Understanding HD

Your comprehensive guide to
High Definition on a budget

Part One

Avid

1

Chapter 1

Video formats and sampling

- Chroma sub-sampling
- Active picture
- Aliasing
- Anamorphic
- Anti-aliasing
- Aspect ratio
- CIF
- Colour space
- Component video
- Co-sited sampling

Avid

Perhaps one of the most baffling areas of HD and SD, is the shorthand jargon used to describe sampling and colour space, such as RGB 4:4:4, and Y,Cr,Cb 4:2:2. Also the video formats such as 1080/24P sound strange until you get to know them. For a quick initiation, or reminder, about sampling ratios, please read the piece directly below.

4:2:2 etc (Chroma sub-sampling)

The sampling rates used in digital television are described by shorthand that has, in some ways, only a tenuous connection to what it is used to describe. The numbers denote ratios of sampling rates, not absolute numbers, and they need a little interpretation to understand them all. Sometimes these ratios are referred to as 'chrominance (chroma) sub-sampling'.

In most instances the first number refers to luminance (Y), the last two refer to chrominance – the exceptions are 4:4:4, or 4:4:4:4 (more later). The first number is nearly always a 4 and that means that the luminance is sampled once for every pixel produced in the image. There are a very few instances where a lower sample rate is used for luminance. An example is HDCAM, which is generally considered to use 3:1:1 sampling. Sampling at a lower rate than the final pixel rate is known as sub-sampling

The second two numbers describe the sampling frequencies of the two pure colour digitised components of (Red-Y) and (Blue-Y), called Cr and Cb. In line with television's practise of taking advantage of our eye's response which is more acute for luminance than for pure colour, cuts to reduce data tend to be made in the chrominance sampling rather than luminance. The most common studio sampling system is 4:2:2 where each of the two colour components is sampled coincidently with every second luminance sample along every line.

Line 1 →	• Y CrCb	• Y	• Y CrCb	• Y	• Y CrCb	• Y	• Y CrCb	• Y
Line 2 →	• Y CrCb	• Y	• Y CrCb	• Y	• Y CrCb	• Y	• Y CrCb	• Y

4:2:2 sampling of luminance and colour difference signals

4:1:1, used in some DV formats and DVCAM, makes Cr and Cb samples at every fourth Y sample point on every line – but still carries more chrominance detail than PAL or NTSC.

Line 1 →	Y CrCb	Y	Y	Y	Y CrCb	Y	Y	Y	Y
Line 2 →	Y CrCb	Y	Y	Y	Y CrCb	Y	Y	Y	Y

4:1:1 sampling

Then another argument says that if the chrominance is sub-sampled horizontally, as in 4:1:1, why not do the same vertically to give a more even distribution of colour information? So instead of sampling both Cr and Cb on every line, they are sampled on alternate lines, but more frequently on each line (at every other Y). This is 4:2:0 sampling (4:2:0 on one line and 4:0:2 on the next) and it is used in MPEG-2 and most common JPEG compression schemes.

Line 1 →	Y Cr	Y	Y Cr	Y	Y Cr	Y	Y Cr	Y	Y Cr
Line 2 →	Y Cb	Y	Y Cb	Y	Y Cb	Y	Y Cb	Y	Y Cb

4:2:0 provides equal colour resolution vertically and horizontally if using square pixels

In many cases it is very useful to have a key (or alpha) signal associated with the pictures. A key is essentially a full image but in luminance only. So then it is logical to add a fourth number 4, as in 4:2:2:4.

Technically 4:4:4 can denote full sampling of RGB or Y, Cr, Cb component signals – but it is rarely used for the latter. RGB may have an associated key channel, making 4:4:4:4.

Occasionally people go off-menu and do something else – like over-sampling which, with good processing can improve picture quality. In this case you might see something like 8:8:8 mentioned. That would be making two samples per pixel for RGB.

This sampling ratio system is used for both SD and HD. Even though the sampling is generally 5.5 times bigger, 4:2:2 sampling is the standard for HD studios.

Why 4?

Logic would dictate that the first number, representing a 1:1 relationship with the pixels, should be 1 but, for many good (and some not so good) reasons, television standards are steeped in legacy. Historically, in the early 1970s, the first television signals to be digitised were coded NTSC and PAL. In both cases it was necessary to lock the sampling frequency to that of the colour subcarrier (SC), which itself has a fixed relationship to line and frame frequencies. NTSC subcarrier is 3.579545MHz and PAL-I's is 4.43361875MHz and the digital systems typically sampled at 4 x NTSC SC or 3 x PAL SC, making 14.3 and 13.3MHz respectively.

Then came the step to use component video Y, B-Y and R-Y (luminance and two pure colour components – known as colour difference signals) that is much easier to process for re-sizing, smooth positioning, standards conversion, compression and all the other 1001 operations that can be applied to pictures today. When a standard was developed for sampling this component video it followed some of the same logic as before, but this time also sort commonality between the two SD scanning systems used around the world: 525/60I and 625/50I. Putting all that together led to what is now the ITU-R BT.601 standard for SD sampling. '601' defines luminance sampling at 13.5MHz (giving 720 pixels per active line) and each of the colour difference signals at half that rate – 6.75MHz.

The final twist in this tale is that someone then noticed that 13.5MHz was nearly the same as 14.3MHz that was 4 x NTSC subcarrier. Had he looked a little further he might have seen a much nearer relationship to 3 x PAL SC and a whole swathe of today's terminology would be that much different! But so it was that the number that might have been 3 and should have been 1, became 4.

As HD sampling rates are 5.5 times faster than those for SD, the commonly used studio 4:2:2 sampling actually represents 74.25MHz for Y and 37.125MHz for Cr and Cb.

1080I

Short for 1080 lines, interlace scan. This is the very widely used HD line format which is defined as 1080 lines, 1920 pixels per line, interlace scan. The 1080I statement alone does not specify the frame rate which, as defined by SMPTE and ITU, can be 25 and 30Hz.

See also: Common Image Format, Interlace, ITU-R.BT 709, Table 3

1080P

TV image size of 1080 lines by 1920, progressively scanned. Frame rates can be as for 1080I (25 and 30Hz) as well as 24, 50, 60Hz.

See also: Common Image Format, Progressive, ITU-R.BT 709, Table 3

13.5MHz

Sampling frequency used in the 601 digital coding of SD video. The frequency was chosen to be a whole multiple of the 525 and 625-line television system frequencies to create some compatibility between the digital systems. The sampling is fast enough to faithfully portray the highest frequency, 5.5MHz, luminance detail information present in SD images. Digital sampling of most HD standards samples luminance at 74.25MHz, which is 5.5 times 13.5MHz.

