JPEG XS – The new mezzanine codec for high quality video production

By Dr. Siegfried Fößel, Dr. Thomas Richter and Dr. Joachim Keinert,
Fraunhofer IIS Erlangen, Germany

Overview

Today, the industry is requesting far more seamless workflows, and this is true in the professional media sector as well. An All-IP[1] approach in the video content production and exploitation chain is therefore a major goal, allowing automatic handling and processing of content from ingest to distribution. To reduce the requirements on the IP infrastructure, the ISO/IEC JTC1 SC29 WG1 committee, better known as JPEG committee, developed the new data compression codec JPEG-XS. The specifications for JPEG-XS define a mezzanine compression codec for images and video data with lowest latency for complete encoding-decoding cycles and require only moderate computational resources while preserving image quality at highest level. The project started in 2016 and by begin of 2019, the first parts of the standard will be published.

JPEG Standardization process and status

The JPEG committee develops still image compression standards since nearly 30 years. In some cases, where intra-frame compression is requested due to latency requirements or complexity issues, these codecs are also used for compression of video signals. Whereas the codec development in the early days targeted a wide area of applications, the JPEG committee today is focusing on codec developments for concrete applications.

In 2016, the committee investigated the potential need for such new application areas and asked the industry for evidence; as a result, the need for standardization of a mezzanine image codec was identified. At present, multiple proprietary codecs are in use in this particular application domain, though the lack of standardization prevented wide adoption of mezzanine codecs in the industry. For example, we find LLVC, VC2, VESA DSC, Lici and Tico in this market segment. In consequence, a new work item called JPEG-XS started, following the regular ISO standardization procedure depicted in Figure I.

After establishing a new work item, the requirements for the new codec had been collected and refined, upon which a call for proposals [1] was published. From the incoming proposals best technology was selected and - in a cooperation phase between the committee members - further optimized. While the standard describing the core coding technology (Part 1) is already under publication of ISO, the JPEG committee is working today on testing procedures, reference software and further refinements of the technology.

JPEG XS System Requirements

In 2016, the JPEG committee identified requirements for the desired new work item. The most relevant corner points are as follows:

- Support of at least RGB/444 and YCbCr 444/422 image formats of up to 12 bits per color component sample precision with the option to extend it to 16 bits in the future
- Visually lossless compression, i.e. no visible degradation, even over multiple compression/decompression cycles. The latter is known under the term “multi-generation robustness”.
- Maximum 32 lines end-to-end (compression-decompression) latency
- Support for multiple platforms e.g. FPGA, ASIC, GPU and CPU, allowing a high degree of parallelism
- Real-time software implementation capability for 4k/60p formats on today’s standard computers
- Low complexity, defined as a maximum percentage of a specific low-cost FPGA[2]

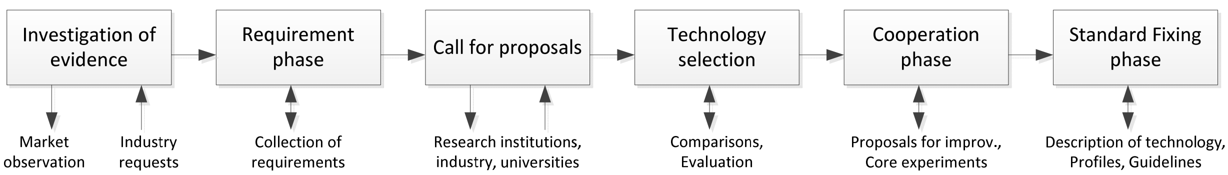

Figure I: JPEG procedure for developing a new standard

1 IP Internet Protocol

2 For details see Call for proposals [1]

- No external frame buffer required in embedded applications. In particular, individual frames shall be decoded independently

A test set of images was selected representing natural images, but also mixed content, i.e. CGI and natural images, including even some screen content which is critical to compress.

To identify target compression rates, Table I lists some of the typical video formats used in the industry, together with the underlying link and its corresponding bandwidth. As expressed in the table, a compression ratio of 6:1 would allow transport over many popular links; with 12:1 nearly all cases are covered. For the purpose of core experiments, the JPEG committee considered bitrates as low as 3bpp, counting all components.

