Bit-Rate Evaluation of Compressed HDR Using SL-HDR1

Ciro Noronha, Kyle Wilken, Ryan Wallenberg
Cobalt Digital Inc.
Champaign, Illinois, USA

Introduction

A number of current HDR standards include transmission of metadata along with the content. HDR encodes absolute luminance information, which may be outside the limits of what a particular monitor can display. Metadata helps the monitor map the incoming content to its capabilities.

The SL-HDR1 standard [1] operates this way. What is unique about it is the fact that the content is actually SDR, and the metadata allows a compatible device to map that SDR content to the original HDR, or to some intermediate level that it can support. Legacy devices with no metadata support can simply display SDR. This is similar to what happened when color TV was introduced – the transmission was a black-and-white image with color information on the side.

Video encoders and decoders are agnostic to HDR. The encoder takes the video samples and converts them to a bit stream; the decoder converts that bit stream back to video samples that are approximately the same as what entered the encoder. Neither device interprets the meaning of the video samples. Metadata, if present, may be carried along the bit stream.

In this article, we attempt to answer the question of whether or not there is a bit rate penalty when one uses SL-HDR1 to transport HDR over a compressed video link. We do that by establishing a baseline with native HDR video, and then switch to SL-HDR1 and determine at what bit rate the quality is the same as the native HDR. This is similar to the work presented in [2].

A Review of HDR Technology

Dynamic range is the ratio between the lowest and highest values of the luminance, or intensity of light emitted, from a display. Essentially, it is the ratio between the “whitest white” and “blackest black.” Conventional 8-bit displays, for example, have a dynamic range of approximately 100:1. Dynamic range is usually expressed in stops, which is calculated as the log base 2 of the ratio. Therefore, 8-bit SDR has a dynamic range of approximately six stops, while professional 10-bit SDR offers approximately ten stops.

The human eye, however, can see about 12-14 stops, which means it can perceive more dynamic range than offered by 10-bit SDR material. The method to pack an HDR signal into a 10-bit display makes use of the fact that the human eye is nonlinear in its response and perceives more detail at lower luminosities. HDR delivers images with improved details by assigning bits to light intensity in a nonlinear manner. More bits are assigned to the lower intensities and fewer bits to the higher intensities to express a higher range. Fundamentally, HDR shows more detail in the bright areas.

Light intensity is measured in candelas per square meter (cd/m2), also known as nits. A standard HDTV can produce luminance at about 100 nits. A typical UHD LCD display can range from 0.05 to 1,000 nits, while a typical UHD OLED monitor can produce 0.0005 to 540 nits. In contrast, HDR can code up to 10,000 nits, which no commercial monitor can reproduce (at the time of this writing).

The luminance encoded in SDR signals is relative – at 100 percent, it tells the display to show its whitest white. In contrast, HDR codes the absolute value of the luminance, using a non-linear transfer function based on what the eye can perceive. This is the SMPTE 2084 Perceptual Quantizer (PQ) transfer [3], illustrated in Figure 1.

