

Using 10-bit AVC/H.264 Encoding with 4:2:2 for Broadcast Contribution

Pierre Larbier
ATEME, Bièvres, France

Abstract

Until now, AVC/H.264 was primarily used at low bit-rates for distribution applications. Its 50% efficiency gains over MPEG-2 allows it to increase the channel density, reach wider distances and reduce the transmission costs. However, as HDTV and digital cinema hold, there is a growing need for production and contribution applications with higher standards of video quality.

4:2:2 10-bit is already the de-facto standard for professional video because it is the way it is captured and transmitted over SDI (Serial Digital Interface). The entire production chain (film scan, video edition, archiving etc.) uses at least 10-bit signals. But when it comes to broadcast contribution, encoders and decoders are still limited to 8-bit, usually with 4:2:0 chroma sub sampling, just 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.

This paper demonstrates the advantages of processing video in its native SDI format using AVC/H.264 4:2:2 10-bit encoding. Maintaining the encoding and decoding stages at 10 bits increases overall picture quality, even when scaled up 8-bit source video is used. 10-bit video processing improves low textured areas and significantly reduces contouring artifacts.

The results presented in this paper were obtained either from ATEME current real-time HD encoders or bit-accurate software models of upcoming real-time products. Comparisons are made over a very wide range of bit-rates in order to illustrate the achieved gains within a great variety of applications.

Introduction

There are multiple High Definition contribution applications sharing common characteristics:

- Relatively high bit-rates: usually between 20Mbps to 60Mbps and sometimes more.

- Low to moderate end-to-end latency: typically less than 1s down to 250ms.
- The need to take into account the fact that the video may be decoded and re-encoded several times before reaching the end customer.

For more than 10 years, 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.

From its early design stages, AVC/H.264 [1] was perceived as an improved replacement of MPEG-2. All features available in MPEG-2 were included, with the notable exception of a simple trans-rating process. 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. Since they can only handle **High Profile**, the encoders must downscale to 4:2:0 8-bit and the decoders must upscale back to

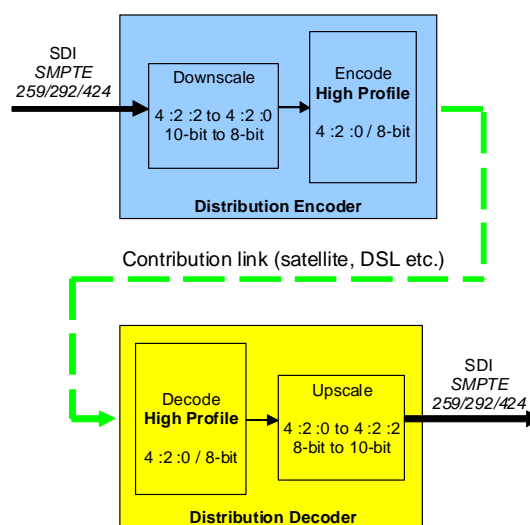


Figure 1 - Architecture with Distribution encoder/decoder

4:2:2 10-bit. Furthermore distribution encoders are limited to less than 30Mbps, which impedes the highest video quality applications in HD.

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

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

Along with algorithmic advances, it provides specific features:

- Significantly better video quality, up to visual transparency through higher attainable bit-rates.
- Optimize video quality in multi-generation

Evaluating AVC/H.264 benefits for contribution

Instead of using the AVC/H.264 reference encoder [2], this paper presents implementation results obtained with ATEME encoders:

- AVC/H.264 4:2:2 8-bit measurements were performed using the already available Kyrion CM3101 contribution encoder.
- AVC/H.264 4:2:2 10-bit measurements were performed using the bit-accurate software model of the upcoming real-time HD encoder, the Kyrion CM4101 (12 and 14-bit evaluations were done with an early version of this model).
- MPEG-2 measurements were performed with a software library included in the latest version of Kyrion File Encoder. All comparisons made so far show that it out-performs available real-time products, which is anticipated since it's slow and brute-force oriented. Considering the algorithms used, this encoder should obtain similar results as in [3]

This method has the great advantage of showing actual results instead of theoretical upper bounds that could never be obtained in real-time.

Using PSNR metrics to evaluate quality

PSNR (peak Signal to Noise ratio) is a commonly used metric to measure the difference between the source and the decoded pictures of a video sequence.

It is a well known fact that PSNR does not correlate well with the human visual perception. For instance, with the same PSNR of 30dB, one sequence could look very good, while another would be visually very poor.

