Keeping Video Quality Pristine throughout the Production Process: Introducing 10-bit 4:2:2 AVC/H.264 encoding

Yasser F. Syed and Pierre Larbier
ATEME (Burbank, USA & Biévres, France)

Abstract
This paper discusses how using the 10-bit 4:2:2 AVC/H.264 encoding profile for transmission can maintain quality throughout the production process workflow. Upto now an MPEG-2 4:2:2 8-bit profile was used, the new AVC/H.264 encoding profile will allow for significantly lower transmission bandwidth or higher quality for the same transmission links. The profile is also compatible with pre-existing native SDI signal infrastructures used already in production processes. Furthermore 4:2:2 10-bit processing can help to preserve quality during multi-generation encodes/decodes that can happen often in the production process. This paper presents empirical results using real-time implementations to demonstrate the advantages of using this profile. Examples are also presented where this new technology can assist in production workflows ranging from remote uplinks of live broadcasts, non-collocated collaboration of dailies, support of playout-to-tape infrastructures, to international distribution & editing of content.

Introduction
The production process has been a collective work of many individuals effort ranging from producing dailies for film shoots, film scanning, editing of live remote event material, CGI production, and final assembly of the work [9]. Increasingly it is becoming a collaborative effort of non co-located groups (interstudio) involved in these processes starting from not only film but from a full digital content workflow, originating with camera capture to eventual re-versioning of the content. Transmitting, storing, editing video in as pristine a quality as possible in this workflow while addressing infrastructure limitations is critical. It becomes especially evident when considering future and transitional infrastructures that will need to be supported.
Keeping video quality pristine throughout the production process: Introducing 10-bit 4:2:2 AVC/H.264 encoding

4:2:2 10-bit processing is already the de-facto standard for professional video because of the way it is captured and transmitted over SDI (Serial Digital Interface). The entire production chain (film scanning, video editing, archiving etc.) uses at least 10-bit signals. Additionally there is already a significant amount of hardware equipment (tape decks, playout servers, file storage) that is capable of handling 4:2:2 10-bit content.

However when transmitting video within an internal system or across a network, encoders and decoders are still limited to 8-bit, usually with 4:2:0 chroma sub-sampling, in the same manner as with consumer video. The result is that when transmitting video from one point to another, picture information can get lost and quality can suffer.

Until recently, the best quality transmission encode that could be achieved was in MPEG-2 using the 4:2:2 8-bit profile at relatively high distribution bit rates, usually between 20-60 Mbps, with mid-to low latency, between 1s to 250 ms. For more than 10 years, this MPEG-2 4:2:2 profile has been used in production and contribution applications [5]. This sub-sampling scheme was motivated by the reduction of chroma artifacts in multi-generation environments [6].

Video coding has improved much from a decade ago and efforts to capture this in a new video coding standard were already underway. From its early design stages, AVC/H.264 has been perceived as an improved replacement of MPEG-2. All features available in MPEG-2 were included with additional improvements. The initial realization of the AVC/H.264 standard in products was more towards the consumer video market [1]. As a result of this first foray, the majority of today’s AVC/H.264 encoders and decoders are limited to relatively low bit-rates and lack specific tools mandated by production and contribution applications.

As illustrated in Figure 1, most of today’s AVC/H.264 broadcast contribution systems are based on existing distribution encoders and decoders. Using these encoders and decoders for production and contribution purposes has limitations. Since they can only handle High Profile, the encoders must downscale to 4:2:0 8-bit and the decoders must upscale back to 4:2:2 10-bit. The result is that when transmitting video from one point to another, picture information can get lost and quality can suffer. Furthermore distribution encoders are limited to less than 30Mbps, which impedes the highest video quality applications in HD.
Keeping video quality pristine throughout the production process:
Introducing 10-bit 4:2:2 AVC/H.264 encoding

As the technology matures, products aiming specifically for production and contribution applications are either already available or about to be released. They all implement the High Profile, a superset of the High 4:2:2 Profile with two new tools designed to avoid the scaling stages shown in Figure 1:

- 4:2:2 processing
- Up to 10-bit pixel bit-depth handling

Along with algorithmic advances, it provides significantly better video quality with optimization of preserving quality in multi-generation encoding environments [1].

