SMPTE REGISTERED DISCLOSURE DOCUMENT

AVC MXF Proxies

The attached document is a Registered Disclosure Document (RDD) prepared by the proponent identified below. It has been examined by the appropriate SMPTE Technology Committee and is believed to contain adequate information to satisfy the objectives defined in the Scope, and to be technically consistent.

This document is NOT a Standard, Recommended Practice or Engineering Guideline, and does NOT imply a finding or representation of the Society.

Errors in this document should be reported to the proponent identified below, with a copy to eng@smpte.org.

All other inquiries in respect of this document, including inquiries as to intellectual property requirements that may be attached to use of the disclosed technology, should be addressed to the proponent identified below.

In addition to Harmonic, the proponents include AmberFin of Basingstoke, United Kingdom, Avid Technology of Burlington, MA, USA, EVS of Seraing, Belgium, and IPV of Cambridge, United Kingdom. Harmonic will assume the role of proponent contact.

Proponent contact information:

Harmonic Inc.
15220 NW Greenbrier Parkway #290
Beaverton, OR
97006
Attention: Video Server Engineering

email: SMPTE-RDD@harmonicinc.com

Approved
April 1, 2014

Table of Contents

Page

Introduction

This RDD documents an MXF application specification for AVC “Long GOP” proxies with AAC audio. Certain restrictions have been applied to the creation of the MXF files as a result of them being used in a Proxy application.

Note that AVC no longer uses the construct or terminology of a “Group of Pictures” (GOP). This document continues to use the term as “generally understood,” namely a Group of Pictures starting with an I or IDR Picture in coded order.

1 Scope

This RDD defines an MXF Application Profile for AVC proxies with MPEG-2 AAC audio per Operational pattern 1A (OP1a). Specific AVC and AAC coding constraints for each application is specified in specific annexes.

2 Reference Documents

Note: All references in this document to other SMPTE documents use the current numbering style (e.g. SMPTE ST 378:2004) although, during a transitional phase, the document as published (printed or PDF) may bear an older designation (such as SMPTE 378M-2004). Documents with the same root number (e.g. 378) and publication year (e.g. 2004) are functionally identical.

The following standards contain provisions which, through reference in this text, constitute provisions of this registered disclosure document. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this registered disclosure document are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

ISO/IEC 13818-7:2006, Information Technology — Generic Coding of Moving Pictures and Associated Audio Information — Part 7: Advanced Audio Coding (AAC)

ISO/IEC 14496-10:2012 | ITU-T Recommendation H.264, Information Technology — Coding of Audio-Visual Objects — Part 10: Advanced Video Coding

ITU-R BT.601-7 (01/07), Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide Screen 16:9 Aspect Ratios

SMPTE ST 12-1:2014, Time and Control Code

SMPTE ST 378:2004, Television — Material Exchange Format (MXF) — Operational Pattern 1a (Single Item, Single Package)

SMPTE ST 379-2:2010, Television — Material Exchange Format (MXF) — MXF Constrained Generic Container

SMPTE ST 381-2:2011, Material Exchange Format (MXF) — Mapping MPEG Streams into the MXF Constrained Generic Container

SMPTE ST 381-3:2013, Material Exchange Format – Mapping AVC Streams into the MXF Generic Container

SMPTE ST 394:2006, Television — Material Exchange Format (MXF) – System Scheme 1 for the MXF Generic Container

SMPTE ST 405:2006, Television — Material Exchange Format (MXF) – Elements and Individual Data Items for the MXF Generic Container System Scheme 1

SMPTE ST 436-1:2013, MXF Mappings for VI Lines and Ancillary Data Packets

SMPTE RP 224, SMPTE Labels Register

ITU-T Rec. T.35, Procedure for the Allocation of ITU-T Defined Codes for Non-Standard Facilities

IPV: *IPV Embedded Metadata Format*, Part No. P06-000122/A, 2011, http://www.ipv.com/smpte-downloads/.

IPV: *IPV SEI Metadata Insertion*, Part No: P06-000121/1.0, 2013, http://www.ipv.com/smpte-downloads/.

