CCITT Study Group XV
Working Party XV/1
Specialists Group on Coding for Visual Telephony

Doc #123
June 1986

Title:

Specification for the Flexible Prototype nx384kbit/s Video Codec

Source: France, FRG, Italy, Netherlands, Sweden, UK

This document is a revised version of the Hardware Specification (Annex 4 to Doc #103R) agreed at the fifth meeting of the Specialists Group in Tokyo.

The amendments and extensions are a result of further studies by the European participants. These changes are indicated by vertical lines in the left hand margin.

Specification for the Flexible Prototype nx384kbit/s Video Codec

CONTENTS

1 SOURCE CODER

1.1 Source format

1.1.1 The format to be coded is 288 lines, 30000/1001 (approximately 29.97) non-interlaced pictures per second - the Common Intermediate Format. The tolerance on the picture frequency is ±50 ppm.

1.1.2 Pictures are coded in component form, these components being luminance (Y) and two colour difference signals (C_R, C_B). These three components and the codes representing their sampled values are as defined in CCIR Rec 601.

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

Codes outside the above ranges will be accepted but may be modified by the coder to avoid emulation of reserved codewords.

1.1.3 The luminance sampling structure is 288 lines per picture, 360 pels per line in an orthogonal sampling arrangement. The colour difference sampling parameters are 144 lines, 180 samples per line, orthogonal. Both C_R and C_B samples are sited such that their block boundaries coincide with luminance block boundaries. (See Table 1 and Fig. 2)

1.1.4 Before coding the number of samples per line is reduced to 352 for luminance and 176 for each colour difference signal by discarding 8 pels and 4 pels respectively from the source.

1.2 Video Coding Algorithm

The video coder algorithm is shown in generalised form in Fig. 3. The main elements are prediction, block transformation, quantisation and classification.

1.2.1 The prediction is inter-picture. The predictor may incorporate movement compensation. The predictor requires a picture store with one pel access for each pel being predicted. (Note: multi-pel prediction may be added later, introducing a requirement for a multiple access picture store configuration.)

1.2.2 Motion compensation is optional at the coder. The decoder will accept one displacement vector for each luminance block of size 8 pels by 8 lines. The maximum displacement vector is ±15 pels and ±15 lines. Only integer values of the horizontal and vertical components of the vector are currently considered. Fractional values will be considered for possible inclusion at the Montreal meeting 1986. (Note: The encoding method for transmission of displacement vectors may restrict the vectors to a subset of these 1089 possible values that the decoder hardware can accomodate. This will be specified later.)

For each pair of C_R and C_B blocks, four motion vectors are derived, one from each of the spatially corresponding luminance blocks. If any of these luminance blocks is not transmitted then a zero motion vector is used in that quadrant. These derived vectors are used for prediction at the coder and the decoder, but are not transmitted in the chrominance block data. (Treatment of fractional chrominance vectors?)

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

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

1.2.3 Coder

The prediction error (inter-picture mode) or the original picture (intra-picture mode) is subdivided into 8 pel by 8 line blocks which are segmented as transmitted or non-transmitted. The method of choosing between inter or intra mode is not defined in this specification, nor is the segmentation criterion, both being left open to equipment designers and may be varied dynamically as part of the data rate control strategy.

Transmitted blocks are coded by a transform based scheme. Coding block size is 8 lines by 8 pels (64 sample values) for Y, C_R and C_B.

1.2.4 Transformer

The transformer shall be implemented in a flexible manner so that a number of different transforms, or configurations of a particular transform in terms of bits per stage, can be investigated. All hardware should be equivalent in performance to classical matrix multiplication.

The 2-D transform is implemented as the equivalent of 2 independent 1-D transforms.

All coefficients in the forward and inverse transforms shall be programmable.

For the purposes of compatibility it is only necessary to specify the inverse transform. The hardware should allow for 12 bit resolution in (ie coefficients) and 9 bits out.

The matrix multiplier implementation of the inverse transform uses TDC1010 or functionally equivalent devices. The matrix coefficients are stored to 16 bits. The most significant 16 bits by truncation are carried between the two 1-D transforms. The output of the inverse transform is rounded to 9 bits giving pel domain values in the range -256 to 255.