This paper demonstrates the advantages of processing video in its native SDI format using AVC/H.264 4:2:2 10-bit encoding. Instead of using the AVC/H.264 reference encoder [2], this paper presents empirical results from an existing real-time implementation obtained with ATEME Kyrion CM4101 encoders (12-bit and ultra-high bitrate evaluations were done with a non-realtime version of this encoder).

Why 4:2:2 video compression?

AVC/H.264 profiles below the High 4:2:2 Profile process the video as 4:2:0. Since the SDI links transport 4:2:2 signals, chroma components need to be sub-sampled vertically prior to encoding and then upsampled after encoding. The drawback of this sub-sampling process is an overall reduction of chroma detail. In consumer video applications, this is not usually a problem since the human eye is not very sensitive to color information.

Unfortunately, the AVC/H.264 standard does not precisely define how the chroma sub-sampling or upsampling has to be performed, leaving this to encoder and decoder manufacturers [1]. Thus there can be an introduction of mismatch between the down-sampling filter in the encoder and the up-sampling filter in the decoder. Besides, wrong interpretation of the progressive or interlaced nature of the video can also lead to faulty decoding of whole chroma planes.
Keeping video quality pristine throughout the production process: Introducing 10-bit 4:2:2 AVC/H.264 encoding

Whether it pertains to production or contribution applications, video quality has to be kept to the highest possible level for each generation in order to handle several encoding-decoding steps. A mismatch in the chroma sampling can introduce color degradations that worsen with each generation even though the original mismatch introduced from the 1st generation of the encode/decode stage may go unnoticed to the visual eye.

After a few encoding-decoding stages, the most common issues are:

- Color bleeding
- Loss of color contrast and details
- Chroma displacement relative to luma
- Creation of interlaced (or progressive) color artifacts on progressive (respectively interlaced) pictures

Figures 2 and 3 give an example of such problems after only five generations.

The solution is 4:2:2 compression

The only way to avoid those artifacts is to process the video in its original color format (4:2:2). This is possible using the AVC/H.264 High 4:2:2 Profile.

The drawbacks in encoding 4:2:2 include a moderate bit-rate increase (for a given quantizer) relative to 4:2:0 encoding. Interestingly through, this bit-rate increase does not lead to a loss of video quality. In fact,
Keeping video quality pristine throughout the production process:
Introducing 10-bit 4:2:2 AVC/H.264 encoding

the perceived quality at the same bit rates is roughly the same except at very high bit-rates where 4:2:2
processing performs slightly better than 4:2:0.

Therefore processing video in 4:2:2 does not exhibit technical disadvantages and can help to avoid
annoying chroma artifacts seen in cascaded encoding-decoding configurations.

Why 10-bit video compression?
Being able to encode pixels directly using a bit-depth above 8-bit is a feature provided by all AVC/H.264
profiles above High Profile, for instance High 10 Profile and High 4:2:2 Profile [1].

The bit-depth increase provides greater accuracy for the miscellaneous prediction processes involved in the
AVC/H.264 compression scheme, including motion compensation, intra-prediction, and in-loop filtering
[4]. Extensive experimentation demonstrates that the coding efficiency gains are highest with videos that
contain shallow textures and low noise. But as shown in Figure 4 and 5, there are also gains to be had with
relatively noisy and textured standard sequences.

![Figure 4](image1.png)
![Figure 5](image2.png)

**Figures 4/5 - Coding efficiency gain, more than 8-bit coding**

These curves clearly show that the gain is smaller as the bit-rate is reduced, but still remains significant.
Interestingly, the PSNR improvements are in the same range of what is achieved with common tools like
8x8 transforms or multiple references.

The PSNR increase that can be achieved using 10-bit encoding is more than **1dB** on some natural
sequences and we measured an average of **0.25dB** at 60Mbps on a varied test set of broadcast HD
sequences. This translates to an average bit-rate saving in the range from 5% to about 20%, while retaining
Keeping video quality pristine throughout the production process: Introducing 10-bit 4:2:2 AVC/H.264 encoding

the same video quality.

