INTERNATIONAL TELECOMMUNICATION UNION

ITU-T H.261

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(03/93)

LINE TRANSMISSION OF NON-TELEPHONE SIGNALS

VIDEO CODEC FOR AUDIOVISUAL SERVICES AT $p \times 64$ kbits

ITU-T Recommendation H.261

(Previously "CCITT Recommendation")

FOREWORD

The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the International Telecommunication Union. The ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Conference (WTSC), which meets every four years, established the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

ITU-T Recommendation H.261 was revised by the ITU-T Study Group XV (1988-1993) and was approved by the WTSC (Helsinki, March 1-12, 1993).

NOTES

1 As a consequence of a reform process within the International Telecommunication Union (ITU), the CCITT ceased to exist as of 28 February 1993. In its place, the ITU Telecommunication Standardization Sector (ITU-T) was created as of 1 March 1993. Similarly, in this reform process, the CCIR and the IFRB have been replaced by the Radiocommunication Sector.

In order not to delay publication of this Recommendation, no change has been made in the text to references containing the acronyms "CCITT, CCIR or IFRB" or their associated entities such as Plenary Assembly, Secretariat, etc. Future editions of this Recommendation will contain the proper terminology related to the new ITU structure.

2 In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

© ITU 1994

All rights reserved. No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the ITU.

CONTENTS

			Page
1	Scope		1
2	Brief specification		1
	2.1	Video input and output	2
	2.2	Digital output and input	2
	2.3	Sampling frequency	2
	2.4	Source coding algorithm	2
	2.5	Bit rate	2
	2.6	Symmetry of transmission	3
	2.7	Error handling	3
	2.8	Multipoint operation	3
3	Source coder		3
	3.1	Source format	3
	3.2	Video source coding algorithm	3
	3.3	Coding control	6
	3.4	Forced updating	6
4	Video multiplex coder		7
	4.1	Data structure	7
	4.2	Video multiplex arrangement	7
	4.3	Multipoint considerations	18
5	Transmission coder		19
	5.1	Bit rate	19
	5.2	Video data buffering	19
	5.3	Video coding delay	20
	5.4	Forward error correction for coded video signal	20
Annex A – Inverse transform accuracy specification			21
Annex B – Hypothetical reference decoder			22
Annex C – Codec delay measurement method			23
Annex D – Still image transmission			24

Recommendation H.261

VIDEO CODEC FOR AUDIOVISUAL SERVICES AT $p \times 64$ kbit/s

(Geneva, 1990; revised at Helsinki, 1993)

The CCITT,

considering

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

(b) that circuits to meet this demand can be provided by digital transmission using the B, H_0 rates or their multiples up to the primary rate or H_{11}/H_{12} rates;

(c) that ISDNs are likely to be available in some countries that provide a switched transmission service at the B, H_0 or H_{11}/H_{12} rate;

(d) 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;

(e) 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;

(f) 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;

(g) 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 the 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.

NOTES

1 Codecs of this type are also suitable for some television services where full broadcast quality is not required.

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 $p \times 64$ 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.

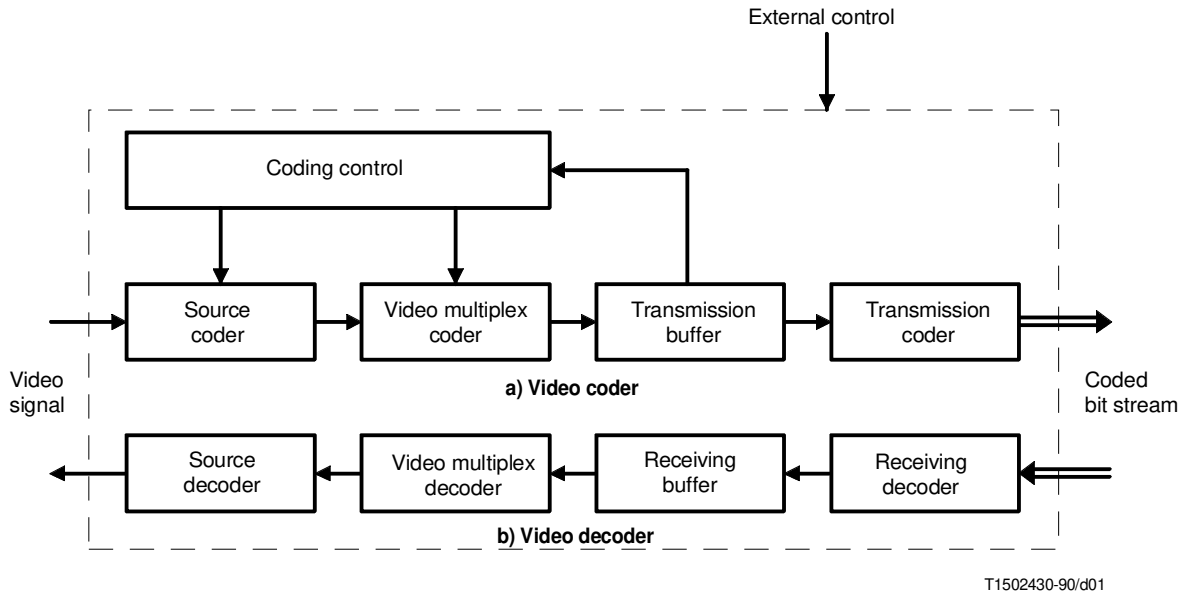


FIGURE 1/H.261

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 Recommendation 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 code (Bose, Chaudhuri and Hocquengham) (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 30 000/1001 (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 352 pels per line, 288 lines per picture in an orthogonal arrangement. Sampling of each of the two colour difference components is at 176 pels per line, 144 lines per picture, 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.

