INTERNATIONAL TELECOMMUNICATION UNION

ITU-T H.120

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(03/93)

LINE TRANSMISSION OF NON-TELEPHONE SIGNALS

CODECS FOR VIDEOCONFERENCING USING PRIMARY DIGITAL GROUP TRANSMISSION

ITU-T Recommendation H.120

(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.120 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 Similarly, in this reform process, the CCIR and the IFRB have been replaced by the Radiocommunication Sector.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 A codec for 625-lines, 50 fields/s and 2048 kbit/s transmission for intra-regional use and capable of interworking with the codec of 2 1
- 1.1 Scope 1
- 1.2 Brief specification 2
- 1.3 Video interface 3
- 1.4 Source coder 3
- 1.5 Video multiplex coding 8
- 1.6 Transmission coding 11
- 1.7 Error correction 12

2 Codecs not requiring separate television standards conversion when used on interregional connections ... 13

A codec for 525-line, 60 fields/s and 1544 kbit/s transmission for intra-regional use and capable of interworking with the codec of clause 1 13
- 2.1 Introduction 13
- 2.2 Brief specification 13
- 2.3 Video interface 14
- 2.4 Source coder 14
- 2.5 Video multiplex coding 15
- 2.6 Transmission coding 15

3 A codec for 525-lines, 60 fields/s and 1544 kbit/s transmission for intra-regional use 16
- 3.1 Introduction 16
- 3.2 Outline of codec 16
- 3.3 Brief specification 16
- 3.4 Video interface 18
- 3.5 Pre- and post-processing 19
- 3.6 Source coding 21
- 3.7 Audio coding 44
- 3.8 Transmission coding 44

Annex A – Graphics option – 625 line 46
- A.1 Introduction 46
- A.2 Graphics codec for videoconferencing graphics – Mode 1 46
- A.3 Graphics codec for videoconferencing-graphics – Mode 2 48

Annex B – Encryption option – 625 line 50

Annex C – Graphics option – 525 line 50
- C.1 Introduction 50
- C.2 Facilities 50
- C.3 Coding 50
- C.4 Transmission and synchronization 51
- C.5 Decoder 51

Annex D – Encryption option – 525 line 52

Annex E – Colour decoding and coding filters 52
- E.1 Configuration 52
- E.2 Basic filter characteristics 54
- E.3 Advanced filter characteristics 54

Annex F – An example of coding control sequence 55

Annex G – Examples of entropy coding 56

Page

Appendix I – Outline description of operation of codecs in clauses 1 and 2 58

Appendix II 59

II.1 Cursor data 59

II.2 Frame structure 60

II.3 Transmission of cursor data 60

References 60

INTRODUCTION

Clause 1 specifies the codec, developed for operation with the 625-line, 50 field/s television standard and the 2048 kbit/s primary digital group. Its architecture has been chosen to permit variations in the detailed design of certain of the functional elements having the greatest influence on the picture quality. This enables future developments, aimed at improving the performance, to be incorporated without affecting the ability of different coders and decoders to interwork. For this reason, no details are given of such items as motion detectors or spatial and temporal filters. The Recommendation confines itself to the details necessary to enable a decoder correctly to interpret and decode the received signals.

The annexes to clause 1 which can be found at the end of this Recommendation give details of some additional optional features which may be provided to supplement the basic design.

Under the general heading of codecs not requiring separate television standards conversion when used on interregional connections, clause 2 describes a version of the codec for 525-line, 60 field/s and 1544 kbit/s operation which also provides automatic television standards conversion when connected to the version of the codec described in 1 via a re-multiplexing unit (to convert between frame structures defined in 2.1 and 2.3/G.704) at the junction of the 2048 and 1544 kbit/s digital paths. This codec is also suitable for use within regions using the 525-line, 60 field/s television standard and 1544 kbit/s transmission.

Other implementations of 2 are to be studied, for example:

- a version of the codec for 625-line, 50 field/s and 2048 kbit/s operation capable of interworking with the codec described in 3;
- a version of the codec for 525-line, 60 field/s and 2048 kbit/s operation capable of interworking with the codec described in 1.

Clause 3 describes a codec for intra-regional use in 525-line, 60 field/s and 1544 kbit/s regions.

The frame structures associated with the codecs described in this Recommendation are to be found in Recommendation H.130.

As the codecs are complex items using combined intraframe and interframe picture-coding techniques which tend to be known only to specialists, Appendix I is provided giving a brief outline of the principles involved in the codecs of 1 and 2.

Recommendation H.120

CODECS FOR VIDEOCONFERENCING USING PRIMARY DIGITAL GROUP TRANSMISSION

(Malaga-Torremolinos, 1984; amended at Melbourne, 1988 and at Helsinki, 1993)

The CCITT,

considering

(a) that there is growing evidence of a customer demand for a videoconference service;

(b) that circuits to meet this demand can, at present, be provided effectively by digital transmission using the primary digital group;

(c) 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;

(d) that the eventual use of switched digital transmission networks should be taken into account,

appreciating

that rapid advances are being made in research and development of video coding and bit-rate reduction techniques which may lead to further Recommendations being proposed for videoconferencing at bit rates which are multiples or submultiples of 384 kbit/s during subsequent study periods so that this Recommendation may be considered as the first of an evolving series of Recommendations,

and noting

that it is a basic objective of the CCITT to recommend a unique solution for international connections as far as possible,

recommends

that the codecs having signal processing and interface characteristics described in clauses 1, 2 and 3, should be used for international videoconference connections.

NOTE – Codecs of types other than those described in this Recommendation are not precluded.

1 A codec for 625-lines, 50 fields/s and 2048 kbit/s transmission for intra-regional[1)] use and capable of interworking with the codec of clause 2

1.1 Scope

Clause 1 defines the essential features of a codec for the digital transmission, at 2048 kbit/s, of signals for videoconference or visual telephone service in accordance with Recommendation H.100. The video input to the coder and output from the decoder in a 625-line, 50 fields/s signal, according to the "Class *a*" standard of Recommendation H.100, or alternatively, the 313-line, 50 fields/s signal of the "Class *b*" standard. Provision is also made for a sound channel and optional data channels. A brief description of the operation of the codec is given in Appendix I.

---

1) The term "intra-regional" is used here to describe connections within a group of countries which share a common television scanning standard and a common digital hierarchy, and may or may not be in geographical proximity. The term "inter-regional" is used here to describe connections between groups of countries which have different television scanning standards and/or different digital hierarchies.

The Recommendation starts with a brief specification of the codec (see 1.2) and a description of the video interface. This is followed by details of the source coder (see 1.4) which provides analogue-to-digital conversion followed by recoding with substantial redundancy reduction in the face-to-face mode. Subclause 1.5 deals with the video multiplex coder which inserts instructions and addresses into the digitized video signal to control the decoder so that it correctly interprets the signals received. Subclause 1.6 describes the transmission coder which arranges the various digital signals (video, sound, data, signalling) into a form compatible with Recommendation G.732 for transmission over 2048 kbit/s digital paths. Subclause 1.7 describes optional forward error correction facilities. Provision is made in the digital frame structure for the inclusion of other optional facilities such as a graphics mode, encryption and multipoint conferencing. Details of such facilities as are at present available are given in the annexes to this Recommendation.

1.2 Brief specification

1.2.1 Video input/output

The video input and output are standard 625-line, 50 field/s colour or monochrome television signals. The colour signals are in, or converted to, component form. Colour and monochrome operation are fully compatible.

1.2.2 Digital output/input

The digital output and input are at 2048 kbit/s, compatible with the frame structure of Recommendation G.704.

1.2.3 Sampling frequency

The video sampling frequency and the 2048 kHz network clock are asynchronous.

1.2.4 Coding techniques

Conditional replenishment coding supplemented by adaptive digital filtering, differential PCM and variable-length coding are used to achieve low bit-rate transmission.

1.2.5 Audio channel

An audio channel using 64 kbit/s is included. At present, coding is A-law according to Recommendation G.711, but provision is made for future use of more efficient coding.

1.2.6 Mode of operation

The normal mode of operation is full duplex.

1.2.7 Codec-to-network signalling

An optional channel for codec-to-network signalling is included. This conforms to emerging ideas in CCITT for switching 2-Mbit/s paths in the ISDN.

1.2.8 Data channels

Optional 2×64 kbit/s and 1×32 kbit/s data channels are available. These are used for video if not required for data.

1.2.9 Forward error correction

Optional forward error correction is available. This is required only if the long-term error rate of the channel is worse than 1 in 10^6.

1.2.10 Additional facilities

Provision is made in the digital frame structure for the future introduction of encryption, a graphic mode and multipoint facilities.

1.2.11 Propagation delay

When the coder buffer is empty and the decoder buffer full, the coder delay is less than 5 ms and the decoder delay is 130 ± 30 ms at 2 Mbit/s or 160 ± 36 ms when only 1.5 Mbit/s are in use[2)].

1.3 Video interface

The normal video input is a 625-line, 50 field/s signal in accordance with CCIR Recommendation 472. When colour is being transmitted, the input (and output) video signals presented to the analogue/digital convertors (and from the digital/analogue convertors) are in colour-difference component form. The luminance and colour-difference components, E'_Y, $(E'_R - E'_Y)$ and $(E'_B - E'_Y)$ are as defined in CCIR Report 624. The analogue video input (and output) interface with the codec may be in the form of colour-difference components, colour components (R, G, B) or as a composite colour signal. The video interface is as recommended in CCIR Recommendation 656.

Optionally, any other video standard which can be converted to give 143 active lines per field may be used.

1.4 Source coder

1.4.1 Luminance component or monochrome

1.4.1.1 Analogue-to-digital conversion

The signal is sampled to produce 256 picture samples per active line (320 samples per complete line). The sampling pattern is orthogonal and line, field and picture repetitive. For the 625-line input, the sampling frequency is 5.0 MHz, locked to the video waveform.

Uniformly quantized PCM with 8 bits/sample is used.

Black level corresponds to level 16 (00010000).

White level corresponds to level 239 (11101111).

PCM code words outside this range are forbidden (the codes being used for other purposes). For the purposes of prediction and interpolation, the final picture element in each active line (i.e. picture element 255) is set to level 128 in both encoder and decoder.

In all arithmetic operations, 8-bit arithmetic is used and the bits below the binary point are truncated at each stage of division.

1.4.1.2 Pre- and post-filtering

In addition to conventional anti-aliasing filtering prior to analogue-to-digital conversion, a digital transversal filtering operation is carried out on the 625-line signal to reduce the vertical definition of the picture prior to conditional replenishment coding. As a result of this process, 143 active lines per field are used instead of the 287½ active lines of the 625-line signal, although the effective vertical definition is greater than one-half of that of a normal 625-line display. An interpolation process in the decoder restores the 625-line signal waveform.

1.4.1.3 Conditional replenishment coding

A movement detector identifies clusters of picture elements which are deemed to be moving. The basic feature is a frame memory which stores 2 fields of 143 lines, each line containing 256 addressable points. The memory is updated at the picture rate and differences between the incoming signal and the corresponding stored values are used to determine the moving area in the coder. A similar frame memory must exist at the decoder and be similarly updated under the control of addressing information received from the coder. It is not necessary to specify the techniques used for movement detection because they do not affect interworking, although they do affect the resultant picture quality.

Detected moving areas are transmitted by differential PCM with a maximum of 16 quantization levels. The first picture element in each moving area is transmitted by PCM. Variable-length coding is used on the DPCM code words.

---

2) These are typical figures. The delays depend on the detailed implementation used.

The first picture element of each cluster and the complete PCM lines, when they are transmitted to provide systematic or forced updating, are coded in accordance with 1.4.1.1.

1.4.1.3.1 DPCM prediction algorithm

The algorithm used for DPCM prediction is:

$$X = \frac{A + D}{2}, \text{ where X is the sample being predicted.}$$

See Figure 1.

For the purpose of prediction, line and field blanking are assumed to be at level 128 (out of 256).

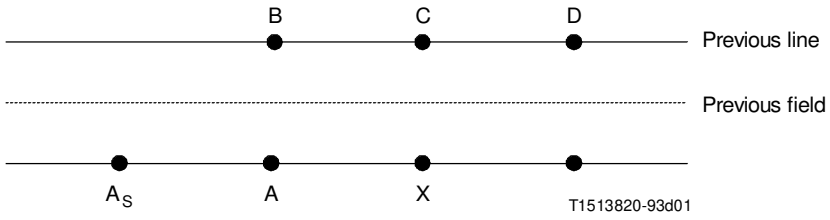


FIGURE 1/H.120

Identification of samples

1.4.1.3.2 Quantization law and variable-length coding

511 input levels are quantized to a maximum of 16 output levels. The quantizer does not assume the use of modulo 256 arithmetic.

The quantization law and associated variable-length codes which are used for both luminance and colour-difference picture elements in moving areas which are not horizontally subsampled are given in Table 1.

Error! Cannot open file.

The end-of-cluster code is 1 0 0 1 and is designated as code number 11. The end-of-cluster code is omitted at the end of the last cluster in a line irrespective of whether it is a luminance cluster or a colour-difference cluster.

1.4.1.4 Subsampling

As the buffer fills, horizontal subsampling and field/field subsampling are introduced.

1.4.1.4.1 Horizontal subsampling

Horizontal subsampling is carried out only in moving areas. Normally, in this mode, only even elements are transmitted on even numbered lines and odd elements on odd numbered lines. This gives rise to a line quincunx pattern in moving areas.

Omitted elements are interpolated in the decoder by averaging the two horizontally adjacent elements.

Interpolated picture elements are placed in the frame stores. A moving area cluster will always start with a PCM value and finish with a transmitted DPCM picture element, even during subsampling. This means that in some instances, the transmitted cluster needs to be extended by one element in comparison with the moving area declared by the movement detector. At the end of the active line, however, this cannot occur as clusters must not extend into blanking, so cluster shortening by one element can be necessary.

TABLE 1/H.120

Code table for non-horizontally-subsampled moving areas

Input levels	Output levels	Variable-length code	Code No.
–255 to –125	–141	1 0 0 0 0 0 0 0 0 1	17
–124 to –95	–108	1 0 0 0 0 0 0 0 1	16
–94 to –70	–81	1 0 0 0 0 0 0 1	15
–69 to –49	–58	1 0 0 0 0 0 1	14
–48 to –32	–39	1 0 0 0 0 1	13
–31 to –19	–24	1 0 0 0 1	12
–18 to –9	–13	1 0 1	10
–8 to –1	–4	1 1	9
0 to 7	+3	0 1	1
8 to 17	+12	0 0 1	2
18 to 30	+23	0 0 0 1	3
31 to 47	+38	0 0 0 0 1	4
48 to 68	+57	0 0 0 0 0 1	5
69 to 93	+80	0 0 0 0 0 0 1	6
94 to 123	+107	0 0 0 0 0 0 0 1	7
124 to 255	+140	0 0 0 0 0 0 0 0 1	8

Adaptive element subsampling allows the transmission of normally omitted elements, either to remove interpolation errors or, to provide a softer switch to subsampling and thus improve the picture quality. The signalling of the extra elements is achieved by using, on horizontally subsampled lines only, 8 quantizing levels for normally transmitted elements and the remaining 8 levels for the extra elements. Also, a cluster can finish either on a normally transmitted element or an "extra" element.

During horizontally subsampled lines, the quantization law and variable-length code shown in Table 2 will be used for both luminance and colour-difference samples in moving areas.

With regard to prediction, if element A is a non-transmitted element in a moving area, it is replaced by A_s (see Figure 1); if element D is part of a subsampled moving area, and not transmitted in the current frame, it is replaced by C.

1.4.1.4.2 Field/field subsampling

Either field can be omitted. In the omitted field, interpolation takes place only in those parts of the picture which are estimated to be moving. "Stationary" areas remain unchanged.

The estimated moving areas are formed from an OR function on the moving areas in the past and future fields, as shown in Figure 2. In the figure, x is a moving element if a OR b OR c OR d are moving.

For the purpose of field interpolation, PCM lines are considered as non-moving and field blanking is assumed to be at a level of 128 out of 256.

TABLE 2/H.120

Quantization law and variable-length code table

Quantization		Variable-length codes			
Input range	Output levels	Normal elements	Code No.	Extra elements	Code No.
–255 to –41	–50	1 0 0 0 0 0 0 1	15	1 0 0 0 0 0 0 0 0 1	17
–40 to –24	–31	1 0 0 0 0 1	13	1 0 0 0 0 0 0 0 1	16
–23 to –11	–16	1 0 1	10	1 0 0 0 0 0 1	14
–10 to –1	–5	1 1	9	1 0 0 0 1	12
0 to +9	+4	0 1	1	0 0 0 1	3
0 to 22	+15	0 0 1	2	0 0 0 0 0 1	5
23 to 39	+30	0 0 0 0 1	4	0 0 0 0 0 0 0 1	7
40 to 255	+49	0 0 0 0 0 0 1	6	0 0 0 0 0 0 0 0 1	8

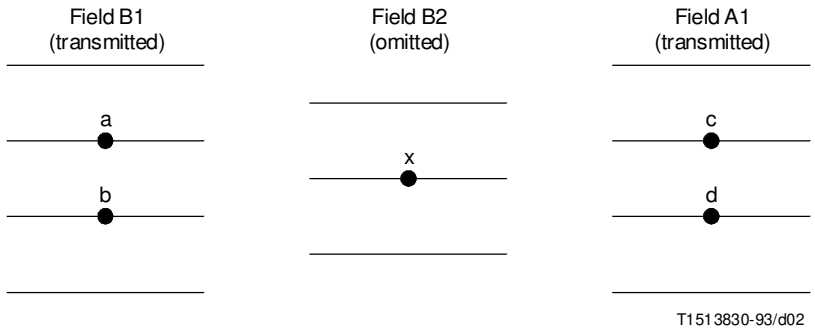


FIGURE 2/H.120

In the interpolator for monochrome or luminance signals, the operations $\frac{a + b}{2}$ and $\frac{c + d}{2}$ are carried out before the combined average is taken. Thus

$$x = \frac{\left[\frac{a + b}{2}\right] + \left[\frac{c + d}{2}\right]}{2}$$

The interpolated values are placed in the frame store.

1.4.2 Colour-difference components

1.4.2.1 Analogue-to-digital conversion

The signal is sampled to produce 52 samples per active line (64 samples per complete line). The sampling pattern is orthogonal and line-, field- and picture-repetitive. For the 625-line input, the sampling frequency is 1.0 MHz, locked to the video waveform.

The $(E'_R - E'_Y)$ and $(E'_B - E'_Y)$ samples are sited so that the centre of the first colour-difference sample on any line is co-sited with the centre of the third luminance sample (addressed as number 2). The $(E'_R - E'_Y)$ and $(E'_B - E'_Y)$ signals are stored and transmitted on alternate lines of the coded picture. The first active line of Field No. 1 contains $(E'_B - E'_Y)$ and the first active line of Field No. 2 contains $(E'_R - E'_Y)$. The colour difference signal not being transmitted during any line is obtained at the decoder by interpolation.

The vertical filtering (see 1.4.2.2) is arranged so that the effective vertical positions of the colour-difference samples in each of the 286 active lines coincide with those of the corresponding luminance samples.

Uniformly quantized PCM with 8 bits/sample is used.

The $(E'_R - E'_Y)$ and $(E'_B - E'_Y)$ signals are quantized using ± 111 steps with zero signal corresponding to level 128. The analogue video signals are amplitude-limited so that the digitized signals do not go outside that range (corresponding to levels 16 to 239). The video levels are set so that a 100/0/75/0 colour bar signal (see CCIR Recommendation 471 for explanation of nomenclature) will occupy levels 17 to 239.

As for the luminance signal, forbidden PCM code words are available for purposes other than transmitting video sample amplitudes.

1.4.2.2 Pre- and post-filtering

In addition to conventional anti-aliasing filtering prior to analogue-digital conversion, a digital transversal filtering operation is carried out on the 625-line signal to reduce the vertical definition of the picture prior to conditional-replenishment coding. As a result of this process, 72 active lines of $(E'_R - E'_Y)$ and 71 active lines of $(E'_B - E'_Y)$ are used in Field No. 2 instead of 287½ active lines per field of a 625-line signal. Similarly, Field No. 1 contains 72 active lines of $(E'_B - E'_Y)$ and 71 lines of $(E'_R - E'_Y)$. An interpolation process in the decoder restores the 625-line signal waveforms.

1.4.2.3 Conditional replenishment coding

Coloured moving areas are detected, coded and addressed separately from the luminance moving areas, but the same principles are employed.

Detected moving areas are transmitted by differential PCM with a maximum of 16 quantization levels. The first picture element in each moving area is transmitted by PCM. Variable-length coding is used on the DPCM code words.