3 MXF Structure

3.1 MXF Physical Structure

The MXF file shall be written as an Operational pattern 1A (OP1a) structured file, compliant with SMPTE ST 378. It should be an interleaved single file, written in a strictly left to right fashion (no post-write random access fix-up), with the value of KLV Alignment Grid (KAG) equal to 1. Further specific constraints are as follows:

Frame wrapped: Each Content Package (CP) shall contain one "Frames worth" of video and audio and shall be interleaved according to the SMPTE ST 379-2 rules:

- When per-frame SMPTE ST 12-1 time code is included in the essence, it shall be encoded in a GC system element, as defined in SMPTE ST 394 and SMPTE ST 405.
- When there is VI or VANC data in the video, then each CP shall contain one SMPTE ST 436-1 KLV item.
- Each CP shall contain one KLV for each video frame as a byte stream NAL Unit structure, starting with an access unit delimiter NAL Unit. Wrapping shall comply with the requirements of SMPTE ST 381-3.
- Each CP shall contain one KLV for each AAC audio frame, which may contain zero, one or two AAC frames. (Note that the durations of the video and audio frames are different and that this may give rise to bitstreams that have zero, one or two AAC frames in a content package). Wrapping shall comply with the requirements of SMPTE ST 381-2.

Index Tables: Every Content Package should be indexed

- The zero Position of the Index table corresponds to the first stored frame in the file.
- The Indexed Position equaling SP::Track(picture)::SourceClip::Origin corresponds to the first displayed frame of video in the file.
- Both video and audio are VBE and the index tables shall therefore contain delta entries for each element and shall contain slices.
- Index tables appear in partitions as detailed below.
- PosTable for AAC tracks is optional.

Header Partition: Shall contain only metadata, no index tables, no essence

Body Partitions: Essence shall be collected into pairs of body partitions containing 10 seconds of essence and 10 seconds of indexing. Body partitions shall contain only essence or only index table segments and shall not contain header metadata. In the case of non-integral frame rates, a duration equivalent to the nearest integral frame count shall be used. Each pair of partitions shall be recorded as follows:

- A Body partition containing the interleaved essence.
- A Body partition containing only one Index Table segment for the same essence.

End of file: At end of file the following sequence shall occur:

- An optional Body partition containing any remaining essence that is of a duration less than 10 seconds.
- A Footer Partition with a closed and complete repetition of the header metadata and the index table segment for the last Body partition.
- A Random Index Pack (RIP):

  Correct construction of a RIP clearly marks which partitions have essence (the RIP will have a non-zero body SID for that partition) and which partitions have indexes (RIP body SID == 0).

3.2 MXF Header Metadata

Time Code: SMPTE ST 12-1 Time code shall be continuous. Decoders shall use the MXF Header Material Package timecode as the lead time code.

Partition: The Picture Essence Container Label shall comply with the requirements of SMPTE ST 381-3. The Sound Element Container Label shall comply with the requirements of SMPTE ST 381-2 (see Section 9.1).

File Descriptor shall be an MPEG Video descriptor (per SMPTE ST 381-3, Tables 3 and 4) with:

Picture Essence Coding Label 06.0E.2B.34.04.01.01.0A. 04.01.02.02.01.31.11.01

31 = profile category (predictive profiles)

11 = profile (AVC Constrained Baseline Profile)

01 = coding variants (AVC unconstrained coding)

Audio Descriptor shall be a MPEG Audio descriptor (per SMPTE ST 381-2, Section10.4) with:

Essence Coding 06.0E.2B.34.04.01.01.03. 04.02.02.02.03.03.01.00 for MPEG-2 AAC audio in ADTS

Note: While this UL is currently in use, there may be a different UL assigned in the future and a backwards compatibility strategy will be created if this happens.

The Material Package and the Source Package shall contain tracks for video, audio, time code, and optional SMPTE ST 436 data, as per SMPTE ST 377-1.

Annex A Bibliography (Informative)

Note: All references in this document to other SMPTE documents use the current numbering style (e.g. SMPTE ST 382:2007) although, during a transitional phase, the document as published (printed or PDF) may bear an older designation (such as SMPTE 382M-2007). Documents with the same root number (e.g. 382) and publication year (e.g. 2007) are functionally identical.

