A Unified Approach to High Dynamic Range for Broadcast and Multiscreen Services

ABSTRACT

High dynamic range (HDR) and wide color gamut (WCG) are increasingly seen as key features of next-generation broadcast and multiscreen services. This white paper describes the technical features of an end-to-end approach that has been deployed to enable HDR and WCG from original production to delivery via broadcast and streaming services, and on to displays on consumer devices. The perceptual quantizer (PQ) transfer curve is introduced as the foundation for the approach, enabling content production to the full fidelity required to meet the expectations of consumers. Considerations for delivery in various key scenarios are also discussed, including over-the-top (OTT), simulcast, and backward-compatible broadcast models. This paper makes the case that a universal Dolby Vision™ decoder enables support for a wide range of broadcast and OTT use cases, and can make use of additional metadata to maintain creative intent across screens with diverse capabilities.

DOLBY
VISION

DEFINITIONS AND BENEFITS OF HIGH DYNAMIC RANGE AND WIDE COLOR GAMUT

HDR and WCG are mechanisms by which more of the contrast and color of the real world can be brought into the electronic-imaging experience. From bright glints of sunlight reflecting off shiny surfaces to all of the rich detail inside of the shadows—these are the things that we see around us every day but that have always been missing from television and cinema. These details may not sound significant at first, but when we see images with these very important visual cues restored, the results are astounding. Together, HDR and WCG bring viewers to a new level of realism and immerse them into an enriched viewing experience that has never before been possible.

On the production side, these new technologies give the creative minds that craft our programs unprecedented freedom to deliver the vision they desire. For live productions, a single shot often contains both brightly lit and deeply shaded material. In the past, difficult decisions had to be made on the fly about which items to preserve and which to crush into black or blow out into white. With HDR, the entire shot can be captured without compromise, not only producing a better end product but requiring less effort to do so. For non-live productions, the range of colors and brightness levels available for artists to choose from is much larger than ever before, removing the creative constraints that have held artists back for so long from achieving their true inspirations.

THE CRT AND THE GAMMA LEGACY

later...

Though we live in an HDR/WCG world, video signals are still constrained to the characteristics of legacy cathode ray tube (CRT) technology. Though CRTs could make fairly dark blacks, their top end was generally limited to about 100 cd/m^2 (candelas per square meter, or more commonly, 100 nits). Likewise, the color primaries for current HDTV (as defined in ITU-R Recommendation BT.709) are limited to the colors that could be created using practical phosphors on the face of the CRTs of the day. These limitations are like viewing the world through a 100-nit window, with color filters that also knock out very saturated colors.

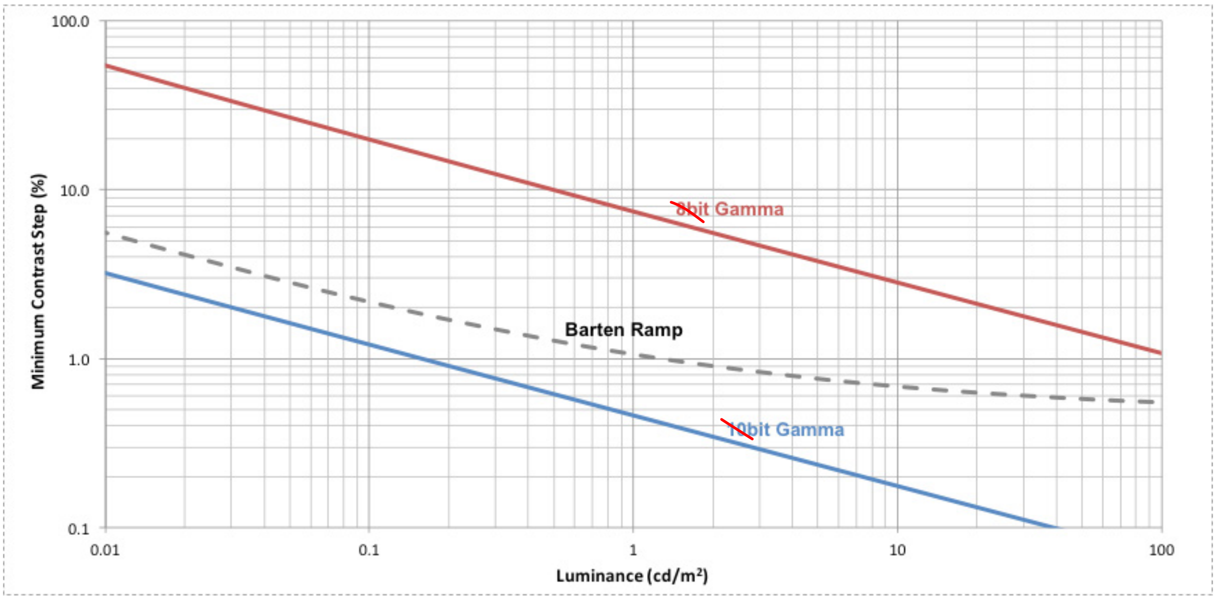
One positive side of these limits for CRTs is that the displays could be calibrated to match each other reasonably well. Consistent images are of paramount importance when it comes to production and distribution workflows. Knowing that the images created in the mobile broadcast truck match those seen in master control and that they match the images seen at the local stations is essential. And so is knowing that the images seen in the color-grading suite are precisely the same as those seen on the client's approval display and that they are the same as the images seen in the service provider's control room.