Complete PCM lines are transmitted to provide systematic and forced updating coincident with luminance PCM lines.

1.4.2.3.1 DPCM prediction algorithm

The algorithm used for colour-difference signals is:

$$x = A$$

See Figure 1.

1.4.2.3.2 Quantization law and variable-length coding

As for luminance component (see 1.4.1.3.2 and 1.4.1.4.1).

1.4.2.4 Subsampling

Horizontal subsampling is carried out in exactly the same way as for the luminance signal, including adaptive element subsampling.

Field/field subsampling of the colour-difference signals is also similar to that of the luminance signal. Either field can be omitted and, in the omitted field, interpolation takes place only in those parts of the picture which are estimated to be moving. Stationary areas remain unchanged.

The estimated moving areas are formed by an OR function on moving areas in past and future fields in the same manner as for luminance (see 1.4.1.4.2).

For colour-difference signals, the interpolated value of x is $\left(\frac{a + c}{2}\right)$ or $\left(\frac{b + d}{2}\right)$ when x is in Field 1 or Field 2, respectively.

Both field and horizontal subsampling take place simultaneously with subsampling of the luminance signal and they are signalled to the decoder in the same way.

1.5 Video multiplex coding

1.5.1 Buffer store

The size of the buffer store is defined at the transmitting end only and is 96 kbit/s. Its delay is approximately equal to the duration of one picture (40 ms).

At the receiving end, the buffer must be of at least this length, but in some implementations of the decoder it may be longer.

1.5.2 Video synchronization

The method used for video synchronization permits the retention of the picture structure. The required information is transmitted in the form of line start and field start codes (LST and FST).

1.5.2.1 Line start code

The line start code includes a synchronization word, a line number code and a digit to signal the presence of element subsampling.

It has the format:

0 0 0 0 0 0 0 0 | 0 0 0 0 1 0 0 0 | "S" | 3-bit line No. code |

"S" is a 1 if horizontal subsampling occurs on the TV line following the line start code. "S" is a "don't care" condition on empty or PCM lines.

The line number code comprises the least three significant digits of the line number, where Line 0 = first active line of Field 1 and Line 144 = first active line of Field 2.

Lines numbered 143 and 287 are non-coded lines, used for field synchronization and line number continuity.

1.5.2.2 Field start code

There are two field start codes, FST-1 and FST-2, where the first line of the field following FST-2 is interlaced between the first two lines of the field following FST-1. FST-1 indicates the start of the first field, starting with line number 0. FST-2 indicates the start of the second field, starting with line number 144, as shown in Figure 3.

Each field start code comprises a line start code, followed by an 8-bit word, followed by the line start code of the first line of the next field.

The field start code is given in Figure 4.

For FST-1, F = 1 and for FST-2, F = 0. A = 0 for normal operation. If required, A = 1 is used to signal that the buffer state is less than 6 kbits (used in switched multipoint applications). S is the subsampling digit as defined in 1.5.2.1.

Field subsampling is signalled by two consecutive field start codes of the same number. For example:

| FST-1 | field of data | FST-1 | field of data |

signifies that field 2 has been omitted and that its moving areas must be interpolated as described in 1.4.1.4.2 and 1.4.2.4.

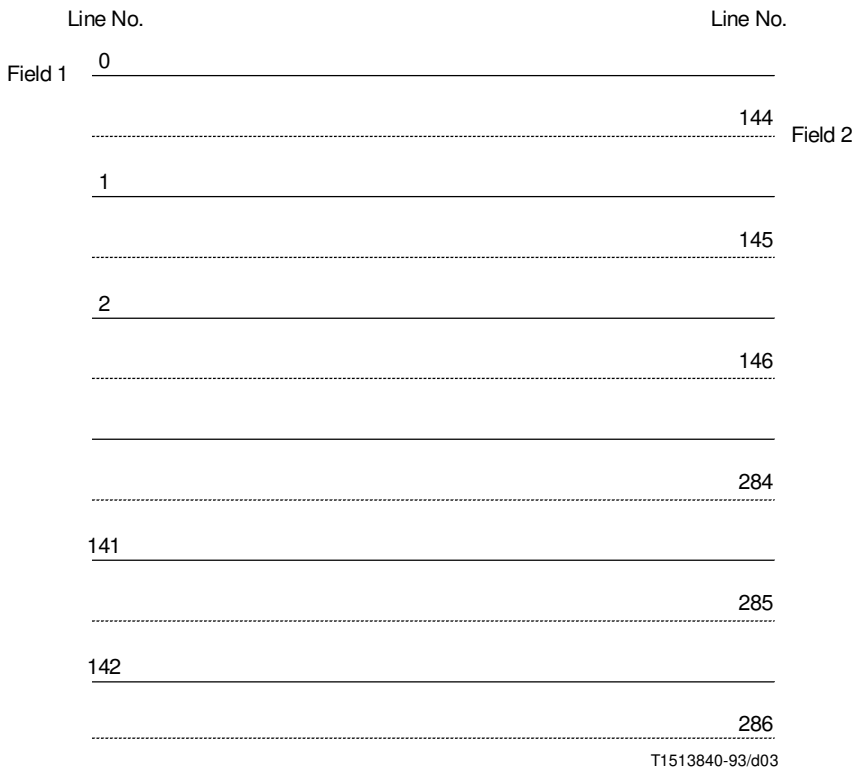


FIGURE 3/H.120

LST					LST			
00000000	00001AAA	F	111	0000F11F	00000000	00001000	S	000

FIGURE 4/H.120

1.5.3 Addressing of moving areas

The positions of the clusters of picture elements along each line which are deemed to form parts of moving areas are addressed by means of an address of the start of the cluster and an "end-of-cluster" code (EOC).

The form of coding is:

LST	PCM value	8-bit address of the PCM picture element	Variable-length DPCM coded moving area	EOC	PCM value	8-bit address	etc.

The PCM value is the amplitude of the first picture element of the cluster. When there is no colour-difference data, the EOC is omitted in the last luminance cluster of every line, i.e. both the LST and FST codes also signify end of cluster.

The EOC is 1001.

The address indicates the sample number along the line belonging to the first picture element of the cluster.

A cluster cannot start at the last element of the line, i.e. (11111111) is a forbidden cluster address, nor can it extend into line blanking even during subsampling.

The minimum gap between the end of one cluster and the start of the next is four picture elements, and the minimum length of a cluster is one picture element.

1.5.4 Addressing of colour-difference data

To permit the insertion of colour-difference data in a line containing moving picture elements, a colour escape code is inserted after the final luminance cluster in the line. This permits addresses to be re-used for colour clusters.

The escape code is 00001001 (an invalid PCM value) and follows the end of cluster code of the last luminance cluster (if any), otherwise it follows the line start code. This is followed by the addresses, variable-length codes and EOC codes of the subsequent coloured clusters, the sequence being terminated by the line start code of the next line.

The form of addressing of colour-difference moving areas is shown in Figure 5.

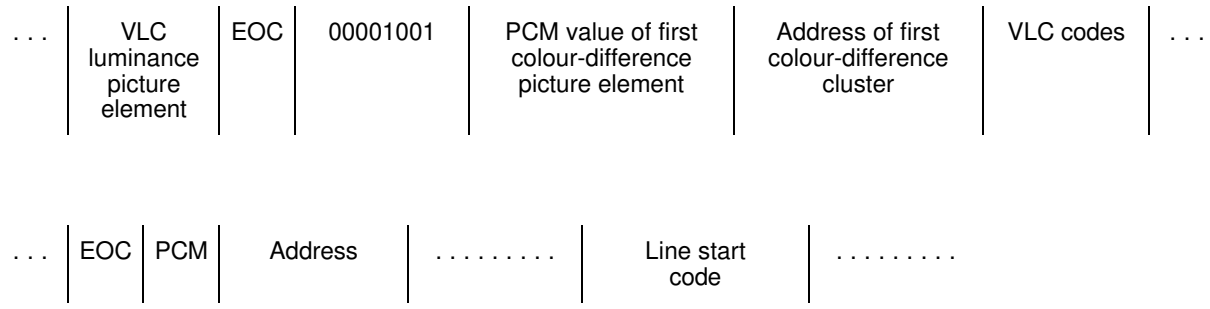


FIGURE 5/H.120

There are 52 colour-difference picture elements per line, the first of which is given an address with numerical value 4. The address range is therefore:

00000100 to 00110111.

A cluster cannot start at address (00110111) nor can it extend beyond this point, even during subsampling. The minimum gap between the end of one colour-difference cluster and the start of the next is 4 picture elements. The minimum cluster length is one picture element. Cluster bridging is not allowed between luminance and colour-difference clusters.

A monochrome decoder will discard the information between the colour escape code and the next line start code.

1.5.5 PCM lines

PCM lines are used for systematic or forced updating and are signalled as shown in Figure 6.

	Invalid PCM code	Invalid cluster address	PCM value of first picture element of line	254 × 8 bit PCM values	
LST	1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1	X X X X X X X X	X X . . .	1 0 0 0 0 0 0 0

FIGURE 6/H.120

With monochrome, all 256 elements in the line are transmitted using 8-bit PCM.

With PCM lines, the subsampling digit "S" is ignored at the receiver. PCM lines cannot be horizontally subsampled.

For the purpose of field interpolation, PCM lines are considered to be non-moving.

With colour signals, the colour-difference data will comprise 52 × 8-bit PCM values following the 256 × 8-bit luminance elements. The colour escape code is not transmitted. A monochrome decoder will discard the colour-difference picture elements.

1.6 Transmission coding

The transmission coder assembles the video, audio, signalling and optional data channels into a 2048-kbit/s frame structure which is in accordance with Recommendation G.704. It also provides justification facilities to enable the video sampling frequency to be independent of the network clock.

1.6.1 Serial data

With all serialized data (video, audio and addressing), the most significant digit is in the leading position. Positive logic is used throughout.

1.6.2 Audio

The audio is coded into 64 kbit/s using A-law PCM, as specified in Recommendation G.711.

In the coder, the delay difference between the coded audio and video when the buffer is empty should be within ± 5 ms. In the decoder, the delays must also be equalized, and the tolerance is under study.

The audio output should be muted in the event of loss of frame alignment.

1.6.3 Transmission framing

1.6.3.1 General

The frame structure is defined in Recommendation H.130 which specifies how the frame is structured and for what purposes the time slots are used. This information need not be repeated here.

Time slot 2 (odd) is allocated to codec-to-codec signalling and the functions of the various bits are specified in Recommendation H.130. In most cases, the action to be taken by the encoder and/or decoder according to whether each of these bits is 0 or 1 is evident from the specified purpose of the bit. In the few cases where this is not so, additional information is given below.

1.6.3.2 Use of certain bits in each octet in the odd frames of time slot 2

Studies to determine the most suitable methods for multipoint conferencing are still in progress, but from the preliminary results, a number of special features and facilities have been identified as being necessary and have therefore been included in the codec and frame structure. In "continuous presence" multipoint conferencing, a transmission channel may, at times, be shared by two codecs in different locations. This requires a reduction of the bit rate of each source so that the total bit rate is within the capacity of the channel. The "facilities bits", i.e. bits (see Recommendation H.130) 3.1.2 and 3.1.7, are used to signal the availability of this facility and the bits 4.9 and 4.15, signal the mode of operation and the active time slots in use at the output of the transmission coder. The details of how the foregoing bits are interpreted are given in Recommendation H.130.

Bits 3.7 to 3.15 also provide facilities whose principal usefulness is likely to be multipoint conferencing. Information on the use of these bits together with details of the use of bits 1 and 2, which are essential to the basic requirements of keeping the encoder and decoder in step, are given below:

Bit 1 – For clock justification

The frequency control arrangements are as follows:

The video sampling clock is locked to the line-scanning frequency of the incoming video signal, which has a permitted tolerance of ± 2 parts in 10^4.

The justification is controlled by a comparison frequency of (22 500/11) kHz, which is locked to the video clock.

The clock for the digital channel has a frequency of 2048 kHz ± 50 parts in 10^6.

The phase of the channel clock is compared with that of the comparison frequency and when the channel-clock phase exceeds that of the comparison frequency by 2π radians, a 1 is transmitted. If the phase difference is less than 2π radians, a 0 is transmitted.

Bit 2 – To signal buffer state

The degree to which the encoder buffer is filled, measured in increments of 1 K (1 K = 1024 bits), is signalled using an 8-bit binary code. The most significant bit (MSB) is in frame 1 of the multiframe, the second MSB in frame 2, etc. The buffer state is sampled at the start of the multiframe in which its state is transmitted.

Bit 3.7 – Fast update request

On receipt of this bit set to 1, the transmitter buffer is forced to decrease its fill and stabilize to a state of less than 6 K by preventing coded picture elements from entering the buffer. Bit A is set to 1 in the next FST. The two following fields are treated as complete moving areas and the encoder uses an arrangement for control of the subsampling modes to make the buffer overflow condition unlikely.

Bit 3.9 – Advance warning of interruption

This bit (set to 1) is used to warn a decoder that its received signal may be interrupted after the start of the next supermultiframe for a period of not more than two seconds. On receipt of bit 3.9 set to 1, a decoder will display a still picture for a period of not more than 2 s, or until an FST code is received with bit A set to 1.

Bit 3.11 – Sound power signal

This bit is used to signal the sound power in the audio channel. The power is integrated over a period of 16 ms (period of the supermultiframe), uniformly quantized to 8 bits and transmitted at supermultiframe rate. It is used during encrypted multipoint operations. In other cases, bit 3.11 is set to zero.

Bit 3.13 – Data distribution

This bit is set permanently to 0 in all encoders. When a 1 is received from the network (introduced, for example, by a multipoint control unit), the encoder will vacate the same time slots in its outgoing signal as signalled on the incoming stream by the settings of the relevant bits 4 (which identify the use of the time slots, see Recommendation H.130). It will confirm the action by transmitting the same bit 4 settings as received. This function should be carried out within 10 supermultiframe periods.

Bit 3.15 – Detection of looped ports

This bit is set to 1 in all codecs. It may be used by a multipoint control unit to detect whether one of its bidirectional 2 Mbit/s ports has been externally looped.

1.7 Error correction

Provision is made for the optional use of forward error correction. This is required if the channel error rate is worse than 1 in 10^6 for significant periods of time. The error corrector used is a (4095, 4035) five-error correcting BCH code (*Bose, Chaudhuri and Hocquengham*). The error-correcting decoder has the ability to correct up to 5 isolated errors and one burst of up to 16 errors in each block. At a channel error probability of 1×10^{-4}, the corrected error rate is 1.25×10^{-8}. The 60 parity bits which are required are obtained by removing the video from time slots 24 to 31 of frame number 15 of each multiframe.

NOTE – The question of whether error correction should be provided on the signal, on the link, or both, should be studied. Also under study is the question of whether the audio should be corrected by the same error corrector or whether a separate error-correcting codec should be used.

2 Codecs not requiring separate television standards conversion when used on interregional connections

A codec for 525-line, 60 fields/s and 1544 kbit/s transmission for intra-regional use and capable of interworking with the codec of clause 1

2.1 Introduction

Clause 2 indicates the changes and additions which must be made to the text of 1 in order to define the version of the codec for use with 525-line, 60 fields/s television standards and transmission at 1544 kbit/s. The two versions are capable of interworking via a re-multiplexing unit which can convert the 2.1/G.704 compatible frame structure on one side to the 2.3/G.704 compatible frame structure (with 6 time slots empty) on the other side.

The two versions of the codec are identical in most respects, the important differences (apart from the obvious ones arising from different input and output signals) being confined to the digital pre- and post-filters and the signals for the control of the buffers. Moreover, the detailed algorithms of the pre- and post-filters do not need to be specified to permit interworking. Only an outline of their mode of operation together with the few necessary specifications are therefore provided.

2.2 Brief specification

2.2.1 Video input/output

The video input and output are standard 525-line, 60 fields/s colour or monochrome television signals. The colour signals are in component form. Colour and monochrome operation are fully compatible.

2.2.2 Digital output/input

The digital output and input are at 1544 kbit/s, compatible with the frame structure of Recommendation G.704.

2.2.3 Sampling frequency

The video sampling frequency and 1544 kbit/s network clock are asynchronous.

2.2.4 Coding techniques

Conditional replenishment coding supplemented by adaptive digital filtering, differential PCM and variable-length coding are used to achieve low bit-rate transmission.

2.2.5 Audio channel

An audio channel using 64 kbit/s is included. At present, coding is A-law according to Recommendation G.711, but provision is made for future use of more efficient coding.

2.2.6 Mode of operation

The normal mode of operation is full duplex.

2.2.7 Codec-to-network signalling

An optional channel for codec-to-network signalling is included.

2.2.8 Data channels

Optional 2×64 kbit/s and 1×32 kbit/s data channels are available. These are used for video if not required for data.

2.2.9 Forward error correction

Optional forward error correction is available. This is required only if the long-term error rate of the channel is worse than 1 in 10^6.

2.2.10 Additional facilities

Provision is made in the digital frame structure for the future introduction of encryption, a graphic mode and multipoint facilities.

2.2.11 When the coder buffer is empty and the decoder buffer full, the coder delay is 31 ± 5 ms and the decoder delay is 176 ± 31 ms[3)] .

2.3 Video interface

The normal video input is a 525-line, 60 fields/s signal in accordance with CCIR Report 624. When colour is being transmitted, the input (and output) video signals are in component form. The luminance and colour-difference components, E'_Y, $(E'_R - E'_Y)$ and $(E'_B - E'_Y)$ are as defined in CCIR Report 624. The video interface is as recommended in CCIR Recommendation 567.

2.4 Source coder

2.4.1 Luminance component or monochrome

2.4.1.1 Analogue-to-digital conversion

The signal is sampled to produce 256 picture samples per active line (320 samples per complete line). The sampling pattern is orthogonal and line, field and picture repetitive. For the 525-line input, the sampling frequency is 5.0 MHz, locked to the video waveform.

Uniformly quantized PCM with 8 bits/sample is used.

Black level corresponds to level 16 (00010000).

White level corresponds to level 239 (11101111).

PCM code words outside this range are forbidden (the codes being used for other purposes). For the purposes of prediction and interpolation, the final picture element in each active line (i.e. picture element 255) is set to level 128 in both encoder and decoder.

In all arithmetic operations, 8-bit arithmetic is used and the bits below the binary point are truncated at each stage of division.

2.4.1.2 Pre- and post-filtering

2.4.1.2.1 Spatial filtering

A digital filter reduces the 242½ active lines-per-field of the 525-line signal to 143 lines-per-field, the same number as in the 625-line version of the codec. In the decoder, the digital post-filter uses interpolation to restore the signal to 525-lines per picture.

2.4.1.2.2 Temporal filtering

A recursive temporal pre-filter with non-linear transfer characteristics is used in the coder to reduce noise in the signal and increase coding efficiency. The frame store used in this filter can also be used as the storage element of a frame interpolator with variable coefficients which is used to reduce the transmitted frame rate to a value less than that of the input video signal. In 525-line to 525-line transmission, the transmitted frame frequency is locked to the video clock and is approximately 29.67 Hz (29.97 Hz times 3057/3088) instead of the nominal video rate of 29.97 Hz. In 525-line to 625-line transmission, the transmitted frame frequency is nominally 25 Hz and is locked to the channel clock.

Because the (television) frames are leaving the coder more slowly than they are entering, the coding process is suspended for one frame every *N*th input frame. *N* is approximately 100 for 525-line to 525-line operation and approximately 6 for 525-line to 625-line operation.

In the decoder, the digital post-filter incorporates a frame store in some versions of the 625-line codec where it is used in the line interpolation process. In the 525-line version, in addition to its use for line interpolation, it is used as a temporal interpolator with variable coefficients to provide an extra output frame during those periods when the decoding is temporarily suspended.

---

3) These are typical figures. The delays depend upon the detailed implementation used.

2.5 Video multiplex coding

2.5.1 Buffer store

The size of the buffer store is defined at the transmitting end only and is 160 kbits. Of this, 96 kbits is used for smoothing the video data in the face-to-face mode and the remainder is used to accommodate the action of the frame interpolator (see 2.5.1.1) and the requirements of the graphics mode.

At the receiving end, the buffer must be at least this length but in some implementations of the decoder, it may be longer.

2.5.1.1 Buffer control

The amount to which the transmitting buffer is filled is used to control various coding algorithms (subsampling, etc.) and is signalled to the decoder to enable it correctly to interpret the received signals. In the 525-line codec, the transmission rate is less than the video input rate and hence the buffer tends to fill more rapidly than would be determined by the movement in the picture, only to empty again when the interpolator suspends the coding process.

