MPEG-2 Technical (and sometimes political) Frequently Asked Questions (FAQ) list.

Draft 3.3 (May 10, 1994)

36. What happened to MPEG-3?
37. What is MPEG-4?
38. What are the scaleable modes of MPEG-2?
39. Why MPEG-2? Wasn't MPEG-1 enough?
40. What did MPEG-2 add to MPEG-1 in terms of syntax/algorithms ?
41. How do MPEG and JPEG differ?
42. How do MPEG and H.261 differ?
43. Is H.261 the de facto teleconferencing standard?
44. What is the TM rate control and adaptive quantization technique ?
45. How does the TM work?
46. What is a good motion estimation method, then?
47. Is exhaustive search "optimal" ?
48. What are some advanced encoding methods?
49. Is so-and-so really MPEG compliant ?
50. What are the tell-tale MPEG artifacts?
51. Where are the weak points of MPEG video ?
52. What are some myths about MPEG?
53. What is the color space of MPEG?
54. Don't you mean 4:1:1 ?
55. Why did MPEG choose 4:2:0 ? Isn't 4:2:2 the standard for TV?
56. What is the precision of MPEG samples?
57. What is all the fuss with cositing of chroma components?
58. How would you explain MPEG to the data compression expert?
59. How does MPEG video really compare to TV, VHS, laserdisc ?
60. What are the typical MPEG-2 bitrates and picture quality?
61. At what bitrates is MPEG-2 video optimal?
62. Why does film perform so well with MPEG ?
63. What is the best compression ratio for MPEG ?
64. Can MPEG be used to code still frames?
65. Is there an MPEG file format?
66. What are some pre-processing enhancements ?
67. Why use these "advanced" pre-filtering techniques?
68. What about post-processing enhancements?
69. Can motion vectors be used to measure object velocity?
70. How do you code interlaced video with MPEG-1 syntax?
71. Is MPEG patented?
72. How many cable box alliances are there?
73. Will there be an MPEG video tape format?
74. Where will be see MPEG in everyday life?
75. What is the best compression ratio for MPEG ?
76. Is there a MPEG CD-ROM format?

1. MPEG is a DCT based scheme, right?

The DCT and Huffman algorithms receive the most press coverage (e.g. _MPEG is a DCT based scheme with Huffman coding_), but are in fact fairly insignificant. The variety of coding modes signaled to the decoder as context-dependent side information are chiefly responsible for the efficiency of the MPEG syntax.

2. What does the MPEG video syntax feature that codes video efficiently?

A. Here are some of the statistical conditions and their syntax counterparts.

Occlusion: forward, backwards, or bi-directional temporal prediction in B pictures.

Smooth optical flow fields: variable length coding of 1-D prediction errors for motion vectors.

Spatial correlation beyond 8x8 sample block boundaries: 1-D prediction of DC coefficients in consecutive group intra-coded macroblocks.

High temporal correlation: variable on/off coding of prediction error at the macroblock (no-coding) or individual block (coded block pattern) level.

Temporal de-correlation: forward, backwards, or bidirectional prediction.

Content dependent quality: locally adaptive quantization

Temporal prediction accuracy: "half-pel" sample accuracy.

High locally correlated signal refresh pictures (I picture) and prediction errors: DCT

Subjective coding: location-dependent quantization of DCT coefficients.

3. What does the syntax provide for error robustness?

1. Byte-aligned start codes in the coded bitstream.
2. End of block codes in coded blocks.
3. Slices.
4. slice_vertical_position embedded as sub-field within slice start codes.
5. slices commencing at regular locations in picture (MPEG-2)

4. What is the significance of each layer in MPEG video ?

Sequence:

Set of pictures sharing same sampling dimensions, bit rate, chromaticy (MPEG-1), quantization matrices (MPEG-1 only).

Group of Pictures:

Random access point giving SMPTE time code within sequence.
Guaranteed to start with an I picture.

Picture:
Samples of a common plane -- "captured" from the same time instant.

Slice:
Error resynchronization unit of macroblocks.
At the commencement of a slice, all inter-macroblock coding dependencies are reset. Likewise, all macroblocks within a common slice can be dependently coded.

Macroblock:
Least common multiple of Y, Cb, Cr 8x8 blocks in 4:2:0 sampling structure.
For MPEG-1, the smallest granularity of temporal prediction.

Block:
Smallest granularity of spatial decorrelation.

5. How does the syntax facilitate parallelism?

A. For MPEG-1, slices may consist of an arbitrary number of macroblocks. The coded bitstream must first be mapped into fixed-length elements before true parallelism in a decoder application can be exploited. Further, since macroblocks have coding dependencies on previous macroblocks within the same slice, the data hierarchy must be pre-processed down to the layer of DC DCT coefficients. After this, blocks may be independently inverse transformed and quantized, temporally predicted, and reconstructed to buffer memory. Parallelism is usually more of a concern for encoders. Macroblock motion estimation and some rate control stages can be processed independently. An encoder also has the freedom to choose the slice structure.

6. I hear the encoder is not part of the standard?

A. The encoder rests just outside the normative scope of the standard, as long as the bitstreams it produces are compliant. The decoder, however, is almost deterministic: a given bitstream should reconstruct to a unique set of pictures. Statistically speaking, an occasional error of a Least Significant Bit is permitted as a result of the fact that the IDCT function is the only non-normative stage in the decoder (the designer is free to choose among many DCT algorithms and implementations). The IEEE 1180 test referenced in Annex A of the MPEG-1 and MPEG-2 specifications spells out the statistical mismatch tolerance between the Reference IDCT, which uses 64-bit floating point accuracy, and the Test IDCT.

7. Are some encoders better than others?

A. Yes. For example, the range over which a compensated prediction macroblock is searched for has a great influence over final picture quality. At a certain point a very large range can actually become detrimental (it may encourage large differential motion vectors). Practical ranges are usually between +/- 15 and +/- 32. As the range doubles, for instance, the search area quadruples.

8. Can MPEG-1 encode higher sample rates than 352 x 240 x 30 Hz ?

A. Yes. The MPEG-1 syntax permits sampling dimensions as high as 4095 x 4095 x 60 frames per second. The MPEG most people think of as "MPEG-1" is actually a kind of subset known as Constrained Parameters bitstream (CPB).

9. What are Constrained Parameters Bitstreams (CPB) for video?

A. MPEG-1 CPB are a limited set of sampling and bitrate parameters designed to normalize decoder computational complexity, buffer size, and memory bandwidth while still addressing the widest possible range of applications. The parameter limits were intentionally designed so that a decoder implementation would need only 4 Megabits of DRAM.

Parameter	Limit
pixels/line	704
lines/picture	480 or 576
pixels*lines	352*240 or 352*288
picture rate	30 Hz
bit rate	1.862million bits/sec
buffer size	40 Kilobytes (327,680 bits)

The sampling limits of CPB are bounded at the ever popular SIF rate: 396 macroblocks (101,376 pixels) per picture if the picture rate is less than or equal to 25 Hz, and 330 macroblocks (84,480 pixels) per picture if the picture rate is 30 Hz. The MPEG nomenclature loosely defines a "pixel" or "pel" as a unit vector containing a complete luminance sample and one fractional (0.25 in 4:2:0 format) sample from each of the two chrominance (Cb and Cr) channels. Thus, the corresponding bandwidth figure can be computed as:

352 samples/line x 240 lines/picture x 30 pictures/sec x 1.5 samples/pixel

or 3.8 Ms/s (million samples/sec) including chroma, but not including blanking intervals. Since most decoders are capable of sustaining

VLC decoding at a faster rate than 1.8 Mbit/sec, the coded video bitrate has become the most often waived parameter of CPB. An encoder which intelligently employs the syntax tools should achieve SIF quality saturation at about 2 Mbit/sec, whereas an encoder producing streams containing only I (Intra) pictures might require as much as 4 Mbit/sec to achieve the same video quality.

