Bit-Rate Evaluation of Compressed HDR Using SL-HDR1

Ciro Noronha, Kyle Wilken, Ryan Wallenberg
Cobalt Digital Inc.
Champaign, Illinois, USA
ciro.noronha@cobaltdigital.com

Abstract – Video encoders and decoders are usually transparent to whether or not their input signal is High Dynamic Range (HDR). Encoders and decoders simply transport the video samples. When it comes to transporting HDR, one has the choice of simply transporting HDR in its native format, or using one of the metadata formats. One such format is SL-HDR1, where the signal is converted to SDR and transmitted with metadata that allows a receiver to reconstruct the HDR signal. This paper focuses on the required bit rates to produce a final HDR signal over a compressed link for a given quality. We compare encoding SMPTE-2084 PQ HDR signals directly versus using SL-HDR1 to generate an SDR signal plus dynamic metadata. The comparison is done objectively by comparing the PSNR of the decoded signal. The SMPTE-2084 HDR signal is used as a reference at a fixed bit rate, and the bit rate of the SL-HDR1 encoded signal is varied until it matches the PSNR, over a range of source material. The evaluation is done for both AVC (H.264) and HEVC (H.265). This is similar to the work described in [1], with different content and metrics.

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 [2], illustrated in Figure 1.

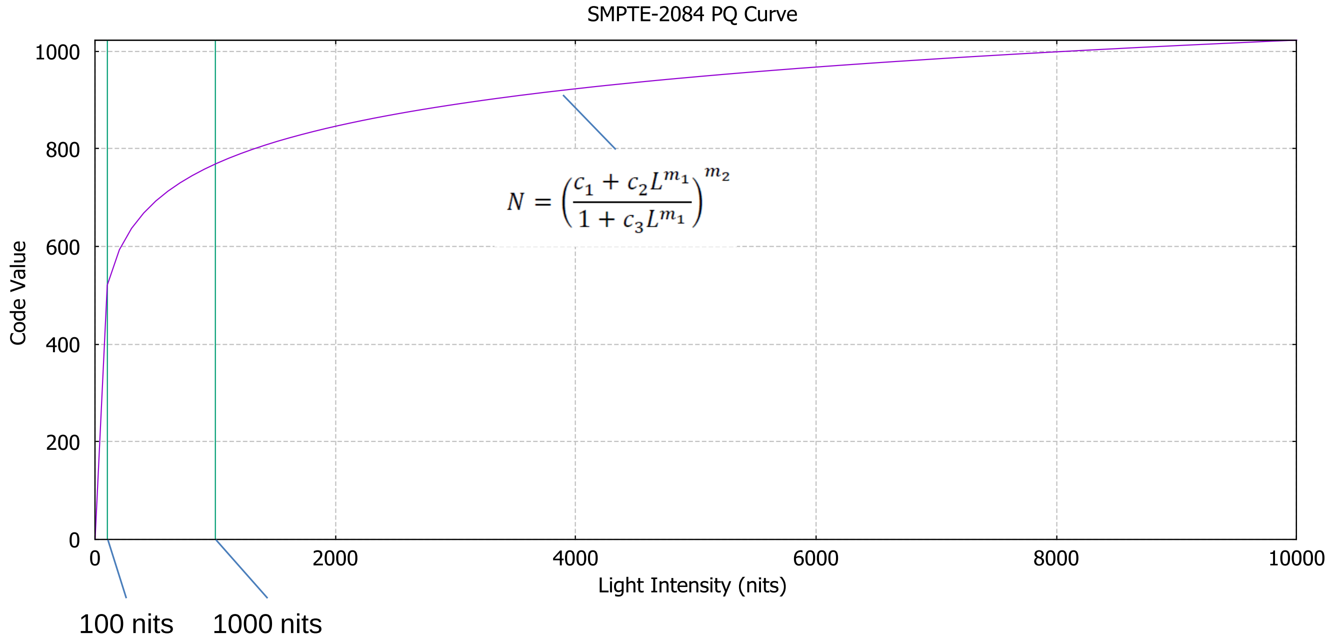

FIGURE 1: SMPTE 2048 PERCEPTUAL QUANTIZER CURVE

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.

