

AVC-163

INTERNATIONAL STANDARDIZATION ORGANIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC2/WG11
CODING OF MOVING PICTURES AND ASSOCIATED AUDIO

ISO/IEC JTC1/SC2/WG11
MPEG 91/227
October 1991

Source: Jean-François VIAL (THOMSON-CSF/LER)
Title: Video coding algorithm proposal #34 for MPEG-2

Purpose: Proposal

Groups: All

Status: Contribution to the 16th meeting

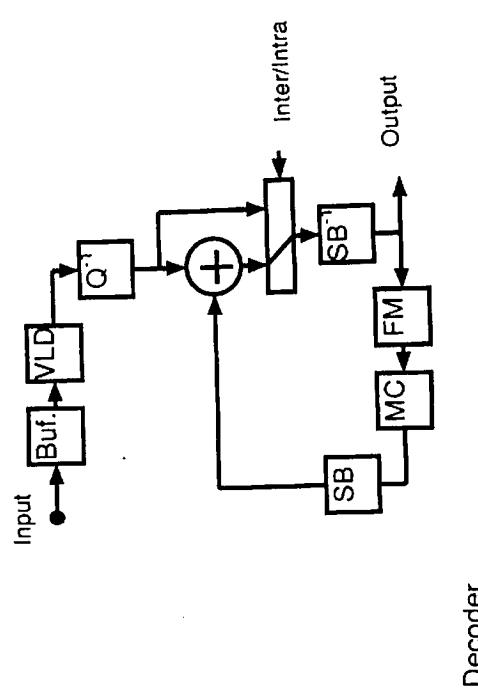
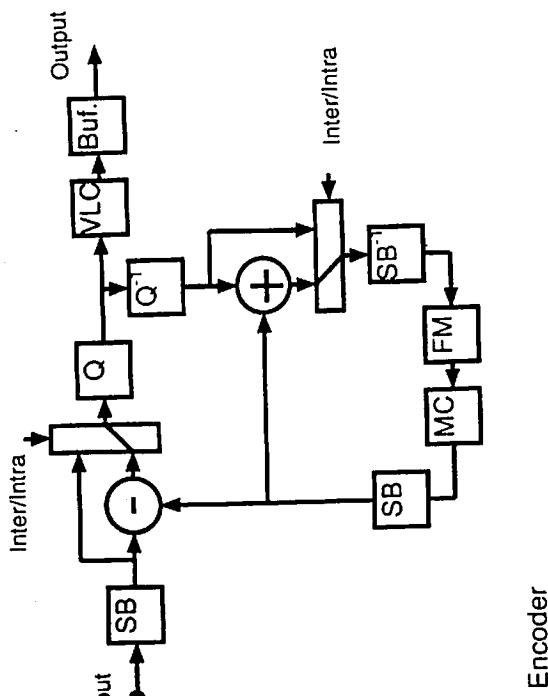
Proposal registration number: 34

This proposal is submitted by THOMSON-CSF/LER.

*It has been developed within the European contract RACE 1018 HIVITS, and
presented in the European EUREKA 625 VADIS / COST 211ter collaboration.*

VIDEO INPUT	Standards	HDP or HDI or EDP or (16/9) 4:2:2 or (4/3) 4:2:2 or 16/9 (VT)	
	Coding (according to EU 95)	HDP : 1920 pixels x 1152 lines HDI : 1920 pixels x 1152 lines EDP : 960 pixels x 576 lines 16/9 4:2:2: 960 pixels x 576 lines 4/3 4:2:2: 720 pixels x 288 lines 16/9 VT : 480 pixels x 288 lines	50 frames/s non interlaced 50 fields/s (interlaced by 2) 50 frames/s 50 fields/s (interlaced by 2) 50 frames/s 50 frames/s
	Interface	Bit parallel or bit serial according to EU95	
VIDEO OUTPUT	Standards	HDP or HDI or EDP or (16/9) 4:2:2 or (4/3) 4:2:2 or 16/9 VT	
	Coding (according to EU95)	HDP HDI 16/9 4:2:2 4/3 4:2:2 16/9 VT	
	Interface	Bit parallel or bit serial according to EU95	
SIGNAL PRE PROCESSING	Interlaced Standards	The interlaced signals (HDI and 4:2:2) require to be processed into progressive like signals before encoding	
	Subsampling	No subsampling of the progressive input signal is performed	
SIGNAL POSTPROCESSING	Interlaced Standards	Displaying the decoded information according to an interlaced standard requires progressive to interlace conversion after decoding.	

PAGE 3



CODING	Signal splitting	The signal is split into subbands according to Annex I before encoding.
	Organisation into blocks	The signal obtained in each subband is organised into blocks consisting of different number of pixels according to the subband considered (Annex II).
	Encoding modes	2 modes are used to encode the blocks (intraframe or interframe motion compensated modes) (Annex III).
	Prediction of the block	In each block processed according to interframe mode the reference block is taken from a previous frame. Its position is derived from the position of the current block by application of a displacement vector (Annex IV)
	Motion compensation	Motion compensation is applied to blocks. A single displacement vector is assigned to each block (see Annex V)
	Quantization	A different quantization characteristic is used for each subband. Its parameters are adapted to the buffer occupancy (See Annex VI).
	Variable lenght coding	VLC's are used to encode the quantized subband pixels and motion information (See Annex VII).
	Video framing	See Annex VIII.
REGULATION		For compatibility with storage some restrictions are introduced in the regulation strategy (See Annex IX).

ANNEX I - SIGNAL SPLITTING

I-1 - SUBBAND FILTERS

Subband splitting is done according to a hierarchical structure based on the specification of an unique elementary cell. Such cells are arranged in a tree structure and some parts of the spectrum may be more or less split according to their informational content.

