study of this issue[6,7].

We use Barten's work to establish perceptual quantization for an opto-electronic conversion function (OECF), and we will confirm that we need 10 to 12 bits coding for HDR. But having given enough bits, other problems come up that cannot be cured with more bits.

Conventional video systems form a "luma" component representative of the achromatic content of the image and two components carrying the "chroma": $(Y''C_BC_R)$. The chroma components are subsampled, that is, spatially lowpass filtered and downsampled, typically in the 4:2:0 scheme where chroma resolution is reduced by a factor of 2:1 in both the horizontal and vertical domains. These calculations are done in the gamma-corrected domain (non-constant luminance); so, the calculations are affected by the choice of (R',G',B') coding (gamma). The scheme works well for moderate (R',G',B') nonlinearity such as the 2.4-power function of BT.1886 for HD. However, we have found that serious chroma subsampling artifacts result when the same calculation is performed on (R',G',B') signals having an HDR OECF.

We propose to convey color using true color science chromaticity coordinates $(u'v')$ instead of (C_BC_R). We have found that $(u'v')$ can be subsampled 4:2:0 without visible impairment. The system we propose has true constant luminance, owing to the fact that one component contains all of the International Commission on Illumination (CIE) luminance.

The suggestion of using $(u'v')$ components for color digital image data dates back many decades[8] prior to the invention of DCT-based compression. Use of log luminance accompanied by $(u'v')$ components has been proposed in recent times, for example by Larson[9]; however, such proposals do not include subsampling of the color components.

EOCF ANALYSIS

We use the following analytical function to approximate Barten's function, where L is absolute luminance [nit] and V is video signal code value (from 0 to 1000):

decode form: $$L = L_{\text{nom}} \cdot \left(\frac{e^{\left[m \cdot \frac{V}{V_{\text{nom}}} \right]} - 1}{e^m - 1} \right)^{\gamma} \quad (1)$$

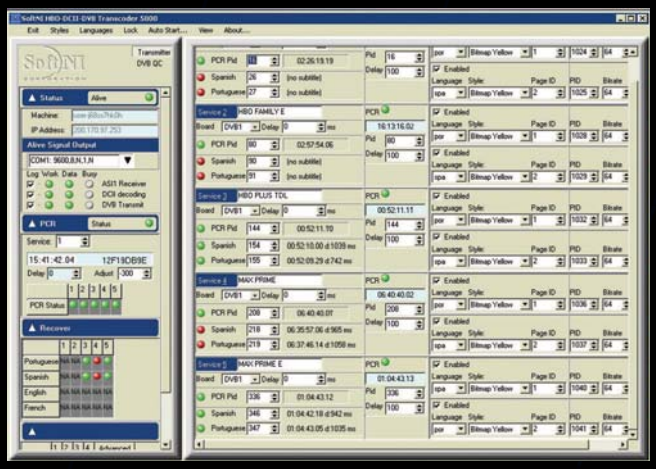
decode form: $$V = V_{\text{nom}} \cdot \frac{\ln\left[\left(\frac{L}{L_{\text{nom}}} \right)^{1/\gamma} \cdot (e^m - 1) + 1 \right]}{m} \quad (2)$$

We found these parameters: L_{nom} = 10000 nit, V_{nom} = 2305.9, m = 4.3365, γ = 2.0676.

SoftNI, the leading developer of advanced subtitling and captioning solutions, presents a SET EXPO 2016 the latest versions of the Subtitler Suite, Casat and Digital subtitling Suites, and LIVE captioning.

CaSat Subtitling Suite™

The Most Advanced and Reliable Subtitling Transmission System for HD/SD ATSC, DVB, ISDB, Motorola®/SCTE-27 and DirecTV® Standards



LIVE Subtitling Suite™

Innovative Solutions for Subtitling and Closed Captioning of Live Events.
High-quality Live Captioning in the same language and Live Subtitling in foreign languages

Digital Subtitling Suite™

Software-based Insertion of Captions and Metadata into Digital Video Files and Transport Streams

Subtitler Suite™

The Most Complete and Multi-purpose All-In-One Subtitling & Captioning System

SoftNI has been selected by Cisco, DirecTV, Embratel, Ericsson, FOX, Globosat/TVGlobo, GrassValley, HBO, NET, SKY, SONY/SPE, Technicolor, Telefonica, Televisa, Thomson, Turner, Warner and many other leading broadcasters and post-production facilities worldwide.

SoftNI Advanced Subtitling & Captioning Systems

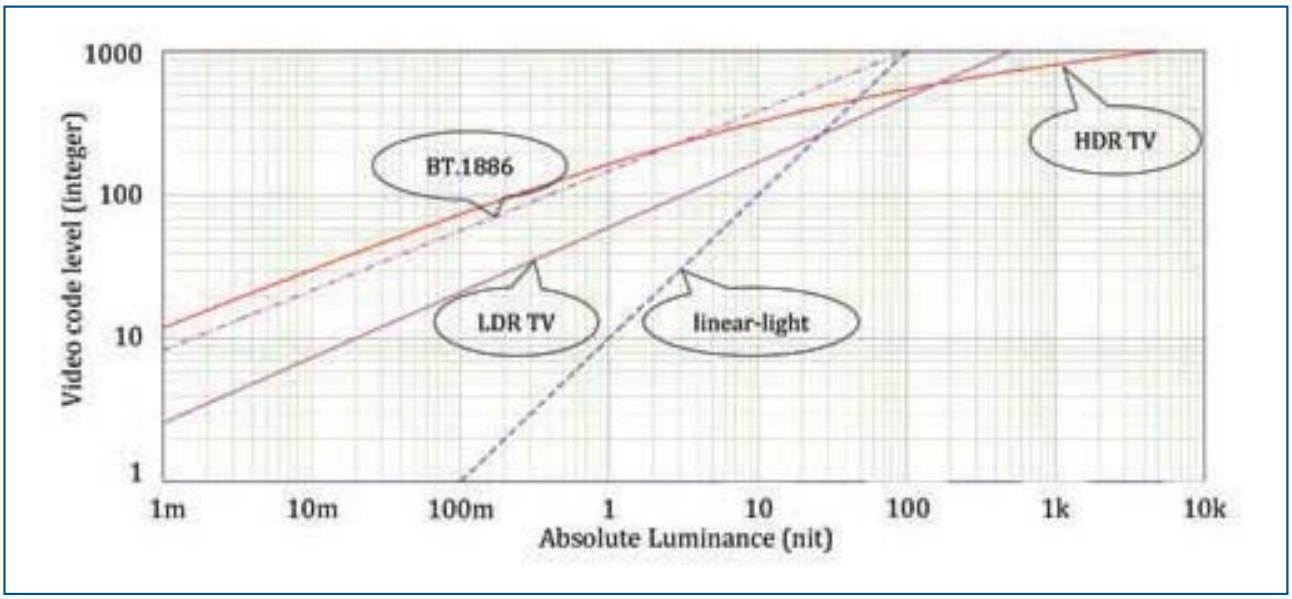


Figure 2 Various EOCFs – log-log scale

The encoder form can be interpreted as a Lightness formula; it predicts number of quantization steps required for a certain luminance range. For luminance range from 0 to 100 nit, it predicts V = 1176 steps (comparable to the 10 bits used in HD studio video). It predicts 1728 steps for 1000 nits, and 2306 steps for 10000 nits (i.e., 11.2 bits). This is a worst-case analysis, assuming perfect luminance adaptation and (noise-free) content that best reveals "banding" artifacts. It is still possible to adequately encode HDR video using only 10 bits.

Video engineers have historically been concerned with gamma at the display. Gamma is the numerical value of a presumed power function that maps the video signal (conceptually from 0 to 1) to light—relative luminance (Y), or tristimulus (R,G,B). This is an electro-optical conversion function (EOCF). In the limited dynamic range of historical video, a gamma function imposed a fair degree of perceptual uniformity. In HDR, we need perceptual uniformity over a much wider range; we need a new perceptual quantizer (an HDR OECF).