See also: 2.25MHz, ITU-R BT.601

2.25MHz

This is the lowest common multiple of the 525/59.94 and 625/50 television line frequencies, being 15.734265kHz and 15.625kHz respectively. Although seldom mentioned, its importance is great as it is the basis for all digital component sampling frequencies both at SD and HD.

See also: 13.5MHz

24P

Short for 24 frames, progressive scan. In most cases this refers to the HD picture format with 1080 lines and 1920 pixels per line (1080 x 1920/24P). The frame rate is also used for SD at 480 and 576 lines with 720 pixels per line. This is often as an offline for an HD 24P edit, or to create a pan-and-scan version of an HD down-conversion. Displays working at 24P usually use the double shuttering technique – like film projectors – to show each image twice and reduce flicker when viewing this low rate of images.

24PsF

24P Segmented Frame. This blurs some of the boundaries between film/video as video is captured in a film-like way, formatted for digital recording and can pass through existing HD video infrastructure. Like film, entire images are captured at one instant rather than by the usual line-by-line TV scans down the image that means the bottom can be scanned 1/24 of a second after the top. The images are then recorded to tape as two temporally coherent fields (segments), one with odd lines and the other with even lines, that are well suited to TV recorders.

The images are a pure electronic equivalent of a film shoot and telecine transfer – except the video recorder operates at film rate (24 fps), not at television rates. The footage has more of a filmic look but with the low frame rate, movement portrayal can be poor.

25PsF and 30PsF rates are also included in the ITU-R BT. 709-4 recommendation.

See also: ITU-R BT. 709

601

See ITU-R BT. 601

709

See ITU-R BT. 709

720P

Short for 720 lines, progressive scan. Defined in SMPTE 296M and a part of both ATSC and DVB television standards, the full format is 1280 pixels per line, 720 lines and 60 progressively scanned pictures per second. It is mainly the particular broadcasters who transmit 720P that use it. Its 60 progressive scanned pictures per second offers the benefits of progressive scan at a high enough picture refresh rate to portray action well. It has advantages for sporting events, smoother slow motion replays etc.

74.25MHz

The sampling frequency commonly used for luminance (Y) or RGB values of HD video. Being 33 x 2.25MHz, the frequency is a part of the hierarchical structure used for SD and HD. It is a part of SMPTE 274M and ITU-R BT.709.

See also: 2.25MHz

Active picture

The part of the picture that contains the image. With the analogue 625 and 525-line systems only 575 and 487 lines actually contain the picture. Similarly, the total time per line is 64 and 63.5µS but only around 52 and 53.3µS contain picture information. As the signal is continuous the extra time allows for picture scans to reset to the top of the frame and the beginning of the line.

Digitally sampled SD formats contain 576 lines and 720 pixels per line (625-line system), and 480 lines and 720 pixels per line (525-line system) but only 702 contain picture information. The 720 pixels are equivalent to 53.3µS.

The sampling process begins during line blanking of the analogue signal, just before the left edge of active picture, and ends after the active analogue picture returns to blanking level. Thus, the digitised image includes the left and right frame boundaries as part of the digital scan line. This allows a gentle roll-on and roll-off the between the blanking (black) and active picture.

HD systems are usually quoted just by their active line count, so a 1080-line system has 1080 lines of active video, each of 1920 samples. This may be mapped onto a larger frame, such as 1125 lines, to fit with analogue connections.

Aliasing

Artefacts created as a result of inadequate or poor video sampling or processing. Spatial aliasing results from the pixel-based nature of digital images and leads to the classic 'jagged edge' (a.k.a. 'jaggies') appearance of curved and diagonal detail and twinkling on detail. This results from sampling rates or processing accuracy too low for the detail. Temporal aliasing occurs where the speed of the action is too fast for the frame rate, the classic example being wagon wheels that appear to rotate the wrong way.

See also: Anti-aliasing

Anamorphic

This generally describes cases where vertical and horizontal magnification is not equal. The mechanical anamorphic process uses an additional lens to compress the image by some added amount, often on the horizontal axis. In this way a 1.85:1 or a 2.35:1 aspect ratio can be squeezed horizontally into a 1.33:1 (4:3) aspect film frame. When the anamorphic film is projected it passes through another anamorphic lens to stretch the image back to the wider aspect ratio. This is often used with SD widescreen images which keep to the normal 720 pixel count but stretch them over a 33-percent wider display. It can also apply to camera lenses used to shoot 16:9 widescreen where the CCD chips are 4:3 aspect ratio.

See also: Aspect ratio

Anti-aliasing

Attempts to reduce the visible effects of aliasing. This is particularly the case with spatial anti-aliasing that typically uses filtering processes to smooth the effects of aliasing which may be noticeable as jaggedness on diagonal lines, or 'twinkling' on areas of fine detail. A better solution is to improve the original sampling and processing and avoid aliasing in the first place.

See also: Aliasing

Aspect Ratio

For pictures, this refers to the ratio of picture width to height. HD pictures use a 16:9 aspect ratio, which also may be noted as 1.77:1. This is a third wider than the traditional 4:3 television aspect ratio (1.33:1) and is claimed to enhance the viewing experience as it retains more of our concentration by offering a wider field of view.

Pixel aspect ratio refers to the length versus height for a pixel in an image. HD always uses square pixels as do most computer applications. SD does not. The matter is further complicated by SD using 4:3 and 16:9 (widescreen) images which all use the same pixel and line counts. Care is needed to alter pixel aspect ratio when moving between systems using different pixel aspect ratios so that objects retain their correct shape.

With both 4:3 and 16:9 images and displays in use, some thought is needed to ensure a shoot will suit its target displays. All HD, and an increasing proportion of SD, shoots are 16:9 but many SD displays are 4:3. As most HD productions will also be viewed on SD, clearly keeping the main action in the middle '4:3' safe area would be a good idea – unless the display is letterboxed.

See also: ARC

Chrominance (or Chroma) sub-sampling

See 4:2:2 etc.

CIF

Common Image Format. An image format that is widely used and denoted 'Common Image Format' by the ITU. The idea is to promote the easy exchange of image information nationally and internationally.

See HD-CIF

Colour space

The space encompassed by a colour system. Examples are: RGB, YCrCb, HSL (hue, saturation and luminance) for video, CMYK for print and XYZ for film. Moving between media, platforms or applications can require a change of colour space. This involves complex image processing so care is needed to get the right result. Also, repeated changes of colour space can lead to colours drifting off.

It is important to note that when converting from YCrCb to RGB more bits are required in the RGB colour space to maintain the dynamic range. For example, if the YCrCb colour space video is 8 bits per component then the RGB colour space video will need to be 10 bits.

