

Analyzing 4K Video Quality

Applications and Practices

Executive Summary

Technology leaders have responded to the challenge of 4K video both by embracing new production methods and by validating compatibility with existing workflows. This validation needs to cover all aspects of the production process including capture, encoding and transcoding, storage and delivery. Although many broadcasters still view 4K as an issue to be addressed in the future, most leading organizations have been testing the technology for years since it has had an impact on current engineering practices. Today's 4K testing results are providing valuable input to help guide important business decisions for future generations of services. As 4K systems and products continue to impact each the market, video quality testing and evaluation will continue to be important, both for processing technology performance benchmarks as well as vendor selection.

Introduction

It is widely acknowledged 4K is playing a significant role in all forms of content delivery and for many television industry participants it is viewed as an issue that needs to be addressed in the future. In reality, however, there has been a major shift to 4K technology use in content creation today that affects quality in the broadcast workflow and it will continue to influence new equipment purchase decisions.

A number of different organizations including program originators, broadcasters, and multi-channel video distributors have been testing 4K as part of everyday workflows and its impact to video quality and system performance. Much of the ongoing testing can be distilled to a set of major areas where 4K technology is expected to make a difference in quality. These subject areas, which are described in detail below, require high-performance test systems that are capable of playing 4K uncompressed video test sequences as well as capturing at the same time to then analyze video quality using modern objective measurement techniques.

4K Testing Priorities

In order for engineers to analyze how 4K technology is impacting picture quality across different aspects of their workflows a significant amount of data needs to be compiled to support decisions to select the right codecs (HEVC, H.264, J2K, JPEG XS, or other), the optimal bit rates, the appropriate file/container formats and the best suppliers for contribution, dis-

tribution and archiving functions. Some organizations are also grappling with the tradeoffs between delivering content to consumers as 1080p60 or 2160p60, and defining whether better quality can be achieved for very low bit rates by relying on up-conversion at the display device. The following sections describe workflow methods, measurements and some of the expected trends in quality scores for each major subject area.

Codec Technology Selection

- **Basis for Test** – Choosing suitable compression formats that will meet a range of financial and functional objectives will be a major business decision for many portions of the 4K broadcast chain. At each point from source to viewer, tradeoffs must be made between bandwidth, delay, cost, and video quality that greatly influence the type of compression algorithms selected. These decisions can have significant financial implications, particularly if a large population of subscriber devices needs to be upgraded to support a new codec such as HEVC. To get a clear picture of the technical benefits and limitations of each candidate standard, functional testing with real devices and software is crucial.

To properly evaluate a candidate codec, testing should be performed with several devices across a range of useful bit rates. The results can help determine which compression standard best fits into a given 4K service rollout plan.