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 generalized form in Figure 3. The main elements are prediction, block transformation and quantization.

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 10.

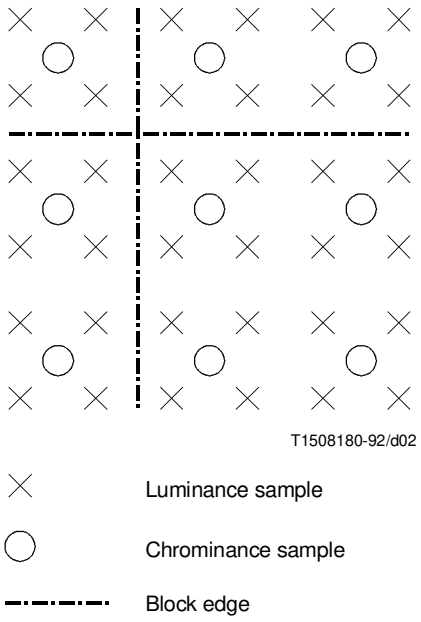


FIGURE 2/H.261

Positioning of luminance and chrominance samples

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 quantized and variable length coded.

3.2.1 Prediction

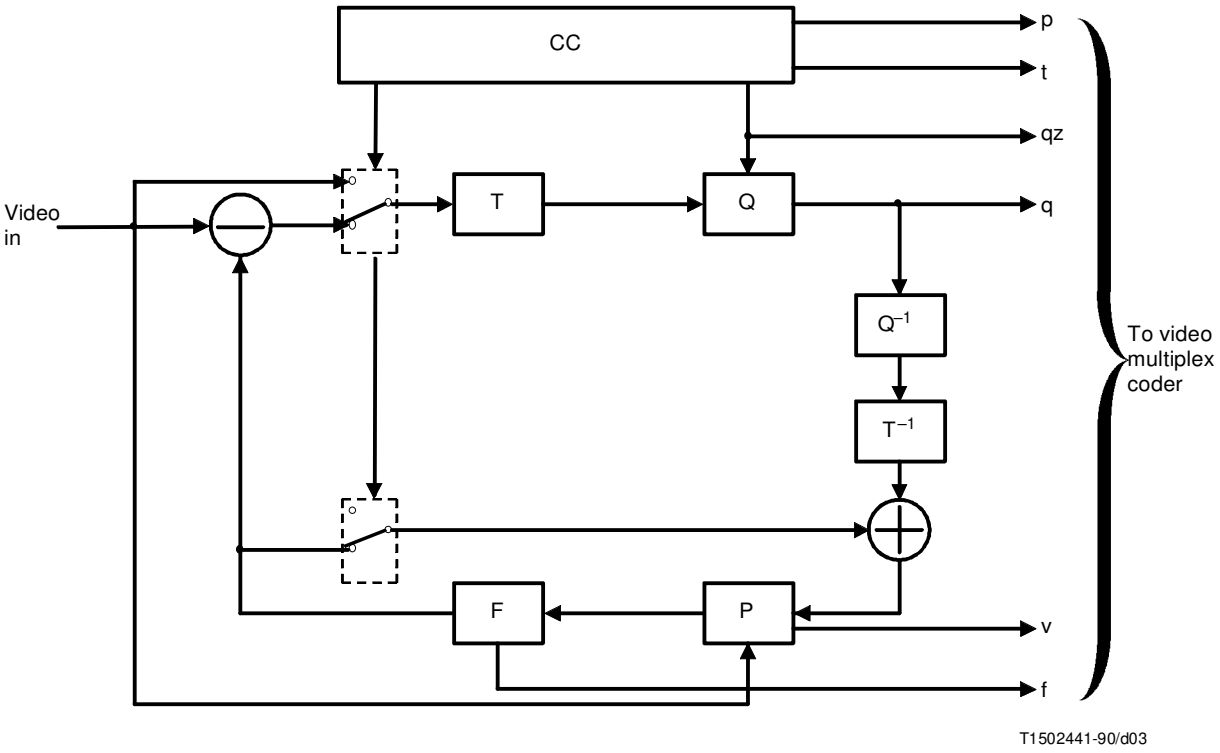
The prediction is inter-picture and may be augmented by motion compensation (see 3.2.2) and a spatial filter (see 3.2.3).

3.2.2 Motion compensation

Motion compensation (MC) 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.