10. Why is Constrained Parameters so important?

A. It is an optimum point that allows (just barely) cost effective VLSI implementations in 1992 technology (0.8 microns). It also implies a nominal guarantee of interoperability for decoders and encoders. Since CPB is a canonical conformance point, MPEG devices which are not capable of meeting SIF rates are usually not considered to be true MPEG.

11. Who uses constrained parameters bitstreams?

A. Applications which are focused on CPB are Compact Disc (White Book or CD-I) and computer video applications. Set-top TV decoders fall into a higher sampling rate category known as _CCIR 601_ or _Broadcast rate._

12. Are there ways of circumventing constrained parameters bitstreams for SIF class applications and decoders ?

A. Yes, some. Remember that CPB limits pictures by macroblock count. 416 x 240 x 24 Hz sampling rates are still within the constraints, but this would only be of benefit in NTSC (240 lines/field) displays. Deviating from 352 samples/line could throw off many decoder implementations which possess limited horizontal sample rate conversion abilities. Some decoders do in fact include a few rate conversion modes, with a filter usually implemented via binary taps (shifts and adds). Likewise, the target sample rates are usually limited or ratios (e.g. 640, 540, 480 pixels/line, etc.). Future MPEG decoders will likely include on-chip arbitrary sample rate converters, perhaps capable of operating in the vertical direction (although there is little need of this in applications using standard TV monitors, with the possible exception of windowing in cable box graphical user interfaces).

13. Are there any other conformance points like CPB for MPEG-1?
A. Undocumented ones, yes. A second generation of decoder chips emerged on the market about 1 year after the first wave of SIF-class decoders. Both LSI Logic and SGS-Thomson introduced CCIR 601 class MPEG-1 decodersto fill in the gap between canonical MPEG-1 and the emergence of MPEG-2. Under non-disclosure agreement, C-Cube had the CL-950.

14. What frame rates are permitted in MPEG?
A. A limited set is available for the choosing in MPEG-1, although "tricks"

could be played with Systems-layer Time Stamps to convey non-standard rates. The set is: 23.976 Hz (3-2 pulldown NTSC), 24 Hz (Film), 25 Hz (PAL/SECAM or 625/60 video), 29.97 (NTSC), 30 Hz (drop-frame NTSC or component 525/60), 50 Hz (double-rate PAL), 59.97 Hz (double rate NTSC), and 60 Hz (double-rate drop-frame NTSC/component 525/60 video).

15. Special prediction switches for MPEG-2

```
              MPEG-2 sequence
             /         \
  progressive            interlaced sequence
  sequence              /                    \
                Field picture            Frame picture
                                     /             \
                           Frame or field pred.     Frame MB prediction only
                            /         \
                       Field dct      Frame dct
```

16. What is MPEG-2 Video Main Profile and Main Level?

A. MPEG-2 Video Main Profile and 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 place limits on coding 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.MP@ML is the conformance point for most cable and satellite systems.

Profiles
======
Simple: I and P pictures only. 4:2:0 sampling ratio. 8,9, or 10 bits DC precision.
Main: I, P, and B pictures. Dual Prime with no B-pictures only. 4:2:0 sampling ratio. 8, 9, or 10 bits sample precision.
SNR profile:
Spatial profile:
High: 8,9,10, or 11 bits sample precision. 4:2:2 and 4:4:4 sampling ratio.

Level
====
Simple: SIF video rate (3.041280 Mhz), 4 Mbit/sec, 0.489472 Mbit VBV buffer, 64 vertical in frame, 32Vertical in field, 1:7 fcode hor.

Main: CCIR 601 video rate (10.368 Mhz), 15 Mbit/sec, 1.835008 Mbit VBV buffer, 128 V in frame, 64 V in field, 1:8 f_code Hor.

High 1440: 1440 x 1152 x 30 Hz (47.0016 Mhz), 60 Mbit/sec. 7.340032 Mbit VBV buffer, 128 V in Fe, 1:9 fcode H.

High: 1920 x 1152 x 30 Hz (62.6688 Mhz), 80 Mbit/sec. 9.787392 Mbit VBV buffer.
1:9 fcode H

17. Does anybody actually use the scalability modes?

A. At this time, scalability has found itself a limited number of applications, although research is definitely underway for its use in HDTV. Experiments have been demonstrated in Europe where, for example, PAL-rate video (720 x 576 x 25 fps) is embedded in the same stream as HDTV rate video (1440 x 1152 x 25 fps). The Nov. 1992 VADIS experiment divided the base layer (PAL) and enhancement into 4 and 16 Mbit/sec channels, respectively. The U.S. Grand Alliance favors HDTV simulcasting (separate NTSC analog and digital HDTV broadcasts). Temporal scalability is the pet scalability mode as the possible future solution for coding 60 Hz progressive sequences while maintaining backwards compatibility with early-wave equipment (e.g. 1920 x 1080 x 30 Hz displays) . To elaborate, the first wave receivers of the late 1990's would be limited to 62at 0 Hz interlaced/30 Hz progressive HDTV decoders. Essentially, 60 interlaced fields would be coded in a, for example, 16 Mbit/sec stream in 1996, and when VLSI processes shift another thousand or so angstroms down the wavelength scale, an 8 Mbit/sec enhancement layer containing the coded "high pass" between 60 Hz progressive and 60 Hz interlaced would be simulcasted or multiplexed. Several corporate mouths have been known to water at the mention of charging the quality conscious subscriber an extra fee for the enhancement layer.

18. What's the difference between Field and Frame pictures?
A. A frame-coded picture consists of samples from both even and odd fields. A
frame picture is coded in progressive order (an even line, then an odd line, etc.) and in the case of MPEG-2, may optionally switch between field and frame order on a macroblock basis. The Display Process, which is *almost* completely outside the scope of the MPEG specification, can chose to re-interlace the picture by displaying the odd and even lines at different times (16 milliseconds apart for 60 Hz displays). In fact, most pictures, regardless of whether they were coded as a Field or Frame, end up being displayed interlaced due to the fact that most TV sets are interlaced.

19. What do B-pictures buy you?

A. Since bi-directional macroblock predictions are an average of two macroblock areas, noise is reduced at low bit rates (like a 3-D filter, if you will). At nominal MPEG-1 video (352 x 240 x 30, 1.15 Mbit/sec) rates, it is said that B-frames improves SNR by as much as 2 dB. (0.5 dB gain is usually considered worth-while in MPEG). However, at higher bit rates, B-frames become less useful since they inherently do not contribute to the progressive refinement of an image sequence (i.e. not used as prediction by subsequent coded frames). Regardless, B-frames are still politically controversial.

B pictures are interpolative in two ways: 1. predictions in the bi-directional macroblocks are an average from block areas of two pictures 2. B pictures _fill in_ or interpolate the 3-D video signal over a 33 or 25 millisecond picture period without contributing to the overall signal quality beyond that immediate point in time. In other words, a B pictures, regardless of its internal make-up of macroblock types, has a life limited to its immediate self. As mentioned before, its energy does not propagate into other frames. In a sense, bits spent on B pictures are wasted.

20. Why do some people hate B-frames?

A. Computational complexity, bandwidth, delay, and picture buffer size are the four B-frame Pet Peeves. Computational complexity in the decoder is increased since a some macroblock modes require averaging between two macroblocks.

Worst case, memory bandwidth is increased an extra 15.2 MByte/s (4:2:0 601 rates, not including any half pel or page-mode overhead) for this extra prediction. An extra picture buffer is needed to store the future prediction reference (bi-directionality). Finally, extra delay is introduced in encoding since the frame used for backwards prediction needs to be transmitted to the decoder before the intermediate B-pictures can be decoded and displayed.

