The dynamics of digital colors: Part two

BY MARCUS BENGTSSON

IN PART ONE, we considered what color is color representation, RGB and Y'CbCr color models, and bit depths. Now we move on to color spaces and dynamic range, along with various ways of transferring or storing data representing a wide dynamic range into a reasonably sized message or file format.

Color spaces

The term *color space* is somewhat ambiguous and is sometimes incorrectly used to describe color models. For example, there is no such thing as a single "Y'CbCr color space," although there are plenty of color spaces that make use of an Y'CbCr color model.

Examples of color spaces are Rec. 2020[1], DCI-P3, Rec. 709[2], and sRGB. It's worth noting at this point that Rec. 2020 and Rec. 709 are international recommendations for television (hence "Rec.") that indeed define color gamuts and color models (and thus define a color space), but both standards go beyond that and also define bit depths, frame rates, and resolutions, topics that are beyond the scope of the sRGB standard.

Part of what makes up a color space is the combination of a color model and a color gamut. As previously discussed, the color model defines a mathematical model that describes how colors are represented with numbers. However, the numbers defined by the color model are not meaningful unless mapped to an underlying color gamut and color space. If I say that I am traveling down the freeway at 100 km/h, that number only makes sense if we have clearly defined the length of a kilometer, and the duration of an hour. Without those definitions, the information provided about our velocity is meaningless.

Similarly, without having a clearly defined relationship (mapping) between the color model, the available range of colors (the color gamut), and an absolute color space, the end result is completely arbitrary and there is no guarantee that a given set of RGB values will ever look the same across different display devices.

Arguably, the best way to visualize a color gamut and its relationship to a given color space is through a chromaticity diagram, such as the **Figure 1** CIE 1931 chromaticity diagram. The outermost boundary is called the *spectral locus* (or monochromatic locus), with wavelengths shown in nanometers,