It is commonly believed that the camera's *opto-electronic conversion function* (OECF) should be the inverse of the display function; but that is not the case, mainly because of the necessity to impose picture rendering[10]. The camera does not play directly in our story — we are not concerned with any actual OECF. However, we are concerned with the inverse of the EOCF, which we should denote EOCF^{-1}. **Figure 1** shows the flow.

Several inverse EOCF^{-1} functions are shown in **Fig. 2**. The horizontal axis is absolute luminance (L) on a log scale from 10^{-3} to 10^{+4} nit. The vertical axis is the digital video signal (V) on a log scale from 1 to 1000 (10 bits).

- The dashed blue line represents a linear-light function (*i.e.*, $L \propto V$), here from 0.1 to 100 nit. Linear-light coding is impractical for image interchange[10].
- The dashed-dotted magenta line represents video decoding according to the BT.1886 standard; here, reference white is 100 nit. The line has a constant slope of $^{1}/_{2.4}$[10].
- The solid magenta line represents a typical consumer television receiver EOCF from 0 to 500 nit. This line represents typical consumer TV receiver behavior; the line has a constant slope of $^{1}/_{2.2}$, which is suitable for reference white luminance considerably higher than 100 nit and surround condition brighter than approval condition.
- The solid red line represents an HDR function from 0 to 5000 nit, proposed by Philips, inspired by Barten's contrast sensitivity function (CSF), and similar to that proposed by Miller.[5] The dark part of this line has a slope of $^{1}/_{2.4}$. The brighter part of the line is a logarithmic function. The logarithmic property allows an efficient expansion of the dynamic range to very high luminance, more so than any realistic power function.

To better appreciate the nonlinearity of various EOCFs, we can plot them on a linear scale **Fig. 3**.

- SDR TV uses a gamma function based on CRT (BT.1886); not very linear (Y').
- HDR TV uses a gamma-log function based on Barten's CSF; very nonlinear (Y'').

A linear-light luminance signal is written as Y; in the gamma domain, the signal is called luma and written as Y'. We need a new notation for the corresponding quan-

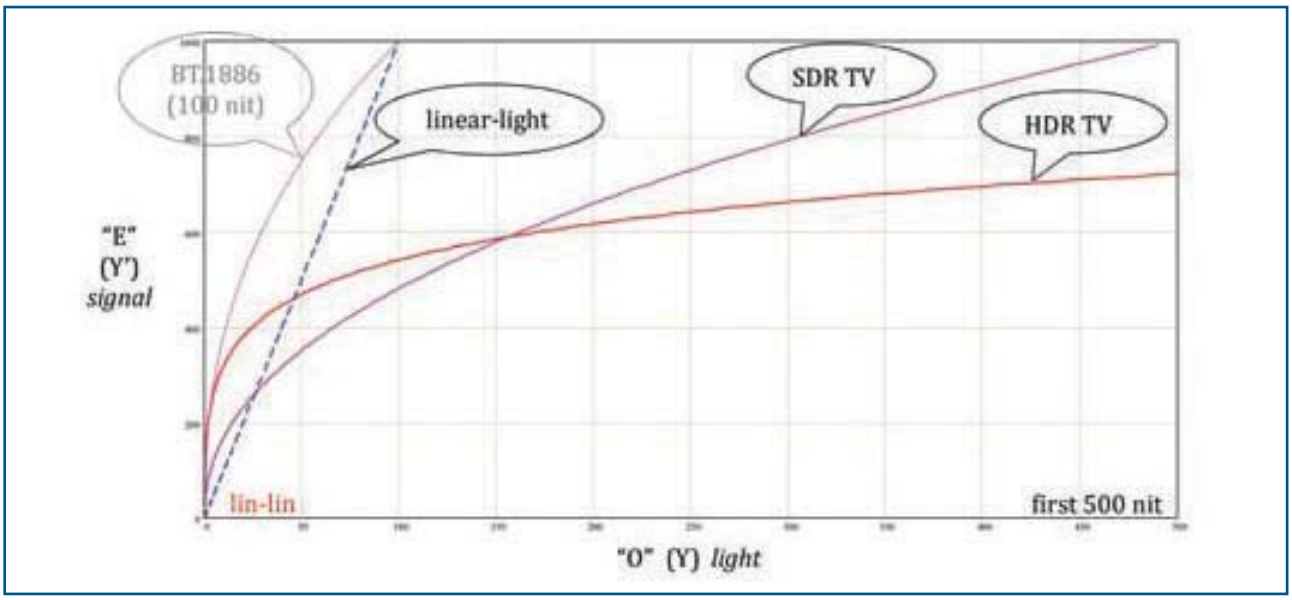


Figure 3 Various EOCFs – lin-lin scale

tities in the HDR perceptually quantized domain. We will write Y'' and (R'',G'',B'').

We can evaluate the visibility of quantization of the V signal for various functions. We plot $\Delta f/L$ against L, on log-log axes, for various functions f. The graph is presented in **Fig. 4**.

We have added, to the four curves in **Fig. 4**, a dashed red line that graphs the reciprocal of Barten's CSF; it represents the just noticeable quantization step across the luminance range. Anything above the red dashed line is liable to be visible as a false contour. The linear-light signal is quantized far too coarsely at the dark end and far too finely at the bright end.

The Barten CSF line shows that the dynamic range can be extended indefinitely at a ratio of 1.004 per step, analogous to the Weber- Fechner "law."

USE OF THE OECF

The OECF is applied by the transmitter. It can be used in two different ways, as seen in **Fig. 5**.

The top part of **Fig. 5** illustrates the "non constant luminance" way of generating $(Y'C_BC_R)$ signals. The OECF is applied to each of the (R,G,B) components before mixing the signals into Y' = luma. This has been the method used in all color TV standards since NTSC (National Television Systems Committee).