Cable television (e.g. -- more like i.e.-- General Instruments) have been particularly adverse to B-frames since, for CCIR 601 rate video, the extra picture buffer pushes the decoder DRAM memory requirements past the magic 8-Mbit (1 Mbyte) threshold into the evil realm of 16 Mbits (2 Mbyte).... although 8-Mbits is fine for 352 x 480 B picture sequence. However, cable often forgets that DRAM does not come in convenient high-volume (low cost) 8-Mbit packages as does the friendly 4-Mbit and 16-Mbit. In a few years, the cost difference between 16 Mbit and 8 Mbit will become insignificant compared to the bandwidth savings gain through higher compression. For the time being, some cable boxes will start with 8-Mbit and allow future drop-in upgrades to the full 16-Mbit.

21. Why was the 16x16 area chosen?

A. The 16x16 area corresponds to the Least Common Multiple (LCM) of 8x8 blocks, given the normative 4:2:0 chroma ratio. Starting with medium size images, the 16x16 area provides a good balance between side information overhead & complexity and motion compensated prediction accuracy. In gist, 16x16 seemed like a good trade-off.

22. Why was the 8x8 DCT size chosen?
A. Experiments showed little improvements with larger sizes vs. the increased complexity. A fast DCT algorithm will require roughly double the arithmetic operations per sample when the transform point size is doubled. Naturally, the best compaction efficiency has been demonstrated using
locally adaptive block sizes (e.g. 16x16, 16x8, 8x8, 8x4, and 4x4) [See Baker and Sullivan]. Naturally, this introduces additional side information overhead and forces the decoder to implement programmable or hardwired recursive DCT algorithms. If the DCT size becomes too large, then more edges (local discontinuities) and the like become absorbed into the transform block, resulting in wider propagation of Gibbs (ringing) and other phenomena. Finally, with larger transform sizes, the DC term is even more critically sensitive to quantization noise.

23. What is motion compensated prediction, and why is it a pain?

A. MCP in the decoder can be thought of as having four stages:

1. Motion vector computation
2. Prediction retrieval
    various predictions are 16x16, 16x8, 8x4, 8x8 plus any half-pel overhead (e.g. 17x16, 17x17, etc).
3. Filtering
    3.1 Forming half-pel predictions through bi-linear interpolation.
    3.2 Averaging two predictions together (B macroblocks, Dual Prime)
4. Combination and ordering
    4.1 combining 1 or 2 predictions from stage three into upper and lower halves (16 x 8, field in frame)
    4.2 interleaving or grouping together odd and even lines in frame picture predictions.

The final, combined prediction is always a 16x16 block of luminance and 8x8 block of chrominance, just like we experience in MPEG-1.

A single motion vector can be associated with each source, hence a macroblock can have as many as 4 motion vectors.

24. What are the various prediction modes in MPEG-2?

24.1 Frame:
Predictions are formed from a 16 x 16 pixel area in a previously reconstructed frame. Identical to MPEG-1. There can be only one source in forward or backward predicted macroblocks, and two sources in bi-directional macroblocks. The prediction frame itself may have been coded as either a frame or two fields, however once a frame is reconstructed, it is simply a frame as far as future predictions are concerned.

24.2 Field predictions in frame-coded pictures:

Separate predictions are formed for the top (8 lines from field 1)and bottom (8 lines from field 2) portions of the macroblock. A total of two motion vectors in forward or backward predictions, four in bi-directional.

24.3 Field predictions in field-coded pictures:

Predictions are formed from the two most recently decoded fields. Prediction sizes are 16x16, however the 16 lines have a corresponding projection onto a 16x32 pixel area of a frame. One motion vector for forward or backward predictions, and two for bi-directional.

24.4 16x8 predictions in field-coded pictures:

Like field macroblocks in frame-coded pictures, the upper and lower 8 lines in this macroblock mode can have different predictions (hence two motion vectors). This mode compensates for the reduced temporal prediction precision of field picture macroblocks (a result of the fact that fields inherently possess half the number of lines that frames do). The field prediction area projected onto a frame is restored to 16 lines. 2 motion vectors for backwards or forwards, 4 for bi-directional.

24.5 Dual Prime prediction in frame and field-coded pictures

Predictions for the current macroblock are formed from the average of two 16 x 8 line areas from the two most recently decoded fields. Dual Prime was devised as an alternative for B pictures in low delay applications, but still offers many of the signal
quality benefits of B-pictures. Dual Prime requires one less prediction picture buffer, but still retains the same instantaneous prediction bandwidth of a B picture system. As an alternative to coding separate motion vectors for each of the upper and lower 16x8 areas, a full motion vector is sent for the first area, and a +1, 0, or -1 differential vector (variable length coded) is specified for the second prediction area. A macroblock will have total of two full motion vectors and two differential vectors in frame-coded pictures. Due to the prediction bandwidth overhead, Main Profile restricts the use of Dual Prime prediction to P picture sequences only. High Profile permits use of Dual Prime in B pictures.

24.6 Field and frame organized macroblocks:

Originally intended as a cheaper means of achieving field-decorrelation in frame-coded pictures without the fussy overhead of separate field prediction estimates, the dct coefficients (quantized prediction error for a given macroblock) may be organized into either a field or frame pattern. Essentially this means that the prediction error for the combined 16x16 macroblock may be grouped into field or frame blocks. A bit in the macroblock header (dct_type) indicates whether the upper and lower portions of the macroblock are to be interleaved (frame organized) or remain separated (field organized).

25. How do you tell a MPEG-1 bitstream from a MPEG-2 bitstream?

A. All MPEG-2 bitstreams must contain specific extension headers that *immediately* follow MPEG-1 headers. At the highest layer, for example, the MPEG-1 style sequence_header() is followed by sequence_extension() exclusive to MPEG-2. Some extension headers are specific to MPEG-2 profiles. For example, sequence_scalable_extension() is not allowed in Main Profile bitstreams.

A simple program need only scan the coded bitstream for byte-aligned start codes to determine whether the stream is MPEG-1 or MPEG-2.

26. What is the reasoning behind MPEG syntax symbols?

A. Here are some of the Whys and Wherefores of MPEG symbols:

Start codes
These 32-bit byte-aligned codes provide a mechanism for cheaply searching coded bitstreams for commencement of various layers of video without having to actually parse variable-length codes or perform any decoder arithmetic. Start codes also provide a mechanism for resynchronization in the presence of bit errors.

Coded block pattern (CBP --not to be confused with Constrained Parameters!) When the frame prediction is particularly good, the displaced frame difference (DFD, or prediction error) tends to be small, often with entire block energy being reduced to zero after quantization. This usually happens only at low bit rates. Coded block patterns prevent the need for transmitting EOB symbols in those zero coded blocks.

DCT_coefficient_first
Each intra coded block has a DC coefficient. With coded block patterns

signaling all possible combinations of all-zero valued blocks, the dct_coef_first mechanism assigns a different meaning to the VLC codeword that would otherwise represent EOB as the first coefficient.

End of Block:
Saves unnecessary run-length codes. At optimal bitrates, there tends to be few AC coefficients concentrated in the early stages of the zig-zag vector. In MPEG-1, the 2-bit length of EOB implies that there is an average of only 3 or 4 non-zero AC coefficients per block. In MPEG-2 Intra (I) pictures, with a 4-bit EOB code, this number is between 9 and 16 coefficients. Since EOB is required for all coded blocks, its absence can signal that a syntax error has occurred in the bitstream.