Another positive of the CRT legacy is that this display technology introduced the concept of nonlinear signal encoding to video. The relationship between input voltage

DOLBY
VISION

and output light of a CRT is not linear; it is a power function with an exponent of 2.4. Power functions are often expressed as $y = x^{gamma}$, so this function has become commonly known as a gamma curve. In the case of television, gamma = 2.4.

As it turns out, the human visual system has a sensitivity curve that very roughly matches the CRT's gamma 2.4 characteristic at low luminance levels. The graph below shows the Barten model of human vision compared with a 100 cd/m^2 gamma 2.4 response, representative of reference CRT displays.

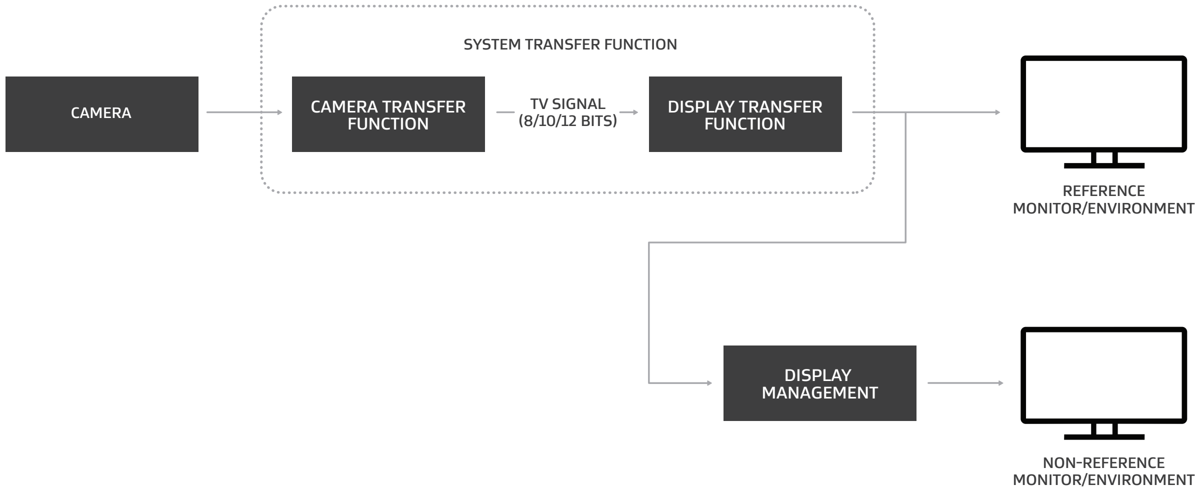


This similarity to human vision allows for efficient use of signal voltage levels in an analog system, and consequently efficient use of bit depth in a digital system. This is the fundamental equation of electronic imaging: the function that describes how to turn data into light. It is known as the electro-optical transfer function (EOTF), or more simply as the display-transfer function. It is this function that creates practical system efficiency and ensures consistent reproduction anywhere and everywhere.

later...

One of the large points of confusion regarding video systems is the fact that even though the display-transfer function is fundamental, it was never rigorously specified until 2011 as ITU-R Recommendation BT.1886. How could this be? Because of the dominance of a select few CRT manufacturers and the tight control of their reference-grade displays, there was effectively no need for a standardized specification. It was only as these manufacturers began to discontinue their reference CRTs that the industry saw the clear need to document the "standard" reference display-transfer function before all of the legacy CRTs were gone.