1.2.5 Classification

The hardware will allow transform coefficients to be transmitted and received in up to 8 sequences, each of which has programmable order and programmable length. Additionally the end of block code (EOB) of 2.2.5 may be used to terminate a sequence before its programmed length.

1.2.6 Quantizer

The number of inverse quantisers provided in the decoder shall be 32. Each quantiser has 12 bits input and up to 12 bits out.
(The same or different for luminance/chrominance? Adaptive/non-adaptive)

1.2.7 VLC

1.3 <u>Data rate control and subsampling modes</u>

The exact method of assessing the encoder data generation rate need not be specified but the specified minimum size of encoder buffer must incorporate an allowance for latency in the assessment and control loop. Hence, any requirement to constrain overall system delay may effectively preclude some schemes.

Control information is carried by side information - not derived recursively from received data.

Horizontal subsampling - picture, group of blocks or block basis?
Vertical subsampling " " " " " "
Temporal subsampling - picture basis only. Interpolated pictures are not placed in the picture memory.
Quantizer selection - see also 1.4.
Block significance criterion - not part of the specification.

The coding algorithm will automatically permit the full quality of the source format specified in section 1.1 to be realised on still pictures.

1.4 <u>Forced Updating</u>

There is no separate coding scheme for forced update. This function is achieved merely by forcing the use of the intra-picture mode of the coding algorithm. Since the decoder cannot distinguish between normal and forced update blocks, then there are no parameters to specify which are unique to forced update. The bitrate allotted to forced updating, the sequence in which blocks are updated etc, are not specified.

In order that the quality of forced update blocks can be sufficiently high at all times, quantiser selection must not be dependent solely on buffer fill state.

2 VIDEO MULTIPLEX CODER

2.1 Tasks

The video multiplex coder has the following tasks:

1. Block address coding
2. Video data formatting and serialising
3. Synchronisation - picture/line
4. Motion vector coding
   (Absolute, differential, run length etc)
5. Side channels for indicating dynamic coding parameters eg subsampling modes, quantizers, buffer state etc. (Permanent or transient channels? Transient channels require care in switched multipoint.)

2.2 Video Multiplex Arrangement

2.2.1 Picture start code (PSC)

Unique sequence of f bits [Buffer state] [Temporal ref.] [Type]

The buffer state is a 6 bit number representing the encoder buffer fullness in 1Kbit units at the beginning of this picture.

The temporal reference is a 3 bit number representing the time sequence, in Common Intermediate Format picture periods, of a particular picture.

Type is a Variable Length Code which allows block attributes to be applied to all blocks within a picture.

All PSCs are transmitted. Two successive PSCs indicate the intervening picture has been dropped.

2.2.2 Group of Blocks Start Code (GBSC)

Unique sequence of g bits [Group Number] [Type]

A group of blocks consists of two lines of luminance blocks, one line of C_R blocks and one line of C_B blocks. See Fig. 4.

Y_U	Y_L	C_R	C_B

Fig 4

Subsampling aspects are open for further consideration.

The group number is a 5 bit number, ranging from 00001_2 to 10010_2, representing the vertical spatial position, in units of groups, of the current group of blocks from the top of the picture.

Type is a Variable Length Code which allows block attributes to be applied to all blocks within a group of blocks.

All GBSCs are transmitted.

2.2.3 [Block Address]

Block Address indicates the position of a transmitted block within a group of blocks. (A decision is urgently needed on how this is achieved. Possible methods include absolute or relative addressing of only transmitted blocks and one bit per block for all blocks.)

The range of absolute block addresses for the upper line of luminance blocks (Y_U) is 0 to 43.

The range of absolute block addresses for the lower line of luminance blocks (Y_L) is 44 to 87.

The range of absolute block addresses for C_R blocks is 88 to 109.

The range of absolute block addresses for C_B blocks is 110 to 131.

2.2.4 Block Type

A Variable Length Code representing the type of the block.

1. Intra-picture coded block
2. Inter-picture coded block
3. Motion compensated block
4. Motion compensated with coded residue
5. Future expansion on
6. Future expansion off

The decoder shall be designed to ignore all data between types 5 and 6 and also between type 5 and the next GBSC.