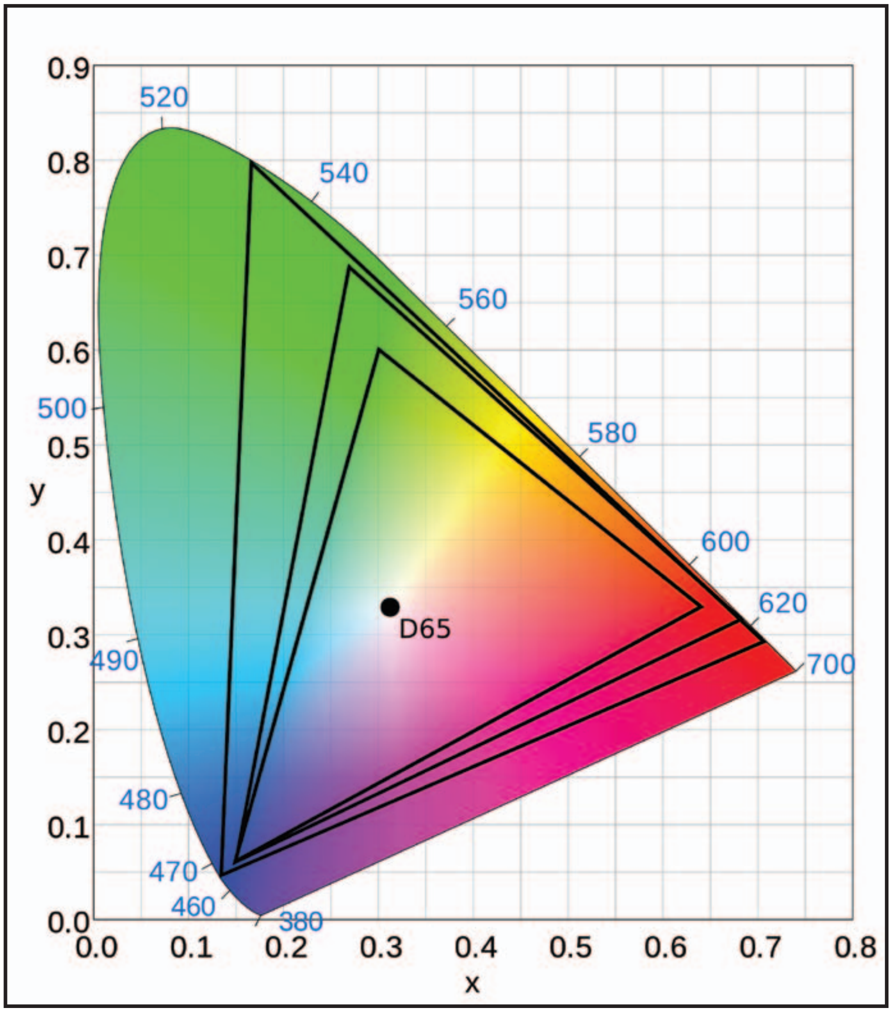

Figure 1 – CIE 1931 chromaticity diagram

depicting the chromaticity visible to an average human eye. The three triangles inside the spectral locus are three different color gamuts; the outer triangle is the color gamut used in Rec. 2020, the center triangle represents DCI-P3, and the innermost triangle is the color gamut used in both Rec. 709 and sRGB.

The available color gamut is then defined by the x/y coordinates of the chromaticity diagram. For the various color spaces, the following coordinates are used:

Rec. 2020
Red: 0.708, 0.292
Green: 0.170, 0.797
Blue: 0.131, 0.046

DCI-P3
Red: 0.680, 0.320
Green: 0.265, 0.690
Blue: 0.150, 0.060

Rec. 709/sRGB
Red: 0.640, 0.330
Green: 0.300, 0.600
Blue: 0.150, 0.060

As we can see on both the chromaticity diagram and the corresponding numbers, Rec. 2020 and DCI-P3 offer a significantly wider color gamut with more saturated primaries compared to Rec. 709 and sRGB. To enable this wider gamut to be efficiently used, Rec. 2020 specifies a minimum of 10-bit color depth and allows up to 12 bits. Rec. 709 and sRGB on the other hand, require a minimum of 8 bits but allow for 10-bit color depths to be used. It's important to point out that the absolute colors (monochromatic wavelengths) of the primaries does not change depending on the bit depth, only the number of steps or variations between the primaries change.

White point

Aside from defining the chromaticity of the primaries, a color space also needs to define the white point, or reference white. The most commonly used reference white is referred to as D65. The coordinates for D65 is the same for both Rec. 2020 and Rec. 709/sRGB at $x = 0.3127$, $y = 0.329$ which equates to a white balance of 6,504 K, close to daylight.

Digital representation

Exactly how x/y coordinates are converted into digital RGB or Y'CbCr values and vice versa involves complicated math beyond the scope of this article. It is, however, important to know that the conversion from x/y coordinates to RGB values is not linear. This stems from the fact that human vision is highly non-linear, almost logarithmic. Changes at low intensity are perceived as much more significant to us than changes at high intensity. Because of this, a light source attenuated to about 22% of its original intensity is perceived as half as bright. The value 255 means full intensity on an 8-bit system; the value 128 (50% of 255) will not be perceived as 50% as bright as 255, but approximately 78% as bright. This mathematical function is generally called an *Electro-Optical Transfer Function*, or EOTF for short. The EOTF describes how the display converts from data to screen brightness. Conversely, an *Opto-Electrical Transfer Function* (OETF) describes how a camera sensor converts scene brightness into data. When these two transfer functions work together in an end-to-end content acquisition, transmission and delivery system, the result is called an *Opto-Optical Transfer Function*, or OOTF.