To add to the confusion, most video system standards actually specify an opto-electronic transfer function (OETF) instead. This can be more simply referred to as the camera-transfer function. The camera-transfer function and the display-transfer function are not simply inverses of each other. Together they create what is called an opto-optical transfer function (OOTF), which is the mapping from scene light to displayed light. This function may be called the system-transfer function, and it involves the adjustment of images captured under typically bright lighting conditions to adapt them to look correct to human eyes in the much dimmer display environment.

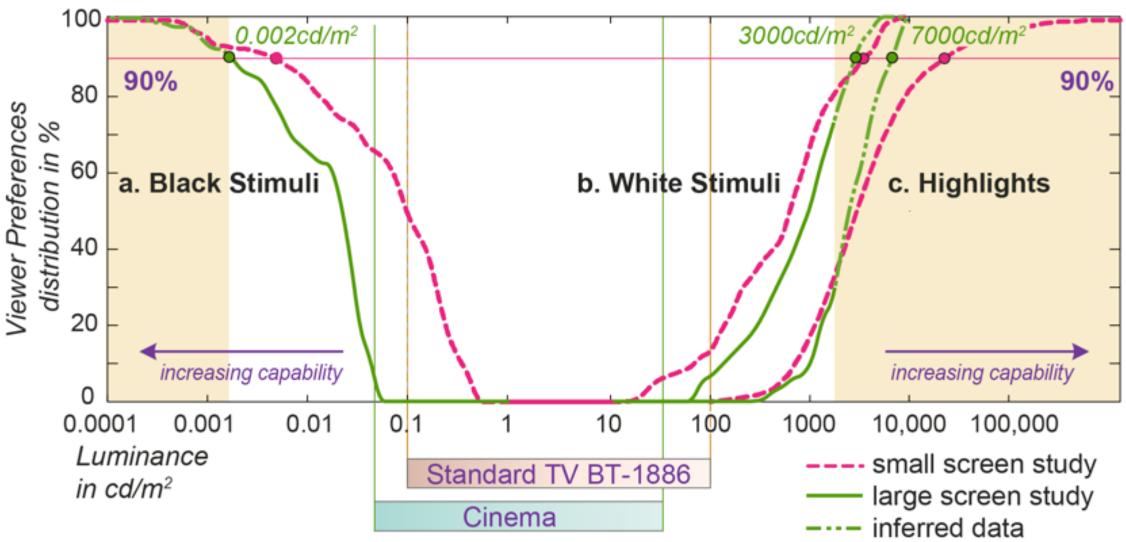


Specifying any two of the display-, camera-, or system-transfer functions will determine the third. But it should be obvious that the one that is essential for any type of program is the display-transfer function. During production, all reference displays must show exactly the same images. Otherwise, conflicting adjustments may be made, and the results will be unpredictable. Many non-live productions are shot with cameras using multiple transfer functions, including raw capture, so the concept of a single reference camera-transfer function makes no sense for these types of productions. Even in live production, once the program has been created and is sent for distribution or archiving, the camera-transfer function and system-transfer function no longer matter. Ultimately, all creative decisions from complex color grading to simple iris control are performed while judging the image on a reference display. Clearly, the reference display and its transfer function are the foundation upon which a practical video system must be built.

