root6
technology for the creative

15th February 2011, 13:00 Neal Kemsley and Rupert Watson

101 Digital media: the size, the shape, and the flavour

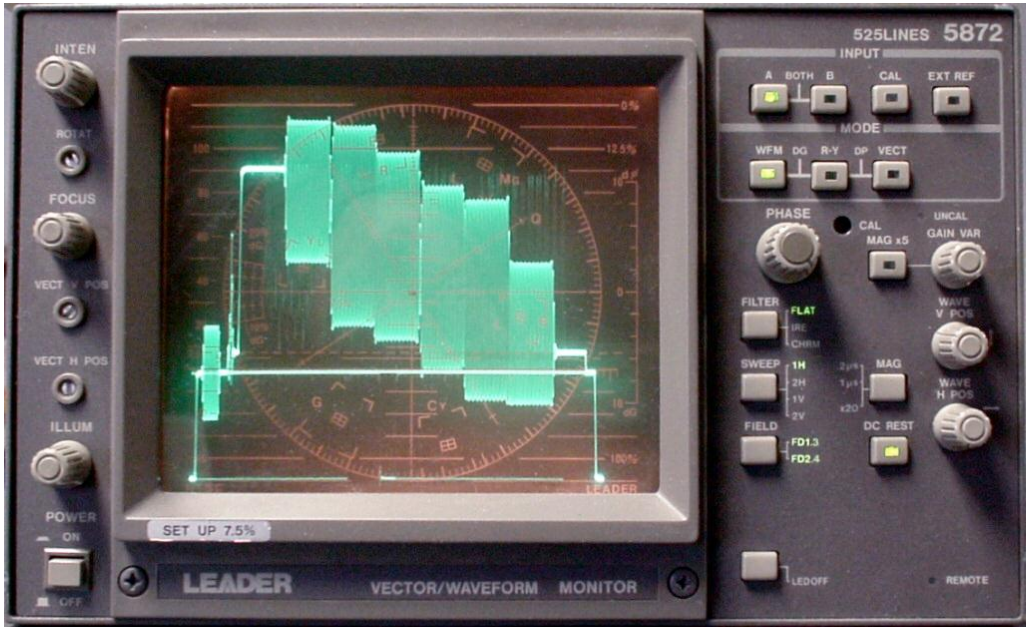
Since the 1970s digital imaging has been slowly making its way into television and film. Early digital video standards were not ideally suited to computer graphics and it's only in the last few 1
years that those problems have been resolved.

- History – what did video look like?
- Computer Graphics Formats – MPEG and beyond
- Editing vs Transmission codecs
- Contemporary MPEG4 codecs
- MXF
- Q&A

This is by its nature an introduction – you can't cover in a few hours what university training takes months over but it will give you a confidence in the basics to start investigating for yourself.

www.root6.com +44 (0) 20 7437 6052

History – what did video look like?



- Until the 1970s video was entirely analogue and didn't have pixels.
- Everything from the monochrome video level (luminance information) to the colour content (chrominance information) and all of the synchronising information was encoded onto an analogue signal.
- Cameras, videotape, telecine, colour-correction, and captioning was all done in the analogue domain with not a hint of digital imaging.

This all changed in the mid 70's with three requirements; Timebase correction, synchronisation and standards conversion.

www.root6.com +44 (0) 20 7437 6052

Timebase Correction

- 1″ and ¾″ Umatic VTRs have an inherently unstable timebase – the stability of the video signal is insufficient to mix with camera sources etc.
- The only way to stabilise and lock to the station reference is to turn the analogue off-tape signal into a digital representation, write it into a store and read it out with the super-stable station genlock.
- In the early seventies being able to store eight lines of video required a large unit that typically sat underneath the VTR.

Synchronisation

- In a television studio all the cameras and VTRs are locked to a common reference which allows those sources to be seamlessly mixed with each other.
- In a studio centre this is easy to achieve – with an outside broadcast contribution or a camera in a helicopter not so much!
- Montreal 1976 - first Olympics that had live coverage from helicopters as well as other remote cameras. This was achieved with early-model synchronisers from Quantel – the DFS 3000.
- A frame-store synchroniser operates much like a TBC but it has a whole video-frame (625 lines, a 25th of a second) of storage.