Testing shows that increasing the bit depth to 12-bit (or even 14-bit) does provide a much smaller coding efficiency gain (up to 1% in bit rate savings) with no quality loss over 8-bit or 10-bit encodes.

Lastly, there is no relation between 10-bit encoding and the frame format: the advantages are the same whether the source video is HD, SD, progressive or interlaced.

Beyond coding efficiency improvements

One noteworthy aspect of 10-bit processing is that it provides perceivable gains in the reduction of three kinds of artifacts:

- Contouring or banding
- Smearing
- Mosquito noise

This gives a better quality aspect to plain surfaces and shallow textured areas (smoke, clouds, sky, sunset etc.) as it slightly improves object edges. The following figures show an example of the improvements that are achieved on ordinary sequences:

![Contouring](Figure 6 – 8-bit coding) ![Mosquito Noise](Figure 7 – 10-bit coding)

With 10-bit processing, the coding gains are provided through increased accuracy of internal computations where improvements can also be observed on 8-bit video sources. Interestingly, the reduction of artifacts provided by 10-bit processing does not require a 10-bit display. It’s perceivable even on standard LCD panels (8-bit or dithered 6-bit)

Given that high bit-rates benefit most from using 10-bit compression, production and contribution
Keeping video quality pristine throughout the production process: Introducing 10-bit 4:2:2 AVC/H.264 encoding

applications are the best candidates for using this tool. Furthermore, it gives the opportunity to keep the original pixel bit-depth all along the processing chain as it avoids scaling from 10-bit to 8-bit at the encoder input and back to 10-bit at the decoder output.

About GOP structures

Choosing the codec and the appropriate tools is only one facet of the compressed chain configuration. Another critical aspect is the GOP structure since it can have a major impact on video quality. But Contribution and Production application may have specific constraints that prevent the use of some GOP structures. The most important are:

- Low end-to-end latency: this avoid the possibility of using B-pictures
- Low recovery time in case of transmission errors which implies a low GOP size
- Easy stream editing capabilities: which could motivate the use of I-picture only coding

Regarding GOP structure, AVC/H.264 offers more options than MPEG-2. And most modern AVC/H.264 encoders provide the ability to analyze in advance the video source to adapt the picture types (I, B, P) on the fly. This permits maximizing video quality at the expense of only a small look-ahead.

Major additions provided by AVC/H.264 are:

- Multi-references pictures: each picture can be used as a reference for several other pictures. This is particularly beneficial to videos that exhibit small global motion like camera vibrations.
- Generalization of the reference concept: each picture can be chosen as being a reference picture or not regardless of its type (I, P or B). For instance, a single picture with a flash could be coded as an I-picture but should then not be used as a reference to the next pictures, because there is no temporal relation with the surrounding pictures.
- Hierarchical GOP structure: B-pictures may be used as reference on a hierarchical basis to improve coding efficiency. It should be noted that this capability is a consequence of the two previous new features of AVC/H.264.

Therefore, a tradeoff must be found between video quality provided by an adaptive GOP structure and broadcasters operational constraints.

Figure 8 illustrates the effects of various GOP structure on coding efficiency. Four configurations were tested using the High422 profile:
Keeping video quality pristine throughout the production process: Introducing 10-bit 4:2:2 AVC/H.264 encoding

- Intra-only pictures
- AVC-I 100: intra-only coding but using CAVLC and a high slice count
- Short GOP of 4 frames with a hierarchy of 3 B-frames
- Adaptive Long GOP structure with a maximum size of 25 frames

![Figure 8: Importance of GOP structure on coding efficiency](image)

On this relatively simple but noisy sequence, the measured results show that the best GOP configuration is the *Adaptive Long GOP* while *Intra-only* is outperformed by about 50% even at 100Mbps. It should be noted that this measurement illustrates why MPEG-2 in long GOP configuration might perform as well than *Intra-only* at the same rate.

Interestingly; the *Short GOP* configuration exhibits only a small loss compared the *Adaptive Long GOP* at Contribution rates. This GOP structure could then serve an efficient alternative to Intra-only coding when very short GOP sizes are required.