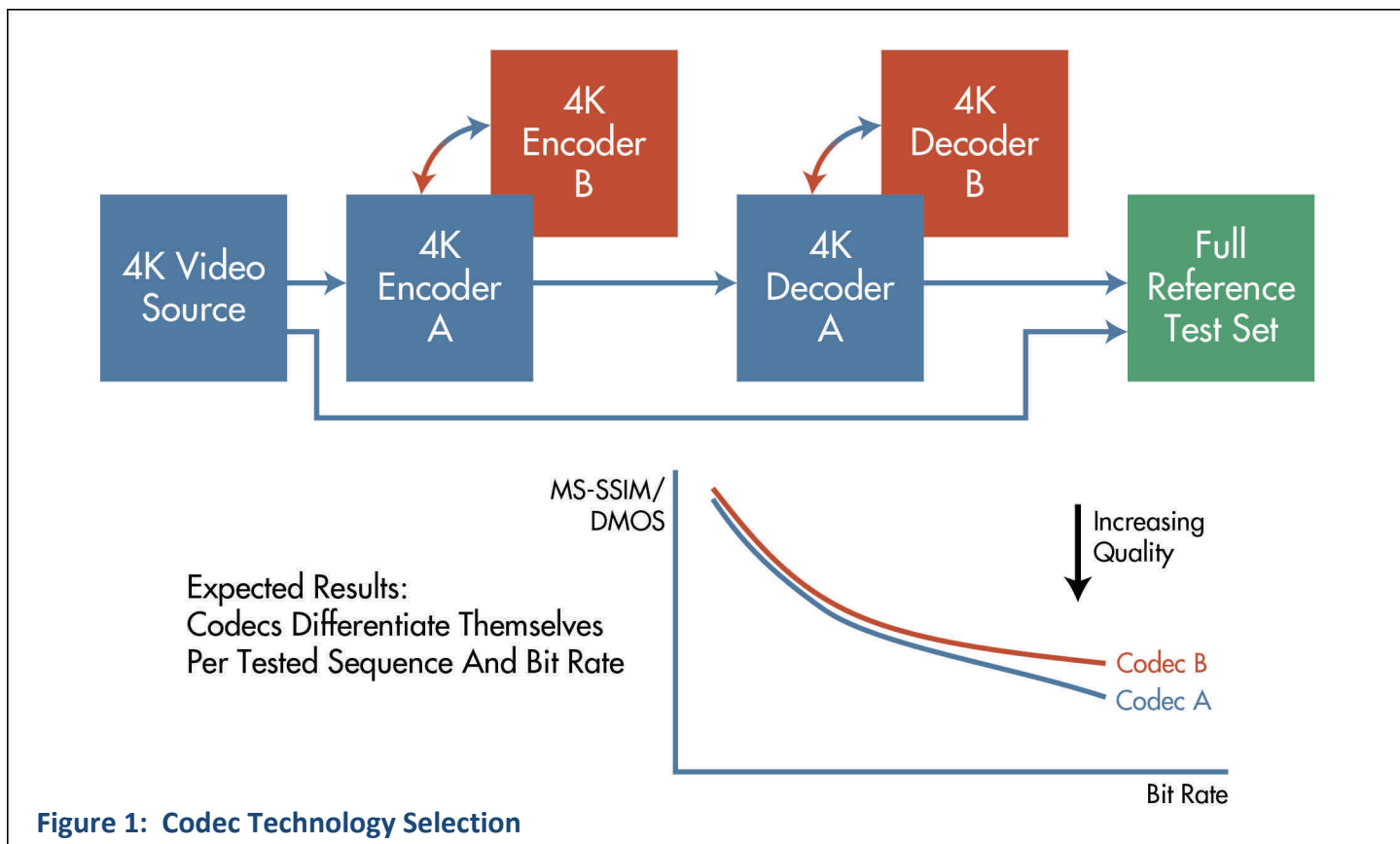


Figure 1: Codec Technology Selection

- Work Flow** – The test setup shown in Figure 1 provides a straightforward platform for comparing various compression systems. A 4K video source supplies uncompressed content to the input of the encoder using a video test sequence that has been chosen to stress the encoding process in some manner. The encoder output or file is then decoded, which recreates the original 4K uncompressed source sequence as closely as possible. This output video sequence is recorded and stored to form one element of a full-reference test sequence set that will be used for picture quality analysis. The test is then repeated at different encoded bit rates, and each captured 4K uncompressed output sequence is added to the full reference test sequence set. More full reference test sequence sets can be created as desired using other source sequences which have different characteristics from the initially chosen video source. After the first codec (A) is evaluated, a second (B) can be inserted into the test work flow, and additional full reference test sequence sets can be produced.
- Measurements** – Image quality trends affecting the selected source content caused by each compression system can be measured in each sequence set using objective perceptual techniques such as MS-SSIM (Multi-Scale Structural Similarity), DMOS (Differential Mean Opinion Score) or JND (Just Noticeable Differences) index as well as subjective evaluation by human viewers. Each of these objective full-reference test measurements will

provide a quality value for each frame of video. These frame-by-frame quality values can be compared directly to the quality values from other elements of the same test sequence set (since they all used the same source sequence) or they can be compared to other sets that have been generated using the same source sequence using another codec. Portions of output sequences where objective measurements vary greatly from one codec to another can be more extensively examined using subjective techniques if desired.