This elementary cell splits the input spectrum in four rectangular bands using a separable approach : Rows and columns are treated separately.

This cell and the corresponding spectrum decomposition are sketched in figure 1.

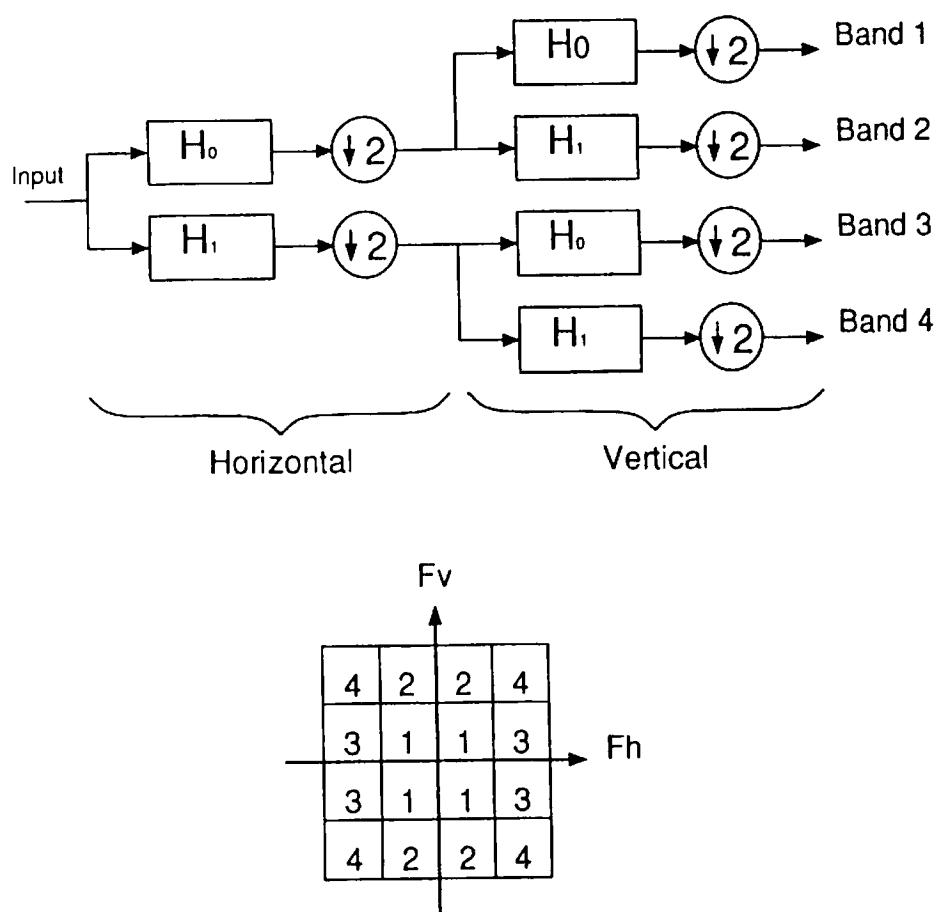


FIGURE 1 - 2D BASIC CELL - SEPARABLE 4-BAND DECOMPOSITION

PROPOSAL 34 FOR MPEG-2

Columns of the input image are filtered in the first stage (followed by subsampling of every second row). The input spectrum is thus first split into two horizontal subbands. Rows are then processed in the second stage.

H_0 and H_1 are respectively a low-pass and a high pass filter, having the same even length L. Coefficients $h_1(n)$ are deduced from H_0 as follows :

$$h_1(n) = (-1)^{L-n} h_0(L-1-n) \text{ for } n = 0 \text{ to } L-1$$

H_0 is a 16-tap linear phase filter (Quadrature Mirror Filter) whose coefficients are listed in table 1.

$h_0(0) =$	0.001050	$= h_0(15)$
$h_0(1) =$	-0.005055	$= h_0(14)$
$h_0(2) =$	-0.002590	$= h_0(13)$
$h_0(3) =$	0.027641	$= h_0(12)$
$h_0(4) =$	-0.009666	$= h_0(11)$
$h_0(5) =$	-0.090392	$= h_0(10)$
$h_0(6) =$	0.097798	$= h_0(9)$
$h_0(7) =$	0.481028	$= h_0(8)$

TABLE 1 : SUBBAND FILTER h_0

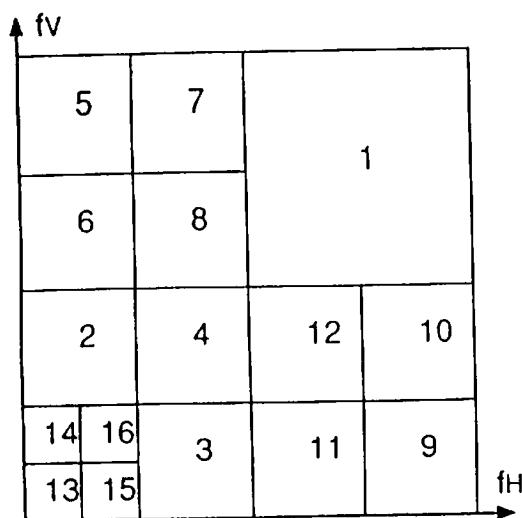
II - SPLITTING ("FORMAT INDEPENDENT SPLITTING" - FIS - SYSTEM)

II-1 - Progressive input signal

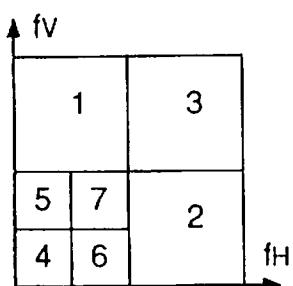
Whatever its format (HDP, EDP or VT) a progressive input image is split in the same way.