In addition, these measurements don’t reflect important subjective observations such as:
- Annoying picture instabilities when Intra-only coding is involved
- Devastating defects on scene-changes when the GOP size is fixed
- Important quality degradation when B-frame coding are imposed on sequences where the temporal correlation between pictures is small

Consequently, Contribution and Production applications should choose the carefully the GOP structure, avoiding fixed and Intra-only configuration, letting the encoder make the picture type decision whenever possible.
Advantages of 4:2:2 10-bit in multi-generations

On major constraint of Contribution applications is the need to re-encode the stream several times before being delivered. Since video compression is a lossy process, each generation adds degradations. The challenge is then to keep this video quality decrease at a minimal level.

As it has been shown before, 4:2:2 chroma sub-sampling limits the most significant artifacts that are produced by a mismatch between the encoder and decoder. Because internal rounding errors are significantly reduced, 10-bit coding also helps visibly to limit the quality drop among multiple generation. Those two advantages make the AVC/H.264 High 422 Profile a candidate of choice when multiple generations are required.

But the most important cause of degradation in a multiple generation process is caused by a GOP misalignment across consecutive generations. If the GOPs are not aligned at each generation, the video quality drops continuously each time the video is re-encoded. On the other hand, if the GOPs are aligned, the degradation can be limited. This effect is illustrated by Figure 9 and 10.

The measurements were performed using three different GOP structures:

- Intra-only pictures
- Short GOP of 4 frames with a hierarchy of 3 B-frames
- Adaptive Long GOP structure with a maximum size of 25 frames

These graphs show that aligned GOP structure permit to preserve the same quality advantage of Adaptive
Keeping video quality pristine throughout the production process:
Introducing 10-bit 4:2:2 AVC/H.264 encoding

Long GOP over Intra-Only. Interestingly, this is not the case when GOPs are not aligned in consecutive re-encodes. For instance, the PSNR drop is less than 3dB after the 7th generation if GOPs are aligned while it’s more than 5dB when GOP are not aligned.

It shall be noted that aligning GOP across several generation does not necessarily require transmitting additional information when the GOP structure is adaptive. If the same algorithm is used to adapt the GOP structures, they will be aligned after the first scene-change. This is achieved for instance by using products by the same vendor.

High 4:2:2 Profile fits all contribution needs
As seen before, processing 4:2:2 10-bit pixels provides the best possible quality and reduces degradations in a multi-generation environment. This capability is offered by the AVC/H.264 High 4:2:2 Profile, designed specifically for production and contribution applications.

This profile works well in production and contribution environments because it enables very high maximum bit-rates for the Video Coding layer (VCL) [1]:
- 525i and 576i (level 3): 40 Mbps
- 720p and 1080i25/30 (level 4.1): 200 Mbps
- 1080p50/60 (Level 4.2): 200 Mbps

HD encoding at around 50 Mbps provides quasi-transparency for the vast majority of Broadcast Contents. However, measurements show that up to 150 Mbps (35 Mbps in SD) might be needed to achieve 43dB which is a common definition of true transparency. Since the High 4:2:2 Profile can even surpass these extremely high bit-rates, it can cover the full range of production and contribution applications, including those that require advanced archiving and mezzanine format support.

AVC/H.264 outperforms MPEG-2

Today, HD contribution is mostly performed using MPEG-2 with 422P@HL. This profile offers 4:2:2 processing but is limited to 8-bit pixel components bit-depth [5]. As illustrated by figure 11, AVC/H.264 High 4:2:2 Profile offers important savings when compared to MPEG-2, even at the highest bit-rates.
Keeping video quality pristine throughout the production process:
Introducing 10-bit 4:2:2 AVC/H.264 encoding

Objective and subjective measurements allow us to draw some conclusions:

- As it’s well known, AVC/H.264 offers a bit-rate gain of roughly 50% at below 15Mbps.

- Above 30Mbps, AVC/H.264 produces results comparable in quality to MPEG-2 with at least 20Mbps increase. For instance, MPEG-2 quality at 60Mbps is achieved by AVC/H.264 at only 40Mbps or less.

- Above the 50Mbps mark, the quality provided by AVC/H.264 increases linearly with the rate. This indicates that most of the encoder “effort” is devoted to coding non-redundant information like noise. Since the human eye is not very sensitive to noise fidelity, this explains why most sequences look quasi-transparent above this rate.