Comparison with other codecs

Based on the requirements, it is easy to explain that existing standards cannot address all needs of film and broadcast applications. JPEG-LS [2] and JPEG [3] as well as its successor JPEG-XT [4], which provides backward compatible support of higher bit depths, make a precise rate control difficult, neither do they allow to limit latency. JPEG 2000 [5] employs a complex entropy coder, which implies that its demands for hardware and software resources makes it unattractive for real-time implementations at UHD resolutions. The encoder complexity of HEVC [6] is even one magnitude higher than that of JPEG 2000; while this is acceptable for a distribution (one-to-many) codec and ASIC implementations, it makes it

Table I: Typical video formats and interfaces

Format and bitrate	12G SDI compr. ratio	3G SDI compr. ratio	HD SDI compr. ratio	40G Ethernet compr. ratio	10G Ethernet compr. ratio	1G Ethernet compr. ratio
HD/2k 60p/422/10 bits 2.7 Gbps	1 x uncompr.	1 x uncompr.	1 x ~2:1 2 x ~4:1	14 x uncompr.	3 x uncompr.	1 x ~3:1 2 x ~ 6:1
HD/2k 60p/444/12 bits 4.8 Gbps	1 x uncompr.	1 x ~2:1 2 x ~4:1	1 x ~3.5:1	7 x uncompr.	2 x uncompr.	1 x ~6:1
UHD-1/4k 60p/422/10 bits 10.8 Gbps	1 x uncompr.	1 x ~4:1	~8:1	3 x uncompr.	1 x uncompr.? 3 x ~4:1	~12:1
UHD-1/4k 60p/444/12 bits 19 Gbps	1 x ~2:1 2 x ~4:1	1 x ~7:1	-	2x uncompr.	1 x ~2:1 2 x ~4:1	-
UHD-2/8k 120p/422/10 bits 85 Gbps	1 x ~8:1	-	-	1 x ~2.5:1 2 x ~5:1	~10:1	-

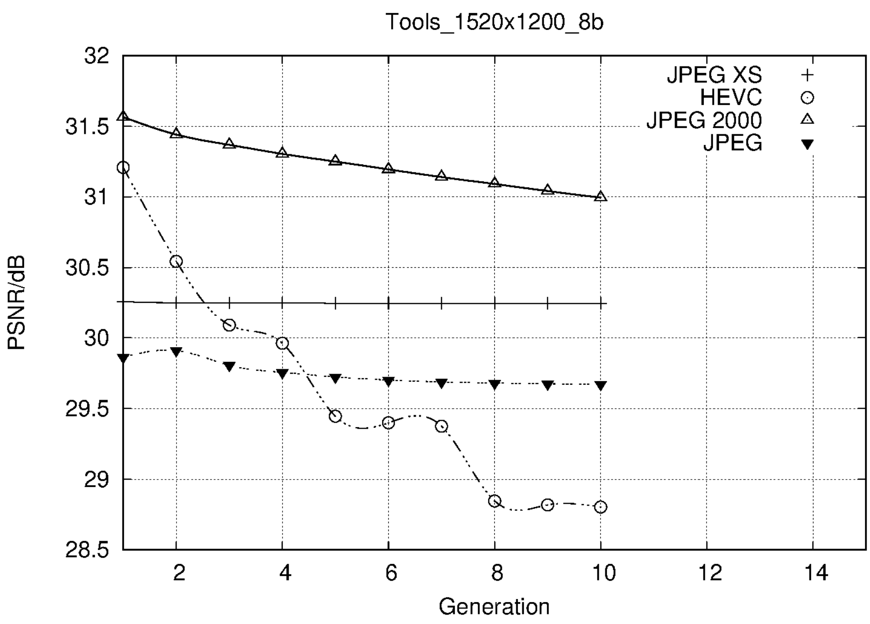

Figure II: Generation loss of JPEG XS compared to JPEG, JPEG 2000 and HEVC on the Tools image unsuitable as a mezzanine codec. Furthermore, as experiments carried out by WG1 members showed, it is not multi-generation robust either.

Figure II shows the bitrate over quality for JPEG XS, JPEG, JPEG 2000 in a low-latency configuration and a restricted variant of HEVC similar to what has been proposed as coding technology for JPEG XS. As seen there, while HEVC offers better quality for the first generation, image quality degrades quickly generation after generation. Similar, but not quite as prominent problems can be seen for the low-latency variant of JPEG 2000.

VC-2 [7] is of low complexity, but image quality was considered insufficient. ProRes as documented by the SMPTE disclosure document [8] is based on macro blocks of 16x16 pixels. This, plus the lack of any sophisticated rate control mechanism, defeats latency bound coding. Moreover, the symbol-per-symbol entropy coding makes fast CPU implementations challenging as it does not offer an inherent parallelism. DSC [9] finally targets ASIC-based display compression, making efficient implementations on FPGAs and GPUs hard to achieve. In addition, most of the codecs are not optimized for fine-grain parallel processing making a fast implementation on multi-core CPU or GPU platforms difficult.

