MPEG-2 Fundamentals
for Broadcast and Post-Production Engineers

A Video and Networking Division White Paper

By Adolfo Rodriguez, David K. Fibush and Steven C. Bilow

July, 1996

INTRODUCTION

This paper will discuss the digital representation of video as defined in the MPEG-2 standard and will examine some crucial, strategic issues regarding the newly adopted 4:2:2 profile at main level. Specifically, we will cover the rationale for this new profile and will illustrate the benefits it provides to the broadcast and post-production communities.

The intent of this paper is to articulate the appropriate applications of the various MPEG profiles and levels, and to demonstrate the fundamental necessity for the new 4:2:2 profile. We will also address the encoder/decoder compatibility required by the standard, the pervasive need for a 4:2:2 video representation, and the method by which the 4:2:2 profile at main level allows MPEG-2 to provide these advantages in a cost-effective manner.

The new MPEG 4:2:2 profile at main level has recently become an official part of the MPEG standard. To broadcasters and production houses that do not accept the low chroma bandwidth options provided by previous MPEG main level profiles, the 4:2:2 profile at main level will prove to be the long-awaited solution for contribution quality MPEG compressed video.

The Tektronix perspective is the result of a 50-year history of instrumentation, broadcast, and video equipment design. It is also the product of our success in helping the MPEG-2 standards body toward the provision of this new, cost-effective support for broadcast-quality video. In the pages that follow, we will share that perspective and experience; in doing so, we hope to contribute to your success in adopting this exciting new technology in future aquisition, storage, distribution, and production systems

VIDEO COMPRESSION FUNDAMENTALS

Background

The issue of video compression has become an extremely hot topic in the past few years. Interestingly, though, the compression of video signals is not at all new. Many people do not realize that the compression of video was first commercialized in the early 1950's with the introduction of color television. Three full resolution images representing red, blue, and green, with a total bandwidth of 15 MHz, were compressed into a single composite signal with a bandwidth of 5 MHz -- a 3:1 reduction in ban dwidth, or, simply put, compression! (Figure 1.)

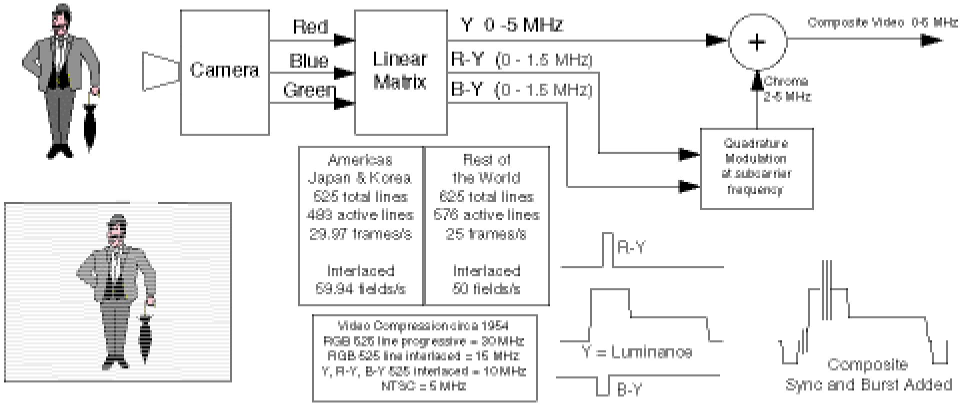
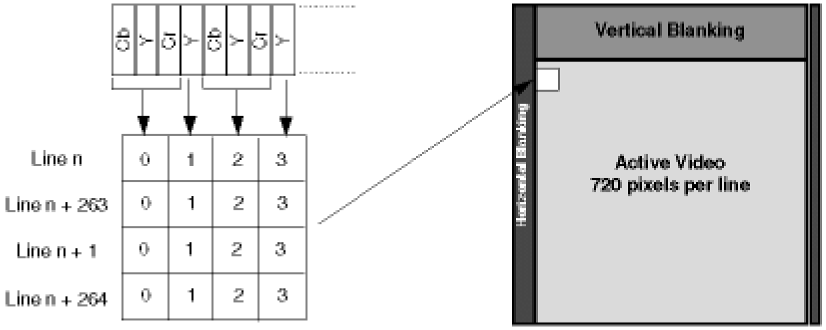


***Figure 1**. Video compression in the 1950's.*

Of course, this early method of compression was accomplished using analog techniques. Today, more consistent quality and much higher compression ratios can be achieved by converting analog video into the digital domain, where some highly sophisticated techniques may be applied. These techniques result in far more efficient compression, as well as more refined methods of handling the compressed images. These digital techniques require tremendous computational power which, until recently, was not economically viable. Today, however, that has changed.

COMPRESSION IS BASICALLY A PROCESS BY WHICH THE INFORMATION CONTENT OF AN IMAGE OR GROUP OF IMAGES IS REDUCED BY EXPLOITING THE REDUNDANCY PRESENT IN VIDEO SIGNALS.

Generally, digital video compression begins with a component video representation where the signal is represented in the form of one luminance and two chrominance components. The most widely used digital component video format is Recommendation 601 which uses a 4:2:2 sampling structure with co-sited samples (Figure 2). When we say "co-sited," we mean that each color pixel is described by 3 samples: one luminance sample and two chrominance samples formed from color differences. Because these 3 samples are coincident in time, they are refered to as "co-sited." In 525-line systems, there are 720 pixels in each active line and 483 active lines in each frame. For 625-line systems there are 576 active lines.



4:2:2 Sampling Recommendation ITU-R BT.601 (Recommendation 601)

"4" = 13.5 MHz "2" = 6.75 MHz

Figure 2. *Component Video*

By associating chrominance samples with every other luminance sample, it is possible to reduce the required bandwidth without impacting the visual quality of the image. The "4" in 4:2:2 comes from the 4 times subcarrier sampling frequency used in composite digital systems. Interestingly, the sampling rate for luminance is actually 13.5 MHz; the use of the number "4" is merely an historical artifact left over from a time when a 14.3 MHz NTSC sampling frequency was under consideration. The rate 13.5 MHz is actually a compromise, chosen because it has an integer relationship with both 525-line systems and 625-line systems. The "2" in the sampling rate for the color difference signals, Cb and Cr, is exactly half of the luminance sampling frequency (6.75 MHz). It is possible to halve this sampling frequency due to the luminance resolution of the human eye. Extensive psycho-visual testing has been performed for the purpose of demonstrating that the color bandwidth achievable with a sampling rate of 6. 75 MHz is sufficient for critical production applications such as chroma key.