Macroblock stuffing
A genuine pain for VLSI implementations, macroblock stuffing was introduced to maintain smoother, constant bitrate control in MPEG-1. However, with normalized complexity measures and buffer management performed a priori (pre-frame, pre-slice, and pre-macroblock) in the MPEG-2 encoder test model, the need for such localized smoothing evaporated. Stuffing can be achieved through virtually unlimited slice start code padding if required. A good rule of thumb: if you find yourself often using stuffing more than once per slice, you probably don't have a very good rate control algorithm. Anyway, macroblock stuffing is now illegal in MPEG-2, so don t start using it if you already haven t.

MPEG's modified Huffman VLC tables
The VLC tables in MPEG are not Huffman tables in the true sense of Huffman coding, but are more like the tables used in Group 3 fax. They are entropy constrained, that is, non-downloadable and optimized for a limited range of bit rates (sweet spots). With the exception of a few codewords, the larger tables were carried over from the H.261 standard of 1990. MPEG-2 added an "Intra table". Note that the dct_coefficient tables assume positive/negative coefficient pmf symmetry.

27. Why bother to research compressed video when there is a standard?
A. Despite the worldwide standard, many areas remain open for research: advanced encoding and pre-processing, motion estimation, macroblock decision models, rate control and buffer management in editing environments, etc. There's practically no end to it.

28. Where can I get a copy of the latest MPEG-2 draft?

A. Contact your national standards body (e.g. ANSI Sales in NYC for the U.S., British Standards Institute in the UK, etc.). A number of private organizations offer ISO documents.

29. What are the latest working drafts of MPEG-2 ?
A. MPEG-2 has reached voting document of the Draft International Standard for :

Information Technology -- Generic Coding of Moving Pictures and Associated Audio. Recommendation H.262, ISO/IEC Draft International Standard 13818-2. [produced March 25, 1994, not yet approved by voting process].

Audio is Part 1, Video Part 2, and Systems is Part 3. A committee draft for Conformance (Part 4) is expected in Novemeber 1994, as well as the Technical Report on Software Simulation (Part 5).

30. What is the latest version of the MPEG-1 documents?

A. Systems (ISO/IEC IS 11172-1), Video (ISO/IEC IS 11172-2), and Audio (ISO/IEC IS 11172-3) have reached the final document stage. Part 4, Conformance Testing, is currently DIS

31. What is the evolution of ISO standard documents?

A. In chronological order:

ISO/Committee notation	Author's notation
Problem (unofficial first stage)	Barroom Witticism
New work Item (NI)	Napkin Item
New Proposal (NP)	Need Permission
Working Draft (WD)	We're Drunk
Committee Draft (CD)	Calendar Deadlock
Draft International Standard (DIS)	Doesn't Include Substance
International Standard (IS)	Induced patent Statements

32. Where is a good introductory paper to MPEG?

A. Didier Le Gall, "MPEG: A Video Compression Standard for Multimedia Applications," Communications of the ACM, April 1991, Vol.34, No.4, pp. 47-58

33. What are some journals on related MPEG topics ?
A.

IEEE Transactions on Consumer Electronics
IEEE Transactions on Broadcasting

IEEE Transactions on Circuits and Systems for Video Technology
Advanced Electronic Imaging
Electronic Engineering Times (EE Times -- more tabloid coverage. Unfortunate columns by Richard Doherty)
IEEE Int'l Conference on Acoustics, Speech, and Signal Processing (ICASSP)
International Broadcasting Convention (IBC)
Society of Motion Pictures and Television Engineers (SMPTE)
SPIE conference on Visual Communications and Image Processing
SPIE conference on Video Compression for Personal Computers
IEEE Multimedia [first edition Spring 1994]

34. Is there a book on MPEG video?

A. Yes, there will be a book published sometime in 1994 by the same authors who brought you the JPEG book (Bill Pennebaker, Joan Mitchell). Didier Le Gall will be an additional co-author, and will insure digressions into, e.g. arithmetic coding aspects, be kept to a minimum :-)

35. Is it MPEG-2 (Arabic numbers) or MPEG-II (roman)?

A. Committee insiders most often use the Arabic notation with the hyphen, e.g. MPEG-2. Only the most retentive use the official designation: Phase 2. In fact, M.P.E.G. itself is a nickname. The official title is: ISO/IEC JTC1 SC29 WG11. The militaristic lingo has so far managed to keep the enemy (DVI) confused and out of the picture.

ISO: International Organization for Standardization
IEC: International Electrotechnical Commission
JTC1: Joint Technical Committee 1
SC29: Sub-committee 29
WG11: Work Group 11 (moving pictures with... uh, audio)

36. What happened to MPEG-3?

A. MPEG-3 was to have targeted HDTV applications with sampling dimensions up to 1920 x 1080 x 30 Hz and coded bitrates between 20 and 40 Mbit/sec. It was later discovered that with some (compatible) fine tuning, MPEG-2 and MPEG-1 syntax worked very well for HDTV rate video. The key is to maintain an optimal balance between sample rate and coded bit rate.

Also, the standardization window for HDTV was rapidly closing. Europe and the United States were on the brink of committing to analog-digital subnyquist hybrid algorithms (D-MAC, MUSE, et al). European all-digital projects such as HD-DIVINE and VADIS demonstrated better picture quality with respect to bandwidth using the MPEG syntax. In the United States, the

Sarnoff/NBC/Philips/Thomson HDTV consortium had used MPEG-1 syntax from the beginning of its all-digital proposal, and with the exception of motion artifacts (due to limited search range in the encoder), was deemed to have the best picture quality of all three digital proponents. HDTV is now part of the MPEG-2 High-1440 Level and High Level toolkit.

37. What is MPEG-4?
A. MPEG-4 targets the Very Low Bitrate applications defined loosely as having sampling dimensions up to 176 x 144 x 10 Hz and coded bit rates between 4800 and 64,000 bits/sec. This new standard would be used, for example, in low bit rate videophones over analog telephone lines.

This effort is in the very early stages. Morphology, fractals, model based, and anal retentive block transform coding are all in the offering. MPEG-4 is now in the application identification phase.

Scaleable modes of MPEG-2

38. What are the scaleable modes of MPEG-2?
A. Scaleable video is permitted only in the High Profiles.

Currently, there are four scaleable 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/or increased signal strength (i.e. higher Carrier-to-Noise ratio or lower Bit Error Rate) than the lower priority channel. 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 part of the ″graceful degradation_ concept. Breaking a video signal into two streams (base and enhancements) has a penalty, however. Usually less than 1.5 dB.

Another purpose of salability is complexity division. A standard TV set need only decode the 720 x 480 channel, thus requiring a less expensive decoder processor than a TV set wishing to display 1440 x 960. This is known as simulcasting.

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

Spatial Scalablity-- 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 (achieved 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 prediction 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 comparable or better picture quality with greater simplicity.

39. Why MPEG-2? Wasn't MPEG-1 enough?

A. 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 cooperation executed well enough 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.(or a particle or wave) ? The answer of course is that some blocks are one or the other at different times, depending on motion activity. The additional man years of experimentation and implementation between MPEG-1 and MPEG-2 improved the method of block-based transform coding.

40. What did MPEG-2 add to MPEG-1 in terms of syntax/algorithms ?
A. 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 scaleable modes [see scalability discussion]

Picture layer:
All MPEG-2 motion vectors are specified to a half-pel sample grid.

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

New scalar quantization matrices may be downloaded once per picture. In High profile, separate chrominance matrices now exist (Y and C no longer have to share)

Concealment motion vectors were added to I-pictures in order to increase robustness from bit errors. I pictures are the most critical and sensitive picture in a group of pictures.

A non-linear macroblock quantization factor providing a wider dynamic range, from 0.5 to 56, than the linear MPEG-1 (1 to 32) range. Both are sent as a 5-bit FLC side information in the macroblock and slice headers.

