UHD INTER-CODEC INTERFERENCE IN REAL LIFE PRODUCTION CHAIN

Asma Jenhani

Mesclado, France

Jean-François Nivart

Image Matters, Belgium

ABSTRACT

The emergence of Beyond HD resolutions has triggered a new competition both in broadcast and cinema industries. ProRes, XAVC, AVC-Ultra, JPEG-2000, HEVC, VP9… All these are 4K/UHD-supporting codecs. Manufacturers' marketing slides show excellent quality guarantees. However, these may concern a single-generation encoding/decoding. Real-life workflows imply more complicated parameters' configurations, especially in post-production.

Mesclado's internal Lab for applied researches simulated a complete media chain, based on different production genres such as sport and fiction. Production, editing, mastering and distribution, each step is potentially being affected by a decoding-processing-encoding generation.

Our goal is to objectively measure distortion levels between an original sample and its distributed version.

We were able to identify good and sub-optimal codecs' combinations. We concluded by recommending good engineering practices to save both bandwidth and storage through the production process and to increase the delivered image quality through the conventional distribution channels.

This work was conducted in partnership with Image Matters and direct involvement of Dwarf Animation Studio.

INTRODUCTION

We are witnessing today the big shifting of audio-visual professionals towards Beyond HD resolutions. A huge excitement is worldwide spread about a new user experience with video and audio immersion. In cinema, it is going fast and many productions are done in 4K. In Broadcast, tests are being performed during global public events (Linkin Park Berlin 2014 concert UHD live broadcast, Roland Garros since 2013, FIFA World Cup, 8K UHD 2020 Tokyo Olympics, etc.).

In parallel, new UHD-supporting codec schemes are being introduced to the market, such as XAVC and HEVC implementations. Many doubts were raised about the maturity of such codecs when involved in complex media workflows.

ORIGIN OF THE IDEA

Mesclado was interested in this issue and launched a measurement campaign in the 1st quarter of 2014. It concerned a single encoding/decoding generation on UHD samples to test the codecs performance. Three 10-second samples were involved for the tests: ballet show (50fps, 3840x2160), France Televisions' series "Plus Belle La Vie" (50fps, 3840x2160) and French Tennis Open Roland Garros (59.94 fps, 3840x2160). Each one has been encoded using the codecs' panel.

The comparison was performed between the raw input reference and the encoded output, subsequently decoded to raw data. We chose to adopt objective measurement using Peak Signal to Noise Ratio (PSNR) as involving human beings would introduce a wider error range and subjective parameters. Figure 1 and Table 1 illustrate the results of the tests conducted on the Roland Garros sample.