A) LUMINANCE

The luminance signal is split into 16 subbands according to the decomposition tree of figure 3. These bands will be referred to as L1 to L16 according to the numbering of figure 3.

**FIGURE 3 : LUMINANCE SPECTRUM DECOMPOSITION****B) CHROMINANCE**

Chrominance signals CR and CB, which are subsampled by two in both directions are split into 7 subbands according to the decomposition tree of figure 4. These bands will be referred to as CR1 to CR7 and CB1 to CB7 according to the numbering of figure 4.

**FIGURE 4 : CHROMINANCE SPECTRUM DECOMPOSITION**

Luminance and chrominance bands with the same spectral position will be processed (inter/intra decision, motion vectors) and transmitted together ; however these bands do not have the same size due to the subsampling of chrominance signals.

II-2 - Interlaced input signal

An interlaced input signal is also processed in the same way regardless of its format (HDI or TV). First it is transformed into a progressive signal. This conversion may be implemented in several ways. At the moment this conversion is implemented by simply mixing two fields together into one frame (improvement are under study).

ANNEX II - ORGANISATION INTO BLOCKS

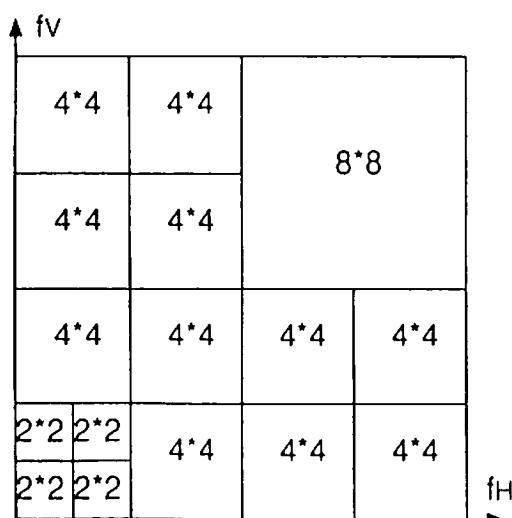
In order to carry out inter/intra decision, which is done at the subband level, the signal obtained in each band is organized in subband blocks. This block structure is deduced from a virtual block splitting of the luminance input image.

The luminance input image, whatever its format, once transformed into progressive, is supposed to be divided into adjacent NxN blocks, with N = 16 (in present simulations).

This block structure further determines that of each subband (luminance or chrominance bands), which is then described with the same number of blocks as the input luminance image. Each block of this input signal corresponds to one block in every subband, whose size is divided according to the subsampling factor of the considered band.

a) Luminance bands

Figure 1 gives block sizes of each luminance subband block for a 16x16 picture block at the input level.



**FIGURE 1 : BLOCK SIZES OF THE LUMINANCE SUBBAND DECOMPOSITION
(progressive input)**

Band 1 : 8 x 8

Bands 2 to 12 : 4 x 4

Bands 13 to 16 : 2 x 2

b) Chrominance bands

Due to the subsampling of input chrominance images, chrominance bands must be described with smaller blocks than corresponding luminance bands, in order to keep the same number of blocks per subband. These block sizes are described in figure 2.

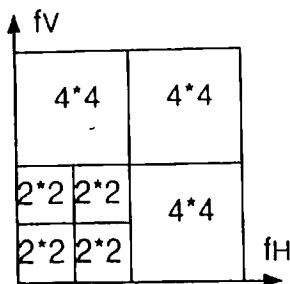


FIGURE 2 : BLOCK SIZES OF THE CHROMINANCE SUBBAND DECOMPOSITION (progressive input)

Bands 1 to 3 : 4 x 4

Bands 4 to 7 : 2 x 2

c) Macroblock

For inter/intra decision the subband blocks are grouped together in subband macroblocks. Two sets of subbands are defined :

- For bands L2 to L4 and L13 to L16 : A macroblock consists of 1 luminance subband block and 2 chrominance subband blocks (1 CR and 1 CB). Naturally, these 3 blocks represent the same area in the input image.
- For bands L1 and L5 to L12 : A macroblock consists of 1 luminance block.

ANNEX III - ENCODING MODES

2 modes are used to encode the subband blocks : intra frame and inter frame motion compensated modes.

Motion compensation is performed at the image level.

