MPEG-2

What is MPEG-2 ?

At a meeting hosted in New York by Columbia University, the Moving Picture Experts Group (MPEG) completed definition of MPEG-2 Video, MPEG-2 Audio, and MPEG-2 Systems. MPEG therefore confirmed that it is on schedule to produce, by November 1993, Committee Drafts of all three parts of the MPEG-2 Standard, for balloting by its member countries.

To ensure that a harmonized solution to the widest range of applications is achieved, MPEG, an ISO/IEC working group designated JTC1/SC29/WG11, is working jointly with the ITU-TS Study Group 15 *Experts Group for ATM Video Coding*. MPEG also collaborates with representatives from other parts of ITU-TS, and from EBU, ITU-RS, SMPTE, and the North American HDTV community.

The final approval of ISO/IEC 13818-1 (MPEG-2 Systems), ISO/IEC 13818-2 (MPEG-2 Video) and ISO/IEC 13818-3 (MPEG-2Audio) as International Standard (IS) was given by the 29th meeting of ISO/IEC JTC1/SC29/WG11 (MPEG) held in Singapore in November 1994.

Why MPEG-2 ?

The MPEG-2 concept is similar to MPEG-1, but includes extensions to cover a wider range of applications. The primary application targeted during the MPEG-2 definition process was the all-digital transmission of broadcast TV quality video at coded bitrates between 4 and 9 Mbit/sec.

However, the MPEG-2 syntax has been found to be efficient for other applications such as those at higher bit rates and sample rates (e.g. HDTV). The most significant enhancement over MPEG-1 is the addition of syntax for efficient coding of interlaced video (e.g. 16x8 block size motion compensation, Dual Prime, et al).

Several other more subtle enhancements (e.g. 10-bit DCT DC precision, non-linear quantization, VLC tables, improved mismatch control) are included which have a noticeable improvement on coding efficiency, even for progressive video. Other key features of MPEG-2 are the scalable extensions which permit the division of a continuous video signal into two or more coded bit streams representing the video at different resolutions, picture quality (i.e. SNR), or picture rates.

MPEG-1 was optimized for CD-ROM or applications at about 1.5 Mbit/sec. Video was strictly non-interlaced (i.e. progressive). The international co-operation had executed so well for MPEG-1, that the committee began to address applications at broadcast TV sample rates using the CCIR-601 recommendation (720 samples/line by 480 lines per frame by 30 frames per second or about 15.2 million samples/sec including chroma) as the reference.

Unfortunately, today's TV scanning pattern is interlaced. This introduces a duality in block coding: do local redundancy areas (*blocks*) exist exclusively in a field or a frame ?
The answer of course is that some blocks are one or the other at different times, depending on motion activity. The additional years of experimentation and implementation between MPEG-1 and MPEG-2 improved the method of block-based transform coding.

What are the typical MPEG-2 bitrates and picture quality ?

Here are some examples of typical frame sizes in bits :

```
                                Picture type
                        I               P               B          Average
MPEG-1 SIF
@ 1.15 Mbit/sec         150,000         50,000          20,000      38,000

MPEG-2 601              400,000         200,000         80,000     130,000
@ 4.00 Mbit/sec

Note: parameters assume Test Model for encoding, I frame distance of 15
(N = 15), and a P frame distance of 3 (M = 3).
```

Of course with scene changes and more advanced encoder models found in any real-world implementation, these numbers can be very different.

When will an MPEG-2 decoder chip be available ?

Several chips are available since late 1993. For reasons of economy and scale in the cable TV application, all are single-chip (not including DRAM and host CPU/controller) implementations.
They are:

```
  SGS-Thomson  STi-3500
        first MPEG-2 chip on market
        multi-tap binary horizontal sample rate converter.
        pan & scanning support for 16:9
        requires external, dedicated microcontroller (8 bit)
        8-bit data bus, no serial data bus.

  LSI Logic L64112 successor (pin compatible)
        serial bus, 15 Mbit coded throughput.
        smaller pin-count version due soon.
```

```
C-Cube CL-9100 for MPEG-2 video
C-Cube CL-9110 for system/demutiplixer.
```

In 1994 there have been announcements from :

Pioneer : single-chip MPEG-2 successor to CD-1100 MPEG-1 chip set.
IBM : single-chip decoder.
Hyundai.

Where will we see MPEG-2 in everyday life ?

Just about wherever you see video today.