JPEG XS System Description

Architecture of the JPEG XS codec

Due to the shortcomings of the existing codec algorithms described above, the JPEG committee has developed the new codec JPEG XS. It is based on technology which was received through the call for proposals and extensions and refinements integrated into it during the standardization phase.

Figure III shows the overall block diagram of the JPEG XS codec. In case of RGB input, the color components are decorrelated by means of a lossless color transform identical to the one used in JPEG 2000. Next, an integer LeGall 5/3 wavelet transform is applied. In order to comply with the latency constraints and to avoid excessive memory requirements, only up to two vertical wavelet decompositions are envisaged. In horizontal direction, up to eight successive decompositions are permitted, though in typical applications, only five stages are used.

The resulting wavelet coefficients are analyzed by a budget computation module that predicts the number of bits required for each possible quantization step. Since heavier quantization implies stronger signal distortion but also lower rate, the rate control algorithm computes the smallest quantization factor that does not exceed the bit budget available for coding the wavelet coefficients. Then the wavelet coefficients are entropy coded, four at a time offering data parallelism to speed up processing. Finally, all data sections are combined into a packet structure and sent to the transmission channel. A smoothing buffer ensures a constant bit rate at the output of the encoder, although the input image might consist of input regions that are easier to compress, and others that require more bits per pixel. A separate part of the standard specifies exactly how large the smoothing buffer is and in which units data enter and leaves it, which has also implications on the rate control algorithm. JPEG XS is the first codec developed by the JPEG committee that specifies buffer control in such detail to allow implementations with limited, but well-controlled resources.

Given that the decoder should be able to process the pixels with a constant clock frequency, the number of bits read per time unit varies depending whether a current wavelet coefficient is easy to compress or not. These rate variations are again compensated by a smoothing buffer at the input of the decoder. A packet parser splits the bit stream into individual data chunks representing parts of a sub-band before the wavelet coefficients are decoded, inversely quantized and transformed back into the spatial pixel domain.

Performance

Figure IV shows the typical PSNR of a natural image over different compression rates. As can be seen from the diagram, a lot of improvements could be achieved in the collaboration phase and JPEG XS can even reach the performance of the tile-based JPEG 2000 codec, despite its considerably lower requirements on computational resources.

Encoding and Decoding Speed

Figure V shows typical frame rates for JPEG XS encoding and decoding of a multithreaded CPU-only implementation on two systems, a powerful desktop machine and an intel NUC mini-PC. What is measured here is the number of frames per second of the Fraunhofer JPEG XS implementation on the 4K

"AlexaDrums" image, also used in Figure IV. As seen, the achievable frame rates scale almost linearly with the number of available cores up to a saturation point where additional effects such as available memory bandwidth or the thermal budget limit the throughput.

Profiles and Levels

Particular application domains may enforce additional constraints on the encoder and decoder, such as even lower complexity, or limitation of buffer or screen sizes. To this end, the second part of the JPEG XS standard introduces profiles, levels and sublevels.

Profiles define restricted tool sets, and hence lower the complexity of encoder and decoder by only offering a confined set of coding tools. Profiles are structured along the expected number of logic elements, the expected memory footprint, and whether chroma subsampling or an alpha channel is required. As such, the current draft spells out eight profiles, which can be classified according to the hierarchy depicted in Figure VI and specified in Table IV:

- Two light profiles, one for 4:2:2 and one for 4:4:4 subsampling. Coding tools have been cut down to a set that only requires a minimal amount of gates on FPGA implementations. In particular, both profiles limit the number of vertical wavelet decompositions to 1, simplifying the wavelet stage, and disallow the uniform quantizer which requires a larger amount of logical elements especially at the dequantizer.
- A 4:2:2 light sub-line profile which limits the amount of buffer memory required even further. Unlike the former profile, uniform quantization is allowed, but the number of vertical wavelet transformations must be zero, avoiding additional wavelet buffers. In this particular configuration, only horizontal transformation remains. The maximum column width in this mode is 2048 luma samples wide. In optimized implementations a latency below a single image line can be achieved.
- Two main profiles, one for 4:2:2 and one for 4:4:4 subsampling. They do not impose any restriction on the quantizer, but allow only one vertical wavelet decomposition. As the wavelet filter requires buffering entire lines anyhow, lines cannot be split horizontally into columns, except if the number of vertical wavelet transformations is zero.
- One high profile which is similar to the 4:4:4 main profile, except that the allowed maximum number of vertical wavelet decompositions is two. This enables higher quality, but also increases the complexity. While the same latency as in the main profile can be achieved, the number of vertical wavelet stages requires additional latency itself which is then not available for rate-allocation purposes.
- Two extensions of the main 4:4:4 and high 4:4:4 profiles that also include a (non-subsampled) alpha channel of the same precision and dimension as the luma channel. These two profiles address use-cases where transparency information needs to be carried as well.