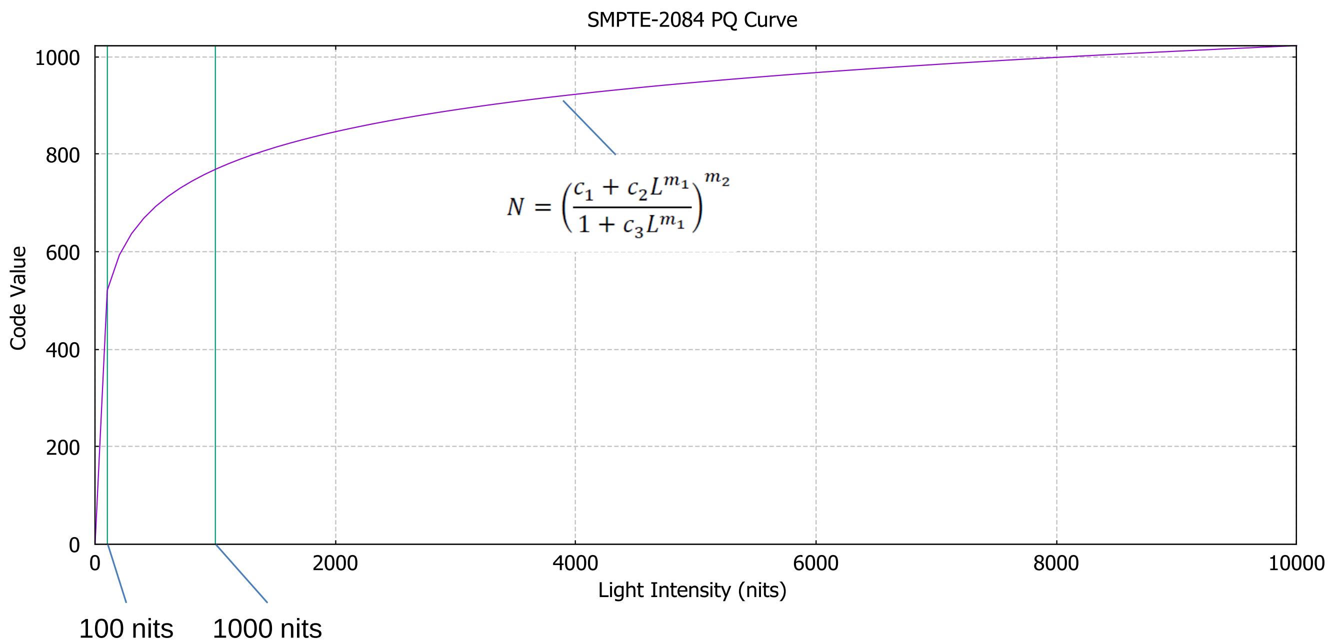

FIGURE 1: SMPTE 2048 PERCEPTUAL QUANTIZER CURVE

Beyond luminance, HDR also features wide color gamut (WCG), which has more color information than standard HD signals. The set of colors a signal uses or a display can show is called the "color space." There are three defined color spaces in use today: ITU-R BT.709, which is considered the standard color space for HD; DCI-P3, which is the standard for digital cinema; and ITU-R BT.2020, which is the UHDTV standard. The color spaces are illustrated in Figure 2.

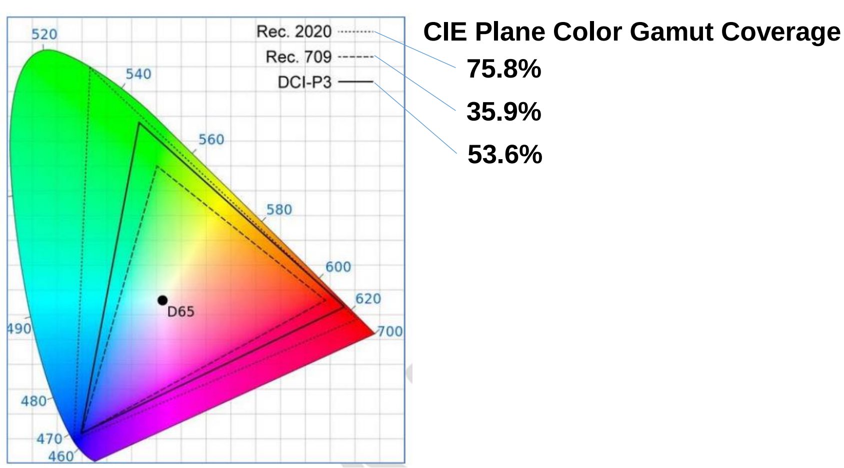

FIGURE 2: COLOR SPACES

No commercial monitor today can display the full Rec. 2020 color space, but a UHDTV with HDR and WCG will show more than the entire Rec. 709 color space and at least 90 percent of the DCI-P3 color space.

Static and Dynamic Metadata

Displays vary widely in their capabilities. A display may be presented with a signal that has luminance and/or color information that are outside its capabilities. The display must create an image as close as possible to the original source material – and to help the display do this job, metadata may be included in the stream.

Metadata can be static or dynamic. Static metadata, as the name implies, is fixed for the duration of the content, and provides only "general" information. The basis for static metadata is SMPTE 2086 [4]. When content is created, it is mastered on a reference display by the director/colorist. The static metadata describes the characteristics of this mastering display, so that the display currently playing it can best approximate the mastering display based on its capabilities. SMPTE 2086 static metadata includes parameters such as the color and luminance range of the mastering display, the color space used, and the transfer characteristics. SMPTE 2086 was augmented by CTA to include additional parameters such as the Maximum Content Light Level (so the display will know the "brightest" pixel in the content) and the Maximum Frame-Average Light Level.