Compression is basically a process by which the information content of an image or group of images is reduced by exploiting the redundancy present in video signals. This can be accomplished by analyzing the statistical predictability of a signal. The majority of signals have some degree of predictability. An extreme example is a sine wave, which is highly predictable because every cycle is identical and represents a single frequency, thus requiring no bandwidth. At the other extreme, noise is completely unpredictable. In reality, of course, all signals fall somewhere between these extremes. Compres-sion techniques generally seek to identify and eliminate this redundancy, thereby reducing the amount of storage and bandwidth required.

The requirement for compression, data reduction, or bit-rate reduction in video applications can be attributed to two fundamental needs:

1) The need to introduce new capabilities within an existing infra-structure. Color television is a good example. Its challange was to integrate additional information (color) into an existing black and white signal.

2) Economic benefits. In communications, the cost of data links is generally proportional to the data rate so the higher the compression factor, the lower the cost. Reduced storage costs represent another economic advantage of compression.

More on Chrominance Sub-sampling

Sub-sampling of the color difference signals which represent chrominance is frequently used to reduce data density; 4:2:2 is an outstanding example. This takes advantage of the fact that human vision is more sensitive to changes in luminance than to changes in chrominance. This is not a new concept in bandwidth both color difference signals. A similar reduction in bandwidth was also used in some early digital effects units. These devices used a 4:1:1 sampling scheme where samples are still co-sited, but the color difference sampling rate is reduced to one-fourth that of the luminance.

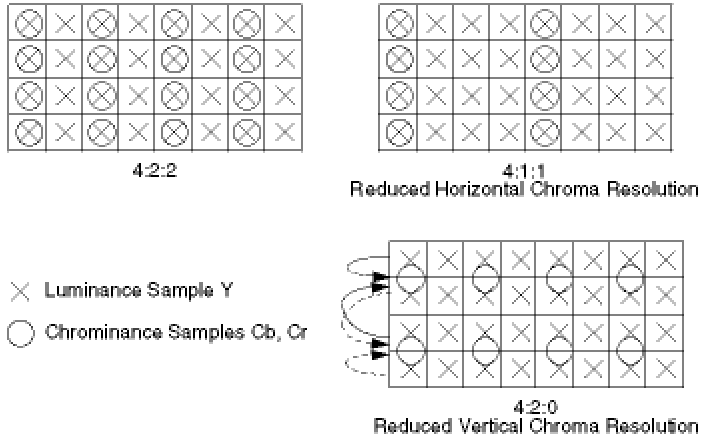
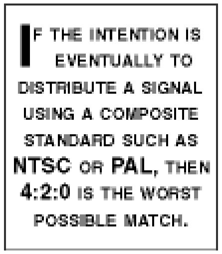


Figure 3. *4:2:0 Chroma Sub-sampling*

A reduction in the chrominance data density equivalent to 4:1:1 can also be achieved using vertical sub-sampling. In this case the objective is to reduce the chrominance resolution by equal amounts in both horizontal and vertical directions. In 4:2:0 sampling, chroma samples from two adjacent lines in a field are interpolated to produce a single chroma sample which is spatially located half-way between one of the original samples and the location of the same line but the opposite field (Figure 3). The line from the other field is that which falls between the two lines where samples are located. The solid arrow in Figure 3 shows samples from frame lines n and n+2 being interpolated to create a sample which is spatially located between frame lines n and n+1.

There are many drawbacks to 4:2:0 sampling. First, vertical resolution is already compromised due to the use of interlace. Second, vertical interpolation of the chrominance samples is complex and requires adaptive filtering. This frequently produces a less pleasing picture than would be available with 4:1:1 sampling, and results in poor multi-generational performance. Finally, the notion that the use of vertical sub-sampling reduces the chroma resolution by equal amounts in both directions has a basic flaw: it fails to consider the difference in sample density between the horizontal and vertical directions. Table 1 summarizes these sample densities for the different video formats including NTSC and PAL.


IF THE INTENTION IS EVENTUALLY TO DISTRIBUTE A SIGNAL USING A COMPOSITE STANDARD SUCH AS NTSC OR PAL, THEN 4:2:0 IS THE WORST POSSIBLE MATCH.

format	chromninance horiz. samples	chrominance vert. samples	normalized horiz. (relative to vert.)	normalized horiz. (adjusted for aspect ratio)	error factor (compared to 1.00)
4:2:2	360	240	1.5	1.125	1.125
4:1:1	180	240	0.75	0.5625	1.777
4:2:0	360	120	3.0	2.25	2.25
NTSC	140	240	0.58	0.44	2.273
PAL	140	288	0.49	0.37	2.703

***Table 1**. Sample densities by video format.*

Obviously, 4:2:0 does not reduce the chroma resolution by equal amounts in both directions. This can be seen by noting the 3:1 ratio of horizontal to vertical chrominance resolution. Even when the 4:3 aspect ratio is factored in, the horizontal to vertical ratio is still 2.25:1. Clearly, the 4:1:1 sampling structure comes much closer to providing balanced resolution in both directions.

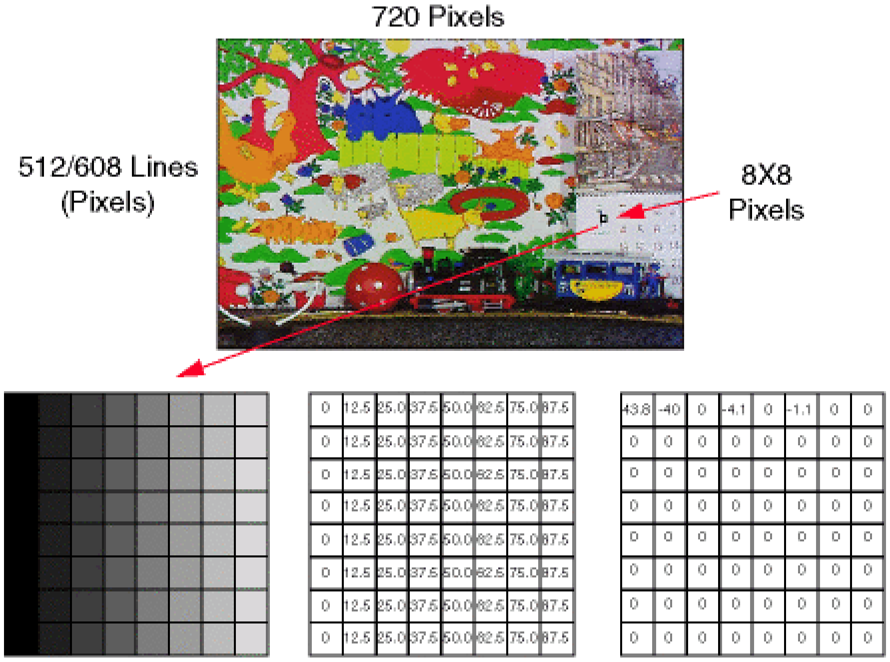
If the intention is eventually to distribute a signal using a composite standard such as NTSC or PAL, then 4:2:0 is the worst possible match. It has a higher horizontal chrominance resolution than either NTSC or PAL, and will thus deliver significantly inferior vertical chrominance resolution than these composite standards can deliver. The use of 4:2:0 will compromise both horizontal and vertical chrominance resolution.

