CCITT SGXV
Working Party XV/1
Specialists Group on Coding for Visual Telephony
Tokyo, 7-10 November 1989

Document #572
Oct. 26, 1989

SOURCE: EDITORIAL GROUP
TITLE : DRAFT REVISION OF RECOMMENDATION H.261

The attached draft has been produced at the Editorial Group meeting held in Ipswich during October 24 - 26, 1989. Study results obtained after the Stuttgart meeting have been reflected in the baseline document which was edited using the texts of the current H.261 in the Blue Book and the Flexible Hardware specification (Annex 4 to Doc. #540R). Substantially modified parts are indicated with bars in the left margin compared with those original texts.

Items to be worked out at the Tokyo meeting are as follows and indicated in { };

1. Quantizer reconstruction levels (§3.2.5)
2. Linked lists for PSPARE and GSPARE (§4.2.1 and §4.2.2)
3. Codewords for QUANT1 or choice of quantizer step sizes (§4.2.2)
4. Video coding delay (§5.3)
5. Re-lock time of FEC framing (§5.4.4)
6. Value of HRD buffer size B (Annex 2)

DRAFT REVISION OF RECOMMENDATION H.261

VIDEO CODEC FOR AUDIOVISUAL SERVICES AT px64 kbit/s

The CCITT,

considering

that there is significant customer demand for videophone, videoconference and other audiovisual services;

that circuits to meet this demand can be provided by digital transmission using the B, H0 rates or their multiples up to the primary rate or H11/H12 rates;

that ISDNs are likely to be available in some countries that provide a switched transmission service at the B, H0, or H11/H12 rate;

that the existence of different digital hierarchies and different television standards in different parts of the world complicates the problems of specifying coding and transmission standards for international connections;

that a number of audiovisual services are likely to appear using basic and primary rate ISDN accesses and that some means of intercommunication among these terminals should be possible;

that the video codec provides an essential element of the infrastructure for audiovisual services which allows such intercommunication in the framework of Recommendation H.200;

that Recommendation H.120 for videoconferencing using primary digital group transmission was the first in an evolving series of recommendations;

appreciating

that advances have been made in research and development of video coding and bit rate reduction techniques which lead to the use of lower bit rates down to 64 kbit/s so that this may be considered as the second in the evolving series of recommendations;

and noting

that it is the basic objective of CCITT to recommend unique solutions for international connections;

recommends

that in addition to those codecs complying to Recommendation H.120, codecs having signal processing and transmission coding characteristics described below should be used for international audiovisual services.

Note 1: Codecs of this type are also suitable for some television services where full broadcast quality is not required.
Note 2: Equipment for transcoding from and to codecs according to Recommendation H.120 is under study.

1. Scope

This Recommendation describes the video coding and decoding methods for the moving picture component of audiovisual services at the rates of px64 kbit/s, where p is in the range 1 to 30.

2. Brief Specification

An outline block diagram of the codec is given in Figure 1.

Figure 1/H.261 Outline block diagram of the video codec

2.1 Video input and output

To permit a single recommendation to cover use in and between regions using 625 and 525 line television standards, the source coder operates on pictures based on a common intermediate format (CIF). The standards of the input and output television signals, which may, for example, be composite or component, analogue or digital and the methods of performing any necessary conversion to and from the source coding format are not subject to recommendation.

2.2 Digital output and input

The video coder provides a self-contained digital bit stream which may be combined with other multi-facility signals (for example as defined in Rec. H.221). The video decoder performs the reverse process.

2.3 Sampling frequency

Pictures are sampled at an integer multiple of the video line rate. This sampling clock and the digital network clock are asynchronous.

2.4 Source coding algorithm

A hybrid of inter-picture prediction to utilize temporal redundancy and transform coding of the remaining signal to reduce spatial redundancy is adopted. The decoder has motion compensation capability, allowing optional incorporation of this technique in the coder.

2.5 Bit rate

This Recommendation is primarily intended for use at video bit rates between approximately 40 kbit/s and 2 Mbit/s.

2.6 Symmetry of transmission

The codec may be used for bidirectional or unidirectional visual communication.

2.7 Error handling

The transmitted bit-stream contains a BCH(511,493) Forward Error Correction Code. Use of this by the decoder is optional.

2.8 Multipoint Operation

Features necessary to support switched multipoint operation are included.

3. Source Coder

3.1 Source format

The source coder operates on non-interlaced pictures occurring 30000/1001 (approximately 29.97) times per second. The tolerance on picture frequency is +/-50 ppm.

Pictures are coded as luminance and two colour difference components (Y, C_B and C_R). These components and the codes representing their sampled values are as defined in CCIR Recommendation 601.

Black = 16
White = 235
Zero colour difference = 128
Peak colour difference = 16 and 240

These values are nominal ones and the coding algorithm functions with input values of 1 through to 254.

Two picture scanning formats are specified.

In the first format (CIF), the luminance sampling structure is 288 lines per picture, 352 pels per line in an orthogonal arrangement. Sampling of each of the two colour difference components is at 144 lines, 176 pels per line, orthogonal. Colour difference samples are sited such that their block boundaries coincide with luminance block boundaries as shown in Figure 2. The picture area covered by these numbers of pels and lines has an aspect ratio of 4:3 and corresponds to the active portion of the local standard video input.

Note: The number of pels per line is compatible with sampling the active portions of the luminance and colour difference signals from 525 or 625 line sources at 6.75 and 3.375 MHz respectively. These frequencies have a simple relationship to those in CCIR Recommendation 601.

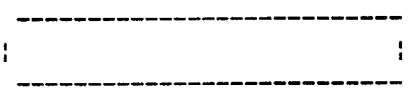


Figure 2/H.261 Positioning of luminance and chrominance samples

The second format, Quarter-CIF (QCIF), has half the number of pels and half the number of lines stated above. All codecs must be able to operate using QCIF. Some codecs can also operate with CIF.

Means shall be provided to restrict the maximum picture rate of encoders by having at least 0, 1, 2 or 3 non-transmitted pictures between transmitted ones. Selection of this minimum number and CIF or QCIF shall be by external means (for example via Recommendation H.221).

3.2 Video source coding algorithm

The source coder is shown in generalised form in Figure 3. The main elements are prediction, block transformation, and quantisation.