The objective of dynamic metadata is the same, but it changes from frame to frame. Static metadata can be seen as sort of an average over the content, while dynamic metadata is tailored to each individual frame. This dynamic metadata is one of the ways the various HDR standards differ.

HDR Standards

There are a number of competing HDR standards available today. From a high-level point of view, they can be classified as static (using static metadata or no metadata) and dynamic (using dynamic metadata). The baseline HDR support starts with the SMPTE 2084 PQ transfer function, using 10-bit or 12-bit samples. This is the basis for most HDR standards. The HDR10 standard is simply the combination of SMPTE 2084 with 10-bit samples and SMPTE 2086 static metadata; this combination is standardized in ATSC A/341 and in other standards.

One interesting static HDR standard in wide use today is Hybrid Log-Gamma (HLG), which is not based on SMPTE 2084 PQ. It is an attempt to use a backward-compatible transfer curve that will "work" with both SDR and HDR monitors without any metadata. At the low luminance levels, it matches SDR, so an HLG signal applied to an SDR monitor will "look OK", while an HDR monitor will show the improved ranges at the higher luminance levels. HLG trades off simplicity (same signal everywhere, no metadata processing) with quality (it is not as good as the dynamic metadata options). HLG is standardized in ARIB STD B-67, ITU-R BT.2100, and ATSC 3.0.

Most of the dynamic HDR standards start from a PQ base layer, with metadata defined in ST 2094-1 [6]. The most relevant ones are SMPTE 2094-10 (Dolby Vision), SMPTE 2094-40 (HDR10+, a dynamic version of HDR10), and SL-HDR1.

The basic operation of various dynamic standards can be understood as transmitting "an image plus instructions" that can be processed by a monitor. What varies is the starting point. In Dolby Vision and HDR10+, the starting point is an HDR image, and the "instructions" allow the mapping of that HDR image to any monitor, all the way down to an SDR monitor. While it is possible for an end device to generate SDR from this HDR signal, that end device needs to understand and process the metadata in order to do so.

With SL-HDR1 [1], the opposite happens. What is transmitted is a standard SDR signal, and the metadata allows a compatible device to reconstruct the original HDR signal (or any intermediate level suitable for its capabilities). This is the ideal way to support legacy SDR devices – they will just ignore the metadata (because they do not support it) and simply display the SDR image. This is conceptually the same as what was done when analog color TV was introduced. The signal was the standard black-and-white content, with the color information added “on the side”. A black-and-white TV would understand the signal and show the black-and-white version, while a newer color TV would extract the color information and show a color picture. SL-HDR1 is standardized in ETSI TS 103 433-1 and included as part of ATSC 3.0 in ATSC A/341.

Bit Rate Evaluation

Test Setup

The test setup is shown in Figure 3. As indicated, there are three test paths:

- **Path 1** (in **blue**) is an end-to-end HDR10 path. A native SDI HDR10 signal is applied directly to the encoder, converted to either AVC or HEVC, and then decoded back to SDI.
- **Path 2** (in **purple**) is an SL-HDR1 path. The native SDI HDR10 signal is routed to an SL-HDR1 encoder, which produces an SDI SDR signal with metadata, carried in the ancillary data space using SMPTE-2108 [5]. This signal is applied to the encoder and converted to either AVC or HEVC. The SL-HDR1 metadata is extracted from the ancillary space and injected in the video bit stream as SEI messages. The decoder produces an SDI signal, and with the metadata restored to the ancillary space. Finally, an SL-HDR1 decoder re-creates the SDI HDR10 signal.
- **Path 3** (in **red**) is an SL-HDR1 path that bypasses the encoder/decoder. It is used to obtain a baseline reading without compression.

In all cases, both the original and decoded signals are captured in YUV format by a video recorder. The files are transferred to a computer, where the quality metrics are calculated. For this evaluation, we selected the following metrics:

- Peak Signal-to-Noise Ratio (PSNR), which measures the absolute difference between each frame in a sequence. It is well-known that PSNR does not correlate well with perceived quality.
- PSNR_DE100: PSNR of mean of absolute deltaE2000 metric, referred to a 100-nit luminance.
- PSNR_L100: PSNR of mean square error of L component of CIELab color space used for the deltaE2000 metric, referred to a 100-nit luminance.