To avoid incorrect changes in coding algorithms, the buffer-state signal is modified to take account of the progressively changing coefficients of the interpolator in the pre-filter. The buffer then operates as though the data is coming from a video source whose frame rate is uniform and the same as the transmitted frame rate.

2.6 Transmission coding

The transmission coder assembles the video, audio, signalling and optional data channels into a 1544 kbit/s frame structure which is compatible with Recommendation G.704.

2.6.1 Serial data

See 1.6.1.

2.6.2 Audio

See 1.6.2.

2.6.3 Transmission framing

The frame structure, compatible with Recommendation G.704 and also compatible with that of the 625-line version in clause 1, is given in 2/H.130.

2.6.3.1 General

See 1.6.3.1.

2.6.3.2 Use of certain bits in each octet in the odd frames of time slot 2

The use of certain of the bits in time slot 2 (odd) differs slightly from that given for the codec in clause 1. The differences are as follows:

Bit 1 – For clock justification

This bit is disregarded in 525-line decoders.

To permit interworking with the 625-line codecs of clause 1, the 525-line coders must transmit a fixed bit-pattern which is used to control the frequency of the video clock in 625-line decoders. The exact form of the repetitive pattern need not be specified but it must contain seven “ones” and four “zeros” in 11 bits, e.g.:

1 0 1 1 0 1 0 1 1 0 1

Bit 2 – To signal buffer state

The degree to which the encoder buffer is filled, after correction for the interpolator (see 2.5.1.1), is measured in increments of 1 K (1 K = 1024 bits), and signalled using an 8-bit binary code. When working to a 525-line decoder, the buffer state is sampled every 3057 channel-clock periods. When working to a 625-line decoder, the buffer state is sampled 10 times during every 525-line field period. When the buffer input is suspended for a frame period, the buffer

sampling is stopped. The sampled values of the buffer state are stored prior to transmission. The store may hold between zero and 23 values which have been modified to take account of the interpolator coefficients at the times of sampling. The modified sample values are read out [as bit 2 of TS2 (odd)] at a uniform rate; the most significant bit (MSB) in frame 1 of the multiframe, the second MSB in frame 2, etc.

Bit 3.7 – Fast update request

On receipt of this bit set to 1, the transmitter buffer is forced to decrease its full and stabilise to a modified state of less than 6 K by preventing coded picture elements from entering the buffer. Bit A is set to 1 in the next FST. The two following fields are treated as complete moving areas and the encoder uses an arrangement for control of the sub-sampling modes to make the buffer overflow condition unlikely.

3 A codec for 525-lines, 60 fields/s and 1544 kbit/s transmission for intra-regional use

3.1 Introduction

A 1.5 Mbit/s interframe codec described under 3, is capable of transmitting and receiving a single NTSC video signal and audio signal using an adaptive predictive coding technique with motion-compensated prediction, background prediction and intraframe prediction.

The aim of this codec is to effectively transmit video telephone and video conferencing signals which have relatively small movements. The video interface of the codec is a 525-line, 60 fields/s standard analogue television signal corresponding to the "Class *a*" standard of Recommendation H.100.

3.2 Outline of codec

The essential parts of the codec block diagram are shown in Figure 7. The coder consists of three basic functional blocks, that is, pre-processing, video source coding and transmission coding.

In the pre-processor, the input analogue NTSC video signal is digitized and colour decoded into one luminance component and two chrominance components. These three components are time division multiplexed into a digital video form, whose noise and unnecessary signal components are removed by the pre-filter.

In the video source coder, the digital video signal is fed to the predictive coder where interframe and intraframe predictive coding techniques are fully utilized for minimizing prediction errors to be transmitted. The prediction error signal is next entropy-coded using its statistical properties to reduce redundancies. Since the coded error information is generated in irregularly spaced bursts, a buffer is used. If the buffer becomes full, the number of prediction error quantizing levels and/or picture elements to be coded is reduced to prevent any overflow.

In the transmission coder, coded video and audio signals are first encrypted on an optional basis. The coded video signal is then forward error correction coded and scrambled. The three signals, coded video, coded audio and optional data signals are multiplexed into a 1544 kbit/s digital format with a frame structure as defined in Recommendation H.130.

The decoder carries out a reverse operation.

3.3 Brief specification

3.3.1 Video input/output

NTSC signals are used for the video input/output signal, with monochrome signals being additionally applicable.

3.3.2 Digital output/input

The interface conditions for the digital output/input signal satisfy Recommendation G.703 specifications. The signal transmission rate is 1544 kbit/s.

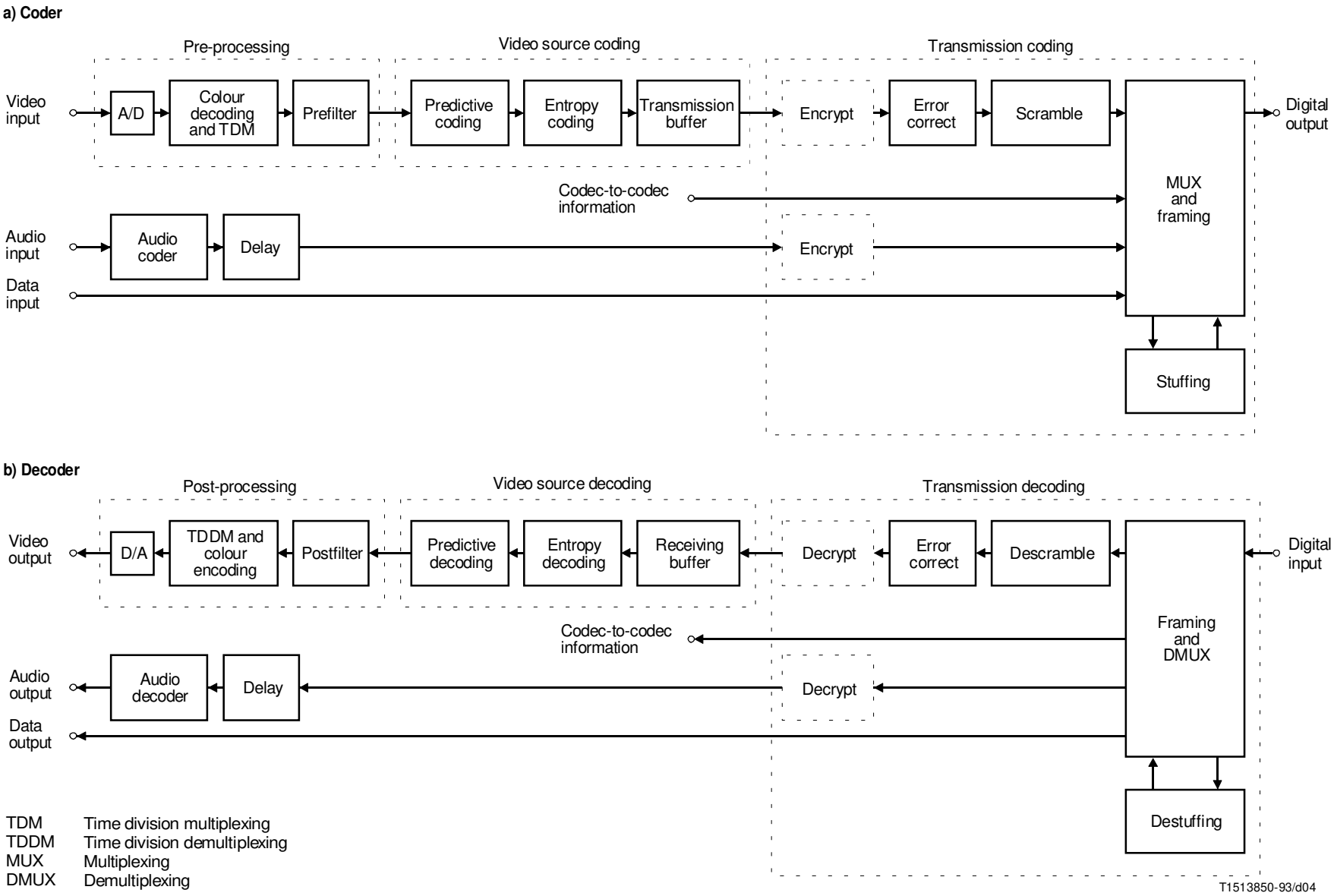


FIGURE 7/H.120

Codec blockdiagram

3.3.3 Sampling frequency

The video sampling frequency is four times the colour sub-carrier frequency (f_{SC}) and asynchronous with the 1544 kHz network clock.

3.3.4 Time division multiplexed (TDM) digital video format

An NTSC signal is separated into a luminance component (Y) and two chrominance components (C_1 and C_2). A time division multiplexed signal composed of Y and time-compressed C_1 and C_2 is employed in the source coding as the standard digital video format.

3.3.5 Coding algorithm

Adaptive predictive coding supplemented by variable word-length coding is used to achieve low bit rate transmission. The following three predictions are carried out adaptively on a pel-by-pel basis:

a) motion-compensated interframe prediction for a still or slowly moving area;

b) background prediction for an uncovered background area; and

c) intraframe prediction for a rapidly moving area.

Prediction errors for video signals and motion vectors are both entropy-coded using the following two techniques:

i) variable word-length coding for non-zero errors; and

ii) run-length coding for zero errors.

3.3.6 Audio channel

An audio channel using 64 kbit/s is included. The audio coding algorithm complies with Recommendation G.722.

3.3.7 Data channel

An optional 64 kbit/s data channel is available, which is used for video if not required for data.

3.3.8 Mode of operation

The normal mode of operation is full duplex, with other modes, e.g. the one-way broadcasting operation mode, also taken into account.

3.3.9 Transmission error protection

A BCH error correcting code is used along with a demand refreshing method to prevent uncorrected errors from degrading the picture quality.

3.3.10 Additional facilities

Provision is made in the digital frame structure for the future introduction of such facilities as encryption, graphics transmission and multipoint communication.

3.3.11 Processing delay

The coder plus decoder delay is about 165 ms without that of a pre-filter and a post-filter.

3.4 Video interface

The video input/output signal of the codec is an analogue NTSC signal (System M) in accordance with CCIR Report 624.

3.5 Pre- and post-processing

3.5.1 Analogue-to-digital and digital-to-analogue conversion

An NTSC signal band-limited to 4.5 MHz is sampled at a rate of 14.3 MHz, four times the colour sub-carrier frequency (f_{SC}), and converted to an 8-bit linear PCM signal. The sampling clock is locked to the horizontal synchronization of the NTSC signal. Since the sampling frequency is asynchronous with the network clock, the justification information is coded and transmitted from the coder to the decoder.

The digital video data is expressed in two's complement form. The input level to the A/D converter is defined as follows:

– sinc tip level (-40 IRE) corresponds to –124 (10000100);

– white level (100 IRE) corresponds to 72 (01001000).

(IRE: Institute of Radio Engineers)

As a national option, a pad can be inserted before the A/D converter if a level fluctuation should be taken into account at analogue transmission lines connecting terminal equipment and codec.

At the decoder, the NTSC signal is reproduced by converting the 8-bit PCM signal to an analogue signal.

3.5.2 Colour decoding and encoding

The digitized NTSC signal is separated into the luminance component (Y) and the carrier band chrominance component (C) by digital filtering. The two baseband chrominance signals (C_1 and C_2) are obtained by digitally demodulating the separated carrier band chrominance component. The effective sampling frequency after colour decoding is converted to 7.2 MHz (2 f_{SC}) and 1.2 MHz (1/3 f_{SC}) for the luminance signal and chrominance signals, respectively.

The replica of the NTSC signal is obtained by digitally modulating the C_1 and C_2 signals and adding to the Y signal at the decoder.

Filter characteristics for colour decoding and encoding are left to each hardware implementation since they do not affect interworking between different design codecs. Examples of recommended characteristics are described in Annex E.

3.5.3 TDM signal

A time division multiplexing (TDM) signal is constructed from the separated component signals.

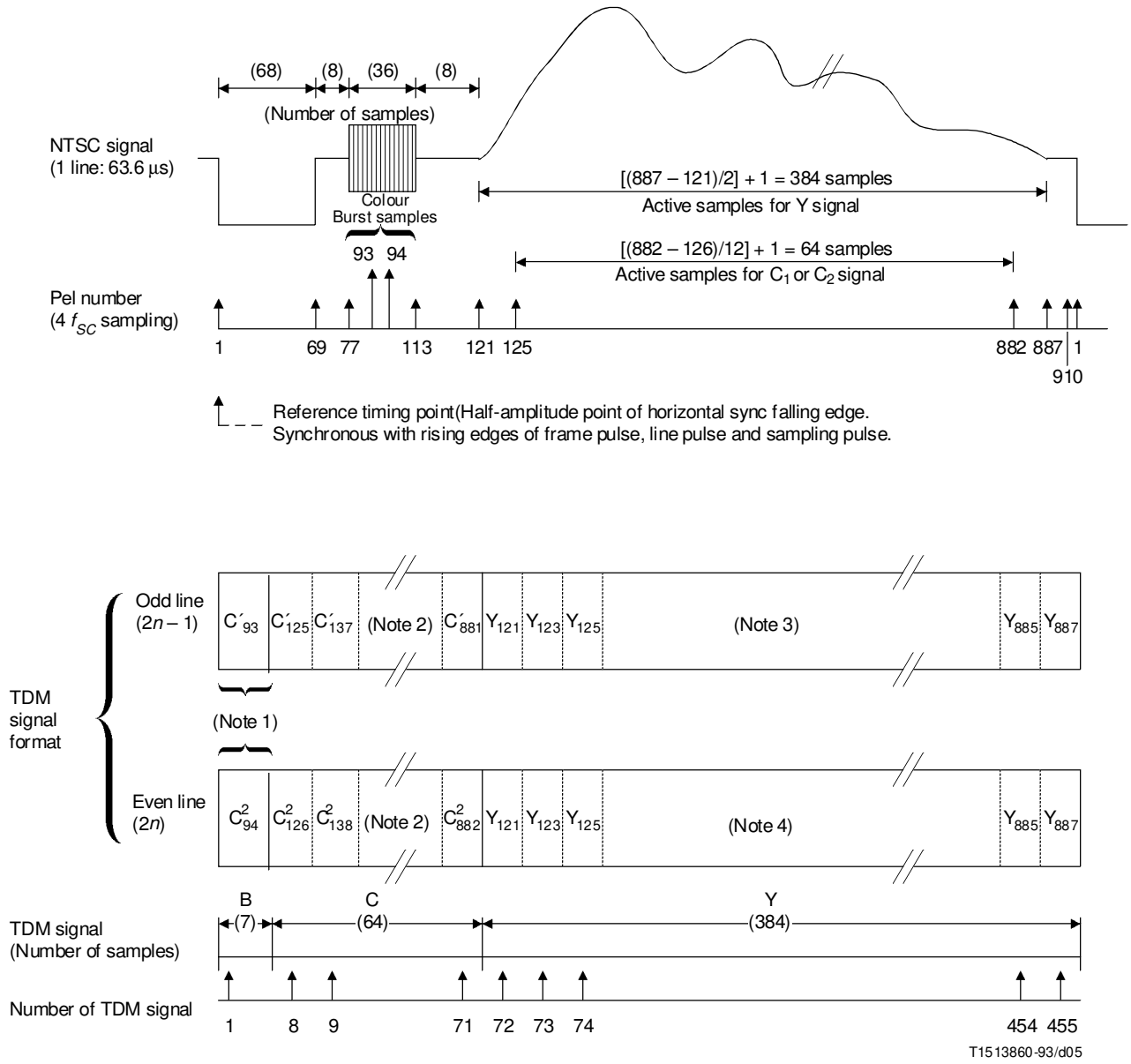
First, the C_1 and C_2 signals are time-compressed to 1/6. Next, each of the time compressed C_1 and C_2 signals, with their horizontal blanking parts removed, is inserted into the Y signal horizontal blanking interval on alternate lines. C_1 is inserted on the first line of the first field and on every other line following throughout the frame, while C_2 is inserted on the second line of the first field and on every other line following throughout the frame.

Active samples for the Y signal are 384 samples/line and 64 samples/line for the C_1 and C_2 signals. The TDM signal is constructed with these active samples and 7 colour burst samples (B), which are inserted into the top of the TDM signal.

As shown in Figure 8, the C_1 and C_2 signal sampling points coincide with that of the Y signal on every sixth sample. The C_1 and C_2 signals of only the odd lines are transmitted to the decoder.

At the decoder, each component signal is again demultiplexed from the TDM signal, and time-expansion processing of 6 times is carried out for the C_1 and C_2 signals.

NOTE – When a pad is inserted before the A/D converter as described in 3.5.1, pre-emphasis (de-emphasis) with a compensating gain for the C_1, C_2 and colour burst signals is recommended at the source coder input (decoder output) to obtain better picture reproduction in coloured parts.



NOTES

1 Odd line samples. A colour burst sample is repeated seven times.

2 Odd line samples.

3 Odd line samples.

4 Even line samples.

FIGURE 8/H.120

TDM signal format

3.5.4 Pre- and post-filtering

In addition to conventional anti-aliasing filtering prior to analogue-to-digital conversion, the following two filtering processes should be used as pre-filtering for source coding:

a) temporal filtering to reduce random noise included in the input video signal;

b) spatial filtering to reduce aliasing distortion in subsampling.

At the decoder, the following three filtering processes should be used as post-filtering in addition to conventional low pass filtering after digital-to-analogue conversion:

i) spatial filtering to interpolate the omitted picture elements in subsampling;

ii) spatio-temporal filtering to interpolate the omitted fields in field repetition;

iii) temporal filtering to reduce noise generated in the course of source coding.

Although these filtering processes are important for improving reproduced picture quality, their characteristics are independent of interworking between different design codecs. Hence, pre- and post-filtering is left to each hardware implementation.

3.6 Source coding

3.6.1 Configuration of source coder and decoder

The video source coder and decoder configuration of this codec is outlined in Figure 9.

The predictive encoder converts the input video signal x into the prediction error signal e, using the motion vector v. This conversion is controlled by the coding mode m.

The variable word-length (VWL) coder codes e and v into the compressed data C using the variable length coding method. The transmission buffer memory (BM) smoothes out the irregularly spaced data C. The coding mode m is also coded.

The frame memory parity information p is used to check the identity of coder and decoder frame memory contents. If any parity error is detected, frame memories of both coder and decoder are reset by the demand refresh information (DR) and the demand refresh confirmation information (DDR).

At the decoder, the variable word-length (VWL) decoder decodes e, v, m and p, and the predictive decoder reproduces the video signal x'.

3.6.2 Predictive coding

3.6.2.1 Coding modes

Five coding modes as summarized in Table 3 are provided. All of the samples are coded and transmitted in normal mode, while half of the samples are omitted in subsampling mode. In field repetition mode, one or more consecutive fields are omitted (called multi-field repetition, see Note 1). If field repetition mode and subsampling mode are used in combination, only a quarter or less of the original picture elements are coded and transmitted.

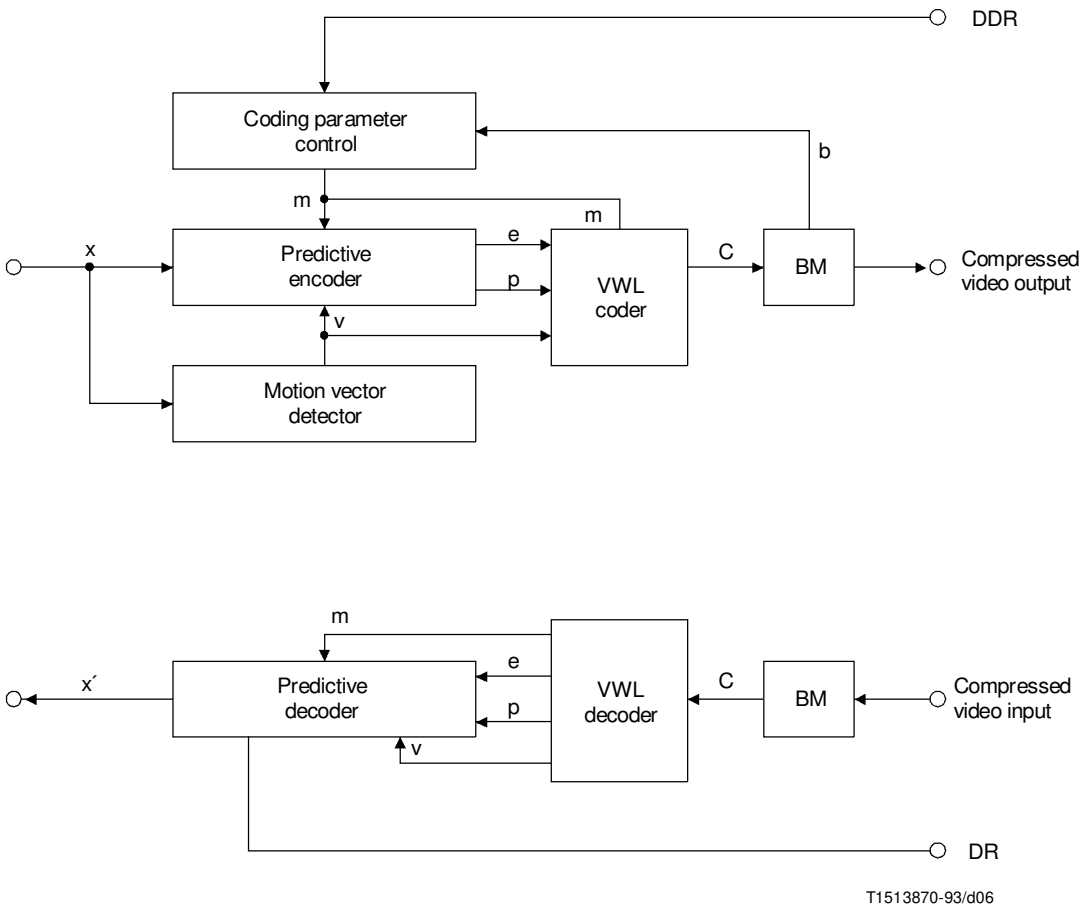
Subsampling is carried out in a quincunx way, namely by transmitting only odd-numbered pels on odd-numbered lines and even-numbered pels on even-numbered lines in each block-line (see Note 2).