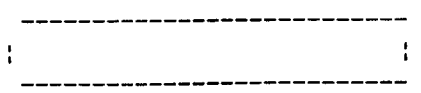


Figure 3/H.261 Source Coder

The prediction error (INTER mode) or the input picture (INTRA mode) is subdivided into 8 pel by 8 line blocks which are segmented as transmitted or non-transmitted. Further, four luminance blocks and the two spatially corresponding colour difference blocks are combined to form a macroblock as shown in Figure 11 of §4.2.4.

The criteria for choice of mode and transmitting a block are not subject to recommendation and may be varied dynamically as part of the coding control strategy. Transmitted blocks are transformed and resulting coefficients are quantised and variable length coded.

3.2.1 Prediction

The prediction is inter-picture and may be augmented by motion compensation (§3.2.2) and a spatial filter (§3.2.3).

3.2.2 Motion compensation

Motion compensation is optional in the encoder. The decoder will accept one vector per macroblock. Both horizontal and vertical components of these motion vectors have integer values not exceeding +/- 15. The vector is used for all four luminance blocks in the macroblock. The motion vector for both colour difference blocks is derived by halving the component values of the macroblock vector and truncating the magnitude parts towards zero to yield integer components.

A positive value of the horizontal or vertical component of the motion vector signifies that the prediction is formed from pels in the previous picture which are spatially to the right or below the pels being predicted.

Motion vectors are restricted such that all pels referenced by them are within the coded picture area.

3.2.3 Loop filter

The prediction process may be modified by a two-dimensional spatial filter which operates on pels within a predicted 8 by 8 block.

The filter is separable into one dimensional horizontal and vertical functions. Both are non-recursive with coefficients of 1/4, 1/2, 1/4 except at block edges where one of the taps would fall outside the block. In such cases the 1-D filter is changed to have coefficients of 0, 1, 0. Full arithmetic precision is retained with rounding to 8 bit integer values at the 2-D filter output. Values whose fractional part is one half are rounded up.

The filter is switched on/off for all 6 blocks in a macroblock according to the macroblock type. (See §4.2.3 TYPEB).

3.2.4 Transformer

Transmitted blocks are first processed by a separable 2-dimensional Discrete Cosine Transform of size 8 by 8. The output from the inverse transform ranges from -256 to +255 and is represented with 9 bits. The transfer function of the inverse transform is given by

$$f(x,y)=1/4 \sum_{u=0}^{7} \sum_{v=0}^{7} C(u)\ C(v)\ F(u,v)\ \cos[Pi(2x+1)u/16]\ \cos[Pi(2y+1)v/16]$$

with u, v, x, y = 0, 1, 2, ...7

where x,y = spatial coordinates in the pel domain
u,v = coordinates in the transform domain

$$C(u),\ C(v) = \begin{cases} 1/SQRT(2) & \text{for } u,v = 0 \\ 1 & \text{otherwise} \end{cases}$$

Note: Within the block being transformed, x=0 and y=0 refer to the pel nearest the left and top edges of the picture respectively.

The arithmetic procedures for computing the transforms are not defined, but the inverse one should meet the error tolerance specified in Annex 1.

3.2.5 Quantisation

The number of quantisers is 32. The decision levels are not defined. The reconstruction levels are as defined below.

{Tables of reconstruction levels should be fixed at the Tokyo meeting.}
{Some words should be added for quantizer assignment?}

Table 1/H.261 Quantizer for INTRA dc component

Table 2/H.261 Quantizers for INTRA ac and INTER components

Note: For the smaller quantisation step sizes, the full dynamic range of the transform coefficients cannot be represented.

3.2.6 Clipping

To prevent quantisation distortion of transform coefficient amplitudes causing arithmetic overflow in the encoder and decoder loops, clipping functions are inserted. In addition to those in the inverse transform, a clipping function is applied at both encoder and decoder to the reconstructed picture which is formed by summing the prediction and the prediction error as modified by the coding process. This clipper operates on resulting pel values less than 0 or greater than 255, changing them to 0 and 255 respectively.

3.3 Coding control

Several parameters may be varied to control the rate of generation of coded video data. These include processing prior to the source coder, the quantiser, block significance criterion and temporal subsampling. The proportions of such measures in the overall control strategy are not subject to recommendation.

When invoked, temporal subsampling is performed by discarding complete pictures.

3.4 Forced updating

This function is achieved by forcing the use of the INTRA mode of the coding algorithm. The update pattern is not defined. For control of accumulation of inverse transform mismatch error a block should be forcibly updated at least once per every 132 times it is transmitted.

4. Video Multiplex Coder

4.1 Data Structure

Unless specified otherwise the most significant bit is transmitted first. This is Bit 1 and is the leftmost bit in the code tables in this document. Unless specified otherwise all unused or spare bits are set to '1'. Spare bits must not be used until their functions are specified by CCITT.

4.2 Video Multiplex arrangement

The video multiplex is arranged in a hierarchical structure with four layers. From top to bottom the layers are:

Picture
Group of Blocks (GOB)
Macroblock (MB)
Block

A syntax diagram of the video multiplex coder is shown in Figure 4.

Figure 4/H.261 Syntax diagram for the video multiplex coder

4.2.1 Picture Layer

Data for each Picture consists of a Picture Header followed by data for GOBs. The structure is shown in Figure 5. Picture Headers for dropped pictures are not transmitted.

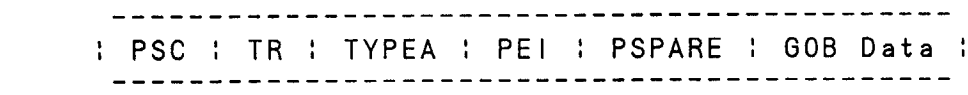


Figure 5/H.261 Structure of the Picture Layer

Picture Start Code (PSC) 20 bits

A word of 20 bits. Its value is 0000 0000 0000 0001 0000

Temporal Reference (TR) 5 bits

A five bit number which can have 32 possible values. It is formed by incrementing its value in the previously transmitted Picture header by 1 plus the number of non-transmitted pictures (at 29.97 Hz) since that last transmitted one. The arithmetic is performed with only the 5 LSBs.

Type Information (TYPEA) 6 bits

Information about the complete picture:

Bit 1 Split screen indicator. '0' off, '1' on.
Bit 2 Document camera indicator. '0' off, '1' on.
Bit 3 Freeze Picture Release. '0' off, '1' on.
Bit 4 Source Format. '0' QCIF, '1' CIF.
Bits 5 to 6 Spare.