- **DBS (Direct Broadcast Satellite)**
  The Hughes/USSB service will use MPEG-2 video and audio. Thomson has exclusive rights to manufacture the decoding boxes for the first 18 months of operation. No doubt Thomson 's STi-3500 MPEG-2 video decoder chip will be featured. Hughes/USSB DBS already begun service in North America in 1994. Two satellites at 101 degrees West share the power requirements of 120 Watts per 27 MHz transponder. Multi-source channel rate control methods is employed to optimally allocate bits between several programs on one data carrier. An average of 150 channels are planned.
- **CATV (Cable Television)**
  Despite conflicting options, the the cable industry has more or less settled on MPEG-2 video. Audio is less than settled. For example, General Instruments (the largest U.S. consumer cable set-top box manufacturer) have announced the planned use of the Dolby AC-3 audio algorithm.
- **DigiCipher**
  The General Instruments DigiCipher I video syntax is similar to MPEG-2 syntax but uses smaller macroblock predictions and no B-frames. The DigiCipher II specification includes modes to support both the GI and full MPEG-2 Video Main Profile syntax. Services such as HBO will upgrade to DigiCipher II in 1994.
  At the European IBC broadcast technology convention, in September 1994, GI demonstrated a prototype DCII encoder which handles both digital encoding standards. Fully configured the encoder will be able to process 16 analogue video inputs, plus 32 stereo audio channels and 32 data channels into a single high speed datastream which can be carried on cable, satellite, microwave or ATM systems.
  DCII technology has now been licensed to Scientific Atlanta and Hewlett Packard (both set-top manufacturers) and to chip manufacturers Motorola, LSI Logic and C-Cube. All these manufacturers already support MPEG2 and plan to incorporate DCII into *dual mode* digital video decoder chips for the set-top terminal market.
- **HDTV**
  The U.S. Grand Alliance, a consortium of companies that formely competed for the U.S. terrestrial HDTV standard, have already agreed to use the MPEG-2 Video and Systems syntax (including B-pictures) . Both interlaced (1440 x 960 x 30 Hz) and progressive (1280 x 720 x 60 Hz) modes will be supported. The Alliance must then settle upon a modulation (QAM, VSB, OFDM), convolution (MS or Viterbi), and error

correction (RSPC, RSFC) specification.

In September 1993, the consortium of 85 European companies signed an agreement to fund a project known Digital Video Broadcasting (DVB) which will develop a standard for cable and terrestrial transmission by the end of 1994. The scheme will use MPEG-2. This consortium has put the final nail in the coffin of the D-MAC scheme for gradual migration towards an all-digital, HDTV consumer transmission standard. The only remaining analog or digital-analog hybrid system left in the world is NHK's MUSE (which will probably be axed in a few years).

What did MPEG-2 add to MPEG-1 in terms of syntax/algorithm ?

Here is a brief summary:

- Sequence layer:

  More aspect ratios. A minor, yet necessary part of the syntax.

  Horizontal and vertical dimensions are now required to be a multiple of 16 in frame coded pictures, and the vertical dimension must be a multiple of 32 in field coded pictures.

  4:2:2 and 4:4:4 macroblocks were added in the Next profiles.

  Syntax can now signal frame sizes as large as 16383 x 16383.

  Syntax signals source video type (NTSC, PAL, SECAM, MAC, component) to help post-processing and display.

  Source video color primaries (609, 170M, 240M, D65, etc.) and opto- electronic transfer characteristics (709, 624-4M, 170M etc.) can be indicated.

  Four scalable modes.

- Picture layer:

  All MPEG-2 motion vectors have half-pel accuracy.
  MPEG-1 allows to code ful-pel vectors, but this does not work well because MPEG-1 does not have a *filter in the loop*, like H.261. So, practically, all MPEG-1 bitstreams use 1/2 pels motion vectors. MPEG-2 does not allow full-pels vectors.

DC precision can be user-selected as 8, 9, 10, or 11 bits.

Concealment motion vectors can be added to I-pictures in order to increase robustness from bit errors since I pictures are the most critical and sensitive in a group of pictures.
Concealment motion vectors can also be added to P and B-frames. They are not useful in B-frames, but they are useful in P-frames.

A non-linear macroblock quantization factor that results in a more dynamic step size range, from 0.5 to 56, than in MPEG-1 (1 to 31).

New Intra-VLC table for dct_next_coefficient (AC run-level events) that is more geared towards I-frame probability distribution. EOB is 4 bits. The old tables are still included.