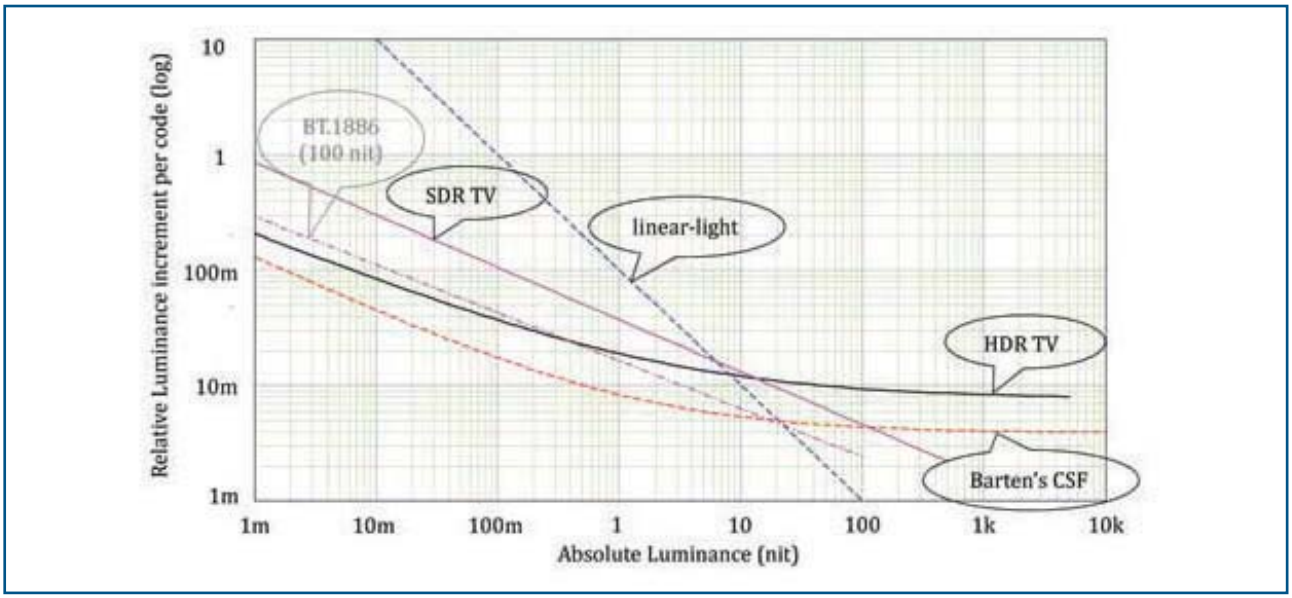


Figure 4 Quantization visibility

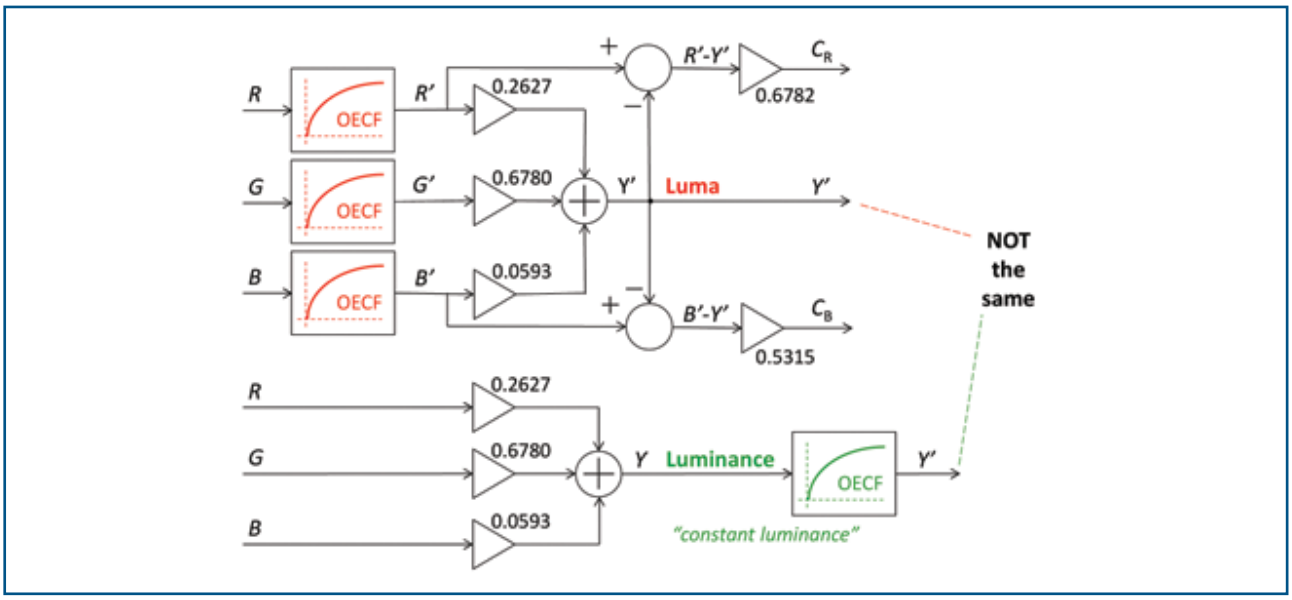


Figure 5 OECF and Luma signals

The bottom part of **Fig. 5** illustrates the "constant luminance" way of generating Y = luminance, and then converting that to a Y' signal by means of a single OECF.

The two Y' signals are generally not the same, except for black and white pixels ($R=G=B$). We shall see that there are important differences for the quality of color signals.

CIE CHROMATICITY AND UNIFORM CHROMACITY SCALE (UCS)

It is standard for (C_B,C_R) signals to have the same bit depth and precision as the associated Y' signal. (Y', C_B,C_R) are always defined for a certain color gamut, like BT.709 (a.k.a. sRGB). If we want to increase the color gamut then we can use some of the "illegal" codes to represent colors outside the standard RGB cube. Alternatively, we can choose more colorful ("saturated") color primaries, as is done in BT.2020. To maintain precision for a larger color space, the range of (Y',C_B,C_R) values in BT.2020 should be increased: both HDR and WCG should use chroma signals having about one additional bit in order to maintain today's color precision.

Requirements for color signals can be relaxed by choosing a more perceptually uniform color space, that is, one with fewer code combinations that are used more efficiently. Instead of *chroma* signals we can choose *chromaticity* signals. On top of that, the latter are independent of dynamic range, which will save even more codes.

Many image coding and video engineers are familiar with CIE (x,y) chromaticity coordinates, formed from a projective transformation of CIE (X,Y,Z). In 1976, the CIE defined[11] a uniform chromaticity scale (UCS) in which the coordinates are much more perceptually uniform **Fig. 6**.

The (u',v') coordinates are formed from a projective transformation of either (X,Y,Z) or (x,y):

$$u' = \frac{4X}{X+15Y+3Z} = \frac{4x}{3-2x+12y}, \quad v' = \frac{9Y}{X+15Y+3Z} = \frac{9y}{3-2x+12y} \qquad (3)$$

The inverses of the (u',v') system are not often found in the literature; we state them here:

$$x = \frac{9u'}{12+6u'+16v'}, \quad y = \frac{4v'}{12+6u'+16v'} \qquad (4)$$

To recover the other two tristimulus linear-light (X,Z) components, the inverses are these:

$$X = Y \cdot \frac{9 \cdot u'}{4 \cdot v'} \qquad Z = Y \cdot \frac{12-3 \cdot u'-20 \cdot v'}{4 \cdot v'} \qquad (5)$$

From (X,Y,Z) it is an easy step—a 3 x 3 matrix—to form any (R,G,B) values for a display. These are projective transforms, so (u',v') is a chromaticity space having coordinates that are invariant with scaling of (X,Y,Z). In our view, this scaling invariance is a critical property of an image code for HDR image data. The (a^*,b^*) coordinates of the CIE LAB system, the (C_B,C_R) components of video, and the (D_Z,D_X) components that have been proposed for HDR, all do not have this property: the chroma components in the latter systems do not have chromaticity diagrams, and the chroma components vary as (R,G,B) or (X,Y,Z) are scaled. Also, (D_Z,D_X) do not have constant values along the grayscale (depending on the chosen white point).

Figure 7 illustrates the relation between the available color space in (u',v') coordinates versus some of the traditional RGB color spaces.

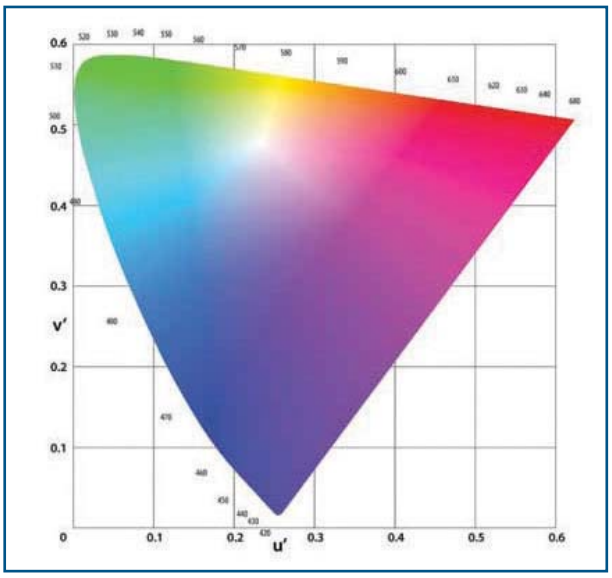


Figure 6 UCS 1976 Chromaticity plane

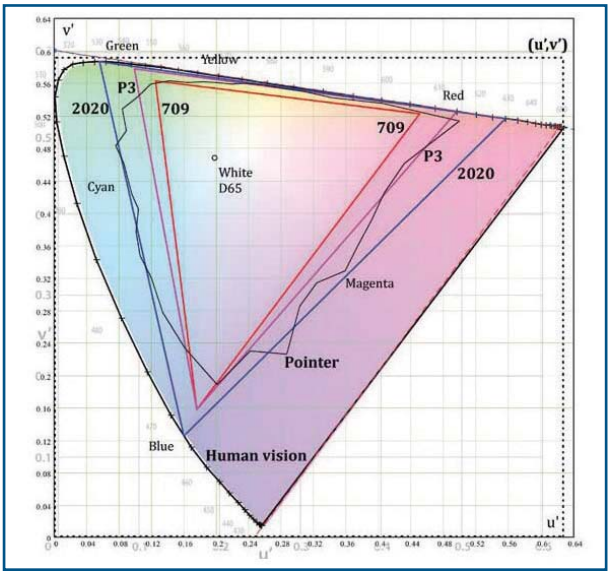


Figure 7 Color spaces

Sizes from small to large:

- BT.709 color gamut & space
- DCI P3 color gamut
- BT.2020 color space
- Pointer set of surface colors
- Entire human color vision (the "horseshoe")
- UCS 1976 (u',v') color space (the dotted square)

The (u',v') color space covers more than all humanly visible colors, so it is forever large enough. The same can be said for the (X,Y,Z) and Academy Color Encoding System (ACES) (R,G,B) color spaces, but they too have some problems for HDR.

Figure 8 illustrates the large XYZ and ACES triangles drawn in **Fig. 7**.

Even in the XYZ and ACES color spaces the red, orange, yellow, and yellow-green colors touch one edge of the color space ($Z=0$, $B=0$, respectively). We will show this to be significant when we construct chroma signals from (X',Y',Z') or (R',G',B') (without using constant luminance).

CHROMA VERSUS CHROMATICITY

Figure 9 illustrates the difference between chroma and chromaticity signals.

The left images are in a (Y',C_B,C_R) 3D color space; the right images are in (Y,x,y).

The top shows a 3D side view of the space; the bottom shows a 2D top view.

A typical color gamut is represented by the 12 edges of an (R,G,B) cube.

In the (Y',C_B,C_R) color space, the cube remains a cube, but it is standing on its black tip.

In the nonlinear (Y,x,y) color space, the cube is totally distorted, though you may still recognize it.

The top views (at the bottom) illustrate the difference between a chroma plane and a chromaticity plane. The chroma plane shows the familiar hexagon shape known from the vectorscope. The chromaticity plane, here it is (x,y), shows the familiar triangle from color gamut comparisons. Here, (u',v'), not shown, is just a different perspective projection from (x,y).

In the chroma plane we see that all of hue, saturation, and brightness are encoded. Specifically, if we move from the center to one of the primary R,G,B colors, then we see that the brightness increases.

This implies that (C_B,C_R) have the same dynamic range as Y' or Y''. We have found that to avoid visibility of quantization artifacts in the most critical circumstances (C_B,C_R) need 12 bits or more for HDR. Y'' would need almost 12 bits. With practical content there is little perceived difference between 10 or 12 bits.

In the chromaticity plane, we see that only hue and saturation are encoded and (u',v') have no dynamic range at all, and no correlation with luminance. The brightness dimension (Y or Y' or Y'') is perpendicular to this plane. We have found that (u',v') need only 9 bits each, and that this does not increase for a higher dynamic range. It depends on the error metric that one uses; consider that CIEDE2000 was not designed for WCG or HDR. It is practical to use 2x 10 bits.

Luminance is transmitted at the full spatial resolution, as a Y'' signal through a high-bandwidth HDR channel. The Y'' signal carries full dynamic range but no color; it needs 10 bits (practical for TV) to 12 bits (perfect).

The chromaticity signals (u',v') are transmitted over two downsampled channels (preferably 4:2:0). With modern digital signal processing hardware the complexity of the conversion is almost trivial. Here, (u',v')

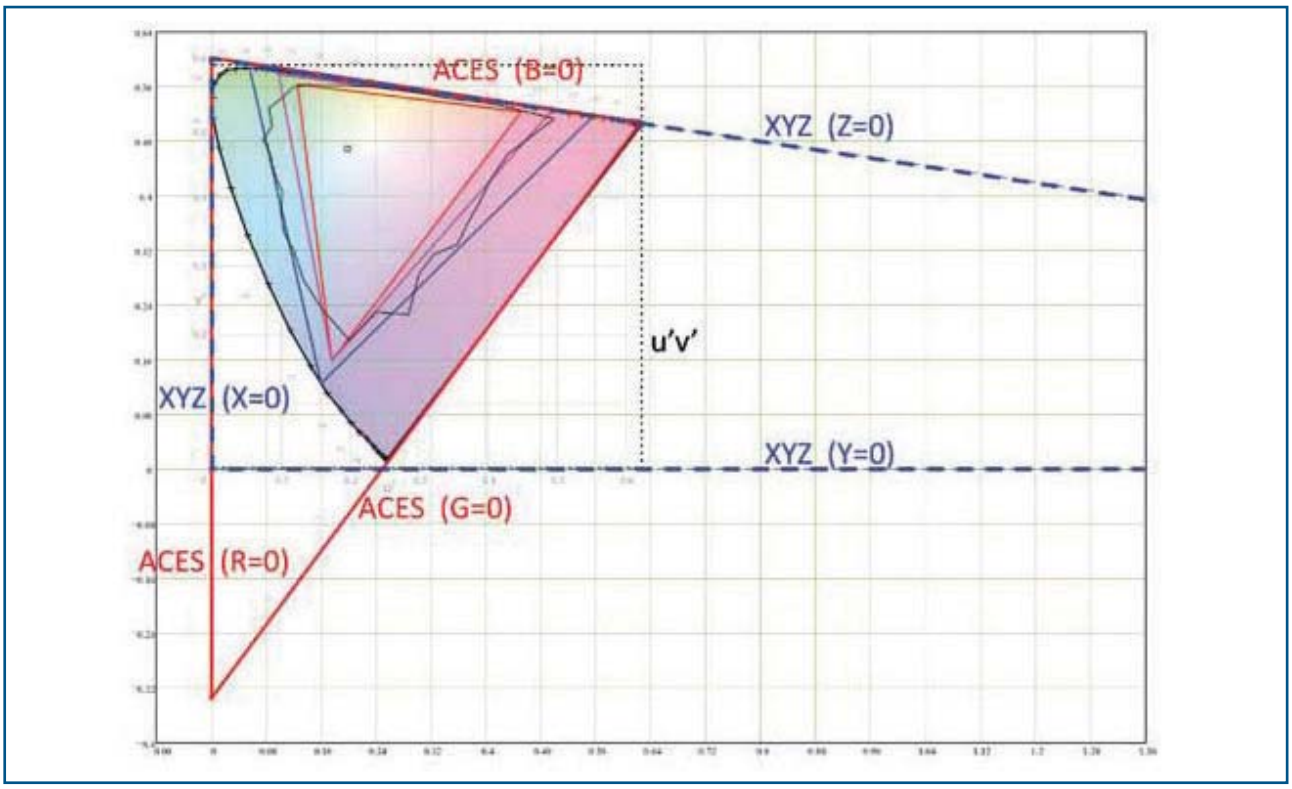


Figure 8 Color spaces including XYZ and ACES

carry the entire theoretical color gamut (no limitations) but no dynamic range; each needs 9 to 10 bits.

The only question that remains is how (u',v') should be sub-sampled to a lower spatial resolution and back, while not introducing significant perceived errors of color or brightness on the display. This is not trivial.