2.2.5 Block Data

Data specifying motion vectors, scanning class, quantiser type, transform coefficients and end of block marker (EOB). Although some of these may not be present, they are always transmitted in the following order:

....
....
....
Transform coefficients
EOB

EOB is one of the codewords in the VLC set used for quantised transform coefficients. It is always present for every transmitted block

2.3 <u>Multipoint Considerations</u>

2.3.1 Freeze Picture Request (FPR)

On receipt of a FPR, conveyed in bit .... of timeslot .... the decoder shall freeze its received video until the next picture attribute indicating intra-picture mode is encountered.

2.3.2 Fast Update Request (FUR)

On receipt of a FUR, conveyed in bit ..... of timeslot...... the encoder shall empty its transmit buffer and encode its next picture in intra-picture mode with such coding parameters as to avoid buffer overflow.

2.3.2 Data continuity

The protocol adopted for ensuring data continuity in a multipoint connection will be handled by the message channel.

3 TRANSMISSION CODER

3.1 The transmission coder assembles all data and interfaces to the digital line transmission system.

3.1.1 The data rate is n x 384kbit/s where n is an integer between 1 and 5, both inclusive.

3.1.2 The codec output clock rate source shall be switchable between either a free running internal source or a source synchronised to the received data from the network. The mechanism for this switching is bit ..... of the application channel.

3.1.3 When in free running mode the tolerance on output clock rate will be ±50 ppm of nominal.

3.1.4 When in synchronised mode the synchronism should be maintained when the frequency of the received data clock is within ±50 ppm of nominal.

3.2 Framing structure (See Fig. 5)

As per CCITT Study Group XV WP XV/1 Doc #58 plus the following coding for the applications channel:

list of codec attributes/facilities/parameters needing transmission from transmitter to receiver. No return path is assumed. Currently this list includes only encryption.

Operability with audioconferencing. Timeslot positioning to CCITT Rec I.431.

3.3 Video data buffering

The size of the transmission buffer at the transmitter is switchable from 8Kbits to 64Kbits, both inclusive, in steps of 8Kbits. (K=1024) Buffer size should be related to the transmission rate (overall - not video) to ensure acceptable system delay.

3.4 Video clock justification

Not provided.

3.5 Optional full spatial resolution mode data for quasi-stationary pictures.

To be specified later if required.

3.6 Audio

As per CCITT Draft Rec. G72X type 2 terminal.

The audio channel is carried by the first time slot.

Flexible testbed hardware need only incorporate?
(64kbit/s A-law
64kbit/s u-law
56kbit/s sub-band ADPCM according to CCITT Draft Rec G72X)

3.7 Error handling

Video coding strategy to be error resilient without internal or external error corrector. Note that demand refresh for error correction requires back channel.

3.8 Encryption

3.9 Data transmission

Framing structure to allow 2 data ports of 64kbit/s each, though picture quality constraints may require only one to be available at 384kbit/s. In this case time slot 4 is used.

3.10 Network Interface

Access will be at the primary rate with vacated timeslots as per CCITT Recommendation I.431.

For 1544kbit/s interfaces the default H0 channel is timeslots 1 to 6.

For 2048kbit/s interfaces the default H0 channel is timeslots 1-2-3-17-18-19.