While profiles select coding features, levels and sublevels limit the buffer sizes, with levels imposing restrictions in the uncompressed image domain and sublevels on the compressed domain. Thus, in particular, levels constrain the frame dimensions and the refresh ratio, in approximately the same way as HEVC levels (e.g. 1920/60p as maximum input format).

File formats

In Part 3 of the JPEG XS standard various file formats are defined. These file formats are based on already existing file format syntax schemata, in particular the JPEG2000 file format, the ISO Base Media File Format MP4 and the High efficiency File format HEIF. The file formats include additional metadata such as color space information and frame rate, allowing color-correct rendering on computers and archival of videos. The already defined formats with its extensions are listed in Table III.

Transport formats

Other ongoing standardization activities are the definition of transport formats for streaming of JPEG XS bit streams. Actual activities are targeting RTP payload over IP, MPEG2-TS[3] and encapsulation of JPEG XS in SMPTE 2110-22.

A complete list of ongoing standardization activities can be found in Table II.

Next Steps

[3] TS Transport Stream as defined in ISO/IEC 13818-1

Potential future activites include the definition of profiles for bit depth up to 16 bits per component or the extension of the core coding standard for efficient compression of Bayer raw data generated by sensors in industrial or automotive applications.

Summary/Conclusion

The new JPEG XS standard offers an optimized solution for a wide range of applications where only a light compression is necessary and requested, but latency and complexity need to be controlled. The typical compression range is between 2:1 and 12:1. For critical content like mixed screen content or Bayer raw data the optimal compression range is between 4:1 and 6:1. For natural RGB images, compression ratios up to 12:1 are feasible. The mezzanine compression codec is optimized for low-latency and for minimal resource consumption allowing the real-time usage even as software codec for UHD video content.

[1] JPEG Committee, JPEG-XS Call for proposals, https://jpeg.org/downloads/jpegxs/wg1n71031-REQ-JPEG_XS_Call_for_proposals.pdf

[2] Information technology – Lossless and near-lossless compression of continuous-tone still images – Baseline. ISO/IEC 14495-1 | ITU-T Rec. T.87. 1998

[3] Information Technology – Digital compression and coding of continuous-tone still images – Requirements and guidelines. ISO/IEC 10918-1 | ITU-T Recommendation T.81. 1992

[4] Information technology -- Scalable compression and coding of continuous-tone still images -- Part 1: Scalable compression and coding of continuous-tone still images. ISO/IEC 18477-1. 2015

[5] Information technology -- JPEG 2000 image coding system: Core coding system. ISO/IEC 15444-1:2004 | ITU-T Rec. T.800. 2015

[6] High efficiency video coding. ISO/IEC 23008-2 | ITU-T Rec. H.265. 2013

[7] VC-2 Video Compression. SMPTE ST 2042-1. 2012.

[8] SMPTE Engineering Project (ANSI). Apple ProRes Bitstream Syntax and Decoding Process. SMPTE registered disclosure document RDD 36. 2015.

[9] Video Electronics Standards Association. VESA Display Stream Compression (DSC). VESA Standard. 2014.