The Compression Process

Compression is basically a process by which the information content of an image or image sequence is reduced by eliminating redundancy in video signals. Compression techniques generally attempt to identify this redundancy and expel a significant amount of it from the bit-stream.

Identification of Redundancy with DCTs

The first step in most compression systems is to identify the spatial redundancy present in each field or frame of video. This is done by applying the Discrete Cosine Transform (DCT) throughout the image. The DCT is a lossless, reversible, mathematical process which converts spatial amplitude data into spatial frequency data. For video compression, this calculation is made on 8 by 8 blocks of luminance samples and the corresponding blocks of color difference samples (Figure 4). The DCT coefficient at the upper left is the dc value for the block. All of the coefficients beneath the dc value represent increasingly higher vertical spatial frequencies. The coefficients to the right of the dc value denote increasingly higher horizontal spatial equencies. Other coefficients designate various combinations of vertical and horizontal spatial frequencies.



Picture

Sample Values

0	12.5	25.0	37.5	50.0	62.5	75.0	87.5
0	12.5	25.0	37.5	50.0	62.5	75.0	87.5
0	12.5	25.0	37.5	50.0	62.5	75.0	87.5
0	12.5	25.0	37.5	50.0	62.5	75.0	87.5
0	12.5	25.0	37.5	50.0	62.5	75.0	87.5
0	12.5	25.0	37.5	50.0	62.5	75.0	87.5
0	12.5	25.0	37.5	50.0	62.5	75.0	87.5
0	12.5	25.0	37.5	50.0	62.5	75.0	87.5

DCT Coefficients

43.8	-40	0	-4.1	0	-1.1	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

***Figure 4**. Discrete cosine transform.*

It is the nature of video that a DCT-based transformation often results in very small values for many higher spatial frequency coefficients. Similarly, it is the nature of human visual perception that many of the non-zero, higher frequency, spatial coefficients can be very coarsely defined (meaning represented by fewer bits) or completely dropped without noticeable picture degradation. The DCT does not reduce the data. In fact, to be properly reversible and lossless, more significant bits must actual ly be carried to ensure that there are no rounding errors in the calculations.

Intra-frame Compression

The actual compression begins with a reduction of the spatial redundancy. This is accomplished using intra-frame (inside the frame) compression (Figure 5). Intra-frame compression uses a combination of lossy and lossless processing to reduce the data for one picture. It does not use any information from previous or future pictures. Note that the use of the word "picture" here is intentional. Some compression techniques, like MPEG, allow either a field or a frame to be used as the fundamental image. In the field-based example, intra-field coding would produce two pictures per frame. This is the reason the word "picture," not "frame," is used when discussing video compression.

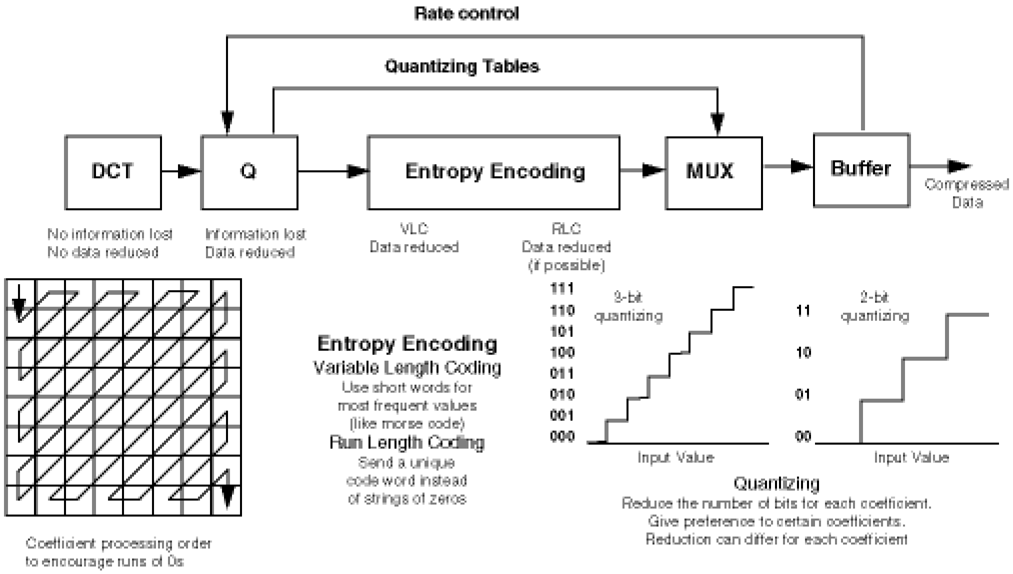


***Figure 5**. Intra-frame compression.*

Quantization

In addition to the modest amount of information lost in the 4:2:2 to 4:2:0 conversion, the power of MPEG compression comes from the clever quantization of the DCT coefficients. Quantizing is simply the process of reducing the number of bits which represent each coefficient.

MPEG HAS THE CAPABILITY TO

Quantization may use up to 11 bits to carry the DC coefficient but significantly fewer bits are used to carry the higher order coefficients. A different quantizing scale may be specified for each macroblock (16 x 16 pixels) or for large groups of macroblocks. This capability to use different quantizing factors for each macroblock is the primary reason that intra-frame-only MPEG has the capability to provide between 10%-20% more efficient compression than the equivalent quality Motion JPEG. In general, it takes substantially less data to send the tables and the heavily quantized coefficients than it would to send the original DCT coefficients.

Lossless Compression

Following quantization, lossless data reduction is applied using VLC (variable length coding) and RLC (run length coding). The order in which the coefficients are sent optimizes the efficiency of this encoding process. Processing the 64 coefficients in the 8X8 blocks using a zigzag pattern maximizes runs of zero values for more efficient compression. Variable length encoding is a process that identifies common patterns (or words) in the data and uses fewer bits to code frequently occurring values and more words to code less frequently occurring values. Morse code is a form of VLC using very short sequences for frequently occurring letters such as "e" (one dot). Another example of VLC is the popular PC program PKZIP which uses the Lempel-Ziv-Welch (LZW) algorithm to compress data files. Like quantization, VLC coding produces tables to map patterns to codes. These tables, combined with the mapped codes, generally take much less data than the original data patterns. Run length encoding is a process by which a unique code word represents a repeating pattern such as zeroes. For example, a string of 25 zero values can be represented by the ESC character followed by the value 25 (the count) followed by the value zero. Thus, 25 bytes are compressed to only 3 bytes. Note that VLC and RLC are lossless encoding processes.