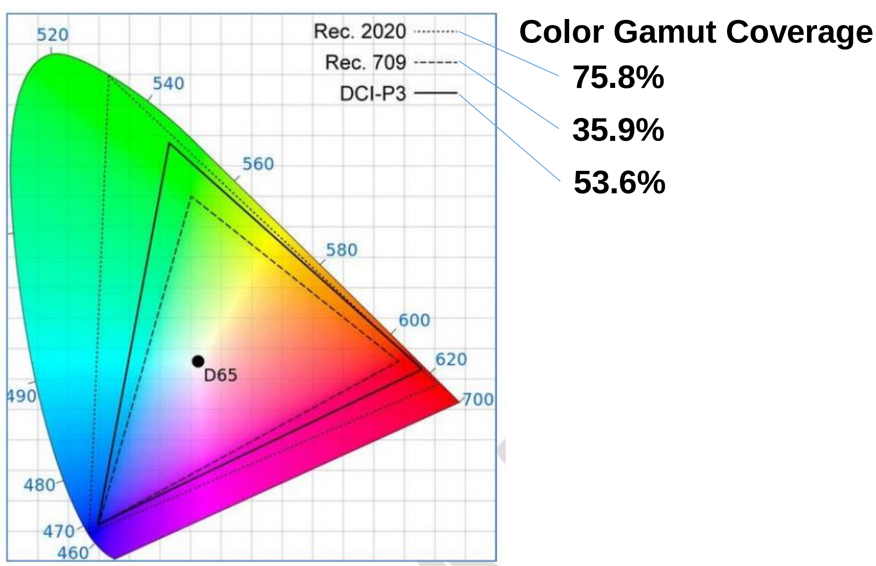

FIGURE 2: COLOR SPACES

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 [3]. 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.

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 [4]. 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 [5], 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.

Bit Rate Measurements

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 HDR path. A native SDI HDR 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 HDR signal is routed to an SL-HDR1 encoder, which produces an SDI SDR signal with metadata, which is carried in the ancillary data space using SMPTE-2108 [6]. 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 bitstream 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 HDR 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.

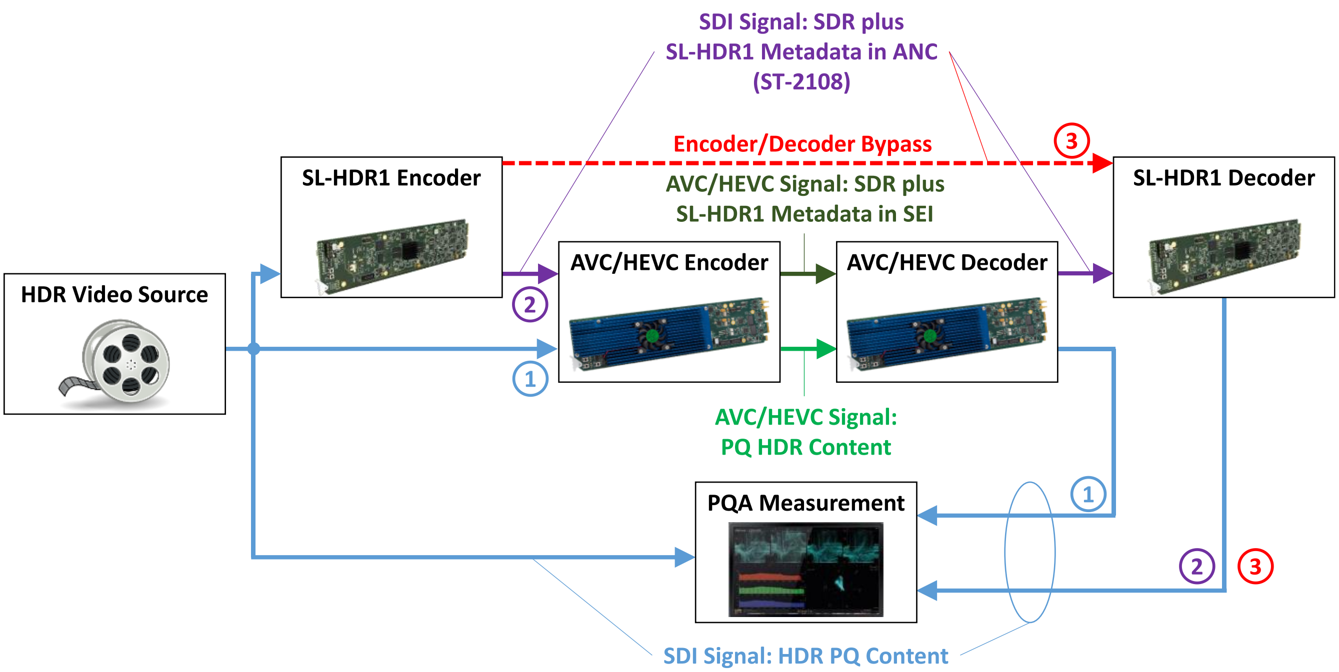

FIGURE 3: TEST SETUP

In both cases, a Picture Quality Analyzer (PQA) compares the original signal with the decoded signal and computes the quality metrics. 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.
- Compensated Signal-to-Noise Ratio (CSNR), which is a custom metric from the PQA used in the evaluation. CSNR attempts to capture the fact that picture degradation artifacts are more

noticeable around object edges and at mid-range brightness levels. A detailed discussion of the algorithm can be found in [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 bit rate B_r for the AVC/HEVC encoder.
3. Run the **Path 1** signal and record the PSNR for the sequence, which we will denote by P_1.
4. Run the **Path 2** signal and record the PSNR for the sequence, which we will denote by P_2.
5. Still running the **Path 2** signal, adjust the encoder bit rate until the PQA measures the same PSNR P_1 as in step 3 above. Record the bit rate B_h at which the PSNR matches.
6. Repeat steps 2-5 for other values of B_r.
7. Repeat steps 1-5 for the CSNR.