New Intra-VLC table for dct_coefficient_next (AC run-level events) that is a better match for the histogram of Intra-coded pictures. EOB is 4 bits. The old table, dct_coef_next, are reserved for use in non-intra pictures (P, B), although they new table can be used for Intra-coded macroblocks in P and B pictures as well.

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 an irregular 3:2 pulldown process (repeat_field_first flag)

Progressive and interlaced frame coding

Syntax to indicate 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 coded image. Vertical pan offset has 1/16th pixel accuracy.

Macroblock layer:
Macroblock stuffing is now illegal in MPEG-2 (hurray!!). If stuffing is really needed, the encoder can pad slice start codes.

Two organizations for macroblock coefficients (interlaced and progressive) signaled by dct_type flag.

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 the summed coefficients. MPEG-2 mismatch control is performed on the transform coefficients, whereas in MPEG-1, it is applied to the quantized transform coefficients.

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.

41. How do MPEG and JPEG differ?

A. The most fundamental difference is MPEG's use of block-based motion compensated prediction (MCP)---a method falling into the general category of temporal DPCM.

The second most fundamental difference is in the target application. JPEG adopts a general purpose philosophy: independence from color space (up to 255 components per frame) and quantization tables for each component. Extended modes in JPEG include two sample precision (8 and 12 bit sample accuracy), combinations of frequency progressive, spatial hierarchically progressive, and amplitude (point transform) progressive scanning modes. Further color independence is made possible thanks to downloadable Huffman tables (up to one for each component.)

Since MPEG is targeted for a set of specific applications, there is only one color space (4:2:0 YCbCr), one sample precision (8 bits), and one scanning mode (sequential). Luminance and chrominance share quantization and VLC tables. MPEG adds adaptive quantization at the macroblock (16 x 16 pixel area) layer. This permits both smoother bit rate control and more perceptually uniform quantization throughout the picture and image sequence. However, adaptive quantization is part of the Enhanced JPEG charter (ISO/IEC 10918-3) currently in verification stage. MPEG variable length coding tables are non-downloadable, and are therefore optimized for a limited range of compression ratios appropriate for the target applications.

The local spatial decorrelation methods in MPEG and JPEG are very similar. Picture data is block transform coded with the two-dimensional orthanormal 8x8 DCT, with asymmetric basis vectors about time (aka _DCT-II_). The resulting 63 AC transform coefficients are mapped in a zig-zag pattern (or alternative scan pattern in MPEG-2) to statistically increase the runs of zeros. Coefficients of the vector are then uniformly scalar quantized, run-length coded, and finally the run-length symbols are variable length coded using a canonical (JPEG) or modified Huffman (MPEG) scheme. Global frame redundancy is reduced by 1-D DPCM of the block DC coefficients, followed by quantization and variable length entropy coding of the quantized DC coefficient.

```
           MCP                    DCT                    ZZ
Q
  Frame -> 8x8 spatial block -> 8x8 frequency block -> Zig-zag scan ->

                  RLC                 VLC
       quanitzation -> run-length coding -> variable length coding.
```

The similarities have made it possible for the development of hard-wired silicon that can code both standards. Even some highly microcoded architectures employing hardwired instruction primitives or functional blocks benefit from JPEG/MPEG similarities. There are many additional yet minor differences. They include:

1. In addition to the 8-bit mode, DCT and quantization precision

in MPEG has a 9-bit and 12-bit mode, respectively, exclusively in non-intra coded macroblocks. A 1-bit expansion takes place in the macroblock difference operation.

2. Mismatch control in MPEG-1 forces quantized coefficients to become odd values (oddification). JPEG does not employ any mismatch mechanism.

3. JPEG run-length coding produces run-size tokens (run of zeros, non-zero coefficient magnitude) whereas MPEG produces fully concatenated run-level tokens that do not require magnitude differential bits.

4. DC values in MPEG-1 are limited to 8-bit precision (a constant stepsize of 8), whereas JPEG DC precision can occupy all possible 11-bits. MPEG-2, however, re-introduced extra DC precision critical even at high compression ratios.

Difference between MPEG and H.261

42. How do MPEG and H.261 differ?

A. H.261, also known as Px64, was targeted for teleconferencing applications where motion is naturally more limited. Motion vectors are restricted to a range of +/- 15 pixel unit displacements. Prediction accuracy is reduced since H.261 motion vectors are specified to only integer-pel accuracy. Other quality syntactic differences include: no B-pictures, inferior mismatch control.

43. Is H.261 the de facto teleconferencing standard?

A. Not exactly. To date, about seventy percent of the industrial teleconferencing hardware market is controlled by PictureTel of Mass. The second largest market controller is Compression Labs of Silicon Valley. PictureTel hardware includes compatibility with H.261 as a lowest common denominator, but when in communication with other PictureTel hardware, it can switch to a mode superior at low bit rates (less than 300kbits/sec). In fact, over 2/3 of all teleconferencing is done at two-times switched 56 channel (~P = 2) bandwidth. ISDN is still expensive. In each direction, video and audio are coded at an aggregate rate of 112 kbits/sec (2*56 kbits/sec). The PictureTel proprietary compression algorithm is acknowledged to be a combination of spatial pyramid, lattice vector quantizer, and an unidentified entropy coding method. Motion compensation is considerably more refined and sophisticated than the 16x16 integer-pel block method specified in H.261.

The Compression Labs proprietary algorithm also offers significant improvement over H.261 when linked to other CLI hardware. Local decorrelation is based on a DCT-VQ hybrid.

Currently, ITU-TS (International Telecommunications Union--teleconferencing Sector), formerly CCITT, is quietly defining an improvement to H.261 with the participation of industry vendors.

Rate control

44. What is the TM rate control and adaptive quantization technique ?

A. The Test model (MPEG-2) and Simulation Model (MPEG-1) were not, by any stretch of the imagination, meant to epitomize state-of-the art encoding quality. They were, however, designed to exercise the syntax, verify proposals, and test the *relative* compression performance of proposals in a timely manner that could be duplicated by co-experimenters. Without simplicity, there would have been no doubt endless debates over model interpretation. Regardless of all else, more advanced techniques would probably trespass into proprietary territory.

The final test model for MPEG-2 is TM version 5b, aka TM version 6. The final MPEG-1 simulation model is version 3. The MPEG-2 TM rate control method offers a dramatic improvement over the SM method. TM adds more accurate estimation of macroblock complexity through use of limited a priori information. Macroblock quantization adjustments are computed on a macroblock basis, instead of once-per-slice.

45. How does the TM work?
A. Rate control and adaptive quantization are divided into three steps:

Step One:Bit Allocation

In Complexity Estimation, the global complexity measures assign relative weights to each picture type (I,P,B). These weights (Xi, Xp, Xb) are reflected by the typical coded frame size of I, P, and B pictures (see typical frame size discussion). I pictures are usually assigned the largest weight since they have the greatest stability factor in an image sequence. B pictures are assigned the smallest weight since B energy do not propagate into other pictures and are usually highly correlated with neighboring P and I pictures.

The bit target for a frame is based on the frame type, the remaining number of bits left in the Group of Pictures (GOP) allocation, and the immediate statistical history of previously coded pictures.

Step Two: 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. If the virtual buffer begins to overflow, the macroblock quantization step size is increased, resulting in a smaller yield of coded bits in subsequent macroblocks. Likewise, if underflow begins, the step size is decreased.   The Test Model approximates that the target
picture has spatially uniform distribution of bits. This is a safe approximation since spatial activity and perceived quantization noise are almost inversely proportional. Of course, the user is free to design a custom distribution, perhaps targeting more bits in areas that contain text, for example.

Step Three: Adaptive Quantization

The final step modulates the macroblock quantization step size obtained in Step 2 by a local activity measure. The activity measure itself is normalized against the most recently coded picture of the same type (I, P, or B). The activity for a macroblock is chosen as the minimum among the four 8x8 block luminance variances. Choosing the minimum block is part of the concept that a macroblock is no better than the block of highest visible distortion (weakest link in the chain).