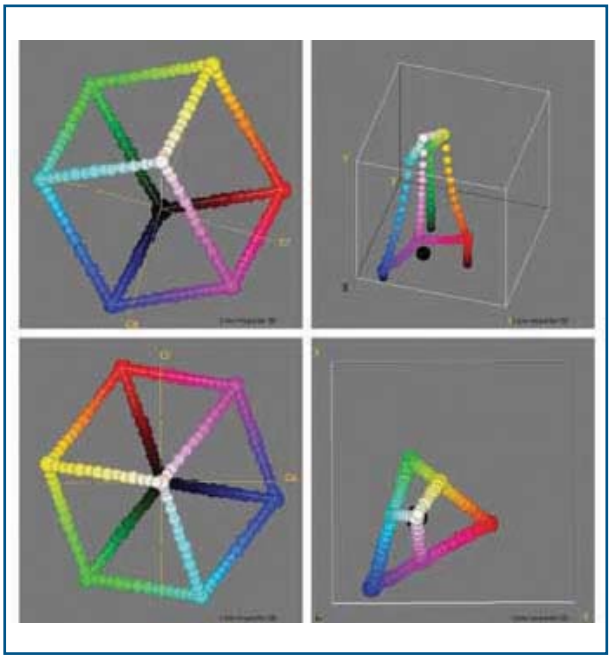


Figure 9 Chroma and chromaticity

The magnitude of the (u',v') signals is totally independent of the luminance of colors. If we try to mix dark and bright colors in the (u',v') domain then we will see an unjustifiable dominance of the dark colors. Only if we mix colors in the linear-light domain do we get the same resulting color as when colors are mixed in our eyes.

The same problem exists for the (C_B,C_R) signals: due to the HDR OECF applied to (R,G,B) the magnitude of (C_B,C_R) for dark colors is much larger than it should be. If we try to mix dark and bright colors in the (C_B,C_R) domain then again we'll see an unjustifiable dominance of the dark colors. For an HDR OECF (almost a logarithmic function) this effect is much worse than for an SDR OECF (gamma function). This is illustrated in the next section.

COLOR CROSSTALK

Figure 10 shows some examples of color mixing going wrong on high-frequency (on-off) textures. The left half of each picture is an original; the right half is after conversion to 4:2:0 and back. The images make it evident why HDR proposals based upon (C_B,C_R) or (D_Z,D_X) fail: the darker of the two colors becomes too dominant.

The proper solution is to do the color mixing, especially the lowpass filters for color down-sampling, in the linear-light (R,G,B) or (X,Y,Z) domain. This fixes the problem, but we maintain the advantages of (u',v').

Figure 11 illustrates the difference in color crosstalk between (Y'',C_B,C_R) and (Y'',u',v') (zoom in if necessary).

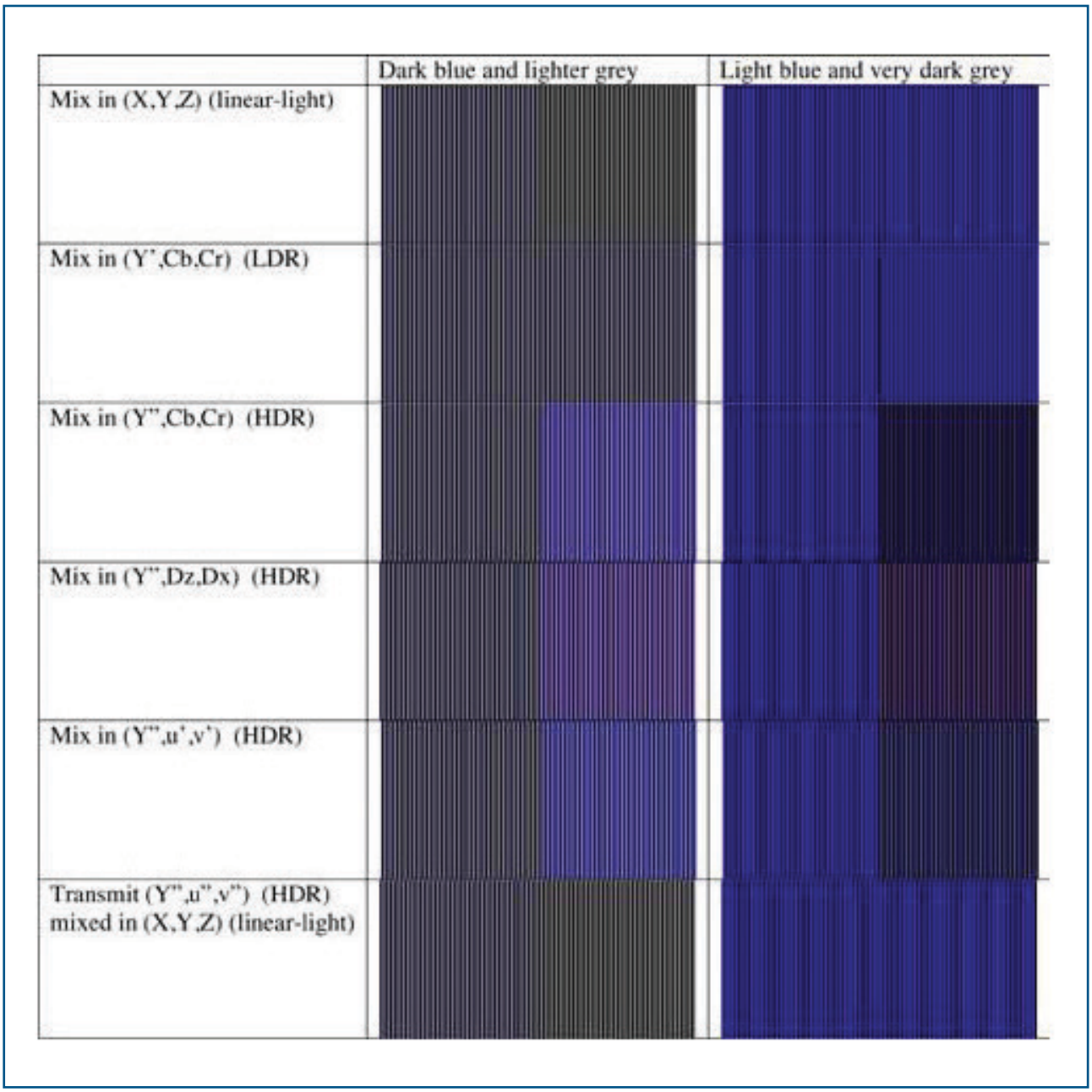


Figure 10 Color crosstalk examples

At the left is (Y'',C_B,C_R) 4:2:0 processed in HDR BT.709, in the middle is (Y'',C_B,C_R) 4:2:0 processed in the larger HDR BT.2020 space, and at the right is (Y'',u',v') 4:2:0, HDR too. The background is not black but a dark blue. As a consequence, we see at the bottom that with (Y'',C_B,C_R) all thin bright lines are polluted by the tint of the dark blue background. Also, some contrast is lost due to the familiar "constant luminance error," which is worse for HDR than for SDR.

At the top we see that for (Y'',C_B,C_R), the letters are less sharp than those on the right. This is because these colors are carried mainly by the lower bandwidth (C_B,C_R) signals. On the other hand, the (Y'',u',v') solution can decode almost 4:4:4 quality, because the sharpness is defined mostly by the higher bandwidth Y'' signal. This is also due to an optimum order of operations, i.e., an asymmetrical block diagram.

COLOR NOISE

Another strange effect happens if we convert saturated colors to (Y'',C_B,C_R) 4:2:0 and back to (R,G,B). If there is some color noise on the input signals then this noise is amplified significantly, but only for colors that approach 100% saturation relative to an RGB color space.