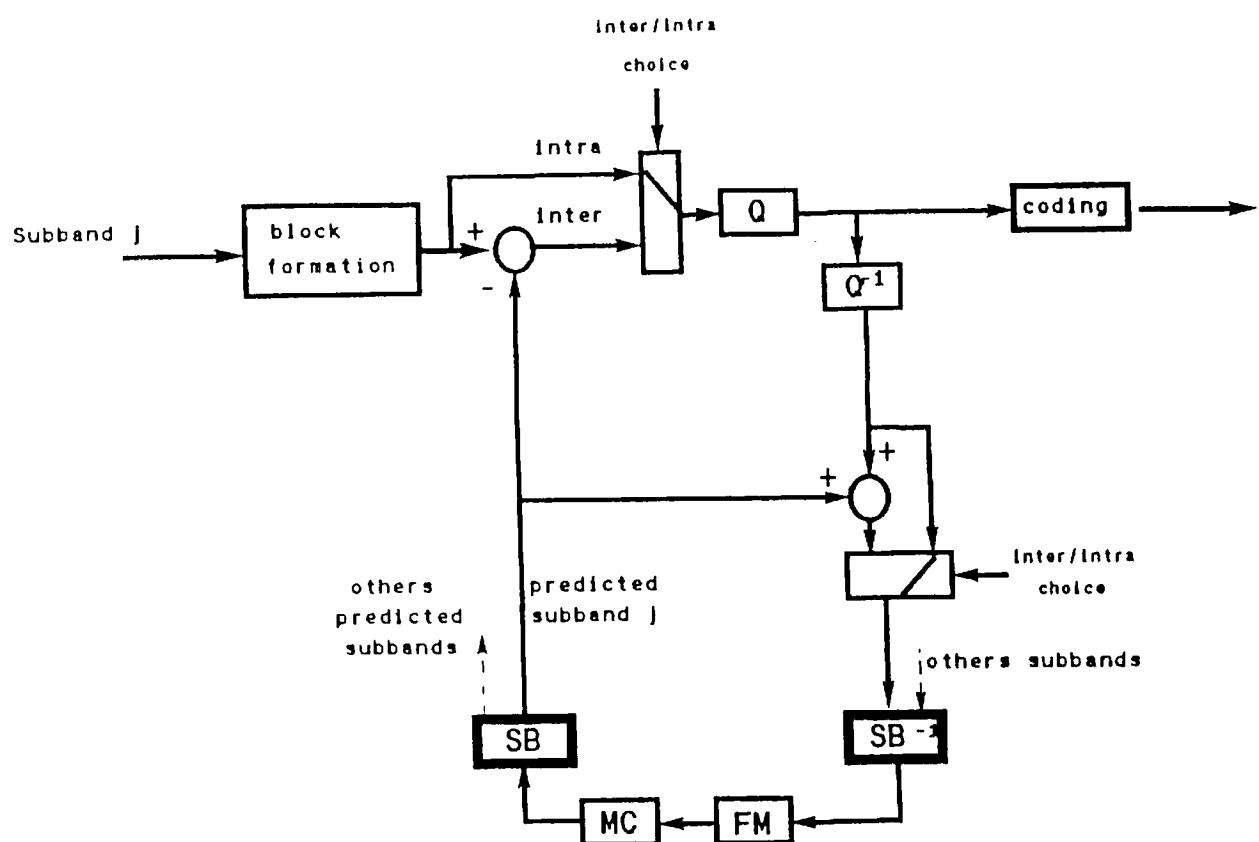


FIGURE 1 : INTRA/INTER ENCODING MODES

BLOCK FORMATION (see annex II)

Q : Quantization (see annex VI)

CODING : see annex VII

Q-1 It builds a subband block from the corresponding transmitted informations by assigning to the coefficients of the block the reconstruction values corresponding to the transmitted quantization level.

FM : Frame memory : It provides storage for the current decoded image which will be used as prediction reference for coding the subbands of a next image.

- MC : Motion compensation = prediction block determination.

It determines the prediction of the block currently processed for the inter-frame mode.
This prediction block is computed with pixels of a previous picture according to a motion vector (see Annex V)
INTER - INTRA CHOICE : In order to reduce side information transmission, the same inter/intra choice is taken for all subband macro-blocks corresponding to the same spatial position in each subband set (see Annex II).

This chosen mode is then transmitted with the subband macro-block description.

Comments :

The choice is based on the energy of the luminance block belonging to the macro blocks of the subband set. Many criteria may be used to compute the energy, but no specification is required since it only concerns the coder side.

To avoid the temporal propagation of transmission errors effects, it is recommended to use an intra-frame refreshing processing. Besides, intra-frame refresh is also an important task for compatibility with storage media purposes as described in Annex IX.

ANNEX IV PREDICTION OF THE BLOCK

According to the block organization described in ANNEX II, a choice between intra frame mode and inter frame motion compensated mode is taken for each subband-block (see ANNEX III).

Two cases must thus be distinguished according to the accuracy of this motion vector :

- 1) When both coordinates of this vector are integer, there is no ambiguity in the definition of the prediction block. It is naturally defined by the translation of the corresponding block in the previous decoded picture.
- 2) If one of the coordinates of the motion vector has a non-zero fractional part, an interpolation scheme has to be used to build the prediction block.

This fractional part may equal 1/2 pel, and a classical bilinear interpolation is used.

In case of interlaced scanning, each field (i.e. one line over two in the block) is predicted by the previous field of same parity. That means that a vector of an odd number of vertical pixels for the frame corresponds to a half pixel vector for the field, and half pixel for the frame corresponds to quarter pixel for the field. Bilinear interpolation is used.

ANNEX V - MOTION COMPENSATION

Motion estimation is performed at the input image level on the luminance signal. This image is divided into NxN blocks ($N = 16$) and one motion vector per picture block is estimated using for instance a full search block matching algorithm. The estimation process is characterized by the following parameters :

- accuracy : 1/2 pel (1/4 pel vertically, see annex IV)
- search area : +/- 16 pels horizontally and vertically

ANNEX VI - QUANTIZATION

Luminance and chrominance informations of each band are quantized on a (subband) block basis according to the block description of ANNEX II.

Linear quantizers with different characteristics are used for each band. Two parameters, the dead zone (d) and the stepsize (s) define each quantizer as described in figure 1.