Alternate scanning pattern that (supposedly) improves entropy coding performance over the original Zig-Zag scan used in H.261, JPEG, and MPEG-1. The extra scanning pattern is geared towards interlaced video.

Syntax to signal 3:2 pulldown process (repeat_field_first flag)

Syntax flag to signal chrominance post processing type (4:2:0 to 4:2:2 upsampling conversion)

Progressive and interlaced frame coding

Field-pictures and frame-pictures (MPEG-1 has only frame-pictures). With field-pictures, I-frames cost a lot less, since only one field is coded intra.

Group of pictures (GOP) are optional, and direct access to a bitstream can be done at any repeated sequence header, even if there is no GOP header there.

Syntax to signal source composite video characteristics useful in post-processing operations. (v-axis, field sequence, sub_carrier, phase, burst_amplitude, etc.)

Pan & scanning syntax that tells decoder how to, for example, window a 4:3 image within a wider 16:9 aspect ratio image. Vertical pan offset has 1/16th pixel accuracy.

- Macroblock layer:

Macroblock stuffing is now illegal in MPEG-2.

Two line modes (*interlaced* and *progressive*) for DCT operation.

Now only one run-level escape code code (24-bits) instead of the single (20-bits) and double escape (28-bits) in MPEG-1.

Improved mismatch control in quantization over the original oddification method in MPEG-1. Now specifies adding or subtracting one to the 63rd AC coefficient depending on parity of summed quantized coefficients.

Quantizer matrices are downloadable before each frame.

The range of the coefficients that can be coded is extended to -2043, +2043 (in MPEG-1 was -255 to +255 only).

Many additional prediction modes (16x8 MC, field MC, Dual Prime) and, correspondingly, macroblock modes.

Overall, MPEG-2's greatest compression improvements over MPEG-1 are: prediction modes, Intra VLC table, DC precision, non-linear macroblock quantization. Implementation improvements *(macroblock stuffing was eliminated)*.

What are the scalable modes of MPEG-2 ?

Scalable video is permitted only in the Main+ and Next profiles. Currently, there are four scalable modes in the MPEG-2 toolkit. These modes break MPEG-2 video into different layers (base, middle, and high layers) mostly for purposes of prioritizing video data. For example, the high priority channel (bitstream) can be coded with a combination of extra error correction information and decreased bit error (i.e. higher Carrier-to-Noise ratio or signal strength) than the lower priority channel.

Another purpose of scalability is complexity division. For example, in HDTV, the high priority bitstream (720 x 480) can be decoded under noise conditions were the lower priority (1440 x 960) cannot. This is *graceful* degradation. By the same division however, a standard TV set need only decode the 720 x 480 channel, thus requiring a less expensive decoder than a TV set wishing to display 1440 x 960. This is simulcasting.

A brief summary of the MPEG-2 video scalability modes:

- Spatial Scalability

  Useful in simulcasting, and for feasible software decoding of the lower resolution, base layer. This spatial domain method codes a base layer at lower sampling dimensions (i.e. *resolution*) than the upper layers. The upsampled reconstructed lower (base) layers are then used as prediction for the higher layers.

- Data Partitioning

  Similar to JPEG's frequency progressive mode, only the slice layer indicates the maximum number of block transform coefficients contained in the particular bitstream (known as the *priority break point*). Data partitioning is a frequency domain method that breaks the block of 64 quantized transform coefficients into two bitstreams. The first, higher priority bitstream contains the more critical lower frequency coefficients and side informations (such as DC values, motion vectors). The second, lower priority bitstream carries higher frequency AC data.

- SNR Scalability

  Similar to the point transform in JPEG, SNR scalability is a spatial domain method where channels are coded at identical sample rates, but with differing picture quality (through quantization step sizes). The higher priority bitstream contains base layer data that can be added to a lower priority refinement layer to construct a higher quality picture.

- Temporal Scalability

  A temporal domain method useful in, e.g., stereoscopic video. The first, higher priority bitstreams codes video at a lower frame rate, and the intermediate frames can be coded in a second bitstream using the first bitstream reconstruction as prediction. In stereoscopic vision, for example, the left video channel can be predicted from the right channel.

Other scalability modes were experimented with in MPEG-2 video (such as Frequency Scalability), but were eventually dropped in favor of methods that demonstrated similar quality and greater simplicity.

What is the TM (Test Model) rate control and adaptive quantization technique ?