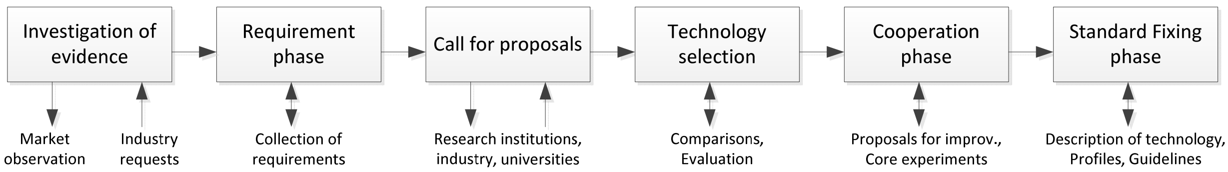

Figure I: JPEG procedure for developing a new standard

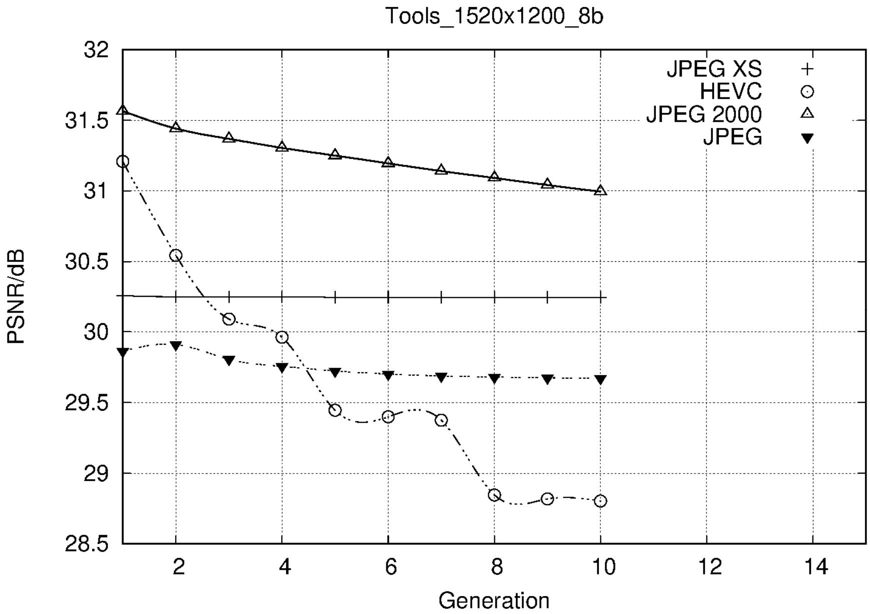

Figure II: Generation loss of JPEG XS compared to JPEG, JPEG 2000 and HEVC on the Tools image