In field repetition mode, either the odd or even fields are omitted. For the omitted fields, both the prediction error e and the motion vector v are set to 0.

NOTES

1 If odd fields and even fields are mixed after field omission, a severe picture degradation takes place. Hence, 1 out of 2, 3 out of 4 or 5 out of 6 field omission is recommended.

2 Each block-line consists of 8 lines as defined in 3.6.2.5.



x	Input video signal
x´	Output video signal
m	Coding mode
e	Prediction error
v	Motion vector
p	Parity check information
C	Compressed data
b	Buffer memory occupancy information
VWL	Variable word length
BM	Buffer memory
DR	Demand refresh request information
DDR	Demand refresh confirmation information

FIGURE 9/H.120

Source coder and decoder configuration

3.6.2.2 Adaptive prediction

Prediction functions are adaptively selected on a pel-by-pel basis as shown in Figure 10. The selection is carried out so as to minimize probable prediction errors. This is accomplished using the two prediction status signals, which are determined by prediction reference signals, for the preceding pels located on the previous and the present lines.

When subsampling and/or field repetition are operated, omitted pels are interpolated in the prediction loop.

TABLE 3/H.120

Coding modes

Coding modes		Abbreviation	Operation
1	Normal	NRM	Full sampling
2	Field repetition	FRP	One or more fields omission
3	Subsampling	SBS	2:1 per omission
4	Stop	STP	Suspension of coding
5	Refresh	RFS	Renewal of frame memory

The notations defined for the i-numbered pel are as follows:

X_i local decoder output;

Y_i interpolator output;

M_i motion compensated interframe prediction value;

B_i background prediction value;

I_i intraframe prediction value;

$*$ logical product; and

$+$ logical sum.

3.6.2.2.1 Motion-compensated interframe prediction/background prediction

Prediction status signal S_{1i} for pel i is determined as

$$S_{1i} = R_1(i-455) * R_1(i-456) + R_1(i-456) * R_1(i-454) + R_1(i-454) * R_1(i-455) \quad (3\text{-}1)$$

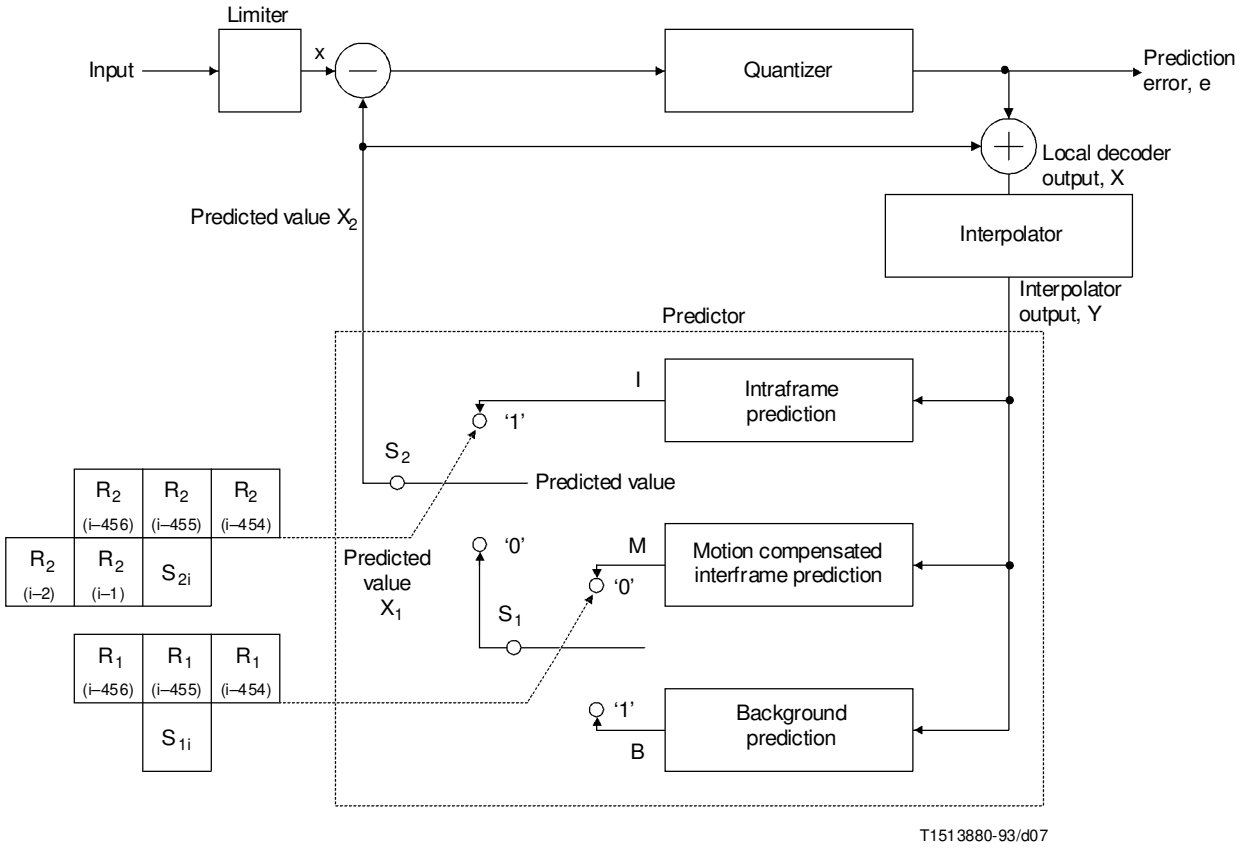
where prediction reference signal $R_1(i)$ is

$$R_1(i) = \begin{cases} 0 \text{ if } \left| Y_i - B_i \right| \geq \left| Y_i - M_i \right|, \\ 1 \text{ if otherwise.} \end{cases} \quad (3\text{-}2)$$

Based on S_{1i}, prediction signal X_{1i} is given as

$$X_{1i} = \begin{cases} M_i, \text{ if } S_{1i} = 0, \\ B_i, \text{ if } S_{1i} = 1. \end{cases} \quad (3\text{-}3)$$

If pel i is either omitted due to subsampling and/or field repetition or forced intraframe coded or in burst B, its corresponding $R_i(i)$ is set to 0 regardless of equation (3-2).



S Prediction status signal
R Prediction reference signal

FIGURE 10/H.120
Adaptive prediction

3.6.2.2.2 Interframe prediction/intraframe prediction

Prediction status signal S_{2i} for pel i is determined as

$$S_{2i} = R_2 (i - 1) * R_2 (i - 455) \tag{3-4}$$

where prediction reference signal R_2 (i) is

$$R_2 (i) = \begin{cases} 0 \text{ if } \left| Y_i - I_i \right| \geq \left| Y_i - X_{1i} \right|, \\ 1 \text{ if otherwise.} \end{cases} \tag{3-5}$$

Based on S_{2i}, prediction signal X_{2i} is given as

$$X_{2i} = \begin{cases} X_{1i} \text{ if } S_{2i} = 0, \\ I_i \text{ if } S_{2i} = 1. \end{cases} \tag{3-6}$$

If pel ($i - 1$) is omitted due to subsampling, R_2 ($i - 2$) is used instead of R_2 ($i - 1$). On the other hand, if pel ($i - 455$) is omitted, R_2 ($i - 454$) $*$ R_2 ($i - 456$) is used instead of R_2 ($i - 455$). If pel i is forced intraframe-coded, its corresponding R_2 (i) is set to 1 regardless of equation (3-5).

If pel i is omitted due to field repetition, its corresponding R_2 (i) is set to 0 regardless of equation (3-5). When pel i is not forced-intraframe coded, R_2 (i) in burst B is set to 0.

3.6.2.3 Background generation

The background prediction value is generated scene adaptively as

$$b_i = b_i^{-f} + v(k)\ \mathrm{sgn}\left(Y_i - b_i^{-f}\right) u\left(Y_i - Y_i^{-f}\right) \tag{3-7}$$

where

$$u\left(Y_i - Y_i^{-f}\right) = \begin{cases} 1 \text{ if } \left| Y_i - Y_i^{-f} \right| \leq L, \\ \\ 0 \text{ if otherwise.} \end{cases} \tag{3-8}$$

$$v(k) = \begin{cases} 1, \text{ for one frame period in every block of } k \text{ frames} \\ 0, \text{ for consecutive } (k-1) \text{ frames following the frame of } v(k) = 1 \end{cases}$$

and

b_i is the background prediction value for the present frame;

b_i^{-f} is the background prediction value for the previous frame;

Y_i is the interpolator output for the present frame;

Y_i^{-f} is the interpolator output for the previous frame;

u is the still area detection function;

k is the background updata control parameter; and

L is the threshold value.

Parameters k and L are set as $k = 8$ and $L = 1$. It is noted that for hardware simplification, b_i^{-f}, instead of b_i, is used as background prediction value B_i (see Figure 11).

3.6.2.4 Forced intraframe prediction

This codec usually used the demand refresh mode to prevent the defected picture due to transmission errors from being left in the decoder frame memory. The demand refresh mode is carried out if BWP (bit 3.15.4 in codec-to-codec information) = 0, which indicates that backward path from decoder to coder is available. However, the cyclic refresh mode is also provided, considering such applications as broadcasting communication where no backward path (from decoder to coder) is available. This mode is carried out when BWP = 1.

For either of the two refresh modes, the prediction function is forcedly set to intraframe prediction.

In the demand refresh mode, the motion frame memory and the background frame memory are updated block-line by block-line within a frame time by writing the interpolator output simultaneously. Once demand refresh starts by receiving DRR in the coder, the following DRR (demand refresh request command) received is disregarded for one second (see Note).

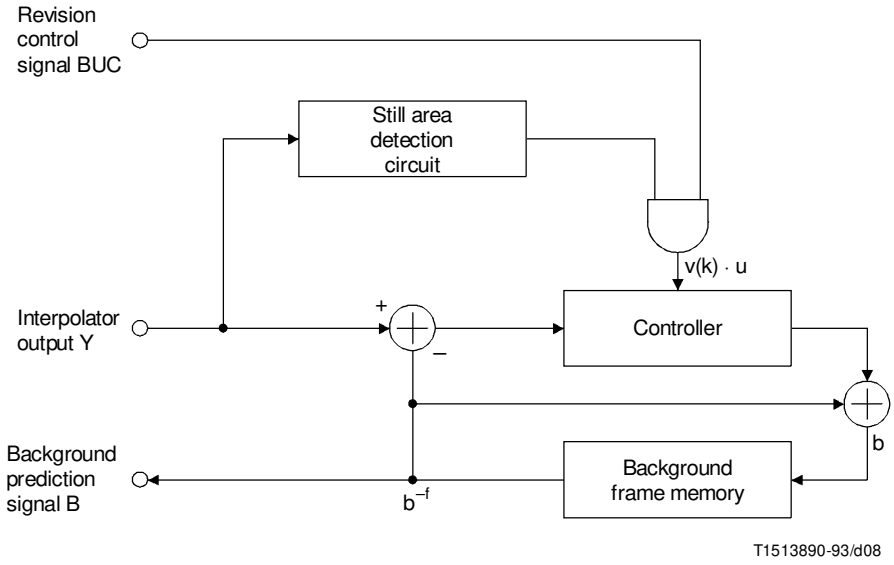


FIGURE 11/H.120

Background generation

In the cyclic refresh mode, the two memories are simultaneously updated two lines at a time by writing the interpolator output. When a field is omitted due to field repetition, the background frame memory is updated by the signal which updates the motion frame memory. It should be noted that the command for the cyclic refresh mode is ignored in block-lines where updating based on the demand refresh mode is carried out.

NOTE – If a transmission error happens on the line from the codec A to the codec B, the decoder of the codec B detects the error occurrence and generates a demand refresh request information (DR). This DR is passed to the coder of the codec B and transmitted as a demand refresh request command (DRR) to the codec A. When the decoder of the codec A receives DRR, a demand refresh confirmation information DDR is passed to the coder of the codec A. Finally, demand refresh mode is operated along with transmission of a demand refresh mode command (DRM) from the codec A to the codec B.

3.6.2.5 Definition of blanking and block-line and edge pels treatment

3.6.2.5.1 The kinds of pels arranged on a horizontal scanning line (see Figure 8), for which prediction functions are defined, are as follows:

Burst B 7 pels;

Colour C 64 pels;

Luminance Y 384 pels.

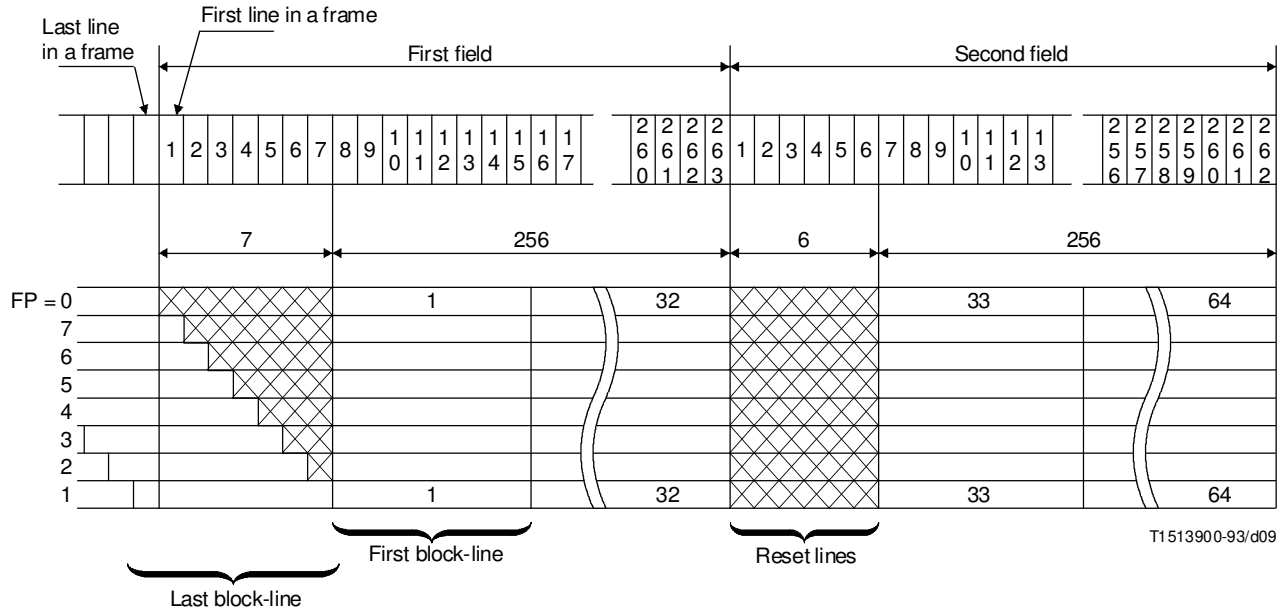
The vertical blanking periods are treated in the same way as active lines.

3.6.2.5.2 Block-line

See Figure 12.

In the first field, 8 lines consisting of the 8th to the 15th lines form the first block-line, with every 8 lines following this forming each a block-line. In the second field, 8 lines consisting of the 7th to the 14th lines form the 33rd block-line. Each field has 32 block-lines.

The last block-line in a frame is defined as the 8 lines which include the last line of the frame, or the line closest to the head line of the frame. The position of the video last line in the last block-line is coded as frame position.



FP Frame position
X Reset line

NOTE – When the last line in a frame falls on the *n*th line of the last block-line, FP = mod(*n*,8).

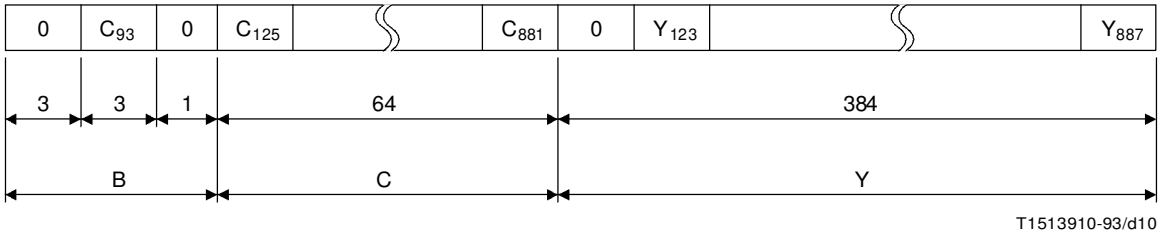
FIGURE 12/H.120

Definition of block-line and reset line

3.6.2.5.3 Reset lines

The lines which are excluded from block-lines are defined as reset lines. These reset lines are clampled to 0 in the predictive coding and decoding loops, or the corresponding prediction values X_2 and prediction errors e in Figure 10 are set to 0. The reset lines are prediction-coded in normal mode with adaptive prediction and by setting $v = 0$.

3.6.2.5.4 Edge pels suffer from crosstalk due to interpolation between B and C, C and Y, and Y and B. In order to prevent such crosstalk, the first 3 pels in B, the last pel in B and the first pel in Y are clampled to 0 at the source coder input IDM signal as in Figure 13.



NOTE – This figure shows an odd line in the first field. See Figure 8 as for pel numbers.

FIGURE 13/H.120

Zero insertion for preventing crosstalks due to interpolation

3.6.2.5.5 Edge pels are not treated specifically in the source coder and decoder. That is, video signals including reset lines and the 3 clamped pels in B are processed as if they were existing continuously (see Note). Consequently, even if a motion vector points to pels outside the active picture area, it functions as a delay control for the input time serial video signal.

NOTE – The right end of each line in the picture is assumed to be connected to the left end of the next line, and the lower end of each frame is assumed to be connected to the upper end of the next frame.

In the forced intraframe prediction mode, the prediction value for the first pel of each line is set to 0.

For the burst signal, no adaptive prediction nor subsampling is applied and no motion vectors are transmitted.

3.6.2.6 Prediction and interpolation functions

Prediction functions and interpolation functions are shown in Table 4 for all of the coding modes. It should be noted that motion vectors for the colour signal can be set to 0 without much loss of coding efficiency.

3.6.2.7 Quantization

Prediction errors for video signals are quantized using one of the four quantizing characteristics indicated in Table 5, that is, Q_0 (57 levels), Q_1 (57 levels), Q_2 (51 levels) and Q_3 (37 levels). The same set of quantizing characteristics are applied regardless of prediction functions.

3.6.2.8 Limiter in prediction loop

No limiter is allocated in the prediction loop. Accordingly, the input signal x for the prediction loop is limited to $-124 \leq x \leq 123$ so that the local decoder output X is maintained in the $-128 \leq X \leq 127$ range.

3.6.2.9 Frame memory parity check

Parity is counted for each bit plane of the interpolator output during a video frame period from the 1st to the 64th block-line as defined in Figure 12. If block-lines are omitted in field repetition mode, parity is not counted during these omitted block-lines.

Eight odd parity bits are sent to the decoder, where they are compared with the parity bits of the decoder interpolator output to detect uncorrected errors. If any difference between received and counted parity bits is found, a demand refresh is requested from the decoder to the coder.