ISO/IEC 13818-1:2007 | ITU-T Rec. H.222, Information Technology — Generic Coding of Moving Pictures and Associated Audio Information: Systems

ISO/IEC 14496-3:2009, Information Technology — Coding of Audio-Visual Objects — Part 3: Audio

ISO/IEC 14496-15:2010, Information Technology — Coding of Audio-Visual Objects — Part 15: Advanced Video Coding (AVC) File Format

SMPTE ST 382:2007, Material Exchange Format — Mapping AES3 and Broadcast Wave Audio into the MXF Generic Container

Amendment 1:2012 to SMPTE ST 382:2007

Amendment 2:2013 to SMPTE ST 382:2007

SMPTE ST 2041-2:2010, Format for Non-PCM Audio in AES3 — MPEG-2 AAC and HE AAC Audio in ADTS

SMPTE ST 2041-3:2010, Format for Non-PCM Audio and Data in AES3 — MPEG-4 AAC and HE AAC Compressed Digital Audio in ADTS and LATM/LOAS Wrappers

Annex B Coding Constraints

B.1 Video Coding

B.1.1 Source Formats

All proxies shall be encoded using progressive video. See Annex C for additional guidance on proxy dimensions and aspect ratios.

Note: AVC essence is to be encoded per ISO/IEC 14496-10 and not per ISO/IEC 14496-15.

B.1.1.1 Frame Rates

24 fps, variable GOP lengths of up to 12 frames

24/1.001 fps, variable GOP lengths of up to 12 frames

25 fps, variable GOP lengths of up to 12 frames

30 fps, variable GOP lengths of up to 15 frames

30/1.001 fps, variable GOP lengths of up to 15 frames

50 fps, variable GOP lengths of up to 12 frames

60 fps, variable GOP lengths of up to 15 frames

60/1.001 fps, variable GOP lengths of up to 15 frames

B.1.1.2 Frame Sizes

B.1.1.2.1 Frame Dimensions

The sample width of the encoded proxy video shall be no greater than 960 and no less than 176. The sample height of the encoded video shall be no greater than 544 and no less that 112. Both the width and the height shall be a multiple of 16.

B.1.1.2.2 Aspect Ratio

The aspect ratio for the encoded video shall be included in both the AVC video (as SAR values in the VUI parameters of the Sequence Parameter Set) and the MXF Wrapper (as Optional Picture Descriptor Properties values for DisplayWidth and DisplayHeight). Note that the MXF Wrapper also holds “Stored” and “Sample” Width and Height values, along with “Sample” and “Display” offsets, thereby allowing a crop to be applied to the decoded video before it is rendered. See Annex C for a detailed discussion of this topic.

B.2 Video Compression Constraints

B.2.1 Coding

Video compression shall be compliant with AVC Constrained Baseline profile, using variable bitrate encoding, selectable between 0.6 and 6 Mbps. No B frames or forward temporal references are permitted.

The following additional provisions shall apply:

1. There shall be one SPS and one PPS for the duration of the stream.

2. The SPS and PPS shall be placed at the beginning of the Content Package containing the first IDR or I-frame.

3. The SPS and PPS should be repeated with every IDR and I-frame.

B.2.2 Pre-charge and Rollout

First displayed image shall occur within the first 2 GOPs.

Origin value of the Source Package shall the number of frames of pre-charge.

Duration value of the Material Package shall be the total number of displayed frames.

Duration value of the Source Package shall be the sum of the pre-charge and displayed frames.

Rollout of one or more frames is allowed. Rollout frames shall not be included in the Material and/or Source Package durations. The existence of rollout frames can be determined by comparing the length of the Index and the length of the Source Package.

See SMPTE ST 377-1 MXF Timing Model for more details.

B.2.3 Encapsulation

Byte stream NAL Units inside each MXF Frame with stream ID = 0, starting with an access unit delimiter NAL Unit.

B.3 SEI Metadata Insertion for Video

Additional video metadata can be inserted in the AVC stream using Supplemental Enhancement Information elements as per ISO/IEC 14496-10. The Metadata SEI should be positioned within the video stream before the first IDR (0x01) or non-IDR (0x05) NAL Unit for each video frame.

B.3.1 NAL Unit Type

Per Table 7-1 in ISO/IEC 14496-10, SEIs use NAL Unit type 0x06. There shall not be any other messages in the NAL unit.

B.3.2 SEI Type

The SEI Type is 0x04 (User Data registered by ITU-T Rec. T.35). The value for itu_t_t35_country_code shall be 0xB5 (United States). The value for itu_t_35_provider_code shall be 0x002B (Harmonic Inc.).

B.3.3 Payload Length

The Payload Length is encoded as a series of FF bytes, the number being the payload length divided by 255, followed by a byte of the payload length modulo 255. The length itself comprises the data which follows the Payload Length field, namely: SEI GUID; and Payload (see below).

B.3.4 UUID

The UUID for IPV Metadata SEI blocks is (all numbers in hex)
56.77.C2.77.F3.58.47.48.89.97.9F.39.EC.FD.F1.B9.