Variable Bit Rates

Rate control is the process which determines how coarsely the DCT coefficients are quantized. An output buffer smoothes the data flow and provides control to the quantizer to limit the data rate or maintain it at a given level. Some MPEG encoders produce a constant bit rate, although variable bit rate is allowed by the standard. A constant bit rate is required when recording compressed data to a video tape machine which, by definition, uses mechanical parts such as a scanner rotating at a constant speed. A variable bit rate, on the other hand, is ideal for a disk recording medium which works fine with bursty packets. In general, a variable bit rate is a better choice for providing a constant quality level. In any case, a fixed data rate is a matter of definition -- data is variable at many points in the process, given line to line variations, frame to frame variations, DCT coefficient variations, and entropy coding variations.

Eliminating Temporal Redundancy

Another property of video signals is temporal redundancy which means that, for a given image sequence, the picture content generally varies little from frame to frame. The calculation of the relative picture content position changes (motion) between frames is an important part of inter-frame (between frame) compression (Figure 6). The motion estimation process in MPEG consists of dividing the picture into macroblocks which are 16 by 16 pixels (four 8X8 blocks) and a search carried out to determine its location in a subsequent frame. Although the samples in the macroblock may have changed somewhat, correlation techniques are used to determine the best location match down to one-half pixel. A successful search will result in a motion vector for that macroblock.

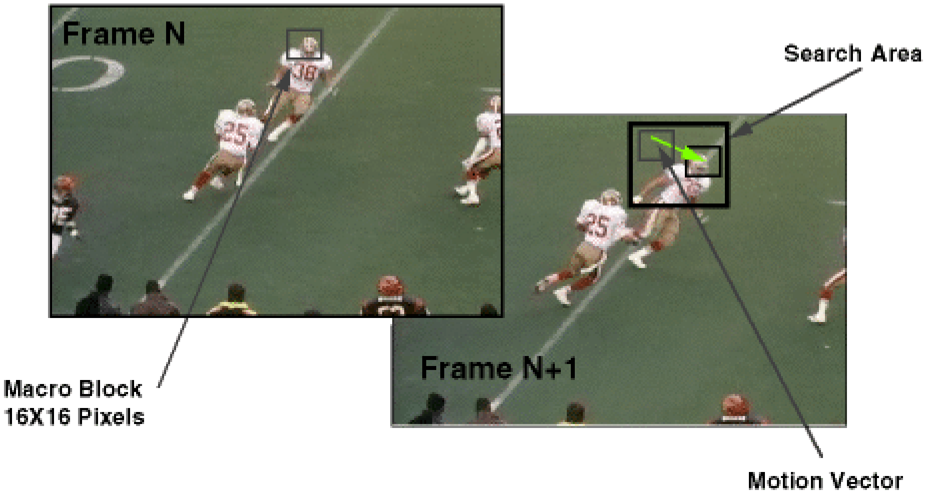


Figure 6. *Motion estimation.*

Inter-frame Compression and Motion Compensation

Inter-frame compression works on the uncompressed picture and is essentially lossless. In Figure 7, the Anchor Frame(s) Store has a full resolution, full data, previous picture. In the Motion Compensation block, vectors are calculated that can best predict the present frame. Since the frames may be different in various manners and only macroblock vectors are allowed, the prediction may not be perfect (the insignificance of imperfect prediction is addressed below).

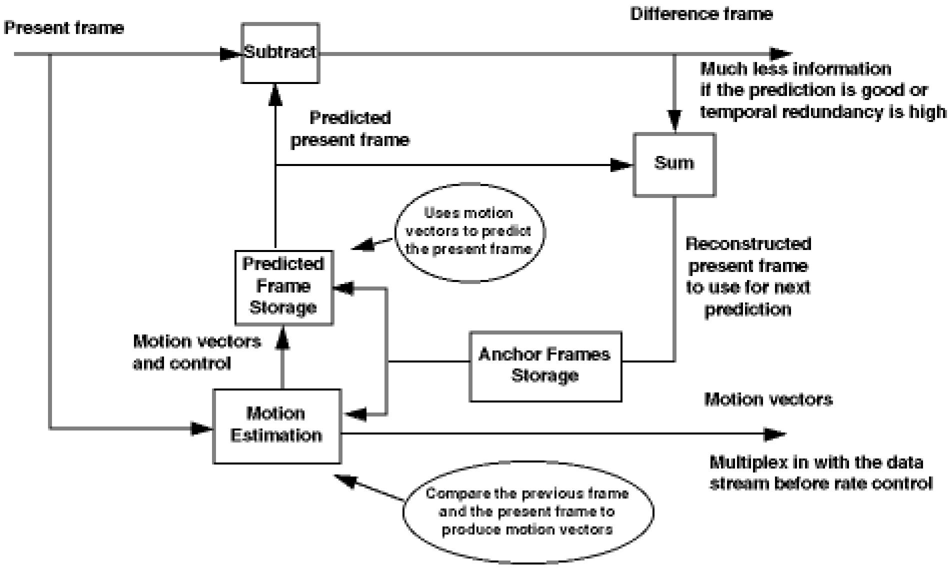


Figure 7. *Inter-frame compression.*

The Predicted Frame Store holds the predicted present frame which has been constructed using the previous frame and the motion vector information. The predicted present picture is then subtracted from the actual present picture and the difference is the output. If there was no motion and no other changes (consider the extreme of a frame repeat), the present frame could be perfectly predicted and the difference frame output would be zero (very easy to compress). When the two frames aren't the same, the difference frame can still have much less information and be easier to compress.

Note that Figure 7 shows how to make both forward predicted frames, known as P-frames, and bi-directionally predicted ones, called B-frames. These are defined in greater detail below. The essential point is that this combination of intra-frame and inter-frame compression forms the foundation of the basic MPEG-2 video compression system (Figure 8). The predicted frame is developed from a DCT quantized, decoded picture. This produces better overall results since the encoder duplicates the decoder process, thus minimizing the artifacts of mathematical operations such as truncation and rounding error.