3.6.2.10 Suspension of coding operation

When the information is generated to the degree that the transmission buffer memory overflows, coding operation is suspended by setting $e = 0$ and $v = 0$. This stop mode is defined only in the coder. Interpolation and prediction functions for this mode are defined as either of NRM, SBS, FRP or RFS modes according to the control of the coding parameter controller.

3.6.3 Motion vector transmission

3.6.3.1 Block size

A block for motion compensation consists of 8 lines (vertical) by 16 pels (horizontal).

3.6.3.2 Maximum tracking range

Motion vectors are tracked in the range of +7 to –7 lines (vertical) and +15 pels (horizontal) at its maximum. The decoder should be able to reproduce any vector in this maximum range.

TABLE 4/H.120

Prediction and interpolation functions

Coding mode		Kind of pel	Prediction functions P(Z) (Note 1)			Interpolation functions I(Z) (Note 2)		
			P_Y (Z)	P_C (Z)	P_B (Z)	I_Y (Z)	I_C (Z)	I_B (Z)
Normal		Coded	Z^{-1}; $S_2 = 1$ Z^{-F+V}; $S_2 = 0$, $S_1 = 0$ P_b (Z) (Note 3); $S_2 = 0$, $S_1 = 1$		Z^{-F}	1		
Subsampling		Coded	Z^{-2} (Note 4); $S_2 = 0$ Z^{-F+V}; $S_2 = 0$, $S_1 = 0$ P_b (Z); $S_2 = 0$, $S_1 = 1$			1		
		Omitted	(Not defined)			$\frac{1}{2}\left\{\frac{1}{2}\left(Z^{-1} + Z^{+1}\right) + \frac{1}{2}\left(Z^{-H} + Z^{+H}\right)\right\}$	$\frac{1}{2}\left(Z^{-1} + Z^{+1}\right)$	
Field repetition		Omitted	(Not defined)			$\frac{1}{2}\left(Z^{-262H} + Z^{-263H}\right)$	Z^{-263H}; first field Z^{-262H}; second field	
Refresh	NRM	Coded	Z^{-1}			1		
	SBS	Coded	Z^{-2} (Note 4)			1		
		Omitted	(Not defined)			$\frac{1}{2}\left(Z^{-1} + Z^{+1}\right)$		

NOTES

1 S_1 and S_2 are prediction status signals defined in 3.6.2.2.

2 To deal with fractions generated by operation of (A + B)/2, (A + B + 1)/2 is executed and the 8 MSBS are used.

3 Background is generated as described in 3.6.2.3.

4 Z^{-1}, if the previous pel is coded.

TABLE 5/H.120

Quantizing characteristics

Q_0		Q_1		Q_2		Q_3	
Input range	Output level	Input range	Output level	Input range	Output level	Input range	Output level
0 to 1	0	0 to 3	0	0 to 4	0	0 to 6	0
2	1	4 to 6	3	5 to 8	5	7 to 11	7
3	2	7 to 8	6	9 to 12	10	12 to 17	14
4 to 5	3	9 to 10	9	13 to 17	15	18 to 24	21
6 to 7	5	11 to 13	12	18 to 22	20	25 to 31	28
8 to 9	7	14 to 16	15	23 to 27	25	32 to 38	35
10 to 11	10	17 to 19	18	28 to 32	30	39 to 45	42
12 to 14	13	20 to 22	21	33 to 37	35	46 to 52	49
15 to 17	16	23 to 26	24	38 to 42	40	53 to 59	56
18 to 20	19	27 to 30	28	43 to 47	45	60 to 66	63
21 to 23	22	31 to 34	32	48 to 52	50	67 to 73	70
24 to 26	25	35 to 39	37	53 to 57	55	74 to 80	77
27 to 29	28	40 to 44	42	58 to 62	60	81 to 87	84
30 to 32	31	45 to 49	47	63 to 67	65	88 to 94	91
33 to 37	35	50 to 54	52	68 to 72	70	95 to 101	98
38 to 42	40	55 to 59	57	73 to 77	75	102 to 108	105
43 to 48	45	60 to 64	62	78 to 82	80	109 to 115	112
49 to 54	51	65 to 69	67	83 to 87	85	116 to 123	119
55 to 60	57	70 to 74	72	88 to 92	90	124 to 255	127
61 to 67	64	75 to 79	77	93 to 97	95		
68 to 74	71	80 to 84	82	98 to 102	100		
75 to 81	78	85 to 89	87	103 to 107	105		
82 to 88	85	90 to 94	92	108 to 112	110		
89 to 95	92	95 to 99	97	113 to 118	115		
96 to 102	99	100 to 104	102	119 to 124	121		
103 to 109	106	105 to 109	107	125 to 255	127		
110 to 116	113	110 to 116	113				
117 to 123	120	117 to 123	120				
124 to 255	127	124 to 255	127				

NOTE – Characteristics are symmetrical with respect to zero.

3.6.3.3 Definition of vector direction

The motion vector v (v_x, v_y) is defined as

$$v_x = x_a - x_b$$

and (3-9)

$$v_y = y_a - y_b$$

where the block positions in the present frame and in the corresponding previous frame are (x_a, y_a) and (x_b, y_b), respectively. The x and y directions are identical to those of the horizontal and vertical scanning. This definition means that delay in the interframe prediction loop increases for v_x, v_y greater than 0.

3.6.3.4 Motion detection method

A motion vector is detected for each block by the interframe block matching method. Detailed detection methods are left to each hardware implementation (see Note).

NOTE – When multi-field repetition is employed, the detected vector for the previous transmitted frame can be utilized as the initial value for vector detection in the present frame to be omitted, and the detected vector for the present frame can be utilized as the initial value for vector detection in the next frame, and so on.

3.6.4 Coding parameter control

3.6.4.1 Control method

Coding control is carried out by selecting quantizing characteristics described in 3.6.2.7 and coding modes described in 3.6.2.1.

3.6.4.2 Control timing

Coding parameters are controlled according to the timing and commands as shown in Table 6.

TABLE 6/H.120

Coding parameter, control unit and commands

Coding parameter	Control unit	Commands
Normal	Frame Block-line (8 lines) Block (8 × 16 pels)	SBC = 1, IFM = 1, FRP = 1 and TRANS (SBS: off)
Quantization	Block-line	QC1 and QC2
Field repetition	Block-line (Note)	FRP = 0
Subsampling	Frame	SBC = 0 and FRP = 1
	Block	TRANS (SBS: on) and FRP = 1
Stop	Arbitrary	Prediction error e = 0 Motion vector v = 0
Demand refresh	Block-lines	DRM = 0 and IFM = 0
Cyclic refresh	Two lines	DRM = 1, IFM = 0 and CRM 1, 2
NOTE – Consecutive 32 block-lines from the first through the 32nd block-lines or from the 33rd through the 64th block-lines are omitted for ordinary field repetition. Other methods are also possible using the FRP command controlled in the unit of block-line.		

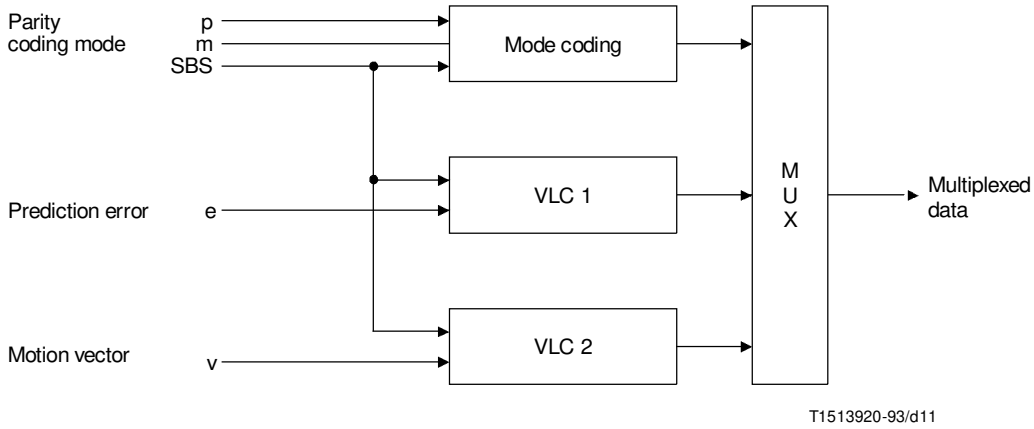
3.6.4.3 Control sequence

The control sequence is determined based on buffer memory occupancy and other control information. Since this sequence does not affect interworking between different design codecs, it is left to each hardware implementation. However, the codec operating principle is that the coder determines all operating modes, which are transmitted with the coded video data to the decoder as a combination of commands. The decoder reproduces the video signal according to the received commands and data. A control sequence example is shown in Annex F.

3.6.5 Entropy coding

3.6.5.1 Configuration of entropy coding

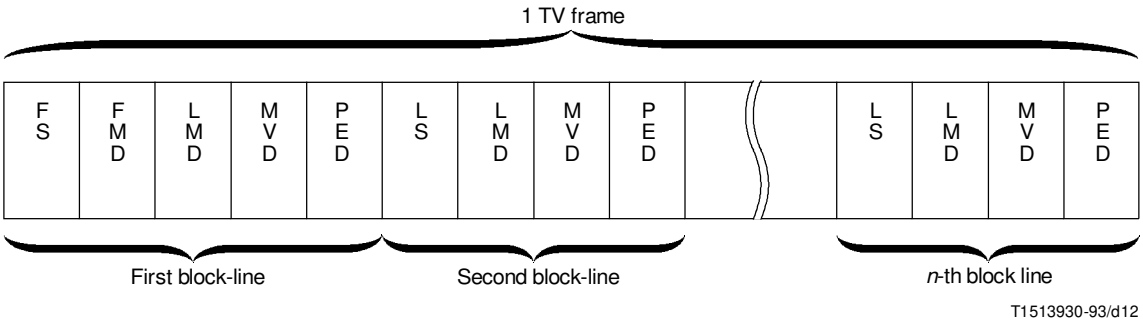
The configuration of entropy coding is shown in Figure 14. The entropy coder compresses the data of the prediction error *e* and motion vector *v*, which are provided by the source coder, using variable length coding. These compressed data are multiplexed with coding mode data *m* and fed into the transmission buffer memory. The multiplexed data format is outlined in Figure 15.



SBS Subsampling
VLC Variable word length coder

FIGURE 14/H.120

Configuration of entropy coding



FS Frame sync
FMD Frame mode data
LS Line sync
LMD Line mode data
MVD Motion vector data
PED Prediction error data

FIGURE 15/H.120

Multiplexed data format

3.6.5.2 Commands for coding modes and data structure

The commands for coding modes and data structure are defined as follows:

3.6.5.2.1 Frame sync (FS)

The frame sync is a unique word to designate the start of a video frame and its value is: 0000000000000010.

3.6.5.2.2 Frame mode data (FMD)

The format of the frame mode data is given in Figure 16.

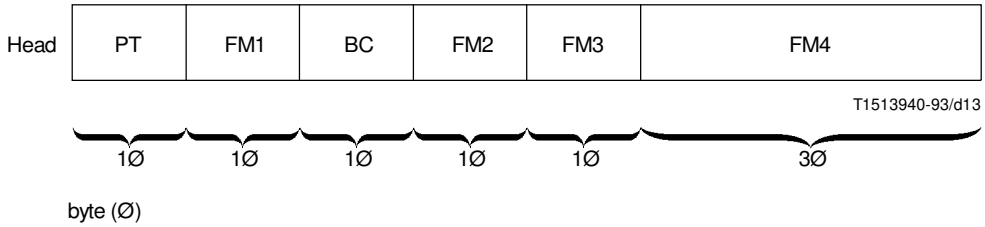


FIGURE 16/H.120

a) *Parity data (PT)*

Odd parity for each of the 8-bit planes of the interpolator output during the previous frame period (MSB first).

b) *Frame mode 1 (FM1)*

The format of the frame mode 1 is given in Figure 17.

SBC	BRC	BUC	DRM	1	FP2	FP1	FP0

FIGURE 17/H.120

i) *Subsample control (SBC)*

When SBC = 0, subsampling is carried out throughout the frame excluding burst signals, reset lines and block-lines with FRP = 0. See 3.6.2.1.

ii) *Background revision control (BRC)*

When BRC = 0, the contents of the motion frame memory are transferred to the background frame memory during this frame period. See 3.6.2.4.

iii) *Background update control (BUC)*

When BUC = 0, the background frame memory is updated. If BRC is operating, it has priority. See 3.6.2.3.

iv) *Demand refresh mode (DRM)*

When DRM = 0, coding is carried out with demand refresh mode. See 3.6.2.4.

v) *Frame position, FP2-FP0* (see Note)

This 3-bit word designates the position of the head line of the video frame or the first line in the first field (MSB first). See Figure 12.

NOTE – FP bits are employed for preventing degradation in the case where the input signals are asynchronously switched to another signal where those signals have different sync phase or different sync frequency. For this purpose, the horizontal sync pulse interval in the codec, namely the picture element numbers per line should be kept as 455 samples, even in the transition period. Furthermore, input signal switching which takes place during the reset line periods should be ignored.

c) *Buffer control (BC)*

The staying time of FS in the transmission buffer memory is coded into an 8 bit word (MSB first). See 3.6.6.1.

d) *Frame mode 2 (FM2)*

The format of frame mode 2 is given in Figure 18.

1	DRR	CMS	CRM1	CRM2	SF1	MAF	1

FIGURE 18/H.120

i) *Demand refresh request (DRR)*

When DRR = 0, the decoder requests a demand refresh to the coder. See 3.6.2.9.

ii) *Colour/monochrone state (CMS)*

Colour (bit = 1) / monochrome (bit = 0) where monochrome is optional and the default mode is colour.

iii) *CRM1, CRM2: cyclic refresh mode*

This 2-bit word designates the position of the two lines in a block-line which is cyclic refreshed. See Figure 19. See also 3.6.2.4.

iv) *Spare frame mode (SF1)*

v) *Mode addition flag (MAF)*

When MAF = 0, FM4 is added.

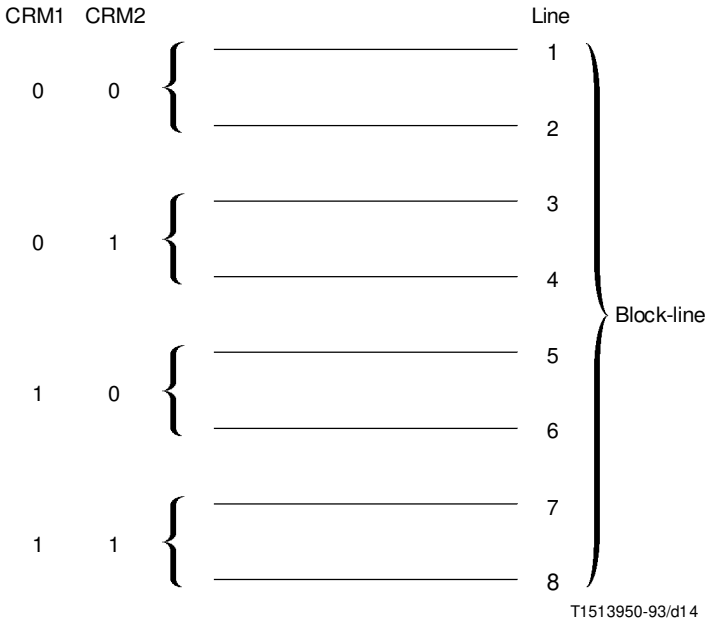


FIGURE 19/H.120

e) *Frame mode 3 (FM3)*

National option 8-bit data. If not used, a code consisting of all ones (11111111) is inserted.

f) *Frame mode 4 (FM4)*

Byte 1	SF2	SF3	SF4	1	SF5	SF6	SF7	SF8
Byte 2	SF9	SF10	SF11	1	SF12	SF13	SF14	SF15
Byte 3	SF16	SF17	SF18	1	SF19	SF20	SF21	SF22

SF2-SF22 Spare frame mode.

3.6.5.2.3 Line sync (LS)

The line sync is a unique word to designate the start of a block-line and its value is: 0000000000000011.

3.6.5.2.4 Line mode data (LMD)

The format of the line mode data is given in Figure 20.

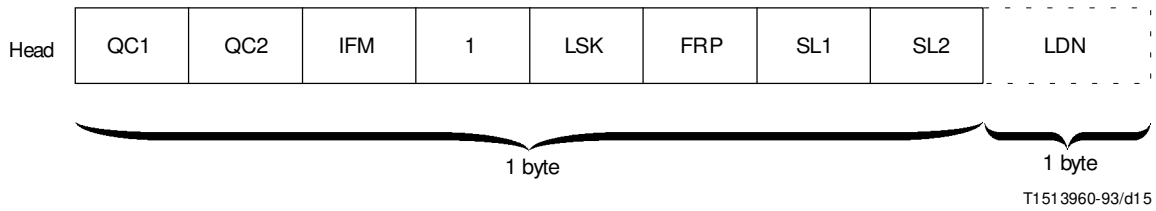


FIGURE 20/H.120

a) QC1, QC2: *Quantizing characteristics*

QC1	QC2	Characteristics (Table 5)
0	0	Q_0
0	1	Q_1
1	0	Q_2
1	1	Q_3

b) *Forced intraframe prediction mode (IFM)*

When IFM = 0, the prediction function is fixed to intraframe prediction, throughout this block-line if DRM = 0, and at the two lines designated by CRM1 and CRM2 if DRM = 1. See 3.6.2.4.

c) *Line skip (LSK)*

When LSK = 0, the following byte (LDN, line data number) designates the number of block-lines which are skipped. See 3.6.5.5. LDN is coded similarly as the number of vector data, VDN. When LDN = n, consecutive (n + 1) block-lines are the same. Therefore $0 \leq n \leq 63$.

d) *Field repetition (FRP)*

When FRP = 0, this block-line is omitted because of field repetition. This is valid even if IFM = 0. See 3.6.2.1.

e) *SL1, SL2: Spare line mode*

3.6.5.2.5 Motion vector data (MVD)

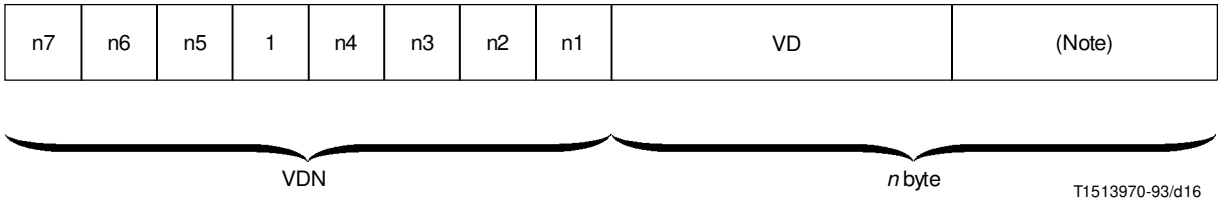
The format of the motion vector data is given in Figure 21.

a) *Vector data number (VDN)*

Designates the byte number of the following vector data (VD) (in natural binary code, MSB first).

b) *Vector data (VD)*

Variable length coded motion vector data.



NOTE – Dummy, see 3.6.5.4.6.

FIGURE 21/H.120

3.6.5.2.6 Prediction error data (PED) (variable word length coded)

The format of the prediction error data is given in Figure 22.

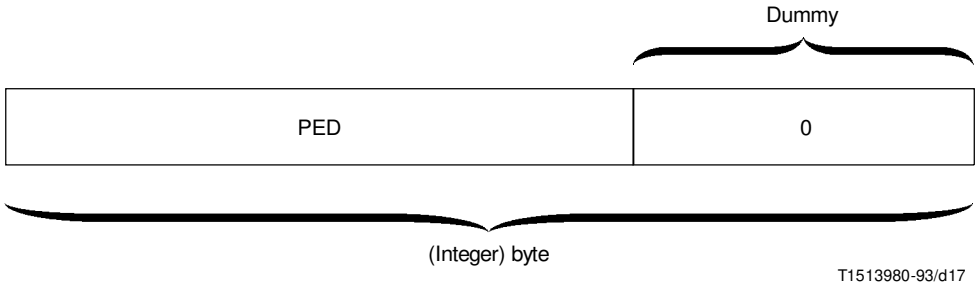


FIGURE 22/H.120

3.6.5.3 Prediction error coding (VLC 1)

See Figure 14.

3.6.5.3.1 Coding method

The quantizing level number corresponding to prediction error e is coded based on its statistical characteristics. For $e \neq 0$, variable word length coding is carried out using code V or F stating the quantizing level number (see Table 7). For $e = 0$, run length code R is used to state a run (RL) of ineffective pels. Note that if RL = 1, a variable word length code V_0 or F_0 to state $e = 0$ is used (see Table 8).