THE PQ TRANSFER FUNCTION FOR HDR

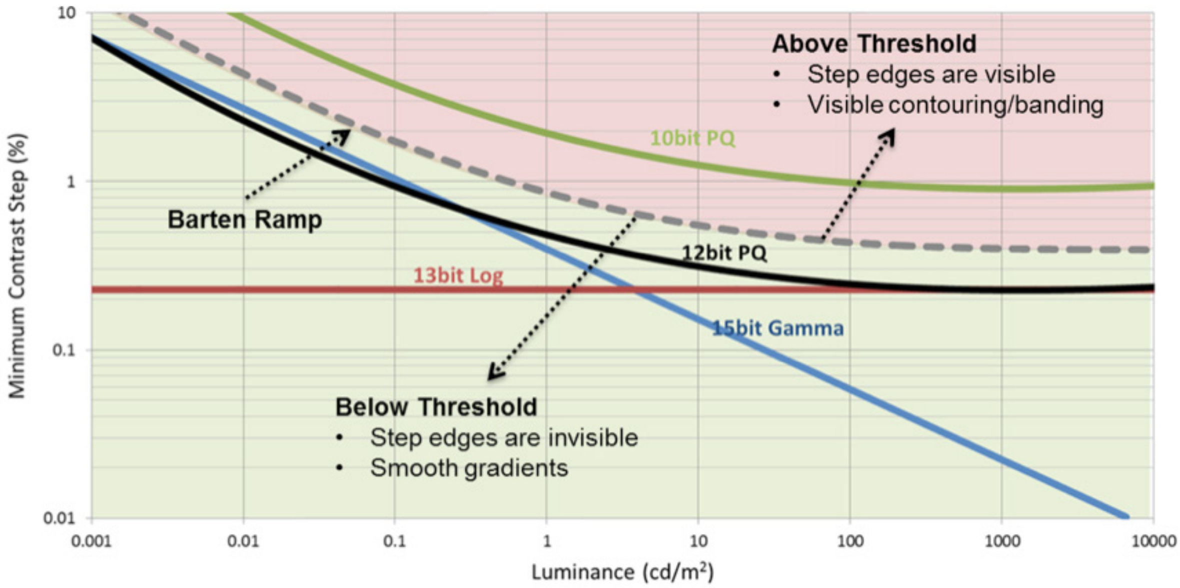
Based on the previous discussion, it was clear that to build a new video system for HDR, our team of engineers at Dolby had to start with the reference display-transfer function or EOTF. But how much dynamic range should be covered by the reference display? We did not want to limit this new system to current technology limitations but instead devise a future-ready system that various industries could grow into at their own pace. So we chose to optimize the new transfer function to the human visual system. The human eye has remarkable range, with the threshold of vision being about 10^{-6} cd/m^2, and the sun roughly 10^9 cd/m^2. This is a potential visual dynamic range of 15 orders of magnitude (or 50 stops in photographic terms). But reproducing eye-damaging levels of light is not very useful in a practical system, so we needed to first figure out what kind of range viewers would prefer to see if we removed the technology constraints.

A series of visual experiments were conducted with laboratory displays that could generate images with high precision anywhere between thousandths of a cd/m^2 and 20,000 cd/m^2. Viewers examined three sets of images, and they were to pick their preferences for black level, white level, and highlight level, respectively. The images were carefully chosen to avoid biases that have been a stumbling block to many previous experiments of this type. Results of the experiments are summarized in the graph below.



In the end, it was determined that a display range between 0.005 and 10,000 cd/m^2 would satisfy most viewers for a wide variety of images on both large and small screens. This is a much more manageable dynamic range of just over 6 orders of magnitude, or 20 stops.

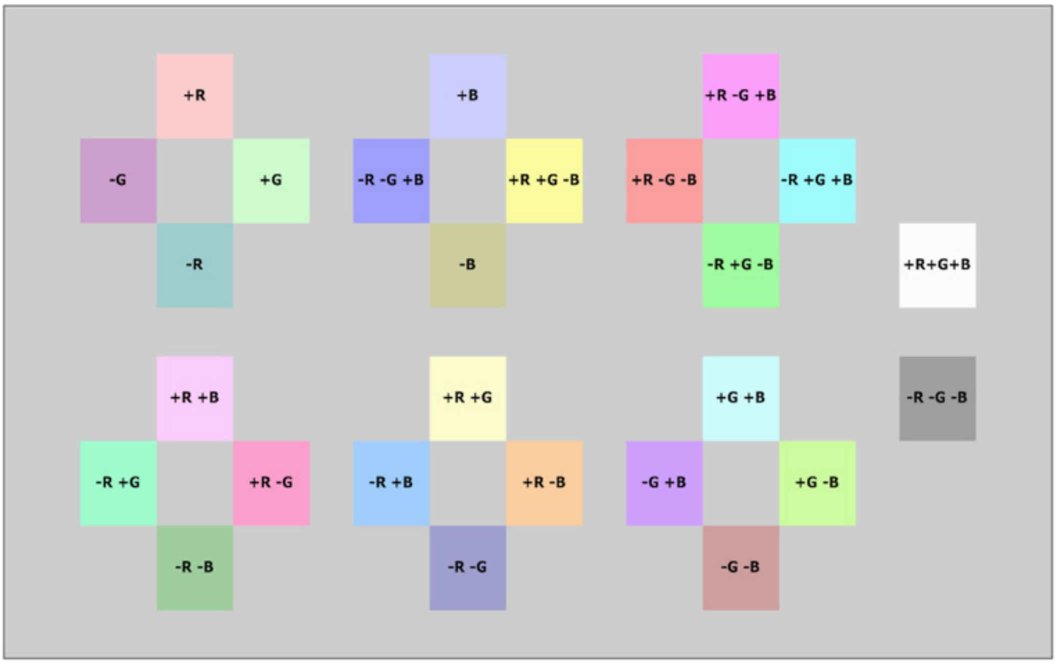
Using the same Barten model for human vision discussed earlier, we can see the response of the human eye over this extended dynamic range, and more importantly, build a new nonlinear transfer function that follows it. The resulting display-transfer function has come to be known as the perceptual quantizer (PQ) and was standardized by SMPTE in ST 2084 in 2014 and more recently by the ITU-R in Recommendation BT.2100. The graph below shows that PQ has enough precision to stay below the visual threshold across this entire extended range with only 12 bits, where traditional gamma encoding would have required 15 bits to stay completely below the Barten Ramp curve.