**Uses in Production**

Transmission of video in a 4:2:2 10-bit format keeps the video quality from production efforts at its highest level possible and is compatible with SDI signal infrastructures. With AVC/H.264 High 4:2:2 profile, this can be done at lower bandwidths than the MPEG-2 4:2:2P and with 10-bit precision instead of 8-bit.

With bandwidth compression advantages and multi-generation encoding benefits, the AVC/H.264 4:2:2 profiles can be used in many places in the production and contribution environments.

**Remote Truck Uplinks**

For live sporting events, it is critical to maintain quality especially with high motion content. But at the same time, there are limitations due to being at a remote site. There are needs to maintain low latency for live events, to economize on space for equipment on the truck, and to integrate to the physical transmission uplink infrastructure that is available at the time for sending the signal back to the studio.
Presently at the remote truck site, SDI signals can be maintained for production purposes but transmission and uplink restrictions can degrade the quality of the signal transferred back to the main studio site [8]. Switching to an encoding system, that can maintain SDI quality through 4:2:2 10-bit at lower bit rates or increasing quality when bandwidth is available, helps to preserve the source quality of the video as much as possible back to the studio.

At the studio, the signal can be decoded and integrated back to the studio SDI infrastructure for more production operations or to be transcoded to consumer grade video. The generation of encode/decodes from a remote site is at least 1 but can often be 2-3 generations before it is viewed.

**Playout to Tape Infrastructure**

Content may be created at a main site, but may need some further production at local sites. This can happen quite often when content is prepared for a different country and in particular for TV episodic material. It is easier to transmit content through a satellite uplink to many sites at once and it reduces delay risks due to multiple country custom issues in transporting the content using multiple tapes or drives. Upto now real-time transmission quality could not be send at the same quality that could be received on tape. With 4:2:2 10-bit transmission, this signal can be received at a quality level similar to tape and be immediately integrated to the local SDI signaling infrastructure. For those sites still operating on a legacy tape infrastructure workflow, the received and decoded 4:2:2-10 bit signals could be again mastered to tape with little or no quality degradation. The need for local production also adds the need to decode and reencode content. With the maintainance of at least 4:2:2 10-bit throughout the workflow, the effects of multi-generation encoding can be minimized.

**Remote Collaboration**

When dealing with a video or film work, content in production often needs to be looked at by multiple people and often at the same time (e.g. digital dailies). This required in the past that people would need to be co-located at the same site. With increasing bandwidth being available between sites, this opens up the possibility of doing this remotely. Initially this was limited to viewing low bit rate 4:2:0 8-bit transcoded outputs in a non-realtime format. Increasing the quality of what can be seen remotely through 4:2:2 10-bit streaming and using the AVC/H.264 4:2:2 profile to lower bit rates can extend further what can be done
Keeping video quality pristine throughout the production process:
Introducing 10-bit 4:2:2 AVC/H.264 encoding through remote collaboration.

Multiple Encode/Decode Generations

Each time content needs to be brought into different production tools, this can create a new generation of encode/decode especially when tools are not co-located on the same platform. At different parts of the workflow there may be additional needs that will require a possible encode/decode generation. Some examples of this could be logo insertion, watermarking, and fingerprinting. The requirement to maintain better chroma subsampling and bit depth precision is critical to maintaining quality throughout multi-generation encodes in the workflow.

Summary

This paper presented the advantages of using AVC/H.264 **High 4:2:2 Profile** for Production and Contribution applications. Exploiting all the available tools within this profile, namely **4:2:2 10-bit coding**, allows us to fulfill three highly desired features:

- Processing the source video in its original format
- Enable even the most demanding applications both in terms of quality and rate.
- Offer a significant gain in quality and/or rate over existing solutions

It has been shown that 4:2:2, 10-bit or the combination of the two, will always present a gain over AVC/H.264 **High Profile** and MPEG-2 422P, as all subjective and objective measurements exhibit a quality increase for the same bit-rate. This gives the opportunity to either:

- Significantly lower transmission costs, keeping the same visual quality - OR -
- Greatly improve the video quality using existing transmission links

References


Keeping video quality pristine throughout the production process:
Introducing 10-bit 4:2:2 AVC/H.264 encoding