Extra Insertion Information (PEI) 1 bit

A bit which signals the presence of the following optional data field.

Spare Information (PSPARE) 0/9/18 ... bits

If PEI is set to '1', then 9 bits follow consisting of 8 bits of data and then another PEI bit to indicate if a further 9 bits follow and so on. Encoders must not insert PSPARE until specified by CCITT. Decoders must be designed to discard PSPARE if PEI is set to 1.

Note: This will allow CCITT to specify future "backward" compatible additions in PSPARE.

(This change needs to be agreed in Tokyo.)

4.2.2 Group of Blocks Layer

Each picture is divided into Groups of Blocks (GOBs). A group of blocks (GOB) comprises one twelfth of the CIF or one third of the QCIF picture areas (see Figure 6). A GOB relates to 176 pels by 48 lines of Y and the spatially corresponding 88 pels by 24 lines of each of C_R and C_B.

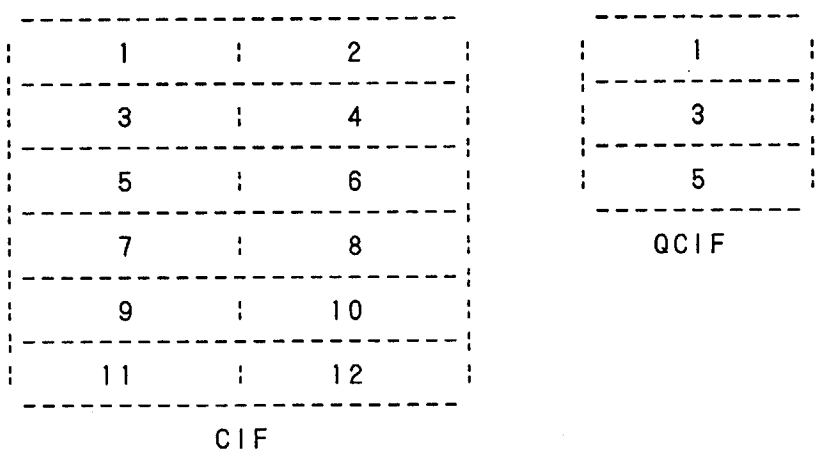


Figure 6/H.261 Arrangement of GOBs in a Picture

Data for each Group of Blocks consists of a GOB Header followed by data for macroblocks. The structure is shown in Figure 7. Each GOB Header is transmitted once between Picture Start Codes in the CIF or QCIF sequence numbered above, even if no macroblock data is present in that GOB.

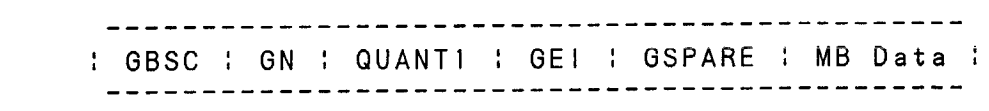


Figure 7/H.261 Structure of Group of Blocks Layer

Group of Blocks Start Code (GBSC) 16 bits

A word of 16 bits, 0000 0000 0000 0001.

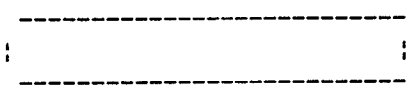
Group Number (GN) 4 bits

Four bits indicating the position of the group of blocks. The bits are the binary representation of the numbers in Figure 6. Group numbers 13, 14 and 15 are reserved for future use. Group number 0 is used in the PSC.

Quantiser Information (QUANT1) 5 bits

A fixed length codeword of 5 bits which indicates the quantiser table to be used in the group of blocks until overridden by any subsequent QUANT2. The codewords are given below.

Table 3/H.261 Codewords for QUANT1



{To be added in Tokyo}

Extra Insertion Information (GEI) 1 bit

A bit which signals the presence of the following optional data field.

Spare Information (GSPARE) 0/9/18 .... bits

If GEI is set to '1', then 9 bits follow consisting of 8 bits of data and then another GEI bit to indicate if a further 9 bits follow and so on. Encoders must not insert GSPARE until specified by CCITT. Decoders must be designed to discard GSPARE if GEI is set to 1. This will allow CCITT to specify future "backward" compatible additions in GSPARE.

{This change needs to be agreed in Tokyo.}

Note: The future specification of GSPARE should ensure that emulation of start codes by the final 9 bits of GSPARE does not occur.

4.2.3 Macroblock Layer

Each GOB is divided into 33 macroblocks as shown in Figure 8. A macroblock relates to 16 lines by 16 pels of Y and the spatially corresponding 8 lines by 8 pels of each of C_B and C_R.

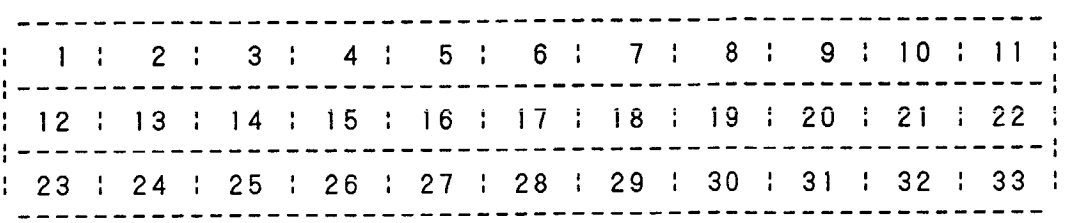


Figure 8/H.261 Arrangement of macroblocks in a GOB

Data for a macroblock consists of a MB Header followed by data for blocks (Figure 9). Elements are omitted when not required.

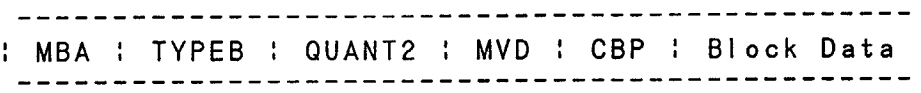


Figure 9/H.261 Structure of macroblock layer

Macroblock Address (MBA) Variable Length

A variable length codeword indicating the position of a macroblock within a group of blocks. The transmission order is as shown in Figure 8. For the first transmitted block in a GOB, MBA is the absolute address in Figure 8. For subsequent macroblocks, MBA is the difference between the absolute addresses of the macroblock and the last transmitted macroblock. The code table for MBA is given in Table 4.