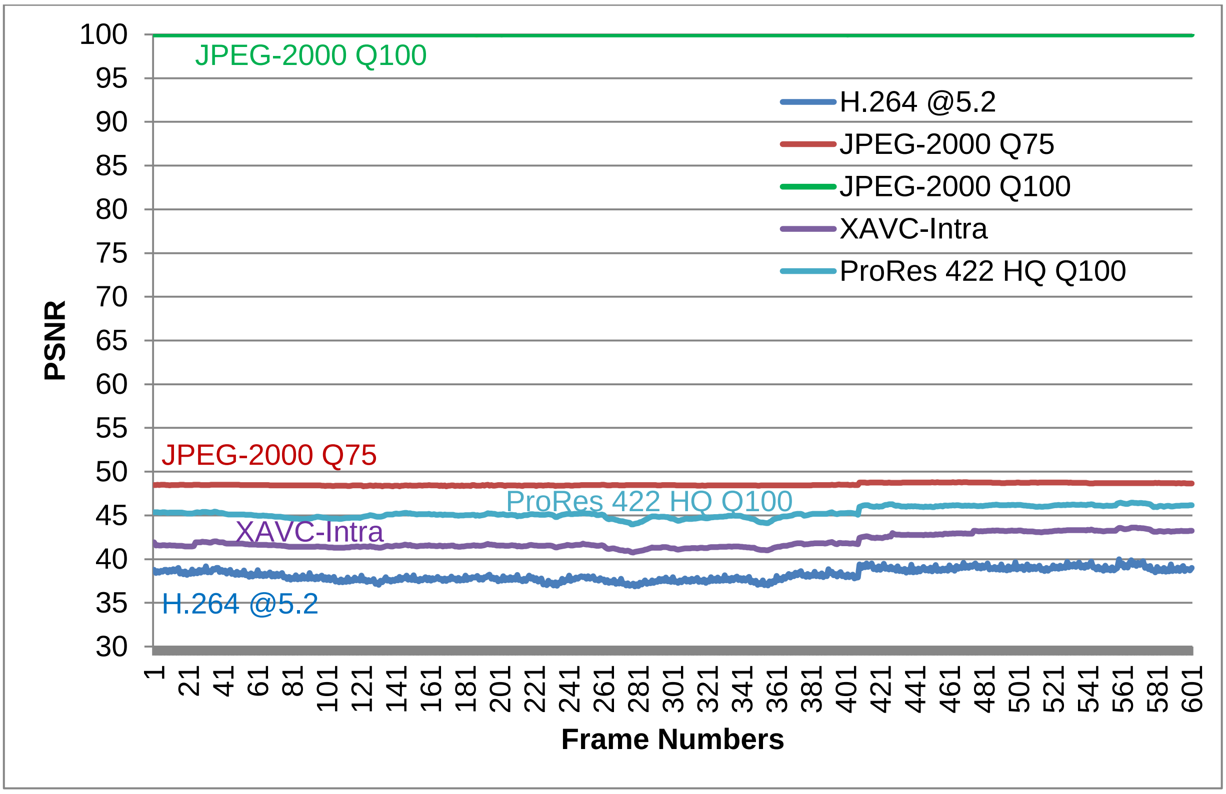

Figure 1 - PSNR comparison for Roland Garros 2013 sample

Codec	Encoding parameters	Average Bitrate (Gbps)	Quality (PSNR average)
JPEG-2000	Quality 100	3,65	100,0 dB
JPEG-2000	Quality 75	1,72	48,5 dB
ProRes	Quality 100	2,00	45,3 dB
XAVC	Intra	0,48	41,9 dB
H.264	High Profile @ Level 5.2	0,02	38,2 dB

Table 1: PSNR comparison of Roland Garros 2013 samples

Source: Quality analysis report, Mesclado

These "one shot" operations were not sufficient to conclude about the true potential of each codec through a complete chain (production, post-production, mastering, distribution). Taking this project to the next level was therefore necessary, using the same PSNR algorithm but different encoding platforms. This time, we wanted to go further with multi-generation encoding, by simulating a complete media chain with these codecs through real-life professional pipelines. The aim is to detect the level of distortion at delivery phases. This issue was submitted by our partner Image Matters, a company involved in JPEG-2000 media workflows.

OBJECTIVE MEASUREMENT

Objective quality measurement delivers an unbiased judgement: it is based on mathematical algorithms strongly correlated to human perception. Many metrics are used such as PSNR, Structural SIMilarity (SSIM) and Difference Mean Opinion Score (DMOS). Although SSIM and DMOS offer the best correlation with human perception, it has been demonstrated that SSIM is not efficient for high bitrate (reference related to our previous work of video quality measurement with a French media group) and that DMOS needs calibration for every test (it is an average between 1 and 100 and for a high video dynamic, we need to calibrate DMOS with a Minkowski variable. This calibration is required for every test campaign. The comparison must take in consideration this calibration.). For this reason, we decided to use PSNR for measuring the video quality.

PSNR (Peak Signal to Noise Ratio): Mathematic comparison between one image and its reference. Whatever the used tool, the algorithm is globally the same and the result values also are. It allows a computation of the level of compression distortion. The algorithm is appropriate when it comes to small levels of signal corruption, as it compares two images on a pixel basis. For an 8-bit source, fair video compression quality starts at PSNR values of 30dB, where higher is better. For 16-bit sources, it has been found that fair video compression quality starts at PSNR values of 60dB (1) (2).

$$PSNR = 10\log_{10}\frac{(2^n - 1)^2}{\sqrt{MSE}}$$

Figure 2: MSE: Mean Square Error

n: quantification in bit per sample

MEASUREMENT WORKFLOW

The four classic media processing phases are Production, Post-production, Mastering and Distribution. We built several use-cases based on some of our customers' inputs and on the available encoding platforms and codecs. The source samples were delivered in an uncompressed DPX format. The output of the chain needs to be decoded into the same uncompressed format for comparison. Figure 3 illustrates the measurement workflow.