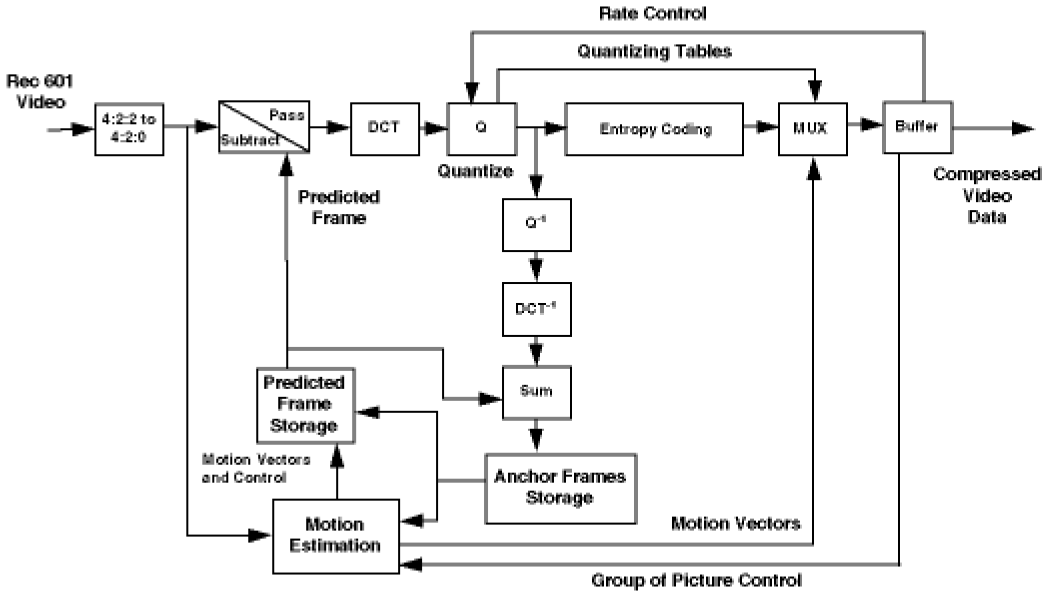


Figure 8. *MPEG video compression.*

A Note on How Anchor Frame Storage Affects Prediction Capabilities

The main difference between an encoder capable of only P pictures and one capable of both P and B pictures lies in the Anchor Frame(s) Storage. Forward prediction requires storage of only the last anchor picture and bi-directional prediction requires storage of the last and a future anchor picture.

I-, B-, and P- Frames

Based on control of the inter-frame compression, three types of frames (pictures) are developed (Figure 9):

* I-frames which are Intra-frame coded only;

* P-frames which are forward predicted from
I-frames or other P-frames;

* B-frames which are bi-directionally predicted from I-frames and/or P-frames.

Figure 9 illustrates the bi-directional prediction for B-frame 1. Frame 2 will be predicted from the same I-frame and P-frame as B-frame 1. Frame 4 will be predicted from P-frame 3 and the next I-frame numbered 0 (the temporal equivalent to T0 in time references).

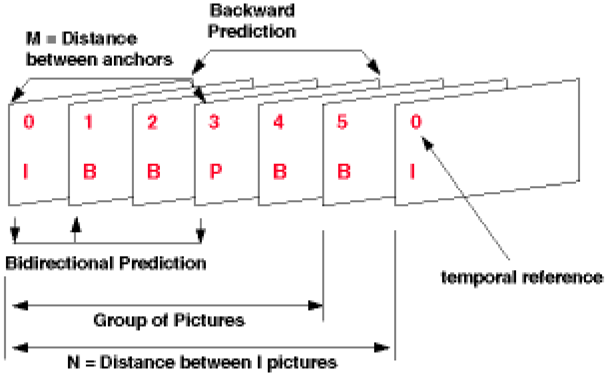


***Figure 9**. MPEG picture structure.*

Groups of Pictures

group-of-pictures (GOP) begins with an I-frame and extends to the last frame before the next I-frame. The GOP sequence shown is known as an open GOP -- the last frame in the GOP uses the first frame in the next GOP as a reference. Another type of GOP is the closed GOP, which has no prediction links to the next GOP and by definition always ends in a P frame. The frame sequence in Figure 9 is shown in the video input order. However, to predict the B-frame, the two anchor frames (I and/or P) must be known. Therefore the order of the frames leaving the decoder would be 0, 3, 1, 2, 0, 4, 5 where the second 0 is the I-frame from the next group of pictures. Only the temporal-reference is added to the compressed video data. Further timing information to help the decoder is added when the transport stream is developed (known as the presentation time stamp). This is MPEG compression.

IT IS CRITICAL TO UNDERSTAND THAT THE **MPEG** STANDARD ONLY DEFINES A STANDARD CONFIGURATION FOR THE *DECODER*.

THE MPEG-2 STANDARD

The MPEG-2 standard was defined by the International Standards Organization. The earlier standard, MPEG-1, focused on lower bandwidth, (and hence lower resolution) video at roughly 1.5 Mb/s. MPEG-2 uses many of the same techniques, but focuses on bitrates in excess of 4 Mb/s (commonly up to 15 Mb/s but potentially much higher).

Elementary and Transport Streams

Audio and video are compressed to form ES (elementary streams). The ES are then used to form PES (packetized elementary streams) which are further packetized to form TS (transport streams). This is illustrated in Figure 10.

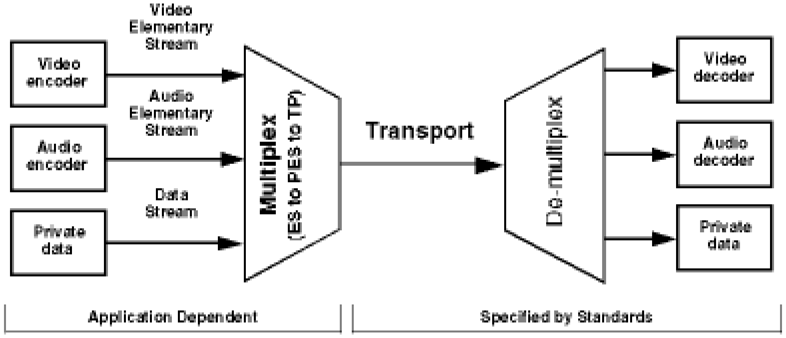
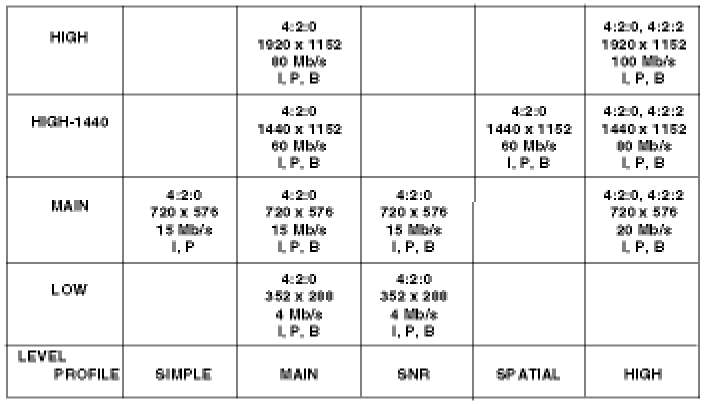


***Figure 10**. MPEG-2 standards.*

Decoder Standardization wer in the encoder increases in the future.