The DE100 and L100 metrics were selected since they have been shown to correlate well with perceived quality [7].

The test procedure was as follows:

1. Take a baseline reading of the PSNR using **Path 3**. This only needs to be done once.
2. Select a target test video bit rate B_r for the AVC/HEVC encoder.
3. Run the **Path 1** signal and compute the selected quality metrics.
4. Run the **Path 2** signal and compute the selected quality metrics.
5. Repeat steps 2-4 for other values of B_r.
6. Perform the BD-rate computation for both PSNR_DE100 and PSNR_L100.

After the PSNR results were obtained, we performed the BD-rate computation [8] in order to evaluate the average increase or decrease in bit rate in the SL-HDR1 case to match the quality metric in the HDR10 case. This process was done only for the metrics that correlate well with perceived quality, namely PSNR_DE100 and PSNR_L100.

The details of the test setup in Figure 3 are:

- SL-HDR1 Encoder and Decoder: Cobalt 9904-UDX
- AVC/HEVC Encoder: Cobalt 9992-ENC
    - GOP size: 100 frames

- Bit Depth: 10 bits
- Chroma Mode: 4:2:0 (consumer grade signals)
- HEVC only: Chroma Qp Offset set to zero.

- AVC/HEVC Decoder: Cobalt 9992-DEC

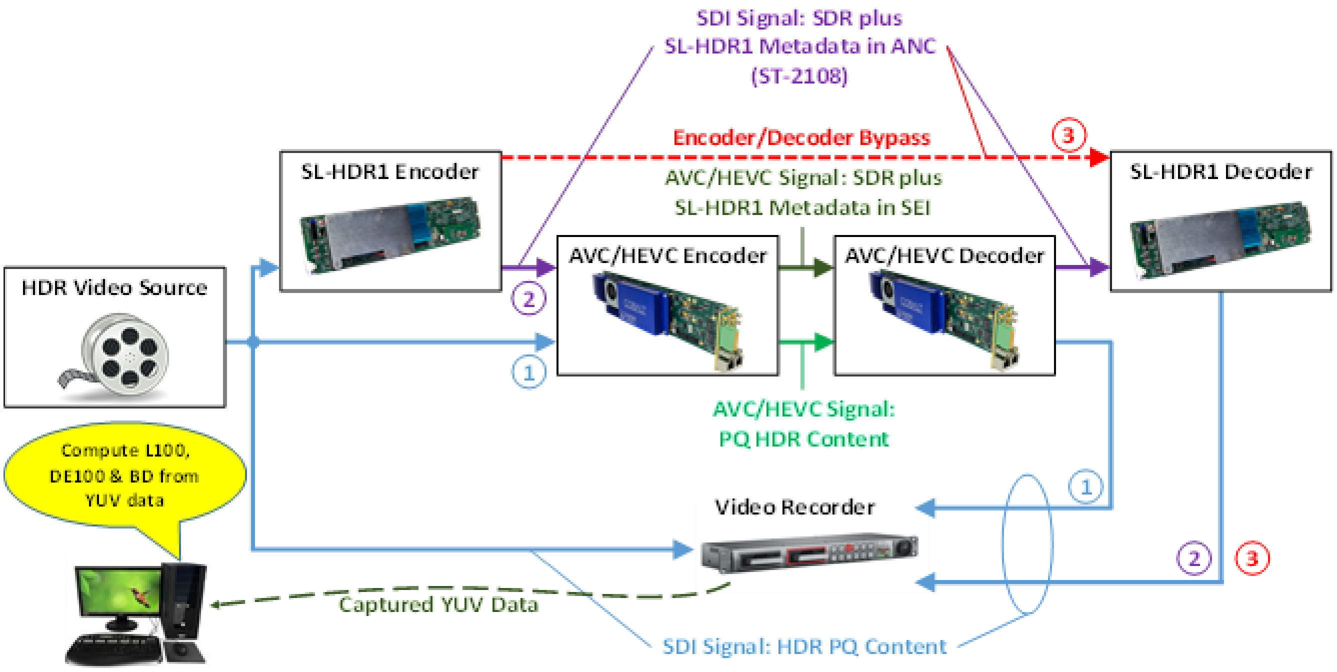

FIGURE 3: TEST SETUP

Test Sequences

The tests were performed with three test sequences. All sequences had the following common parameters:

- Duration: 12 seconds
- Resolution: 1920×1080
- Color Space: BT 2020

Figure 4 shows a representative frame from each sequence. Note that each sequence had a different frame rate, as indicated in the figure.