Extended range or video range

As we've previously discussed, the minimum value for one byte (8 bits) is 0 and the maximum is 255. 0 represents pure black and 255 represents pure white. But that is not always the case. Rec. 709 incorporates a concept called legal levels, also referred to as video range or "studio swing." This refers to the range the digital value has to stay within to be considered standard compliant. In legal levels, video range or studio-swing, black is placed at value 16 and reference white is at 235 (the range is 64 – 940 when using 10-bit color depth). Values from 0 – 16 are called super black and from 236 – 255 are called super white. These values are considered outside of the nominal range and should not be used. This seemingly arbitrary limitation is a remnant from a time when analog video signals were converted into digital signals, a detailed explanation of which is considered outside the scope of this article. This only refers to the digital levels. Reference white is still 100% white and reference black is still 100% black, regardless of which range is being used.

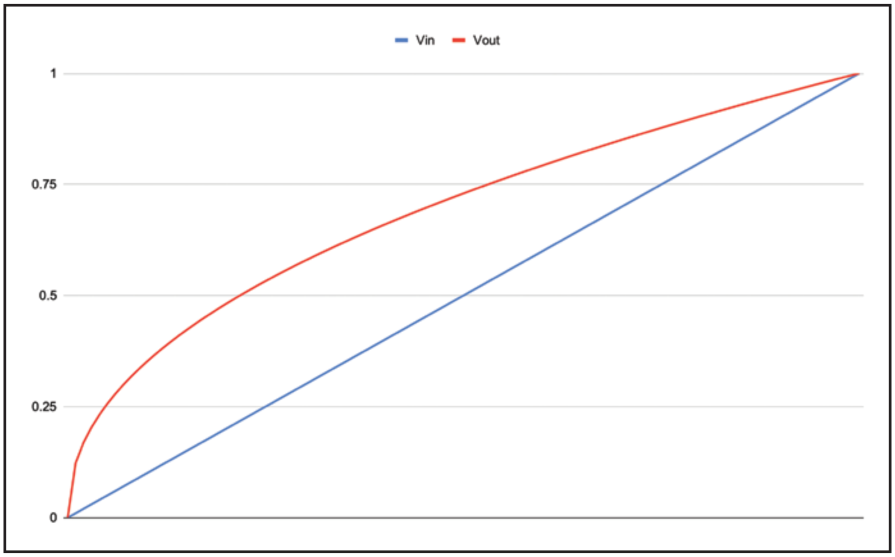

Figure 2 – Linear input vs. non-linear output with a gamma of 2.2

Figure 3 – Left side: SDR exposure, right side: HDR exposure

Dynamic range and HDR

The term HDR is short for *High Dynamic Range*, which simply put is a term that describes images or videos with deeper blacks and brighter whites. HDR provides significant and tangible improvements over SDR, or *Standard Dynamic Range*. But what does it mean and how does it work? To understand that, we need to dissect the term "High Dynamic Range." Let's start with the concept dynamic range.

Dynamic range is the ratio between the brightest color (whitest white) and the darkest color (blackest black) in an image. For example, a printed image on a piece of white paper has an approximate dynamic range of 100:1. The black ink will reflect at least 1% of the incoming light.

Dynamic range is usually measured in so called *stops*. Explaining in detail how stops work is beyond the scope of this article, so let's simplify it to define one stop as double or half. One stop more means double the light; one stop less means half the light. So when describing available dynamic range with stops, we measure how many stops our image (or display device) contains between its brightest and darkest sections as a base 2 logarithmic function of the contrast ratio. The printed image we previously used as an example thus has a dynamic range of $\log2(100) \approx 7$ stops. A standard computer monitor or TV may have somewhere between six and 10 stops of dynamic range and a professional DSLR up to 14 stops. Some top of the line digital cinema cameras can brag about a dynamic range of more than 16 stops. For reference, a healthy human eye has a dynamic range of up to 14 stops.

In essence, this means that when using an HDR pipeline, we can capture and display content that has a larger contrast ratio than we can see. On the **Figure 3** St. Louis Arch image, the left side is a SDR exposure with around eight steps of dynamic range. The right side is a simulated HDR exposure with approximately 14 steps of dynamic range. With the extended dynamic range we are able to maintain detail in the brightest parts of the image while also adding more detail in the darkest sections.