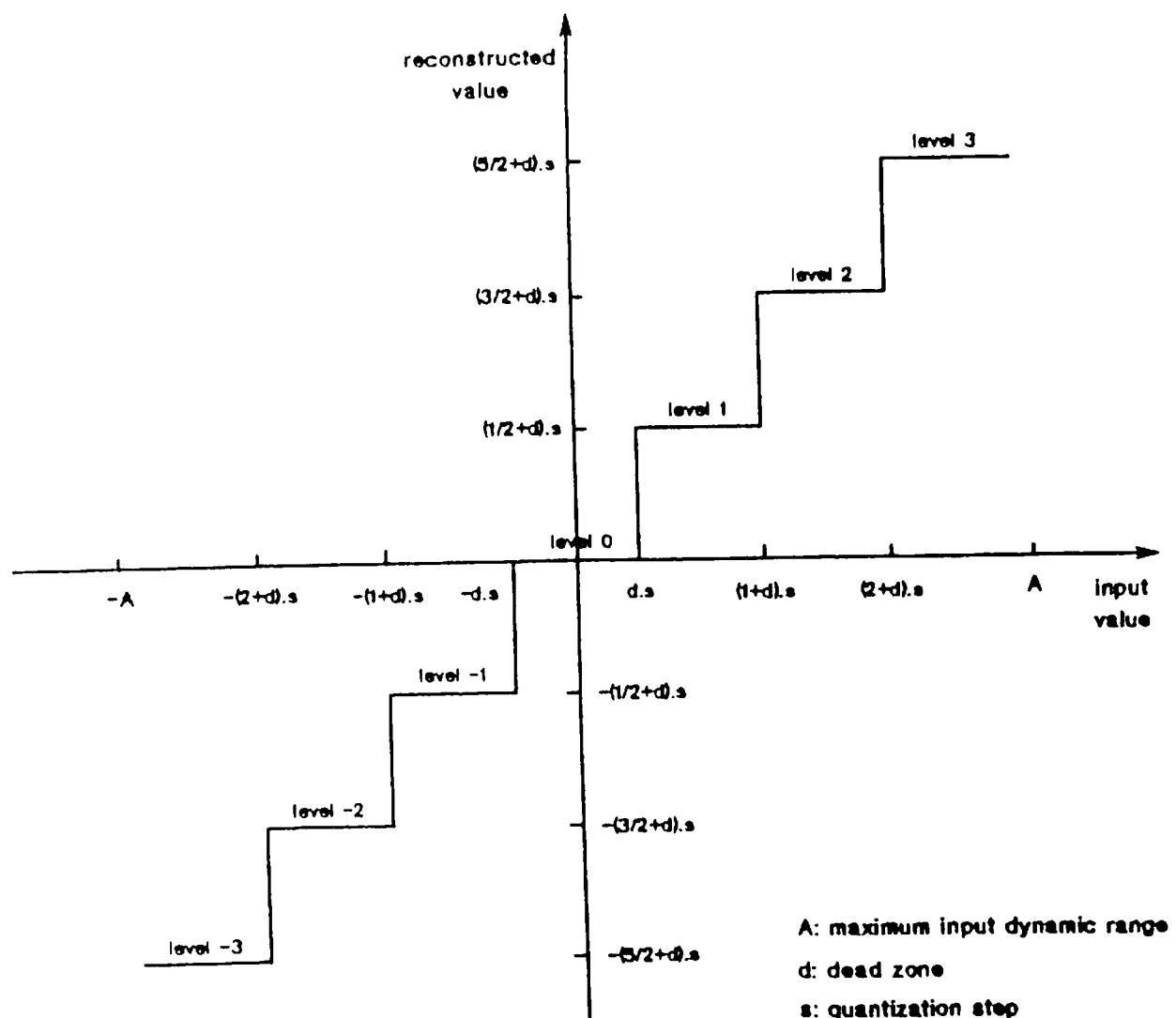


FIGURE 1 : LINEAR QUANTIZER

Furthermore, a regulation factor (f), common to all subbands determines the quantization step applied to each band $q(i)$ according to :

$$q(i) = f \cdot s(i)$$

This regulation factor is computed according to the buffer occupancy (see ANNEX IX).
The same quantization steps are used for inter and intra blocks.

Remark : The quantization process may be improved by use of an adaptive factor (m) common to all subbands but varying from block to block according to the criticality of the corresponding block in the input luminance image (not implemented at the moment).

Dead zones are set to 1 in all bands, both for luminance and chrominance signals.

Step sizes $s(i)$ used for each band are described in figure 2 for a progressive input image.

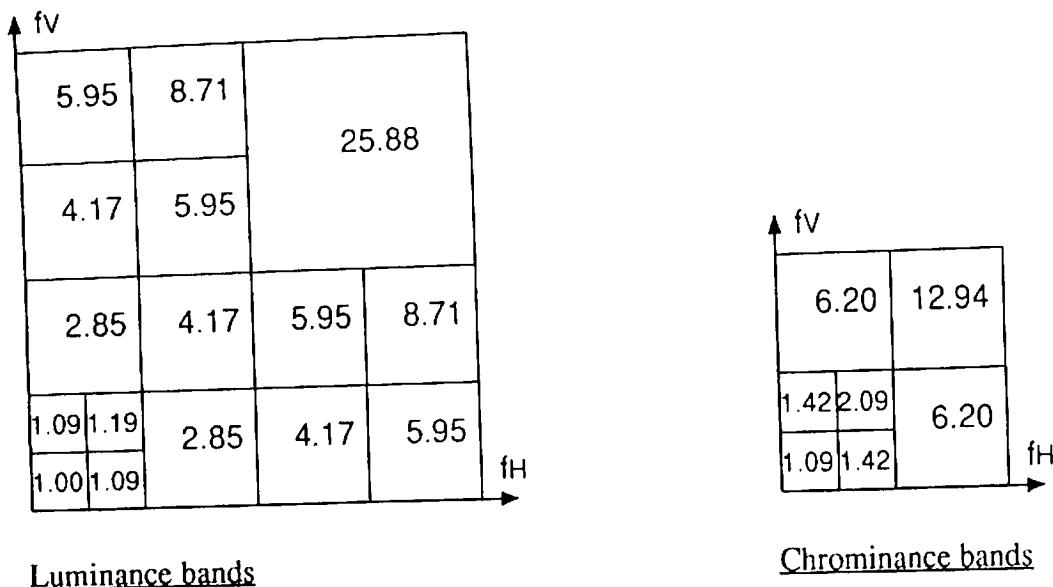


FIGURE 2 : Stepsizes $s(i)$ for a progressive input image

Remark : In order to minimize the reconstruction error variance, stepsizes $s(i)$ should be chosen as :

$$s(i) = 2^{-ki}$$

Where k_i equals the number of times band i has undergone a 4-band split. Thus, further 4-band splitting of a band should divide the corresponding stepsize by 2.
These optimal stepsizes are not used here ; instead, some weighting is applied among the subbands. Such an approach yields in fact better visual quality of reconstructed images.