TABLE 7/H.120

Variable length code for non-zero amplitude prediction errors

Level number	Code length	V code		Level number	Code length	F code
V_0	4	0 1 1 1		F_0	4	0 0 0 1
1	2	1 S		1	6	1 1 1 1 1 S
2	5	0 1 1 0	S	2	6	1 1 1 1 0 S
3	7	0 1 0 1	1 1 S	3	6	1 1 1 0 1 S
4	7	0 1 0 1	1 0 S	4	6	1 1 1 0 0 S
5	8	0 1 0 1	0 1 1 S	5	6	1 1 0 1 1 S
6	8	0 1 0 1	0 1 0 S	6	6	1 1 0 1 0 S
7	8	0 1 0 1	0 0 1 S	7	6	1 1 0 0 1 S
8	8	0 1 0 1	0 0 0 S	8	6	1 1 0 0 0 S
9	9	0 1 0 0	1 1 1 1 S	9	6	1 0 1 1 1 S
10	9	0 1 0 0	1 1 1 0 S	10	6	1 0 1 1 0 S
11	9	0 1 0 0	1 1 0 1 S	11	6	1 0 1 0 1 S
12	9	0 1 0 0	1 1 0 0 S	12	6	1 0 1 0 0 S
13	9	0 1 0 0	1 0 1 1 S	13	6	1 0 0 1 1 S
14	9	0 1 0 0	1 0 1 0 S	14	6	1 0 0 1 0 S
15	9	0 1 0 0	1 0 0 1 S	15	6	1 0 0 0 1 S
16	9	0 1 0 0	1 0 0 0 S	16	6	1 0 0 0 0 S
17	10	0 1 0 0	0 1 1 1 1 S	17	6	0 1 1 1 1 S
18	10	0 1 0 0	0 1 1 1 0 S	18	6	0 1 1 1 0 S
19	10	0 1 0 0	0 1 1 0 1 S	19	6	0 1 1 0 1 S
20	10	0 1 0 0	0 1 1 0 0 S	20	6	0 1 1 0 0 S
21	10	0 1 0 0	0 1 0 1 1 S	21	6	0 1 0 1 1 S
22	10	0 1 0 0	0 1 1 1 0 S	22	6	0 1 0 1 0 S
23	10	0 1 0 0	0 1 0 0 1 S	23	6	0 1 0 0 1 S
24	10	0 1 0 0	0 1 0 0 0 S	24	6	0 1 0 0 0 S
25	10	0 1 0 0	0 0 1 1 1 S	25	6	0 0 1 1 1 S
26	10	0 1 0 0	0 0 1 1 0 S	26	6	0 0 1 1 0 S
27	10	0 1 0 0	0 0 1 0 1 S	27	6	0 0 1 0 1 S
28	10	0 1 0 0	0 0 1 0 0 S	28	6	0 0 1 0 0 S

NOTE – S denotes sign. S = 0 for positive, S = 1 for negative.

TABLE 8/H.120

Run length code R for zero amplitude prediction errors

RL (Note 1)	Code length	Code word R				Remark
2	5	0 0	0 0 1			
3	5	0 0	0 0 0			
4	6	0 0	1 0 1 0			
5	6	0 0	1 0 0 1			
6	6	0 0	1 0 0 0			
7	7	0 0	1 0 1 1 1			
8 to 11	7	0 0	1 1 0 X X			X = 11 – RL
12	8	0 0	1 1 1 1 0 1			
13	8	0 0	1 1 1 1 0 0			
14 to 17	8	0 0	1 1 1 0 X X			X = 17 – RL
18 to 25	9	0 0	0 1 1 1 X X	X		X = 25 – RL
26 to 33	10	0 0	0 1 1 0 0 X	X X		X = 33 – RL
34 to 37	10	0 0	0 1 0 1 0 0	X X		X = 37 – RL
38 to 64	12	0 0	0 1 0 0 1 X	X X X X		X = 69 – RL
MK1	13	0 0	1 0 1 1 0 Y	Y Y Y Y Y		
MK2	14	0 0	1 1 1 1 1 1	Y Y Y Y Y Y		Y = 0 to 63
MK3	14	0 0	1 1 1 1 1 0	Y Y Y Y Y Y		
MK4 to 7	15	0 0	0 1 1 0 1 X	X Y Y Y Y Y Y		X = 7 – MK
MK8 to 15	16	0 0	0 1 0 1 1 X	X X Y Y Y Y Y Y		X = 15 – MK
MK16 to 19	16	0 0	0 1 0 1 0 1	X X Y Y Y Y Y Y		X = 19 – MK
MK20 to 34	18	0 0	0 1 0 0 0 1	X X X X X Y Y Y	Y Y	X = 35 – MK
MK35 to 49	19	0 0	0 1 0 0 0 0	1 X X X X Y Y Y	Y Y Y	X = 50 – MK
MK50 to 56	19	0 0	0 1 0 0 0 0	0 1 X X X Y Y Y (Note 2)	Y Y Y	X = 57 – MK

NOTES

1 RL = 64 × (MK number) + 1 + Y, 0 ≤ Y ≤ 63.

2 The maximum run length is (455 – 3) × 8 = 3616. Corresponding MK and Y turn out to be 56 and 31, respectively. Hence, 0 ≤ Y ≤ 31 for MK = 56.

3.6.5.3.2 Scanning sequence

Entropy coding for a video frame is carried out from the first to the last block lines, excluding reset lines. The frame sync (FS) and frame mode data (FMD) are codes in the first block-line. When the last line falls on the *n*th line of the last block-line, the frame position is set to FP = md (*n*, 8). FP is transmitted to the decoder as a part of the frame mode data (see Note 1).

Since the first three pels of each line are clamped to 0 in the predictive source coder, and the reset lines are so defined as described in 3.6.2.5.3, the pels to be entropy coded can be indicated as in Figure 23 (see Note 2).

The scanning sequence is a block scan as shown in Figure 24. The first block after scan conversion consists of 4 pels × 8 lines = 32 pels.

NOTE

1 Without asynchronous switching of input video signals, the last block-line coincides with the 64th block-line and FP = 0.

2 Entropy coding is not carried out for the reset lines defined in Figure 12. The number of reset lines in the first field varies according to the FP value.

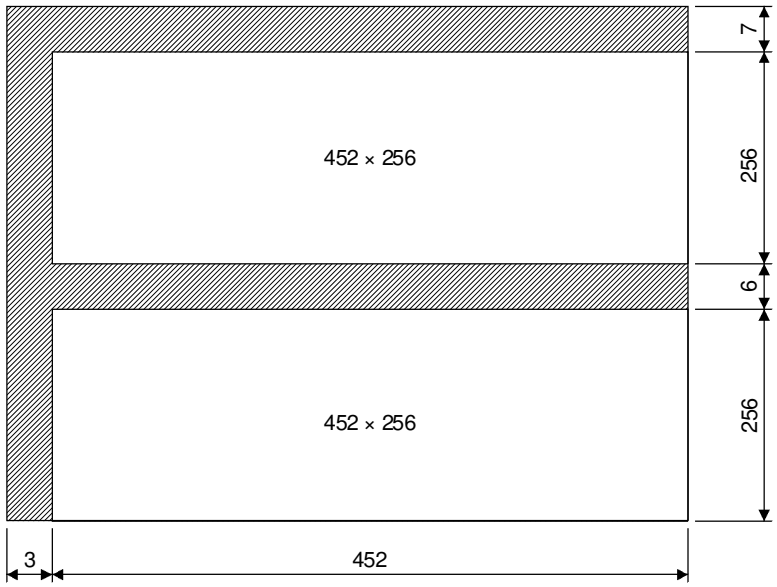


FIGURE 23/H.120

Entropy-coded pels

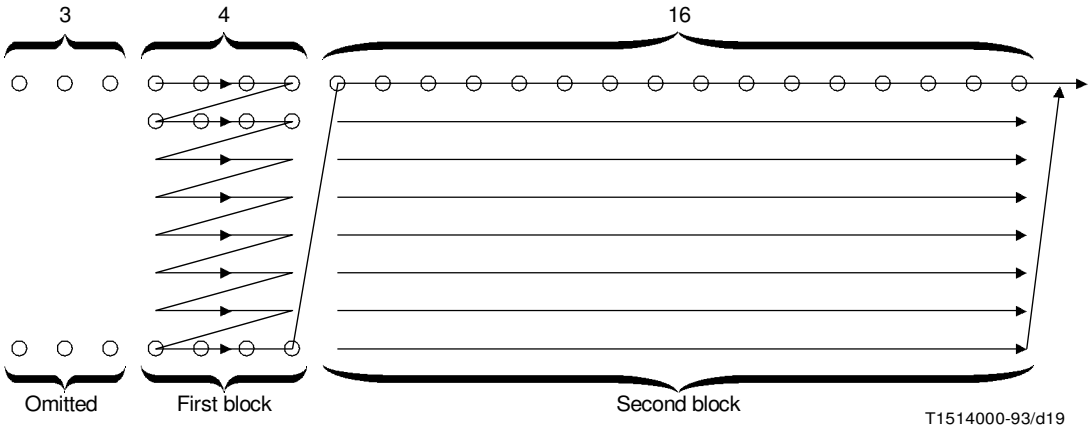


FIGURE 24/H.120

Scanning sequence

3.6.5.3.3 Group of codes

See Table 9.

TABLE 9/H.120

	Symbol	Number of codes	Length of codes
Amplitude code No. 1	F	57	4, 6
Amplitude code No. 2	V	57	2-10
Run length code	R	3615	5-19

F Pseudo fixed length code for stating quantizing level number. This code is introduced to shorten the maximum code length.

V Variable length code for stating quantizing number.

R Variable length code for stating run length of ineffective pels for RL ≥ 2.

3.6.5.3.4 Code transition rule

The rule is shown in Figure 25 and a prediction error coding example is given in Annex F.

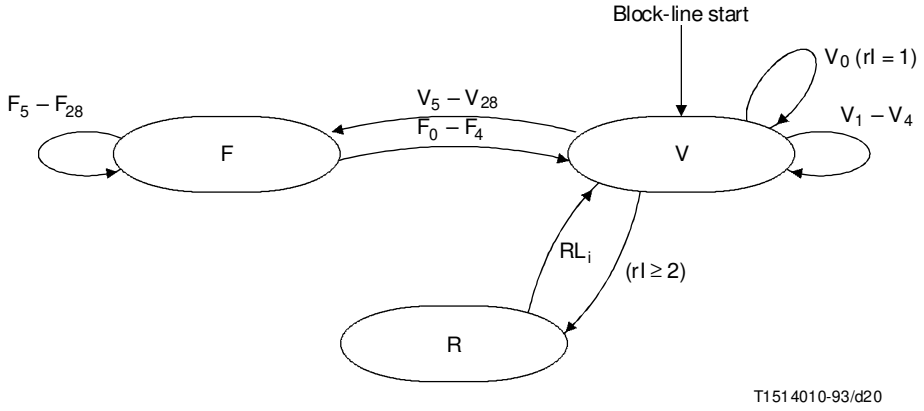
The following points should be noted:

a) the starting code is either the V or R code;

b) the last run in a block-line may not be transmitted since the LS or FS command can be utilized as a termination of the last run;

c) coding is carried out assuming that omitted pels due to subsampling do not exist;

d) some 0s are filled at the tail tail-end of the PED as dummies to make the total number of bits for the block-line data a multiple of 8.



NOTES

1 RL is a length of the run to be coded, while rl is a number of continued pels whose e = 0.

2 Prediction error data start with an R or V code. R code is used if rl ≥ 2. Otherwise V code is used.

3 The code can shift to V even if RL ≥ 2 to prevent a buffer memory underflow.

FIGURE 25/H.120

Code transition rule for prediction error data

3.6.5.3.5 Code assignments for F and V

See Table 7.

Code assignments are common to the four quantizing characteristics Q_0, Q_1, Q_2 and Q_3.

3.6.5.3.6 Code assignments for R

See Table 8.

3.6.5.4 Motion vector coding (VLC 2)

3.6.5.4.1 Coding method

A motion vector v, is first coded with predictive coding whose output Δv, is variable length coded throughout a block-line.

3.6.5.4.2 Predictive coding

The prediction algorithm is the previous block prediction which is

$$\Delta v = v - v_1 \tag{3-10}$$

where v and v_1 represent the present and the previous block vectors. The operation is carried out for each x and y component in two's complement form. The operated results are expressed with 5 bits for the x component and 4 bits for the y component neglecting carries (MSB first). Note that the decoder carries out the inverse operation $v = v_1 + \Delta v$ in two's complement from neglecting carries.

The motion vector for the first block (horizontal blanking) is set to (0,0).

3.6.5.4.3 Variable length coding

For $\Delta v = (0,0)$, a run length of zero is coded. For $\Delta v \neq (0,0)$, variable length coding is applied with their code lengths shown in Figure 26.

The coding of Δv is carried out for the 28 vectors of the 2nd to the 29th blocks.

The last run of $\Delta v = (0,0)$ may not be transmitted since VDN states the total bits of VD.

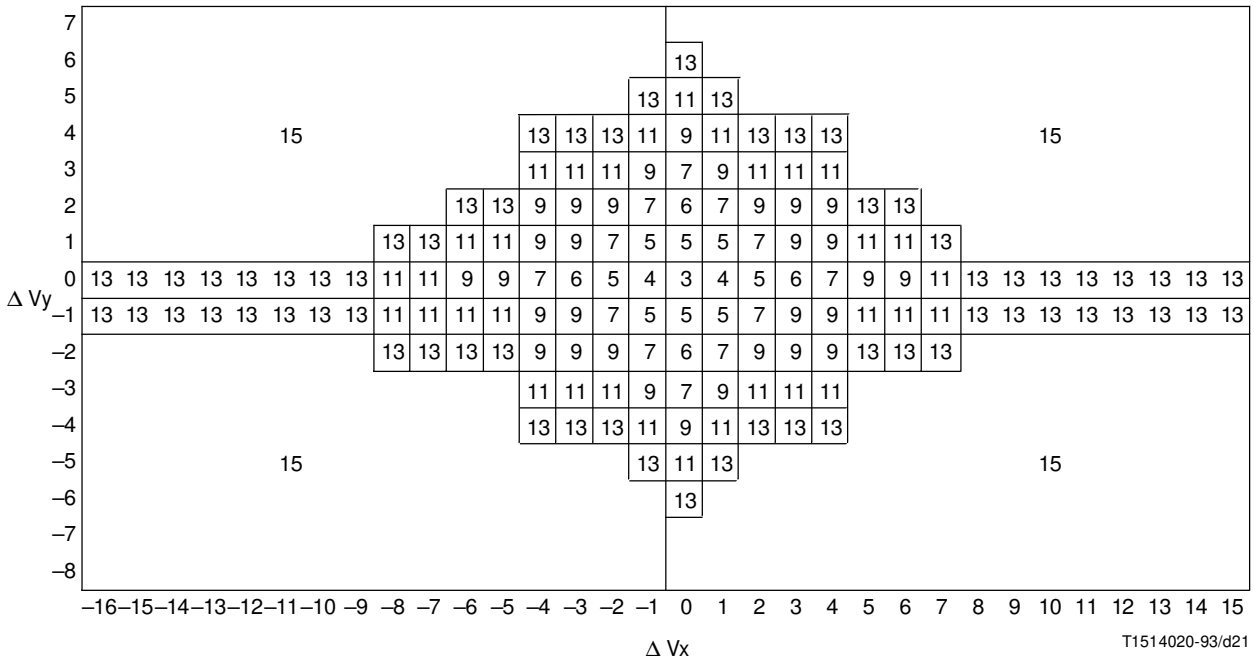


FIGURE 26/H.120

Word length for motion vector prediction errors

3.6.5.4.4 Code assignments

The codes are assigned as shown in Table 10, where the maximum code length is 15. The variable length codes consist of 541 codes, or 512 codes for Δv, 28 codes for run length and one TRANS code for transition of subsampling ON/OFF.

TABLE 10/H.120

Variable length code and run length code for motion vector data

ΔVx	ΔVy	Code length	Code word	Number of codes
± 1	0	4	0 0 1 Sx	2
± 1	± 1	5	1 1 1 SxSy	8
0	± 1	5	1 1 0 1 Sy	
± 2	0	5	1 1 0 0 Sx	
0	± 2	6	1 0 1 1 1 Sy	4
± 3	0	6	1 0 1 1 0 Sx	
± 1	± 2	7	1 0 0 1 1 SxSy	12
± 2	± 1	7	1 0 0 1 0 SxSy	
0	± 3	7	1 0 0 0 1 1 Sy	
± 4	0	7	1 0 0 0 1 0 Sx	
± 3	± 1	9	1 0 1 0 1 1 1 SxSy	30
± 1	± 3	9	1 0 1 0 1 1 0 SxSy	
± 2	± 2	9	1 0 1 0 1 0 1 SxSy	
± 3	± 2	9	1 0 1 0 1 0 0 SxSy	
± 4	± 1	9	1 0 1 0 0 1 1 SxSy	
± 4	± 2	9	1 0 1 0 0 1 0 SxSy	
± 5	0	9	1 0 1 0 0 0 1 1 Sx	
± 6	0	9	1 0 1 0 0 0 1 0 Sx	
0	± 4	9	1 0 1 0 0 0 0 1 Sy	
–8 to 7	–5 to +5 (see Figure 26)	11	1 0 0 0 0 1 X X X X Sy	[X] = ΔVx	32
–16 to 15	–6 to +6 (see Figure 26)	13	0 1 0 0 0 0 1 X X X X X Sy	[X] = ΔVx	64
–16 to 15	–8 to +7 (see Figure 26)	15	1 0 0 0 0 0 X X X X X Y Y Y Y	[X] = ΔVx [Y] = ΔVy	359

RL	Code length	Code word	Number of codes
1	3	0 0 0	1
2	4	0 1 1 1	1
3 to 6	6	0 1 1 0 X X X X = 6 – RL	4
7 to 12	7	0 1 0 1 X X X X X = 12 – RL	6
13 to 20	8	0 1 0 0 1 X X X X X = 20 – RL	8
21 to 28	9	0 1 0 0 0 1 X X X X X = 28 – RL	8
TRANS	6	0 1 0 1 1 1	1

NOTES

1 Sx and Sy denote signs: $S_i = 0$ for positive, $S_i = 1$ for negative.

2 XX..X and YY..Y are expressed in two's complement form (MSB first).

3.6.5.4.5 Transition code for subsampling (TRANS)

The code TRANS indicates transition between ON and OFF for subsampling (SBS). For the first block in a block-line, SBS is set to OFF. Subsampling is then set to ON at the block just after the first TRANS code is inserted, and returned to OFF at the block just after the second TRANS code is inserted. The same sequence follows on. The TRANS code is expressed as a 6-bit word. When SBS = 0, transition code is disregarded in the decoder.

3.6.5.4.6 Dummy code insertion

When a vector data (VD) for a block-line does not have exactly 8 multiple bits, a dummy code consisting of 1 to 7 bits is inserted at the tail of the vector data.

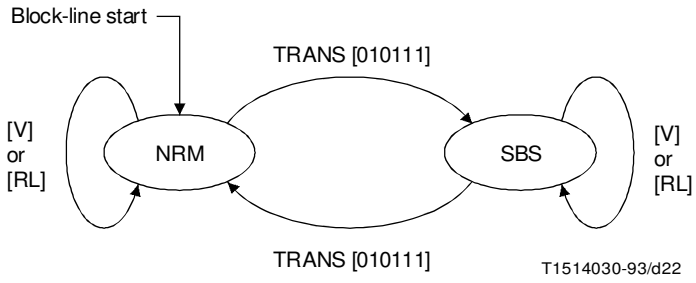
The dummy code has 1 as the head, 0s as the body and 1 as the tail (see Table 11).

TABLE 11/H.120

Number of dummy bits	Dummy code
1	1
2	11
3	101
4	1001
5	10001
6	100001
7	1000001

3.6.5.4.7 Code transition rule

The rule is indicated in Figure 27 with a motion vector coding example given in Annex F.



NOTE – The same motion vector coding method is applied for both NRM and SBS modes.

FIGURE 27/H.120

Code transition rule for motion vector data and normal/subsampling mode change

3.6.5.5 Block-line skipping

If block-lines continue, in which all the prediction error e, data and motion vector v data are 0, and whose line mode data (QC1, QC2, IFM, FRP, SL1, SL2) are identical, their number is run-length coded with natural binary code as skipped block-lines. A run ends when it encounters FS or a block-line with new line mode data or some $e \neq 0$ or some $v \neq 0$. A run also ends when variable length code V_0 appears due to underflow prevention.

