UNIVERSITÀ DEGLI STUDI
DI MODENA E REGGIO EMILIA

Lecture notes for
Multimedia Data Processing

H.261 Standard

Last updated on: 27/03/2020

H.261 Standard

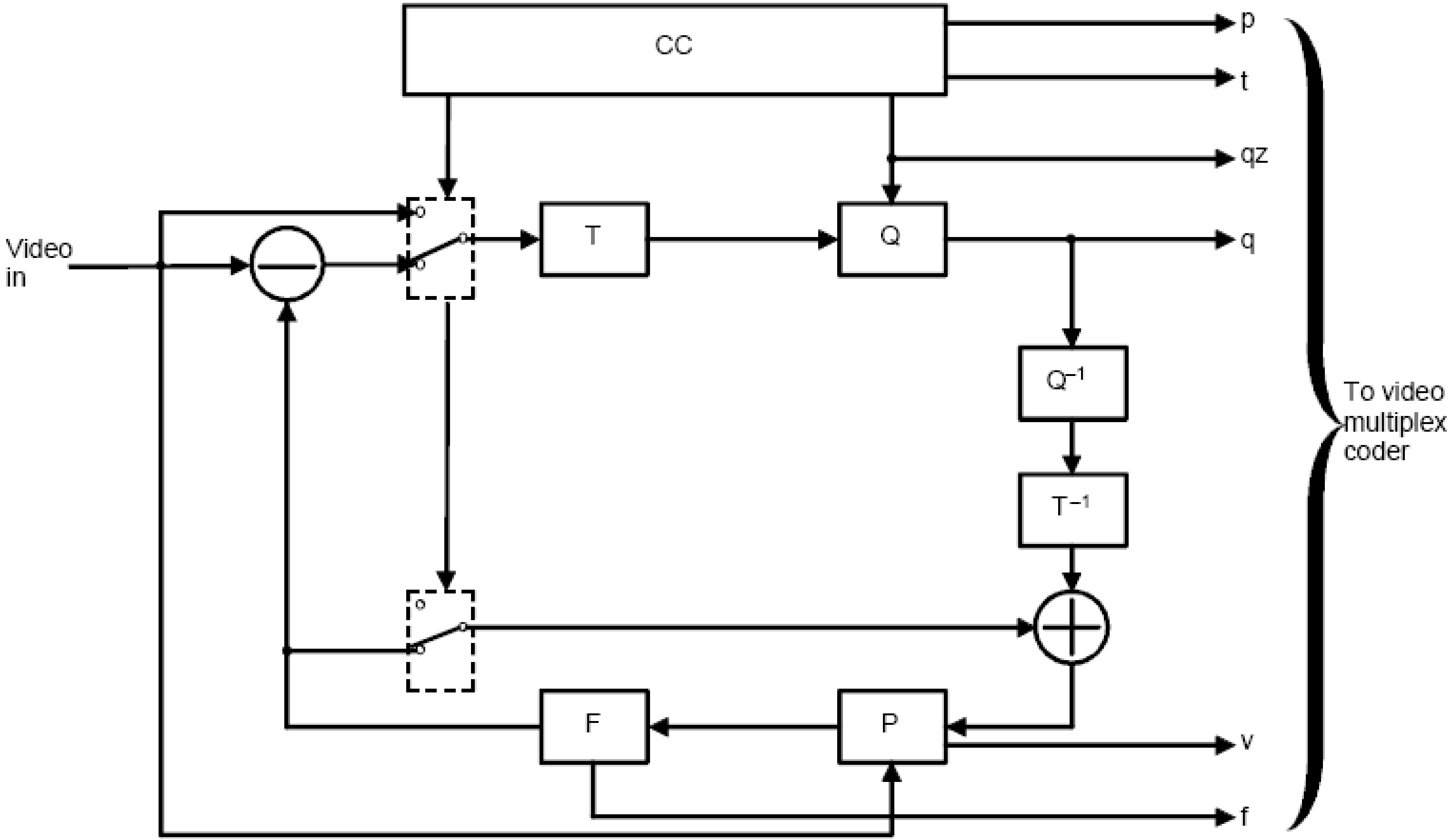
- The ITU-T Recommendation H.261 standard (introduced in 1990) is the first "modern" standard for video encoding which had sufficient spread, and which introduced some concepts used by all subsequent standards.
- The standard title is "Video codec for audiovisual services at p x 64 kbit/s". It was originally designed for transmission over ISDN lines, the first digital lines available for the public. Each ISDN line allowed bi-directional traffic at 64 kbps, therefore by using several of them in parallel it was possible to obtain a band of "$p \times 64$ kbits", with p the number of lines (max 30).
- Video telephony was the goal of the standard, so the whole standard ignores the use of reference points for random access search on a stream (file), unlike the subsequent standards.
- The standard supports two video frame sizes: Common Interchange Format (CIF) 352×288 pixels and quarter-CIF (QCIF) 176×144 pixels.
- The term Common refers to the fact that both PAL and NTSC analog signals can be easily sampled at this resolution.