Component video

Most traditional digital television equipment handles video in the component form: as a combination of pure luminance Y, and the pure colour information carried in the two colour difference signals R-Y and B-Y (analogue) or Cr, Cb (digital). The components are derived from the RGB delivered by imaging devices, cameras, telecines, computers etc.

Part of the reasoning for using components is that it allows colour pictures to be compressed. The human eye can see much more detail in luminance than in the colour information (chrominance). The simple task of converting RGB to Y, (R-Y) and (B-Y) allows exclusive access to the chrominance only, so its bandwidth can be reduced with negligible impact on the viewed pictures. This is used in PAL and NTSC colour coding systems and has been carried through to component digital signals both at SD and HD.

For the professional digital video applications, the colour difference signals are usually sampled at half the frequency of the luminance - as in 4:2:2 sampling. There are also other types of component digital sampling such as 4:1:1 with less colour detail (used in DV), and 4:2:0 used in MPEG-2.

Co-sited sampling

Where samples of luminance and chrominance are all taken at the same instant. This is designed so that the relative timing (phase) of all signal components is symmetrical and not skewed by the sampling system. Sampling is usually co-sited but there is a case of 4:2:0 sampling being interstitial – with chrominance samples made between the luminance samples.

See also: 4:2:2 etc.

DTV

Digital Television. This is a general term that covers both SD and HD digital formats.

Gamut (colour)

The range of possible colours available in an imaging system. The red, blue and green phosphors on television screens and the RGB colour pick-up CCDs or CMOS chips in cameras, define the limits of the colours that can be displayed – the colour gamut. Between the camera and viewer's screen there are many processes, many using component 4:2:2 video. However, not all component value combinations relate to valid RGB colours (for example, combinations where Y is zero). Equipment that generates images directly in component colour space, such as some graphics machines, can produce colours within the component range but that are invalid in RGB, which can also exceed the limits allowed for PAL and NTSC.

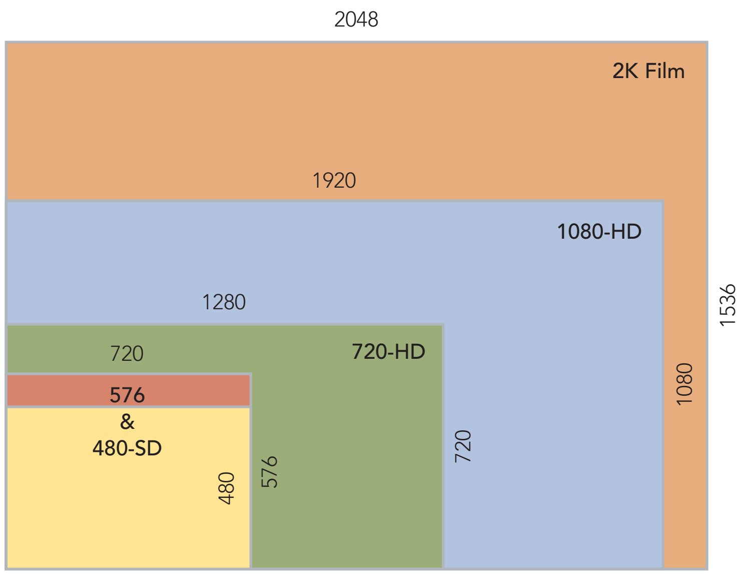
There is potential for overloading equipment – especially transmitters which may cut out to avoid damage! There is equipment that clearly shows many areas of out-of-gamut pictures, so that they can be adjusted before they cause problems.

HD

High Definition Television. This has been defined in the USA by the ATSC and others as having a resolution of approximately twice that of conventional television (meaning analogue NTSC – implying 486 visible lines) both horizontally and vertically, a picture aspect ratio of 16:9 and a frame rate of 24fps and higher. This is not quite straightforward as the 720-line x 1280 pixels per line, progressive scan format is well accepted as HD. This is partly explained by the better vertical resolution of its progressive scanning. Apart from the video format, another HD variation on SD is a slightly different colorimetry where, for once the world agrees on a common standard.

As HD's 1080 x 1920 image size is close to the 2K used for film, there is a crossover between film and television. This is even more the case if using a 16:9 window of 2K as here there is very little difference in size. It is generally agreed that any format containing at least twice the standard definition format on both H and V axes is high definition.

After some initial debate about the formats available to prospective HD producers and television stations, the acceptance of 1080-HD video at various frame rates, as a common image format by the ITU, has made matters far more straightforward. While television stations may have some latitude in their choice of format, translating, if required, from the common image formats should be routine and give high quality results.



2K, HD and SD images sizes

See also: Common Image Format, Interlace Factor

PAL and NTSC

PAL and NTSC do not exist in HD. They do not exist in modern SD digital television either – although it was digitised in early digital VTR formats. PAL, means Phase Alternating Line and is an analogue system for coding colour that is still widely in use. Similarly NTSC (National Television Standards Committee) describes an analogue system. Confusingly PAL and NTSC are still used to describe frame rates and formats that relate in some way with their analogue world. So 1080 PAL might be 1080/50I.

Quantization

Quantization refers to sampling: the number of bits used in making digital samples of a signal. For video, 8 bits is quite common in consumer and prosumer products such as DV. HDV also uses 8 bits. Note that the 8 bits can define 2^8 or 256 numbers or levels that, for converting analogue video into digits, are assigned to levels of image brightness.

For more accuracy and to withstand multiple levels of complex post production processing, studio video applications often use 10-bit sampling – providing 1024 levels.

Usually the distribution of the levels between brightest and darkest is linear (even) but in the case of scanning film negative for input to a digital intermediate chain, then a logarithmic distribution is often used that progressively squashes the levels into the darker areas of picture. This is because film negative has to carry a very wide range of contrast information from the original scene, and the levels in the dark/shadow areas are more significant and visible than those in bright areas. The 'log' sampling suitably redistributes the available digital levels – hence 10-bit log. This is considered to be as useful as 13-bit linear quantization.

NB: Quantization has another meaning.

See section: Video Compression 1

RGB

Red, Green and Blue. Cameras, telecines and most computer equipment originate images in this colour space. For digital sampling, all three colours are sampled in the same way at full bandwidth – hence 4:4:4. images may offer better source material for the most critical chroma keying, but they occupy 50 percent more data space than 4:2:2 and as no VTRs record 4:4:4, data recorders or disks must be used to store them. Also, there are no television means to connect them, so IT-based networking technology is used.

Often 4:4:4 is only used in post production areas and is converted to 4:2:2 when material is more widely distributed.