Test model was not by any stretch of the imagination meant to be the show-stopping, best set of algorithm. It was designed to exercise the syntax, verify proposals, and test the *relative* performance of proposals in a way that could be duplicated by co-experimentors in a timely fashion. Otherwise there would be more endless debates about model interpretation than actual time spent in verification.

The MPEG-2 Test Model (TM) rate control method offers a dramatic improvement to the Simulation Model (SM) method used for MPEG-1. TM's improvements are due to more sophistication pre-analysis and post-analysis routines. Rate control and adaptive quantization are divided into three steps:

1. Bit Allocation

   In Complexity Estimation, the global complexity measures assign relative weights to each picture type. These weights (Xi, Xp, Xb) are reflected by the typical coded frame size of I, P, and B pictures. I pictures are assigned the largest weight since they have the greatest stability factor in an image sequence. B pictures are assigned the smallest weight since B data does not propagate into other frames through the prediction process.

Picture Target Setting allocates target bits for a frame based on the frame type and the remaining number of frames of that same type in the Group of Pictures (GOP).

2. Rate Control

   Rate control attempts to adjust bit allocation if there is significant difference between the target bits (anticipated bits) and actual coded bits for a block of data.

3. Adaptive Quantization

   Recomputes macroblock quantization factor according to activity of block against the normalized activity of the frame.

   The effect of this step is to roughly assign a constant number of bits per macroblock (this results in more perceptually uniform picture quality).

What is MPEG-2 VIDEO ?

MPEG-2 Video is a generic method for compressed representation of video sequences using a common coding syntax defined in the document ISO/IEC 13818 Part 2 (CD: Nov. 1993, DIS: March 1994) by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), in collaboration with the International Telecommunications Union (ITU) as Recommendation H.262.

The MPEG-2 Video Standard specifies the coded bit stream for high-quality digital video. As a compatible extension, MPEG-2 Video builds on the completed MPEG-1 Video Standard (ISO/IEC IS 11172-2), by supporting interlaced video formats and a number of other advanced features, including features to support HDTV.

As a generic International Standard, MPEG-2 Video is being defined in terms of extensible profiles,each of which will support the features needed by an important class of applications. At the Sydney MPEG meeting, the MPEG-2 Main Profile was defined to support digital video transmission in the range of about 2 to 15 Mbits/sec over cable, satellite, and other broadcast channels, as well as for Digital Storage Media (DSM) and other communications applications. Building on this success at the New York meeting, MPEG experts from participating countries in Asia, Australia, Europe, and North America further defined parameters of the Main Profile and Simple Profile suitable for supporting HDTV formats.

MPEG experts also extended the features of the Main Profile by defining a hierarchical/scalable profile. This profile aims to support applications such as compatible terrestrial TV/HDTV, packet-network video systems, backward compatibility with existing standards (MPEG-1 and H.261), and other applications for which multi-level coding is required. For example, such a system could give the consumer the option of using either a small portable receiver to decode standard definition TV, or a larger fixed receiver to decode HDTV from the same broadcast signal.

The technical definition of MPEG-2 Video has been completed. This was a critical milestone, and MPEG-2 Video was scheduled for a Committee Draft in November 1993.

What are MPEG-2 VIDEO Main Profile and Main Level ?

It is inappropriate to talk about MPEG-2 profiles without also talking about levels. The 4 profiles define the colorspace resolution, and scalability of the bitstream.
The levels define the maximum and minumum for image resolution, and Y (Luminance) samples per second, the number of video and audio layers supported for scalable profiles, and the maximum bit rate per profile.

The combination of a profile and a level produces an architecture which defines the ability of a decoder to handle a particular bitstream.

MPEG-2 Video Main Level is analogous to MPEG-1's CPB, with sampling limits at CCIR-601 parameters (720 x 480 x 30 Hz). Profiles limit syntax (i.e. algorithms), whereas Levels limit parameters (sample rates, frame dimensions, coded bitrates, etc.). Together, Video Main Profile and Main Level (abbreviated as MP@ML) normalize complexity within feasible limits of 1994 VLSI technology (0.5 micron), yet still meet the needs of the majority of application users.

```
  Level      Max. sampling     Pixels/  Max.     Significance
             dimensions   fps  sec      bitrate
  ---------  ----------------  -------  -------  -------------------------
  Low         352 x  240 x 30   3.05 M   4 Mb/s  CIF, consumer tape equiv.
  Main        720 x  480 x 30  10.40 M  15 Mb/s  CCIR 601, studio TV
  High 1440  1440 x 1152 x 30  47.00 M  60 Mb/s  4x 601, consumer HDTV
  High       1920 x 1080 x 30  62.70 M  80 Mb/s  production SMPTE 240M std

Note 1: pixel rate and luminance (Y) sample rate are equivalent.
     2: Low Level is similar MPEG-1's Constrained Parameters Bitstreams.

  Profile  Comments
  -------  -----------------------------------------------------------
  Simple   Same as Main, only without B-pictures.  Intended for software
           applications, perhaps CATV.
  Main     Most decoder chips, CATV, satellite. 95% of users.
  Main+    Main with Spatial and SNR scalability
  Next     Main+ with 4:2:2 macroblocks
```