46. What is a good motion estimation method, then?

A. When shopping for motion vectors, the three basic characteristics are: Search range, search pattern, and matching criteria. Search pattern has the greatest impact on finding the best vector. Hierarchical search patterns first find the best match between downsampled images of the reference and target pictures and then refine the vector through progressively higher resolutions. When compared to other fast methods, hierarchical patterns are less likely to be confused by extremely local distortion minimums as being a best match. Also note that _subsampled search_ and _hierarchical search_ are not synonymous.

Q. Is there a limit to the length of motion vectors?

The search area is unlimited, but the reconstructed motion vectors must not:

a. point beyond the picture boundaries   (1 <= MV_x <= luminancewidth - 16) and (1 <= MV_y <= luminanceheight - 16). The _- 16_ is due to the fact that the motion vector origin is the upper left hand corner of a macroblock)

b. In Constrained Parameters MPEG-1, the motion vector is limited to a range of [-64,+63.5] luminance samples with half-pel accuracy, and [-128,+127.5] with integer pel accuracy. Break the constrained parameters rules and your video sequence will not likely display on many hardware devices.

c. In MPEG-2 Video Main Profile at Main Level, the motion vectors are always on a half-pel co-ordinate grid, and the vertical range is restricted to [-64, +63.5], and the horizontal limit is [-256,+255.5].

d. in MPEG-1, the syntactic limit of the motion vector is [-1024,+1023] integer pel, horizontal and vertical.

e. in MPEG-2, the syntactic limit of the motion vector is [-2048,+2047.5] horizontal, [-1024,+1023.5] vertical.

47. Is exhaustive search "optimal" ?

A. Definitely not in the context of block-based MCP video. Since one motion vector represents the prediction of 256 pixels, divergent pixels within the macroblock are misrepresented by the "global" vector. This leads back to the general philosophy of block-based coding as an approximation technique. In their ICASSP'93 paper, Sullivan discusses ways in which block-based prediction schemes can solve part of this problem.

Exhaustive search may find blocks with the least distortion (displaced frame difference) but will not produce motion vectors with the lowest entropy.

48. What are some advanced encoding methods?

Quantizer feedback: determine the dependent quantization stepsize by modeling quantization error propagating over multiple pictures. [Uz/et al ICASSP _93, Ortega/Vetterli/et al ICASSP _93]

Smoothness constraint placed on local activity measures. immediate blocks outside target macroblock are considered when selecting macroblock quantization stepsize .[Thomson/Savitier patent]

Horizontal variance: measure variance between columns of pixels in addition to the traditional measure of variance along rows (lines) when making field/frame macroblock prediction decision.

DFD energy: examine DFD energy/variance when making Intra/Non-intra macroblock decision.

Activity measures: use total bits from a first-pass encoding of a picture or macroblock as a measure of the activity. Coded bits is a more accurate reflection of local complexity than variance. [Thomson/Savitier patent]

motion vector cost: this is true for any syntax elements, really. Signaling 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, some Lagrangian operator. In summary, any compression system with side information, there is a optimum point between signaling 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 desirable 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 a form of windowing. [Warning: watch out for eigenimage/basis vector orthoganality.]

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. Suppressing isolated small amplitudes is popular.

Temporal spreading of quantization error:
This method is similar to the original intent behind color subcarrier phase alternation by field in the NTSC, PAL, and SECAM analog TV standards: 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: analyze picture activity several pictures into the future, looking for scene changes or motion statistics.

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.

49. Is so-and-so really MPEG compliant ?

A. At the very least, there are two areas of conformance/compliance in MPEG: 1. Compliant bitstreams 2. compliant decoders. Technically speaking, video bitstreams consisting entirely of I-frames (such as those generated by Xing software) are syntactically compliant with the MPEG specification. The I-frame sequence is simply a subset of the full syntax. Compliant bitstreams must obey the range limits (e.g. motion vectors limited to +/-128, frame sizes, frame rates, etc.)and syntax rules (e.g. all slices must commence and terminate with a non-skipped macroblock, no gaps between slices, etc.).

Decoders, however, cannot escape true conformance. For example, a decoder that cannot decode P or B frames are *not* legal MPEG. Likewise, full arithmetic precision must be obeyed before any decoder can be called "MPEG compliant." The IDCT, inverse quantizer, and motion compensated predictor must meet the specification requirements... which are fairly rigid (e.g. no more than 1 least significant bit of error between reference and test decoders). Real-time conformance is more complicated to measure than arithmetic precision, but it is reasonable to expect that decoders that skip frames on reasonable bitstreams are not likely to be considered compliant.

Artifacts

50. What are the tell-tale MPEG artifacts?

A. If the encoder did its job properly, and the user specified a proper balance between sample rate and bitrate, there shouldn't be any visible artifacts. However, in sub-optimal systems, you can look for:

Gibbs phenomenon/Ringing/Aliasing (too few AC bits, not enough pre-processing)

Blockiness (not considering your neighbors before quantizing)

Posterization (too few DC bits)

Checkerboards (DCT eigenimages as a result of too few AC coefficients)
Colorbleeding (not considering color in encoder cost model, not subtracting color at edges of objects, etc.)

51. Where are the weak points of MPEG video ?
A.

Texture patterns (rapidly alternating lines)
sharp edges (especially text)

[installment 3]

52. What are some myths about MPEG?
A. There are a few major myths that I am aware of:

1. Block displacements: macroblock predictions are formed out of arbitrary 16x16 (or 16x8/16x16 in MPEG-2) areas from previously reconstructed pictures. Many people believe that the prediction macroblocks have boundaries that fall on interchange boundaries (pixel 0, 15, 31, 53... line 0, 15, 31, 53... etc.). In fact, motion vectors represent relative translations with respect to the target reconstruction macroblock coordinates. The motion vectors can point to half pixel coordinates, requiring that the prediction macroblock to be formed via bi-linear interpolation of pixels.

2. Displaced frame (macroblock) difference construction: the prediction error formed as the difference between the prediction macroblock and source macroblock is coded much like an Intra macroblock. The prediction may come from different locations (as in bi-directional prediction--or in MPEG-2--16x8, field-in-frame, and Dual Prime), but the DFD is always coded as a 16x16 unit.

3. Compression ratios

You hear 200:1 and 100:1 in the media. Utter rubbish. The true range is between 16:1 and 40:1. Spreading misinformation about compression ratios in public will catch the attention of the infamous _MPEG Police._ They say mild-mannered Michael Barnsley will snap, without warning, into violent rage if he doesn't get the upper bunk bed.

4. Picture coding types all consist of the same macroblocks

Macroblocks within I pictures are strictly intra-coded. Macroblocks within P pictures can be either predicted or intra-coded, and B pictures they can be bi-directional, forward, backward, or intra. Additional macroblock modes switches include: predicted with no motion compensation, modified macroblock quantization, coding of prediction error or not. The switches are concatenated into the macroblock_type side information and variable length coded in the macroblock header.

53. What is the color space of MPEG?

MPEG strictly specifies the YCbCr color space, not YUV or YIQ or YPbPr or YDrDb or any other color difference variations. Regardless of any bitstream parameters, MPEG-1 and MPEG-2 Video Main Profile specify 4:2:0

chroma ratio, where the color difference channels (Cb, Cr) have half the _resolution_ or sample grid density in both the horizontal and vertical direction
with respect to luminance.

MPEG-2 High Profile includes an option for 4:2:2 and 4:4:4 coding. Applications
for this are likely to be broadcasting and contribution equipment.