See also: 4:4:4, Gamut

Segmented Frame

See: 24PsF

Square pixels

Square pixels are the pixel aspect ratio where the pixels describe a square area of the displayed image. This is the case with HD broadcast standards, as the picture formats describe line length (number of pixels per line) and number of lines, in exact 16:9 ratios – which is also the display aspect ratio of the pictures.

There are places in HD where pixels are not square. The very widely used HDCAM sub-samples the 1920-pixel HD line lengths with 1440 luminance samples. This is only an internal function of the recorder; the inputs and outputs use square pixels. In a similar way the 1080I HDV(2) format also uses 1440 samples per line.

Generally, computers generate images with square pixels but digital SD television images are not square. This means that any applications or equipment used needs to take this into account when transferring between applications, or performing image manipulations to maintain correct image aspect ratios (so circles remain circular).

See also: Anamorphic, Aspect ratio

Sub-sampling

In a digital sampling system, taking fewer samples of an analogue signal than the number of pixels in the digital image is called sub-sampling. Generally sub-sampling is used to reduce the amount of data used for an image. In the widely used 4:2:2 sampling system for studio quality video, each luminance sample corresponds to one pixel – denoted by the '4'. The two chrominance signals are each sampled at half the rate, making one per two pixels. This is known as chrominance sub-sampling – a term that is sometimes more generally ascribed to the sampling ratios – such as 4:2:2, 4:1:1, etc.

See also: 4:2:2 etc

System nomenclature

A term used to describe television standards. The standards are mostly written in a self-explanatory form but there is room for confusion concerning vertical scanning rates. For example, 1080/60I implies there are 60 interlaced fields per second that make up 30 frames. Then 1080/30P describes 30 frames per second, progressively scanned.

The general rule appears to be that the final figure always indicates the number of vertical refreshes per second. However, Table 3 (below) uses a different method. It defines frame rates (numbers of complete frames) and then defines whether they are interlaced or progressive. So here the 'frame rate code 5' is 30Hz which produces 30 vertical refreshes when progressive, and 60 when interlaced. Be careful!

See also: Interlace, Progressive

Table 3

The video formats allowed for broadcast in the ATSC DTV standard are listed in Table 3 of document Doc. A/53A.

Table 3 Compression Format Constraints

Vertical_ size_ value	Horizontal_ size_ value	aspect_ ratio_ information	frame_ rate_ code	progressive_ sequence
1080	1920	1,3	1,2,4,5	1
			4,5	0
720	1280	1,3	1,2,4,5,7,8	1
480	704	2,3	1,2,4,5,7,8	1
			4,5	0
	640	1,2	1,2,4,5,7,8	1
			4,5	0

Legend for MPEG-2 coded values in Table 3
aspect_ratio_information 1 = square samples 2 = 4:3 display aspect ratio 3 = 16:9 display aspect ratio
Frame_rate_code 1 = 23.976 Hz 2 = 24 Hz 4 = 29.97 Hz 5 = 30 Hz 7 = 59.94 Hz 8 = 60 Hz
Progressive_sequence 0 = interlaced scan 1 = progressive scan

This table lists no fewer than 18 DTV formats for SD and HD. Initially, this led to some confusion about which should be adopted for whatever circumstances. Now most HD production and operation is centred on the 1080-line formats either with 24P, 25P or 60I vertical scanning, and 720-line formats at 50P and 60P.

Truncation (a.k.a. Rounding)

Reducing the number of bits used to describe a value. This is everyday practice; we may say 1,000 instead of 1024 in the same way we leave off the cents/pence when talking about money. There is also the need to truncate the digits used in digital video systems. With due care, this can be invisible, without it degradation becomes visible.

Decimal:	186	x 203	= 37758
Binary:	10111010	x 11001011	= 1001001101111110

It is the nature of binary mathematics that multiplication, which is commonplace in video processing (e.g. mixing pictures), produces words of a length equal to the sum of the two numbers. For instance, multiplying two 8-bit video values produces a 16-bit result – which will grow again if another process is applied. Although hiways within equipment may carry this, ultimately the result will have to be truncated to fit the outside world which, for HD, may be a 10-bit HD-SDI interface or 8-bit MPEG-2 encoder.

In the example, truncating by dropping the lower eight bits lowers its value by 01111110, or 126. Depending on video content, and any onward processing where the error is compounded, this may, or may not be visible. Typically, flat (no detail) areas of low brightness are prone to showing this type of discrepancy as banding. This is, for example, sometimes visible from computer generated images.

Inside equipment, it is a matter of design quality to truncate numbers in an intelligent way that will not produce visible errors – even after further processing. Outside, plugging 10-bit equipment into 8-bit needs care. Intelligent truncation is referred to as Rounding.

Universal Format

1080/24P is sometimes referred to as the Universal Format for television. The reason is its suitability for translation into all other formats to produce high quality results in all cases.

See also: HD-CIF, Universal Master

Universal Master

The 1080/24P format has well defined and efficient paths to all major television formats and is capable of delivering high quality results to all. An edited master tape in this format is sometimes referred to as a Universal Master.

See also: HD-CIF

Y, Cr, Cb

This signifies video components in the digital form. Y, Cr, Cb is the digitised form of Y, R-Y, B-Y.

Y, R-Y, B-Y

See component video

YUV

De-facto shorthand for any standard involving component video. This has been frequently, and incorrectly, used as shorthand for SD analogue component video – Y, R-Y, B-Y. Y is correct, but U and V are axes of the PAL colour subcarrier which are modulated by scaled and filtered versions of B-Y and R-Y respectively. Strangely, the term is still used to describe component analogue HD. This is double folly. Although Y is still correct, all HD coding is digital and has nothing to do with subcarriers or their axes. So forget it!

2

Chapter 2

Video Compression: Concepts

- Blocks
- Codec
- Compression ratio
- Compression-friendly
- DCT
- GOP
- I-Frame
- Inter-frame compression
- Interlace
- Interlace Factor

Avid

Video compression reduces the amount of data or bandwidth used to describe moving pictures. Digital video needs vast amounts of data to describe it and there have long been various methods used to reduce this for SD. And as HD has up to a six times bigger requirement of 1.2Gb/s and requiring 560GB per hour of storage, the need for compression is even more pressing.

Intro Compression – General

Exactly which type and how much compression is used depends on the application. Consumer delivery (DVD, transmission, etc) typically uses very high compression (low data rates) as the bandwidth of the channels is quite small. For production and online editing use much lighter compression (higher data rates) are used as good picture quality needs to be maintained though all the stages leading to the final edited master.