Remark : the weighting strategy described above applies to progressive originated pictures. A different (adaptive ?) strategy should be designed when pictures are obtained from interlace format.

ANNEX VII - VARIABLE LENGTH CODING

I - GENERALITIES

Arithmetically computed variable length codes (ACVLC's) are used.

I-1 - ACVLC - Codeword Structure

An ACVLC (Arithmetically Computed Variable Length Code) codeword possesses the following structure :

- It is composed of a number of groups of bits. For a given codeword let k denote the number of groups constituting the codeword.
- The first $k-1$ groups (the prefix) of the codeword contain only "1"s. If $k=1$ the prefix is absent.
- The k^{th} group (the suffix) contains at least one "0".

Let l_i denote the number of bits in the i^{th} group of a codeword.

I-2 - ACVLC - Definition Of A Code

An ACVLC code is completely defined by specifying a maximum number of groups k as well as the group lengths $l_1 \dots l_k$. The complete code is then the set of all ACVLC codewords which can be derived using these parameters. This code is uniquely decodable (no codeword is the prefix of an other codeword) and can be represented by a binary tree.

I-3 - ACVLC - Assignment Of Codewords

The assignment of a code word to a positive value n is as follows :

$$1) \text{ if } 0 \leq n < S_1 = 2^{l_1} - 1$$

the ACVLC codeword is as follows :

- * no prefix
- * the suffix is equal to n coded on l_1 bits
- * the total length (prefix + suffix) is l_1

PROPOSAL 34 FOR MPEG-2

2) if $S_1 \leq n < S_2 = S_1 + 2^{l_2} - 1$

the ACVLC codeword is as follows :

- * the prefix is composed of l_1 "1"s
- * the suffix is equal to $n - S_1$ coded on l_2 bits
- * the total length (prefix + suffix) is $l_1 + l_2$

and, more generally

3) if $S_i \leq n < S_{i+1} = S_i + 2^{l_{i+1}} - 1$

where

$$S_i = \sum_{j=1}^i (2^{l_j} - 1)$$

The ACVLC codeword is as follows :

* the prefix is composed of $\sum_{j=1}^i l_j$ "1"s

* the suffix is equal to $n - S_i$ coded on l_{i+1} bits.

In all cases :

The ACVLC codeword always ends with the LSB (Least Significant Bit) of the suffix.

This algorithm defines a one-to-one mapping between the set of all codewords of a given ACVLC code and the set of all integers n satisfying the condition :

$$0 \leq n < S_k$$

$$\text{where } S_k = \sum_{j=1}^k (2^{l_j} - 1)$$

The ACVLC codeword assigned to an integer value n using this algorithm will be denoted in what follows by $\text{ACVLC}(n)$.

I-4 - Variable Length Coding (VLC) Modes

ACVLC is used to encode different parameters needed for the description of a block e.g.

- * number of transmitted coefficients
- * relative addresses
- * absolute values of the coefficients
- * motion vectors.

Due to the different ranges involved 3 different modes of encoding an integer λ are distinguished :

I - λ can be positive or zero
 λ is encoded as ACVLC(λ).

II - λ can be positive or negative but not zero
If $\lambda > 0$, it is encoded as ACVLC ($\lambda - 1$), + "0"
If $\lambda < 0$, it is encoded as ACVLC ($|\lambda| - 1$), + "1"

III - λ can be positive or negative or zero
If $\lambda \geq 0$, it is encoded as ACVLC (λ), + "0"
If $\lambda < 0$, it is encoded as ACVLC ($|\lambda| - 1$), + "1"

in modes II and III an integer is encoded using an ACVLC word and the sign bit shown.

To each parameter an ACVLC tree (i.e. a set of group lengths $l_1 \dots l_k$) and an encoding mode are associated (see paragraph 4.5)

I-5 - Transmission Rules

- * Fixed length codeword

The MSB (Most Significant Bit) is transmitted first.

- * Variable length codeword

The prefix is always transmitted first, followed by the suffix and (if present) the sign bit.

I-6 - Ranges Of ACVLC Tree Parameters

- * The maximum number of groups (k) cannot exceed 8.
- * The group length (l_i) of any group cannot exceed 10.

Consequently 32 bits (8 group lengths, each specified on 4 bits) are needed to completely define an ACVLC tree.

I-7 - Number Of Trees

For a given application 16 predefined ACVLC trees are sufficient, requiring $16 \times 8 \times 4 = 512$ bits to fully specify all the VLC parameters.