54. Don’t you mean 4:1:1 ?

A. No, here is a table of ratios:

CCIR 601 (60 Hz) image			Chroma sub-sampling factors	
format	Y	Cb, Cr	Vertical	Horizontal
4:4:4	720 x 480	720 x 480	none	none
4:2:2	720 x 480	360 x 480	none	2:1
4:2:0	720 x 480	360 x 240	2:1	2:1
4:1:1	720 x 480	720 x 120	none	4:1
4:1:0	720 x 480	180 x 120	4:1	4:1

3:2:2, 3:1:1, and 3:1:0 are less common variations.

55. Why did MPEG choose 4:2:0 ? Isn’t 4:2:2 the standard for TV?

A. At least three reasons I can think of:

1. 4:2:0 picture memory requirements are 33% less than the size of 4:2:2 pictures.
MPEG-1 decoder are able to snugly fit all 3 SIF pictures (1 reconstruction & display, 2 prediction) into 512 KBytes of buffer space. CCIR 601 is a tighter fit into 2 Mbytes.

2. The subjective difference between 4:2:0 and 4:2:2 is minimal, when considering consumer display equipment and distribution compression ratios.

3. Vertical decimation increases compression efficiency by reducing syntax overhead posed in an 8 block (4:2:0) macroblock structure.

4. You re compressing the hell out of the video signal, so what possible difference can the 0:0:2 high-pass make?

Interlacing and the 62 microsecond gap between successively scanned lines introduces some discontinuities, but most of this can be alleviated through pre-processing.

56. What is the precision of MPEG samples?

A. By definition, MPEG samples have no more and no less than 8-bits uniform sample precision (256 quantization levels). For luminance (which is unsigned) data, black corresponds to level 0, white is level 255. However, in CCIR recommendation 601 chromaticy, levels 0 through 14 and 236 through 255 are reserved for blanking signal excursions. MPEG currently has no such clipped excursion restrictions, although decoder might take care to insure active samples do not exceed these limits. With three color components per pixel, the total combination is roughly 16.8 million colors (i.e. 24-bits).

57. What is all the fuss with cositing of chroma components?

A. It is moderately important to properly co-site chroma samples, otherwise a sort of chroma shifting effect (exhibited as a _halo_) may result when the reconstructed video is displayed. In MPEG-1 video, the chroma samples are exactly centered between the 4 luminance samples (Fig 1.) To maintain compatibility with the CCIR 601 horizontal chroma locations and simplify implementation (eliminate need for phase shift), MPEG-2 chroma samples are arranged as per Fig.2.

```
 Y   Y   Y   Y           Y   Y   Y   Y        YC  Y   YC  Y
   C       C             C       C
 Y   Y   X   Y           Y   Y   Y   Y        YC  Y   YC  Y

 Y   Y   Y   Y           Y   Y   Y   Y        YC  Y   YC  Y
   C       C             C       C
 Y   Y   Y   Y           Y   Y   Y   Y        YC  Y   YC  Y

 Fig.1 MPEG-1             Fig.2  MPEG-2          Fig.3 MPEG-2 and
4:2:0 organization       4:2:0 organization        CCIR Rec. 601
                                               4:2:2 organization
```

MPEG for the data compression expert

58. How would you explain MPEG to the data compression expert?

A. MPEG video is a block-based video scheme.

59. How does MPEG video really compare to TV, VHS, laserdisc ?
A. 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 luminance lines compared to 240)...
but when taking into account (supposedly such things as) interfield magnetic tape crosstalk and the TV monitor Kell factor, the perceptual vertical advantage is not all that significant. VHS is prone to such inconveniences as timing errors (an annoyance addressed by 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 speed), opening up additional avenues for artifacts. In gist, MPEG-1 at its nominal parameters can match VHS's sexy low-pass-filtered look.

With careful coding schemes, 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 should be treated with bit rates more like 5 and 6 Mbit/sec, respectively. Laserdisc is a tough one to compare. Laserdisc's are encoded with composite video (NTSC or PAL). Manufacturers of laser disc players make claims of up to 425 TVL (or 567 samples/line) response. Thus it could be said the laserdisc has a 567 x 480 x 30 Hz "potential 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 with MPEG-1 at 1.15 Mbit/sec (SIF rates), especially for those areas of medium detail (low spatial activity) in the presence of uniform motion. This may be why some people say MPEG-1 video at 1.15 Mbit/sec looks almost as good as Laserdisc or Super VHS at times.

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

	Picture type			
	I	P	B	Average
MPEG-1 SIF @ 1.15 Mbit/sec	150,000	50,000	20,000	38,000
MPEG-2 601 @ 4.00 Mbit/sec	400,000	200,000	80,000	130,000

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, among differing source material, scene changes, and use of advanced encoder models... these numbers can be significantly different.

61. At what bitrates is MPEG-2 video optimal?
A. 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) (better)	2 Mbit/sec	Half horizontal 601. Looks almost NTSC broadcast quality, and is a good 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.

[these numbers subject to change at whim of MPEG Test subgroup]

62. Why does film perform so well with MPEG ?
A. 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 amenable to the transform and quantization methods of MPEG.

63. What is the best compression ratio for MPEG ?

A. The MPEG sweet spot is about 1.2 bits/pel Intra and .35 bits/pel inter. Experimentation has shown that intra frame coding with the familiar DCT-Quantization-Huffman hybrid algorithm achieves optimal performance at about an average of 1.2 bits/sample or about 6:1

compression ratio. Below this point, artifacts become noticeable.

64. Can MPEG be used to code still frames?

A. Yes. There are, of course, advantages and disadvantages to using MPEG over JPEG:

Disadvantages:

1. MPEG has only one color space
2. MPEG-1 and MPEG-2 Main Profile luma and chroma share quanitzation and VLC tables
3. MPEG-1 is syntactically limited to 4k x 4k images, and 16k x 16k for MPEG-2.

Advantages:

1. MPEG possesses adaptive quantization

2. With its limited still image syntax, MPEG averts any temptation to use unnecessary, expensive, and academic encoding methods that have little impact on the overall picture quality (you know who you are).

Philips' CD-I spec. has a requirement for a MPEG still frame mode, with double SIF image resolution. This is technically feasible mostly thanks to the fact that only one picture buffer is needed to decode a still image instead of three buffers.

65. Is there an MPEG file format?

A. Not exactly. The necessary signal elements that indicate image size, picture rate, aspect ratio, etc. are already contained within the sequence layer of the MPEG video stream. The Whitebook format for Karoke and CD-I movies specify a range of (time-division) multiplexing strategies for audio and video bitstreams. A directory format listing scenes and their locations on the disc is associated with the White Book specification.

66. 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 correlation between two immediate macroblock fields, and if the correlation is high enough, uses an average of both fields to form a frame macroblock. Otherwise, a field area from one field (usually 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 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 (i.e. below the sweet spot), 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 sequence) or locally adaptive. More complex methods will again use multirate/multidimensional methods.

One straightforward concept from multidimensional/multirate e-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.

These filters emphasize the fact that most information content is contained in the lower harmonics of a picture anyway. VHS is hardly considered to be a _sharp cut-off_ medium, tragically implying that "320 x 480 potential" of VHS is never truly realized.

67. Why use these "advanced" pre-filtering techniques?

A. Think of the DCT and quantizer as an A/D converter. Think of the DCT/Q 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 transform coefficient. Judging on the normalized activity measured in the pre-analysis stage of video encoding (assuming you even have a pre-analysis stage), and the target buffer size status, you have a fairly good idea of how many bits can be spared for the target macroblock, for example.

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.).

68. What about post-processing enhancements?

Some research has been carried out in this area. Non-linear interpolation methods have been published by Wu and Gersho (e.g. ICASSP _93), convex hull projections for MAP (Severinson, ICASSP _93), and others. Post-processing unfortunately defies the spirit of MPEG conformance. Decoders should produce similar reconstructions. Enhancements should ideally be done during the pre-processing and encoding stages.