The lower portion of Figure 1 shows estimated trends of MS-SSIM sequence measurement averages as a function of bit rate from two different codecs that used the same source sequence. Note that a decrease in the MS-SSIM score indicates improved video image quality.

Target 4K Bit Rate Designation

- Basis for Test** – Selecting an appropriate bit rate for delivering 4K signals is a billion-dollar question for broadcasters around the globe. Delivery networks will only be cost-effective if the bit rates chosen for 4K delivery allow enough channels to be offered to consumers while still remaining within the capacity limits of the network. For example, a cable provider might require at least two 4K signals within each 6 MHz channel slot. Using 256 QAM that generates a usable bandwidth of 38 Mbps in a channel, each 4K channel would need to consume less than 19 Mbps. Similar decisions are faced by satellite, over-the-air and IPTV providers.

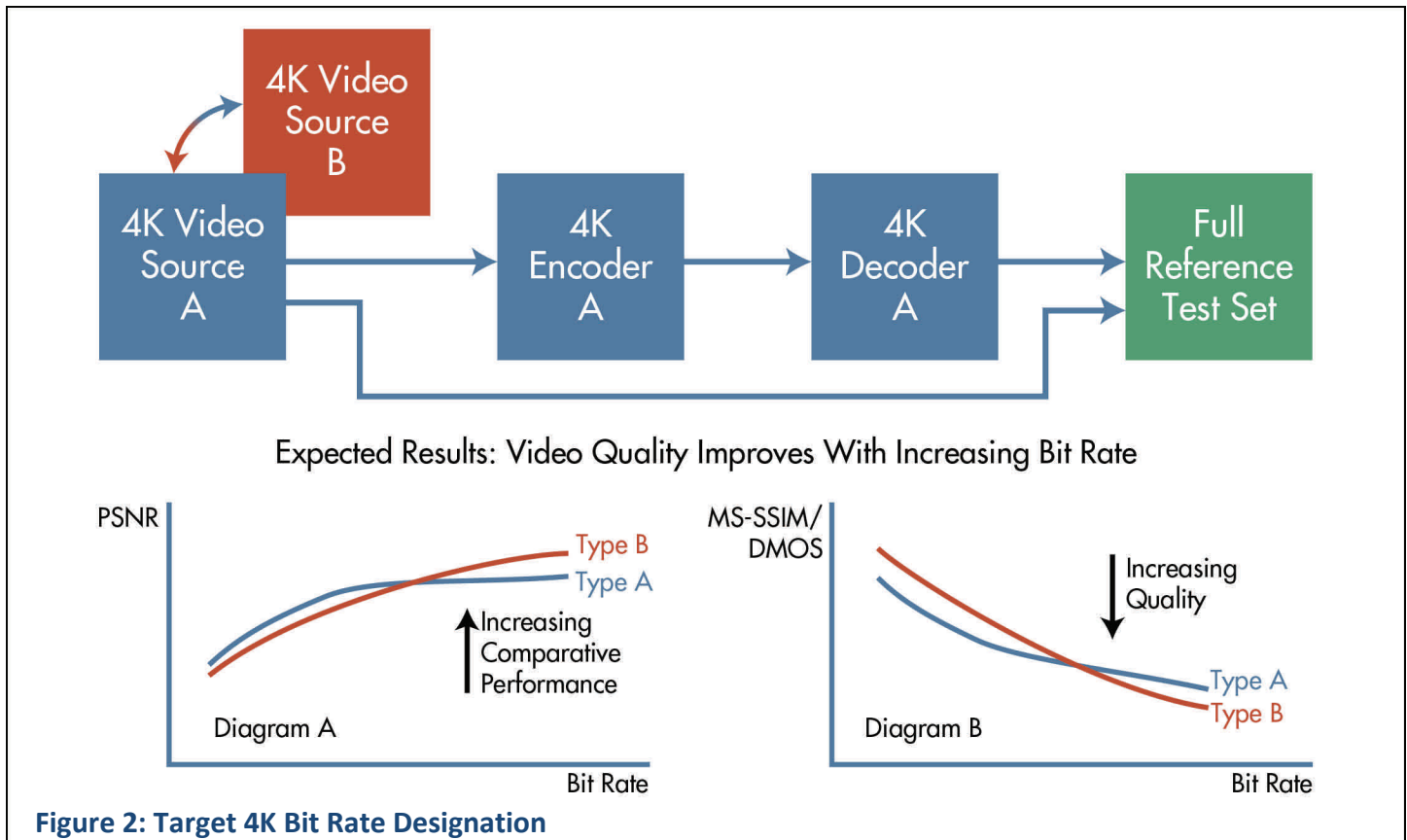


Figure 2: Target 4K Bit Rate Designation

- Work Flow** – Testing can easily be performed to determine an optimal transmission bit rate using the test configuration shown in Figure 2. A 4K video source playing a video source sequence feeds an uncompressed signal to the encoder's input. The compressed output of the encoder can be fed directly to the decoder's input. The 4K output of the decoder captured in real time to form an element of a full-reference test sequence set. Additional outputs are then captured and added to the sequence set by making small step changes in the encoder bit rate. More full reference test sequence sets can be made by using different source content and repeating the process of capturing and storing the decoder output sequences. This work flow can also be performed as a file process in place of real-time play/record functions.