Sequence 1 (50fps)

Sequence 2 (59.94fps)

Sequence 3 (24fps)

FIGURE 4: TEST SEQUENCES

Table 1 presents the quality metrics of the SL-HDR1 process before encoding and decoding (Path 3 in Figure 3). The SL-HDR1 process is not exact – there is a small impact in the metrics, as the image is not exactly reconstructed. Note that the YUV PSNR is provided as reference; since these metrics are fundamentally different, the absolute values should not be compared between them.

TABLE 1: SL-HDR1 METRICS WITH ENCODER/DECODER BYPASSED

Sequence	Path 3 YUV PSNR	Path 3 DE100	Path 3 L100
Sequence 1	59.26 dB	38.54 dB	51.80 dB
Sequence 2	58.72 dB	38.16 dB	51.53 dB
Sequence 3	59.07 dB	38.32 dB	51.67 dB

AVC (H.264) Results

Figure 5 shows the AVC PSNR_DE100 test results over a range of representative bit rates. The behavior varies depending on the sequence and the bit rate:

- For Sequence 1, there is a clear quality advantage for SL-HDR1 for the whole bit rate range.
- For Sequence 2, the overall quality level is lower than sequences 1 and 2, and there is almost no difference between SL-HDR1 and HDR10.
- For Sequence 3, SL-HDR1 has a large advantage at lower bit rates, but falls below HDR10 at higher bit rates.

Figure 6 shows the AVC PSNR_L100 test results over the same range of bit rates as Figure 5. SL-HDR1 shows a clear quality advantage for Sequences 1 and 3. For Sequence 2, again there is almost no difference between SL-HDR1 and HDR10.