Level	Profile: Simple	Main	Main+	Next
High	illegal		illegal	4:2:2 chroma
High-1440	illegal		With spatial Scalablity	4:2:2 chroma
Main		90% of users	Main with SNR scalability	4:2:2 chroma
Low	illegal		Main with SNR scalability	illegal

At what bitrates is MPEG-2 video optimal ?

The Test subgroup has defined a few examples :

Sweet spot sampling dimensions and bit rates for MPEG-2:

Dimensions	Coded rate	Comments
352x480x24 Hz (progressive)	2 Mbit/sec	Half horizontal 601. Looks almost NTSC broadcast quality, and is a good (better) substitute for VHS. Intended for film src.
544x480x30 Hz (interlaced)	4 Mbit/sec	PAL broadcast quality (nearly full capture of 5.4 MHz luminance carrier). Also 4:3 image dimensions windowed within 720 sample/line 16:9 aspect ratio via pan&scan.
704x480x30 Hz (interlaced)	6 Mbit/sec	Full CCIR 601 sampling dimensions.

There are two separate reasons for the popularity of 544 pixel/line :

1. The sample rate is near the Nyquist limit, or better *bandlimit*, of terrestrial bandlimited signals such as PAL (5.4 MHz luminance including blanking) and NTSC (4.2 MHz). It can also be said that 544 pixels/line captures the full *glory* (or at least 405 out of the claimed 425 *TVL* or TV lines) of analog video laserdiscs.

   ```
   (4.2 MHz) * (80 samples/line/MHz) * (4/3 aspect ratio) <
   ```

```
544 samples/line <
(5.4 MHz) * (80 samples/line/MHz) * (4/3 aspect ratio)
```

544 is a nice compromise between PAL and NTSC terrestrial broadcast bandlimits. Besides NTSC D-2 signals, being sampled at 4 times color subcarrier, do much better than 4.2 MHz anyway and probably has around 5.4 MHz worth of luminance bandwidth.

2. When converting a signal that has been coded in the 16:9 aspect ratio to a 4:3 display using the Pan&Scan method (as opposed to letterboxing), 544 samples have to be extracted from the full 720 per line and then interpolated to 720 (or 704 if you prefer) to meet the requirements of subsequent display devices such as NTSC signal generators that accept video at only one sampling rate (namely CCIR-601, or in some cases "Square NTSC: 640x480").
   Since :

   720 * (9/16)(4/3) = 540

   the nearest multiple of 16 (the horizontal and vertical dimensions of macroblocks in both MPEG-1 and MPEG-2) is 544.

Since it is difficult to implement *polyphase* filters (filters that map one arbitrary sample rate to another) in inexpensive hardware that must also implement MPEG decoding and display processes, the industry has settled on a few popular line rates :

```
352             4:3  SIF video (352x240x30 or 352x288x25)
384             16:9 SIF video (384x216x 24 frames/sec)
480             NTSC bandlimit (also popular with General Instruments)
544             PAL/Laserdisc, NTSC D-2 , BetaCam SP bandlimit.
640             Square NTSC.  SECAM 6 MHz bandlimit.
704 or 720      CCIR 601 (component studio "D-1" video).
```

Notice that converting all of the above rates to 720 (or 704) pixels/line is fairly cheap to do (e.g. 720/540 = 4/3) with low-cost padding and smoothing (shifts and adds but no multiplies).

How does MPEG video really compare to TV, VHS, laserdisc ?