An extra codeword is available in the table for bit stuffing immediately after a GOB header or a coded macroblock (MBA Stuffing). This codeword should be discarded by decoders.

The VLC for start code is also shown in Table 4.

Table 4/H.221 Macroblock Addressing

MBA	CODE	MBA	CODE
1	1	17	0000 0101 10
2	011	18	0000 0101 01
3	010	19	0000 0101 00
4	0011	20	0000 0100 11
5	0010	21	0000 0100 10
6	0001 1	22	0000 0100 011
7	0001 0	23	0000 0100 010
8	0000 111	24	0000 0100 001
9	0000 110	25	0000 0100 000
10	0000 1011	26	0000 0011 111
11	0000 1010	27	0000 0011 110
12	0000 1001	28	0000 0011 101
13	0000 1000	29	0000 0011 100
14	0000 0111	30	0000 0011 011
15	0000 0110	31	0000 0011 010
16	0000 0101 11	32	0000 0011 001
		33	0000 0011 000
		MBA Stuffing	0000 0001 111
		Start code	0000 0000 0000 0001

MBA is always included in transmitted macroblocks.

Macroblocks are not transmitted when they contain no information for that part of the picture.

Type Information (TYPEB) Variable Length

Variable length codewords giving information about the macroblock and which data elements are present. Macroblock types, included elements and VLC words are listed in Table 5.

TYPEB is always included in transmitted macroblocks.

Table 5/H.261 VLC Table for TYPEB

Prediction	QUANT2	MVD	CBP	TCOEFF	VLC
Intra				x	0001
Intra	x			x	0000 001
Inter			x	x	1
Inter	x		x	x	0000 1
Inter + MC		x			0000 0000 1
Inter + MC		x	x	x	0000 0001
Inter + MC	x	x	x	x	0000 0000 01
Inter + MC + FIL		x			001
Inter + MC + FIL		x	x	x	01
Inter + MC + FIL	x	x	x	x	0000 01

Note 1: 'x' means that the item is present in the macroblock
Note 2: It is possible to apply the filter in a non-motion compensated macroblock by declaring it as MC + FIL but with a zero vector.

Quantiser (QUANT2) 5 bits

QUANT2 is present only if so indicated by TYPEB.

A codeword of 5 bits signifying the quantiser table to be used for this and any following blocks in the group of blocks until overridden by any subsequent QUANT2.

Codewords for QUANT2 are the same as for QUANT1.

Motion Vector Data (MVD) Variable length

Motion Vector Data is included for all MC macroblocks. MVD is obtained from the macroblock vector by subtracting the vector of the preceding macroblock. For this calculation the vector of the preceding macroblock is regarded as zero in the following three situations:

1) Evaluating MVD for macroblocks 1, 12 and 23.
2) Evaluating MVD for macroblocks in which MBA does not represent a difference of 1.
3) TYPEB of the previous macroblock was not MC.

MVD consists of a variable length codeword for the horizontal component followed by a variable length codeword for the vertical component. Variable length codes are given in Table 5.

Table 5/H.261 VLC Table for MVD

MVD	CODE
-16 & 16	0000 0011 001
-15 & 17	0000 0011 011
-14 & 18	0000 0011 101
-13 & 19	0000 0011 111
-12 & 20	0000 0100 001
-11 & 21	0000 0100 011
-10 & 22	0000 0100 11
-9 & 23	0000 0101 01
-8 & 24	0000 0101 11
-7 & 25	0000 0111
-6 & 26	0000 1001
-5 & 27	0000 1011
-4 & 28	0000 111
-3 & 29	0001 1
-2 & 30	0011
-1	011
0	1
1	010
2 & -30	0010
3 & -29	0001 0
4 & -28	0000 110
5 & -27	0000 1010
6 & -26	0000 1000
7 & -25	0000 0110
8 & -24	0000 0101 10
9 & -23	0000 0101 00
10 & -22	0000 0100 10
11 & -21	0000 0100 010
12 & -20	0000 0100 000
13 & -19	0000 0011 110
14 & -18	0000 0011 100
15 & -17	0000 0011 010

Advantage is taken of the fact that the range of motion vector values is constrained. Each VLC word represents a pair of difference values. Only one of the pair will yield a macroblock vector falling within the permitted range.

Coded Block Pattern (CBP) Variable length

CBP is present if indicated by TYPEB. The codeword gives a pattern number signifying those blocks in the macroblock for which at least one transform coefficient is transmitted. The pattern number is given by

$$32*P_1 + 16*P_2 + 8*P_3 + 4*P_4 + 2*P_5 + P_6$$

where P_n is 1 if any coefficient is present for block n, else 0. Block numbering is given in Figure 10.

The code words for CBP is given in Table 7.

Table 7/H.221 VLC Table for CBP

CBP	CODE	CBP	CODE
60	111	35	0001 1100
4	1101	13	0001 1011
8	1100	49	0001 1010
16	1011	21	0001 1001
32	1010	41	0001 1000
12	1001 1	14	0001 0111
48	1001 0	50	0001 0110
20	1000 1	22	0001 0101
40	1000 0	42	0001 0100
28	0111 1	15	0001 0011
44	0111 0	51	0001 0010
52	0110 1	23	0001 0001
56	0110 0	43	0001 0000
1	0101 1	25	0000 1111
61	0101 0	37	0000 1110
2	0100 1	26	0000 1101
62	0100 0	38	0000 1100
24	0011 11	29	0000 1011
36	0011 10	45	0000 1010
3	0011 01	53	0000 1001
63	0011 00	57	0000 1000
5	0010 111	30	0000 0111
9	0010 110	46	0000 0110
17	0010 101	54	0000 0101
33	0010 100	58	0000 0100
6	0010 011	31	0000 0011 1
10	0010 010	47	0000 0011 0
18	0010 001	55	0000 0010 1
34	0010 000	59	0000 0010 0
7	0001 1111	27	0000 0001 1
11	0001 1110	39	0000 0001 0
19	0001 1101		

4.2.4 Block Layer

A macroblock comprises four luminance blocks and one of each of the two colour difference blocks.