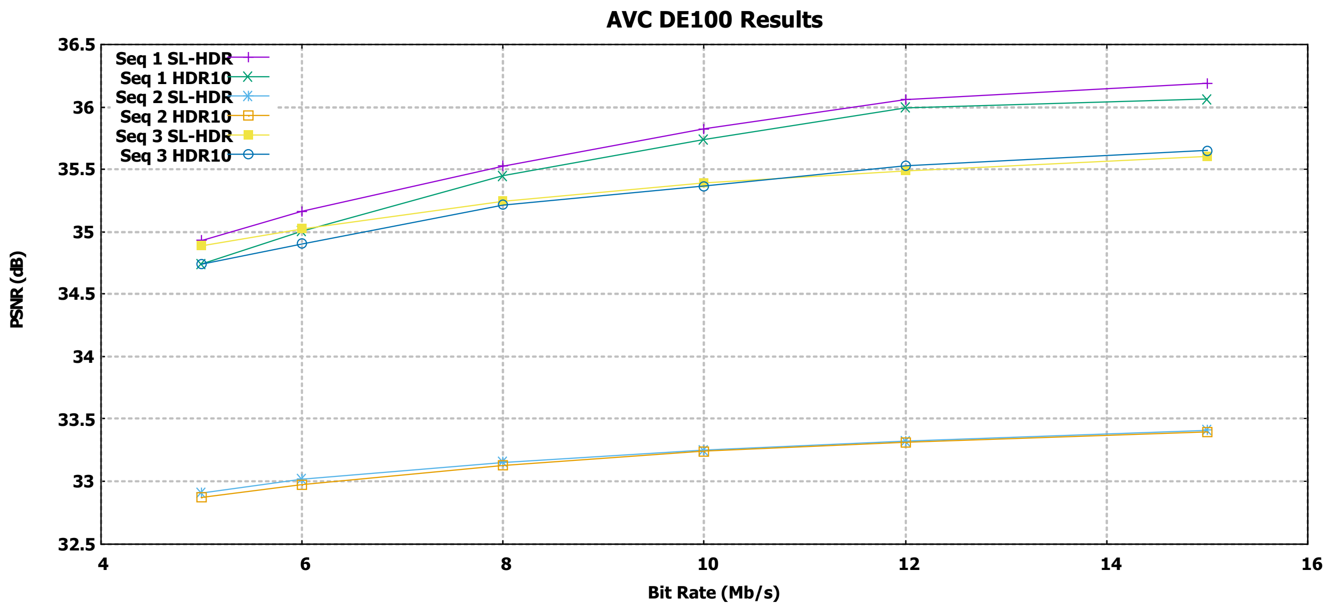

FIGURE 5: AVC PSNR_DE100 RESULTS

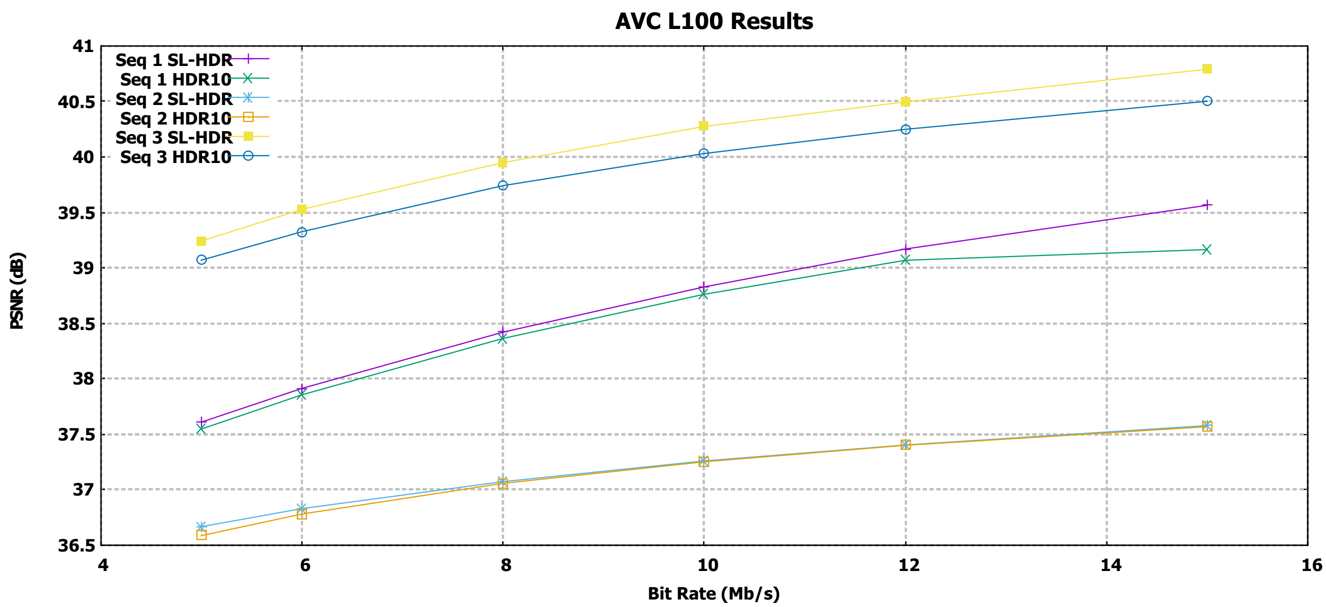

FIGURE 6: AVC PSNR_L100 RESULTS

HEVC (H.265) Results

Figure 7 shows the HEVC PSNR_DE100 results for the same sequences, using a suitable bit rate range. In this case, the HDR10 and the SL-HDR1 results are very close. Depending on the sequence and the bit rate range, there is a small advantage for one or the other, but it is not significant.

Figure 8 shows the PSNR_L100 results for HEVC, using the same bit rate ranges as Figure 7. Using this metric, there is a clear advantage for SL-HDR1 for all sequences and all bit rates.