VHS picture quality can be achieved for source film video at about 1 million bits per second (with proprietary encoding methods). It is very difficult to objectively compare MPEG to VHS. The response curve of VHS places -3 dB at around 2 MHz of analog luminance bandwidth (equivalent to 200 samples/line). VHS chroma is considerably less dense in the horizontal direction than MPEG source video (compare 80 samples/ line to 176!). From a sampling density perspective, VHS is superior only in the vertical direction (480 lines compared to 240), but when taking into account interfield magnetic tape crosstalk and the TV monitor Kell factor, not by all that much. VHS is prone to timing errors (which can be improved with time base correctors), whereas digital video is fully discretized. Pre-recorded VHS is typically recorded at very high duplication speeds (5 to 15 times real time playback), which leads to further shortfalls for the format that has been with us since 1977.

Broadcast NTSC quality can be approximated at about 3 Mbit/sec, and PAL quality at about 4 Mbit/sec. Of course, sports sequences with complex spatial-temporal activity need more like 5 and 6 Mbit/sec, respectively.

Laserdisc is a tough one to compare. Disc is composite video (NTSC or PAL) with up to 425 TVL (or 567 samples/line) response. Thus it could be said that laserdisc has 567 x 480 x 30 Hz *resolution*. The carrier-to-noise ratio is typically better than 48 dB. Timing is excellent. Yet some of the clean characteristics of laserdisc can be achieved at 1.15 Mbit/sec (SIF rates), especially for those areas of medium detail (low spatial activity) in the presence of uniform motion. This is why some people say MPEG-1 video at 1.15 Mbit/sec looks almost as good as Laserdisc or Super VHS.

Regardless of the above figures, those clever proprietary encoding algorithms can push these bitrates even lower.

Why film does so well with MPEG ?

Several reasons, really:

1. The frame rate is 24 Hz (instead of 30 Hz) which is a savings of some 20%.
2. the film source video is inherently progressive. Hence no fussy interlaced spectral frequencies.
3. the pre-digital source was severely oversampled (compare 352 x 240 SIF to 35 millimeter film at, say, 3000 x 2000 samples). This can result in a very high quality signal, whereas most video cameras do not oversample, especially in the vertical direction.
4. Finally, the spatial and temporal modulation transfer function (MTF) characteristics (motion blur, etc) of film are more ameniable to the transform and quantization methods of MPEG.

What are some pre-processing enhancements ?

- Adaptive de-interlacing:

  This method maps interlaced video from a higher sampling rate (e.g 720 x 480) into a lower rate, progressive format (352 x 240). The most basic algorithm measures the variance between two fields, and if the variance is small enough, uses an average of both fields to form a frame macroblock. Otherwise, a field area from one field (of the same parity) is selected. More clever algorithms are much more complex than this, and may involve median filtering, and multirate/ multidimensional tools.

- Pre-anti-aliasing and Pre-blockiness reduction:

  A common method in still image coding is to pre-smooth the image before compression encoding. For example, if pre-analysis of a frame indicates that serious artifacts will arise if the picture were to be coded in the current condition, a pre-anti-aliasing filter can be applied. This can be as simple as having a smoothing severity proportional to the image activity. The pre-filter can be global (same smoothing factor for whole image) or locally adaptive. More complex methods will use multirate/multidimensional tools again.

The basic idea of multidimensional/multirate pre-processing is to apply source video whose resolution (sampling density) is greater than the target source and reconstruction sample rates. This follows the basic principles of oversampling, as found in A/D converters.

Most detail is contained in the lower harmonics anyway. Sharp-cut off filters are not widely practiced, so the *320 x 480 potential* of VHS is never truly realized.

Why use advanced pre-filtering techniques ?

Think of the DCT and quantizer as an A/D converter. Think of the pre-filter as the required anti-alias prefilter found before every A/D. The big difference of course is that the DCT quantizer assigns a varying number of bits per sample (transform coefficient).

Judging on the normalized activity measured in the pre-analysis stage of video encoding, and the target buffer size status, you have a fairly good idea of how many bits can be spared for the target macroblock, for instance.

Other pre-filtering techniques mostly take into account: texture patterns, masking, edges, and motion activity. Many additional advanced techniques can be applied at different immediate layers of video encoding (picture, slice, macroblock, block, etc.).

What are some advanced encoding methods ?

- Quantizer feedback [Thomson patent]

- Horizontal variance

- Motion vector cost:

  this is true for any syntax elements, really. Signalling a macroblock quantization factor or a large motion vector differential can cost more than making up the difference with extra quantized DFD (prediction error) bits. The optimum can be found with, for example, a Lagrangian process. In summary, any compression system with side information, there is a optimum point between signalling overhead (e.g. prediction) and prediction error.