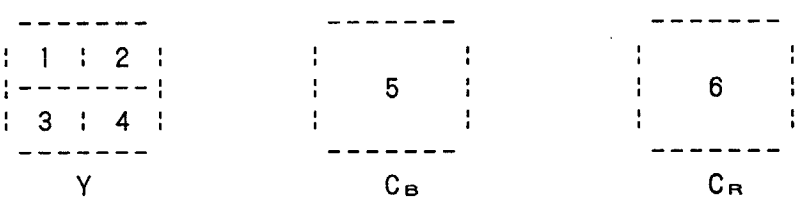


Figure 10/H.261 Arrangement of blocks in a macroblock

Data for a block consists of codewords for transform coefficients followed by an end of block marker (Figure 11). The order of block transmission is as in Figure 10.

Figure 11/H.261 Structure of Block Layer

Transform Coefficients (TCOEFF)

Transform coefficient data is always present for all 6 blocks in a macroblock when TYPEB indicates INTRA. In other cases TYPEB and CBP signal which blocks have coefficient data transmitted for them. The quantised transform coefficients are sequentially transmitted according to the sequence given in Figure 12.

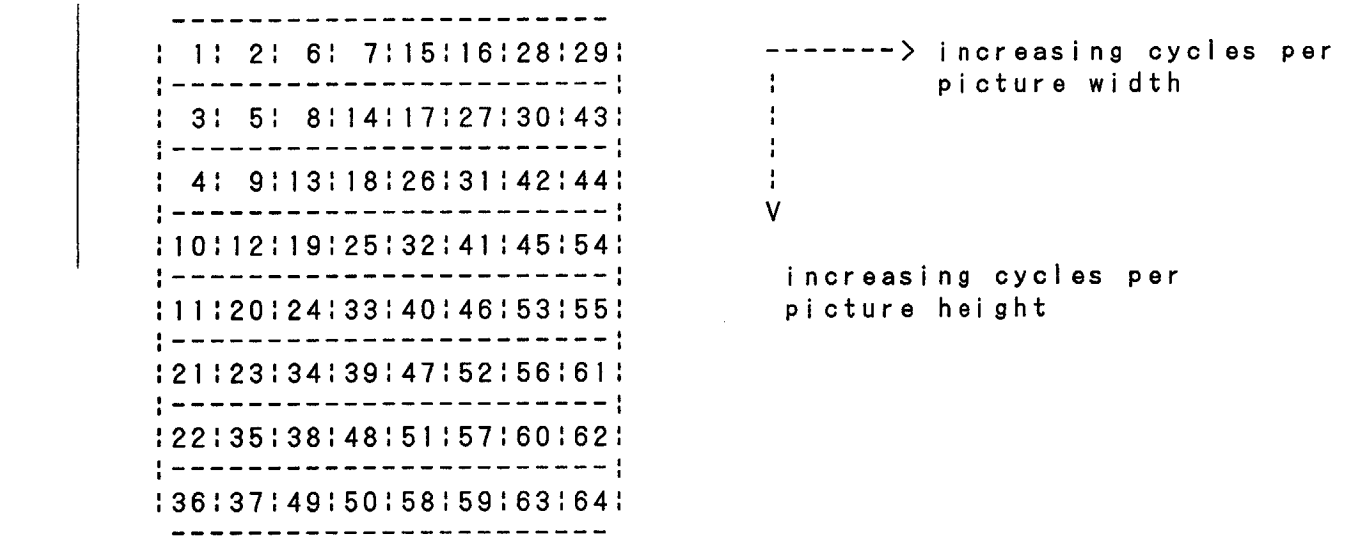


Figure 12/H.261 Transmission Order for Transform Coefficients

The most commonly occurring combinations of successive zero coefficients (RUN) and the following value (LEVEL) are encoded with variable length codes. Other combinations of (RUN, LEVEL) are encoded with a 20 bit word consisting of 6 bits ESCAPE, 6 bits RUN and 8 bits LEVEL. For the variable length encoding there are two code tables, one being used for the first transmitted LEVEL in INTER and INTER + MC blocks, the second for all other LEVELS except the first one in INTRA blocks which is fixed length coded with 8 bits.

Codes are given in Table 8/H.261.

Table 8/H.261 VLC Table for TCOEFF

The most commonly occurring combinations of zero-run and the following value are encoded with variable length codes as listed in the table below. End of Block (EOB) is in this set. Because CBP indicates those blocks with no coefficient data, EOB cannot occur as the first coefficient. Hence EOB can be removed from the VLC table for the first coefficient.

The last bit 's' denotes the sign of the level, '0' for positive
'1' for negative.

RUN	LEVEL	CODE
EOB		10
0	1	1s IF FIRST COEFFICIENT
0	1	11s NOT FIRST COEFFICIENT
0	2	0100 s
0	3	0010 1s
0	4	0000 110s
0	5	0010 0110 s
0	6	0010 0001 s
0	7	0000 0010 10s
0	8	0000 0001 1101 s
0	9	0000 0001 1000 s
0	10	0000 0001 0011 s
0	11	0000 0001 0000 s
0	12	0000 0000 1101 0s
0	13	0000 0000 1100 1s
0	14	0000 0000 1100 0s
0	15	0000 0000 1011 1s
1	1	011s
1	2	0001 10s
1	3	0010 0101 s
1	4	0000 0011 00s
1	5	0000 0001 1011 s
1	6	0000 0000 1011 0s
1	7	0000 0000 1010 1s
2	1	0101 s
2	2	0000 100s
2	3	0000 0010 11s
2	4	0000 0001 0100 s
2	5	0000 0000 1010 0s
3	1	0011 1s
3	2	0010 0100 s
3	3	0000 0001 1100 s
3	4	0000 0000 1001 1s
4	1	0011 0s
4	2	0000 0011 11s
4	3	0000 0001 0010 s
5	1	0001 11s
5	2	0000 0010 01s
5	3	0000 0000 1001 0s
6	1	0001 01s
6	2	0000 0001 1110 s
7	1	0001 00s
7	2	0000 0001 0101 s
8	1	0000 111s
8	2	0000 0001 0001 s
9	1	0000 101s

9	2	0000 0000 1000 1s
10	1	0010 0111 s
10	2	0000 0000 1000 0s
11	1	0010 0011 s
12	1	0010 0010 s
13	1	0010 0000 s
14	1	0000 0011 10s
15	1	0000 0011 01s
16	1	0000 0010 00s
17	1	0000 0001 1111 s
18	1	0000 0001 1010 s
19	1	0000 0001 1001 s
20	1	0000 0001 0111 s
21	1	0000 0001 0110 s
22	1	0000 0000 1111 1s
23	1	0000 0000 1111 0s
24	1	0000 0000 1110 1s
25	1	0000 0000 1110 0s
26	1	0000 0000 1101 1s
ESCAPE		0000 01

The remaining combinations of (RUN, LEVEL) are encoded with a 20 bit word consisting of 6 bits ESCAPE, 6 bits RUN and 8 bits LEVEL.