T Transform
Q Quantizer
P Picture memory with motion compensated variable delay
F Loop filter
CC Coding control
p Flag for INTRA/INTER
t Flag for transmitted or not
qz Quantizer indication
q Quantizing index for transform coefficients
v Motion vector
f Switching on/off of the loop filter

FIGURE 3/H.261

Source coder

3.2.3 Loop filter

The prediction process may be modified by a two-dimensional spatial filter (FIL) 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 six blocks in a macroblock according to the macroblock type (see 4.2.3, MTYPE).

3.2.4 Transformer

Transmitted blocks are first processed by a separable two-dimensional discrete cosine transform of size 8 by 8. The output from the inverse transform ranges from –256 to +255 after clipping to be represented with 9 bits. The transfer function of the inverse transform is given by:

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

with $u, v, x, y = 0, 1, 2, \ldots, 7$

where x,y = spatial coordinates in the pel domain,

u,v = coordinates in the transform domain,

$C(u) = 1/\sqrt{2}$ for $u = 0$; otherwise 1,

$C(v) = 1/\sqrt{2}$ for $v = 0$; otherwise 1.

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 A.

3.2.5 Quantization

The number of quantizers is 1 for the INTRA dc coefficient and 31 for all other coefficients. Within a macroblock the same quantizer is used for all coefficients except the INTRA dc one. The decision levels are not defined. The INTRA dc coefficient is nominally the transform value linearly quantized with a stepsize of 8 and no dead-zone. Each of the other 31 quantizers is also nominally linear but with a central dead-zone around zero and with a step size of an even value in the range 2 to 62.

The reconstruction levels are as defined in 4.2.4.

NOTE – For the smaller quantization step sizes, the full dynamic range of the transform coefficients cannot be represented.

3.2.6 Clipping of reconstructed picture

To prevent quantization distortion of transform coefficient amplitudes causing arithmetic overflow in the encoder and decoder loops, clipping functions are inserted. The clipping function is applied 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 quantizer, block significance criterion and temporal sub-sampling. The proportions of such measures in the overall control strategy are not subject to recommendation.

When invoked, temporal sub-sampling 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 macroblock 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 Recommendation. Unless specified otherwise all unused or spare bits are set to "1". Spare bits must not be used until their functions are specified by the 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. Abbreviations are defined in later subclauses.

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.

4.2.1.1 Picture start code (PSC) (20 bits)

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

4.2.1.2 Temporal reference (TR) (5 bits)

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

4.2.1.3 Type information (PTYPE) (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;

Bit 5 Optional still image mode HI_RES defined in Annex D; "0" on, "1" off;

Bit 6 Spare.

4.2.1.4 Extra insertion information (PEI) (1 bit)

A bit which when set to "1" signals the presence of the following optional data field.