Video compression methods are all based on the principle of removing information that we are least likely to miss – so-called 'redundant' picture detail. This applied to still images as well as video and cinema footage. This takes the form of several techniques that may be used together. Digital technology has allowed the use of very complex methods which have been built into low cost mass produced chips.

First, our perception of colour (chroma) is not as sharp as it is for black and white (luminance), so the colour resolution is reduced to half that of luminance (as in 4:2:2). This is used in colour television (NTSC, PAL and digital). Similarly, fine detail with little contrast is less noticeable than bigger objects with higher contrast. To access these a process called DCT resolves 8 x 8 pixel blocks of digital images into frequencies and amplitudes to make it possible to scale (down), or 'quantize', the DCT coefficients (frequencies and amplitudes) and so reduce the data. This applies most of the digital video compression schemes in use today including AVR, DV, HDV, JPEG (but not JPEG2000) and the *I frames* of MPEG-1, 2 and 4, and Windows Media 9. A further reduction is made using Huffman coding, a purely mathematical process that reduces repeated data.

MPEG-2 and the more recent MPEG-4 add another layer of compression by analysing what changes form frame to frame by analysing the movement of 16 x 16-pixel *macro blocks* of the pictures. Then it can send just the movement information, called motion vectors, that make up predictive (B and P) frames and contain much less data than I frames, for much of the time. Whole pictures (I frames, more data) are sent only a few times a second. MPEG-2 compression is used in all forms of digital transmission and DVDs as well as for HDV. The more refined and efficient MPEG-4 is now being introduced for some HD services, and is set to become widely used for new television services.

Each of these techniques does a useful job but needs to be applied with some care when used in the production chain. Multiple compression (compress/de-compress) cycles may occur while moving along the chain, causing a build-up of compression errors. Also, as many compression schemes are designed around what looks good to us, they may not be so good in production, post production and editing. This particularly applies in processes, such as keying and colour correction, that depend on greater image fidelity than we can see, so disappointing results may ensue from otherwise good-looking compressed originals.

See also: AVR, Component video, DV, DNxHD, Huffman coding, JPEG, JPEG2000, MPEG-2, MPEG-4

Blocks

See DCT

Codec

Codec is short for coder/decoder – usually referring to a compression engine. Confusingly, the term is often misused to describe just a coder or decoder.

Compression ratio

This is the ratio of the uncompressed (video or audio) data to the compressed data. It does not define the resulting picture or sound quality, as the effectiveness of the compression system needs to be taken into account. Even so, if used in studio applications, compression is usually between 2:1 and 7:1 for SD (and D1 and D5 uncompressed VTRs are also available), whereas compression for HD is currently approximately between 6:1 and 14:1 – as defined by VTR formats, and is I-frame only. For transmission, the actual values depend on the broadcaster's use of the available bandwidth but around 40:1 is common for SD and somewhat higher, 50 or 60:1 for HD (also depending on format). These use both I-frames and the predictive frames to give the greater compression.

HDV records data to tape at 19-25 Mb/s – a rate comparable with HD transmission and a compression ratio of around 40:1, depending on the standard used.

Transmission and video recorders in general work at a constant bit rate so, as the original pictures may include varying amounts of detail, the quality of the compressed images varies. DVDs usually work on a constant quality/variable bit rate principle. So the compression ratio slides up and down according to the demands of the material, to give consistent results. This is part of the reason why DVDs can look so good while only averaging quite low bit rates – around 4 Mb/s.

Compression-friendly

Material that looks good after compression is sometimes referred to as 'compression friendly'. This can become important in transmission where very limited data bandwidth is available and high compression ratios have to be used. Footage with large areas of flat colour, little detail and little movement compress very well: for example, cartoons, head-and-shoulder close-ups and some dramas. As, MPEG-2 compression looks at spatial detail as well as movement in pictures and an excess of both may show at the output as poor picture quality. This often applies to fast-moving sports – for instance football.

Poor technical quality can be compression unfriendly. Random noise will be interpreted as movement by an MPEG-2 or MPEG-4 encoder, so it wastes valuable data space conveying unwanted movement information. Movement portrayal can also be upset by poor quality frame-rate conversions that produce judder on movement, again increasing unwanted movement data to be transmitted at the expense of spatial detail. Such circumstances also increase the chance of movement going wrong – producing 'blocking' in the pictures.

Errors can be avoided by the use of good quality equipment throughout the production chain. Also, the choice of video format can help. For example, there is less movement in using 25 progressively scanned images than in 50 interlaced fields, so the former compress more easily. The efficiency increase is typically 15-20 percent.

DCT

Discrete Cosine Transform is used as a first stage of many digital video compression schemes including JPEG and MPEG-2 and –4. It converts 8 x 8 pixel blocks of pictures to express them as frequencies and amplitudes. This may not reduce the data but it does arrange the image information so that it can. As the high frequency, low amplitude detail is least noticeable their coefficients are progressively reduced, some often to zero, to fit the required file size per picture (constant bit rate) or to achieve a specified quality level. It is this reduction process, known as quantization, which actually reduces the data.

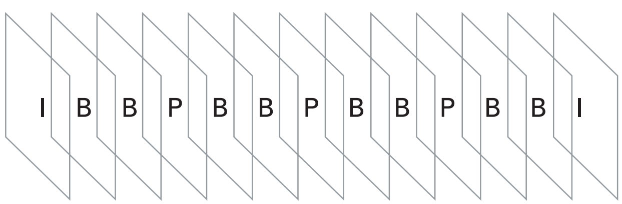
For VTR applications the file size is fixed and the compression scheme's efficiency is shown in its ability to use all the file space without overflowing it. This is one reason why a quoted compression ratio is not a complete measure of picture quality.

DCT takes place within a single picture and so is intra-frame (I-frame) compression. It is a part of the currently most widely used compression in television.

See also: AVR, Compression ratio, DV, JPEG, MPEG-2, MPEG-4

GOP

Group Of Pictures – as in MPEG-2 and MPEG-4 video compression. This is the number of frames to each integral I-frame: the frames between being predictive (types B and P). 'Long GOP' usually refers to MPEG-2 and 4 coding. For transmission the GOP is often as long as half a second, 13, or 15 frames (25 or 30fps), which helps to achieve the required very high compression ratios.



A typical group of pictures

Cutting long GOP MPEG is not straightforward as its accuracy is limited to the GOP length unless further processing is applied – typically decoding. HDV uses long GOP MPEG-2 of 6 or 15 frames for HDV1 or HDV2 respectively making it editable at 1/4 or 1/2 second intervals. A GOP of 1 indicates 'I-frame only' video, which can be cut at every frame without need of processing.