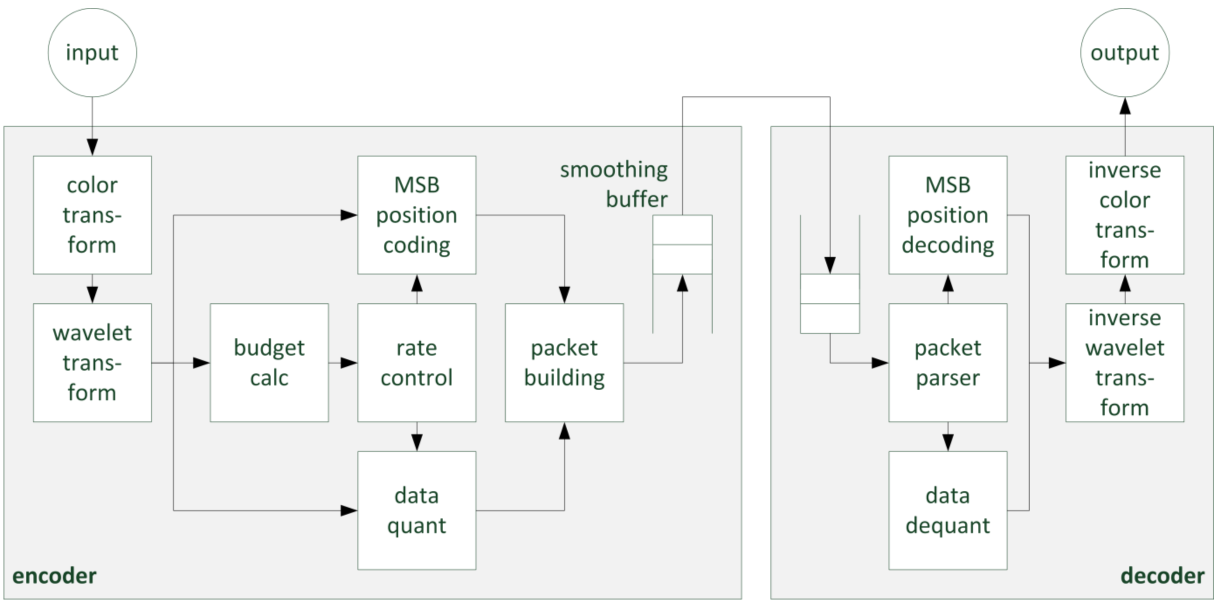

Figure III: JPEG block diagram for encoder and decoder

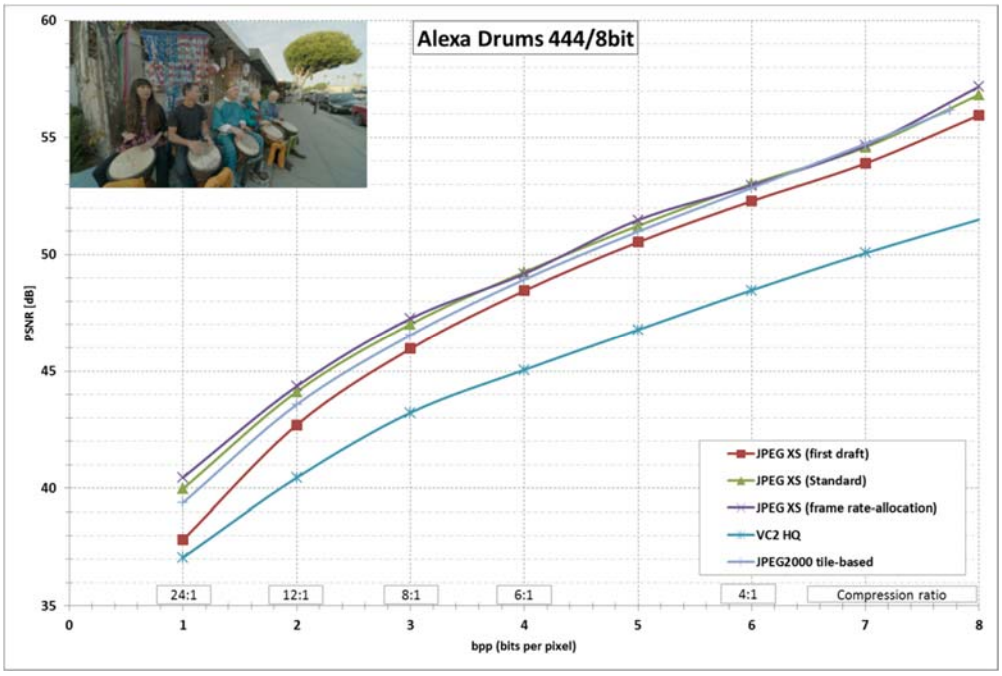

Figure IV: PSNR values as quality example for a 24bit natural test image