- Liberal Interpretations of the Forward DCT

Borrowing from the concept that the DCT is simply a filter bank, a technique that seems to be gaining popularity is basis vector shaping. Usually this is combined with the quantization stage since the two are tied closely together in a rate-distortion sense. The idea is to use the basis vector shaping as a cheap alternative to pre-filtering by combining the more desiderable data adaptive properties of pre-filtering/ pre-processing into the transformation process... yet still reconstruct a picture in the decoder using the standard IDCT that looks reasonably like the source. Some more clever schemes will apply windowing. [Warning: watch out for eigenimage/basis vector orthogonality.]

- Frequency-domain enhancements:

  Enhancements are applied after the DCT (and possibly quantization) stage to the transform coefficients. This borrows from the concept: if you don't like the (quantized) transformed results, simply reshape them into something you do like.

- Temporal spreading of quantization error:

  This method is similar to the original intent behind color subcarrier phase alternation by field in the NTSC analog TV standard: for stationary areas, noise does not *hang* in one location, but *dances* about the image over time to give a more uniform effect. Distribution makes it more difficult for the eye to *catch on* to trouble spots (due to the latent temporal response curve of human vision). Simple encoder models tend to do this naturally but will not solve all situations.

- Look-ahead and adaptive frame cycle structures:

  Scene changes

- Post-processing

  (non-linear) Interpolation methods (Wu-Gersho) Convex hull projections Some ICASSP '93 papers, etc.

- Conformance vs. post-processing:

  Post-processing makes judging decoder output for conformance testing near impossible.

It is easy to spot encoders that do not employ any advanced encoding techniques: reconstructed video usually contains ringing around edges, color bleeding, and lots of noise.

What is MPEG-2 AUDIO ?

MPEG is developing the MPEG-2 Audio Standard for low bitrate coding of multichannel audio. MPEG-2 Audio coding will supply up to five full bandwidth channels (left, right, center, and two surround channels), plus an additional low frequency enhancement channel, and/or up to seven commentary/multilingual channels. The MPEG-2 Audio Standard will also extend the stereo and mono coding of the MPEG-1 Audio Standard (ISO/IEC IS 11172-3) to half sampling-rates (16 kHz, 22.05 kHz, and 24 kHz), for improved quality for bitrates at or below 64 kbits/s, per channel.

MPEG produced an updated version of the MPEG-2 Audio Working Draft, and is on track for achieving a Committee Draft specification by the November MPEG meeting.

The MPEG-2 Audio multichannel coding Standard will provide backward-compatibility with the existing MPEG-1 Audio Standard (ISO/IEC IS 11172-3). Together with ITU-RS, MPEG is organizing formal subjective testing of the proposed MPEG-2 multichannel audio codecs and up to three non backward compatible (NBC) codecs. The NBC codecs are included in order to determine whether an NBC mode should be introduced as an addendum to the standard. If the results show clear evidence that an NBC mode improves the performance, a formal call for NBC proposals will be issued by MPEG, with a view to incorporate these features in the audio syntax.

MPEG-2 audio attempts to maintain as much compatibility with MPEG-1 audio syntax as possible, while adding discrete surround-sound channels to the original MPEG-1 limit of 2 channels (Left, Right or matrix center and difference). The main channels (Left, Right) in MPEG-2 audio will remain backwards compatible, whereas new coding methods and syntax will be used for the surround channels.

A total of 5.1 channels are included that consist of the two main channels (L,R), two side/rear, center, and a 100 Hz special effects channel (hence the ".1" in "5.1").

At this time, non-backwards compatible (NBC) schemes are being considered as an ammedment to the MPEG-2 audio standard. One such popular system is Dolby AC-3.

What is MPEG-2 SYSTEMS ?

MPEG is developing the MPEG-2 Systems Standard to specify coding formats for multiplexing audio, video, and other data into a form suitable for transmission or storage. There are two data stream formats defined: the Transport Stream, which can carry multiple programs simultaneously, and which is optimized for use in applications where data loss may be likely, and the Program stream, which is optimized for multimedia applications, for performing systems processing in software, and for MPEG-1 compatibility.

Both streams are designed to support a large number of known and anticipated applications, and they retain a significant amount of flexibility such as may be required for such applications, while providing interoperability between different device implementations. The Transport Stream is well suited for transmission of digital television and video telephony over fiber, satellite, cable, ISDN, ATM, and other networks, and also for storage on digital video tape and other devices. It is expected to find widespread use for such applications in the very near future.