4.2.1.5 Spare information (PSPARE) (0/8/16 . . . bits)

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

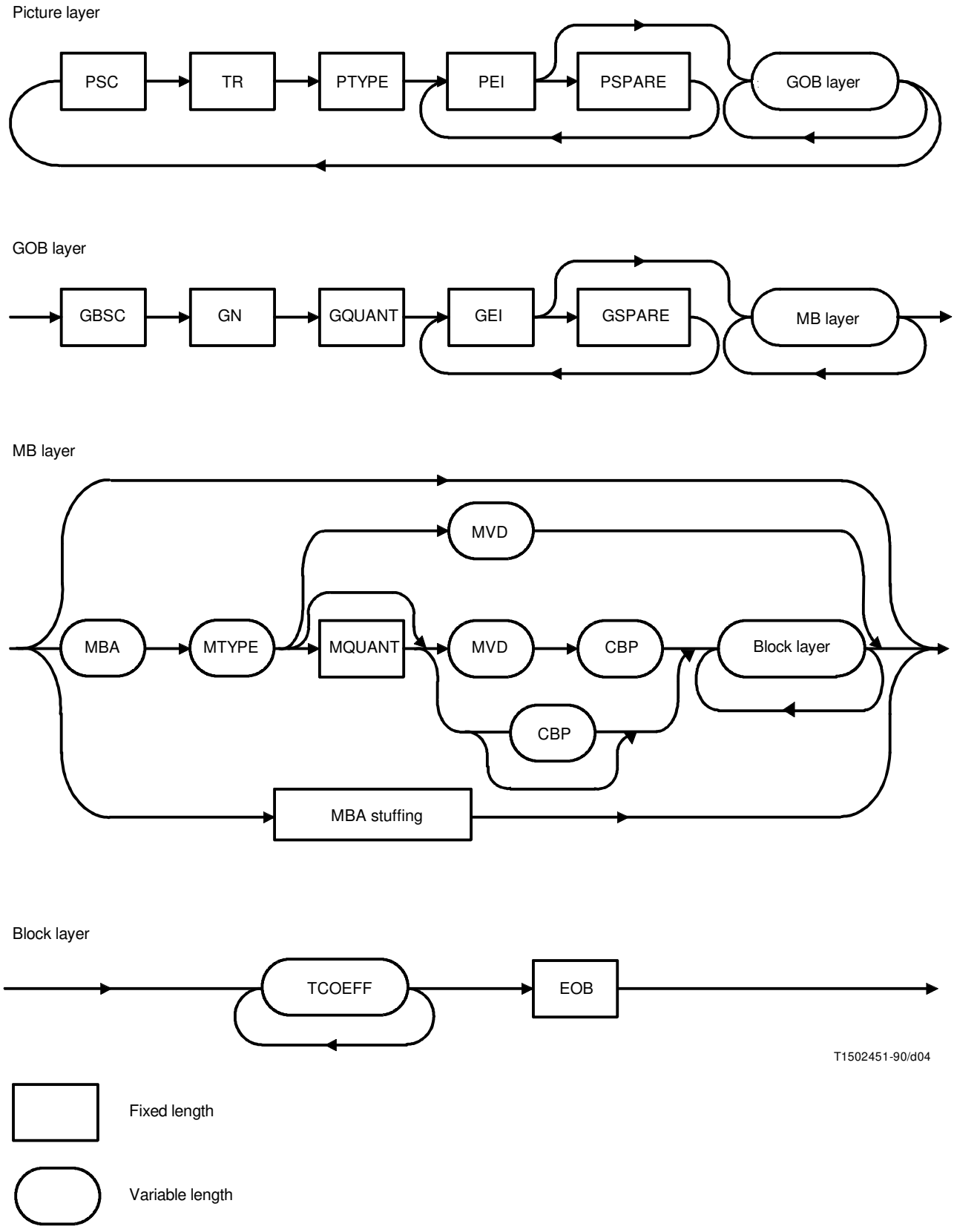


FIGURE 4/H.261

Syntax diagram for the video multiplex coder

FIGURE 5/H.261

Structure of picture layer

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_B and C_R.

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 in Figure 6, even if no macroblock data is present in that GOB.

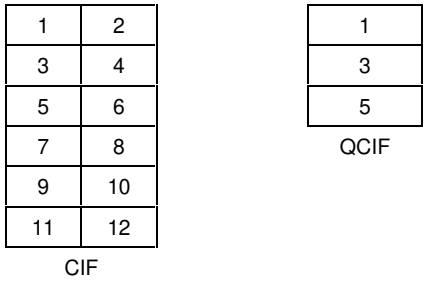


FIGURE 6/H.261

Arrangement of GOBs in a picture

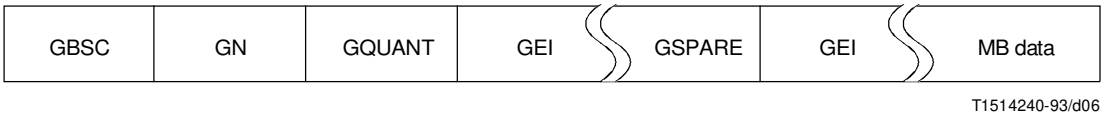


FIGURE 7/H.261

Structure of group of blocks layer

4.2.2.1 Group of blocks start code (GBSC) (16 bits)

A word of 16 bits, 0000 0000 0000 0001.

4.2.2.2 Group number (GN) (4 bits)

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

4.2.2.3 Quantizer information (GQUANT) (5 bits)

A fixed length codeword of 5 bits which indicates the quantizer to be used in the group of blocks until overridden by any subsequent MQUANT. The codewords are the natural binary representations of the values of QUANT (see 4.2.4) which, being half the step sizes, range from 1 to 31.

4.2.2.4 Extra insertion information (GEI) (1 bit)

A bit which when set to "1" signals the presence of the following optional data field.

4.2.2.5 Spare information (GSPARE) (0/8/16 . . . bits)

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

NOTE – Emulation of start codes may occur if the future specification of GSPARE has no restrictions on the final GSPARE data bits.

4.2.3 Macroblock layer

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

Data for a macroblock consists of an MB header followed by data for blocks (see Figure 9). MQUANT, MVD and CBP are present when indicated by MTYPE.

1	2	3	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	20	21	22
23	24	25	26	27	28	29	30	31	32	33

FIGURE 8/H.261

Arrangement of macroblocks in a GOB

MBA	MTYPE	MQUANT	MVD	CBP	Block data

FIGURE 9/H.261

Structure of macroblock layer

4.2.3.1 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 macroblock 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 1.

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 1.

MBA is always included in transmitted macroblocks.

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

TABLE 1/H.261

VLC table for 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

4.2.3.2 Type information (MTYPE) (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 2.

MTYPE is always included in transmitted macroblocks.

4.2.3.3 Quantizer (MQUANT) (5 bits)

MQUANT is present only if so indicated by MTYPE.

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

Codewords for MQUANT are the same as for GQUANT.

TABLE 2/H.261

VLC table for MTYPE

Prediction	MQUANT	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

NOTES

1 "x" means that the item is present in the macroblock.

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.

4.2.3.4 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) MTYPE 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 3.

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.