RUN is a 6 bit fixed length code

RUN	CODE
0	0000 00
1	0000 01
2	0000 10
.	.
.	.
63	1111 11

LEVEL is an 8 bit fixed length code

LEVEL	CODE	
-127	1000 0001	
.	.	
.	.	
-2	1111 1110	
-1	1111 1111	
0	0000 0000	FORBIDDEN
1	0000 0001	
2	0000 0010	
.	.	
127	0111 1111	

For INTRA blocks the first coefficient is the transform dc value linearly quantised with a stepsize of 8 and no dead-zone. The resulting values are represented with 8 bits. A nominally black block will give 0001 0000 and a nominally white one 1110 1011. The code 0000 0000 is not used. The code 1000 0000 is not used, the reconstruction level of 1024 being coded as 1111 1111. (See Table 9)

Table 9/H.261 Quantiser for INTRA-mode DC component

FLC	Reconstruction level into inverse transform
0000 0001 (1)	8
0000 0010 (2)	16
0000 0011 (3)	24
.	.
.	.
0111 1111 (127)	1016
1111 1111 (255)	1024
1000 0001 (129)	1032
.	.
.	.
1111 1101 (253)	2024
1111 1110 (254)	2032

Note: The decoded value corresponding to FLC 'n' is 8n except FLC 255 gives 1024.

Coefficients after the last non-zero one are not transmitted. EOB (End of Block code) is always the last item in blocks for which coefficients are transmitted.

4.3 Multipoint considerations

The following facilities are provided to support switched multipoint operation.

4.3.1 Freeze Picture Request

Causes the decoder to freeze its displayed picture until a picture freeze release signal is received or a timeout period of at least 6 seconds has expired. The transmission of this signal is via external means (for example by H.221).

4.3.2 Fast Update Request

Causes the encoder to encode its next picture in INTRA mode with coding parameters such as to avoid buffer overflow. The transmission method for this signal is via external means (for example by H.221).

4.3.3 Picture Freeze Release

A signal from an encoder which has responded to a Fast Update Request and causes a decoder to exit from its picture freeze mode and display decoded pictures in the normal manner. This signal is transmitted by Bit 3 of TYPEA. (See §4.2.1)

5. Transmission Coder

5.1 Bit rate

The transmission clock is provided externally (for example from an I.420 interface).

5.2 Video Data Buffering

The encoder must control its output bitstream such that the buffer of the hypothetical reference decoder defined in Annex 2 does not overflow.

When operating with CIF the number of bits created by coding any single picture must not exceed 256 Kbits.

When operating with QCIF the number of bits created by coding any single picture must not exceed 64 Kbits.

In both the above cases the bit count includes the Picture Start Code and all other data related to that picture including PSPARE, GSPARE and MBA Stuffing. K = 1024.

Video data must be provided on every valid clock cycle. This can be ensured by the use of either dummy blocks in the error corrector block framing (See Figure 13) or MBA Stuffing.

5.3 Video Coding Delay

(For Tokyo meeting)

5.4 Forward Error Correction for Coded Video Signal

5.4.1 Error correcting code

The transmitted bit-stream contains a BCH(511,493) Forward Error Correction Code. Use of this by the decoder is optional.

5.4.2 Generator polynomial

$g(x)=(x^9+x^4+1)(x^9+x^6+x^4+x^3+1)$

Example: For the input data of '01111...11' (493 bits) the resulting correction parity bits are '011011010100011011' (18 bits).

5.4.3 Error Correction Framing

Independent framing is used with multiframe of 8 frames, where a 512-bit frame consists of 1-bit framing, 493-bit data and 18-bit Error Correction Parity. The frame alignment pattern is

$(S_0S_1S_2S_3S_4S_5S_6S_7)=(00011011)$,

See Figure 13 for the frame arrangement. Note that the Error Correction Parity is calculated against the 493-bit including Fill Indicator.

Figure 13/H.221 Error correcting frame

5.4.4 Re-lock time for error corrector framing

Three consecutive error correction framing sequences (24 bits) should be received before frame lock is deemed to have been achieved. The decoder should be designed such that frame lock will be re-established within .... bits after an error corrector framing phase change.

{Specific value should be fixed at the Tokyo meeting}

Note: This assumes that the video data does not contain 3 correctly phased emulations of the error correction framing sequence during the re-locking period.

END

Annex 1 to H.261

Inverse Transform Accuracy Specification

1. Generate random integer pel data values in the range -L to +H according to the random number generator given below ('C' version). Arrange into 8 by 8 blocks. Data sets of 10,000 blocks should each be generated for (L=256, H=255), (L=H=5) and (L=H=300).

2. For each 8 by 8 block, perform a separable, orthonormal, matrix multiply, Forward Discrete Cosine Transform using at least 64-bit floating point accuracy.

3. For each block, round the 64 resulting transformed coefficients to the nearest integer values. Then clip them to the range -2048 to +2047. This is the 12-bit input data to the inverse transform.

4. For each 8 by 8 block of 12-bit data produced by step 3, perform a separable, orthonormal, matrix multiply, Inverse Discrete Transform (IDCT) using at least 64-bit floating point accuracy. Round the resulting pels to the nearest integer and clip to the range -256 to +255. These blocks of 8 by 8 pels are the "reference" IDCT output data.

5. For each 8 by 8 block produced by step 3, apply the IDCT under test to perform an IDCT and clip the output to the range -256 to +255. These blocks of 8 by 8 pels are the "test" IDCT output data.

6. For each of the 64 IDCT output pels, and for each of the 10,000 block data sets generated above, measure the peak, mean and mean square error between the "reference" and the "test" data.

7. For any pel, the peak error should not exceed 1 in magnitude.
   For any pel, the mean square error should not exceed 0.06.
   Overall, the mean square error should not exceed 0.02.
   For any pel, the mean mean error should not exceed 0.015 in magnitude.
   Overall, the mean error should not exceed 0.0015 in magnitude.