Profiles and Levels

The MPEG specification is intended to be generic in the sense that it serves a wide range of applications. Bitrates of up to 400 Gb/s and picture sizes up to 16,000 by 16,000 pixels can be defined. In order to put practical bounds on the many parameters for real applications a system of "profiles" and "levels" is defined (Table 2). A profile is a subset of the entire bit stream syntax. As an example the syntax allows for scalable operation, either SNR or spatial. The main and simple profiles do not use that part of the syntax and do not provide that feature. A level constrains the parameters within the allowed syntax. As an example the main profile, main level allows a maximum bit rate of 15 Mb/s whereas the main profile, high level allows a maximum bit rate of 80 Mb/s.



HIGH		4:2:0 1920 x 1152 80 Mb/s I, P, B			4:2:0, 4:2:2 1920 x 1152 100 Mb/s I, P, B
HIGH-1440		4:2:0 1440 x 1152 60 Mb/s I, P, B		4:2:0 1440 x 1152 60 Mb/s I, P, B	4:2:0, 4:2:2 1440 x 1152 80 Mb/s I, P, B
MAIN	4:2:0 720 x 576 15 Mb/s I, P	4:2:0 720 x 576 15 Mb/s I, P, B	4:2:0 720 x 576 15 Mb/s I, P, B		4:2:0, 4:2:2 720 x 576 20 Mb/s I, P, B
LOW		4:2:0 352 x 288 4 Mb/s I, P, B	4:2:0 352 x 288 4 Mb/s I, P, B		
LEVEL PROFILE	SIMPLE	MAIN	SNR	SPATIAL	HIGH

***Table 2**. MPEG-2 profiles and levels.*

ScalabilityMPEG decoders and encoders do not need to be of comparable quality to work together. This is because the specification allows for decoders to be scalable. Scalability is a method whereby a simple, inexpensive decoder can posess the ability to decode only a portion of the complete bit stream and produce a lower quality picture than a full bit stream decoder. The standard allows for scalability along an SNR tradeoff or spatial resolution tradeoff.

Signal-to-Noise (SNR) scalability means that a quality tradeoff is made against SNR performance. A low bit rate decoder will have full resolution but will have less signal to noise ratio than a high bit rate one.

Spatial scalability makes this tradeoff against spatial resolution. The low bit rate receiver produces a picture with less resolution than the full bit rate one.

To be compliant with the standard, decoders operating in one of the eleven defined profiles/levels must be able to decode any defined profile/level to the left and/or lower quadrants in the chart. This is known as a fully-compliant MPEG-2 decoder. Most product development today is for main profile, main level with some work on the main profile, high level.

MPEG-2 4:2:2 Profile @ Main Level The MPEG-2 standard was initially designed for video distribution, providing an acceptable quality level to the home viewer. However, primarily due to the limited data rates (up to 15 Mb/s) and 4:2:0 sampling, this quality is not deemed adequate for professional production and post-production applications. The 4:2:0 sampling structure does not produce the picture quality required for studio applications over multiple generations due to the interpolation of the chrominance data. Using data rates up to 15 Mb/s with a small GOP (required for editability) does not produce the required picture quality. Large GOPs are potentially difficult to handle in any kind of studio vertical interval switching situation.

> 4:2:2@ML extends extends the bit rates to 50 Mb/s and supports both 4:2:0 and 4:2:2 chrominance sampling formats. Like the MP@ML, it does not carry the complexity burden of supporting the scalability modes of the higher profiles and levels.

In 1994, a number of manufacturers and users requested that the MPEG committee define a 4:2:2 profile which would allow the use of that sampling structure for 525 and 625 line systems at bit rates up to 50 Mb/s in order to accommodate higher quality for professional applications. At that time, Tektronix, Inc. took the initiative of organizing an ad-hoc consortium of a number of companies to propose a new profile to the MPEG-2 standard. Even before the 4:2:2 profile proposal, Tektronix worked actively in the MPEG standards community to ensure that features were included in the basic syntax such as 11 bit DC coefficients, separate luminance and chrominance quantization tables, etc. in order to allow the addition of a higher quality profile in the future.

The Need for a 4:2:2 Profile

In examining Table 2, the need for a new 4:2:2 profile may not be directly observable. After all, bit rates up to 100 Mb/s are covered by the high profile at high level (HP@HL) definition and a 4:2:2 video representation can be handled by any profile above main at high level. Why, then, is there a need for yet another profile?

The answer to this question lies in the economics of implementation. In other words, the expense involved in building a fully compliant decoder. Remember, to be fully compliant, a decoder must handle any profile/level combination in the quadrants to its left; and any below it. This means that in order to be compliant, a decoder at high profile/main level (HP@ML) would need to decode the SNR scalable profile at main and low level. Scalability is very complex and generally not being pursued by equipment manufacturers. One reason for this is that in simple terms, the encoding and the decoding process will double the number of encoding and decoding circuits (Figures 11 and 12) basically doubling the cost of a scalable decoder. In addition, HP@ML has a maximum bit rate of 20 Mb/s which at short GOP structures does not provide the quality needed in professional applications. Thus, HP@ML requires a higher level implementation which in turn requires spatial scalability functionality in order to be compliant. In either case, professional applications will be burdened with additional costs which can easily be avoided with the addition of the 4:2:2 profile at main level (MPEG-2 4:2:2@ML).