4.2.3.5 Coded block pattern (CBP) (Variable length)

CBP is present if indicated by MTYPE. 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:

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

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

The codewords for CBP are given in Table 4.

TABLE 3/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

4.2.4 Block layer

A macroblock comprises four luminance blocks and one of each of the two colour difference blocks (see Figure 10).

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

4.2.4.1 Transform coefficients (TCOEFF)

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

The most commonly occurring combinations of successive zeros (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, INTER+MC and INTER+MC+FIL blocks, the second for all other LEVELs except the first one in INTRA blocks which is fixed length coded with 8 bits.

TABLE 4/H.261

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		

FIGURE 10/H.261

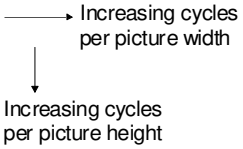
Arrangement of blocks in a macroblock

TCOEFF	EOB

FIGURE 11/H.261

Structure of block layer

1	2	6	7	15	16	28	29
3	5	8	14	17	27	30	43
4	9	13	18	26	31	42	44
10	12	19	25	32	41	45	54
11	20	24	33	40	46	53	55
21	23	34	39	47	52	56	61
22	35	38	48	51	57	60	62
36	37	49	50	58	59	63	64



T1 514100-93/d07

FIGURE 12/H.261

Transmission order for transform coefficients

Codes are given in Table 5.

The most commonly occurring combinations of zero-run and the following value are encoded with variable length codes as listed in the table 5. 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 and "1" for negative.

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. Use of this 20-bit word form encoding the combinations listed in the VLC table is not prohibited.

TABLE 5/H.261

VLC table for TCOEFF

Run	Level	Code
EOB		10
0	1	1s[a] If first coefficient in block
0	1	11s Not first coefficient in block
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

[a] Never used in INTRA macroblocks.

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
−128	FORBIDDEN
−127	1000 0001
.	.
−2	1111 1110
−1	1111 1111
0	FORBIDDEN
1	0000 0001
2	0000 0010
.	.
127	0111 1111

For all coefficients other than the INTRA dc one, the reconstruction levels (REC) are in the range −2048 to 2047 and are given by clipping the results of the following formuls:

$$\left.\begin{aligned} REC &= QUANT \cdot (2 \cdot level + 1);\ level > 0 \\ REC &= QUANT \cdot (2 \cdot level - 1);\ level < 0 \end{aligned}\right\} \quad QUANT = \text{“odd”}$$

$$\left.\begin{aligned} REC &= QUANT \cdot (2 \cdot level + 1) - 1;\ level > 0 \\ REC &= QUANT \cdot (2 \cdot level - 1) + 1;\ level < 0 \end{aligned}\right\} \quad QUANT = \text{“even”}$$

$$REC = 0;\ level = 0$$

NOTE – QUANT ranges from 1 to 31 and is transmitted by either GQUANT or MQUANT.

Reconstruction levels (REC)

	QUANT											
Level	1	2	3	4	·	8	9	·	17	18	·	30	31
−127	−255	−509	−765	−1019	·	−2039	−2048	·	−2048	−2048	·	−2048	−2048
−126	−253	−505	−759	−1011	·	−2023	−2048	·	−2048	−2048	·	−2048	−2048
.	.	.	.	.	.	.	.	.	.	.	.	.	.
−2	−5	−9	−15	−19	·	−39	−45	·	−85	−89	·	−149	−155
−1	−3	−5	−9	−11	·	−23	−27	·	−51	−53	·	−89	−93
0	0	0	0	0	·	0	0	·	0	0	·	0	0
1	3	5	9	11	·	23	27	·	51	53	·	89	93
2	5	9	15	19	·	39	45	·	85	89	·	149	155
3	7	13	21	27	·	55	63	·	119	125	·	209	217
4	9	17	27	35	·	71	81	·	153	161	·	269	279
5	11	21	33	43	·	87	99	·	187	197	·	329	341
.	.	.	.	.	.	.	.	.	.	.	.	.	.
56	113	225	339	451	·	903	1017	·	1921	2033	·	2047	2047
57	115	229	345	459	·	919	1035	·	1955	2047	·	2047	2047
58	117	233	351	467	·	935	1053	·	1989	2047	·	2047	2047
59	119	237	357	475	·	951	1071	·	2023	2047	·	2047	2047
60	121	241	363	483	·	967	1089	·	2047	2047	·	2047	2047
.	.	.	.	.	.	.	.	.	.	.	.	.	.
125	251	501	753	1003	·	2007	2047	·	2047	2047	·	2047	2047
126	253	505	759	1011	·	2023	2047	·	2047	2047	·	2047	2047
127	255	509	765	1019	·	2039	2047	·	2047	2047	·	2047	2047

NOTE – Reconstruction levels are symmetrical with respect to the sign of level except for 2047/−2048.

For INTRA blocks the first coefficient is nominally the transform dc value linearly quantized with a step size 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 6).

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 freeze picture release signal is received or a timeout period of at least six seconds has expired. The transmission of this signal is via external means (for example by Recommendation 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 Recommendation H.221).