The details of the test setup in Figure 3 are:

- SL-HDR1 Encoder and Decoder: Cobalt 9904
- AVC/HEVC Encoder: Cobalt 9992-ENC
    - GOP size: 60 frames
    - Bit Depth: 10 bits
    - Chroma Mode: 4:2:0 (consumer grade signals)
- AVC/HEVC Decoder: Cobalt 9992-DEC
- PQA Tester: Omnitek OTR 1003

Test Sequences

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

- Duration: 12 seconds
- Resolution: 1920×1080
- Frame Rate: 50 frames/second, progressive
- Color Space: BT 2020

Figure 4 shows a representative frame from each sequence.

The content of each sequence was:

- Sequence 1: “basejump”, extreme sports in mountains
- Sequence 2: “baseball”, baseball game at night
- Sequence 3: “nature”, nature scenes with dark and bright lighting

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.

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

Sequence	Path 3 PSNR	Path 3 CSNR
Sequence 1	66.00 dB	65.90 dB
Sequence 2	60.21 dB	68.99 dB
Sequence 3	56.54 dB	52.63 dB

Sequence 1

Sequence 2

Sequence 3

FIGURE 4: TEST SEQUENCES

AVC (H.264) Results

The first step in the evaluation is to compare the quality metrics for Paths 1 and 2 in Figure 3 at the same encoder bit rate (values P_1 and P_2). This comparison is presented in Figure 5 for PSNR and in Figure 6 for CSNR.

As indicated before, applying SL-HDR1 processing to a native HDR signal has some PSNR impact. Figure 5 demonstrates that the same results are observed when compression is used – at any given bit rate, there is reduction in PSNR as the original PQ HDR signal is reconstructed. However, the magnitude of the PSNR change is relatively small, and content-dependent.

The AVC CSNR results are presented in Figure 6. The figure indicates that, for Sequences 1 and 2, the CSNR is actually **better** for SL-HDR1 at the same bit rate, indicating that, for these sequences, one could **reduce** the bit rate to achieve the same quality, as measured by this indicator.

The next step in the evaluation was to change the encoder bit rate in the SL-HDR1 case until the metrics matched the PQ case. The results of this test are presented in Figure 7 for PSNR and Figure 8 shows the bit rate change B_h/B_r, expressed as a percentage of B_r.

Figure 7 indicates that, depending on the sequence, large changes in bit rate are required to achieve small improvements in PSNR. However, as indicated in Figure 8, for some sequences the bit rate can actually be reduced to achieve the same CSNR. These results are very content-dependent. For Sequence 2, the bit rate can be reduced by about 30% across the whole range; for Sequence 1, the gains are smaller, and, for Sequence 3, the bit rate has to be increased, similar to the PSNR results. These results are similar to what is reported in [1].