8. All zeros in must produce all zeros out.

9. Rerun the measurements using exactly the same data values of step 1, but change the sign on each pel.

'C' Program for random number generation

```
/* L and H must be long, that is 32 bits */
long rand(L,H)
long      L,H;
{
  static long  randx = 1;  /* long is 32 bits */
  static double  z = (double)0x7fffffff;

  long i,j;
  double x;                    /* double is 64 bits */

  randx = (randx * 1103515425) + 12345;
  i = randx & 0x7ffffffe;        /* keep 30 bits */
  x = ( (double)i ) / z;         /* range 0 to 0.99999... */
  x *= (L+H+1);                  /* range 0 to < L+H+1 */
  j = x;                         /* truncate to integer */
  return( j - L );               /* range -L to H */
}
```

END

Annex 2 to H.261

Hypothetical Reference Decoder

The hypothetical reference decoder (HRD) is defined as follows:

1. The clock frequency of the HRD and that of the source coder are the same CIF rate and operate synchronously.

2. The HRD receiving buffer consists of a fixed size B and a variable size X. The value of B is defined as:

$$B = m * R * k/29.97$$

where R is the video channel rate,
29.97/k denotes the maximum picture rate defined in §3.1,
X is the number of bits of one coded picture.

{The value of m is to be fixed in Tokyo}

3. The HRD buffer is initially empty.

4. All the bits for one coded picture are removed instantaneously from the buffer at the earliest possible 1/29.97 interval.

5. The buffer occupancy after the data for the N-th coded picture is removed is b_N and is given by:

$$b_N = \sum_{n=1}^{N} (R*k/29.97 - d_n)$$

6. To avoid buffer overflow, the number of bits for the (N+1)th coded picture must satisfy:

$$d_{N+1} \geqq b_N + R*k/29.97 - B$$

This is a requirement on the coder output bitstream.

END

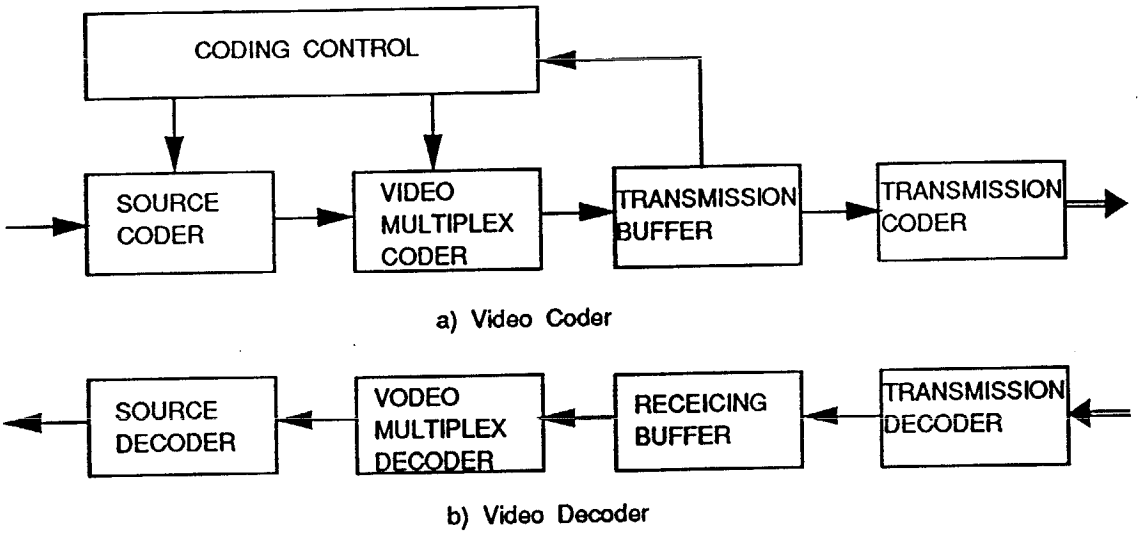


Figure 1/H.261 Outline block diagram of the video codec

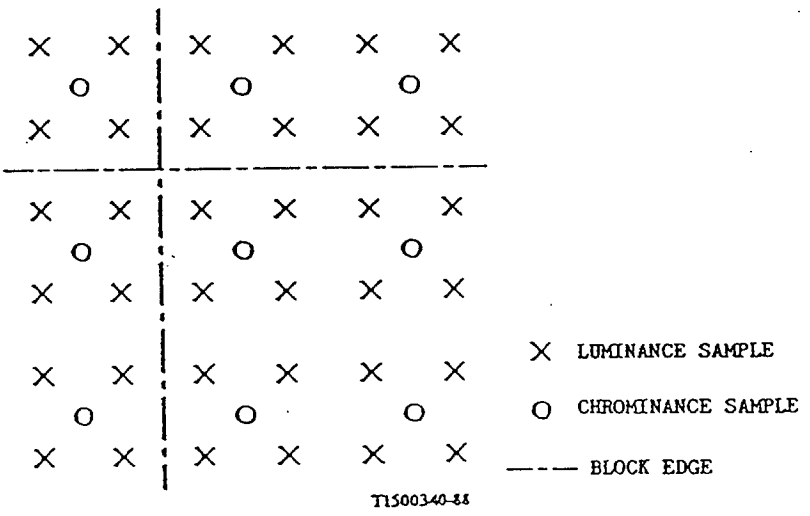
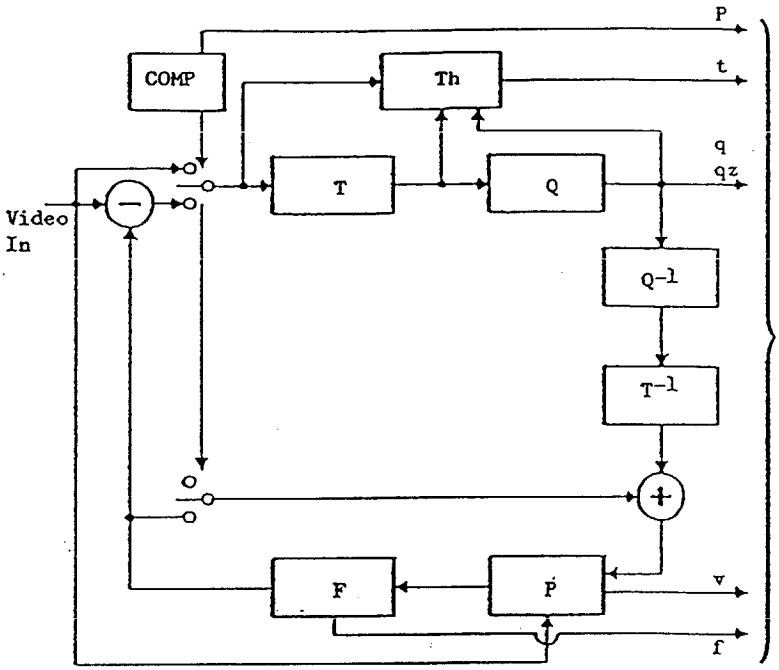