Montréal 1976

Standards Conversion

Different territories around the world use different standards for their television;

- PAL – common in Europe, 625 lines per frame, 25 interlaced frames per second
- NTSC – Common in the Americas, 525 lines per frame, 30 interlaced frames per second

It turns out you need eight frames of storage to do good quality standards conversion.

Lots of material being produced in the US that needed to be transmitted/consumed in non-NTSC territories (and vice-versa).

Digital Video Effects, Painting systems, Slow Motion machines.

In subsequent years all of these devices started to be used in television production and post but it wasn't until the introduction of D1 VTRs in the late eighties that it became common place to interconnect equipment digitally (using the seminal rec-601 system) rather than via their analogue i/o.

Until the 601 standard different manufacturer's equipment operated internally at whatever raster the designer had landed on.

CCIR rec 601 – Standard Definition Digital Video

Originally the 1982 standard defined;

- 4 x 3 aspect ratio
- 720 pixels x 576 lines – enough pixels for 5.5Mhz video
- Y Cb Cr luminance/colour encoding at 4:2:2 data rate – half res colour difference sampling.

There are several things worth noting;

- 4 x 3 display with 720 x 576 gives non-square pixels (almost square, but not quite)
- When 16 x 9 came along pixels got very non-square, same 720 x 576 resolution.
- 4:2:2 COLOUR SAMPLING is a form of compression used to cram two colour difference signals into the same data channel space of a single chroma signal (the Human eye is tuned to be more sensitive to Luma detail and hence the Luma signal is not down sampled)
- Colour space & sampling structure unlike graphics formats (which sample the Red, Green, and Blue aspect of each pixel hence RGB)

Remember – at this point Photoshop was still pre v.1 and 601 served the needs of TV images.

Uncompressed video and codecs

- The data rate of uncompressed standard def video is 270Mbits^{-1}
- High def comes in at 1.48Gbits^{-1} and 3Gbits^{-1}
- These data rates are too high to record on videotape or send over a network, and expensive to store on hard drives both in terms of data capacity and storage bandwidth.
- Using mathematical techniques, the digital data that represents pixels – colour and luminance values – are transformed into a description that allows the pixels to be re-constituted and hence occupies much less space
- Depending on the application, video can be compressed to 10% or less of its original size.
- The particular mathematical function used to achieve this is called a CODEC - a combination of the words COmpressor – DECompressor.
- Different codecs have pros and cons depending on application (shoot, edit, TX etc)

- **Early computer-based graphics and video formats**

With the exception of MPEG most computer video formats tend to;

- Have square pixels (because computer monitors do)
- Use RGB for their colour representation
- Use varying frame rates (from 12fps up)

None of these lend those early computer video systems to television!

image: Wikipedia

The same can be said of computer still image formats – TIFF, Targa, BMP, etc – also;

- They may use CMYK colour space (used by the print industry)
- Graphics software may work in DPI rather than absolute resolutions
- Varying degrees and quality of anti-aliasing

How on Earth has any of this been reconciled?!

Video Compression techniques fit into two groups:

- **Intra-Frame** compression reduces the data rate for every pixel sampled in the video frame but maintains a complete video picture per frame. Examples of this type of compression scheme are Avid's JFIF mechanism (which is in turn based on JPEG technology) and DV.
- **Inter-Frame** compression which not only reduces the "per-pixel" data rate but it also, conditionally, reduces the number of pixels sampled in each frame. Examples of this type of compression scheme are D10 IMX, AVC & H.264 which are based on MPEG technology

www.root6.com +44 (0) 20 7437 6052

Software Codecs vs Hardware Codecs

The codec function can be performed by modules of software in an application or they might be performed by dedicated hardware modules.

- Software codecs have the advantage of typically being straightforward to update as code gets more efficient or as techniques produce cleaner results. Processing efficiency may scale depending on efficiency of the software code and the number of CPU cores available in the computer host.
- Hardware codecs might be less flexible (although firmware updates may be possible depending on the architecture of the module) but can offer realtime or faster processing.
- The trade-off with any video compression scheme is that the heavier the codec compresses a video signal, the less data is consumed by the resulting data stream however the subjective quality of the resulting video frame data typically also reduces.
- A compromise must usually be found in terms of the efficiency of the codec selected for a particular purpose, the quality of the resulting video output as result of use of the codec, and the processing time and power required to apply the codec.