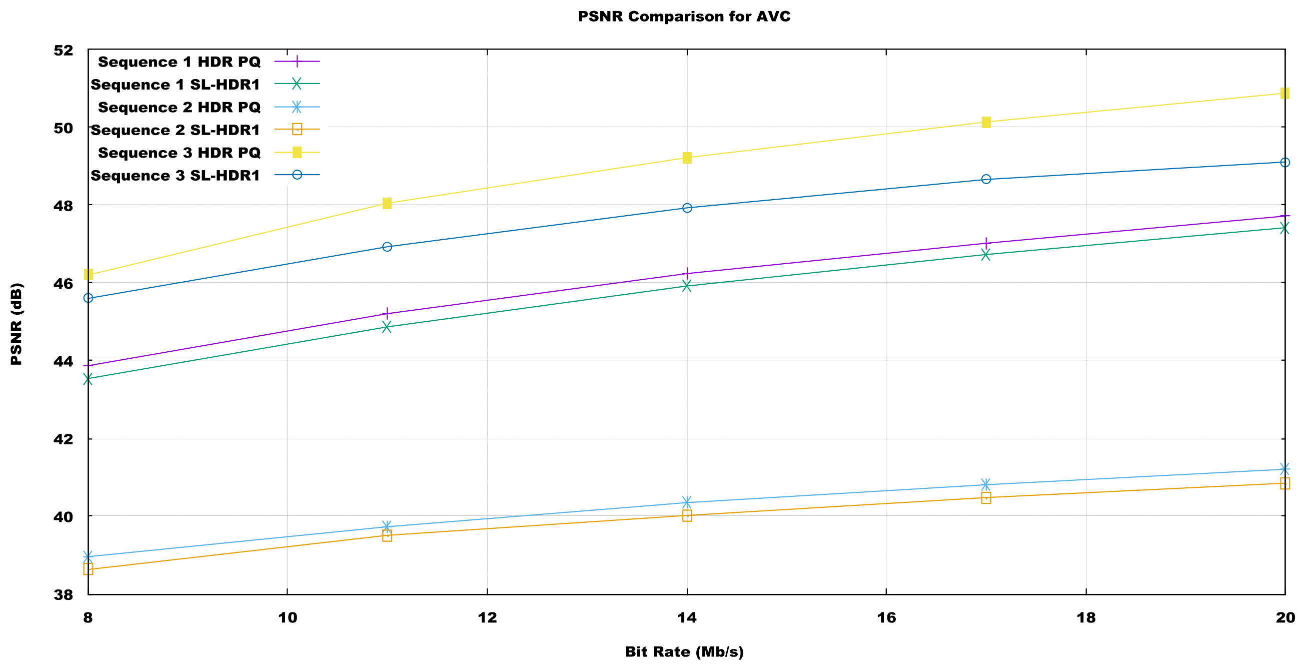

FIGURE 5: SL-HDR1 PSNR IMPACT FOR AVC

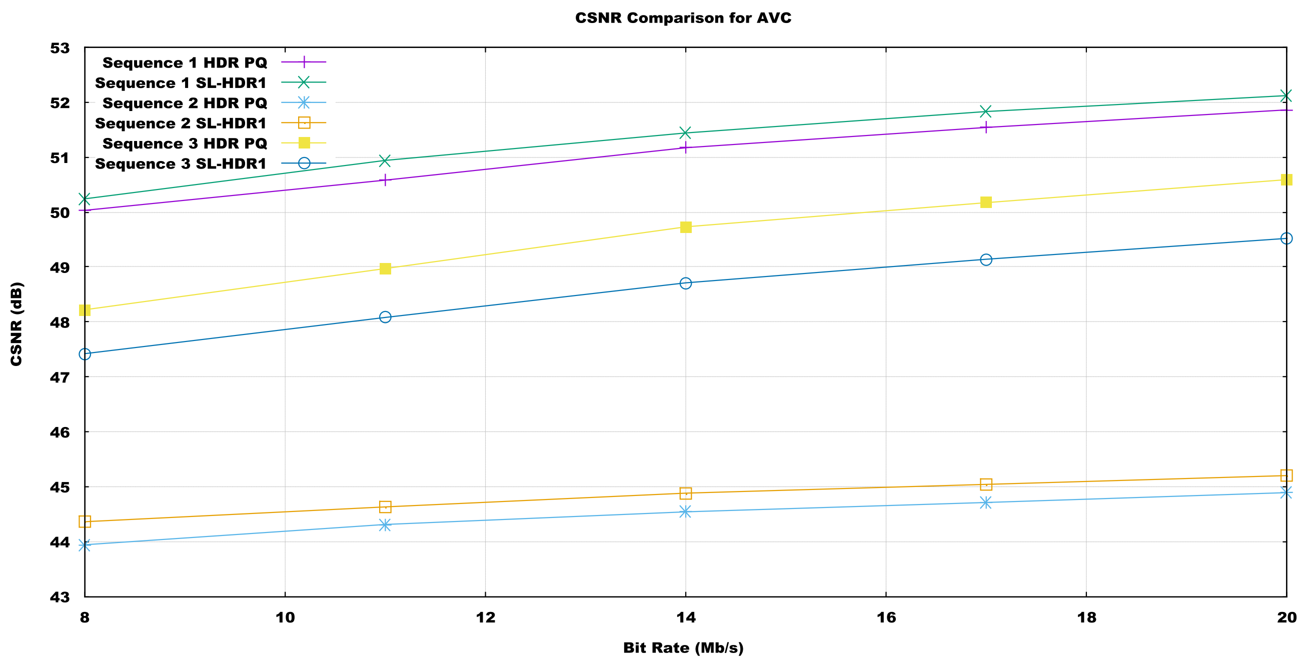

FIGURE 6: SL-HDR1 CSNR IMPACT FOR AVC

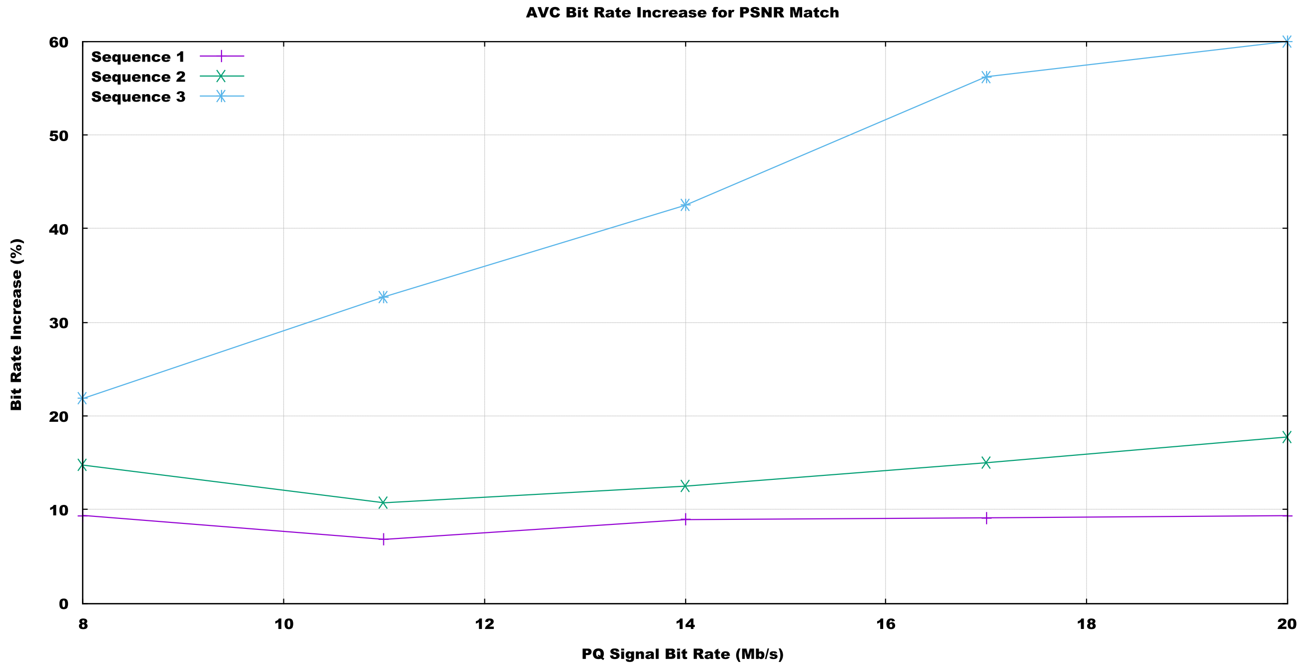

FIGURE 7: AVC BIT RATE INCREASE FOR MATCHING PSNR