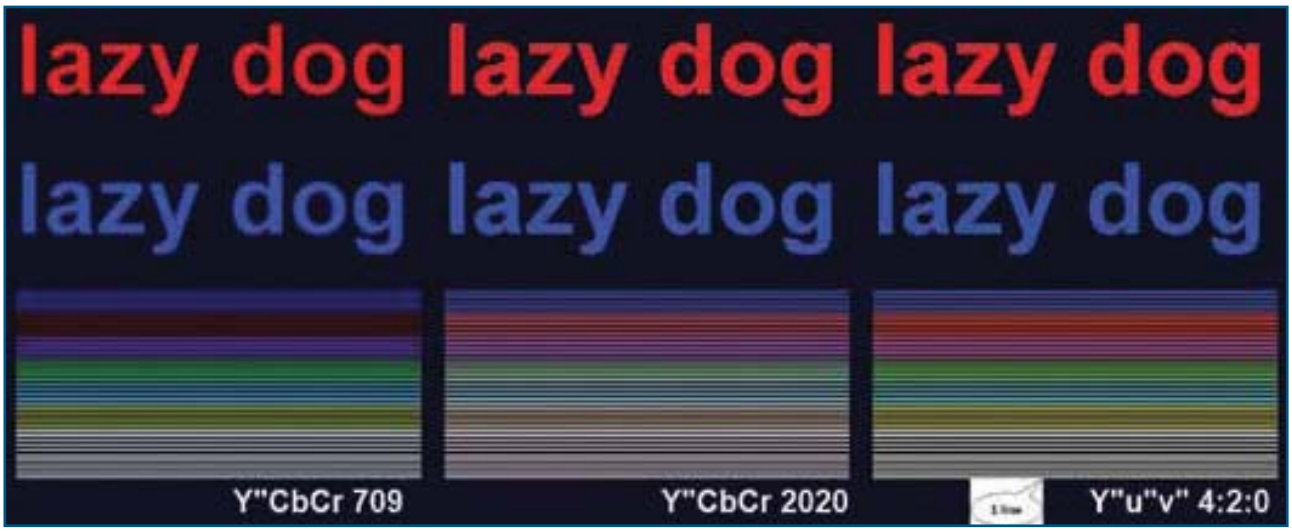


Figure 11 Another color crosstalk example

This is caused by the way that (Y'',C_B,C_R) values are calculated by means of three OECFs, starting with:

$$Y'' = 0.2627 \cdot \text{OECF(R)} + 0.6780 \cdot \text{OECF(G)} + 0.0593 \cdot \text{OECF(B)}^{6}$$

It is made worse by the HDR-type OECF. It is assumed that the transmitter OECF is the inverse of the receiver EOCF, so any noise reducing effects of color grading are not included in this reasoning. **Figure 12** shows an example of what happens (zoom in if necessary).

Top left: an original input image in BT.709 color gamut, with a bit of colored noise. In the left half of the image, the color saturation is reduced from top to bottom. In the right half of the image, the color saturation is 100%, and the brightness is reduced from top to bottom.

Top right: the same picture passed through (Y'',C_B,C_R) 4:2:0 with an HDR OECF-EOCF, and processed in BT.709 color space. The noise is clearly more visible (zoom in if necessary).

Bottom right: the same picture has been passed through (Y'',C_B,C_R) 4:2:0 with an HDR OECF-EOCF, but first it was converted to BT.2020 color space (standard for UHD) on the channel. In this color space, the numerical value of the color saturation is lower.

The noise is not amplified as much. Bottom left: the same picture passed through (Y'',u',v') 4:2:0 with an HDR OECF-EOCF, but now with the signals converted to (u',v') with "constant luminance." In this color space the numerical value of the color saturation is irrelevant. Now the noise is not amplified at all.

For the *saturation value* we use the definition of S from the (H,S,V) color space: S = 1.0 − MIN(R,G,B) / MAX(R,G,B). S = 1.0 if at least one of R,G,B is 0. This depends very much on the chosen RGB color space: the same perceived color has a lower *saturation value* in a larger color space. Likewise,

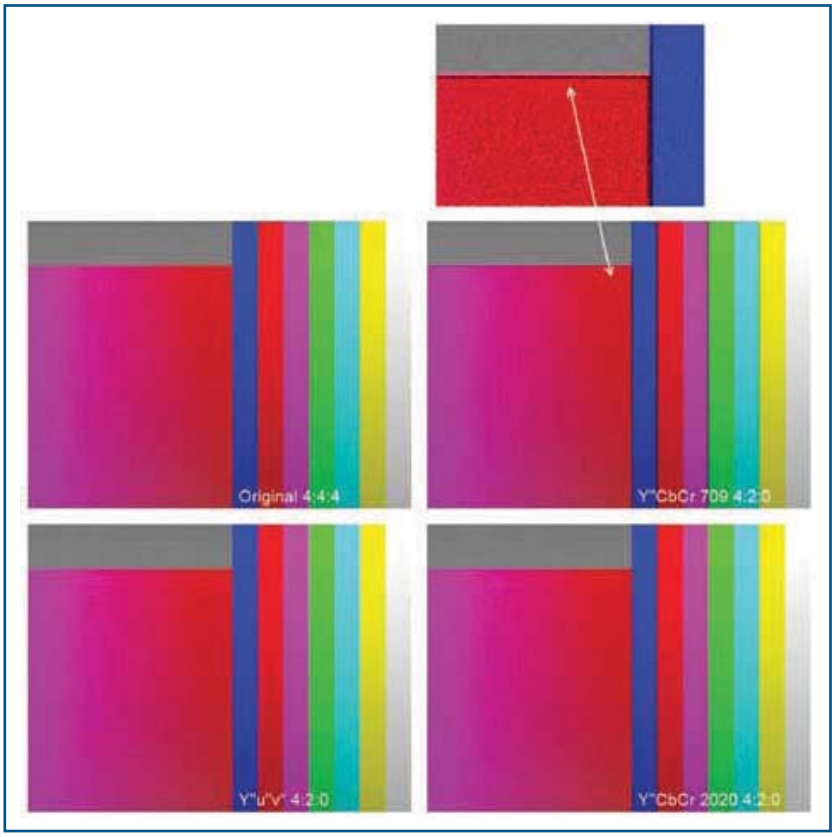


Figure 12 Color noise

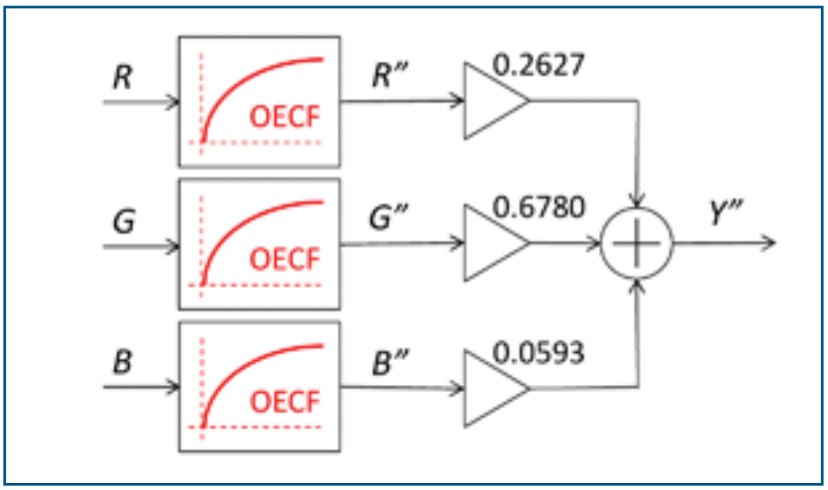


Figure 13 OECF in the Y″ calculation

MIN(R,G,B) is closer to 0 in a smaller RGB color space.

The explanation for the noise amplification has to do with the high *steepness* of the HDR OECF(x) for x close to 0, together with the fact that the OECF is applied to R,G,B separately **Fig. 13**.