TABLE 6/H.261

Reconstruction levels for INTRA-mode dc coefficient

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 8*n* except FLC 255 gives 1024.

4.3.3 Freeze picture release

A signal from an encoder which has responded to a fast update request and allows a decoder to exit from its freeze picture mode and display decoded pictures in the normal manner. This signal is transmitted by bit 3 of PTYPE (see 4.2.1) in the picture header of the first picture coded in response to the fast update request.

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 to comply with the requirements of the hypothetical reference decoder defined in Annex B.

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. The bit count does not include error correction framing bits, fill indicator (Fi), fill bits or error correction parity information described in 5.4.

Video data must be provided on every valid clock cycle. This can be ensured by the use of either the fill bit indicator (Fi) and subsequent fill all 1's bits in the error corrector block framing (see Figure 13) or MBA stuffing (see 4.2.3) or both.

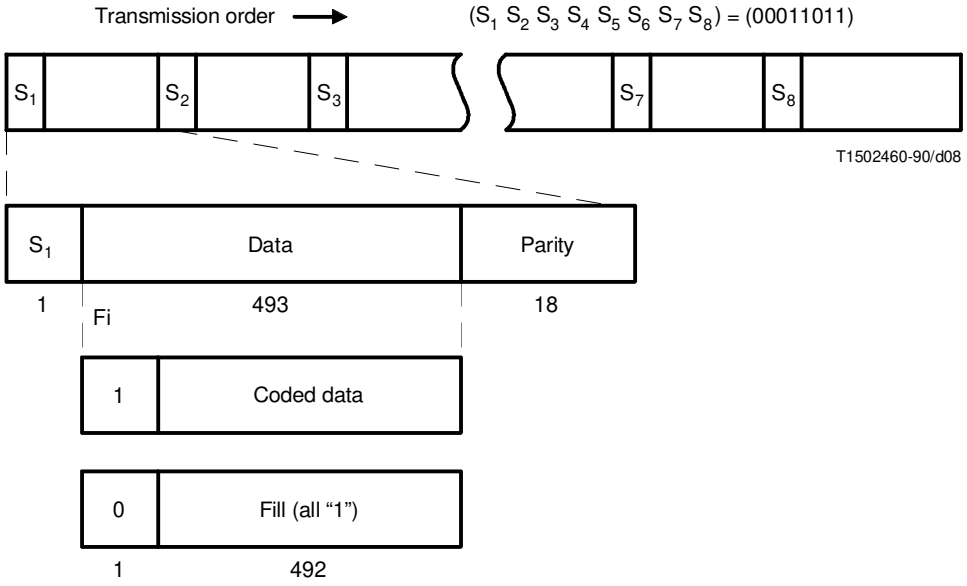


FIGURE 13/H.261

Error correcting frame

5.3 Video coding delay

This item is included in this Recommendation because the video encoder and video decoder delays need to be known to allow audio compensation delays to be fixed when H.261 is used to form part of a conversational service. This will allow lip synchronization to be maintained. Annex C recommends a method by which the delay figures are established. Other delay measurement methods may be used but they must be designed in a way to produce similar results to the method given in Annex C.

5.4 Forward error correction for coded video signal

5.4.1 Error correcting code

The transmitted bitstream 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

To allow the video data and error correction parity information to be identified by a decoder an error correction framing pattern is included. This consists of a multiframe of eight frames, each frame comprising 1 bit framing, 1 bit fill indicator (Fi), 492 bits of coded data (or fill all 1s) and 18 bits parity. The frame alignment pattern is:

$$(S_1S_2S_3S_4S_5S_6S_7S_8) = (00011011).$$

See Figure 13 for the frame arrangement. The parity is calculated against the 493-bits including fill indicator (Fi).

The fill indicator (Fi) can be set to zero by an encoder. In this case only 492 consecutive fill bits (fill all 1s) plus parity are sent and no coded data is transmitted. This may be used to meet the requirement in 5.2 to provide video data on every valid clock cycle.

5.4.4 Relock 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 34 000 bits after an error corrector framing phase change.

NOTE – This assumes that the video data does not contain three correctly phased emulations of the error correction framing sequence during the relocking period.

Annex A

Inverse transform accuracy specification

(This annex forms an integral part of this Recommendation)

A.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 set of 10 000 blocks should each be generated for (L = 256, H = 255), (L = H = 5) and (L = H = 300).

A.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.

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

with $u, v, x, y = 0, 1, 2, \ldots, 7$

where x,y = spatial coordinates in the pel domain,

u,v = coordinates in the transform domain,

$C(u) = 1/\sqrt{2}$ for u = 0; otherwise 1,

$C(v) = 1/\sqrt{2}$ for v = 0; otherwise 1.

A.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.

A.4 For each 8 by 8 block of 12-bit data produced by A.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 × 8 pels are the reference IDCT input data.

A.5 For each 8 by 8 block produced by A.3, apply the IDCT under test and clip the output to the range –256 to +255. These blocks of 8 × 8 pels are the test IDCT output data.