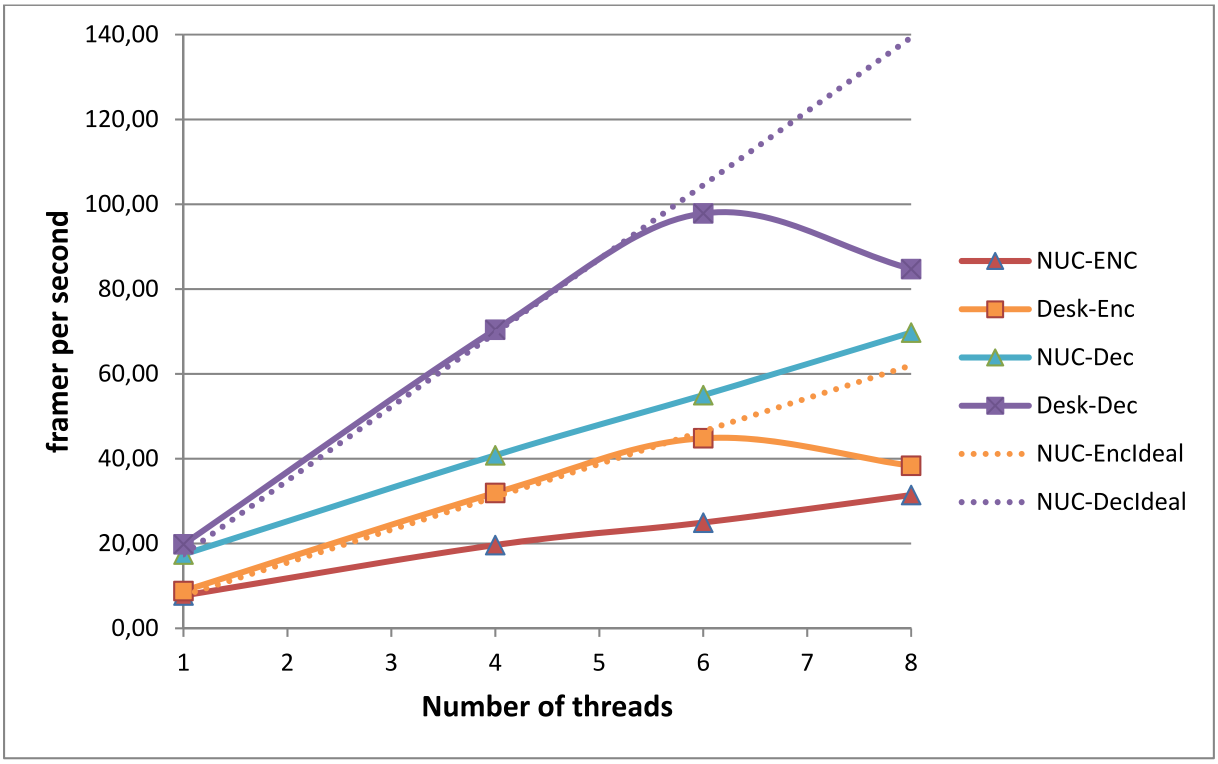

Figure V: Performance of JPEG XS encoders and decoders on the 4K ARRI “AlexaDrums” image (same as in Figure IV) by the number of used threads