B.3.5 Payload

The Payload for IPV Metadata SEI blocks can be generated either by the *Metadata Library* or, manually, in accordance with the *Embedded Metadata Format* specification.

B.3.6 Example (all numbers in hex)

The following is an example of a Metadata SEI. See the *Embedded Metadata Format* specification for an explanation of the payload data. The bytes are presented in byte stream order.

`00 00 00 01 06`	NAL Unit
`04`	Payload Type (User data registered by T.35)
`B5 00 2B`	T.35 identifiers
`3E`	Payload Length (3+16+43)
`56 77 C2 77 F3 58 47 48 89 97 9F 39 EC FD F1 B9`	UUID
`00 49 00 50 04 56 D2 01 80 3F 80 01 01 00 00 80`	Payload [0..15]
`90 01 91 02 92 03 93 04 1E C1 C2 C3 C4 B1 B2 B3`	Payload [16..31]
`B4 1F 85 8A 8F 94 99 9E A3 A8 AD`	Payload [32..42]

B.4 Audio Coding

Coding: MPEG-2 AAC LC profile 16 bit samples @ 64, 96 or 128 Kbps

Sample rate: Audio shall have an Audio Sampling Rate of 48 kHz

Configuration: 1 to 8 stereo pairs. Each pair may contain stereo or mono audio.

Wrapping: ADTS

B.5 Ingest Considerations (Informative)

Ideally, the audio codec should free run, using silence as the input, pending the presence of non-silent user signals. Care needs to be taken in implementation to maintain audio and video synchronization during ingest and ensure the resulting proxy files can be properly aligned by editors or playout devices.

The preferred method for recording audio offsets uses the index table in ST 377-1, however, as an alternative. IPV currently records an audio offset element in a 'dark' key (`0x06,0x0E,0x2B,0x34, 0x01,0x01,0x01, 0x07,0x0E,0x08,0x04,0x01,0x01,0x01,0x02,0x00`) immediately after the CP System Item element. It simply specifies (for each audio channel) the number of audio ticks that precede the video frame.

B.6 Playout Considerations (Informative)

In addition to baseband video and audio, playout devices might be expected to support MPEG-2 TS (ISO/IEC 13818-1) compressed streams. In such a case, there might be customer desire for re-wrapping the AAC audio into LATM/LOAS in some circumstances, as well as signaling MPEG-4 AAC (audio coding is the same).

Annex C Proxy Sizes and Aspect Ratios (Informative)

C.1 Common Source Sizes and Aspect Ratios

The following table gives the common source video size and aspect ratio values. In principle, each size is available in both interlaced and progressive forms, although this has no effect on the dimensions or aspect ratios. Note that the table gives the sample width and height of the encoded material: this is not to be confused with the display width and height, which only affect the way the video is eventually rendered. The final aspect ratio is given in one of two ways: DAR (display aspect ratio), which defines the shape of the window into which the video is to be rendered; or SAR (sample aspect ratio), which defines the shape of each pixel. Using SAR has the advantage that it is unaffected by cropping and/or addition of extra lines (e.g. time code underburn), whereas DAR needs to be altered if the image dimensions are changed.

Sample Width	Sample Height	DAR Width	DAR Height	SAR
1920	1080	16	9	1.0000
1440	1080	16	9	1.3333
1280	1080	16	9	1.5000
1280	720	16	9	1.0000
720	576	4	3	1.0909
720	576	16	9	1.4545
720	480	4	3	0.9091
720	480	16	9	1.2121

C.2 Proxy Encodings

Because proxy material is usually scaled down compared to the source, it will almost inevitably end up being encoded in a progressive manner. This is because the source will either have been through a de-interlace procedure as part of the capture process, or because only one of the original fields is actually encoded. The encoding format also imposes restrictions, such as maximum/minimum dimensions or specific pixel alignment criteria. As a result, proxy recordings often need to be either cropped or padded in order to comply with the encoding standard.

AVC encodes video using 16×16 pixel macroblocks, so the width and height values are always multiples of 16. Because this is not always convenient (1080 is not a multiple of 16, for example), AVC also provides a mechanism to crop the rounded dimensions (1920×1088) back to the original values (1920×1080). Again, using SAR rather than DAR removes any ambiguity when it comes to handling the rounding of image sizes.

Per the requirements of Annex B.1.1.2.2, only the DAR values are placed in the MXF header metadata.

C.3 Unrounded Scaling

The following table gives the common video dimensions when applying simple scales of 2 and 4 to both the width and the height. No rounding has been done.