A.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.

A.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 error should not exceed 0.015 in magnitude.

Overall, the mean error should not exceed 0.0015 in magnitude.

A.8 All zeros in must produce all zeros out.

A.9 Re-run the measurements using exactly the same data values of A 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 * 1103515245) + 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 */
}
```

Annex B

Hypothetical reference decoder

(This annex forms an integral part of this Recommendation)

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

B.1 The HRD and the encoder have the same clock frequency as well as the same CIF rate, and are operated synchronously.

B.2 The HRD receiving buffer size is (B + 256 kbits). The value of B is defined as follows:

B = $4R_{max}/29.97$ where R_{max} is the maximum video bit rate to be used in the connection.

B.3 The HRD buffer is initially empty.

B.4 The HRD buffer is examined at CIF intervals (≈ 33 ms). If at least one complete coded picture is in the buffer then all the data for the earliest picture is instantaneously removed (e.g. at t_{n+1} in Figure B.1). Immediately after removing the above data the buffer occupancy must be less than B. This is a requirement on the coder output bitstream including coded picture data and MBA stuffing but not error correction framing bits, fill indicator (Fi), fill bits or error correction parity information described in 5.4.

To meet this requirement the number of bits for the (n+1)th coded picture d_{n+1} must satisfy:

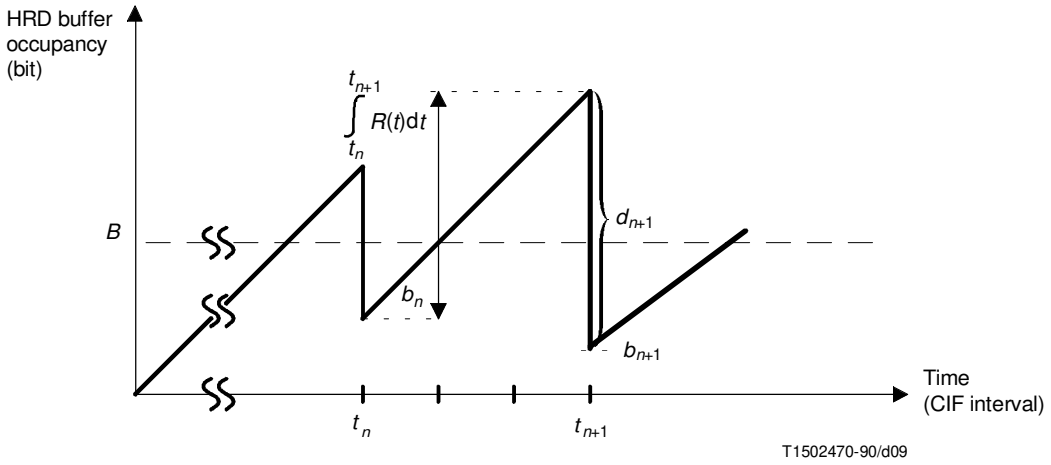
$$d_{n+1} \geq b_n + \int_{t_n}^{t_{n+1}} R(t)dt - B$$

where

b_n is buffer occupancy just after the time t_n;

t_n is the time the nth coded picture is removed from the HRD buffer;

$R(t)$ is the video bit rate at the time t.



NOTE – Time $(t_{n+1} - t_n)$ is an integer number of CIF picture periods (1/29.97, 2/29.97, 3/29.97, ...).

FIGURE B.1/H.261

HRD buffer occupancy

Annex C

Codec delay measurement method

(This annex forms an integral part of this Recommendation)

The video encoder and video decoder delays will vary depending on implementation. The delay will also depend on the picture format (QCIF, CIF) and data rate in use. This annex specifies the method by which the delay figures are established for a particular design. To allow correct audio delay compensation the overall video delay needs to be established from a user perception point of view under typical viewing conditions.

Point A is the video input to the video coder. Point B is the channel output from the video terminal (i.e. including any FEC, channel framing, etc.). Point C is the video output from the decoder.

A video sequence lasting more than 100 seconds is connected to the video coder input (point A) in Figure C.1 above. The video sequence should have the following characteristics:

- it should contain a typical moving scene consistent with the intended purpose of the video codec;
- it should produce a minimum coded picture rate of 7.5 Hz at the bit rate in use;
- it should contain a visible identification mark at intervals throughout the length of the sequence. The visible identification should change every 97 video input frames and be located within the picture area represented by the first GOB in the picture. For example, the first block in the picture could change from black to white at intervals of 97 video frame periods. The identification mark should be chosen so that it can be detected at point B and does not significantly contribute to the overall coding performance.

The codec and video sequence should be arranged so that the bitstream contains less than 10% stuffing (MBA stuffing + error correction fill bits).

The encoder delay is obtained by measuring the time from when the visible identification changes at point A to the time that the change is detected at point B. Similarly, the decoder delay is obtained by taking measurements at points B and C.

Several measurements should be made during the sequence length and the average period obtained. Several tests should be made to ensure that a consistent average figure can be obtained for both encoder and decoder delay times.