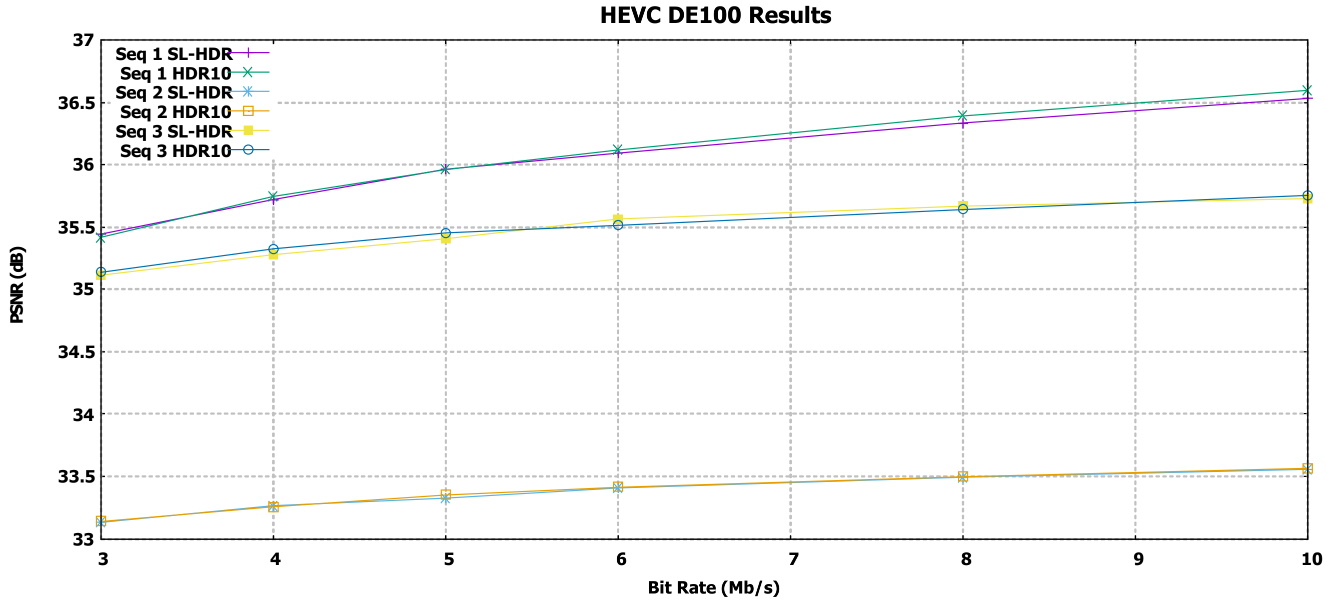

FIGURE 7: HEVC PSNR_DE100 RESULTS

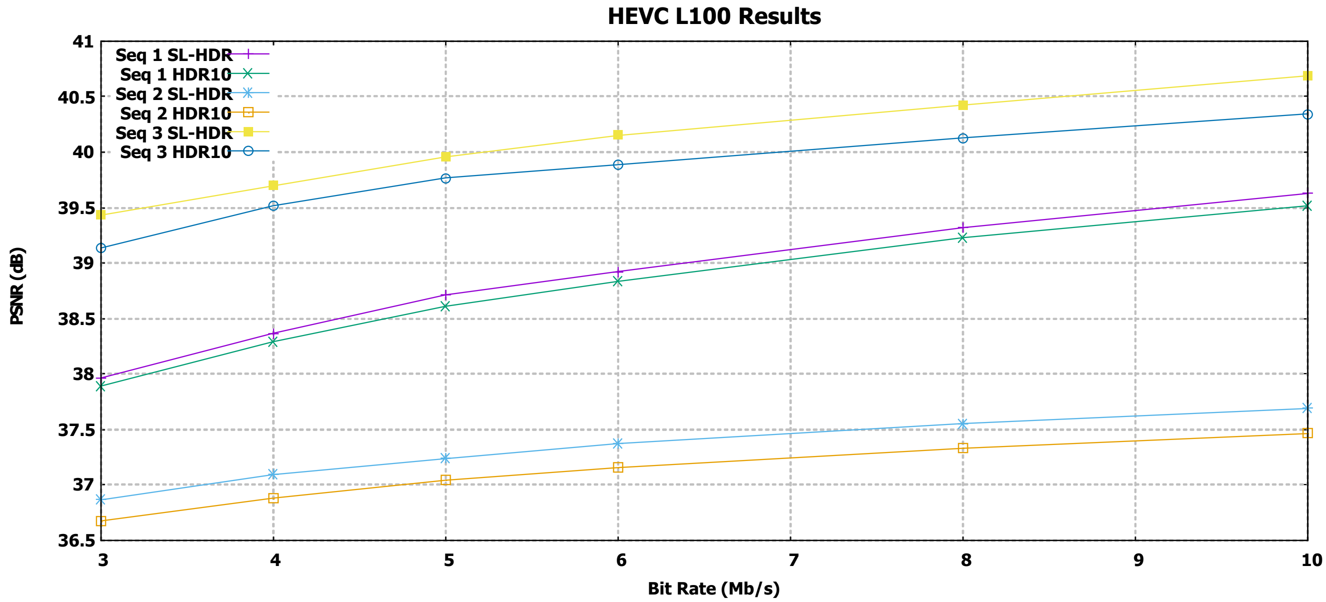

FIGURE 8: HEVC PSNR_L100 RESULTS

Bit Rate Evaluation

The remaining question pertains to quantifying the bit rate advantage. If one is seeking to achieve a certain target quality for an HDR link and has the option of either transmitting HDR10 natively or SL-HDR1, which method will yield the lowest bit rate and by how much? The standard way of answering this question is by the use of BD-rate [8], which produces an “average” value over the tested range. The relevant numbers are provided in Table 2 below. In this table, positive values indicate that the SL-HDR1 bit rate for the same quality is higher, and negative values indicate that it is lower.

TABLE 2: BD-RATE VALUES

Sequence	HEVC		AVC	
	DE100	L100	DE100	L100
Sequence 1	0.90%	-6.92%	-6.07%	-3.65%
Sequence 2	5.61%	-25.90%	-5.84%	-2.42%
Sequence 3	7.27%	-20.25%	-2.58%	-13.07%

Table 2 indicates that, for most combinations, a lower bit rate is required for SL-HDR1 transport as compared with a straight HDR10 link, sometimes significantly so. This confirms the conclusions presented in [2], using a different commercial encoder.

Notes on Metadata Bit Rate

One final item for discussion pertains to the SL-HDR1 metadata bit rate. When transporting HDR10, there is no metadata requirement, but SL-HDR1 adds per-frame metadata that is included in SEI messages inside the video elementary stream. The maximum amount of metadata per frame is 61 bytes. Therefore, an upper bound on the SL-HDR1 metadata bit rate is 24.4 kb/s for a 50-fps signal, and 29.3 kb/s for a 60-fps signal. This bit rate increase is negligible (less than one audio channel), and, in many encoders, is absorbed by a slight adjustment in the NULL packet rate, so the overall bit rate is unchanged.