- Measurements** – Several types of measurements can be made using a complete full-reference test sequence set using objective measurement tools such as MS-SSIM, DMOS, and JND. Figure 2 A shows an estimated trend of sequence score averages that would be expected if quality is measured by using PSNR (Peak Signal to Noise Ratio), where better comparative performance through less image degradation is indicated by a higher value. Figure 2 B shows sequence score averages and trend that would be observed if quality is measured using MS-SSIM/DMOS; again, in these indices a lower score indicates better picture quality. For any given source content type some points in the curve of the encode/

decode process may perform differently at different bit rates. This same test configuration can be used to evaluate video quality performance for codecs used for different functions (say contribution vs. distribution) at different points along the broadcast chain.

4K vs. 2K Delivery Comparison

- Basis for Test** – There is currently a question in the industry about the optimal format for delivering content intended to be viewed on 4K devices. On one hand, taking the original 4K content, encoding it, and then decoding in the viewer's home for 4K delivery to a display should provide a good result. On the other hand, down converting a 2160p60 signal to 1080p60 (i.e. 4K to 2K) prior to compression, encoding the signal at 2K resolution, and then decoding it to deliver a 2K signal that is up converted for the consumer's 4K display has also been theorized to deliver a high-quality visual experience, particularly if the viewer is watching from a distance greater than 1.5 picture-heights. Under these circumstances, and if the delivery infrastructure requires very low bit rates, it is possible that the 2K signal may deliver a higher image quality than the 4K signal.

One key question that is under investigation is the crossover point where up/down conversion and 2K compression yields a comparable image to viewers versus a true 4K end-to-end system over a range of different encoded bit rates. At high bit rates, quality is likely to be limited by the performance of the up and down convert-