Discrete Cosine Transform (DCT)

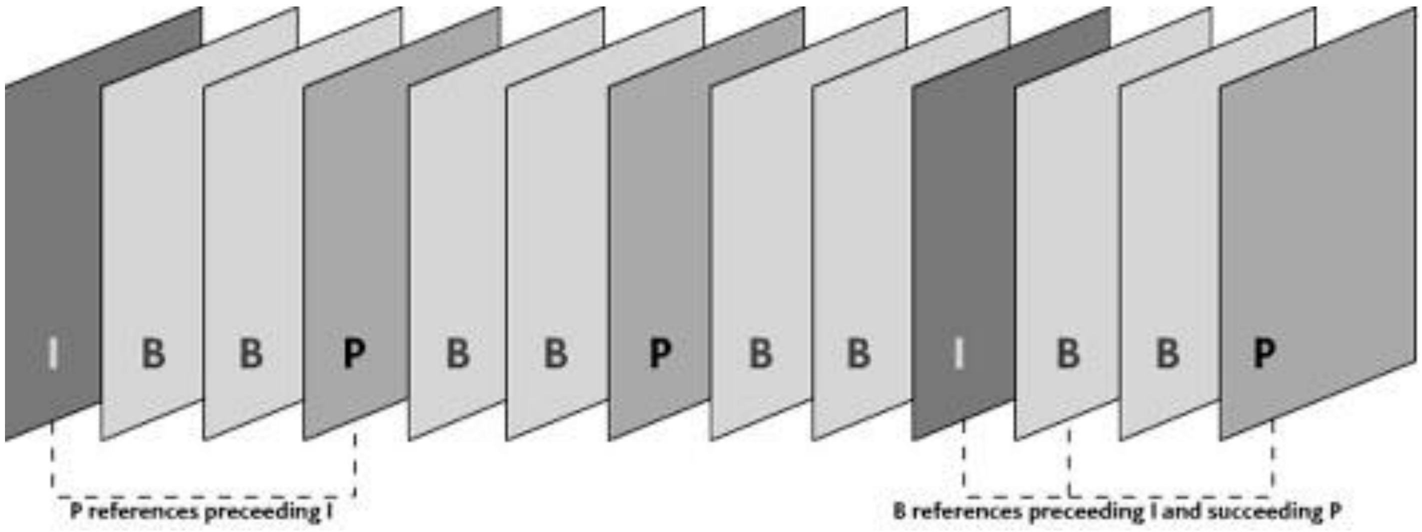
Intra-Frame compression is also known as "Spatial" compression and is based on a mathematical model called Discrete Cosine Transform (DCT). A codec using DCT will analyse each individual video frame and reduce the amount of data needed to describe the video frame by eliminating redundancies in the data. Typically high frequency and resolution detail data is eliminated from the frame based on work done in human visual perception studies to determine what the eye can detect particularly in moving images.

MPEG – Motion Picture Expert Group

MPEG technology uses a technique of analysing a video stream for redundant information across a sequence of video frames and is also known as "Temporal" compression since its analysis is performed on video frames over time.

- This compression scheme divides the video stream into frame sequences known as a Group of Pictures (or GOP), normally using the starting frame as a reference frame called an Intra-Frame (or I-Frame) to mark the beginning of the GOP.
- DCT compression is first applied to this reference frame to reduce the data capacity required to describe its data.

- Once this I-Frame is created, subsequent video frames are similarly compressed. However these subsequent frames are also compared to the I-Frame to determine whether information in these following frames has changed or whether it is repeated from the previous frame. If the information is static and has been repeated in the subsequent frame, it is considered to be redundant and can be eliminated. Detail motion prediction and frame re-ordering techniques ensure that the material can be accurately decoded.