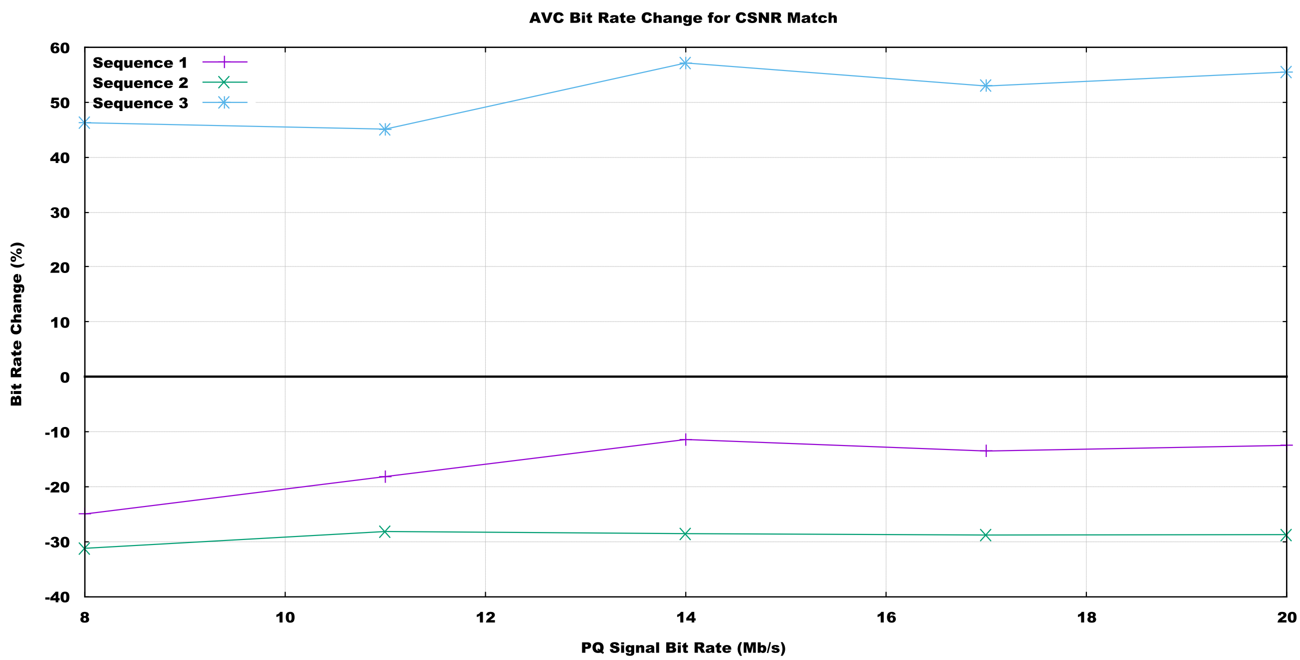

FIGURE 8: AVC BIT RATE CHANGE FOR MATCHING CSNR

HEVC (H.265) Results

The same steps were repeated for HEVC (H.265) encoding. The quality metric impact at the same bit rate is shown in Figure 9 for the PSNR and Figure 10 for the CSNR.