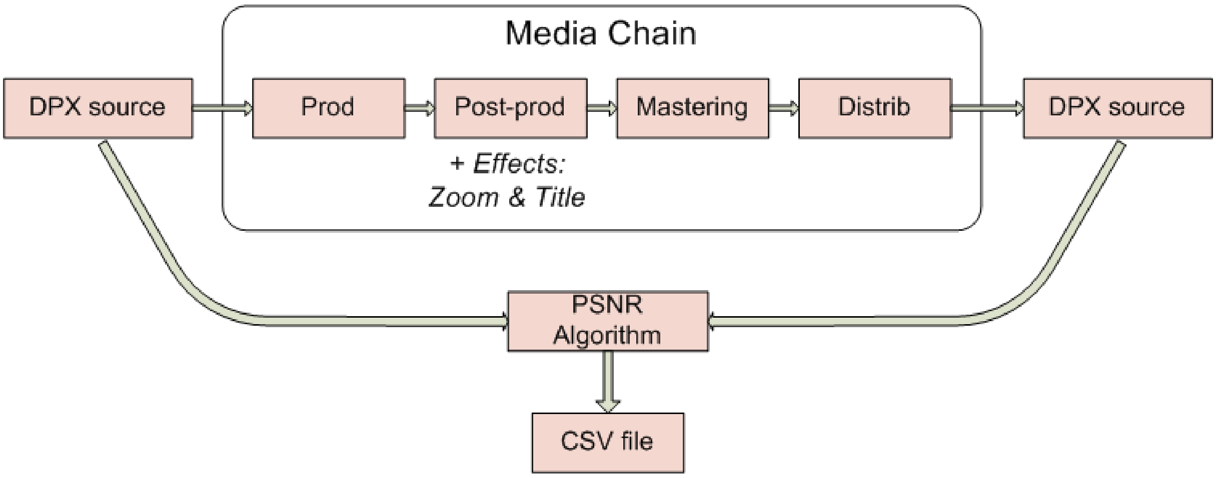

Figure 3: measurement workflow

Effects were also applied:

- A title "TEST RG" with a 1 sec rectangular background at the start of the sample
- A linear zoom starting from 8.30 second (≈ frame 497) till the end of the sequence.

At each phase, several codecs are possible. For example, Fiction is increasingly shot in Raw uncompressed formats. Then, content is transcoded in a high quality format for Post-production such as ProRes, mastered in the same format and distributed in a much lighter one such as H.264. Table 2 states our choices for this study.

	Production	Post-production (PLUS effects)	Mastering	Distribution
Codec	Uncompressed	Uncompressed	JPEG2000-1600Mbps	JPEG-2000-400Mbps
				HEVC
				H.264
	XAVC	XAVC	XAVC	HEVC
				H.264
	Apple ProRes	Apple ProRes	Apple ProRes	HEVC
				H.264
	XAVC	XAVC	JPEG2000-1600Mbps	JPEG-2000-400Mbps
				HEVC 100Mbps
				H.264 120Mbps
	Apple ProRes	Apple ProRes	JPEG2000-1600Mbps	JPEG-2000-400Mbps
				HEVC 100Mbps
				H.264 120Mbps

Table 2: Test scenarios suggested by Mesclado and Image Matters

Table 3 details each codec configuration used during this study:

Codec	Configuration
XAVC	**Container:** MXF **Codec:** XAVC class Intra **Bit rate:** Maximum 480Mbps **Implementation:** Media Encoder (Adobe)
ProRes	**Container:** Quick Time **Codec:** ProRes 422 Quality 90 (default) **Implementation:** Media Encoder (Adobe)
H.264	**Container:** MP4 **Codec:** QHD H264 10 bit **Profile:** High 422 **Bitrate:** 120Mbps **Implementation:** Transkoder (Colorfront)
HEVC	**Container:** MP4 **Codec:** HEVC 4K UHD 10 bit (h265) **Bitrate:** 100Mbps **Profile:** Main 10 **Level:** 5 **Max GOP Length:** 48 **Implementation:** Transkoder (Colorfront)
JPEG-2000	**Container:** No container, Native format j2k **Codec:** JPEG-2000 Intopix **Bitrate:** Up to 1600Mbps **Pix Type:** RGB 444 **Implementation:** Image Matters