An example of the Group Of Pictures (GOP) Frame Sequencing Used in MPEG Compression

- **I-frame:** An intra-frame, or I-frame, is a video frame which has been encoded without any reference to any other frame. An MPEG video file will always start with an I-frame and will have subsequent I-frames added at regular intervals. I-frames are also known as key-frames and are important for random access of video files such as rewind, fast-forward and seek operations. The downside to an I-frame is that they are the largest in terms of size as the whole video frame is encoded every time.
- **P-frame:** A predictive inter-frame, or P-frame uses previous I or P-frames as a reference when encoding. This means a P-frame will analyze a previous I or P-frame for any static elements which do not change between frames. Any areas which do not change are not encoded therefore a P-frame only stores video which registers movement making them much smaller than I-frames. The downside to P-frames is that they are sensitive to transmission errors because of their dependency on earlier frames.
- **B-frame:** A bi-predictive inter frame, or B-frame makes reference to both a preceding reference frame as well as a future reference frame. Using B-frames improves the prediction and ultimately the quality of decoded video but it also increases the processing requirements and latency.

- The Inter-Frame analysis is performed in small square subsections of the video frame called Macro Blocks and these are created by dividing the video frame into a mosaic of tiled pixel analysis areas (this perhaps explains why MPEG compression artifacts created when for example a broadcast is interrupted, cause the distinctive "tiling" picture break up).

MPEG data streams can be of a single track type such as video only, or they may have other tracks embedded in the data. In the case of the former, this kind of presentation is known as an Elementary Stream. In the case of the latter, where audio tracks are included in the stream, this kind of presentation will be known as a Multiplexed (or Mux'd) stream.

MPEG Profiles and Levels

MPEG Technology uses a range of Compression Schemes organised by "Profiles" and "Levels"

- **Profiles** specify the compression technique used and elements of syntax
- **Levels** specify what Bit Rate is used for the data stream

For example DVD is described as "Main Profile at Main Level" and you might see this written as (MP @ ML). A distribution quality stream such as that used by US broadcasters is Main Profile at High Level (MP @ HL) and IMX recorders use the 4:2:2 Profile at Main Level (or 4:2:2 P @ ML).

	Profile	Simple	Main	SNR	Spatial	High	4:2:2
	Frame Type	I & P	I, P & B	I, P & B	I, P & B	I, P & B	I, P & B
	Chroma Sampling	4:2:0	4:2:0	4:2:0	4:2:0	4:2:0 & 4:2:2	4:2:0 & 4:2:2
High Level	Samples/Line Line/Frame Frames/Sec Max Bit-rate (Mbs)		1820 1152 60 80 **HDTV**			1920 1152 60 100	1920 1088 30 300 **HDTV**
High 1440 Level	Samples/Line Line/Frame Frames/Sec Max Bit-rate (Mbs)		1440 1152 60 60		1440 1152 60 60	1440 1152 60 80	
Main Level	Samples/Line Line/Frame Frames/Sec Max Bit-rate (Mbs)	720 576 30 15	720 576 30 15 **SDTV**	720 576 30 15		720 576 30 20	720 576 30 50 **SDTV**
Low Level	Samples/Line Line/Frame Frames/Sec Max Bit-rate (Mbs)		352 288 30 4				

MPEG Profiles and Levels (From a Sony IMX Brochure)

Progression of MPEG Technologies

- **MPEG-1**, 1993 – The basis for CDi and VideoCD (early DVD predecessors) – only 1.5Mbit/s and a quarter-screen resolution of 350x288 pixels. No interlaced video, 4x3 and stereo audio only. Long GOP
- **MPEG-2**, 1995 – Full standard def video, basis of DVD and DVB-T, C & S. Variable data rate and full 601 resolution using Y, Cr, Cb colour sampling. 16x9 or 4x3 video, up to 5.1 audio via AC3. 23.976, 25, and 29.97 FPS. Long GOP or i-frame (TX vs editing).
- **MPEG-4**, 1998 –The basis of most modern video encoding for acquisition, post and delivery. Multi-resolution, multi-framerate, multi-audio standards, editing or TX variants.

technology for the creative

Domestic delivery -> TX -> Editing

Horses for Courses

The Right Tool for the Right Job!

www.root6.com +44 (0) 20 7437 6052

technology for the creative

Editing vs Transmission codecs cont.