Thus two PSNR measurements on two different sequences are almost meaningless when it comes to video quality.

However, two PSNR measurements on the *same* sequence performed with *PSNR optimized* configurations tell a lot about the relative potential of two encoders (or two different encoding conditions with a single encoder). In this case the encoder capable of providing the highest PSNR will also be able to provide the best quality video results. Indeed a higher coding efficiency gives room for visual enhancements that will ultimately lead to better quality.

For this reason, the PSNR metric is an invaluable tool for evaluating the gains achieved with various tools or for optimizing an encoder.

When evaluating coding efficiency, it is customary to use the PSNR of the luma component only. If chroma has to be taken into account, a combined PSNR metric is often used:

$$\text{Combined}_{\text{PSNR}} = 0.8 * Y_{\text{PSNR}} + 0.1 * U_{\text{PSNR}} + 0.1 * V_{\text{PSNR}}$$

The weighting factors might be questionable but our experience tells us that this metric can give a fairly good idea of the overall coding efficiency while maintaining the importance of the chroma.

Encoder configurations

The AVC/H.264 encoders were configured either in **High Profile** or **High 4:2:2 Profile** using the following tools:

- Inter prediction modes 16x16, 16x8, 8x16, 8x8
- Intra prediction modes 16x16, 8x8 and 4x4
- Adaptive GOP structure with at most 3 consecutive B-frames
- At most 2 frame or 4 field references
- Maximum GOP size of 1s
- MBAFF and PAFF coding
- In-loop filtering
- CABAC
- Fixed quantizer or CBR with a CPB duration of 1s
- Fixed chroma delta quantizers

Adaptive quantization and scaling matrixes were disabled along with other non-normative algorithms aimed at improving visual quality at the expense of PSNR.

Why 4:2:2 video compression?

AVC/H.264 profiles below the **High 4:2:2 Profile** processes 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 up-sampled after decoding. This 25% reduction in information was originally intended to:

- Simplify encoder and decoder designs
- Lower the bit-rate needed to transmit compressed video

The drawback of this sub-sampling process is an overall reduction of chroma detail. However, this is usually not a problem since the human eye is not very sensitive to color information.

Even though the AVC/H.264 standard allows six possible locations for the chroma samples relative to the luma samples, only the standard MPEG location is widely used. As shown Figure 2, two schemes are available to handle progressive and interlaced sources:

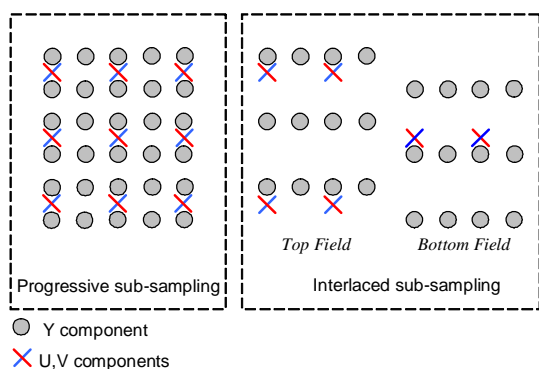


Figure 2 - Chroma 4:2:0 sub-sampling locations

Artifacts introduced by 4:2:2 ↔ 4:2:0 conversions

Unfortunately, the AVC/H.264 standard does not precisely define how the chroma sub-sampling or up-sampling has to be performed, leaving this to encoder and decoder manufacturers. Thus there can be a mismatch between the down-sampling filter in the encoder and the up-sampling one in the decoder. Besides, misinterpretation of the progressive or interlaced nature of the video can lead to faulty decoding of whole chroma planes.

In distribution applications, these problems are usually of secondary importance since the low bit-rate of the video can introduce even more bothersome artifacts.

Whether it pertains to production or contribution applications, video quality has to be kept to the highest possible level in order to handle several encoding-decoding steps. A mismatch in the chroma sampling can introduce color degradations that worsen with each generation.

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

It has to be noted that an interlaced (alternatively progressive) chroma artifact might confuse encoders in the cascading process which in turn significantly reduces their coding efficiency. Therefore, chroma issues may also induce degradation in luma.

Figures 3 and 4 give an example of such problems after only five generations. The only introduced defect was a mismatch in the chroma resampling filters: polyphase bicubic down-sampler before encoding, simple tent up-sampler (with an incorrect phase) after decoding.