The Program Stream is similar to the MPEG-1 Systems standard (ISO/IEC 11172-1). It includes extensions to support new and future applications. Both the Transport Stream and Program Stream are built on a common Packetized Elementary Stream packet structure, facilitating common video and audio decoder implementations and stream type conversions. This is well-suited for use over a wide variety of networks with ATM/AAL and alternative transports. In New York, MPEG completed definitions of the features, syntax, and semantics of the Transport and Program Streams, enabling product designers to proceed. Among other items, the Transport Stream packet length was fixed at 188 bytes, including the 4-byte header. This length is suited for use with ATM networks, as well as a wide variety of other transmission and storage systems.

What about the Grand Alliance ?

The Grand Alliance was formed in May 1993 by seven organizations (AT&T Corp., General Instrument Corp. (GI), Massachusetts Institute of Technology (MIT), Philips Consumer Electronics, David Sarnoff research center, Thomson, Zenith Electronics Corp.) to evaluate technologies and to decide on key elements that will be at the heart of the *best of the best* HDTV system.

The video compression and transport technologies selected by the Grand Alliance are based on the MPEG-2 standards. The scanning formats selected are focused primarily on computer-friendly progressive scanning, while offering and interlaced mode important to some broadcasters.

They agreed to use the MPEG-2 Video and Systems syntax, including B-pictures. Both interlaced (1440 x 960 x 30 Hz) and progressive (1280 x 720 x 60 Hz) modes will be supported. The Alliance then had to settled upon a modulation (QAM or VSB), convolution (MS or Viterbi), and error correction (RSPC, RSFC) specification.

Laboratory tests in early 1993 showed better performance for a variant of VSB modulation and broadcast and cable carriage of digital HDTV signals via 8-VSB and 16-VSB modulation were tested under field conditions in Charlotte, North Carolina USA.

The audio technology selected is a six-channel, compact-disc-quality digital surround sound system. The last major technical decision, the broadcast and cable transmission subsystem, is expected in early 1994 following testing of competing technologies.

The Grand Alliance, now called ATSC (Advanced Television Systems Committee), suggests positions to the Department of State for their use in international standards organizations. ATSC proposes standards to the Federal Communications Commission.

On April 12, 1995, ATSC Members approved the Digital Television Standard for HDTV Transmission.

MPEG-2 references

The Institution of Electrical Engineers organised a one day colloquium in London entitled *MPEG-2 - what it is and what it isn't* in January 1995. The digest from the colloquium (Digest No: 1995/012) includes the following eight papers:

- *MPEG2 - Where did it come from and what is it?*, O.J.Morris (Philips)
- *MPEG2 - Video compression tutorial*, P.N.Tudor (BBC)
- *The ISO/MPEG audio musicam family*, Rault, Dehery, Lever (CCETT)
- *MPEG2 - A tutorial introduction to the systems layer*, P.A.Sarginson (BBC)
- *MPEG2 over ATM*, M.Nilsson (BT)
- *MPEG2 for DVB and cable*, G.M.Drury (NTL)
- *Switching MPEG2*, S.Defrance (Thomson)
- *Application of MPEG2 in the receiver*, W.Fletcher & P.Ardron (Sony)

Copies of the digest are available from:
The Institution of Electrical Engineers (Publication Sales Dept),
Michael Faraday House
Six Hills Way
Stevenage, Herts. SG1 2AY
UK
Tel: +44 1438 313311
Fax: +44 1438 742792

MPEG-2 sample bitstreams

There is an 8mm tape that contains lots of MPEG-2 video bitstreams, designed to test the conformance of video decoders. This include bitstreams with various types of scalability, as well as the corresponding reconstructed frames. This tape is too big (1 gig) to be made available by ftp.

This tape was intended to be used only by members of the MPEG commitee, but you may be able to obtain a copy of it by contacting the following people who have volonteered to distribute the tape:

```
North America (west coast):  Bill Helms (helms@divi.com)
North America (east coast):  S.J. Huang (sjhuang@dvs.com)
Europe:                      Jan De Lameillieure (Jan.DeLameillieure@barco.com)
Asia:                        Katsumi Tahara (tahara@av.crl.sony.co.jp)
```

This 8mm exabyte tape (tar format) is entitled *MPEG2 Video Bitstreams Lausanne'95*.

Another source is the *Centro di Studio per la Televisione* (CSTV) which also archives many MPEG-2 sample bitstreams.