The requirements of editing and transmission differ somewhat;

- Editing requires immediate access to each video frame and should not have to build a complete frame by looking at the frames that surround it, so typically either codecs based on Intra-Frame compression are used or Inter Frame compression based on an **I-Frame** only (or 'short-GOP' codec) are used.
- Transmission would rather leverage the additional image quality available to a long-GOP system and so a 12-frame (typical) GOP is used.

Rule of five;

- Uncompressed standard definition video ~ 250Mbits/sec
- I-Frame editing codec, MPEG2 ~ 50Mbits/sec
- Long GOP transmission codec, MPEG2 ~ 10Mbits/sec
- Statistical Multiplexed DVB stream to the home ~ 2Mbits/sec

Contemporary standards – HD and MPEG4

HD video makes things considerably easier – the Rec 709 standard defines;

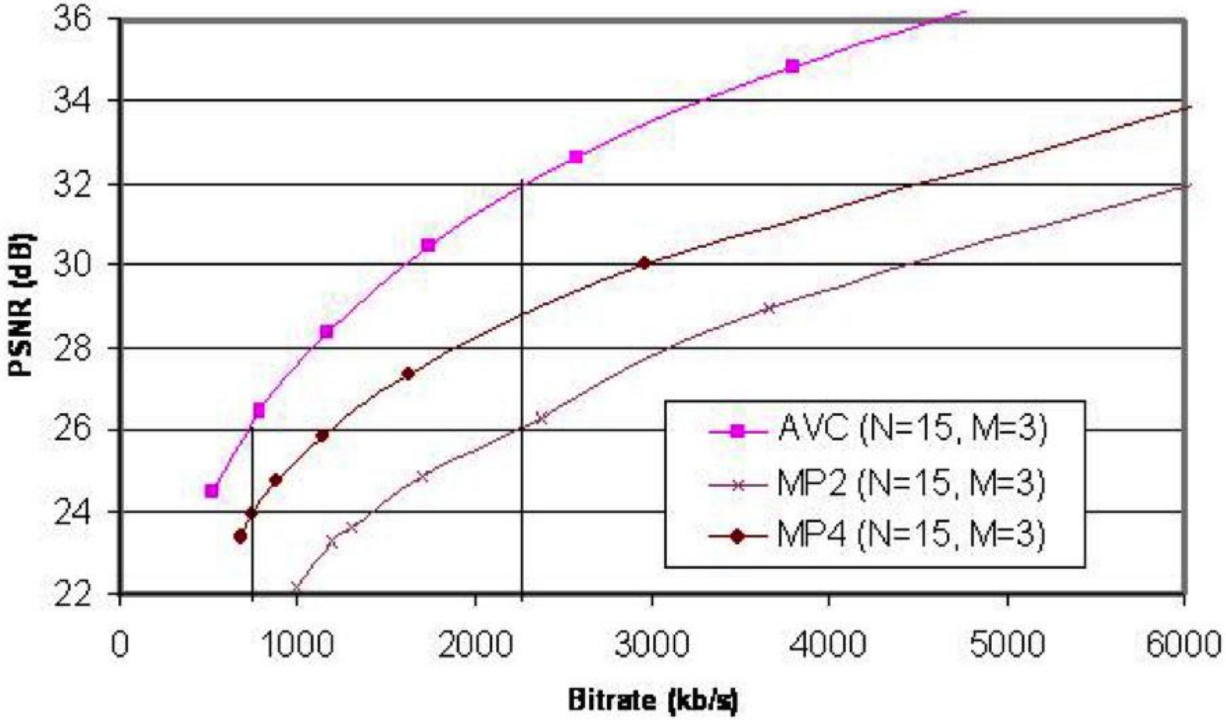
- 1920x1080 resolution at 16x9 now has square pixels!
- Wider variety of frame rates (including 24 PsF)

MPEG4 is the basis for most acquisition, editing and transmission codecs currently.

- MPEG4 improves the performance of MPEG2 by using variable sized macroblocks.
- MPEG4-part 10 (aka 'H.264' or AVC) further improves performance by allowing macroblocks to the referenced across I-Frame boundaries.