Figure 2/H.261 Positioning of luminance and chrominance samples



COMP Comparator for intra/inter
Th Threshold
T Transform
Q Quantizer
P Picture memory with motion compensated variable delay
F Loop filter

p Flag for intra/inter
t Flag for transmitted or not
q Quantizing index for transform coefficients
qz Quantizer indication
v Motion vector
f Switching on/off of the loop filter

Figure 3/H.261 Source Coder

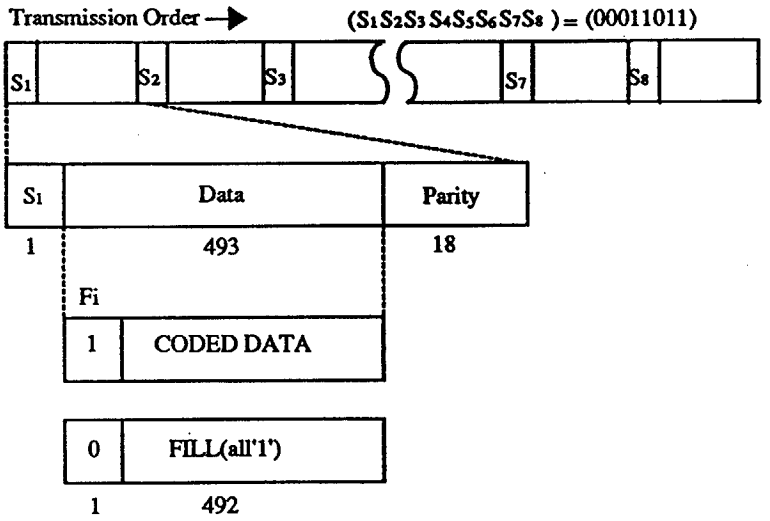
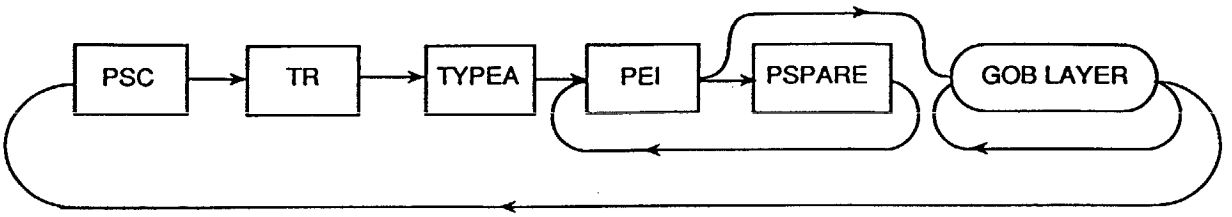
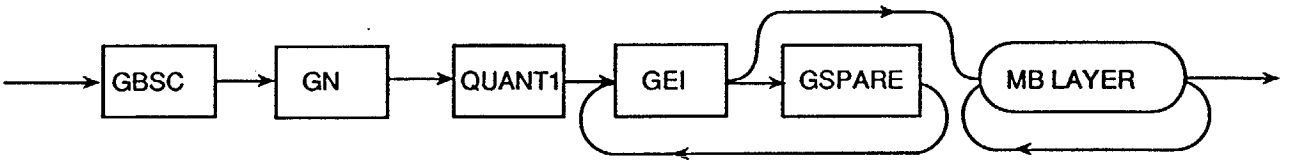


Figure 13/H.221 Error correcting frame

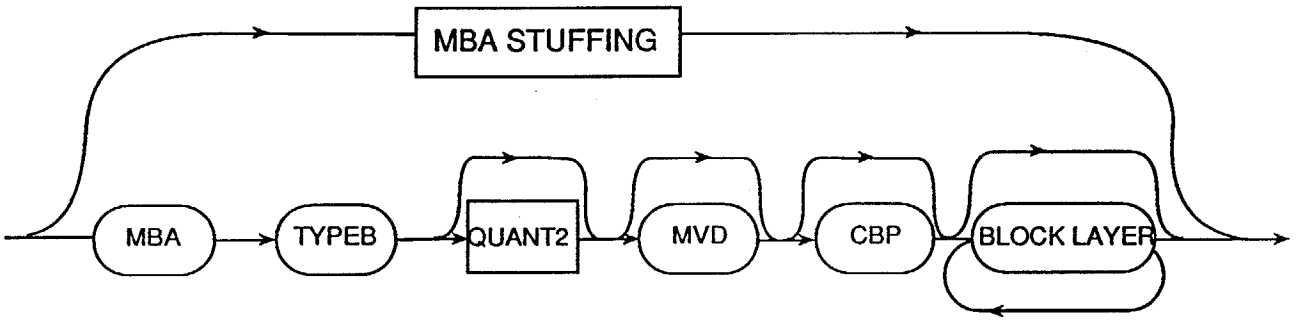
PICTURE LAYER



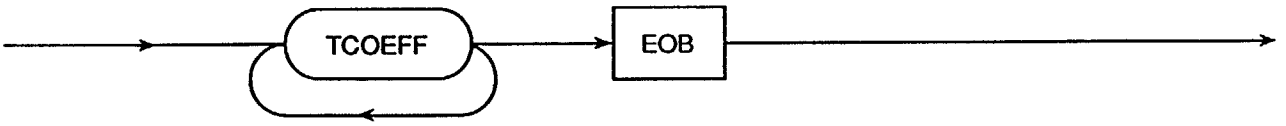
GOB LAYER



MB LAYER



BLOCK LAYER



Fixed length

Variable length

Figure 4/H.261 Syntax diagram for the video multiplex coder