Conclusions

The impact of SL-HDR1 in compressed bit streams is a function of the content, and can be quite significant. When using quality metrics that are better correlated with human perception, it is often possible to actually decrease the link bit rate while keeping the same quality, as originally reported in [2]. We found that the changes are more pronounced in AVC than HEVC.

References

[1] European Telecommunications Standards Institute, “High-Performance Single Layer High Dynamic Range (HDR) System for use in Consumer Electronics devices; Part 1: Directly Standard Dynamic Range (SDR) Compatible HDR System (SL-HDR1),” *ETSI TS 103 433-1 V1.2.1*, 2017

[2] Touze, D., and Kerkhof, L., “Single-Layer HDR Video Coding with SDR Backward Compatibility”, *2017 SCTE-ISBE CABLE-TEC EXPO*

[3] Society of Motion Picture and Television Engineers, “High Dynamic Range Electro-Optical Function of Mastering Reference Displays,” *SMPTE ST 2084*, 2014

[4] Society of Motion Picture and Television Engineers, “Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images,” *SMPTE ST 2086*, 2014

[5] Society of Motion Picture and Television Engineers, “HDR/WCG Metadata Packing and Signaling in the Vertical Ancillary Data Space,” *SMPTE ST 2108-1*, 2018

[6] Society of Motion Picture and Television Engineers, “Dynamic Metadata for Color Volume Transform – Core Components,” *SMPTE ST 2094-1*, 2016

[7] Hanhart, P., Řeřábek, M., and Ebrahimi, T., "Towards high dynamic range extensions of HEVC: subjective evaluation of potential coding technologies", *Proc. SPIE 9599, Applications of Digital Image Processing XXXVIII, 95990G (22 September 2015)*

[8] Bjontegaard, G., "Improvements of the BD-PSNR model", *ITU-T SG16/Q6 VCEG 35th meeting, Berlin, Germany, 16–18 July, 2008, Doc. VCEG-AI11*