3.6.6 Buffer memory

3.6.6.1 Receiving buffer control

The staying period of FS in the transmission buffer is counted with a 1/16 input video line frequency clock and transmitted to the decoder as BC command. The staying period is represented in eight bit binary code. Similarly, the staying period in the receiving buffer is counted and the operation of the receiving buffer is so controlled as to make the total delay time caused by the two buffer memories constant.

NOTE – This control method is applicable even when the read out speed for the transmission buffer varies.

3.6.6.2 Memory size

Transmission buffer memory size B_S is defined as 180 kbits, while receiving buffer memory size B_R should be more than 220 kbits considering the variation of the transmission buffer read out speed.

NOTE – The delay time due to the transmission and receiving buffer memories becomes about 165 ms for B_S = 180 kbits and B_R = 220 kbits.

3.6.6.3 Underflow prevention

If the occupancy of the transmission buffer decreases to a threshold, the run length coding for the prediction error is prohibited and variable length code V_0 is used.

3.6.6.4 Overflow prevention

If the occupancy of the transmission buffer increases to another threshold, stop mode for setting forcedly all the prediction error and motion vector data to 0 is applied.

3.7 Audio coding

An audio channel using 64 kbit/s is included. The audio coding algorithm complies with Recommendation G.722.

Since video coding and decoding introduces a significant delay as described in 3.3.11, the encoded audio signal should be delayed by the corresponding time in the coder and decoder to obtain the proper synchronization between video and audio at the decoder. The delay inserted in the audio coder should be the sum of a half of the buffer memory delay and other video coding process delay, while the delay inserted in the audio decoder should be the sum of a half of the buffer delay and other video decoding process delay.

3.8 Transmission coding

3.8.1 General

The transmission coder assembles the video, audio, optional data and codec-to-codec information channels into a 1544 kbit/s digital stream. With all serialized data, the most significant digit leads.

3.8.2 Encryption

Video and audio signals can be independently encrypted on an optional basis. Their algorithms are under study. Keys and other control information can be transmitted through the message channel provided in the codec-to-codec information channel.

3.8.3 Error correction

An encoded (and encrypted) video signal is forward error corrected by a (255, 239) two-error correcting BCH code with the following polynomial generator:

$$g(x) = (1 + x^2 + x^3 + x^4 + x^8)\ (1 + x + x^2 + x^4 + x^5 + x^6 + x^8)$$

One framing bit is added to each 255 bit error correction frame, and 16 such frames are assembled into one large frame as shown in Figure 28. The frame alignment pattern is 0001101*y* (*y*: for future multiframe alignment signal use). The other 8 bits are used for controlling purposes, whose protocol is under study.

To correct a burst error of up to 32 bits, 16 phase interleaving is employed. The bit allocation rule is also indicated in Figure 28. Note that framing bits are excluded from interleaving.

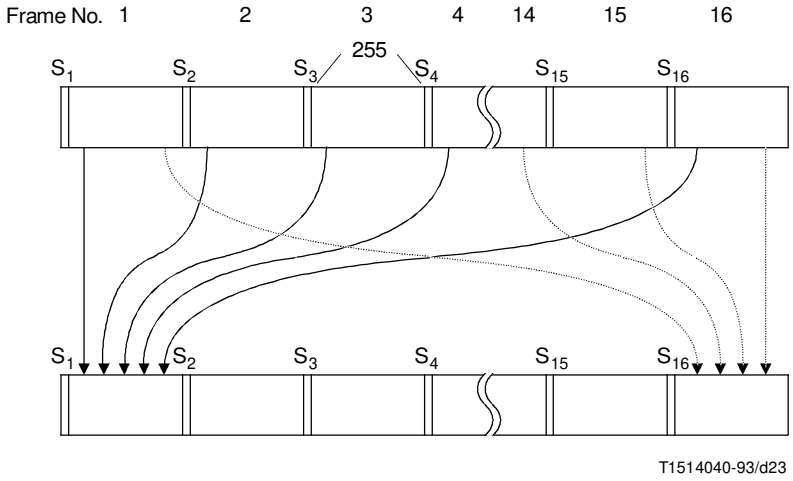


FIGURE 28/H.120

Error correction frames and interleaving

3.8.4 Scrambling

An error corrected video signal is scrambled with an 8-stage pseudo-random pulse generator to reduce stuffing required by the network restrictions. At each error correction frame bit, the scrambler is reset. The polynomial generator and the scrambled output pattern following the reset pulse for the input of all zeros are as follows:

$$1 + x^4 + x^5 + x^6 + x^8,$$

1111010011…1001111011.

3.8.5 Frame structure and stuffing

Clause 3/H.130 is applied.

Annex A

(referring to clause 1 of Recommendation H.120)

Graphics option – 625 line

(This annex forms an integral part of this Recommendation)

A.1 Introduction

In order to comply with the requirements of Recommendation H.100, an optional graphics mode may be provided giving improved definition at the expense of ability to convey movement. Two suitable arrangements are as follows:

A.2 Graphics codec for videoconferencing graphics – Mode 1

A.2.1 Facilities

The graphics mode provides, for still pictures, a capability of full 625-line luminance definition and colour definition which is better than that of the PAL and SECAM systems. It provides a limited capability for conveying movement, sufficient to permit pointing at items under discussion on the display. When the codec is in or adjacent to, the conference room, an alternative frozen-frame mode enables the face-to-face picture frozen for about 1.5 s while the graphics picture is being transmitted, to restart while the frozen graphics picture was displayed on another monitor.

The definition in the graphics mode is sufficient to permit good reproduction of one-half an A4 page of typescript.

A.2.2 Coding

The luminance and colour-difference signals are sampled at 12.5 MHz and 12.5/3 MHz respectively, the sampling frequency being locked to the television line-scanning frequency.

The samples are converted to PCM with 6 bits/sample. The luminance signal has a two-level half-sampling-frequency dither signal added to it which reduces the quantization distortion to approximately that of 7-bit coding.

Only the active picture area is sampled. Thus, there are 639 luminance samples along the line and two fields of 288 lines.

One colour-difference sample occurs every three luminance samples. Two of the 6 bits from a colour-difference sample are added to each of the three associated luminance samples, giving three 8-bit words for three luminance samples plus one colour-difference sample.

The $(E'_R - E'_Y)$ component is associated with the first, third, fifth, etc., active lines of Field No. 1, with the $(E'_B - E'_Y)$ component on the intervening lines, the pattern being reversed in Field No. 2.

The colour-difference samples are delayed with respect to the luminance samples to which they are attached, so that when decoded, they are coincident with the luminance output. The centre of the first colour-difference element on a line is co-sited with the second luminance element. Similarly, the centre of the 213th colour-difference element is co-sited with the 638th luminance element.

The luminance signal is amplitude-limited so that its PCM values are confined to the range:

Black level 000000

White level (700 mV):
transmission between ..... 100111 and 111000

Maximum level (750 mV) 111011

The colour-difference signals are limited to the range: 000000 to 111111 (0 to 63) with black level at 100000 (32). The 100/0/75/0 colour bar signal (see CCIR Recommendation 471 for explanation of nomenclature) fills the range: 000100 to 111100 (4 to 60). Prior to transmission, the colour difference codes are converted to the two's complement form by inversion of the most significant bit. This gives the range: 100000 to 011111 (–32 to 31) with the black level at 000000. The colour bar signal then occupies the range: 100100 to 011100 (–28 to 28).

A.2.3 Transmission and synchronization

A.2.3.1 General

The PCM words, formed as above, are transmitted to give a continuous update of the picture store in the receiver. The update pattern, which has been chosen to give smooth transitions on a changing picture, transmits every 19th luminance sample (with associated colour-difference data). The sequence of every 19th sample continues from one line to the next, as if the 639 elements in one active line are followed immediately (that is, without a gap for line blanking) by the 639 active elements in the next line. The use of this continuous sequence of samples makes line addressing unnecessary. A field synchronization code, followed by the address (in the range of 0 to 18) of the first luminance sample of the first active line, provides all the necessary synchronizing information.

The field synchronization code comprises 8 bytes of the form 11110011 or 11111100, which are invalid PCM values. The order of the last two pairs of bits, 0011 or 1100, in each of the first seven bytes, representing 0 and 1 respectively, signals the address of the first element of the field. In the 8th byte, 1100 signals a first field (starting line 23) and 0011 a second field (starting line 336).

The sequence in which the fields are transmitted by the address of the first luminance sample of the first line need not be specified because the decoder reconstructs the picture from the received addresses. A sequence which has been found satisfactory, giving no patterns with moving objects (e.g. a pointing finger), is as follows with the number in parentheses indicating a first or second field:

1 (2), 13 (1), 6 (2), 18 (1), 3 (2), 10 (1), 15 (2), 4 (1), 0 (2), 8 (1), 12 (2), 5 (1), 14 (2), 9 (1), 17 (2), 2 (1), 11 (2), 7 (1), 16 (2),

followed by:

1 (1), 13 (2), 6 (1), . . . in the same sequence as above, but with the field number interchanged.

After 38 fields, the complete picture has been replenished and the sequence repeats itself from the beginning.

A.2.3.2 Data structure

In each transmitted field, the data comprises 8 field synchronization bytes followed by 9685 or 9686 bytes of picture data (the total number of picture elements per field, 639×288, is not divisible by 19). Fields where the address of the first element is in the range 0 to 16 have 9686 transmitted bytes, while those with the first element address 17 or 18, have 9685 transmitted bytes.

Each byte of picture data comprises 6 bits of luminance data plus 2 bits of colour-difference data. The most significant bits of the colour-difference sample are transmitted first and the pairs of bits of the colour-difference data are placed in the least significant positions of the picture data byte. The data is arranged so that the luminance sample for the first picture element of a line carries with it the two most significant bits of the 19th colour-difference element belonging to the following line. The centre bits and the least significant bits of this colour-difference element are attached to the 20th and 39th along-the-line addresses respectively, these being the next two transmitted samples.

There are no colour-difference data transmitted for the first line of the picture and, on the second line, the first 18 colour-difference elements of the second line cannot be reconstructed at the decoder.

A.2.3.3 Data output

The graphics data are generated at a nominal rate of 3.74 Mbit/s and are transmitted via a buffer store whose capacity is in excess of 160 kbits. The output to the transmission channel is at less than 2 Mbit/s, the actual value depending on the number of time slots allocated to video. When the buffer level at the end of a field exceeds 160 kbits, the sampling is suspended for two complete fields to allow the buffer to empty.

If the level is then still in excess of 160 kbits, sampling is suspended for a further two fields.

The octet structure of the output data must be aligned with the time slot structure of the primary rate interface.

The approximate transmission time for a complete picture is therefore in the range of 1.6 to 4.6 s.

A.2.4 Decoder

The received data are associated with addresses derived from the field synchronization code and assembled in a picture store having 639 × 576 addressable positions with 8-bit capacity. The data are stored in the multiplexed form (luminance and colour-difference) used for transmission. The contents of this store are read out sequentially, the luminance and colour-difference components demultiplexed, and the colour-difference components line-interpolated to give the $(E'_R - E'_Y)$ and $(E'_B - E'_Y)$ components simultaneously and coincident with their associated luminance.

A.3 Graphics codec for videoconferencing-graphics – Mode 2

A.3.1 Facilities

The graphics mode provides, for still pictures, the capability of full 625-line luminance and colour definition. It allows transmission for still pictures with studio quality, defined in CCIR Recommendation 601. The graphics codec can be operated in two modes. In the single shot mode, the face-to-face picture is frozen for about 4 s, while the graphics picture is being transmitted, to restart while the graphics picture is displayed on another monitor. In the continuous mode, the face-to-face picture is frozen during the graphics presentation. The graphics picture is continuously transmitted in order to reproduce slow movement, e.g. for blackboard presentations. When the graphics picture has stabilized or when the presentation is finished, the graphics picture is frozen and the face-to-face picture is restarted.

The definition of the graphics mode 2 is better than that of the PAL, SECAM and NTSC systems and is sufficient to permit good reproduction of one-half of an A4 page typescript.

A.3.2 Coding

The luminance signal (E'_Y) and colour-difference signals $(E'_R - E'_Y, E'_B - E'_Y)$ are sampled at 13.5 MHz and 6.75 MHz respectively, according to the encoding parameters of digital television for studios given in CCIR Recommendation 601. The sampling frequencies are related in the ratio 4 : 2 : 2. The sampling structure is orthogonal, line, field and picture repetitive. The samples of the colour-difference signals are co-sited with the first, third, fifth, etc. luminance sample in each line. All samples are uniformly quantized PCM-values with 8 bits/sample.

Only the active picture area is sampled. There are 720 luminance samples along the line and two fields of 288 lines.

The luminance signal is amplitude-limited so that its PCM values are confined to the range:

Black level: 16;

Peak level: 235.

Each colour-difference signal exhibits 225 quantization levels in the centre part of the quantization scale with zero signal corresponding to level 128.

Further details are given in CCIR Recommendation 601.

A.3.3 Transmission and synchronization

A.3.3.1 General

Luminance and colour-difference samples of each line are arranged as sets for four samples:

$$\left[(E'_B - E'_Y)_n, (E'_Y)_n, (E'_R - E'_Y)_n, (E'_Y)_{n+1}\right]$$

where

n = 0, 2, 4, 6, . . ., 718. Each set consists of four words with 8 bit word length. There are 360 sets in each line of the picture.

The PCM sets are transmitted to give a continuous update of the picture store in the receiver. The update pattern which has been chosen transmits every 19th set. The sequence of every 19th set continues from one line to the next, as if the 360 sets in one active line are followed immediately (that is, without a gap for line blanking) by the 360 sets of the next line. The use of this continuous sequence of samples makes line addressing unnecessary. A field synchronization code, followed by the address of the first set of the first active line, provides all the necessary synchronizing information. The address is in the range 0 to 18.

The field synchronization code comprises eight bytes of the form 11110011 or 11111100. These two codewords are not allowed for the coded video signal. The order of the last two pairs of bits, 0011 or 1100, in each of the first seven bytes, representing 0 and 1 respectively, signals the address of the first set of the field. In the 8th byte, 1100 signals a first field (starting line 23) and 0011 a second field (starting line 336).

The sequence in which the fields are transmitted is defined by the address of the first set of the first line and need not be specified, because the decoder reconstructs the picture from received addresses.

A.3.3.2 Data structure

In each transmitted field, the data comprises 8 field synchronization bytes followed by the sets of picture data. In each set, $(E'_B - E'_Y)_n$ is transmitted first, followed by $(E'_Y)_n$, $(E'_R - E'_Y)_n$ and $(E'_Y)_{n+1}$. For transmission, a parallel-to-serial conversion takes place at the encoder. The most significant bits in the transmitted bit stream are leading.

A.3.3.3 Data output

The output bit-rate to the transmission channel is at less than 2 Mbit/s, the actual value depending on the number of time slots allocated to video.

The octet structure of the output data must be aligned with the time slot structure of the primary rate interface.

The approximate transmission time for a complete picture is 4 s.

A.3.4 Decoder

The received data are associated with addressed derived from the field synchronization code and assembled in a picture store having a capacity of 6.6355 Mbits. The contents of this store are read out sequentially.

A.3.5 Interface

A.3.5.1 Video interface

i) *Analogue interface* – An RGB-interface rather than a composite signal interface (PAL, SECAM) is recommended to maintain a high quality video signal.

ii) *Digital interface* – The structure of sets defined in A.3.3.2 allows for definition of a digital interface according to CCIR Recommendation 656 for E'_Y, $E'_R - E'_Y$ and $E'_B - E'_Y$.

A.3.5.2 Digital interface for transmission signal

The graphics mode may be internal or external to the face-to-face codec. An external device may have a digital interface according to Recommendations X.21 and V.11 (leased circuits). The picture data must be delayed at least 40 ms with respect to the control signal C specified in Recommendation X.21.

A.3.6 Signalling of graphics mode 2

The graphics mode 2 is signalled in the codec-to-codec information with bit 3.1.5 set to 1. For nomenclature of the bits, see Recommendation H.130.

A.3.7 Compatibility with graphics mode 1

In the graphics encoder and decoder, additional means are incorporated to make graphics mode 2 compatible with mode 1. They are signalled by setting bit 3.1.0 to 1 in the codec-to-codec information. If bit 3.1.5 of the codec-to-codec information is received set to 0 and bit 3.1.0 is received set to 1, then the graphics codec automatically switches into mode 1.

Annex B

(referring to clause 1 of Recommendation H.120)

Encryption option – 625 line

Under study.

Annex C

(referring to clause 2 of Recommendation H.120)

Graphics option – 525 line

(This annex forms an integral part of this Recommendation)

C.1 Introduction

The 525-line version of this graphics mode is very similar to the 625-line version for mode 1 specified in Annex A. It uses the same systematic replenishment technique and since the receiver is totally asynchronous from the transmitter, no adjustment for the differing picture rates is necessary. Instead of any form of standards conversion, the interworking between the 525-line and 625-line versions is provided by creating a small change in picture size. In 525 to 525-line transmission, the displayed picture size is the same as produced by the transmitting camera. In 525 to 625-line transmission, the displayed picture is reduced in size and is surrounded by a small black border (about 8%). In 625 to 525-line transmission, the displayed picture is expanded (equivalent to an overscan of about 8.5% in each border) so that a small amount of the transmitted picture area is not displayed.

Most of the details of this graphics mode are identical to those of the 625-line version for mode 1 in Annex A, so that only the difference need be specified in this Annex.

C.2 Facilities

The facilities are essentially the same as for the 625-line version.

C.3 Coding

The luminance and colour-difference sampling frequencies are 10.08 and 10.08/3 MHz respectively, the sampling frequency being locked to the television line-scanning frequency.

The arrangements for PCM coding are identical with the 625-line version but an area greater than the active picture area is sampled. There are 639 samples per line, the same number as for the 625-line version; 494 or 516 lines per picture are sampled. When a 525-line signal is being sampled at 10.08 MHz, only about 537 samples are required for the active line. The excess 102 samples, set to black level, are placed evenly on each side of the active line samples.

For 525-line transmission, samples from the first active line of Field No. 1 (line 14) form the $(E'_B - E'_Y)$ component, while those from the first active line of Field No. 2 (line 277) form the $(E'_R - E'_Y)$ component. For 625-line transmission, samples from the first active line of Field No. 1 (line 9) form the $(E'_R - E_Y)$ component and those from the first active line of Field No. 2 (line 272) form the $(E'_B - E'_Y)$ component.

C.4 Transmission and synchronization

C.4.1 General

The systematic replenishment algorithm based on the consecutive transmission of every 19th sample is also used in the 525-line version. However, since the samples extend over almost the complete line period, the divide-by-nineteen clock is suspended for only one luminance sample period during line-blanking. During transmission to a 625-line decoder, 5 additional lines per field are included before the picture area starts together with 6 additional lines per field after the end of the picture area, increasing the lines per field from 247 to 258. The luminance and colour-difference values on the extra lines are set to black level. Also, the divide-by-nineteen clock is changed to divide-by-five while samples are being selected from the added lines. This deceives the 625-line decoder into believing that there are 19 lines (per field) of black above the picture and 22 lines below the picture, making the total number of lines per field equal to 288, the same value as in Annex A.

The field synchronization code and the method of identifying fields are identical to those described in Annex A (except that the first line of Field No. 1 may be either line 14 or 9 and that of Field No. 2 either line 277 or 272).

C.4.2 Data structure

In 525 to 525-line transmission, each transmitted field comprises 8 field-synchronization bytes followed by 8307 bytes of picture data.

In 525 to 625-line transmission, each transmitted field comprises 8 field-synchronization bytes followed by 9685 or 9686 bytes of picture data, exactly the same as in the 625-line version in Annex A. In the 525-line coder the picture data are assembled from:

5 lines of 639 samples, every 5th sample . . . . . 639 bytes;

247 lines of 639 samples, every 19th sample . . . . . 8207 bytes;

6 lines of 639 samples, every 5th sample . . . . . 766 bytes.

The number of bytes required form the 6 lines at the bottom of the picture is 739 or 740. The excess bytes (all at black level), arising from non-integral results of division, are discarded.

Other details of the data structure are as in Annex A.

C.4.3 Data output

The graphics data are generated at a nominal rate of about 4 Mbit/s and fed into buffer store. The output from the buffer is at less than 2 Mbit/s (depending on the number of time slots allocated to video). When the buffer level at the end of a field exceeds 160 kbits, the sampling is suspended for two complete fields to allow the buffer to empty. If the level is then still in excess of 160 kbits, sampling is suspended for a further two fields.

The octet structure of the output data must be aligned with the time slot structure of the primary rate interface.