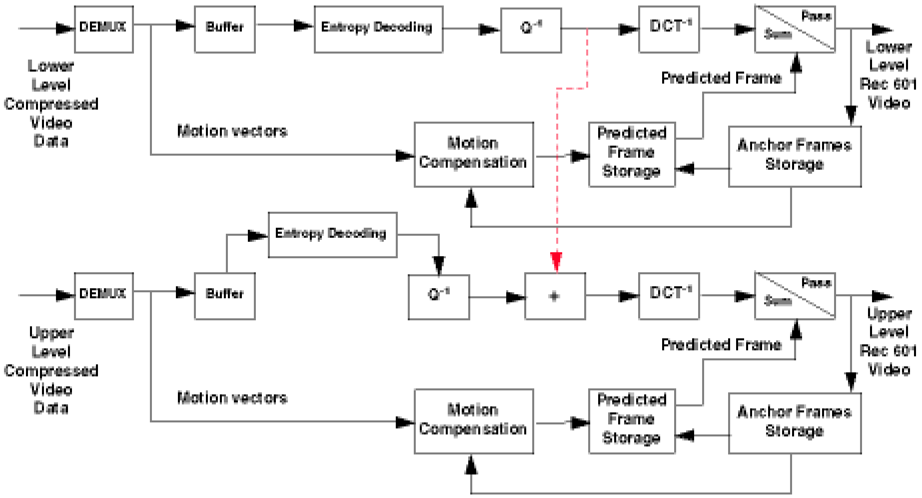


***Figure 11**. SNR scalable decoder*

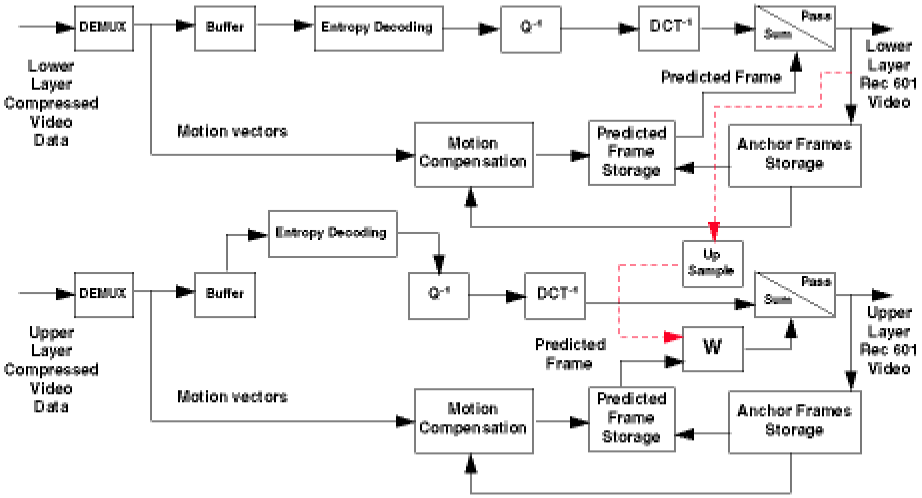


Figure 12. *Spatial scalable decoder*

In January 1996, MPEG approved the 4:2:2 profile at main level making it an international standard. As can be seen in Figure 13, MPEG-2 4:2:2@ML extends the capabilities of the MP@ML in several ways. It extends the bit rates to 50 Mb/s and supports both 4:2:0 and 4:2:2 chrominance sampling formats. Like the MP@ML, it does not carry the complexity burden of supporting the scalability modes of the higher profiles and levels.

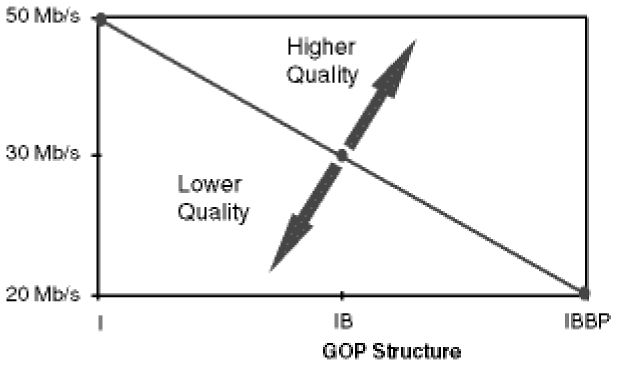
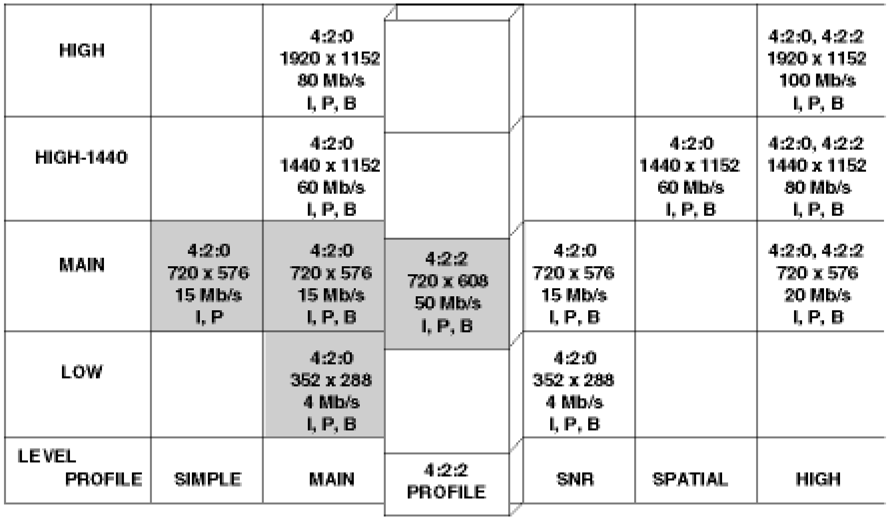


Figure 13. *Achieving quality.*

The highlighted profiles and levels in Table 3 can all be decoded by a compliant 4:2:2 decoder. A fully compliant MPEG-2 4:2:2@ML will decode data rates up to 50 Mb/s and any combination of I, P, and B pictures as well as MP@ML up to 15 Mb/s with any combination of I, P, and B pictures, MP@LL (low level) with any combination of I, P, and B pictures, and

SP@ML (simple profile) with any combination of I and P pictures. Any 4:2:2 decoder limited to a fixed rate (say 18 Mb/s) and a subset of the picture types (say I and B only), and no capability to handle 4:2:0 is not a fully compliant 4:2:2@ML decoder. Also note that a decoder which complies with SNR profile @ Main level does not need to decode the 4:2:2 profile to be considered compliant.



LEVEL / PROFILE	SIMPLE	MAIN	4:2:2 PROFILE	SNR	SPATIAL	HIGH
HIGH		4:2:0 1920 x 1152 80 Mb/s I, P, B				4:2:0, 4:2:2 1920 x 1152 100 Mb/s I, P, B
HIGH-1440		4:2:0 1440 x 1152 60 Mb/s I, P, B			4:2:0 1440 x 1152 60 Mb/s I, P, B	4:2:0, 4:2:2 1440 x 1152 80 Mb/s I, P, B
MAIN	4:2:0 720 x 576 15 Mb/s I, P	4:2:0 720 x 576 15 Mb/s I, P, B	4:2:2 720 x 608 50 Mb/s I, P, B	4:2:0 720 x 576 15 Mb/s I, P, B		4:2:0, 4:2:2 720 x 576 20 Mb/s I, P, B
LOW		4:2:0 352 x 288 4 Mb/s I, P, B		4:2:0 352 x 288 4 Mb/s I, P, B		

***Table 3**. MPEG-2 4:2:2 profile @ mail level.*

MPEG-2 4:2:2@ML provides cost-effective compression in applications which require:

* High level of interoperability and flexibility

* Higher quality than MP@ML

* Better chroma resolution than MP@ML

* Post processing after compression and decompression

* Multiple generations of compression and decompression

* Short Group of Pictures (GOP) for editability

* Capability to represent all active lines of video

* Capability to represent vertical blanking interval information.