Studio applications of MPEG-2 have very short GOPs, Betacam SX has a GOP of 2, IMX has 1, (i.e. I-frame only – no predictive frames) which means cutting at any frame is straightforward. Other formats such as DV, DVCPRO HD and HDCAM, D5-HD do not use MPEG but are also I-frame only.

See also: MPEG-2, MPEG-4

I-frame only (aka I-frame)

Short for *intra-frame* only.

Inter-frame compression

Video compression that uses information from several successive video frames to make up the data for its compressed 'predictive' frames. The most common example is MPEG-2 with a GOP greater than 1. Such an MPEG-2 stream contains a mix of both I-frames and predictive B and P (Bi-directional predictive and Predictive) frames. Predictive frames cannot be decoded in isolation from those in the rest of the GOP so the whole GOP must be decoded. This is an efficient coding system that is good for transmission but it does not offer the flexibility needed for accurate editing as it can only be cut at the GOP boundaries. It also requires estimation of the movement from picture to picture, which is complex and not always accurate – leading to 'blockiness'.

See also: GOP, MPEG-2, MPEG-4

Interlace

A method of ordering the lines of scanned images as two (or more) interlaced fields per frame. Most television uses 2:1 interlacing; alternate fields of odd lines 1,3,5, etc., followed by a field of even lines 2, 4, 6, etc. This doubles the vertical refresh rate as there are twice as many interlaced fields as there would be whole frames. The result is better portrayal of movement and reduction of flicker without increasing the number of full frames or required signal bandwidth. There is an impact on vertical resolution and care is needed in image processing.

See also: Interlace factor, Progressive

Interlace Factor

Use of interlaced, rather than progressive, scans has no effect on the vertical resolution of still images. However, if anything in the image moves the resolution is reduced by the Interlace Factor, which may be 0.7 or less. This is due to the time displacement between the two fields of interlace which will produce detail that is jagged, line-by-line, during the movement and it appears as an overall slight softening of vertical resolution.

Intra-frame compression (a.k.a. I-frame compression)

Video compression which takes information from one video frame only. This way, all the information to re-create the frame is contained within its own compressed data and is not dependent on other adjacent frames. This means that I-frame compressed video is easily edited as it can simply be cut at any picture boundary without the need for any decoding and recoding. I-frame only video can be edited and the result output as first generation material. Any other operations such as wipes, dissolves, mixes, DVE moves etc., can only be performed on the baseband signal, requiring that the video is first decompressed.

See also: AVR, DV, JPEG, MPEG-2,

Macroblock

A 16 x 16 pixel block, comprising four adjacent DCT blocks – macroblocks are used to generate motion vectors in MPEG-2 coding. Most coders use a 'block matching' technique to establish where the block has moved and so generate motion vectors to describe the movement. This works most of the time but also has its well-known moments of failure. For example, slow fades to black tend to defeat the technique, making the resulting misplaced blocks quite visible. Better technologies are available for use in movement estimation, such as phase correlation.

Motion Vectors

Used in MPEG-2 and MPEG-4 compression systems, motion vectors describe the direction and distance that macroblocks (16 x 16 pixels) move between frames. Sending this movement information requires much less data than sending an I frame, and so further reduces the video data.

Progressive (scan)

Sequence for scanning an image where the vertical scan progresses from line 1 to the end in one sweep. In HDTV there are a number of progressive vertical frame (refresh) rates allowed and used. 24Hz is popular for its compatibility with motion pictures and its ability to be easily translated into all of the world's television formats. 25 and 30Hz correspond with existing SD frame rates (although they use interlaced scans). 50 and 60Hz are also allowed for, but, due to bandwidth restrictions, these are limited in picture size, e.g. 720/60P and 720/50P.

Today, progressive scanning is most commonly found in computer displays and all the modern panel TV displays are progressive. Progressive images are rock steady making the detail easy to see. For the equipment designer progressive images are easier process as there is no difference between the two fields of a frame to contend with.

Progressive scans have the disadvantage of a slow vertical refresh rate. Thus, for the lower rates of 24, 25 and 30Hz, which can be used in HD television with the larger 1080-line formats, there would be considerable flicker on displays, unless there were some processing to show each picture twice (as in double shuttering in cinema projectors). Besides flicker, the other potential problem area is that of fast action or pans, as the lower refresh rate means that movement will tend to stutter. Interlace, with its two vertical refreshes per frame, has advantages here.