Qualitative comparison of MPEG2, MPEG4 & H.264

Video reams at the same resolution, encoded three times.



root6
technology for the creative

Contemporary editing & acquisition formats

- Avid DNX HD – I-frame only (i.e. editing) **codec**.
- Quicktime – Apple's wrapper format, especially **ProRes codec**.
- MXF –a 'universal' wrapper format that can encapsulate different codecs
- MPEG2 – the **codec** used in HDV, XDCam
- DV – I-Frame only **codec**, initially domestic cameras but extends up to HD (100Mbits^{-1})

You have to distinguish between ***codecs*** (that mathematical function that changes raw pixels into a file-description of how to re-create the pixels; compression) and ***container*** formats (AKA wrapper).

XDCAM formats

Format name	Container	Video coding	Bit depth	Color sampling	Frame size	Frame rate and scanning type	Video bit rate, Mbit/s	Audio coding
DVCAM	MXF, DV-AVI	DV	8	4:2:0	720x576	25i, 25p	25 (CBR)	PCM 4 ch/16 bit/48 kHz
				4:1:1	720x480	29.97i, 29.97p		
MPEG IMX	MXF	MPEG-2 422P@ML	8	4:2:2	720x576	25i, 25p	30 (CBR), 40 (CBR), 50(CBR)	PCM 8 ch/16 bit/48 kHz, or 4 ch/24 bit/48 kHz
					720x480	29.97i, 29.97p, 23.98p		
MPEG HD	MXF, MP4	MPEG-2 MP@H14/HL	8	4:2:0	1920x1080	29.97i, 25i, 29.97p, 25p, 23.98p	35 (VBR)	PCM 4 ch/16 bit/48 kHz
					1440x1080	29.97i, 29.97p, 23.98p (pulldown), 25p	18 (VBR), 25 (CBR), 35 (VBR)	
					1280x720	59.94p, 29.97p, 23.98p, 50p, 25p	25 (CBR), 35 (VBR), 19 (CBR)[1]	
MPEG HD422	MXF	MPEG-2 422P@HL	8	4:2:2	1920x1080	29.97i, 25i, 29.97p, 25p, 23.98p	50 (CBR)	PCM 8 ch/16 bit/48 kHz, or 4 ch/24 bit/48 kHz
					1280x720	59.94p, 50p, 23.98p (pulldown)		
Proxy AV	?	MPEG-4 Part-2 (ASP)	8	?	CIF (50i - 352x288)	?	1.5	A-Law 4 ch/8 bit/8 kHz

image: Wikipedia

- CBR = Constant Bit Rate, VBR = Variable Bit Rate
- various codecs (DV, MPEG2, MPEG4)
- data rates (18Mbits^{-1} to 50Mbits^{-1})
- resolution/colour sampling (352x288 -> 1920x1080)
- wrappers (MXF, AVI, Quicktime)

As used in the Sony XDCam product line.

Professional Disc

image: Wikipedia

Just for illustration;

- Blue Ray recordable disk
- SxS Memory cards
- SD memory cards

SxS memory card

Recording formats as used in the Sony XDCam product line.

SDHC memory cards

All of these – codec, media, edit system govern how usable your edit workflow will be.

Rich Media – production is not about just audio and video any more – it is about increasing the value of material by adding and maintaining meta-data for the material.

What the heck is meta-data?

Well in a nutshell it is information about information. A good example of meta data is the Index in a book – it is fairly useless information in isolation but is very helpful when looking for a specific topic in the book. Another perhaps more topical example of meta-data is time code.

In the past, production data was written on the tape label or on an index card slid into the tape case, and if you were lucky, this got transferred into the tape library database. With contemporary file formats it is possible to embed meta-data into a media file. This might be limited to basic information such as programme start and end times, but could extend to sub-title information, field capture information such as camera exposure data, Look Up Table (LUT) information, or shoot location GPS data etc – the list is practically endless.

What is MXF?

In 2002 Bruce Devlin of the AAF organisation defined MXF thus:

"The Material eXchange Format (MXF) is an open file format targeted at the interchange of audio-visual material with associated data and metadata. It has been designed and implemented with the aim of improving file based interoperability between servers, workstations and other content creation devices".

There have been several competing formats that aimed to deliver similar levels of interoperability such as Grass Valleys's GXF format however MXF is arguably the most widely accepted file interchange format in use today by broadcasters.