Figure 3 - Source picture (Mobile & Calendar)



Figure 4 - After five 4:2:2 ↔ 4:2:0 conversions

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**.

As illustrated in Figure 5, the drawbacks in encoding 4:2:2 include a moderate bit-rate increase (for a given quantizer) relative to 4:2:0 encoding.

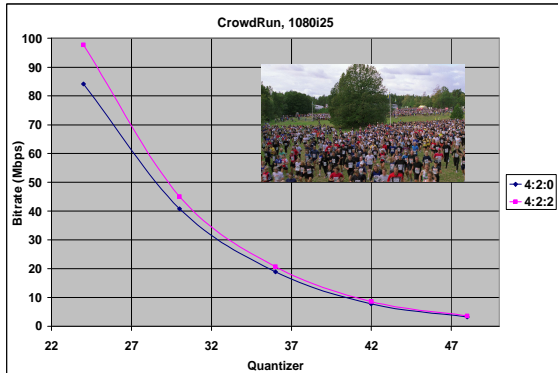


Figure 5 - 4:2:2 vs 4:2:0 Bit-rate Comparison

Surprisingly, this bit-rate increase does not lead to a loss of video quality with the first generation. In fact, the perceived quality is roughly the same except at very high bit-rates where 4:2:2 processing performs slightly better than 4:2:0. As shown in Figure 6, an objective measurement like PSNR reflects this subjective perception.

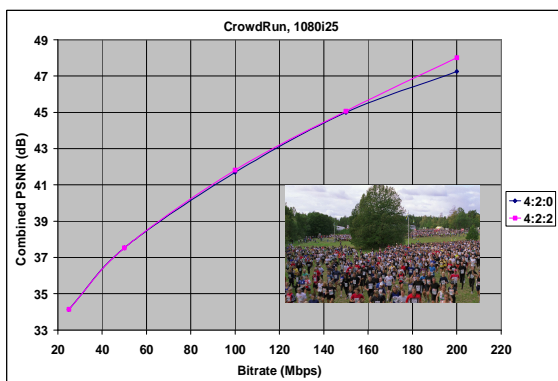


Figure 6 - 4:2:2 vs 4:2:0 Quality Comparison

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.

Taking these advantages into consideration, using 4:2:2 chroma sub sampling can meet the needs of production and contribution applications.

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**:

- **High 10 Profile:** 8-bit up to 10-bit
- **High 4:2:2 Profile:** 8-bit up to 10-bit
- **High 4:4:4 Predictive Profile:** 8-bit up to 14-bit
- **High 10 Intra Profile:** 8-bit up to 10-bit
- **High 4:2:2 Intra Profile:** 8-bit up to 10-bit
- **High 4:4:4 Intra Profile:** 8-bit up to 14-bit
- **CAVLC 4:4:4 Intra Profile:** 8-bit up to 14-bit

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]. Figure 7 illustrates the gains that can be achieved using higher than 8-bit processing (this measurement is performed in 4:2:0 with an 8-bit source up-scaled to 10, 12 or 14-bit).

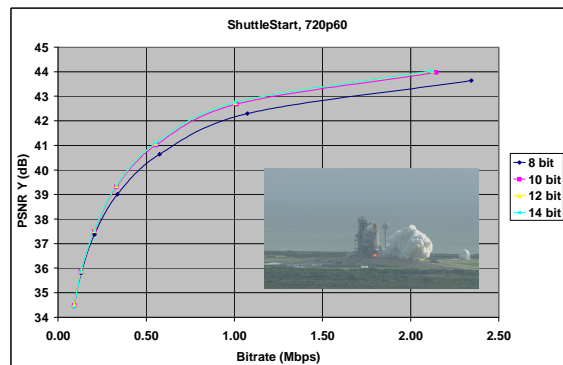


Figure 7 - Coding efficiency gain using more than 8-bit

Extensive experimentation demonstrates that the coding efficiency gains are highest with videos that contain shallow textures and low noise. But as shown in Figure 8 there are also gains to be had with more “significant” sources.