END

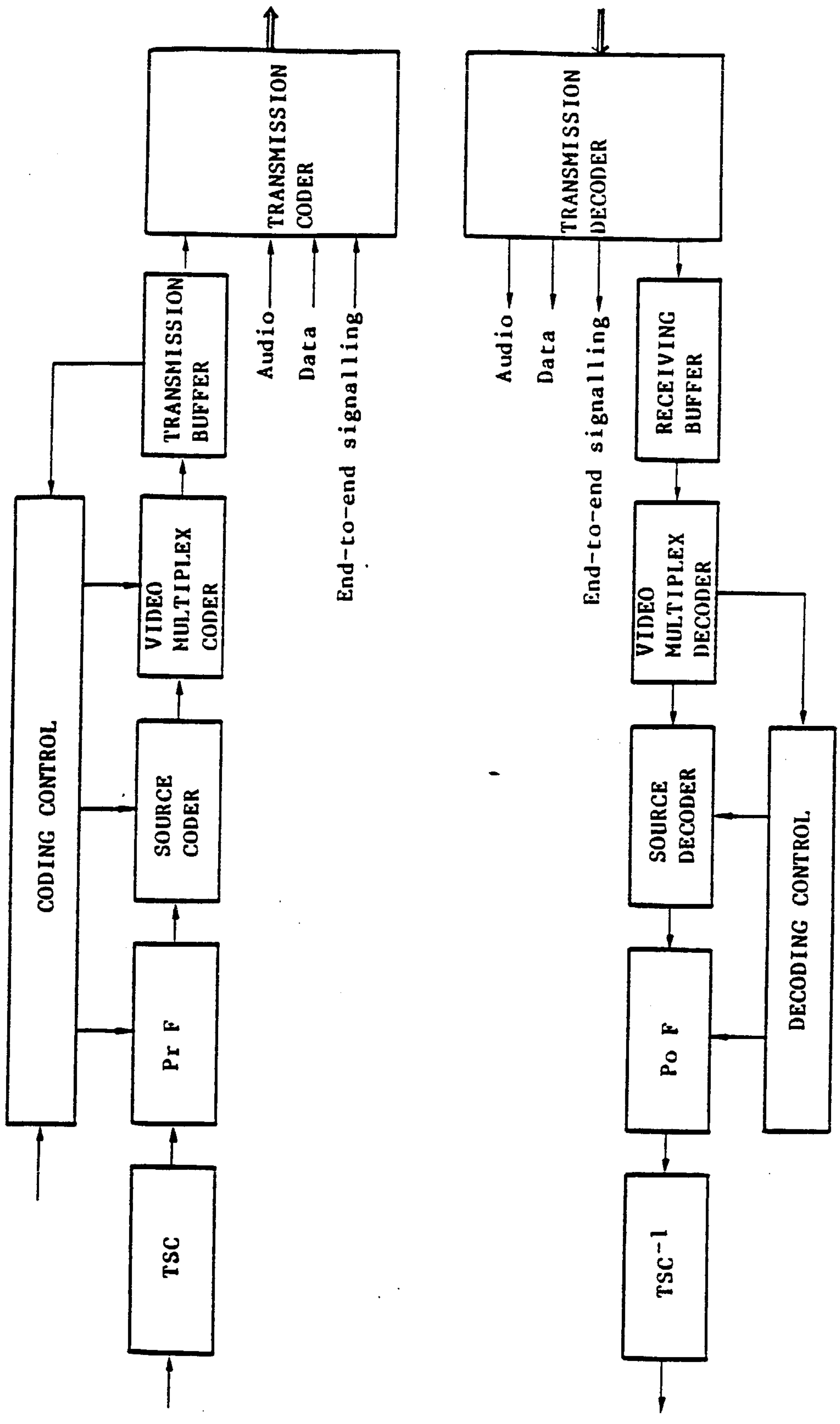

TRANSMISSION CODER
TRANSMISSION BUFFER
VIDEO MULTIPLEX CODER
SOURCE CODER
Pr F
TSC
CODING CONTROL
Audio
Data
End-to-end signalling
TRANSMISSION DECODER
RECEIVING BUFFER
VIDEO MULTIPLEX DECODER
SOURCE DECODER
Po F
TSC-1
DECODING CONTROL

Fig. 1

Table 1. Basic parameters for the new generation n x 384 kbit/s CODEC

Items	Parameters
1. Reference point	Point B in Fig. 1/Annex 1 to COM XV-R 4
2. Baseband signals and their levels	Y, R-Y, B-Y, as defined in CCIR Rec. 601
3. Number of pels per line	Y: 360 (Note 1) R-Y: 180 B-Y: 180
4. Number of lines per field	Y: 288 (Note 2) R-Y: 144 B-Y: 144
5. Field frequency	Y, R-Y, B-Y: 29.97 Hz
6. Interlace	Y, R-Y, B-Y: 1:1
7. Sampling structure	Y, R-Y, B-Y: orthogonal, positioning of R-Y and B-Y samples share the same block boundaries with Y samples as shown in Fig. 2

Note 1: Active line duration is approximately 53 us.

Note 2: Active field duration is approximately 18.4 ms (for 625/50 systems) and approximately 15.2 ms (for 525/60 systems).

Note 3: The common intermediate format defines the maximum attainable spatial and temporal resolution in the codec. Effective resolution may eventually be reduced by some coding operating modes.

Note 4: The common intermediate format is a logical specification to ensure compatibility among codecs. Hence, it might not appear at the physical interface points in the codec.