The MXF standard is classed as a "Wrapper" format because it is compression format independent. An MXF file can be seen as being a container holding (Video and or Audio) data tracks or essences stored in the file along with extensive metadata information. Each of the tracks have the capability of having Unique Media Identifier tags appended to them known as UMIDs. The concept is illustrated in the following slide:

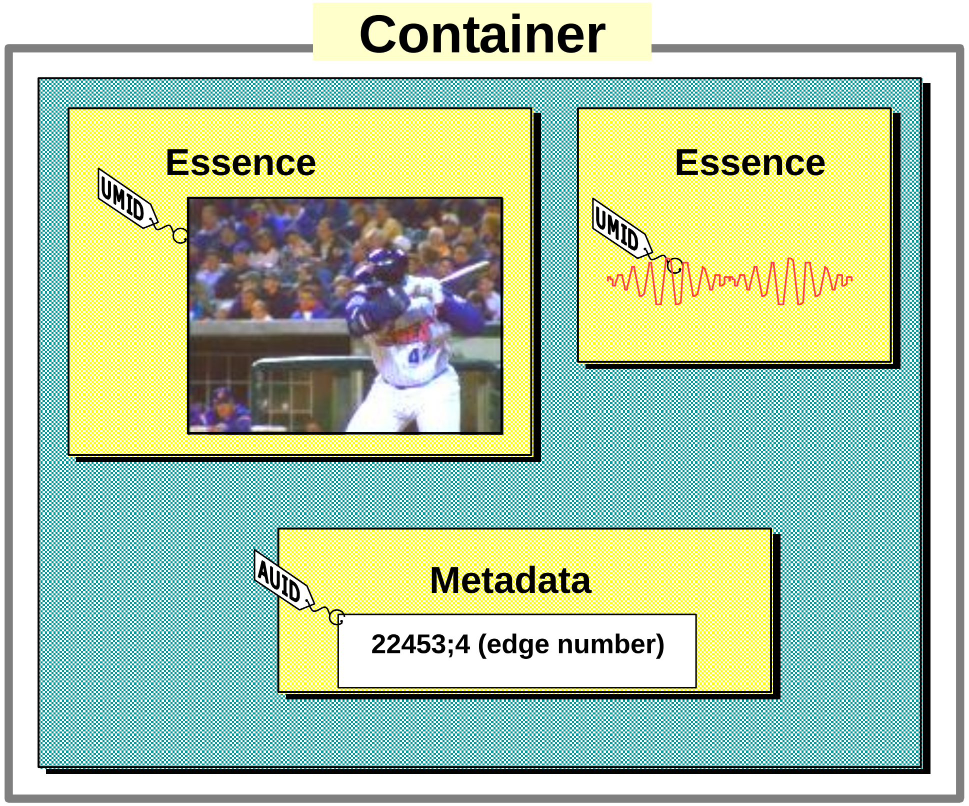
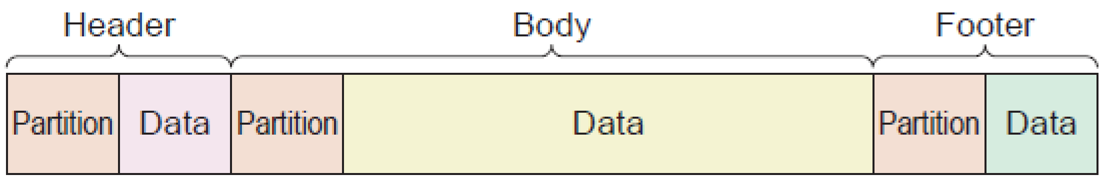


Image courtesy of the AAF Organisation

www.root6.com +44 (0) 20 7437 6052

MXF File Structure

MXF Files have a distinct structure including both a "Header" and a "Footer" section where meta-data relating to the media data "body" or payload can be embedded.



An Example of an MXF File Structure Illustrating the "Header", Data "Body" and "Footer" information separated by Partition Information

- The advantage of the Header and Footer data is that a file can be created at the beginning of capture and the header loaded with whatever meta-data is available at that time.
- As the file grows as the capture progresses in time, devices can then access the growing file (perhaps to start editing the material with non-destructive editors).
- Once the capture is complete, any additional meta-data for the file can then be written into the file footer.