69. Can motion vectors be used to measure object velocity?

A. Motion vector information cannot be reliably used as a means of determining object velocity unless the encoder model specifically set out to do so. First, encoder models that optimize picture quality generate vectors that typically minimize prediction error and, consequently, the vectors often do not represent true object translation. Standards converters that resample one frame rate to another (as in NTSC to PAL) use different methods (motion vector field estimation, edge detection, et al) that are not concerned with optimizing ratios such as SNR vs bitrate. Secondly, motion vectors are not transmitted for all macroblocks anyway.

70. How do you code interlaced video with MPEG-1 syntax?
A. Two methods can be applied to interlaced video that maintain syntactic compatibility with MPEG-1 (which was originally designed for progressive frames only). In the field concatenation method, the encoder model can carefully construct predictions and prediction errors that realize good compression but maintain field integrity (distinction between adjacent fields of opposite parity). Some pre-processing techniques can also be applied to the interlaced source video that would, e.g., lessen sharp vertical frequencies.

This technique is not efficient of course. On the other hand, if the original source was progressive (e.g. film), then it is more trivial to convert the interlaced source to a progressive format before encoding. (MPEG-2 would then only offer superior performance through greater DC block precision, non-linear mquant, intra VLC, etc.) Reconstructed frames are re-interlaced in the decoder Display process.

The second syntactically compatible method codes fields as separate pictures. This approach has been acknowledged not to work as well.

71. Is MPEG patented?
A. Yes and no. Many encoding methods are patented. Approximately 11 blocking patents, that is, patents that are general enough to be unavoidable in any implementation have been recently identified.

A patent pool is being formed within MPEG where a single royalty fee would be split among the 31 patent-holding companies.

72. How many cable box alliances are there?

A. Many. To start with:

Scientific Atlanta (SA), Kaledia, and Motorola:
SA will build the box, Motorola the chips, and Kaleida the O/S and user interface (using ScriptX of course).

Silicon Graphics (SGI), Scientific Atlanta, and Toshiba
For the Time Warner's Orlando trial, SGI will provide the RISC (MIPS R4000) and software, SA will do the box again, and Toshiba will provide the chips.

General Instruments (GI) and Microsoft:
GI will make the box and Intel will supply the special low-cost 386SL processor on which a 1MB flash EPROM executable core of Microsoft windows and DOS will run. Microsoft will develop the user interface.

Hewlett Packard (HP):
HP will manufacture and/or design low cost, open architecture set-top decoder boxes (not a part of the Eon wireless deal). The CPU will explicitly not use a 80x68 based processor.

CLI and Philips:
Compression Labs will provide the encoder technology and Philips will provide the decoder techology for an ADSL system whose transport structure will be put together by Broadband Technologies.

["These alliances subject to change at the whim of PR departments and market forces."]

73. Will there be an MPEG video tape format?

A. Not exactly. A consortium of international companies are co-developing a consumer digital video 6 millimeter wide, metal particle tape format. Due to the initial high cost of MPEG encoders, a JPEG-like compression method will be used for inexpensive encoding of typical consumer source video (broadcast PAL, NTSC). The natural consequence of still image methods is less efficient use of bandwidth: 25 Mbit/sec for the same subjective real-time playback quality achieved at 6 Mbit/sec possible with MPEG-2. A second bit rate mode, 50 Mbit/sec, is designated for HDTV.

Pre-coded digital video from, e.g., broadcast sources will be directly recorded to tape and "passed-through" as a coded bitstream to the video decompression box upon tape playback. Assuming if linear tape speed is to be proportional to bit rate, the recording time of a pre-compressed MPEG-2 program at the upper limit of 5 Mbit/sec for broadcast quality video, the recording time would be over 20 hours. Channel coding schemes (error correction, convolution coding, etc.), however, will most likely be optimized for the tape medium and therefore may differ from the channel methods for cable, terrestrial, and satellite. (A Zenith-Goldstar S-VHS based experiment did, however, directly record the 4-VSB broadcast baseband signal of the old Zenith/AT&T HDTV proposal).

More specs: (Summarized from EE Times July 5, 1993 article)

tape width: 6.35 mm
Audio: two channel 48 KHz 16-bit audio, or 4 channel at 32 KHz at 12-bit
Tape format: metal evaporated tape, 13.5 microns thick

Cassette dimensions: (millimeters)   Recording times:

Size	Width	Height	Depth	525/625 (25Mb/sec)	HDTV (50 Mb/s)
Standard	125	78	14.6	4h30min	2h15min
Small	66	48	12.2	1 hour	30min

Linear tape speeds: 18.812 mm/s (60Hz), 18.831 mm/s (50 Hz)
Video compression: DCT based

Participants: Matsushita, Sony, Philips, Thomson, Hitachi, Mitsubishi, Sanyo, Sharp, Toshiba, JVC.

MPEG in everyday life

74. Where will be see MPEG in everyday life?
A. Just about wherever you see video today.

DBS (Direct Broadcast Satellite)
The Hughes/USSB DBS service will use MPEG-2 video and audio. Thomson has exclusive rights to manufacture the decoding boxes for the first 18 months of operation. Hughes/USSB DBS will begin its U.S. service in April 1994. Two satellites at 101 degrees West will share the power requirements of 120 Watts per 27 MHz transponder over a total of 32 transponders. Multi source channel rate control methods will be employed to optimally allocate bits between several programs normalized to one 22 Mbit/sec data carrier. Bit allocation adapts to instantaneous co-channel
spatial and co-channel temporal activity. An average of 150 channels are planned with the addition of a second set of satellites augmenting the power level of each transponder to 240 Watts. The coded throughput of each transponder will increase to 30 Mbit/sec.

CATV (Cable Television)
Despite conflicting options, 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 exclusive use of Dolby AC-3. The General Instruments DigiCipher I video syntax is similar to MPEG-2 syntax, but employs smaller macroblock predictions and no B-frames. The DigiCipher II specification will include modes to support both the GI and full MPEG-2 Video Main Profile syntax. Digicipher-I services such as HBO will upgrade to DigiCipher II in 1994.

HDTV

The U.S. Grand Alliance, a consortium of companies that formerly competed to win the U.S. terrestrial HDTV standard, have already agreed to use the MPEG-2 Video and Systems syntax---including B-pictures. Both interlaced(1920 x 1080 x 30 Hz) and progressive (1280 x 720 x 60 Hz) modes will be supported. The Alliance has also settled upon a modulation method (VSB) convolution coding (Viterbi), and error correction (Reed-Soloman) 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 as soon as it appears to be politically secure thing to do).

75. What is the best compression ratio for MPEG ?

A. The MPEG sweet spot is about 1.2 bits/pel Intra and .35 bits/pel inter. Experimentation has shown that intra frame coding with the familiar DCT-Quantization-Entropy hybrid algorithm achieves optimal performance at about an average of 1.2 bits/sample or about 6:1 compression ratio. Below this point, artifacts become noticeable.

76. Is there a MPEG CD-ROM format?

A. Yes, a consortium of international companies (Matsushita, Philips, Sony, JVC, et al) have agreed upon a specification for MPEG video and audio. 2 hour long movies are stored on two 650 MByte compact discs. The video
rate is 1.15 Mbit/sec, the audio rate is either 128 kbit/sec or 192 kbit/sec Layer I or Layer II.(this seems to contradict the Philips 224 kbit/s audio spec?). Although the Video, Systems, and Audio syntax are identical, the CD-I movie format and the White Book format are not compatible.

Researchers are busy experimenting with denser and faster rate CD formats, perhaps using green or blue laser wavelengths. One demonstration stretched the pit and track density to its limits, improving areal density by almost 2 fold.