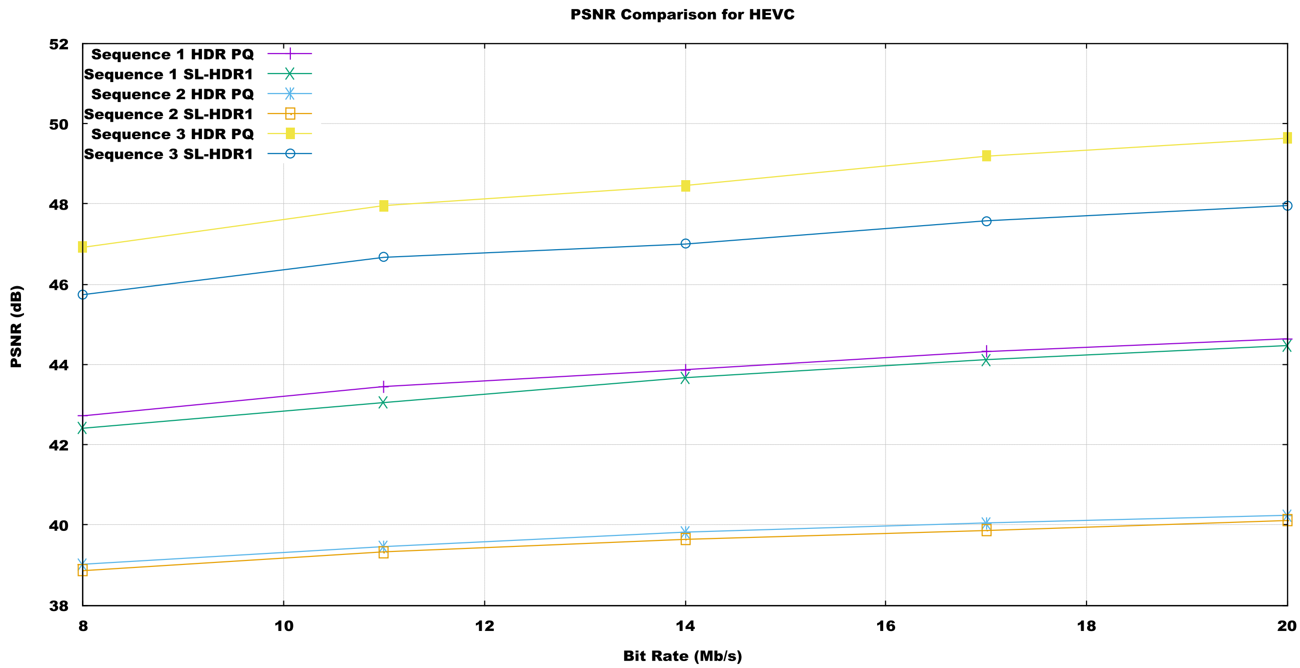

FIGURE 9: SL-HDR1 PSNR IMPACT FOR HEVC

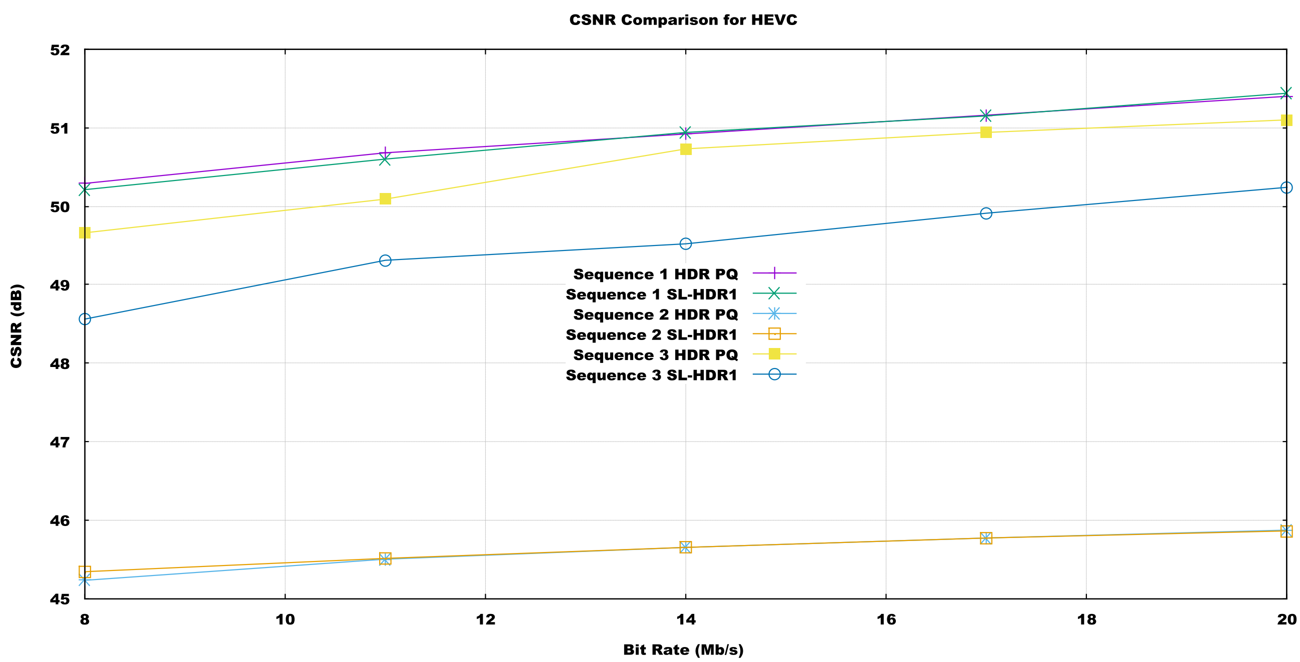

FIGURE 10: SL-HDR1 CSNR IMPACT FOR HEVC

The HEVC PSNR results in Figure 9 are very similar to the AVC results in Figure 5. However, for the CSNR, the use of SL-HDR1 makes very little difference for Sequences 1 and 2. Similarly to AVC, Sequence 3 shows a large difference when SL-HDR1 is applied.

The bit rate changes required to match the metrics are shown in Figure 11 for PSNR and Figure 12 for CSNR.