The HDR OECF is very steep at the bottom, illustrated for the first 100 nits in **Fig. 14**.

Low values for (R,G,B) occur not only for black, but also for 100% saturated colors. Any noise on the near-zero signals is amplified before the signal contributes to the Y'' signal. The contribution to the (C_B,C_R) channels that might compensate for the noise in the receiver is attenuated by the 4:2:0 filtering. This explains why the noise won't go away after decoding.

If we first convert the (R,G,B) signals to a larger color space, e.g., BT.709 colors in a BT.2020 color space, then MIN(R,G,B) is no longer almost 0, and the steepness of the OECF in the path is much lower.

If we convert from (R,G,B) to (Y'',u',v'), then the OECF is applied only once to the Y path. Y is 0 only for black, but in that case, the receiver EOCF is very flat, and that will attenuate any noise.

So, can we avoid this noise problem of (Y'',C_B,C_R) when we use the larger BT.2020 color space where the saturation value is lower?

NO, not if we want to pass colors from the also larger DCI P3 color gamut. P3 red lies on the B=0 edge of the BT.2020 color space, so for saturated red, the noise on the B channel will still be amplified.

The same is true for the ACES color space and even the (X,Y,Z) color space. A (Y'',D_Z,D_X) signal would be just as problematic: For DCI P3 red, the Z value becomes 0, and the noise on the Z channel will be amplified. The anti-noise on a DZ channel would be attenuated by the 4:2:0 filtering, and the noise on Y'' remains. A (Y'',u',v') color space with constant luminance and no arbitrary edges close to actual colors does not have this problem.

U″,V″ INSTEAD OF U′,V′

In practice, we will transmit (u'',v'') instead of (u',v'). Below a certain luminance (approx. 5 nit) the (u',v') signals are attenuated towards gray in proportion to Y'' by the transmitter, and amplified back by the receiver. This sends less dark color noise to the MPEG codec and has an insignificant effect on the perceived accuracy of dark

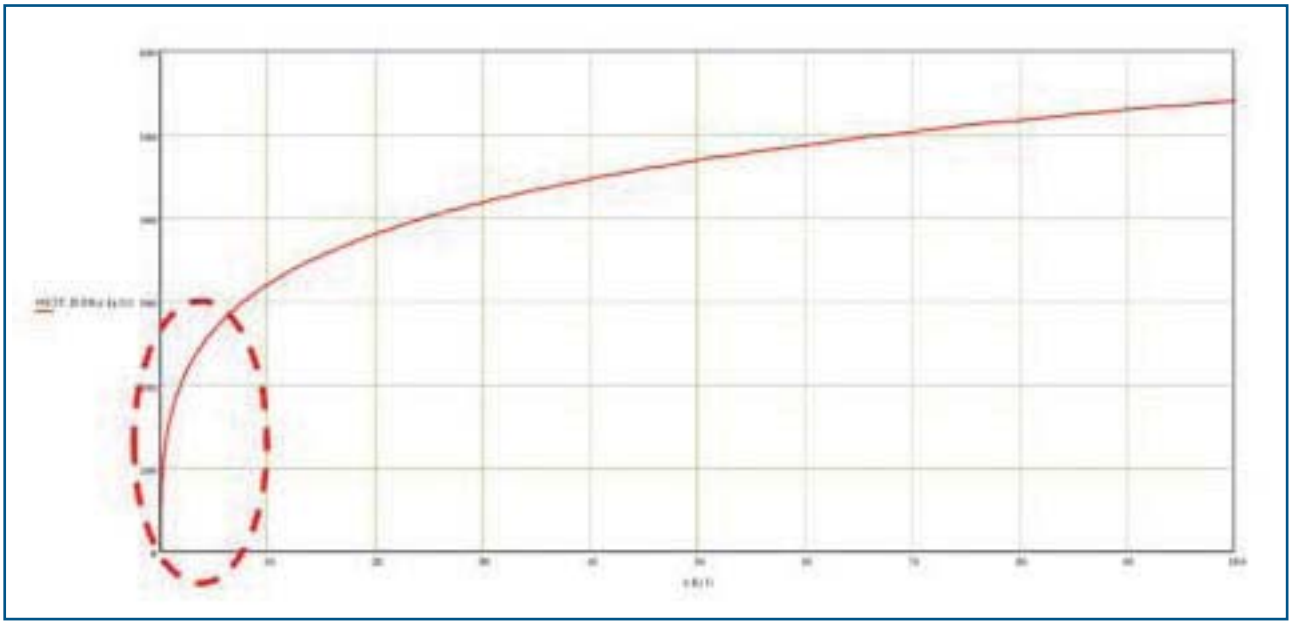


Figure 14 OECF steepness

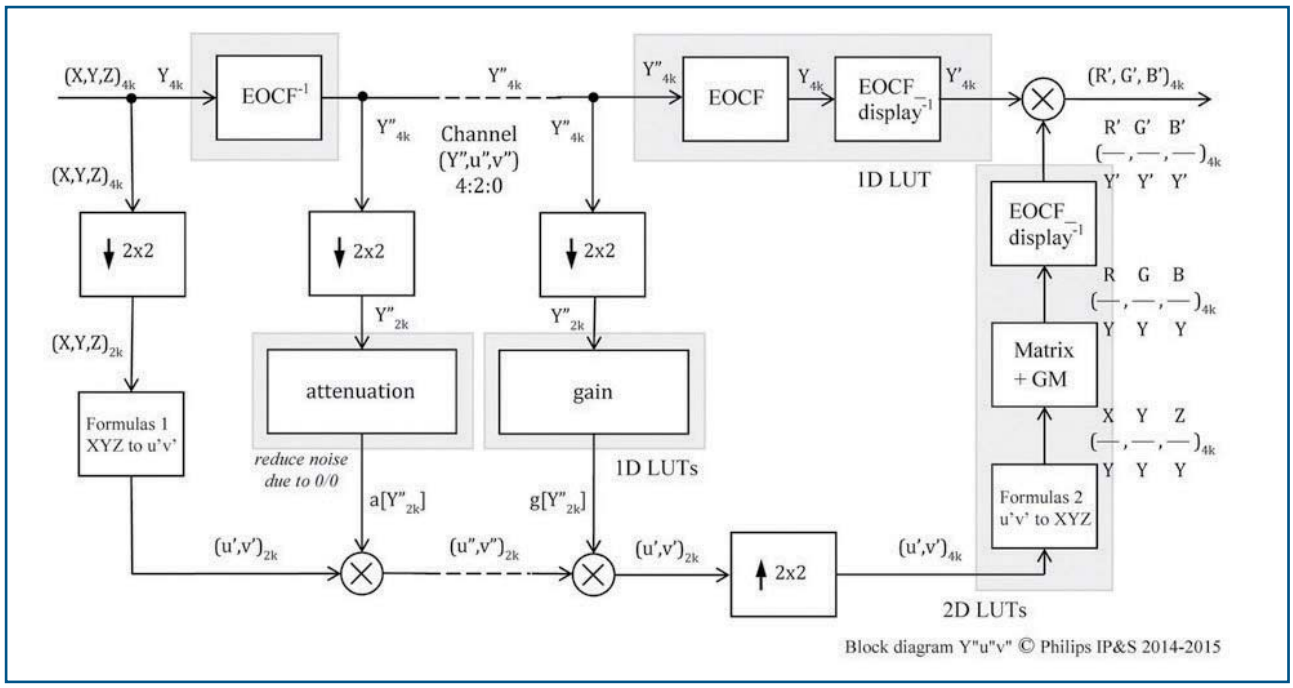


Figure 15 Block diagram of encoding and decoding

colors. Said in another way: in this region, (u'',v'') scale down with brightness just like 12-bit (C_B,C_R) signals do, so the accuracy is never worse.