Average results should be obtained for each combination of picture format and bit rate within the capability of the particular codec design.

NOTE – Due to pre- and post-temporal processing it may be necessary to take a mid-level for establishing the transition of the identification mark at points B and C.

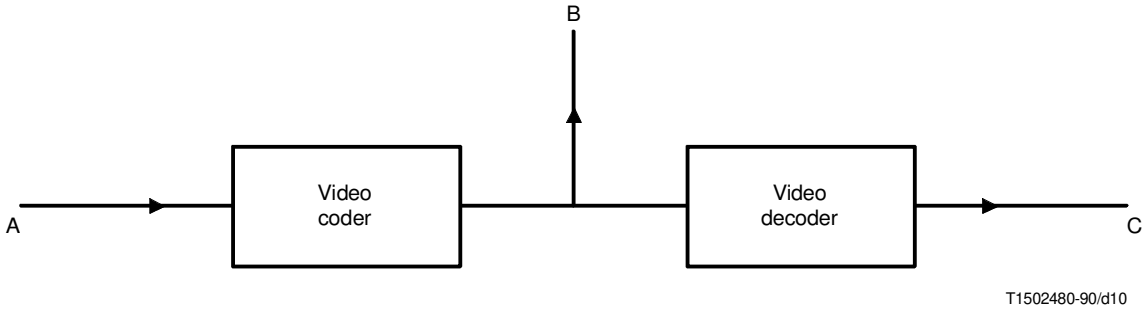


FIGURE C.1/H.261

Measuring points

Annex D

Still image transmission

(This annex forms an integrat part of this Recommendation)

D.1 Introduction

This annex describes the procedure for transmitting still images within the framework of this Recommendation. This procedure enables an H.261 video coder to transmit still images at four times the normal video resolution by temporarily stopping the motion video. Administrations may use this optional procedure as a simple and inexpensive method to transmit still images. However, Recommendation T.81 (JPEG) is preferred when the procedures for using T.81 within audiovisual systems are standardized.

This procedure can provide high quality image transmission with effects similar to those of progressive and hierarchical schemes. Minimal changes to H.261 (low cost), backward compatibility with existing terminals, and flexibility in image quality versus transmission speed were the key considerations in its development.

NOTE – The encoder would set a previously unused bit in PTYPE to "0" when it transmits a still image (unused bits should be set to "1"). A decoder that ignores this bit would receive the image as normal video. A decoder that goes into an error condition when this bit is "0" would most likely freeze the previous video frame, and resume when this bit is reset to "1". A decoder having this new capability could display the image in a higher resolution, transfer the image to a separate graphics display and hold the image when video resumes, print and/or save the image, etc.

D.2 Still image format

The still image format is four times the currently transmitted video format. If the video format is QCIF, then the still image is a CIF frame. If the video format is CIF, which contains 352 × 288 luminance samples, then the still image contains 704 × 576 luminance samples, and a corresponding increase in the number of chrominance samples (a CCIR-601 frame).

For transmission using H.261, the still image is sub-sampled 2:1 horizontally and vertically into four sub-images in the currently transmitted video format. Figure D.1 shows the sub-sampling pattern on the still image. The samples labelled 0, 1, 2 and 3 form the four sub-images 0, 1, 2 and 3, respectively.

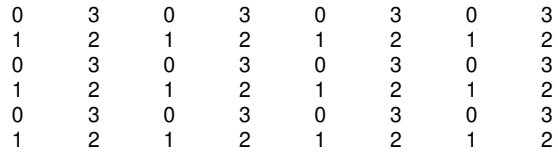


FIGURE D.1/H.261

Sub-sampling pattern

D.3 Picture layer multiplex

When HI_RES is "0", the two lower bits of the temporal reference (TR) identify one of the four sub-images 0, 1, 2 or 3. The three higher bits of the TR shall be set to "0".

The encoder transmits a still image by setting HI_RES to "0" and transmitting the four sub-images 0, 1, 2 and 3 in sequential order. It is allowed to transmit more than one frame for each sub-image, but should not go back once it starts transmitting the next sub-image. The encoder is allowed to resume motion video at any time by setting HI_RES back to "1".

NOTE – The reference memory for the current frame is always the previous frame, regardless of whether a frame is motion video or still image.

D.4 Multipoint considerations

A still image transmitted within the video bit-stream can be broadcast on a multipoint connection by broadcasting the video. The MCV (multipoint command visualization-forcing) and Cancel-MCV commands defined in Recommendation H.230 provide for this capability. A terminal could force an MCU to broadcast its video by sending MCV, and then return to the previous mode of operation by sending Cancel-MCV. MCUs are required to implement these commands, but they are optional for terminals.

D.5 Other considerations

- All the video coding modes are allowed (intra-frame, inter-frame, motion compensation, etc.);
- the multiplex arrangement below the picture layer remains the same (group of blocks, macroblocks, etc.);
- the maximum number of bits allowed per frame (sub-image) should not be exceeded (256 kbit/s for CIF and 64 kbit/s for QCIF);
- forward error correction is not affected.