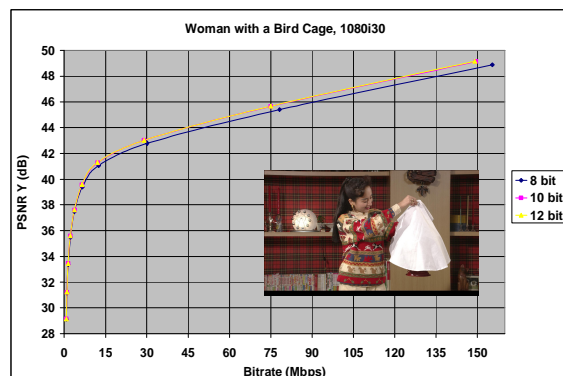


Figure 8 - Coding efficiency on a noisy & textured source

Figure 9 and 10 illustrate PSNR improvements obtained from increasing the bit-depth to 10 or 12 bits on relatively noisy and textured standard sequences.

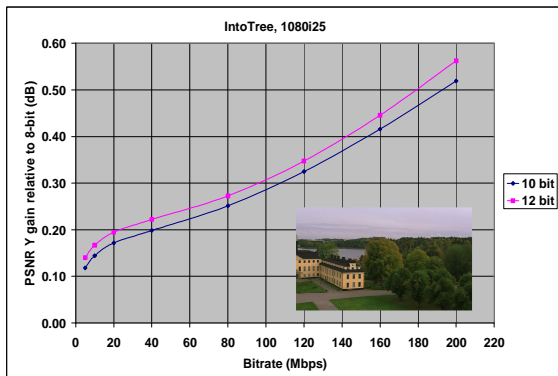


Figure 9 - Coding efficiency gain, more than 8-bit coding

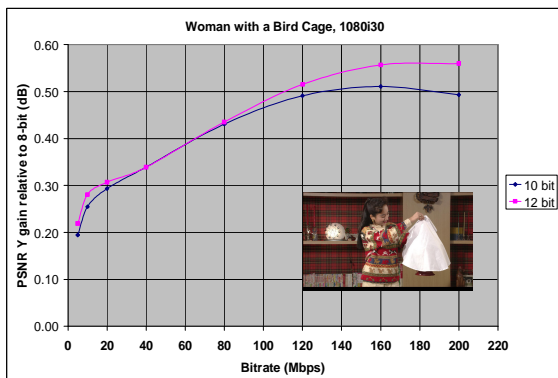


Figure 10 - Coding efficiency gain, more than 8-bit coding

Those curves clearly show that the gain is smaller as the bit-rate is reduced, but that it remains important even at lower bit-rates. Interestingly, those PSNR improvements are in the range of what is achieved with common tools like 8x8 transform or multiple references. This gives reason to use this feature even for low bit-rate applications.

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 of about 5% and up to 20%, while retaining the same video quality.

However, further testing shows that increasing the bit-depth to 12-bit (or even 14-bit) does provide a much smaller coding efficiency gain (up to about 1% in bit-rate saving), but again, no loss over 8 or 10-bit.

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
- Smearing
- Mosquito noise

This gives a better 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.



Figure 11 – Close-up of the Crew sequence 8-bit encoded



Figure 12 - Close-up of the Crew sequence 10-bit encoded

These impairments are otherwise difficult to reduce using traditional tools:

- If the source is not too noisy and the plain areas not too large relative to the picture surface, lowering the quantizer locally produces an effect close to the one achieved with 10-bit processing. Unfortunately, it requires a reduction of around 6 relative to the mean picture quantizer. This has several negative impacts, the most important one being potentially a strong reduction of the coding efficiency and a degradation of the rate-control stability.
- Another approach is to hide the defects by adding noise during the encoding process. The problem is that the amount of added noise needed to achieve the same visual improvement is of significant importance. Even at high bit-rates, this can lead to an unacceptable reduction in coding efficiency.

Since gains are provided through increased accuracy of internal computations, 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).

10-bit is key for contribution

Given that high bit-rates benefit most from using 10-bit compression, production and contribution 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.

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.

Furthermore, this profile enables very high maximum bit-rates for the Video Coding Layer (VCL):

- 525i and 576i (Level 3): 40Mbps
- 720p and 1080i25/30 (level 4.1): 200Mbps
- 1080p50/60 (Level 4.2): 200Mbps

HD encoding at around 50Mbps provides quasi-transparency for the vast majority of Broadcast contents. However, measurements show that up to 150Mbps (35Mbps 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. As illustrated by figures 12, 13 and 14, AVC/H.264 **High 4:2:2 Profile** offers important savings when compared to MPEG-2, even at the highest bit-rates.