This is a good match for existing cinema infrastructures, and many video-on-demand systems are capable of supplying 12-bit signals as well. Even though 12 bits are required for completely imperceptible steps with noise-free images, the 10-bit PQ curve is only slightly above the Barten Ramp line. In practical systems that include low levels of noise, including sensor noise from cameras, 10-bit PQ has proven itself to be a good match.

Many other experiments have also been conducted over the last few years to confirm the lack of color contouring artifacts using PQ. These tests used many subjective images as well as objective test charts like the JND Cross pattern shown below. All of these tests, conducted in both RGB and YCC domains, have confirmed PQ's ability to represent signal values without any perceptual differences using 12 bits, and with minimal perceptual differences (always masked by noise in normal system operation) using 10 bits.

DOLBY
VISION



DISPLAY MANAGEMENT

Hollywood and television directors, live-broadcast producers, and game producers all want to be sure that viewers see what they intended. Dolby has worked with Hollywood and episodic directors, and is working with the broadcast and game industries, to ensure that creative intent can be consistently delivered to consumer televisions and other devices.

To ensure the best possible end result for content like movies, episodic content, and games, it is of prime importance to ensure that the signal be as high a quality as it can, accurately and efficiently representing the full dynamic range and color gamut of the camera or rendering system, and that metadata flows through the system describing both the production environment and the content characteristics. Live and broadcast ecosystems have different requirements that can make it difficult to carry metadata, but representing the full dynamic range and colors captured by the camera without compromise is still critical.

This full-range signal is encoded and transmitted to the consumer television or other display and decoded. Each display can be very different, from tablets to large flat-screen televisions, to dedicated home theater systems. Dolby Vision includes technology in the display that takes the full-range, uncompromised signal and maps the signal to the capabilities of that display through a process called display management.

For offline and rendered environments, Dolby® has, with industry professionals, defined a set of metadata standards. These standards represent both static information about the display the content was mastered on (SMPTE ST 2086) and dynamic information about the content itself (SMPTE ST 2094). Display management can be done without any of this metadata, but when this information is available, Dolby® display management technology is able to truly optimize the quality of the content on the consumer's particular display. Dolby Vision metadata describes the content and how it was made, rather than just sending specific mapping information for a generic display type (or a couple of generic types) as some other systems do. This allows intelligent displays to dynamically create their own customized mappings, so they can fully capitalize on the strengths of their own distinct imaging capabilities. For live-broadcast environments, this data may be automatically generated in real time directly at the final emission point. This way, the broadcast signal may still be fully optimized for the end-user display, without the burden of carrying any metadata through the entire broadcast plant.

For movies and episodic content that is color-graded offline, Dolby has created systems that automatically generate content metadata and allow directors to see how their content will look on consumer devices. They can then add "trims" to allow them to adjust how the automated algorithms work, to better capture their intended vision on more limited displays.

For live content, Dolby is working with camera manufacturers and broadcasters to create equipment that will allow producers to create high-quality PQ signals that can be transmitted, and (optionally) to include content metadata at the emission encoder, just before the signal is transmitted.