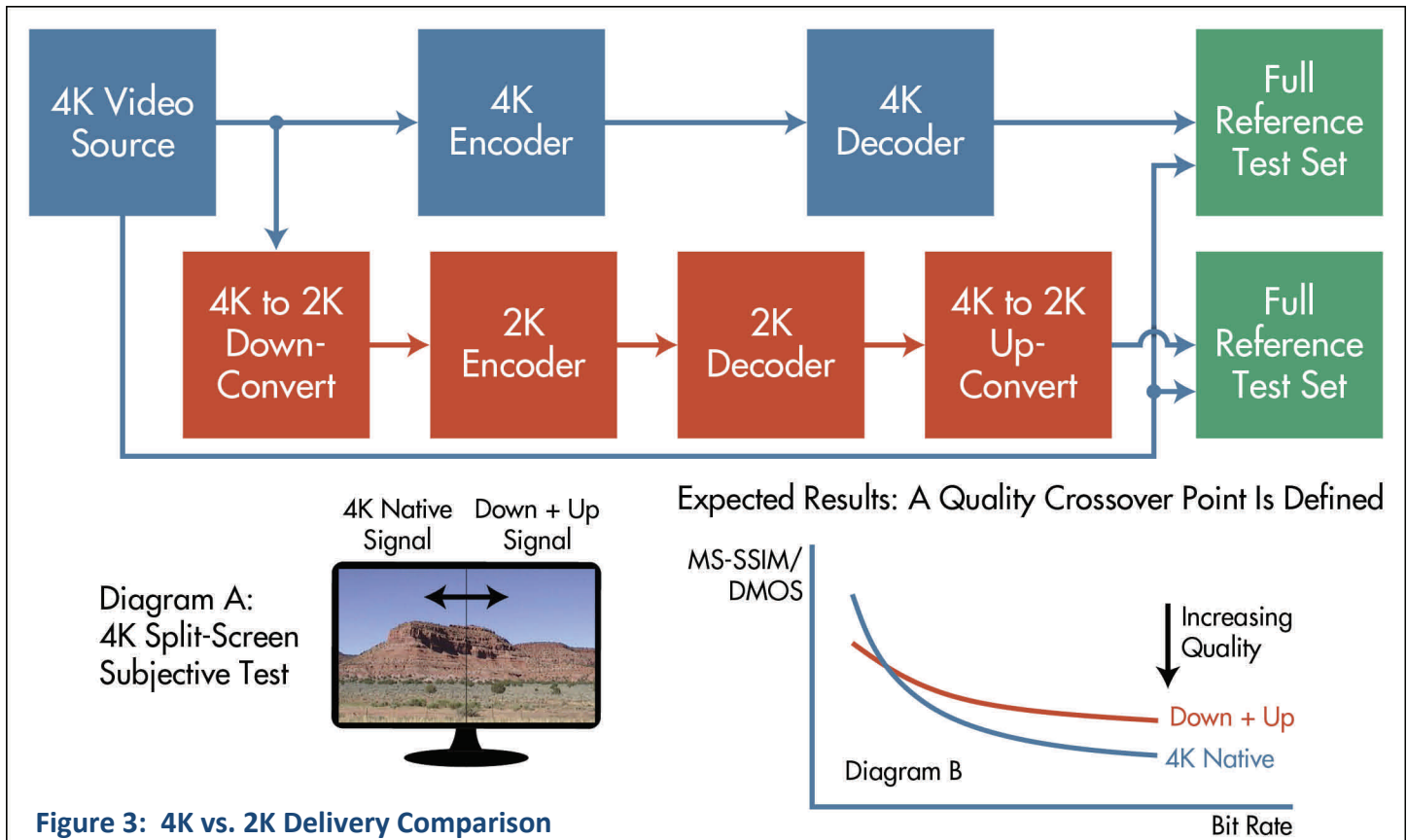


Figure 3: 4K vs. 2K Delivery Comparison

ers, so the 4K native signal would be expected to have better quality. At very low bit rates, the reduced load on the compression system of a distributed 2K signal would be expected to produce comparable picture quality given the viewing criteria discussed earlier.

- Work Flow** – To answer these questions, a test setup consisting of two different signal paths can be used, as shown in Figure 3. One path uses native 4K encoding and decoding. The second path uses a down converter prior to the input to a 2K encoder and an up converter to process the output of the 2K decoder. Note that both sets of equipment are fed with exactly the same signal, and that the decoder outputs are captured in 4K resolution to create a full-reference test sequence set for each path that covers a range of encoded bit rates.
- Measurements** – To make a complete comparison between the 2K and 4K paths, both subjective analysis and MS-SSIM or DMOS measurements can be performed on a full reference test sequence set. Subjective testing should be carried out over a range of different viewing distances from the display; theory predicts that viewing distances beyond 1.5-2 picture heights will make it difficult for viewers to see any differences between 4K images and up converted 2K images.