Table 3: Codecs' configuration

The following is a summary of the performed operations necessary to the quality measurement:

- ✓ We used Adobe Media Encoder to package our reference raw sample into AVI format
- ✓ We used Adobe Premiere Pro to add the effects (zoom and title) in the second generation
- ✓ We used Media Encoder for ProRes and XAVC encoding. ProRes is only available on Mac. We encoded the samples in ProRes using a Mac with X Yosemite OS with the same version of Media Encoder. XAVC encoding was performed on the HP Z800
- ✓ We initially used DVS Rohde&Schwarz's Clipster to encode our references into its JPEG-2000 implementation to prove our end-to-end workflow
- ✓ We used Image Matters' Intopix implementation for JPEG-2000 codec where it is available with high bitrates
- ✓ We used Colorfront's Transkoder for HEVC and H.264 codecs.
- ✓ We used FFmpeg in the intermediate generations to decode samples to DPX files to interface Media Encoder and Transkoder. We did the same in the last generation to decode the final resulting samples of Transkoder into DPX format
- ✓ After getting final DPX files, we used Media Encoder to encapsulate them in AVI format
- ✓ Last step is to compare the two end-to-end samples using PSNR algorithm to get CSV files filled by per-frame PSNR values

TESTS & RESULTS

The first tests have been conducted on the XAVC scenario, introducing effects before the second generation of encoding. Figure 3 clearly shows the quality degradation between the first and the second generation. The quality seems however to be more stable in the second generation, which announces the proximity to the quality limits of the codec.

The fourth generation involved both H.264 and HEVC. Both results are very close. H.264 seems to better handle the title insertion than HEVC. However, there is a small quality performance for HEVC during the zoom effect at the end of the sequence.

These results confirm the quality destruction character of H.264-based codecs from one generation to the other. To find out the limit of this destruction, we need to pursue the tests with further generations and establish if that limit exists and its localisation on the PSNR scale. A subjective test should validate the visual quality, as the fourth generation shows a PSNR close to 30dB, which is already under the acceptable range.