Original Width	Original Height	Scale/2 Width	Scale/2 Height	Scale/4 Width	Scale/4 Height
1920	1080	960	540	480	270
1440	1080	720	540	360	270
1280	1080	640	540	320	270
1280	720	640	360	320	180
720	576	360	288	180	144
720	576	360	288	180	144
720	480	360	240	180	120
720	480	360	240	180	120

C.4 Round Down Scaling

The following table gives the common video dimensions when scaled by factors of 2 and 4, but rounding the width and height down to the nearest multiple of 16. Encodings made in this manner would need to be cropped relative to the unrounded scaling, and this is usually done in a symmetrical manner. For example, 480×256 material would have 14 lines cropped from the height compared to the nominal size of 480×270 —7 lines each from the top and the bottom. Again, using SAR rather than DAR means that the aspect ratio is unaffected by any crop operation.

Original Width	Original Height	Scale/2 Width	Scale/2 Height	Scale/4 Width	Scale/4 Height
1920	1072	960	528	480	256
1440	1072	720	528	352	256
1280	1072	640	528	320	256
1280	720	640	352	320	176
720	576	352	288	176	144
720	576	352	288	176	144
720	480	352	240	176	112
720	480	352	240	176	112

C.5 Round Up Scaling

The following table gives the common video dimensions when scaled by factors of 2 and 4, but rounding the width and height up to the nearest multiple of 16. Encodings made in this manner would need to be padded relative to the unrounded scaling, and this is usually done with black pixels and in a symmetrical manner. For example, 480×272 material would have 2 black lines added to the height compared to the nominal size of 480×270 —1 line each to the top and the bottom. Again, using SAR rather than DAR means that the aspect ratio is unaffected by any pad operation.

Original Width	Original Height	Scale/2 Width	Scale/2 Height	Scale/4 Width	Scale/4 Height
1920	1088	960	544	480	272
1440	1088	720	544	368	272
1280	1088	640	544	320	272
1280	720	640	368	320	192
720	576	368	288	192	144
720	576	368	288	192	144
720	480	368	240	192	128
720	480	368	240	192	128

Annex D Initial User Requirements (Informative)

The following user requirements were captured in order to scope the Proxy design.

Wrapper: MXF OP1a (supports play while record)

Video: AVC Constrained Baseline profile

no B-Frames

constant Bitrate (1000, 1500 and 2000 kBit/s, preferred flexible)

resolution 640 x 360 px (preferred flexible)

25 fps (preferred flexible)

Audio: 4 Stereo Channels

AAC LC-profile

16 bit,

192 kbps

Latency: System latency of sub-400ms between user-GUI action and video response is desired.

Annex E Growing Proxies

In play-while-record workflows, it is possible to handle growing proxy files. In that case, the recorder is writing the MXF AVC proxy file on a shared storage device and the reader is reading the file while it is recording.

To enable an efficient seek and playback, the growing MXF file shall be compliant with the following requirements:

- The Duration in the Header Metadata of the Header Partition shall be set to the distinguished value = -1.
- All Best Effort properties within the MPEG Video Descriptor of the Header Metadata in the Header Partition shall contain actual values and shall not contain distinguished values.
- The optional properties "BitRate", "MaxGOP" and "Picture Essence Coding" within the MPEG Video Descriptor of the Header Metadata in the Header Partition shall contain actual values.
- The MXF file shall be partitioned according to Section 3.1 of this document.
- Index table segments must be distributed in the MXF file according to Section 3.1 of this document.
- Each pair of body partitions shall contain 10 seconds of essence and 10 seconds of indexing according to Section 3.1 of this document.

Annex F Unresolved Comments

During the Ballot process, there were several comments received that were unable to be resolved to the satisfaction of the commenter. According to the provisions of the Standards Operations Manual, a consensus decision of Technology Committee 31FS recommended to the Standards Vice President to append these comments.

The rationale expressed by the commenter for submitting the comments is as follows.

"An RDD by definition is a manufacturer's implementation of a standard that may deviate from the normative provisions of that standard. Although the title of this RDD is "AVC MXF Proxies" and the Scope refers to an MXF Application Profile, this RDD defines an AAC mapping that differs from the MXF mapping that is planned to be standardized by SMPTE. The resulting output file might, therefore, not be interoperable with future standardized AAC MXF files."

Accordingly, the commenter requested that “MXF” be removed from the Title and the Scope and the RDD proponents declined.