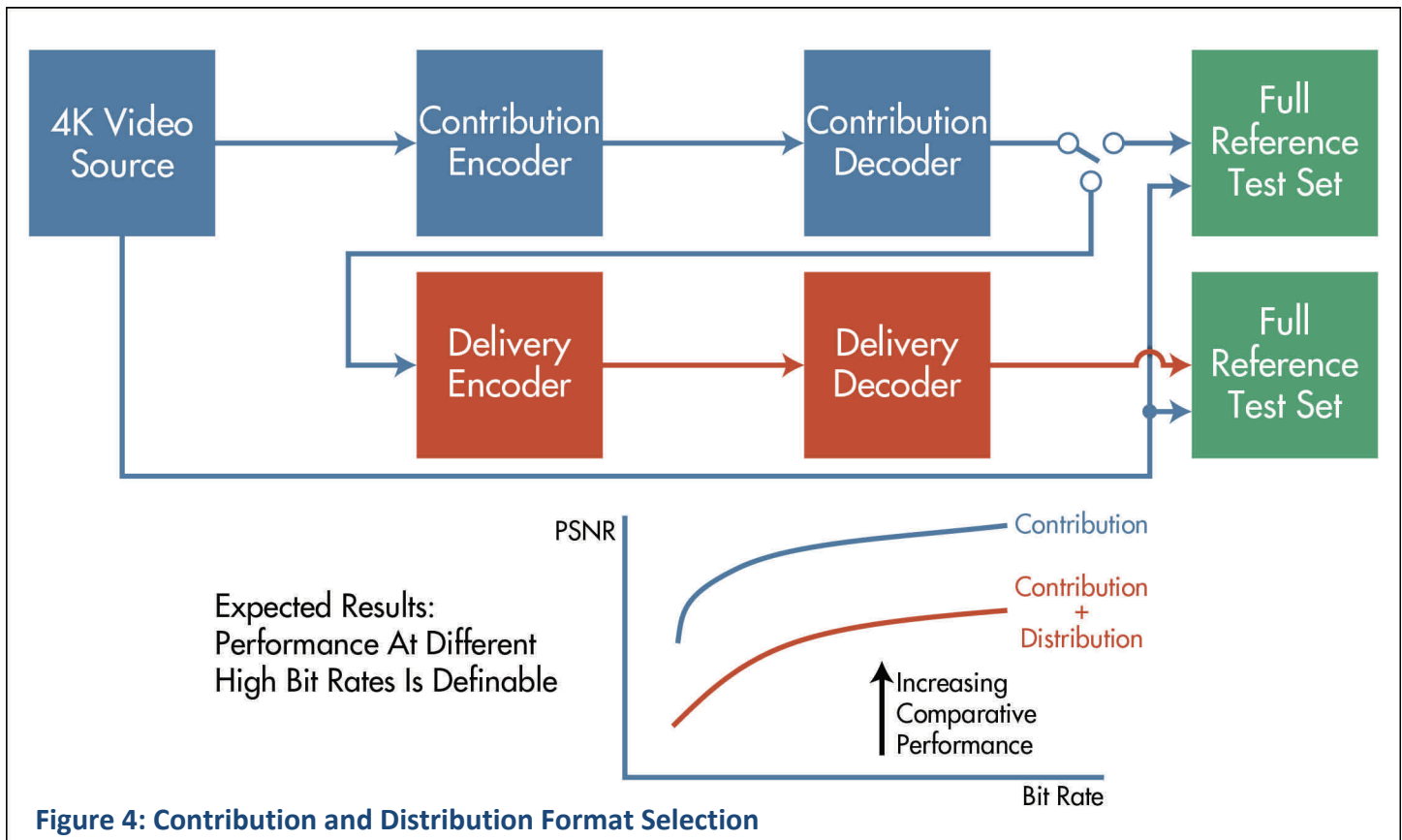
An effective way to perform engineering lab subjective testing and to eliminate differences caused by display to display variation is to use a device, such as a Video Clarity ClearView system, that allows a 4K display

to simultaneously show synchronized, full-resolution portions of two different video images side-by-side, split mirror or other comparative view modes. Typically, the original video source sequence would be compared to the captured output of a decoder, or two decoder output sequence files could be compared to each other. This display arrangement is shown in diagram A of Figure 3.