XAVC chain

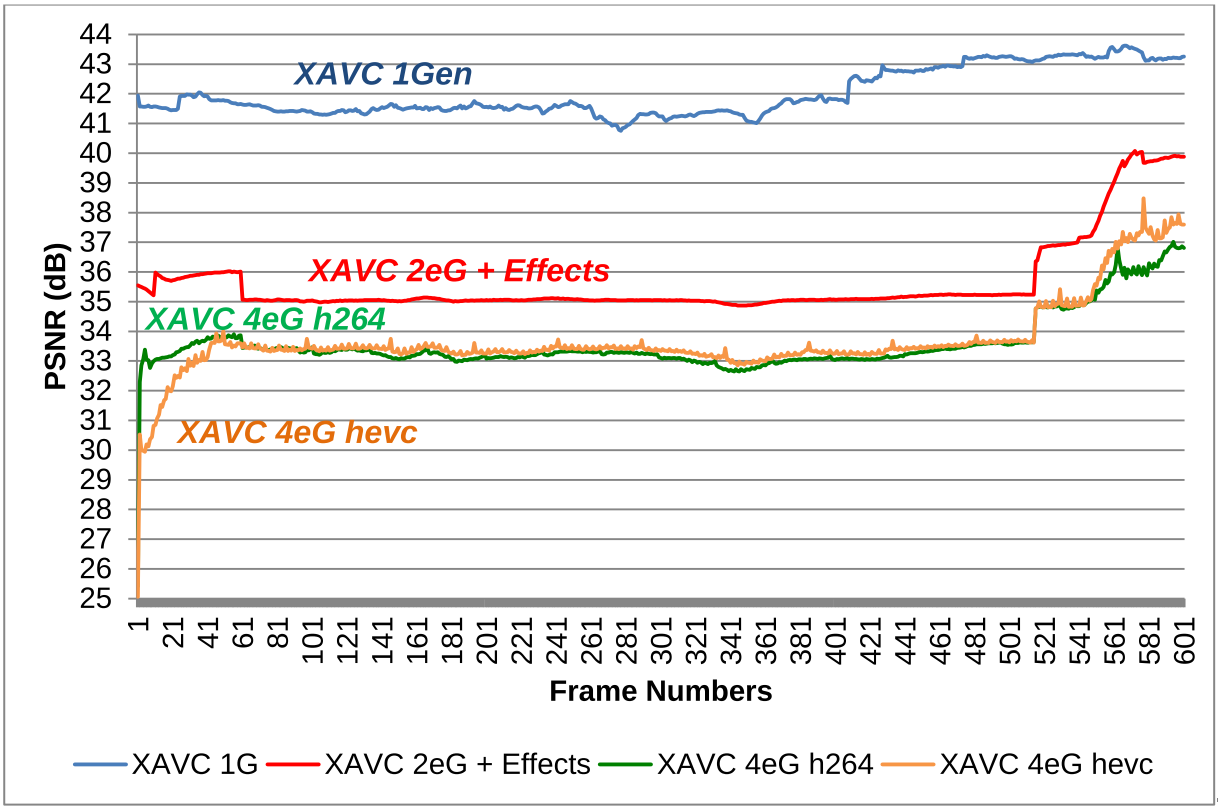

Figure 3: 1st & 2nd generation comparison with XAVC and JPEG-2000

Another interesting perspective is the comparison of XAVC, a DCT-based codec, and JPEG-2000, a wavelet transform one. As the highest profile of XAVC is 480 Mbps, we chose to set this value for JPEG-2000 and compare two generation encoding operations.

Figure 4 clearly shows that, after going through a second generation, JPEG-2000 wins in terms of quality and handles even better the zoom effect, almost with the same way XAVC does. The quality difference between the two codecs seems to be stable, where JPEG-2000 remains 5dB higher than XAVC at the same bitrate.