Source Format

- The images are coded as one luminance component (Y) and two color difference components Cb and Cr.
- The Y component should go from 16 (black) to 235 (white), while Cb and Cr should be between 16 and 240, with 128 zero difference value. All of them can be represented with 8 bits. These values are nominal and everything works even with values between 1 and 254.
- The chrominance components are subsampled by half in width and height, therefore every 4 values of Y there are 1 value of Cb and 1 of Cr (subsampling 4:2:0 - details on http://en.wikipedia.org/wiki/Chroma_subsampling).
- As mentioned in Motion Vector, each 16×16 pixels macro block contains four 8×8 blocks of Y and one 8×8 block respectively for Cb and Cr.

Block Diagram of H.261 Encoder

- T: Transform, Q: Quantizer, P: Predicted image memory, F: Loop filter, CC: Coding control, p: Flag for INTRA/INTER, t: Flag per transmission or not, qz: Level of quantization, q: Quantized coefficients, v: Motion vector, f: Flag for ON/OFF of the loop filter.



Differences From the Block Diagram

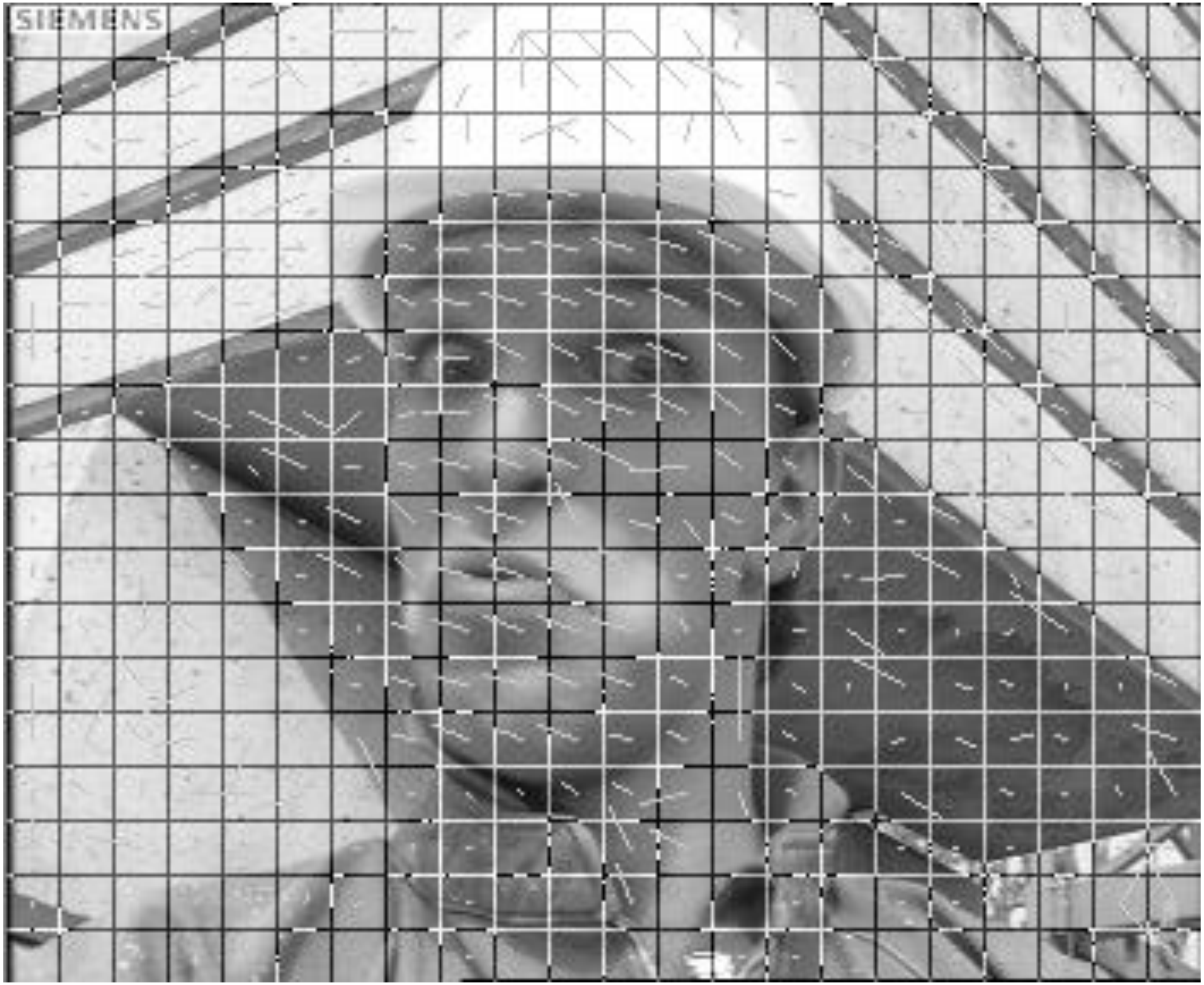
- The H.261 does not work with entire frames, but the fundamental unit to which the previous scheme is applied is the Macroblock.
- **Therefore, there are no INTRA or INTER frames!**
- In each frame there can be INTRA or INTER macroblocks. Furthermore, some macroblocks (or part of them) transmission can be avoided if these are sufficiently similar to those of the previous frame.