I-8 - Range Of Values Encoded

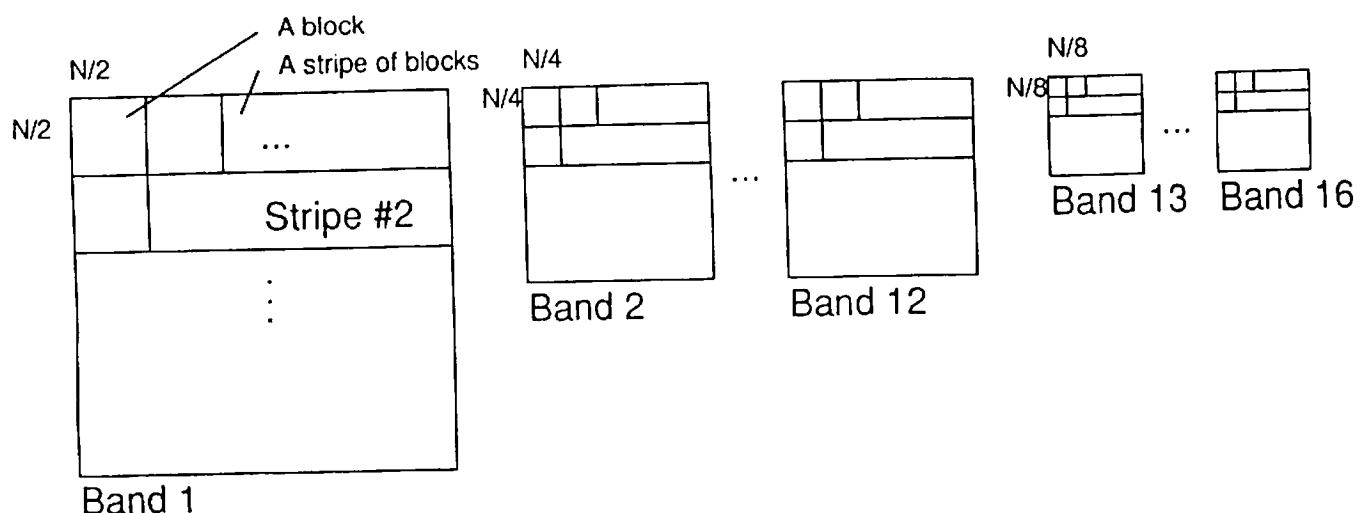
The use of VLC is restricted to values belonging to the set $[-1023, +1023]$ (10 bits + sign).

I-9 - Maximum Codeword Length

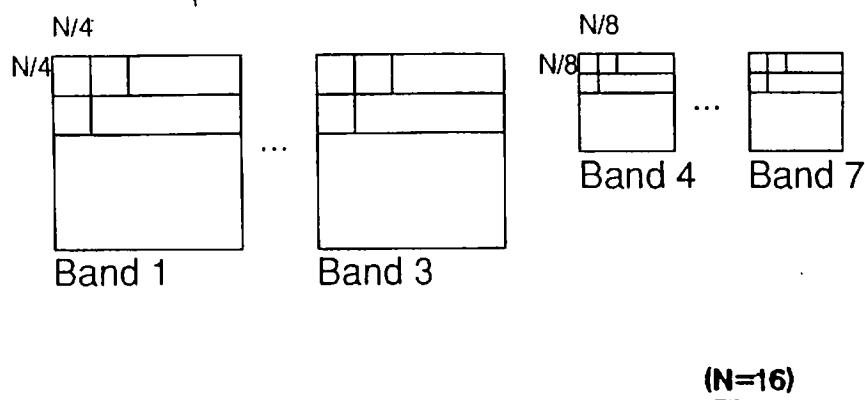
No value may be encoded on more than 32 bits.

II - SUBBAND DESCRIPTION

Once quantized, each subband is encoded using variable length codes and run length description. For this purpose, each band is split into (subband) stripes of blocks (see Annex VIII) according to the block description of annex II. This description is recalled in figure 1 for luminance and chrominance bands.



LUMINANCE BANDS



CHROMINANCE BANDS

Figure 1 : Band stripe decomposition

Each subband stripe of blocks is then encoded using a run length description and VLC's described above. The stripes are scanned (zero run description) according to the horizontal path shown in figure 2.

1 2 3 4	5 480
481 482 960
961 1440
1441 1920

Figure 2 ; Stripe scanning (band 2 of a HDP signal i.e. with 1920 pels per line)

The number of coefficients in a (subband) stripe thus depend upon the considered subband. It may be expressed as :

$$\text{Nbcoef}(s) = N \times \text{NPT}/S^2$$

where N is the picture blocksize in the progressive input image ($N = 16$), NPT is the number of pels per line of the input image and S is the subsampling factor (in both directions) of the considered subband.

The same holds for the maximal length of zero runs which varies according to the considered band (and which indeed equals $\text{Nbcoef}(s)$).

VLC description of a stripe

A subband stripe of blocks is described with the help of three kinds of information :

- The number of transmitted (non zero) coefficients
- The relative address

The relative address is defined as the number of consecutive zeros that precede the non zero coefficient (from the last non zero coefficient) in the 1-D description of the stripe of figure 2.

- The transmitted coefficient (non zero) amplitudes

The data framing is as follows : (for each stripe)

Nb [A] [VLC]

Where :

Nb : Number of transmitted (non zero) coefficients

The relative addresses of the transmitted coefficients (a VLC description may be used)

VLC's of transmitted coefficients

III - DESCRIPTION OF OTHER PARAMETERS

Some benefit can be obtained by encoding the motion vectors with variable length codes. Such a description highly depends on the motion estimation algorithm. For instance when motion estimation is performed on the input image (see ANNEX V), one motion vector is estimated for each 16x16 block in the input image. In order to have a zero mean probability distribution (ie to obtain an efficient VLC encoding) the motion vectors are transmitted in differential mode.

The prediction is the previous vector when it exists and zero when it does not (at the beginning of a stripe).

ANNEX VIII - VIDEO FRAMING

I - GENERAL STRUCTURE

The general structure of the video multiplex is based on the concept of subband stripe of blocks (SB) : each band is split in (subband) stripes of blocks according to the block structure of ANNEX II.

See annex VII for the decomposition into subband stripes of blocks.

The following notations will be used to designate a subband stripe of blocks :

SSBL(i)(j) will designate the stripe n°j in the luminance band n° i.