See also: 24PsF, Interlace

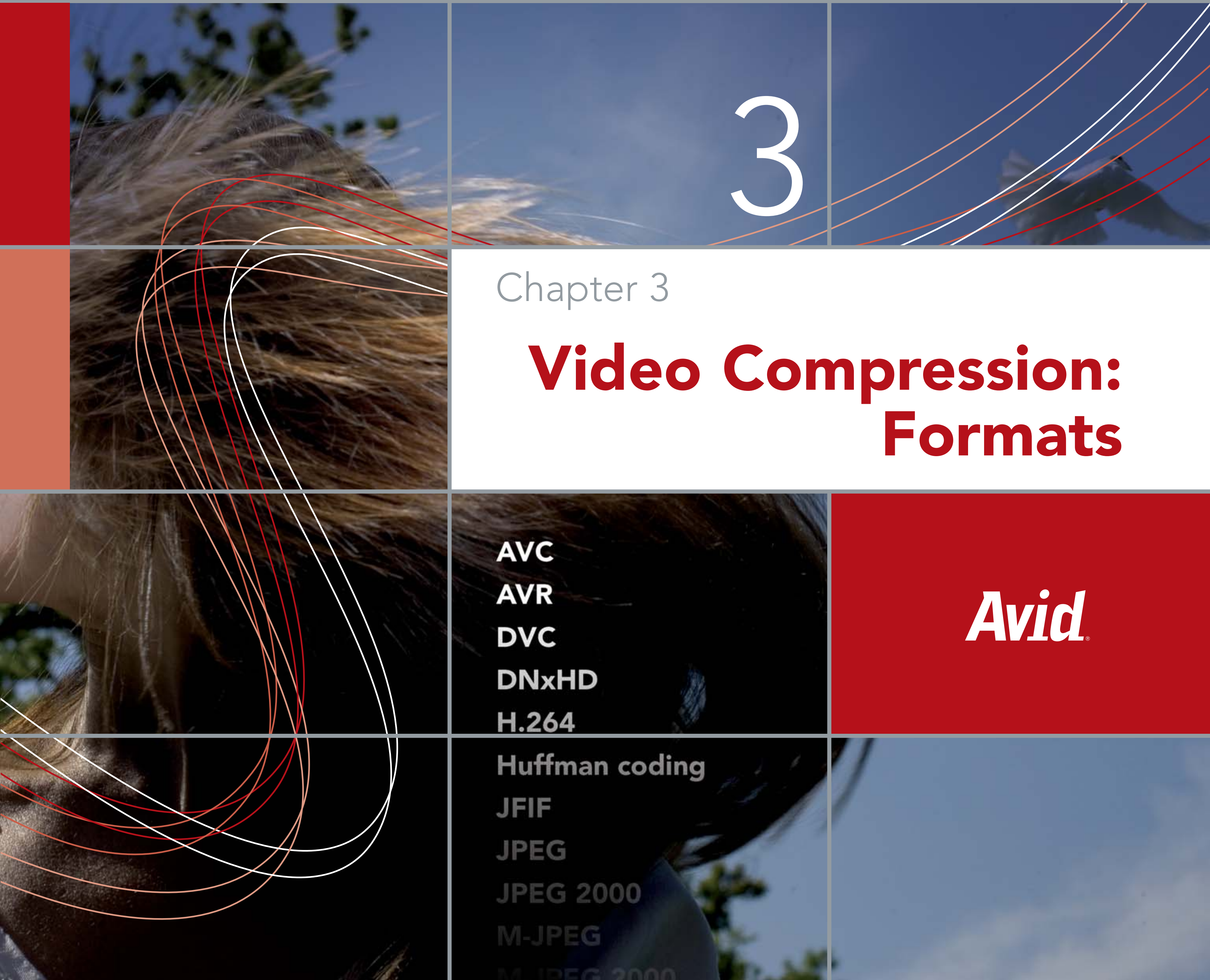
Quantization

Quantizing is the process used in DCT-based compression schemes, including AVC, JPEG, MPEG-2 and MPEG-4, to reduce the video data in an I frame. DCT allows quantizing to selectively reduce the DCT coefficients that represent the highest frequencies and lowest amplitudes that make up the least noticeable elements of the image. As many are reduced to zero significant data reduction is realised.

Using a fixed quantizing level will produce a constant quality of output with a data rate that varies according to the amount of detail in the images. Alternatively quantizing can be varied to produce a constant data rate, but variable quality, images. This is useful where the data must be fitted into a defined size of store or data channel – such as a VTR or a transmission channel. The success in nearly filling, but never overflowing, the storage is one measure of the efficiency of DCT compression schemes.

NB: Quantization has a second meaning.

See Video Formats section



Chapter 3

Video Compression: Formats

AVC
AVR
DVC
DNxHD
H.264
Huffman coding
JFIF
JPEG
JPEG 2000
M-JPEG

Avid

This is the practical side of compression showing the systems and formats that are used. Some are proprietary, in which case the company involved is mentioned.

AVC

See MPEG-4

AVR

AVR is a range of Motion-JPEG video compression schemes devised by Avid Technology for use in its ABVB hardware-based non-linear systems. An AVR is referred to as a constant quality M-JPEG resolution since the same quantization table (of coefficients) is applied to each frame of a video clip during digitization. For any given AVR, the actual compressed data rate will increase as the complexity of the imagery increases. For example, a head shot typically results in a low data rate while a crowd shot from a sporting event will yield a high data rate. To avoid system bandwidth problems, AVRs utilize a mode of rate control called rollback which prevents the compressed data rate from increasing beyond a preset limit for a sustained period. So, when the data rate exceeds the rollback limit on a given frame, high spatial frequency information is simply discarded from subsequent frames until the rate returns to a tolerable level.

See also: DCT, JPEG

DVC

DVC is the compression used in DV equipment that is standardised in IEC 61834. It is a DCT-based, intra-frame scheme achieving 5:1 compression so that 8-bit video sampling of 720 x 480 at 4:1:1 (NTSC) or 720 x 576 4:2:0 (PAL) produces a 25 Mb/s video data rate. The same is used for DV, DVCAM, Digital8 and DVCPRO (where PAL is PAL 4:1:1). It achieves good compression efficiency by applying several quantizers at the same time, selecting the nearest result below 25Mb/s for recording to tape.

DNxHD

Avid DNxHD encoding is designed to offer quality at significantly reduced data rate and file size and it is supported by the family of Avid editing systems. Engineered for editing, it allows any HD material to be handled on SD-original Avid systems. Any HD format can be encoded edited, effects added, colour corrected and the project finished.

There is a choice of compression image formats to suit requirements. Some of the formats are:

Format	DNxHD 220x	DNxHD 185x	DNxHD 185	DNxHD 145	DNxHD 120
Bit depth	10 bit	10 bit	8 bit	8 bit	8 bit
Frame rate	29.92 fps	25 fps	25 fps	25 fps	25 fps
Data rate	220 Mb/s	184 Mb/s	184 Mb/s	135 Mb/s	220 Mb/s

Avid DNxHD maintains the full raster, is sampled at 4:2:2 and uses highly optimised coding and decoding techniques, so image quality is maintained over multiple generations and processes. When you’re ready, master to any format you need.

DNxHD efficiency enables collaborative HD workflow using networks and storage designed to handle SD media. So, for example, Avid Unity shared media networks are HD-ready today! Cost-effective, real-time HD workflows can be designed with Media Composer Adrenaline HD and Avid DS Nitris systems. You can even edit HD on a laptop.

For more information see www.avid.com/dnxhd/index.asp

H.264

See MPEG-4

Huffman coding

A method of compressing data by recognizing repeated patterns and assigning short codes to those that occur frequently, and longer codes to those that are less frequent. The codes are assigned according to a Huffman Table. Sending the codes rather than all the original data can achieve as much as a 2:1 lossless compression and the method is often used as a part of video compression schemes such as JPEG and MPEG.

JFIF

JPEG File Interchange Format – a compression scheme used by Avid Technology in its Meridien hardware-based non-linear systems. A JFIF M-JPEG resolution is termed constant rate since compressing clips of varying complexity results in a fixed data rate. Each JFIF resolution is defined by a target data rate and a base quantization table. When digitizing, the quantization table is linearly scaled (known as rolling Q) to conform the actual compressed data rate to the target rate. Due to the flexibility of this approach, imagery compressed by a JFIF resolution generally looks better than that compressed by an AVR of comparable average data rate.

JPEG

Joint (ISO and ITU-T) Photographic Experts Group; JPEG is a standard for compressing still picture data. It offers *compression ratios* of between two and 100 times and there are three levels of processing available: baseline, extended and lossless encoding.