- There is also a *loop* filter which, after the creation of the aforementioned image, allows you to "smooth" the variations within the block. The filter is low-pass with values that can be easily implemented with integer arithmetic and shifts. It is separable, that is, the block is first filtered horizontally with coefficients (¼, ½, ¼) and then filtered again vertically. If the convolution leaves the block, only the edge value is used.
- The two-dimensional version is represented by the matrix:
- Very close to a Gaussian filter with $\sigma = 0.8$.

1/16	1/8	1/16
1/8	1/4	1/8
1/16	1/8	1/16

Previous Frame ($Q = 64$)

Original Image

Motion Vector



Predicted Image

Predicted Image and Loop Filter

Further Details

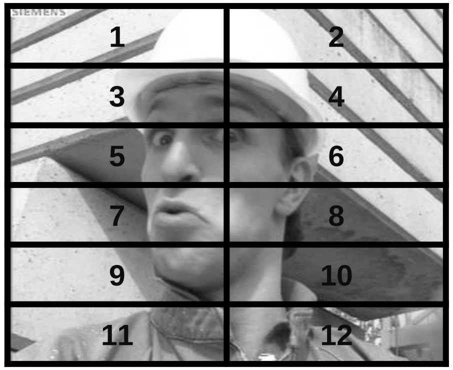
- The introduction of the loop filter has a very low computational cost and allows to reduce the "blocky" effect due to compression. This allows you to slightly reduce the differences. In the previous example the differences (without DCT) had an entropy of 5.14 which becomes 5.08 with the loop filter. After DCT encoding and quantization with Q = 32, entropy is 0.53 which becomes 0.50 with the loop filter. Nothing exceptional, but still contributes to the result.
- The standard also provides that you can skip up to 3 frames per transmitted frame to reduce the frame rate.
- In addition, to prevent the overall image quality from being excessively low, each macroblock must be transmitted INTRA at least once every 132 times it is transmitted.
- There is no indication on when to transmit INTRA or INTER, nor on the level of quantization to use or how to choose it, nor on when a block may not be transmitted.