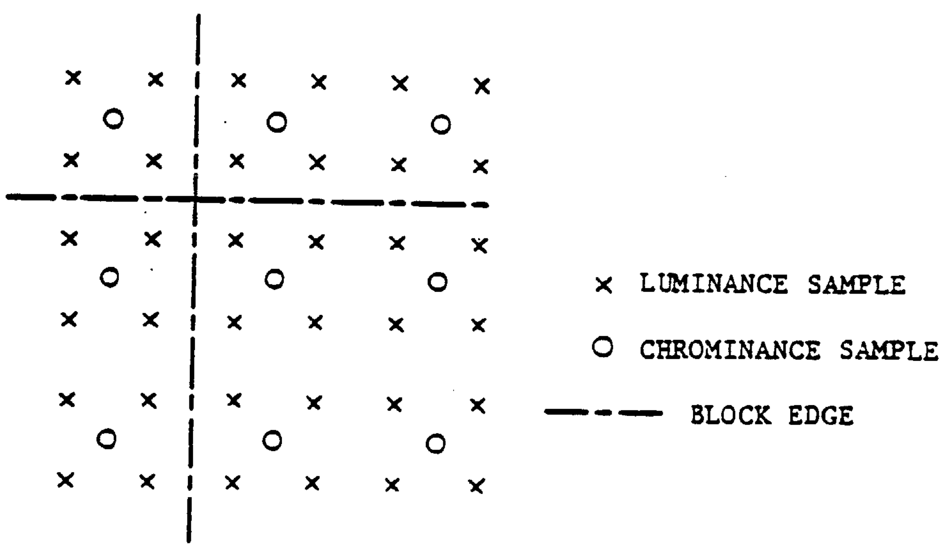
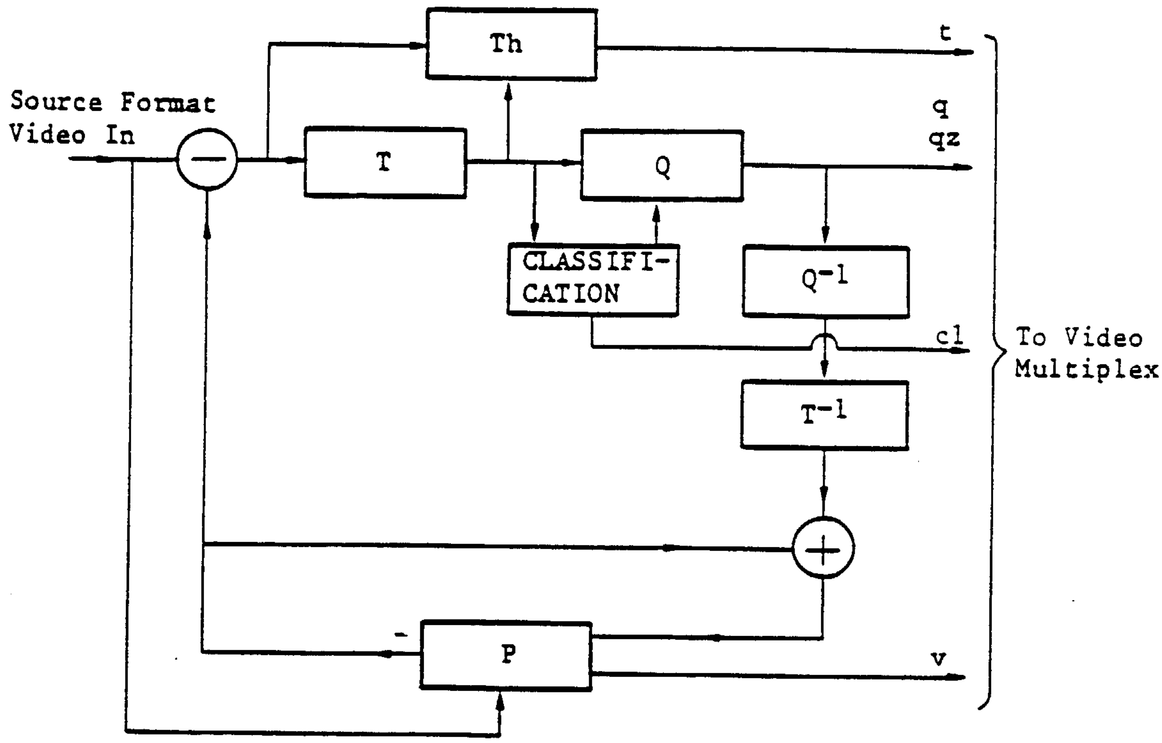
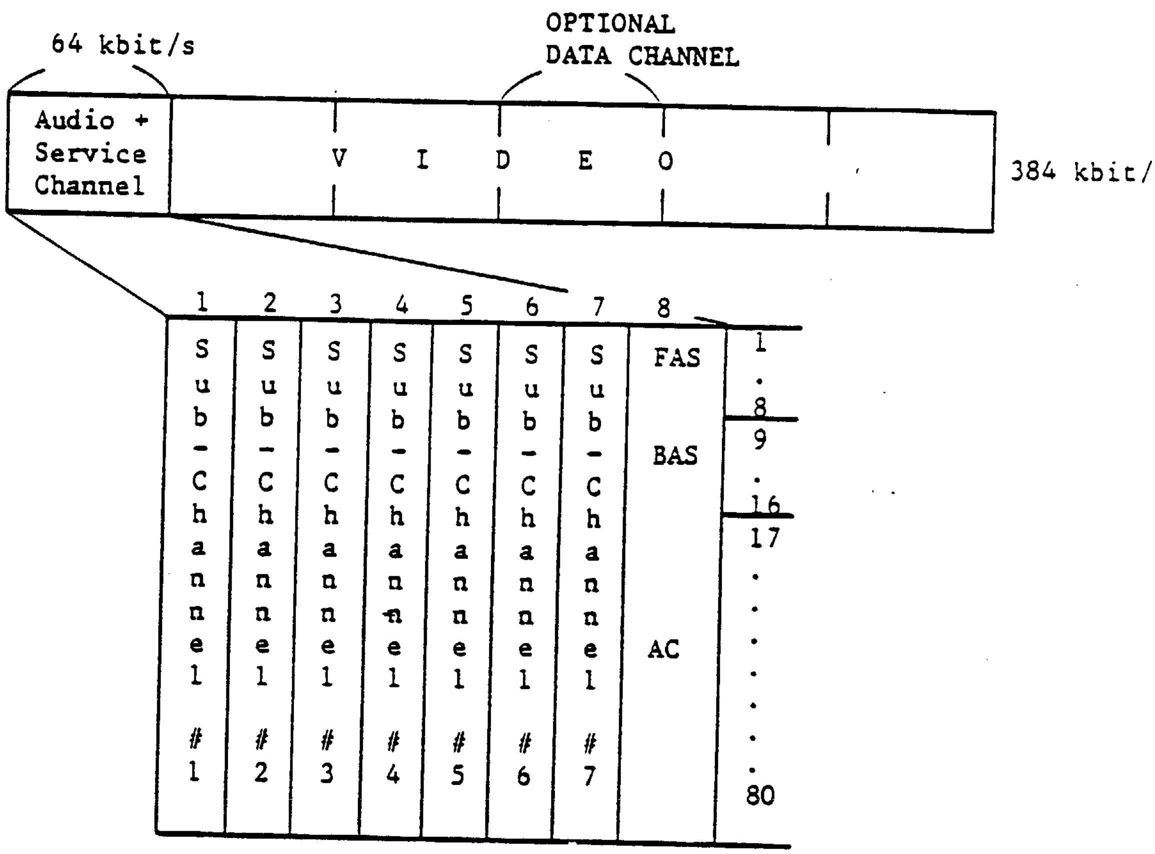


Fig. 2 Positioning of Luminance and Chrominance Samples



Th	Threshold
T	Transform
Q	Quantizer
Q^{-1}	Inverse Quantizer
T^{-1}	Inverse Transform
P	Predictor
t	Flag for transmitted/not transmitted block
q	Quantizing index for transform coefficients
qz	Quantizer indication
v	Motion vector
cl	Classification index

Figure 3



FAS: Frame Alignment Signal (note 1)
BAS: Bitrate Allocation Signal
AC: Applicaion Channel

Note 1: The block termed as FAS contains also other information than for frame alignment purposes.

Fig. 5 Frame Structure for n x 384 kbit/s codec (in case of n = 1)