BLOCK DIAGRAM

Encoding and decoding of (Y'',u'',v'') can be implemented in many variations; **Fig. 15** shows one possibility.

Fig. 13. OECF in the Y'' calculation.

The block diagram shows the important parts of the proposed (Y'',u'',v'') encoder and decoder. The transmitter on the left side does linear-light processing. The receiver on the right side does the processing first in chromaticity space and then in the gamma domain. The asymmetry is deliberate. Some advantages accrue to postponing the multiplication to restore (X,Y,Z); we perform this multiplication with Y' in the display space. If the display EOCF closely resembles a power function, the result is comparable to multiplying in linear-light space. If the display EOCF does not resemble a power function, another conversion function is required.

The Y path is one-dimensional and the part in the receiver can be implemented by a 1D lookup table or a linear interpolator. This could include tone mapping between various dynamic ranges. (NB: a good implementation of tone mapping requires more than just a Y signal.)

The (u',v') path is two-dimensional and the part in the receiver can be implemented by a 2D lookup table or a bi-linear interpolator. This should include color space conversion (from X,Y,Z to R,G,B primaries) and gamut mapping (from a larger color gamut to a color gamut as it is limited by the display). The data that defines the 2D mapping function could be defined by the transmitter and sent as metadata, not to be left to the discretion of the TV set-maker.

At the bottom the (u',v') signals are first up-sampled to full-resolution, with short and cheap filters. A higher quality variant exists, but the extra cost of that may not outweigh the benefits.

Next, the two (u',v') chromaticity signals are converted to three new $(X/Y, Y/Y, Z/Y)$ chromaticity signals by applying only a part of the textbook formulas. These signals do not require many bits yet, about 12 to 14 bits each. Then we can apply the usual color space conversion and gamut mapping to these signals and come out with $(R/Y, G/Y, B/Y)$ chromaticity signals. They have full resolution, but not full bandwidth. If desired, an EOCF_display-1 (= OECF) can be applied to $(R/Y, G/Y, B/Y)$, but only if its shape is an exact power law (like for example a square root).

The (u',v') path is two-dimensional and these three operations can be implemented in a 2D LUT or bi-linear interpolator, e.g., with 10 bits inputs and 3 x 14 bits outputs. This should include color space conversion (from X,Y,Z to R,G,B primaries) and gamut mapping (from a larger color gamut to a color gamut as it is limited by the display). The data that defines the 2D mapping function for some standard (BT.709) display could be defined by the transmitter and sent as metadata, not to be left to the discretion of the TV set-maker.

Last, we multiply $(R/Y, G/Y, B/Y)$ with the full-resolution Y signal and we get (R,G,B). Or, if all 4 signals

have been converted to the gamma domain, the multiplication $(R'/Y', G'/Y', B'/Y') * Y'$ can be implemented in the cheaper gamma domain (maximum 14 bits instead of minimum 24 bits). The output signals (R,G,B) or (R',G',B') can go to the display via the usual back-end processing and optional tone mapping to a different dynamic range.

The asymmetry between transmitter and receiver is deliberate. Some advantages accrue to postponing the multiplication to restore (X,Y,Z) or (R,G,B):

- The chromaticity signals at the bottom have no dynamic range, and they use fewer bits.
- The bottom part needs only a 2D lookup function, not the much more expensive 3D.
- The multiplication of luminance x chromaticity restores a lot of sharpness, this is a true constant luminance solution.
- We can perform the final multiplication with Y' in the cheaper gamma domain, meaning that we'll never need to process true linear-light signals (≥ 24 bits) in the receiver.

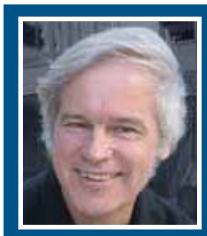
CONCLUSION

We conclude that (Y'',C_B,C_R) or (Y'',D_Z,D_X) is not optimum for HDR. (Y'',u'',v'') is suitable. Compared to other schemes that have been proposed, a (Y'',u'',v'') 4:2:0 transmission signal for HDR promises a lower bit-rate, better color reproduction at high spatial frequencies, less noise at 100% color saturation, and full coverage of the entire humanly visible color gamut. ■

References

1. C. Poynton, J. Stessen, and R. Nijland (2014), "Deploying Wide Colour Gamut and High Dynamic Range in HD and UHD" IBC 2014, Amsterdam, The Netherlands.
2. Scott Daly, et al. (2013), "Preference limits of the visual dynamic range for ultra high quality and aesthetic conveyance," Proc. SPIE, 8651:86510J, 2013.
3. Helge Seetzen, et al. (2004), "High Dynamic Range Display Systems," ACM Transactions on Graphics (Proc. SIGGRAPH '04), 23(3):760–768, 2004.
4. Charles Poynton, (2012). Digital Video and HD Algorithms and Interfaces, Elsevier/Morgan Kaufmann: Burlington, MA, 2012.
5. Scott Miller (2014), "A Perceptual EOTF for Extended Dynamic Range Imagery," ‹https://www.smpte.org/sites/default/files/2014-05-06-EOTFMiller-1-2-handout.pdf›.
6. Peter G.J. Barten (1999), "Contrast Sensitivity of the Human Eye and its Effects on Image Quality," PhD thesis (HV Press Knegsel), http://alexandria.tue.nl/extra2/9901043.pdf.
7. Peter G.J. Barten (2004), "Formula for the Contrast Sensitivity of the Human Eye," Proc. SPIE, 5294:231–238, 2004.
8. Robert D. Solomon, Color Coding For A Facsimile System, PhD diss., MIT, 1975.
9. Greg Ward Larson, [a.k.a. Greg Ward] (1998), "LogLuv Encoding for Full Gamut, High Dynamic Range Images," in J. Graphics Tools 3 (1): 15–31.
10. Charles Poynton and Brian Funt (2014, Feb.), "Perceptual Uniformity in Digital Image Representation and Display," Color Research and Application, 39(1):6–15, Feb. 2014.
11. CIE S 014-5 (2004), Colorimetry, Part 5, CIE 1976 $L^*u^*v^*$ Colour Space and u', v' Uniform Chromaticity Space. Also published as ISO 11664-5:2009.

THE AUTHORS



Charles Poynton is an independent contractor specializing in the physics, mathematics, and engineering of digital colour imaging systems, including digital video, HD/UHD, and digital cinema (D-cinema). He does technology forecasting, systems modelling, algorithm development, video signal processing architecture, color characterization and calibration, and image quality assessment. He lives and works in Toronto, Canada.

Jeroen Stessen was born in 1962, and became interested in electronics around the age of 8. In 1987 he graduated from Eindhoven Technical University with an M.Sc. in electronics, and a specialization in control systems. He then began working in the television laboratory of Philips Consumer Electronics. During more than 25 years with Philips he has developed analog and digital electronics and various image processing algorithms. In 2012, he was transferred along with the television business to TP Vision. In 2014, he began consulting for Philips Intellectual Properties and Standards, in a project about High Dynamic Range TV. In this role, he is developing solutions for HDR transmission and for HDR SDR conversion. He has put his name on approximately 10 external publications and 40 patent applications. He lives in Eindhoven, The Netherlands, with his wife and 3 children.

Rutger Nijland graduated from the Polytechnic College in Enschede in 1999. That same year, he joined the Philips Advanced Systems and Application Labs Eindhoven, The Netherlands, where he worked in the Flat television department on digital video processing and matrix display systems. He was involved in a wide range of projects in the area of plasma and LCD TV development, ranging from algorithm development to definition, implementation, and application of integrated circuits. Currently, he is a research scientist involved in high dynamic range video encoding and tone mapping. Nijland has been involved in video research activities including sharpness, scaling, contrast manipulation, and color theory.