Bitstream Structure

- The fields are bit-coded and are transmitted with the most significant bit first.
- The structure of the stream is divided into 4 levels or layers: Picture, Group of blocks (GOB), Macroblock (MB) and Block. Each level is usually preceded by a header, except for the block.
- The Picture layer is composed of:
  - Picture start code (PSC) of 20 bits: is 0000 0000 0000 0001 0000. Note that this **is not byte aligned**. Therefore, a search for a PSC requires scan it bit by bit.
  - Temporal reference (TR) of 5 bits: a value from 0 to 31 which is increased by one each time and which tracks any skipped frames.
  - Type information (PTYPE) of 6 bits: a bitmapped field, in which only the bit 4 is of interest, since it is 0 if the picture is QCIF or 1 if it is CIF.
  - Extra Insertion information (PEI) of 1 bit: if it is 1 it indicates that there is a byte of additional information (PSPARE) to follow. The standard says not to insert them and to ignore those that meet, until they are defined (which will never happen by now). Continue with PEI until this is 0.

Group of Blocks Layer

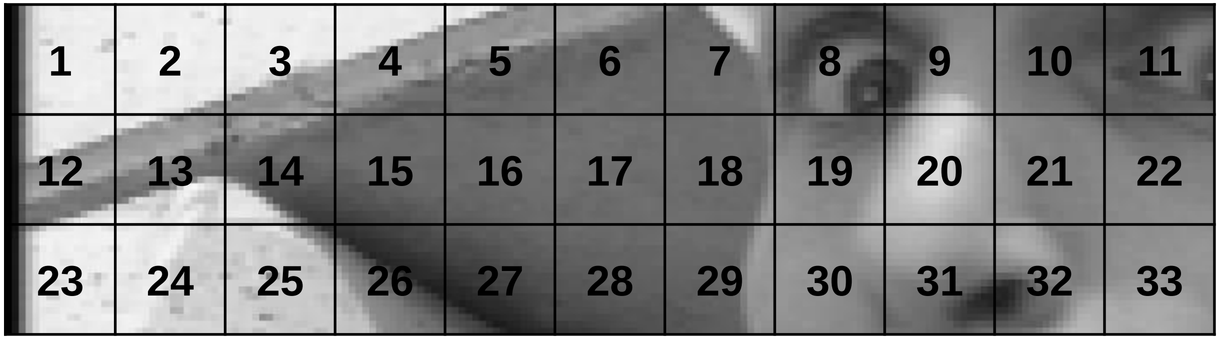
- Each Picture is divided into 12 GOB as in the figure (for the CIF, while for the QCIF they are only 1,3,5).



- Each GOB layer is composed of:
  - 16-bit Group of Blocks Start Code (GBSC): the value is 0000 0000 0000 0001.
  - 4-bit Group Number (GN): a number indicating the GOB number, as in the figure. The 0 corresponds to the PSC, while 13,14,15 are not used.
  - 5-bit Quantizer information (GQUANT): indicates the quantization value to use for all blocks in the GOB.
  - Extra insertion information (GEI) 1 bit: with the same mechanism as before it is also possible to add information (GSPARE) here, but the specification has never been completed.

Macroblock Layer

- The GOB is divided into 11×3 macroblocks of size 16×16.



- Each macroblock starts with a Macroblock address (MBA) which is coded at variable length (variable length code or VLC).
- For the first MB transmitted, its absolute address is sent (from 1 to 33), while for subsequent ones the difference between the current address and that of the previously transmitted MB is sent (prediction of the MBA).
- At this stage it is possible that, instead of meeting the VLC of an MBA, a start code (PSC or GBSC) is encountered. This means that the current GOB has ended (it may not even contain macroblocks).
- There is also a VLC that allows you to "enlarge" the bit stream, for example if you want to obtain exactly a certain bit rate, in cases where compression reduces it excessively (obviously this is used very rarely).

VLC Table for MBAs

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

Macroblock Layer

- The second variable length field is the *Type information* (MTYPE), which provides information on the macroblock and the elements present.

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.