How Good is MPEG-2 4:2:2@ML?

In February 1995, Tektronix helped SMPTE conduct expert viewing tests of 4:2:2 MPEG-2 at bit rates between 20 Mb/s and 50 Mb/s. These viewing tests verified that MPEG-2 4:2:2 met the requirements of professional applications. Further subjective viewing tests were conducted by MPEG after the 4:2:2 profile parameters were agreed upon. Tektronix once again played a key role in these tests by supplying some of the test sequences, assisting with compression simulations, editing the 525/60 test tapes, an d working with SMPTE to organize subjective viewing sessions. The results of this last set of tests are illustrated in Table 4.

There were separate subjective assessment but fewer combinations of data rates, GOP structures, and number of generations. The test included compression and decompression from first to eighth generations (a generation here represents a compression/de-compression cycle). Both spatial and temporal shifting were applied in order to account for the effects of picture repositioning (which might occur in a DVE) and picture alignment over multiple generations. Spatial shifting means that the picture was shifted horizontally and vertically by two pixels and two spatial lines between first and second generations, and back again between the fifth and sixth generations.

The program material included both typical sequences as well as difficult compression test material. Viewers were asked to rank the quality of both an original and a compressed sequence. The rankings were converted to a 0 to 100 scale where 0 means no degradation and 100 means worst possible degradation.

The results from these tests generally show the same trends and as would be expected, the expert viewers generally gave more critical B and still maintain excellent quality. over multiple generations. At 30 Mb/s it is necessary to use GOP structures of IB to maintain excellent quality, over multiple generations. At 20 Mb/s, excellent quality can be achieved only for a limited number of generations and longer GOPs such as IBBP. This suggests that lower data rates such as 20 Mb/s may be used for acquisition but serious multiple generation work should be done at higher data rates such as 30 or 50 Mb/s.

The flexibility of data rate and GOP structure is the key to achieving the right tradeoff of cost and performance for each application D1 without the nightmare of supporting multiple tape formats.

Significance of MPEG-2 to Broadcast and Post-Production

To the Broadcast and Post-Production industries, the MPEG-2 standard has the potential of finally solving the age-old multiple format dilemma. To achieve this potential, we need to keep in perspective the requirements and characteristics of the application. Consider the example applications illustrated in Figure 14.

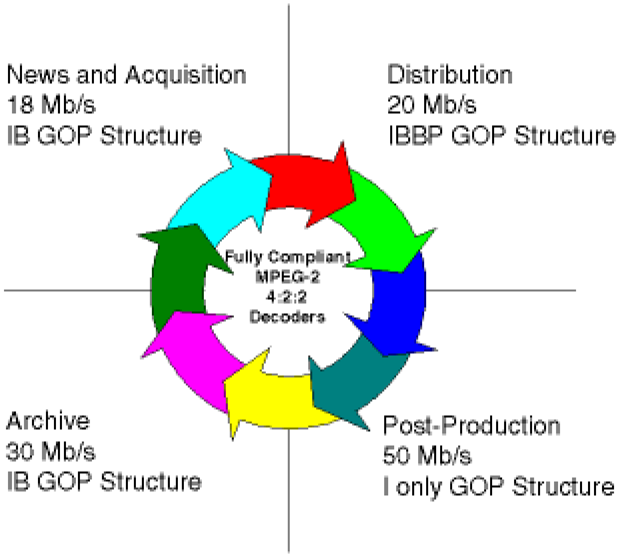


***Figure 14**. Application examples.*

News and Acquisition

This application requires high quality field video acquisition and limited number of generations (news applications). It will more than likely use a camcorder which needs to be portable, lightweight, battery operated (with a reasonable battery life), and reasonably priced (we may need to replace it periodically when we drop one in the river or one falls off the truck). All of these characteristics can be accommodated with a low data rate (less power consumption, less expensive circuits), simple GOP structure (less expensive circuits, easy to edit).

Archive

This application requires high quality or at least no loss of quality over the original material. Here, one might choose a higher data rate and an IB GOP structure for excellent quality.

Post-Production

Distribution This application requires efficient storage and broadcast quality. Here a data rate of 20 Mb/s (or 15 Mb/s to feed an NTSC/PAL transmitter or even lower such as DBS) combined with a long GOP provides both storage efficiency and excellent quality.

This concept of different data rates and GOP structures for different applications is analogous to shooting on Betacam SP, posting on D1, distributing on VHS or U-matic, and archiving on Betacam SP. At any point in time it may be necessary to use material from one application in another application and this is where the fully compliant MPEG-2 4:2:2@ML decoder can help solve the multiple format dilemma. In essence, the fully-compliant decoder has the potential to replace several different VTR formats, allowing a single piece of hardware to perform in any of these applications. You no longer have to choose formats.

CONCLUSIONS

Tektronix has played a significant role in the development of the original MPEG standard and in the 4:2:2 profile. Tektronix proposed and contributed significantly to the identification of the need for the 4:2:2 profile and later, chaired the MPEG committee on the definition of this standard. Committed to both the MPEG-2 standard and the 4:2:2 profile, Tektronix expects to be one of the first manufacturers to offer MPEG-2 4:2:2@ML products. In fact, Tektronix demonstrated early hardware prototypes at the 1996 National Association of Broadcasters (NAB) tradeshow.

It is from this perspective and background that Tektronix offers the following conclusions:

* Compliance with the standard is beneficial for manufacturers and users. It is good for manufacturers such as Tektronix because it opens the door to cost efficient implementations of providing the best solutions to our customers. To video equipment users, compliance with the standard means that equipment choices can be based on who provides the best solution. You are not tied to any one manufacturer because all equipment plays together much like it does today with composite and component video standards.

* The MPEG-2 standard represents a significant step toward interoperability.

* The 4:2:0 format (MPEG-2 MP@ML) is the best choice for delivering MPEG-2 streams directly to the home via Direct Broadcast Satellite and Cable. However, if the intent is to deliver a broadcast quality signal to an NTSC or PAL transmitter, MPEG-2 4:2:2@ML is the more appropriate choice.

* Subjective tests have demonstrated that, with proper choices of data rate and GOP structure, MPEG-2 4:2:2@ML can meet and even exceed professional video requirements.

* Flexibility in terms of data rate and GOP structure is the key to achieving quality. The tradeoffs between these two parameters are determined by the application.

Year 2000 | Contact Support | Documentation | Search

webmaster@grassvalleygroup.com