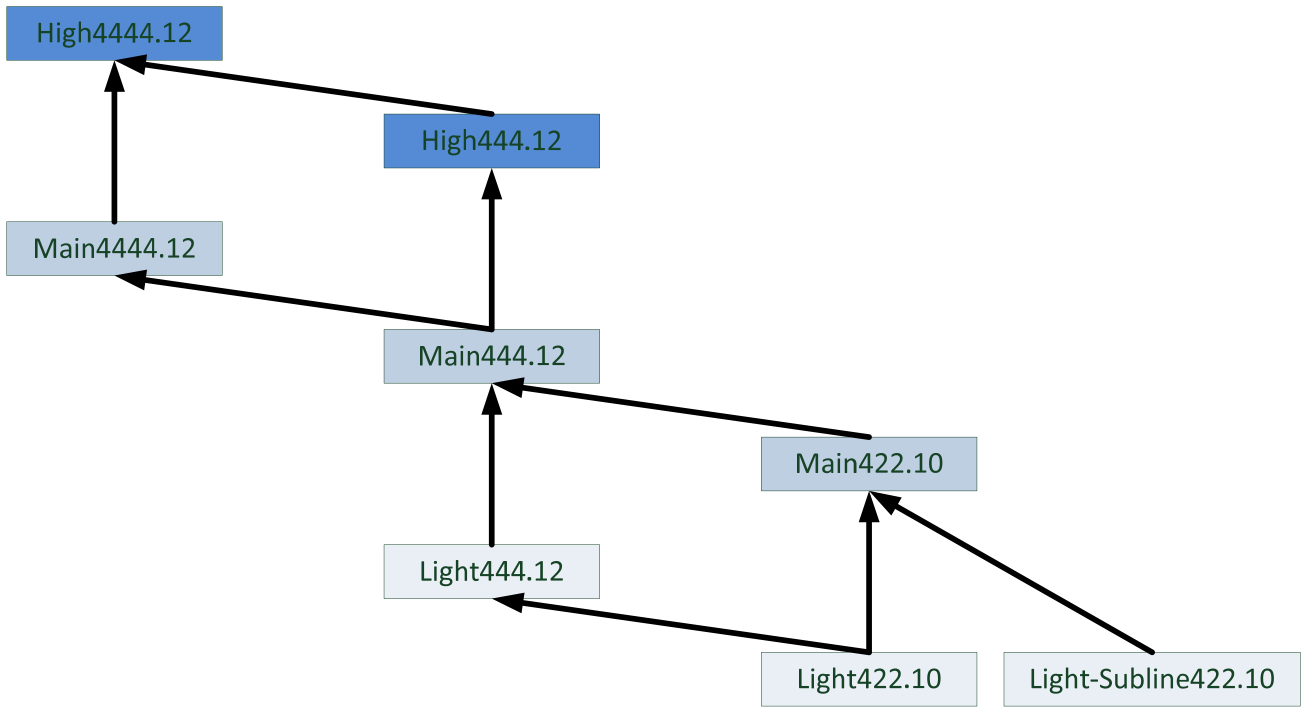

Figure VI: Hierarchy of JPEG XS profiles

Table I: Typical video formats and interfaces						
Format and bitrate	**12G SDI compr. ratio**	**3G SDI compr. ratio**	**HD SDI compr. ratio**	**40G Ethernet compr. ratio**	**10G Ethernet compr. ratio**	**1G Ethernet compr. ratio**
HD/2k 60p/422/10 bits 2.7 Gbps	1 x uncompr.	1 x uncompr.	1 x ~2:1 2 x ~4:1	14 x uncompr.	3 x uncompr.	1 x ~3:1 2 x ~ 6:1
HD/2k 60p/444/12 bits 4.8 Gbps	1 x uncompr.	1 x ~2:1 2 x ~4:1	1 x ~3.5:1	7 x uncompr.	2 x uncompr.	1 x ~6:1
UHD-1/4k 60p/422/10 bits 10.8 Gbps	1 x uncompr.	1 x ~4:1	~8:1	3 x uncompr.	1 x uncompr.? 3 x ~4:1	~12:1
UHD-1/4k 60p/444/12 bits 19 Gbps	1 x ~2:1 2 x ~4:1	1 x ~7:1	-	2x uncompr.	1 x ~2:1 2 x ~4:1	-
UHD-2/8k 120p/422/10 bits 85 Gbps	1 x ~8:1	-	-	1 x ~2.5:1 2 x ~5:1	~10:1	-

Table II: Standard Documents related to JPEG XS low-latency lightweight image coding system			
Doc. No.	**Part Title**	**Decription**	**Document Status**
ISO/IEC 21122-1	Part 1: Core coding system	Specifies syntax of JPEG XS bitstream and algoritm for decompression	under publication as IS
ISO/IEC 21122-2	Part 2: Profiles and buffer models	Specifies application profiles and buffer model for low-latency	under publication as IS
ISO/IEC 21122-3	Part 3: Transport and container formats	Defines file formats and container for JPEG XS bitstreams	FDIS
ISO/IEC 21122-4	Part 4: Conformance testing	Specifies method for conformance testing of JPEG XS bitsreams	DIS
ISO/IEC 21122-5	Part 5: Reference Software	Reference software for JPEG XS	WD
ISO/IEC 13818-1 ed.7 AMD1	Carriage of associated CMAF boxes for audio-visual elementary streams and JPEG XS in MPEG-2 TS	Definition of JPEG XS video stream in MPEG2-TS	DAMD
IETF draft	RTP Payload Format for ISO/IEC 21122 (JPEG XS)	Definition of RTP payload format for JPEG XS video	Draft
SMPTE 2110-22	Compressed Video Essence	Encapsulation of compressed video streams in SMPTE 2110	FCD
Notes: Document Status is as of 4th April 2019 Abbr.: WD Working Draft, CD Committee Draft, FCD Final Committee Draft, DIS Draft International Standard, FDIS Final Draft International Standard, IS International Standard, AMD Amendment, DAMD Draft Amendment			

Table III: File formats for JPEG XS			
Container	**Type**	**Description – Main purpose**	**extension**
JXS	JPEG XS Fileformat (defined in ISO/IEC 21122-3 Annex A)	For storing of single images – JPEG 2000 syntax based	.jxs
MP4	ISO Base Media File format (ISOBMFF) (defined in ISO/IEC 21122-3 Annex B)	For storing of video – ISOBMFF syntax based	.mp4
HEIF	High Efficiency Image File Format (defined in ISO/IEC 21122-3 Annex C)	For storing of mixed image and video content	.heif
MXF	Not yet defined	For storing of video - MXF syntax based	.mxf
Notes: Status is as of 4th April 2019			

Table IV: Profiles with allowed formats and maximum decoded bpp								
Profile	**Main 422.10**	**Main 444.12**	**Main 4444.12**	**Light 422.10**	**Light 444.12**	**Light-Subline 422.10**	**High 444.12**	**High 4444.12**
Profile Indicator (Hex)	0x3540	0x3A40	0x3E40	0x1500	0x1A00	0x2500	0x4A40	0x4E40
Bit depth	8, 10	8, 10, 12	8, 10, 12	8, 10	8, 10, 12	8, 10	8, 10, 12	8, 10, 12
Chroma sampling formats	4:0:0 4:2:2	4:0:0 4:2:2 4:4:4	4:0:0 4:2:2 4:4:4 4:2:2:4 4:4:4:4	4:0:0 4:2:2	4:0:0 4:2:2 4:4:4	4:0:0 4:2:2	4:0:0 4:2:2 4:4:4	4:0:0 4:2:2 4:4:4 4:2:2:4 4:4:4:4
Max decoded bpp	20	36	48	20	36	20	36	48
Vertical Wavelet Decomp.	0-1	0-1	0-1	0-1	0-1	0	0-2	0-2