Macroblock Layer

- Following the MTYPE, the following fields may be present:
  - Quantizer (MQUANT) of 5 bit: indicates that from here on the quantization value must be the one indicated.
  - Motion vector data (MVD): if a block is of the MC type, two VLCs are provided, one for the horizontal component and one for the vertical component. The values are the difference from the previous one. The previous one is considered (0,0) if we are at 1,12,23 MB or if the previous one was skipped (MBA>1), or if the previous one was not MC. Furthermore, a further optimization is carried out: the difference between the components of two MVs (which are limited between ± 15) lies in the range ± 30. However, knowing the value of the previous component, only 31 of the 61 values are possible. Using this observation, in the VLC table for MVDs, two different values correspond to the same VLC. The correct value is the one that makes the component fall within the range ± 15.

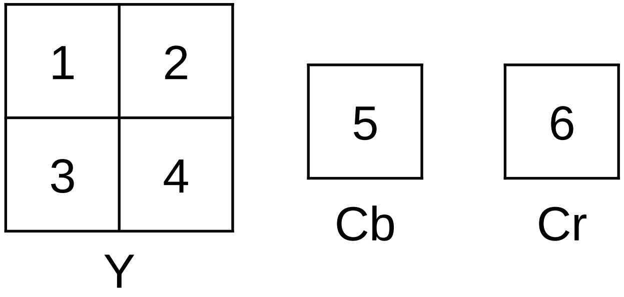
    Example: the component x for the previous MV is -5 and VLC 0000 0111 is received. Looking at the table on the next page, this corresponds to -7 & 25. However -5 + 25 = 20, which is an impossible value, therefore -5 + (- 7) = -12 is the value for the x component of the current MV.

VLC Table for MVDs

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

Macroblock Layer

- The last field of the Macroblock layer is the *Coded Block Pattern* (CBP). This field is a 6-bit bitmap indicating which blocks are present in the MB. The blocks are numbered like this:



- The value 60, for example, in binary is 111100 and this means that blocks 1,2,3,4 are present, but not the 5 and the 6 (that is, there are only the luminance components and not the chrominance ones). Note that the bits are numbered from left to right starting at 1.
- Each CBP is associated with a VLC in order to further compress this field too.

VLC table for CBPs

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		

Block Layer

- In the Block layer the coefficients of the transform (TCOEFF) of the blocks indicated in the CBP are coded according to the zig-zag order. If the MB is INTRA, all the blocks are present.
- All the coefficients of the transforms, both INTRA and INTER, except the first one, are represented as a pair (run,level), which indicates the number of null coefficients preceding the non-zero level that is being coded. Typically the pairs (run,level) are present in the following table. Those that are not present are encoded with a 20-bit word: escape, run, level. *escape* is code 0000 01, *run* is a 6-bit unsigned integer and *level* is an 8-bit two's complement signed integer. Obviously, *level* cannot be zero.
- At the end of each block there is an *end of block* (EOB) 10 code, which indicates that all the other coefficients are null.
- The first coefficient for INTRA blocks is simply coded with 8 bits, except for the value 128 (1000 0000 in binary) which is replaced by 255 (1111 1111) in order not to create false start codes.
- The first coefficient for INTER blocks uses the same table as the others, but since it cannot be an EOB, it shortens the combination by one bit (run=0, level=± 1) (see table).

VLC table for TCOEFFs

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.

The last bit "s" indicates the level sign: "0" if positive, "1" if negative.

Quantization

- Quantization in H. 261 is defined in terms of reconstruction, i.e. the inverse operation that must be performed to switch from quantized to dequantized (reconstructed) values. The formula is as follows:

$$\left.\begin{array}{l} REC = QUANT \cdot (2 \cdot level + 1); level > 0 \\ REC = QUANT \cdot (2 \cdot level - 1); level < 0 \end{array}\right\} QUANT \text{ is odd}$$

$$\left.\begin{array}{l} REC = QUANT \cdot (2 \cdot level + 1) - 1; level > 0 \\ REC = QUANT \cdot (2 \cdot level - 1) + 1; level < 0 \end{array}\right\} QUANT \text{ is even}$$

$$REC = 0; level = 0$$

- The $2 \cdot level$ in the reconstruction formulas is present because the DCT coefficients are divided by 2×QUANT during the coding. So the quantization goes from 2 to 62 (QUANT is always between 1 and 31).

- A simpler fixed division by 8 is used for the DC coefficient of the INTRA blocks.