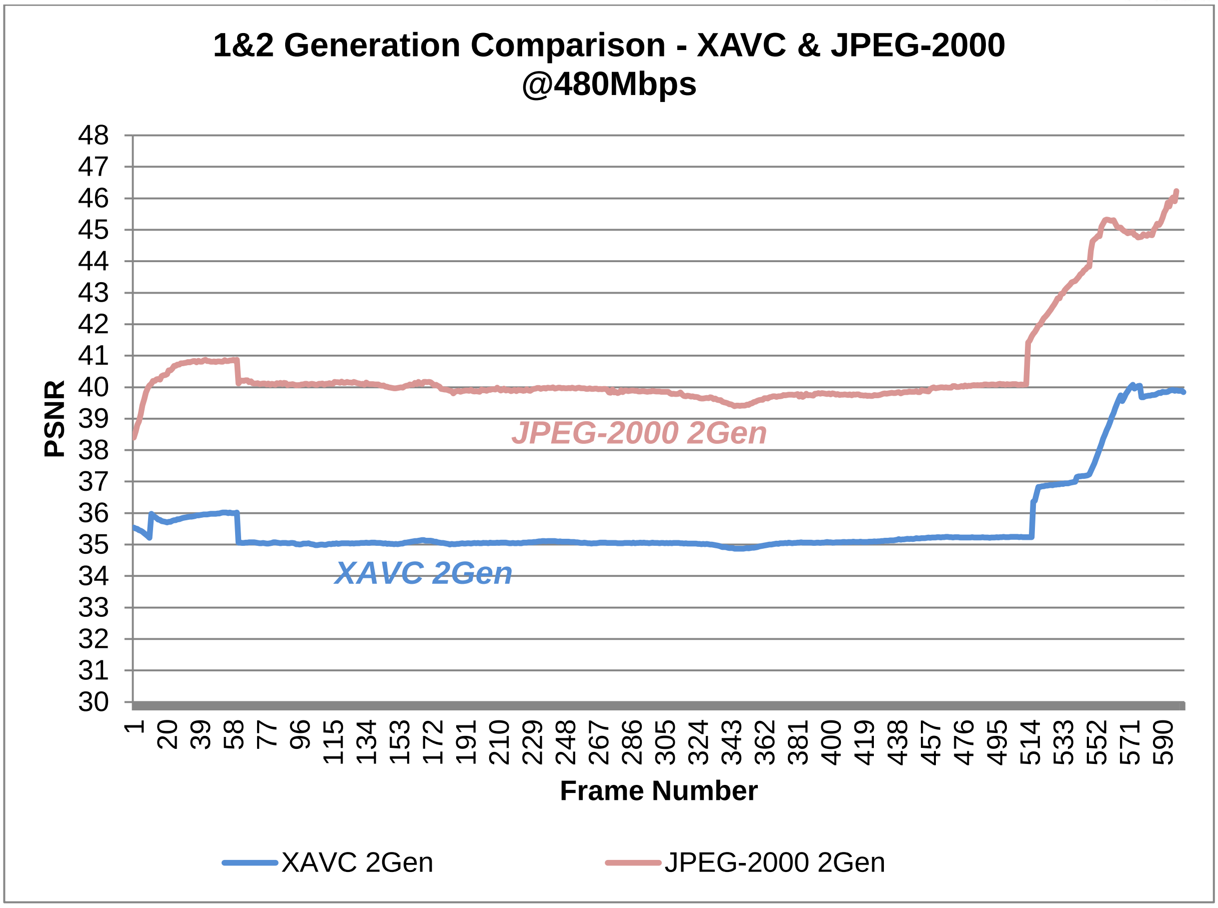

Figure 4: 1&2 Generation Comparison - XAVC & JPEG-2000 @480Mbps

These results confirm that JPEG-2000 has a better performance when used at high bitrates. Cinema and Broadcast industries talk about 600 to 800 Mbps in order to get an excellent quality. The results also confirm that JPEG-2000 can be a good alternative for quality constraints in workflows that do not tolerate destruction. This fact may explain the choice of the Society of Motion Picture and Television Experts (SMPTE) 35PM-50 Workgroup of JPEG-2000 for their Interoperable Master Format (IMF) first applications. Most DCT-based codecs used today do not offer such features for high quality requiring workflows.

ProRes chain

The ProRes chain is similar to the XAVC one. If we compare the two test series, we notice that both codecs deliver nearly the same shapes. It seems however that ProRes better resists to the quality degradation with a 3dB higher compared to XAVC. This difference goes up to 7dB in the second generation, in favour of ProRes.

When comparing XAVC and ProRes chains in distribution (4th generation), these differences go away: both chains exhibit the same quality level. This observation indicates that both H.264 and HEVC respond in the same way during the fourth generation, with either ProRes or XAVC upstream.

A comparison with a similar JPEG-2000 chain would be interesting to analyse as it might show that HEVC and H.264 at the end of the chain have the same response. Thus, such workflows would simply require the less expensive codec between XAVC, ProRes and JPEG-2000, as ultimately the final HEVC or H.264 4th generation quality would be the same. This would only apply to low-bitrate distribution.