BACKWARD COMPATIBILITY

To deliver an uncompromised experience, a system must consider the entire ecosystem—"glass to glass," from the camera to the display—and should consider both existing and future technologies. Cinema and broadcast cameras today can capture HDR and WCG content, even if their current target is standard dynamic range (SDR). Content creators must ensure that the signal they capture and archive is of the highest quality—once compromised, the original signal is impossible to recover. In offline environments, content creators archive in OpenEXR, which is a 16-bit floating-point format. In live environments, the format archived must be as high in quality as possible. PQ (SMPTE ST 2084) is one format that allows efficient storage of uncompromised original signals. To serve existing SDR televisions, a very high-quality SDR signal can be derived from the PQ format for simulcast applications.

In situations where simulcast is not viable, professionals often express the desire for a

DOLBY
VISION

backward-compatible option. Multiple ways to achieve backward compatibility are currently proposed, but we need to be very specific when we ask what exactly backward compatibility really means. Claims of direct signal compatibility between HDR and SDR displays are often implied for one such proposal known as hybrid log gamma (HLG), but this should not be confused with backward compatibility to current HD or SD televisions. All HDR/WCG systems being proposed today utilize the Ultra HD™ (UHD; Rec. 2020) color space, which allows many more colors to be encoded and displayed. But any system with Rec. 2020 color displayed directly on a legacy (Rec. 709 color) display will show very distorted colors.

So if HLG is not backward compatible to existing HD or SD televisions, what is it useful for? The claimed backward compatibility is only for SDR UHD televisions that natively understand Rec. 2020 color. Even if considered in this narrow use case, it should be obvious that if a single signal is to serve both HDR and SDR displays, the image on either the HDR or the SDR display (or both) must be compromised.

If true backward compatibility is required to feed legacy HD televisions directly, one method of achieving it in the market today is the Dolby Vision dual-layer solution. This system includes an uncompromised SDR Rec. 709 signal as a base layer that may be natively decoded by legacy decoders, with an enhancement layer to create the uncompromised HDR signal when a display with HDR/WCG capability is doing the decoding.

RECEIVER AND DISPLAY CONSIDERATIONS

see pdf
also see BT.2390

Standardized solutions are the key to enabling HDR and WCG technologies to spread quickly throughout the television industry. Dolby has been an active participant in international standards for video, contributing to the new ITU-R BT.2100 standard for HDR TV, as well as actively standardizing core technologies such as PQ and metadata in SMPTE (SMPTE ST 2084, ST 2086, and ST 2094). Dolby also participates in MPEG, DVB, ATSC, and 3GPP video standardization efforts, and is specifying its technology for reshaping and composing HDR signals in ETSI.

Hollywood and episodic content directors today choose Dolby Vision (PQ with dynamic and static metadata) and HDR10 (PQ with static metadata only) to deliver their content to consumers. Dolby has developed a universal decoder that handles both of these systems for maximum flexibility. This decoder has been implemented for both TV and set-top box (STB) architectures, and it is supported in current-generation silicon already widely deployed by TV manufacturers.

UHD televisions are now in the reach of the mass market—prices for these televisions now range as low as US$499. To reach this price point, manufacturers must omit

note: Samsung supports "HDR10+" extension to HDR10
(see HDR10_Ecosystem.pdf)
Also, Google's "HDR+" exists, for mobile phones

features that make it easier to deliver a great HDR and WCG experience, such as local dimming, quantum dots, or OLEDs. Producing a high-quality experience on these TVs (and on other devices like tablets) requires advanced technology that understands the content and can adapt the edge-lit backlight to deliver the best experience possible. The dynamic content metadata in Dolby Vision technology makes highly accurate HDR possible even in mass-market devices.

WHAT'S NEXT?

Most Hollywood studios today deliver their content in PQ with Dolby Vision metadata, and Dolby is working with broadcast, episodic, and game studios to deliver new content. Our partners in the television industry are delivering innovative Dolby Vision enabled TVs across a wide price range worldwide. We look forward to working with partners in other industries to deliver exciting new HDR/WCG experiences to new display devices.

Dolby Vision was created to effectively address the challenges to delivering HDR and WCG experiences through a robust future-ready solution. Dolby Vision is leading the drive toward next-generation imaging technology to enable spectacular visual experiences for audiences across the globe.