These HD examples allow us to draw some conclusions verified by subjective measurements:

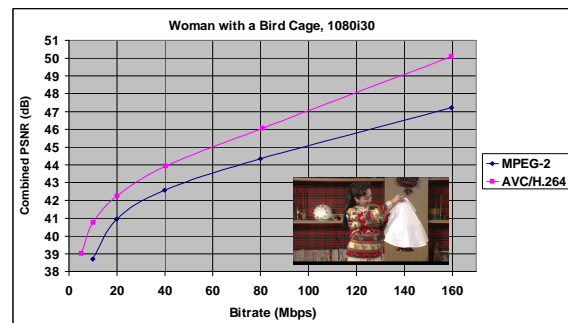


Figure 12 - AVC/H.264 H422P compared to MPEG-2 H422P

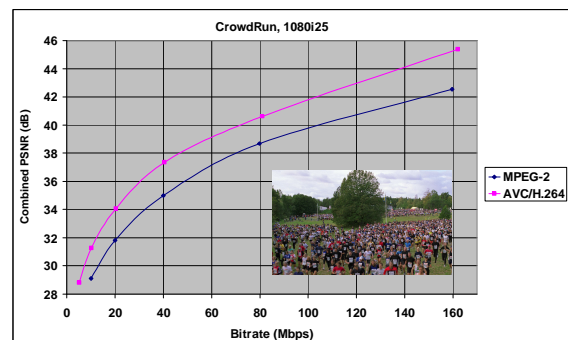


Figure 13 - AVC/H.264 H422P compared to MPEG-2 H422P

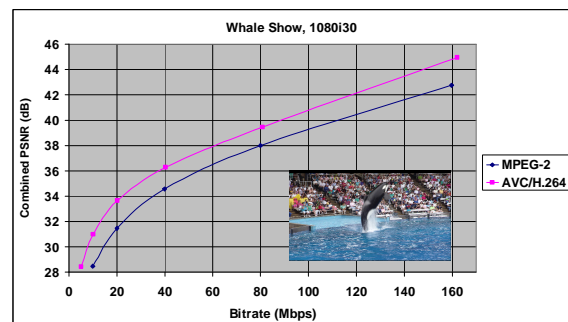


Figure 14 - AVC/H.264 H422P compared to MPEG-2 H422P

- As it's well known, AVC/H.264 offers a bit-rate gain of roughly 50% at below 15Mbps. This gain is lower at higher bit-rates.
- Above 30Mbps, AVC/H.264 produces results comparable in quality to MPEG-2 with a 20Mbps increase. For instance, MPEG-2 quality at 60Mbps is achieved by AVC/H.264 at only 40Mbps or less. At very high bit-rates, this rate saving can sometimes be even greater since the slopes of the rate-distortion curves are slightly different between the two codecs.
- 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.

Summary

This paper presented the advantages of using AVC/H.264 **High 4:2:2 Profile** for 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 **High Profile** as all subjective and objective measurements exhibit a quality increase for the same bit-rate.

Comparisons with MPEG-2 show that AVC/H.264 **High 4:2:2 Profile** can offer important rate savings even at the highest bit-rates. 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

This year will be a turning point for contribution applications as encoders and decoders exploiting the full potential of **High 4:2:2 Profile** become commercially available for the first time. Furthermore, relying on highly standardized bitstream syntax guarantees that products from different manufacturers are interoperable.

References

- [1] "Advanced video coding for generic audiovisual services", ITU-T Recommendation H.264, November 2007
- [2] A. Tourapis, K. Sühring and G. Sullivan, "H.264/MPEG-4 AVC Reference Software Manual", Geneva, Switzerland, 24th JVT Meeting, Doc. JVT-X072, July 2007
- [3] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. Sullivan "Rate-Constrained Coder Control and Comparison of Video Coding Standards", IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 688-703, July 2003

[4] T. Chujoh and R. Noda, "Internal bit depth increase for coding efficiency", Marrakech, Morocco, Doc. VCEG-AE13, January 2007

[5] A. Caruso, L. Cheveau and B. Flowers, "MPEG-2 4:2:2 Profile – its use for contribution/collection and primary distribution", EBU Technical Review N°276, Summer 1998

Acknowledgements

Thanks to Adi Kouadio from the European Broadcasting Union and my fellow DVB members for their useful insights on contribution issues.

I would also like to thank my colleagues Marc Baillavoine and Mathieu Monnier of ATEME for the numerous and invaluable discussions we had during the preparation of this paper.