For objective testing, PSNR may be used to show a trend but MS-SSIM is now the most widely accepted measurement for picture quality on either DMOS or MS-SSIM scales. Plots can be made of the quality of both the up converted 2K and the 4K image quality for each tested bit rate. Figure 3 B shows the general trend that would be expected for sequence average MS-SSIM measurements, where the 4K signal has better performance at higher bit rates, and the signal that is down and up converted to/from 2K is estimated to have better relative quality at some low bit rate threshold.

Contribution and Distribution Format Selection

- Basis for Test** – To form a complete ecosystem using 4K, signals will need to be transmitted from remote venues back to broadcasters. Following current best practices, the formats used for 4K contribution networks are normally much different from those used for delivery to consumers. This is because contribution links have different technical requirements, including lower delay, better support for editing and post production, and



occupying much greater network bandwidths. For example, JPEG2000 with 10-bit, 4:2:2 color sampling is often used for contribution, consuming over 100 Mbps for HD video (as compared to a range of 5 through 19.4 Mbps used for HD delivery to consumers). Decisions about 4K contribution codecs will need to be made by broadcasters in the near future, as more production work migrates to using 4K.

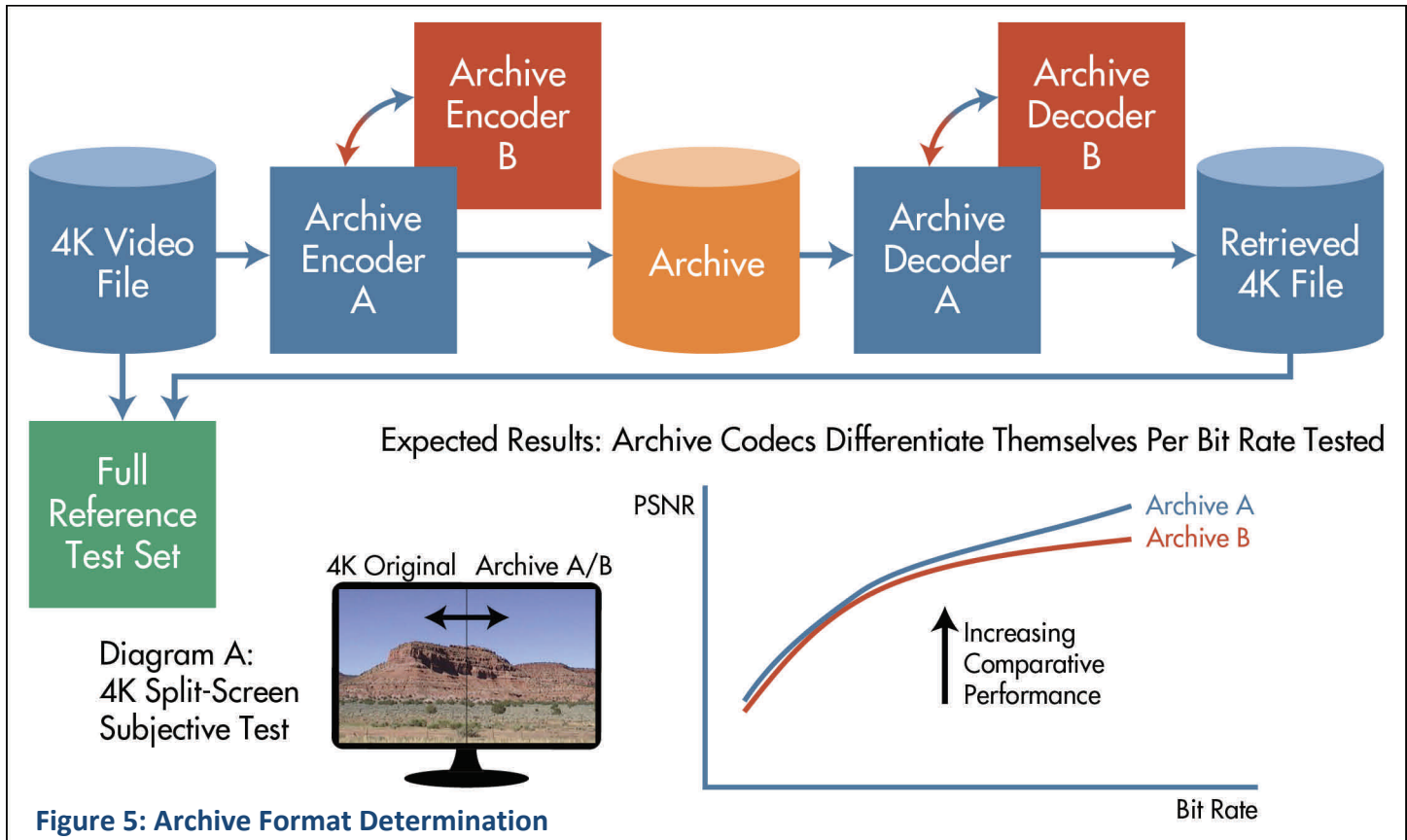
- **Work Flow** – For contribution networks, it makes sense not only to test the effect of compression in the contribution portion of the network, but also to explore what effects, if any, the contribution system has on the signal delivered to the viewer. To that end, the configuration shown in Figure 4 allows the contribution link to be tested in isolation, and also allows the effects of concatenated compression to be tested. This second measurement is achieved by connecting the uncompressed output of the contribution link receiver/decoder to the input of the delivery encoder. Measurements of the end-to-end system can then be made by comparing the captured output of the delivery decoder to the original uncompressed source file.
- **Measurements** – Contribution systems are specifically designed to achieve very high quality levels, and as a result defects are very hard to quantify. The most reliable and repeatable measurement of small image defects is PSNR (Picture Signal to Noise Ratio). Figure 4 shows a