The resulting transmission time for a complete picture is of the order of 1.7 to 3 s.

C.5 Decoder

The received data are associated with addresses derived from the field synchronization code and assembled in a picture store having 639 × 494 addressable positions with 8-bit capacity. The data are stored in the multiplexed form (luminance and colour-difference) used for transmission. The contents of the store are read out sequentially, the luminance and colour-difference components line-interpolated to give the $(E'_R - E'_Y)$ and $(E'_B - E'_Y)$ components simultaneously and coincident with the associated luminance.

The length of the line in the picture store is 639 elements; for a 525-line picture sampled at 10.08 MHz, the active line requires only 537 elements. When television blanking is applied to the output signals, the 102 extra elements are suppressed and the standard 525-line signal results.

When receiving a signal from a 625-line terminal, 639 elements per line are received and stored. However, the first 19 active lines and the last 22 active lines of each field of the 625-line signal are not read into the store and are discarded. This, together with the effect of the line blanking on the horizontal output from the store, provides a 525-line display corresponding to the 625-line input picture after a border about 8% wide has been trimmed from all four edges.

Annex D

(referring to clause 2 of Recommendation H.120)

Encryption option – 525 line

Under study.

Annex E

(referring to clause 3 of Recommendation H.120)

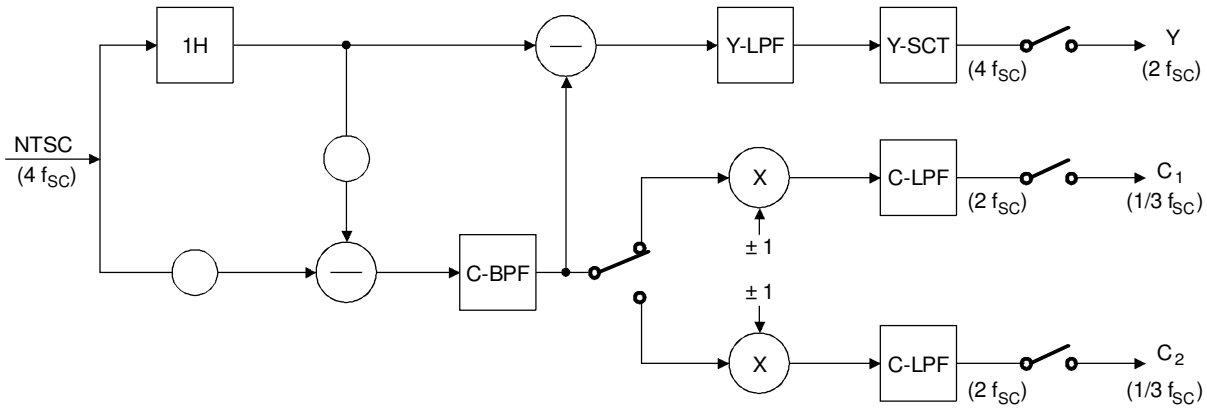
Colour decoding and coding filters

(This annex forms an integral part of this Recommendation)

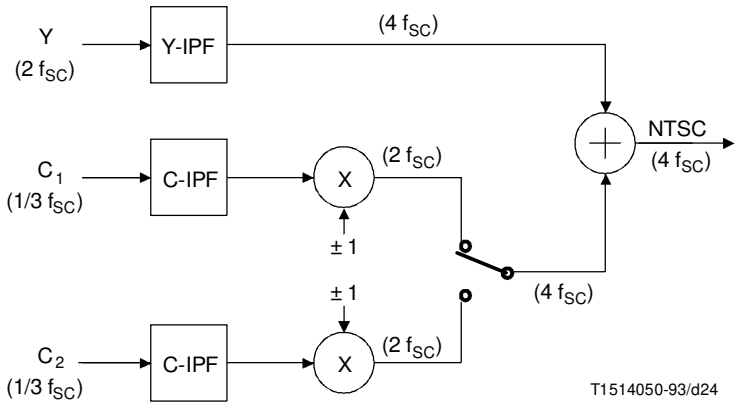
E.1 Configuration

See Figure E.1.

a) Digital colour separation circuit



b) Digital colour composition circuit



H	Line delay
Y-LPF	Low pass filter for Y signal
Y-SCT	Subcarrier trapper
C-BPF	Band pass filter for C signal
C-LPF	Los pass filter for C signal
f_{SC}	Colour subcarrier frequency
Y-IPF	Interpolation filter for Y signal
C-IPF	Interpolation filter for C signal

FIGURE E.1/H.120

E.2 Basic filter characteristics

See Table E.1.

TABLE E.1/H.120

Filter	Transfer function H(z)
C-BPF	$(-Z^{-2} + 2 - Z^{2})/4$
Y-LPF	$(-3Z^{-3} + 19Z^{-1} + 32 + 19Z - 3Z^{3})/64$
Y-SCT	$(Z^{-5} - 3Z^{-3} + 10Z^{-1} + 10Z - 3Z^{3} + Z^{5})/16$
C-LPF	$(Z^{-4} + 3Z^{-2} + 4 + 3Z^{2} + Z^{4})/12$
Y-IPF	$(-3Z^{-3} + 19Z^{-1} + 32 + 19Z - 3Z^{3})/64$
C-IPF	$(Z^{-2} + 1 + Z^{2})(Z^{-1} + 2 + Z)(-Z^{-8} - 2Z^{-6} + 2Z^{-4} + 6Z^{-2} + 6 + 6Z^{2} + 2Z^{4} - 2Z^{6} - Z^{8})/192$

E.3 Advanced filter characteristics

See Table E.2.

TABLE E.2/H.120

Filter	Transfer function H(z)
C-BPF	$(Z^{-8} - 9Z^{-6} + 17Z^{-4} - 23Z^{-2} + 28 - 23Z^{2} + 17Z^{4} - 9Z^{6} + Z^{8})/128$
Y-LPF	$(-Z^{-7} + 4Z^{-5} - 10Z^{-3} + 39Z^{-1} + 64 + 39Z - 10Z^{3} + 4Z^{5} - Z^{7})/128$
Y-SCT	$(Z^{-5} - 3Z^{-3} + 10Z^{-1} + 10Z - 3Z^{3} + Z^{5})/16$
C-LPF	$(Z^{-4} + 3Z^{-2} + 4 + 3Z^{2} + Z^{4})/12$
Y-IPF	$(-Z^{-7} + 4Z^{-5} - 10Z^{-3} + 39Z^{-1} + 64 + 39Z - 10Z^{3} + 4Z^{5} - Z^{7})/128$
C-IPF	$(Z^{-2} + 1 + Z^{2})(Z^{-1} + 2 + Z)(-Z^{-8} - 2Z^{-6} + 2Z^{-4} + 6Z^{-2} + 6 + 6Z^{2} + 2Z^{4} - 2Z^{6} - Z^{8})/192$

Annex F

(referring to clause 3 of Recommendation H.120)

An example of coding control sequence

(This annex forms an integral part of this Recommendation)

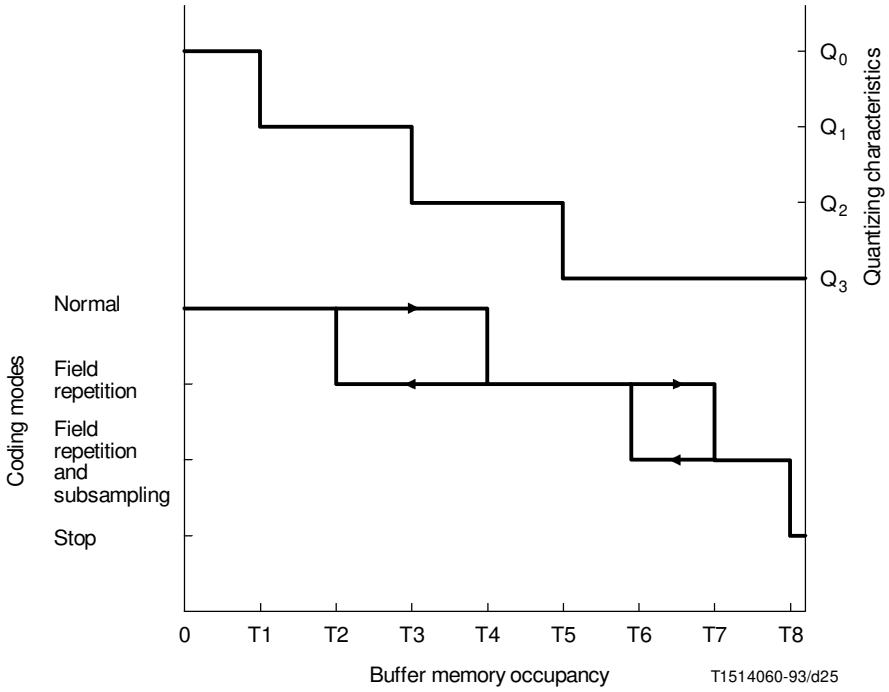


FIGURE F.1/H.120

Annex G

(referring to clause 3 of Recommendation H.120)

Examples of entropy coding

(This annex forms an integral part of this Recommendation)

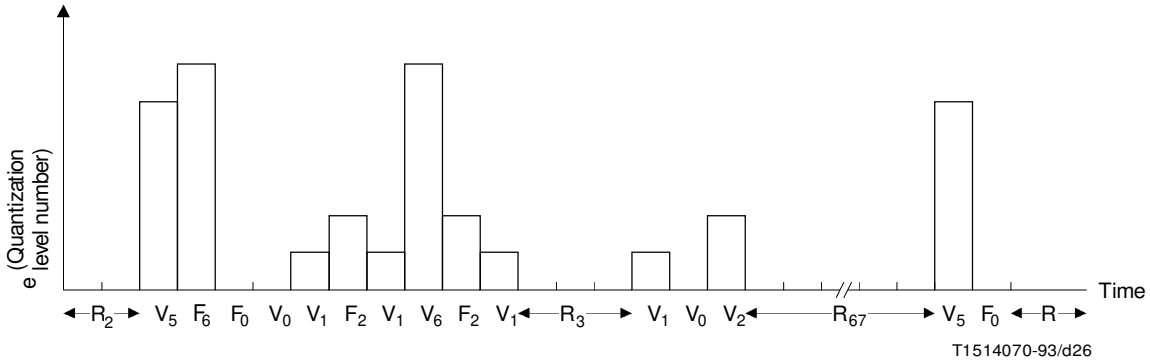
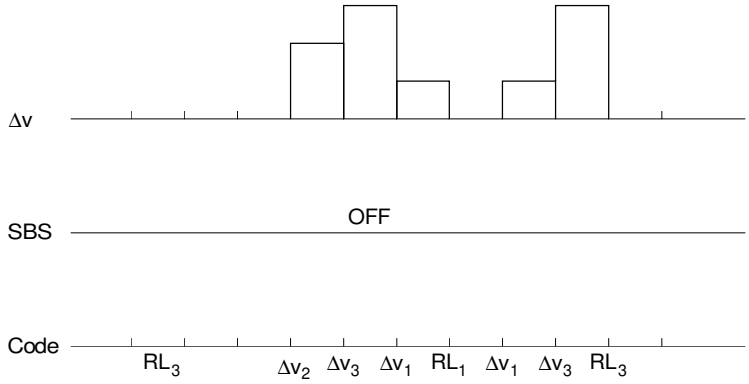


FIGURE G.1/H.120

Coding of prediction error e

a) When SBS is off



b) When SBS is on

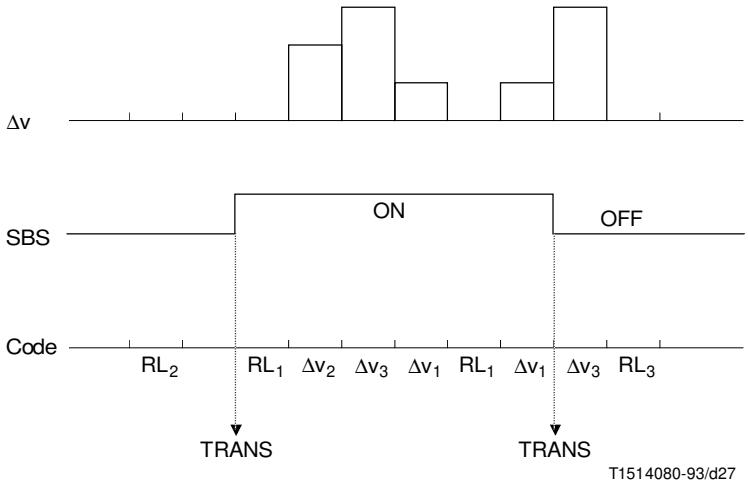


FIGURE G.2/H.120

Coding of motion vector v

Appendix I

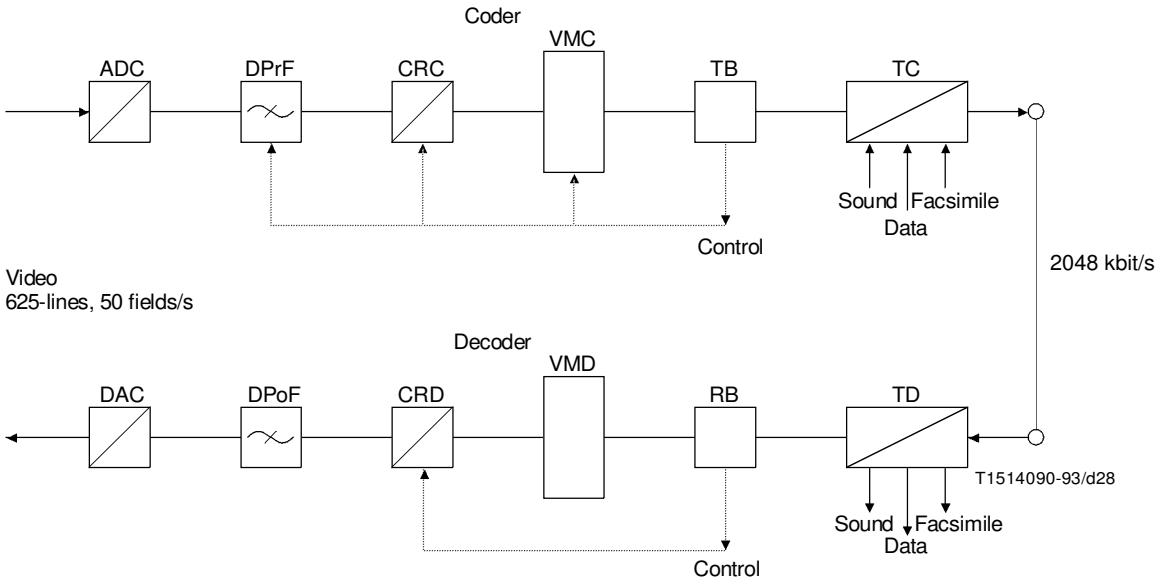
Outline description of operation of codecs in clauses 1 and 2

(This appendix does not form an integral part of this Recommendation)

Since the conditional replenishment codec is a complex and unfamiliar item, this simplified outline of its method of operation is included to make the Recommendation more easily comprehensible. More complete descriptions are to be found in published papers [1], [2].

A conditional replenishment codec operates by transmitting only those parts of a picture which differ significantly from one television frame to the next. This normally gives rise to data being generated in spurts separated by gaps in which no data are being generated. To match the non-uniform data generation to a channel transmitting at a uniform rate, a buffer is used to smooth out short-term fluctuations while, for longer-term variations, the coding algorithm is adaptively modified to change the rate of generation. In the event of too much data, caused for example by a lot of movement, the definition of the transmitted moving area is decreased, taking advantage of the reduced ability of the eye to perceive detail as the rate of movement increases. When little movement is present, the moving-area data are supplemented by data from non-moving areas in such a way that the whole picture is replenished over several picture periods. Picture stores are required at both transmitter and receiver and the objective is to make the content of the receiving store follow that of the transmitting store as closely as possible.

The codec can be regarded as comprising three basic sections: the source codec, the video multiplex codec and the transmission codec. Figure I-1 shows an outline of the arrangement.



Coder		Decoder	
ADC	Analogue-digital converter	DAC	Digital-analogue converter
DPrF	Digital pre-filter	DPoF	Digital post-filter
CRC	Conditional replenishment coder	CRD	Conditional replenishment decoder
VMC	Video multiplex coder	VMD	Video multiplex decoder
TB	Transmission buffer	RB	Receiving buffer
TC	Transmission coder	TD	Transmission decoder

FIGURE I.1/H.120

Outline block diagram of codec

In the source codec, the video signal is first digitized and optionally pre-filtered. When used, the pre-filter conditions the signal for further processing by reducing noise to improve the performance of the subsequent movement detector and to reduce the subjective effects of subsampling. The movement detector, in conjunction with the picture store, determines which areas in the picture are deemed to be moving. Noise introduces uncertainties in this decision and, when two or more groups of picture elements along a scanning line are deemed to be moving but are separated by small numbers of non-moving picture elements (probably caused by noise), the moving groups and separating elements are combined to form a single cluster, thus minimizing the addressing information which is required. Clusters of moving picture elements are then coded using DPCM followed by variable-length (entropy) coding where the shortest codes are allocated to the most frequently-occurring DPCM prediction errors.

The video multiplex codec adds to the video data the line- and field-synchronization signals together with addressing and other information (for example, whether PCM or DPCM is being transmitted) which must be transmitted in close association with the video to ensure that the decoder responds correctly.

The buffer, which strictly is part of the source coder, accepts the irregularly spaced bursts of data and delivers them at a uniform rate for transmission. The extent to which the buffer is filled is monitored and this is used to modify the rate of data generation by the source coder. It can reduce the data rate by modifying the pre-filter response and the thresholds in the movement detector, and by initiating element and field subsampling. On the other hand, if the buffer tends to empty, it may initiate the generation of complete PCM coded lines to provide systematic updating of the picture stores.

The transmission codec accepts the video data, adds a 64 kbit/s channel for sound, a 32 kbit/s channel for codec-codec signalling and optional additional data channels for facsimile, signalling or other data. It assembles the various signals into a frame structure, defined in Recommendation H.130 which is compatible with Recommendation G.732 and therefore suitable for transmission on 2048 kbit/s digital paths. In doing so, it provides the justification facilities to enable the clock for video processing to be independent of the network clock.

Appendix II

(referred to in Annexes A and C)
(This appendix does not form an integral part of this Recommendation)

II.1 Cursor data

In order to point out to items in the frozen graphics picture during face-to-face mode, a cursor has to be provided. The cursor is defined by a horizontal address, a vertical address and bits for cursor specification.

The horizontal address consists of 8 bits:

Horizontal address	*Codeword*
1st picture element	0
4th picture element	1
7th picture element	2
. .	
. .	
. .	
715th picture element	238
718th picture element	239

The vertical address consists of 8 bits:

Vertical address	*Codeword*	
1st line	0	The line numbers indicate lines in a full frame.
5th line	1	
9th line	2	
.		
.		
.		
569th line	142	
573rd line	143	

For the horizontal address codewords of the form 1111xxxx, for the vertical address codewords of the form 11xxxxxx (x = 0 or 1) are forbidden.

For cursor specification, 6 bits are used as follows: XCCSSS.

SSS Shape
000 = small circle
001 = arrow to the top
010 = arrow to the bottom
011 = arrow to the right
100 = arrow to the left
101 = cross

CC Colour
00 = black
01 = red
10 = green
11 = white

X on/off
0 = cursor off
1 = cursor on

The cursor defines the centre of the small circle, the centre of the cross, or the peak of an arrow.

The graphics codecs have been designed for the transmission of half an A4 page. If eight bits each are used for transmission of the horizontal and of the vertical address of the cursor, a resolution of the cursor positioning results which is 0.8 mm in the horizontal direction, and 1 mm in the vertical direction.

II.2 Frame structure

The cursor data are embedded in a pulse frame of length 32 bits as follows:

0111111111	X CC SSS	HHHHHHHH	VVVVVVVV
frame alignment word	cursor specification	horizontal address	vertical address

The frame alignment word cannot be simulated by the cursor data, as some codewords for the horizontal and vertical address are forbidden.

For transmission, the frame alignment word is leading, followed by the data for cursor specification. In the horizontal and vertical addresses the most significant bits are transmitted first.

II.3 Transmission of cursor data

The cursor data for the graphics codec is transmitted in time slot 17 of the transmission frame.

References

[1] DUFFY (T. S.) and NICOL (R. C.): A codec for visual teleconferencing, *Communications 82*, IEE Conference Publication No. 209, 1982.

[2] NICOL (R. C.), CHIARIGLIONE (L.) and SCHAEFER (P.): The development of the European Videotele-conference Codec, *Globecom 82*, IEEE global telecommunications conference, 1982.