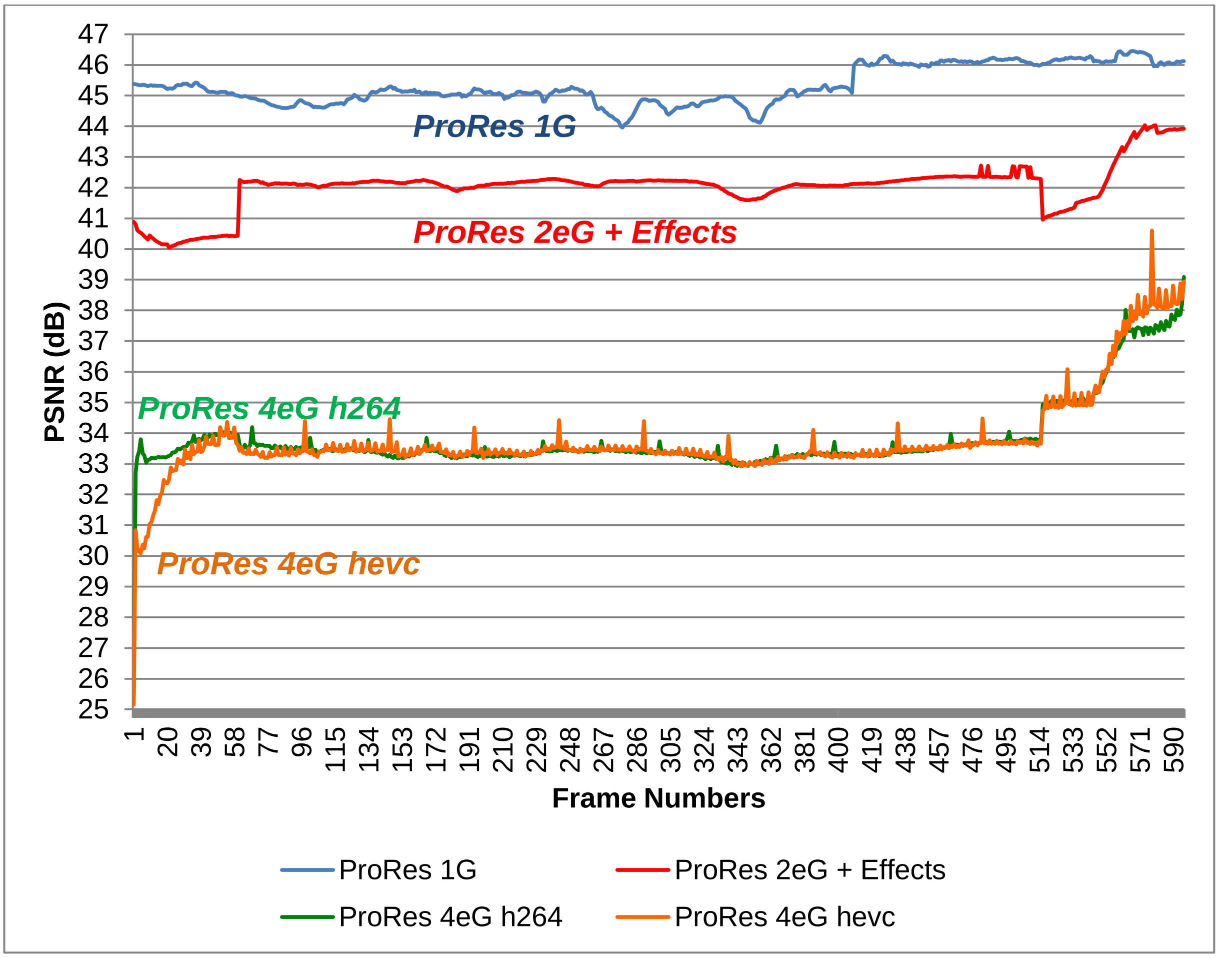

Figure 5: ProRes Chain

Uncompressed chain with JPEG-2000 distribution

There is also the need to compare different encoding bitrates for JPEG-2000. For this, we used an uncompressed chain that concludes its 4th generation for distribution with a JPEG-2000 encoding, ranging from 100Mbps to 1.5Gbps.

Figure 6 clearly shows a degradation proportional as the bitrate decreases. The shape of this decrease seems to be almost linear, which allows determining the required bitrate for a requested PSNR level. However, it is also clear that JPEG-2000 best performances are closely linked to high bitrates, which directly impact on the workflow complexity.