To effectively make use of HDR, the entire signal chain and pipeline must be created with HDR in mind. If you have a camera with a very wide dynamic range, that camera needs to be configured to capture imagery in a format that can later on be processed as HDR content so that it can be displayed on an HDR compatible display. When recording HDR content, it is common to use Log-Format Recording or Log Video to maximize the camera sensor.

Log-Format Recording

Traditional digital sensors used in compact cameras, DSLRs, camcorders, and cell phones work in a linear fashion. If the intensity of the light that hits a pixel on the sensor increases, the voltage that is being outputted for that particular pixel also increases. This value is read, converted into a digital exposure value, and stored somewhere. In most cases, this does not pose any problems, as the dynamic range of the sensor is too narrow for the limitations of this process to become apparent. However, with the introduction of wider dynamic range camera sensors that can record up to 16 stops, recording incoming light linearly presents some issues. Cameras measure exposure in f-stops, and we've already concluded that a stop is a doubling or halving of light intensity. If we count backwards from our brightest white (255 on an 8-bit system), this represents the first stop, taking up 50% of the available data (values 128 – 255). The second stop gets 25% of data (values 64 – 127); the third stop 12.5% of data, and so on. When we're down to our sixth stop, we have less than 1% of our available data range left, which makes it next to impossible to extract any shadow detail.

To combat this, camera manufacturers have introduced different log curves to better use the dynamic range of each camera sensor. What the log curve does is apply a logarithmic transfer function to the A/D conversion that happens after the camera sensor, so each stop in the camera's dynamic range gets allocated roughly equal amounts of data. The end result is a washed out image without much contrast, but this is something we easily fix by applying an HDR transfer function or LUT. In this context a LUT, or Lookup Table, is usually a file or a predefined software function that statically maps one range of values to another range of values. For a given input color, applying a LUT treatment will produce a known output color.

> . . . wider color gamuts . . . and HDR workflows are some of the most exciting technical innovations that have emerged in the last couple of years.

HDR standards and transfer functions

There is one international HDR standard (Rec. 2100) and several open or proprietary distribution and transfer systems, of which the four most common are:

- Hybrid Log-Gamma
- Perceptual Quantizer
- HDR10 (and HDR10+)
- Dolby Vision

Hybrid-Log Gamma

Hybrid-Log Gamma, or "HLG" for short, is a standard that was developed by BBC and NHK mainly for broadcast purposes. As the name suggests, it defines a hybrid non-linear transfer function in which a gamma curve is used for the lower half of the signal values and a logarithmic curve for the upper half. The reason for this, as with most things broadcast, is backwards compatibility. An HLG encoded signal can be displayed on an SDR screen and does not require any exchange of metadata, which makes it suitable for live event production.

Perceptual Quantizer

Perceptual Quantizer (PQ) is an HDR standard and Electro-Optical Transfer Function curve that was defined by Dolby Laboratories and later standardized as *SMPTE ST 2084*. PQ differs significantly from HLG as it was not designed for broadcast, but for cinema. Thus, there was no need to maintain backwards compatibility; instead the standard is designed purely to maximize image quality based on the characteristics of human vision. PQ is unique in that it is a so-called display-referred EOTF, meaning it encodes absolute luminance values for each pixel based on the capabilities of the device, not just relative brightness.

HDR10/HDR10+

HDR10 and is an open HDR standard created by Consumer Technology Association in 2015. The standard uses PQ as its transfer function and defines the minimum requirements that a device must meet to qualify as HDR10 capable: Rec. 2020 color space and a 10-bit color depth. The device must also be capable of transmitting and decoding static metadata used to define characteristics for the whole video. HDR10+ was introduced in 2017 and improves upon HDR10 by increased color depth to 12 bits, and it includes scene-based metadata that allows for brightness and color adjustments to be made to the display on a scene-by-scene or frame-by-frame basis. HDR10+ also supports luminance levels of up to 10,000 nits whereas HDR10 was limited to 4,000 nits.