JPEG baseline coding, which is the most common for television and computer applications, starts by applying *DCT* to 8 x 8 pixel blocks of the picture, transforming them into frequency and amplitude data. This itself may not reduce data but then the generally less visible high frequencies can be divided by a high *quantizing* factor (reducing many to zero), and the more visible low frequencies by a lower factor. The quantizing factor can be set according to data size (for constant bit rate) or picture quality (constant quality) requirements – effectively adjusting the compression ratio. The final stage is *Huffman coding* which is a lossless mathematical treatment that can further reduce data by 2:1 or more.

Baseline JPEG coding creates .jpg files and is very similar to the I-frames of MPEG-1, -2 and -4, the main difference being they use slightly different Huffman tables.

See also: Compression, Compression ratio, DCT, DV, Huffman coding, JFIF, M-JPEG,

WWW http://www.jpeg.org

JPEG 2000

JPEG 2000 is an advanced image coding (compression) system from the Joint Photographic Experts Group. Like 'normal' JPEG, this is intra-frame compression and it is suitable for a wide range of uses from portable digital cameras, to scientific and industrial applications.

Rather than using the established DCT, it employs state-of-the-art techniques based on wavelet technology. Requiring more processing than MPEG, JPEG 2000 has, until recently been too costly for wide use in television applications. Now new chips have lowered the price barriers and JPEG 2000's use in TV and D-cinema is expected to rapidly expand as it has distinct advantages for the high quality large images. It is already recommended for D-cinema and Grass Valley have adopted it for HD compression in their new Infinity range of cameras.

As it does not analyse images block-by-block but in a circular area-by-area pattern, there are no 'blocky' artefacts, instead problem areas tend to become a little softer – which is much less noticeable. JPEG 2000 continues to improve as more bits are used for the images. As a result, at high bit rates of 200-300Mb/s HD and D-cinema images are displayed with 'visually lossess' quality. It is also scalable, so image sizes different to the encoded size, can be extracted directly without decoding.

WWW http://www.jpeg.org

M-JPEG

Motion JPEG refers to JPEG compression applied to moving pictures. As the detail contained within each frame varies, so some decision is required as to whether to use a constant bitrate scheme or constant quality.

See also: AVR, JPEG

M-JPEG 2000

JPEG 2000 used for moving pictures.

MPEG

Moving Pictures Expert Group. A group of industry experts involved with setting standards for moving pictures and sound. These are not only those for the compression of video and audio (such as MPEG-2 and MP3) but also include standards for indexing, filing and labelling material.

WWW http://www.mpeg.org

MPEG-2

ISO/IEC 13818-1. This is a video compression system primarily designed for use in the transmission of digital video and audio to viewers by use of very high compression ratios. Its importance is huge as it is currently used for nearly all DTV transmissions worldwide, SD and HD, as well as for DVDs and many other applications where high video compression ratios are needed.

The Profiles and Levels table (below) shows that it is not a single standard but a whole family which uses similar tools in different combinations for various applications. Although all profile and level combinations use MPEG-2, moving from one part of the table to another may be impossible without decoding to baseband video and recoding.

Profile / Level	Simple 4:2:0 I, B	Main 4:2:0 I, B, P	422P 4:2:2 I, B, P	SNR* 4:2:0 I, B, P	Spatial* 4:2:0 I, B, P	High 4:2:0,4:2:2 I, B, P
High		1920x1152 80 Mb/s				1920x1152 100 Mb/s
High-1440		1440x1152 60 Mb/s			1440x1152 60 Mb/s	1440x1152 80 Mb/s
Main	720x570 15 Mb/s	720x576 15 Mb/s	720x608 50 Mb/s	720x576 15 Mb/s		720x576 20 Mb/s
Low		352x288 4 Mb/s		352x288 4 Mb/s		

MPEG-2 profiles and levels
*SNR and Spatial are both scalable

Profiles outline the set of compression tools used. Levels describe the picture format/quality from High Definition to VHS. There is a bit rate defined for each allocated profile /level combination. In all cases, the levels and bit rates quoted are maximums so lower values may be used. Combinations applicable to modern HD are highlighted.

MPEG-2 is deliberately highly asymmetrical in that decoding is far simpler than the encoding – so millions of viewers enjoy reasonable prices while a few broadcasters incur the higher unit costs. Coding has two parts. The first uses DCT-based intra-frame (I-frame) compression and application of quantizing, to reduce the data – almost identically to JPEG. The second involves inter-frame compression – calculating the movement of macroblocks and then substituting just that information for the pictures between successive I-frames – making a GOP. The movement is conveyed as motion vectors, showing direction and distance, which amounts to far less data than is needed for I-frames. Motion vector calculation is not an exact science so there can be huge difference in quality between different MPEG compressors. Decompression is deterministic so all decompressors (decoders) should be the same.

The encoding process necessarily needs to look at several frames at once and so introduces a considerable delay. Similarly, the decoder delays pictures. For transmissions this can add up to over a second. MPEG-2 is sometimes used on broadcast contribution circuits, this becomes noticeable when news reporters appear to delay answering a question.

To fit HD video and audio down a transmission 'data pipe' requires very high compression. Uncompressed 10-bit HD requires up to 1244Mb/s. But this is 10-bit data and sampled at 4:2:2. MPEG-2 is 8-bit sampled at 4:2:0 – bringing the data down to 746Mb/s. However, the data pipes for ATSC (19.2Mb/s) or DVB (20Mb/s, depending on channel width, parameters etc.) imply the need for around 40:1 compression.

See also: DCT, GOP, Intra-frame compression, Inter-frame compression. Macroblock

MPEG-4

MPEG-4 (ISO/IEC 14496) was developed by MPEG (Moving Picture Experts Group) and is a wide standard covering many subjects but its importance in television production is mostly related to its video compression scheme. MPEG-4 Part 10, AVC (Advanced Video Coding) and H.264 all refer to the same compression system. This is another DCT-based system that builds on MPEG-2 to produce a more efficient codec again using intra and inter-frame techniques. Coding is more complex than MPEG-2 but it can produce extra data savings of around 30 percent – or more. Some of the latest television services are planned to use MPEG-4. This is especially true with HD where more bandwidth is required. It will enable the delivery of better image quality to viewers, or more channels to be delivered within a given bandwidth. It is said to be similar to, but not the same as, WM 9.

WWW http://www.chiariglione.org/mpeg

VC-1

VC-1 is a video compression codec specification that is currently being standardised by SMPTE (SMPTE 421M) and implemented by Microsoft as Windows Media Video (WMV) 9 Advanced Profile.

See: WMV 9

WMV 9

Windows Media Video 9 is a video and audio compression system (codec) developed by Microsoft. It is said to be similar to MPEG-4 AVC and to have as good or slightly better performance giving lower data rates and claims to be a less complex process. It's applications are seen as being for content delivery such as HD DVD.