MXF Operational Patterns

MXF specifies a number of different capabilities for a variety of applications called Operational Patterns. These define what MXF features, compression types and metadata structures are supported. Two commonly seen examples are as follows:

OP1a – a single MXF file containing the digital representations of video and audio tracks for one or more programmes and any associated meta-data. The Audio and Video tracks or Essence tracks are embedded in the file by the creating device in such a way that when the file is opened by a replay device, the data is seen to be interleaved or multiplexed together into one transport track. This method is commonly used as with archiving systems and as a delivery format prior to transmission and is sometimes known as a "transport" format.

Op Atom – in this format only one essence track (and its associated meta-data) are encapsulated in the MXF file. For example an acquisition device might record video and four audio tracks. In this case 5 OP Atom files would be created (1 file for the video data and 1 file for each of the four Audio data tracks). In 2002 Avid adopted this mechanism for the storage of media files in Editor storages. Avid's application allows the operator to choose the most appropriate resolution for the work he or she is carrying out (JFIF 15:1s to Uncompressed 10 bit HD), and the resulting captured data will be stored as OP Atom MXF media files.

Standard – Get Your Standard...

It was hoped that the MXF file format would become a "Standard" format for file exchange, and initially, as various manufacturers adopted the format it was found that each "Interpreted" the format for their own use. Given the number of variants in the format, most manufacturers of equipment can still say that they have adhered to the "standard" however clearly this has led to some mis-understanding on the part of those poor souls who have to use the equipment in anger.

A consortium of UK broadcasters including the BBC, BskyB, ITV, Channel 4, Channel 5 and S4C called the Digital Production Partnership (http://www.digitalproductionpartnership.co.uk/) have met around a table to agree on a common set of rules for file exchange for both Standard Definition material and High Definition material. This has been published (and is available to download from the above web site).

The standard required for file delivery is the AS-11 specification which is in turn set out by the Advanced Media Workflow Association (http://www.amwa.tv).

- HD must be delivered as a single MXF OP1a file conforming to AMWA spec AS-11 v1.0.
- The AS-11 file must use the 'UK DPP HD shim specification'.
- Active Picture size of 1920 by 1080 pixels recorded at 50 fields per second
- Encoding based on nominal bit rate of 100 Mbit AVC Intra High 4:2:2 Intra profile@level 4.1.
- 4 or 16 track audio PCM encoded at 48Khz to a sampling depth of 24 bits. Frame Interleaved with video data.

root6
technology for the creative

Other Topics To Explore In More Detail

AVC Intra, H.264

Raw Formats – Red, Viper, Sony, Alexa, DPX, CINEON

Audio Formats - WAV, Broadcast WAV, AC 3, MP3, PCM

Audio Sample Rates - 44.1 Khz, 48 Khz, 96 Khz, 192 Khz

Composition Formats – OMF, OMFI, AAF, .XML

Closed caption Formats

Digital Cinema Formats – DCI, DCP, DCDM, Interoperable Media Format (IMF)

More information on MXF can be found here:

http://tech.ebu.ch/docs/techreview/trev_2010-Q3_MXF-1.pdf

http://tech.ebu.ch/docs/techreview/trev_2010-Q3_MXF-2.pdf

www.root6.com +44 (0) 20 7437 6052

In Conclusion

Things were a lot simpler in the days of SD!

- Fixed frame size
- Only two frame rates – PAL 25fps & NTSC 29.97fps
- Videotape – several formats but all usable in an SDi environment
- Uncompressed or only a handful of compression codecs.

However, with careful consideration of shooting format & media, edit workflow and delivery spec a file-based workflow will bring immense benefits. Always best to ask someone who has used your workflow already – there are countless examples of people who have assumed "..it'll just work" (like it would ten years ago) and discovered to their expense that it's more involved.

technology for the creative

Please see `www.root6.com` for details of all of root6's training packages.

If you get **your badge scanned** at the root6 booth **(G30)** then you will be eligible for a discount when attending the full version of this taster session.

TCP/IP for Broadcast Engineers

Audio 101

Video 101

Television QC101

http://www.root6.com/blog
download today's notes

www.root6.com +44 (0) 20 7437 6052