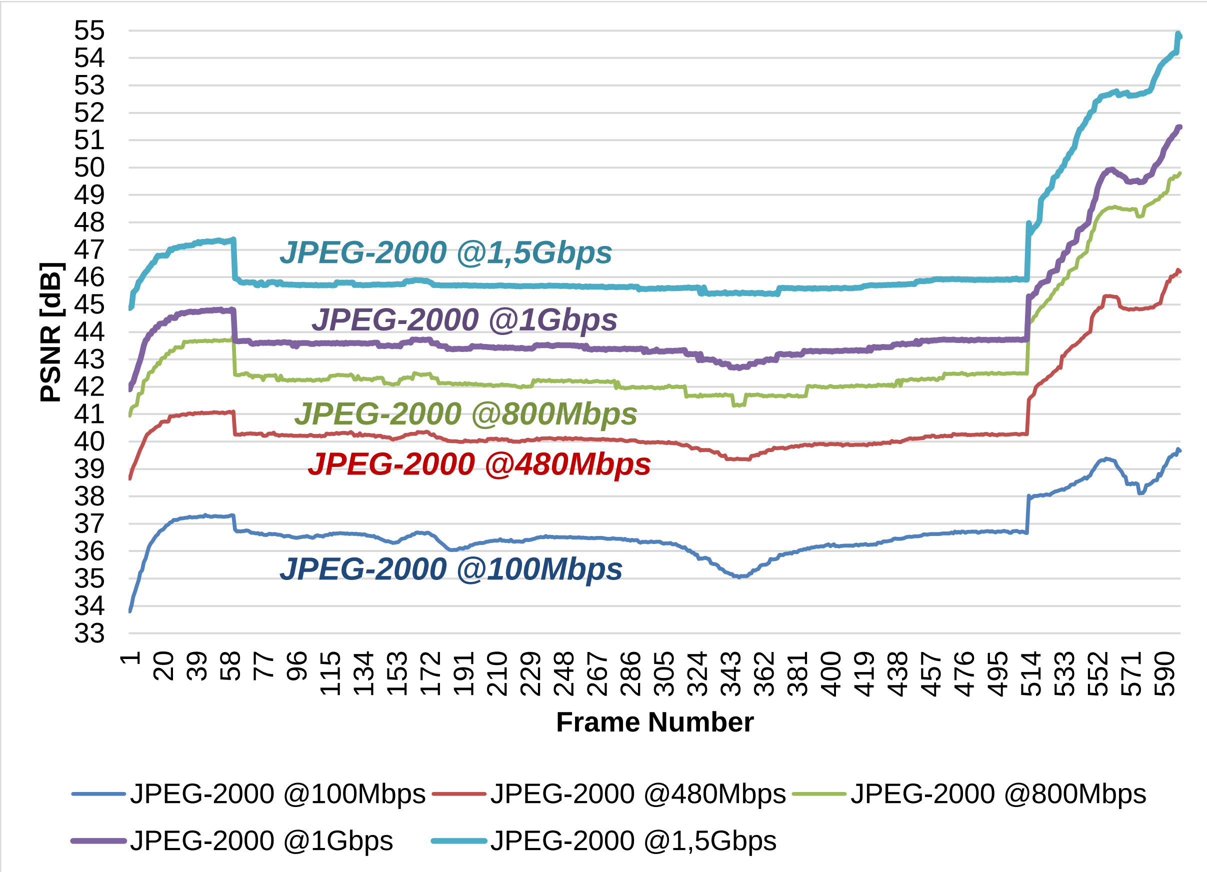

Figure 6: A complete uncompressed chain with JPEG-2000 distribution

CONCLUSIONS

We observed several levels of signal distortions caused by multi-generation encoding, involving codecs not enough mature for Ultra High resolutions. JPEG-2000 lossless version is very efficient. However, it requires high bitrates. HEVC demonstrates coding efficiency but it is still under testing by broadcasters. XAVC is starting to be used in post-production despite weak performances in multi-generation coding. ProRes on the contrary shows good resilience to cascading encoding operations.

Our tests concluded that codecs' performances hardly depend on the coding parameters, especially at bitrates greater than 1Gbps and bitdepths of 10+ bits per sample. However, this will deeply impact the IT infrastructure, requiring a migration to high performance 10+ Gig networks, sufficient storage capacities, etc. We are waiting for new codecs' implementations that will offer a larger set of possibilities.

The ecosystem seems to be at the same point HD codecs were early 2000. It is simpler for Cinema. However, when will it be possible to receive a UHD signal at home and display it on a large TV set fitting a traditional living room size?

REFERENCES

(1)Welstead, Stephen T. (1999). Fractal and wavelet image compression techniques. SPIE Publication. pp. 155–156

(2) Raouf Hamzaoui, Dietmar Saupe (May 2006). Barni, Mauro, ed. "Fractal Image Compression". Document and image compression (CRC Press) pp. 168–169.

ACKNOWLEDGEMENTS

We would like to thank François Abbe, Cherif Ben Zid and Sandrine Hamon (Mesclado) and Vincent Matton (Image Matters) for their contribution to this paper.
We would also like to thank the International Broadcasting Convention for permission to publish this paper.