Both HDR10 and HDR10+ are supported by a wide range of manufacturers and content producers, but the standards are mainly used for home entertainment devices and online streaming services.

Dolby Vision

The main competitor to HDR10+ is Dolby Vision. From a technical point of view they are almost identical; devices have to conform to Rec. 2020 color space, use the PQ transfer function, display a minimum color depth of 12 bits and support scene or frame based dynamic metadata. It also supports the same luminance levels as HDR10+, up to 10,000 nits. The main difference is that Dolby Vision is a proprietary standard developed and maintained by Dolby Laboratories. As such, any manufacturer or content provider who wishes to integrate Dolby Vision must pay royalty fees and undergo strict quality assurance and calibration procedures to become Dolby Vision Certified.

Rec. 2100

Finally, Rec. 2100, or "ITU-R Recommendation BT.2100" as it is officially called, is an international standard that specifies several aspects of HDR video such as display resolution, frame rate, chroma subsampling, bit depth, color space, and transfer functions. It uses the same color space as Rec. 2020 with a color depth of 10 or 12 bits for resolutions up to 8K and 120p refresh rates. Both HLG and PQ optical transfer functions are allowed.

Conclusion

The aforementioned standards and recommendations are mostly focused around cinema or home entertainment where HDR is now in available in every living room, every streaming service, and even in most handheld devices. No, that is not a typo. If your smartphone is less than five-years old, there's a good chance it is HDR capable. In the world of live entertainment, we are for once a few years behind the rest of the world. Only very recently have we seen professional media servers that can playback video with wide color gamuts, 10-bit color depths, apply color profiles in real time, and work with professional HDR standards. Recent innovations in laser projection is also coming close to reaching full Rec. 2020 color gamut support, but LED display technology is still lagging in that regard, with few if any production-ready products capable of accurately displaying the full color gamut required for Rec. 2020. However, with these technologies slowly but surely making their way into live shows, it remains to be seen if the existing standards and recommendations are applicable to live entertainment, or if we need to develop a whole new set of standards specific to our industry.

Some obvious challenges are bandwidth and infrastructure: SDI is not able to transport any kind of HDR metadata. Both DisplayPort 1.4 and HDMI 2.1 can transport dynamic metadata, but transporting and maintaining signal integrity over long distances will pose an entirely new set of challenges. In addition, there are no (to my knowledge) professional DisplayPort 1.4 or HDMI 2.1 compatible routers available—something that is absolutely critical in a live show scenario where redundancy and automatic failover systems are necessities. So what about Video over IP? Well, *SMPTE ST 2110* supports HDR using both PQ and HLG, but mainstream adoption still looks to be a few years away. Content delivery is another challenge; the codecs we typically use for delivery of video files today do not incorporate any of the aforementioned HDR standards. Image formats such as OpenEXR do what we need, but lack the necessary manufacturer support, not to mention the inherent challenges that come with using image sequences.

Nevertheless, it is my opinion that wider color gamuts, higher than 8-bit color depths, and HDR workflows are some of the most exciting technical innovations that have emerged in the last couple of years. How can we use them to enhance the next generation of live shows? ■

Marcus Bengtsson is a Technical Solutions Specialist for disguise, based in sunny Los Angeles. Marcus holds a degree in Software Engineering and Management and is a published author, award-winning engineer, inventor of several patents, and challenger of the status quo. When he is not trying to solve impossible problems, Marcus enjoys rock climbing, mountain biking, scuba diving, good food, and good wine. He can be reached at marcus.bengtsson@disguise.one .

Endnotes

1 Officially "ITU-R Recommendation BT.2020"

2 Officially "ITU-R Recommendation BT.709"