For chrominance bands, L is simply replaced by CR or CB according to :

SSBCR(i)(j) and SSBCB(i)(j)

The general structure is then described as follows (figure 2) :

FSW 00 ST VF VBR BO NSB* NBS*

FSW 01 ST VF VBR BO NSB* NBS*

FSW 10 ST VF VBR BO NSB* NBS*

SSW0 [PSB1] SSW0 [PSB2] ... SSW0 [PSBk]

Where : FSW : Frame synchronisation word.

ST : System type (1 bit)

0 = 50 Hz

1 = 60 Hz

VF : Video format (3 bit)

HDP	000
HDI	001
EDP	010
TV	011
VT	100

PROPOSAL 34 FOR MPEG-2

VBR : Video bit rate (4 bits)

BO : Buffer occupancy (16 bits)
Indicates the buffer occupancy at the coder at the end of the previous active frame

NSB : Number of stripes of blocks per band
NBS : Number of blocks per stripe
SSW : Stripe synchronisation word
[PSBi] : Picture stripe of blocks n°i

Each band (luminance or chrominance) possesses the same number of subband stripes of blocks. A PSB contains the descriptions of the subband stripes of blocks located at the same picture area in all the bands.

[PSBi] = SN BOS TF [MI₁] [MI₂] [MV_x] [MV_y]

[SSBL(1)(i)]
...
[SSBL(16)(i)] [SBCR(16)(i)] [SBCB(16)(i)]

CRC

With the following notations :

SN : Stripe number (or PSB number)
(from 1 to NSB)

BOS : Buffer occupancy.
Indicates the buffer occupancy at the coder at the end of the previous active group of stripes of blocks.

TF : Transmission factor

[MI_k] = Mode identification (intra or inter) of the blocks of the considered subband stripe (same identification for luminance and chrominance blocks).

[MV_x] : Motion vectors in the horizontal direction estimated for the blocks of the corresponding stripe at the input image level (there are NBS such vectors)

[MV_y] : Motion vectors in the vertical direction estimated for the blocks of the corresponding stripe at the input image level (there are NBS such vectors)

CRC : Cyclic Redundancy Code to be applied to all bits of a Picture stripe of blocks.

[SSBL(k)(i)] : Description of the luminance stripe of blocks
[SBCR(k)(i)] : Description of chrominance stripes of blocks
[SBCB(k)(i)] : Description of chrominance stripes of blocks

II - DESCRIPTION OF A SUBBAND STRIPE OF BLOCKS

According to the VLC description of a subband stripe of blocks (see ANNEX VII), the video framing structure is as follows :

$$[\text{SSBL}(k)(i)] = N_L [A_L] [\text{VLC}_L]$$

Where,

N_L : Number of transmitted (non zero) coefficients in the corresponding stripe

$[A_L] : A_L^1 \dots A_L^{NL}$: Relative addresses of the transmitted coefficients of the stripe

$\text{VLC}_L^1 \dots \text{VLC}_L^{NL}$: VLC's of the transmitted coefficients of the stripe.

The same holds for chrominance stripes of blocks :

$$[\text{SSBCR}(k)(i)] = N_{CR} [A_{CR}] [\text{VLC}_{CR}]$$

$$[\text{SSBCB}(k)(i)] = N_{CB} [A_{CB}] [\text{VLC}_{CB}]$$

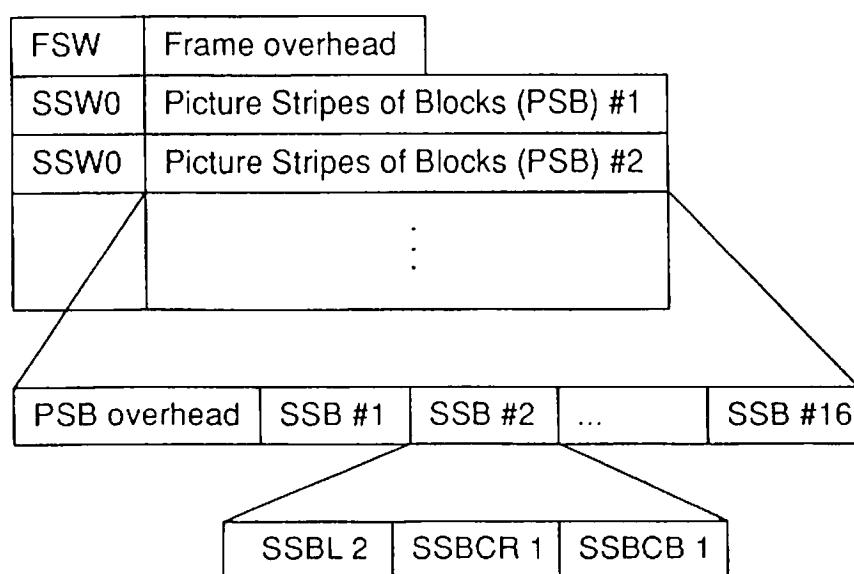


FIGURE 2 : VIDEO FRAMING GENERAL DESCRIPTION

ANNEX IX - BUFFER MEMORY

For storage applications, random access and fast forward and backward playback are made possible through the periodic coding of full pictures in intra mode.
The coded data buffer must be large enough to accept the high amount of bits needed for an entirely intra-coded picture. Five time the average number of bits per image seems sufficient :

800 Kbit at 4 Mbit/s
1.8 Mbit at 9 Mbit/s

The regulation strategy (which does not need to be part of the standard description), consists in computing the regulation factor (see Annex VI) that empties enough the buffer before intra-coded picture while avoiding buffer overflow for this picture while keeping the regulation factor as constant as possible.

ESTUARINE CLOUDS

TANDEM TETRAHYDRATE

TABLE TEEN