sequence average trend that would be expected for PSNR measurements as a function of bit rate for both the contribution link alone and the cascaded contribution and delivery codecs. To fully assess the combined impact of concatenated contribution and distribution systems, it is also common practice to make perceptual quality measurements including MS-SSIM/DMOS to ensure that the best possible video quality is delivered to viewers.

Archive Format Determination

- **Basis for Test** – Choosing an appropriate archiving codec for 4K material will require tradeoffs between lifecycle storage costs and video image quality. On one hand, since 4K sequences may consume eight times the storage per minute as HD content, using a lossy compression codec could be considered an economic necessity. On the other hand, using lossy compression will have an impact on video quality, potentially reducing the future value of archives. Testing archive systems in advance to select suitable compression formats and bit rates will help ensure that the extra expense and effort expended in capturing 4K source materials are not wasted, particularly for content captured today that are intended to support future rebroadcasting or re-purposing.

Continued on next page >>>



- Work Flow** – Figure 5 shows the basic workflow that can be used for testing and measuring a 4K video archiving process. This test begins with an uncompressed video file, which is then converted into an archive format using the archiving process that is under test. The resulting file is then retrieved from the archive and converted back into an uncompressed video file. Along with the original source file, the retrieved file is analyzed using a full-reference testing methodology that supports picture quality measurement and subjective analysis.
- Measurements** – PSNR is a good metric for evaluating high-quality, high bit rate file compression because it has accuracy and repeatability in detecting even small image and color distortions created per frame in the archive versus an original uncompressed version. Different archiving formats at multiple compression ratios can be effectively compared by using this full-reference testing method. Figure 5 shows the general trend of sequence test averages that would be expected in PSNR measurements as a function of bit rate for archive codecs. A seamless split subjective viewing test of the original versus each archive sequence version is also useful, to ensure that the archive process maintains a high level of visual image quality as evaluated by a human observer.

Conclusion

Although 4K broadcasting is not yet widely deployed due to bandwidth and technology limitations, forward-thinking content originators, television networks and multi-channel video delivery providers are already deeply involved with evaluating new equipment and processes that support this new format. Key decisions about when and if to make the transition to 4K require information about what the technology can deliver, and what performance levels can be achieved. Claims of “4K support” made by various suppliers can be verified well in advance of major purchases. And even if the decision is made to postpone 4K adoption, knowledge gained through the process of exercising and quantifying new equipment should prove valuable as the market for 4K broadcast and content delivery continues to evolve.

For information about Video Clarity’s ClearView Analyzer and Player products that support full-resolution 4K recording, quality measurement and playback today, please visit videoclarity.com.