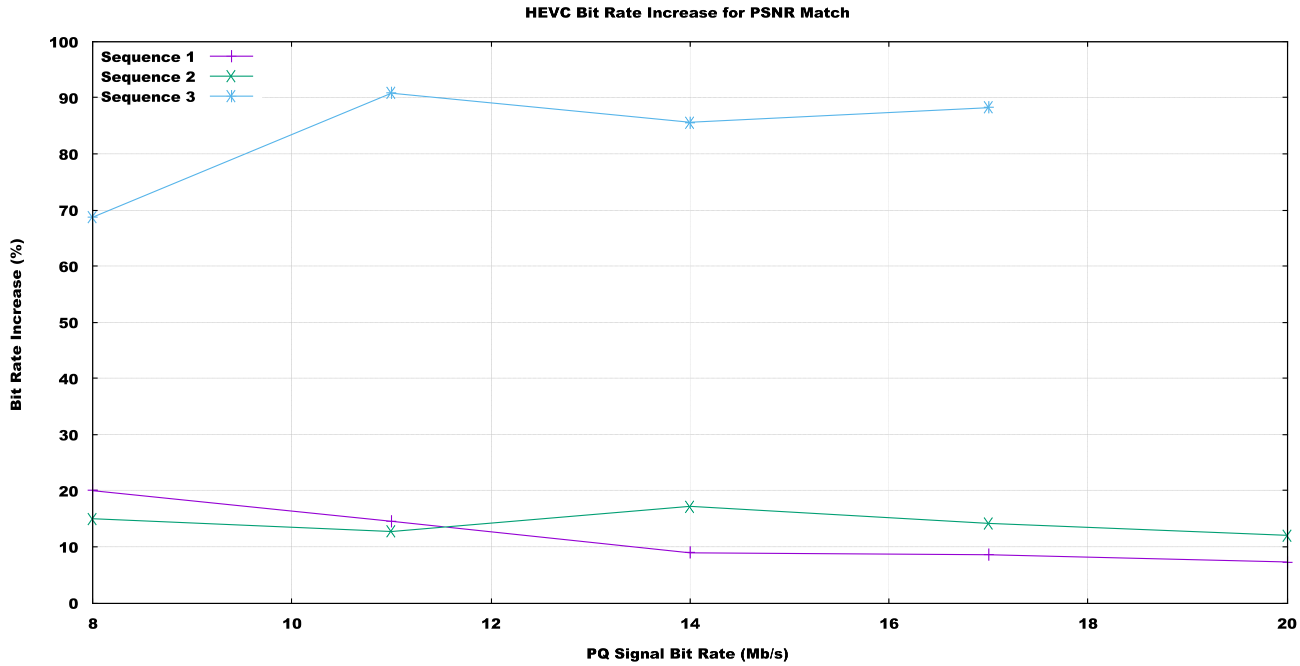

FIGURE 11: HEVC BIT RATE INCREASE FOR MATCHING PSNR

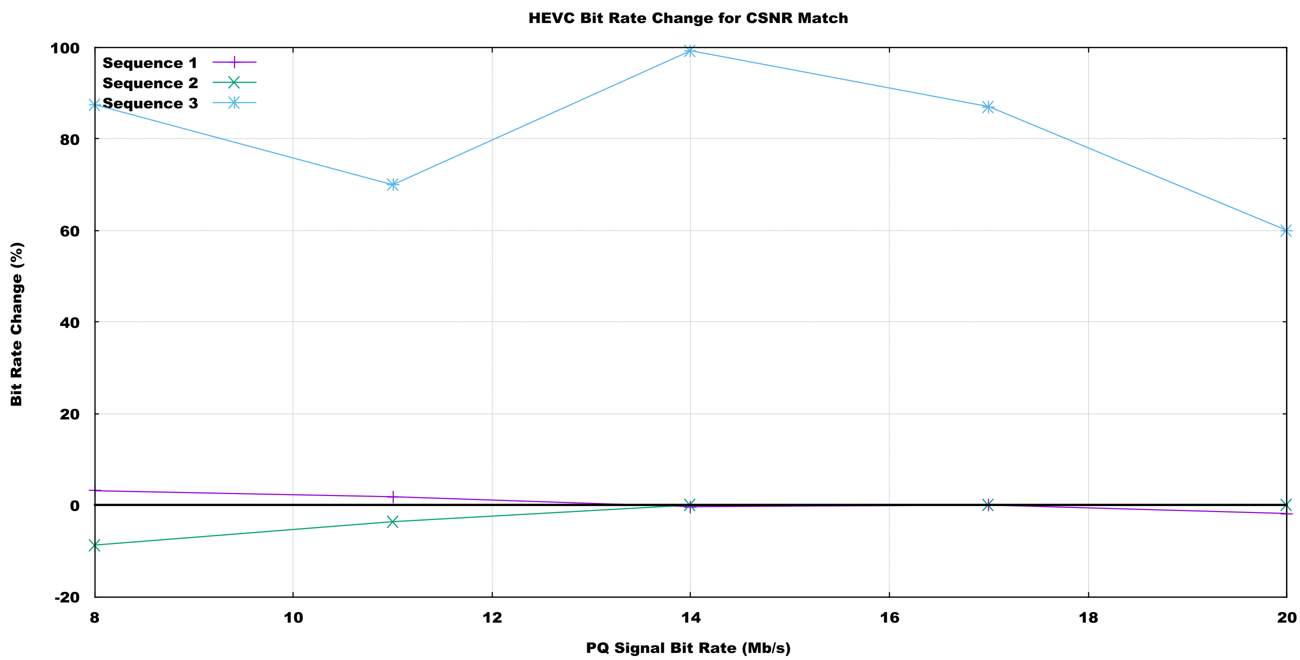

FIGURE 12: HEVC BIT RATE CHANGE FOR MATCHING CSNR

The HEVC results are generally similar to the AVC results. However, there is a much lower spread for Sequences 1 and 2 – the SL-HDR1 impact is almost non-existent, and the bit rate changes are minor.

The Sequence 3 results, though, are very similar – there is a reduction on CSNR that requires a large bit rate change to compensate.

Metadata Bit Rate Discussion

In the previous results, we ignored increase in the video elementary stream bit rate due to the inclusion of the SL-HDR1 metadata. As indicated in section 5.3.5 of [6], the maximum SEI message per frame shall not exceed 251 bytes. Since there is one such message per frame, an upper bound on the metadata bit rate is about 100 kb/s for 50-fps content, and 120 kb/s for 59.94-fps content. This is negligible in relation to the video bit rate; the encoder accommodates this increase by a slight decrease in the NULL packets added to the transport stream.

Conclusions and Future Work

The use of SL-HDR1 produces an SDR signal with metadata that allows reconstruction of the original HDR signal. This makes SL-HDR1 compatible with legacy 10-bit SDR devices that do not understand metadata. The quality impact, as evaluated here, is minimal. Whether the quality increases or decreases is a function of the metric used for quality measurement, which indicates that the changes are not significant.

The authors intend to extend this work to other quality measures and a wider set of content, in order to have a more comprehensive performance evaluation of this technique.

Acknowledgements

We thank Calvin Chen for assistance in running the actual measurements and collecting the PSNR/CSNR data.

References

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

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

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

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

[5] 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

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

[7] OmnitTek, "Picture Quality Analysis on the OminiTek OTM and OTR," pages 34-35, retrieved from https://www.omnitek.tv/wp-content/uploads/2018/07/otxpqa.pdf