

Implementation of
EN ISO/IEC 11172-2:1995

Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s —

Part 2: Video

Committees responsible for this British Standard

The preparation of this British Standard was entrusted by the Information Systems Technology Standards Policy Committee (IST/-) to Technical Committee IST/37, upon which the following bodies were represented:

- British Broadcasting Corporation
- British Federation of Printing Machinery and Supplies Ltd.
- British Telecommunications plc
- GEC Plessey Semiconductors Ltd.
- Information Systems Committee of the Universities Funding Council
- Institution of Electrical Engineers
- International Computers Limited
- National Transcommunications Ltd.
- Philips Research Laboratories
- Pioneer Digital Design Centre Ltd.

The following bodies were also represented in the drafting of the standard through subcommittees and panels:

- British Computer Society
- CCTA (the Government Centre for Information Systems)
- Department of Trade and Industry (IT Standards Unit)
- Digital Equipment Co. Ltd.
- IBM United Kingdom Ltd.
- Kodak Limited
- Level 7 Ltd.
- Ministry of Defence

This British Standard, having been prepared under the direction of the Information Systems Technology Standards Policy Committee, was published under the authority of the Standards Board and comes into effect on 15 October 1993

© BSI 03-2000

The following BSI references relate to the work on this standard:
Committee reference IST/37
Draft for comment 92/62857 DC

ISBN 0 580 22593 3

Amendments issued since publication

Amd. No.	Date	Comments
8798	September 1995	Indicated by a sideline in the margin

Contents

	Page
Committees responsible	Inside front cover
National foreword	ii
Foreword	2
Foreword	v
Text of ISO/IEC 11172-2	1

National foreword

This British Standard reproduces verbatim ISO/IEC 11172-2:1993 and implements it as the UK national standard.

In 1995 the European Committee for Standardization (CEN) accepted ISO/IEC 11172-2:1993 as European Standard EN ISO/IEC 11172-2:1995. As a consequence of implementing the European Standard this British Standard is renumbered as BS EN ISO/IEC 11172-2 and any reference to BS ISO/IEC 11172-2:1993 should be read as a reference to BS EN ISO/IEC 11172-2:1995.

This British Standard is published under the direction of the Information Systems Technology Standards Policy Committee whose Technical Committee IST/37 has the responsibility to:

- aid enquirers to understand the text;
- present to the responsible international committee any enquiries on interpretation, or proposals for change, and keep UK interests informed;
- monitor related international and European developments and promulgate them in the UK.

NOTE International and European Standards, as well as overseas standards, are available from Customer Services, Publications, BSI, Linford Wood, Milton Keynes, MK14 6LE.

A British Standard does not purport to include all the necessary provisions of a contract. Users of British Standards are responsible for their correct application.

Compliance with a British Standard does not of itself confer immunity from legal obligations.

Summary of pages

This document comprises a front cover, an inside front cover, pages i and ii, the EN ISO/IEC title page, page 2, the ISO/IEC title page, pages ii to vi, pages 1 to 110 and a back cover.

This standard has been updated (see copyright date) and may have had amendments incorporated. This will be indicated in the amendment table on the inside front cover.

ICS 35.040

Descriptors: Data processing, moving pictures, video data, video recording, data storage devices, digital storage, coded representation, coding (data conversion)

English version

Information technology — Coding of moving pictures and
associated audio for digital storage media at up to about
1,5 Mbit/s —
Part 2: Video

(ISO/IEC 11172-2:1993)

Technologies de l'information — Codage de
l'image animée et du son associé pour les
supports de stockage numérique jusqu'à
environ 1,5 Mbit/s —
Partie 2: Vidéo
(ISO/IEC 11172-2:1993)

Informationstechnik — Codierung von
bewegten Bildern und damit verbundenen
Tonsignalen für digitale Speichermedien bis zu
1,5 Mbit/s —
Teil 2: Video
(ISO/IEC 11172-2:1993)

This European Standard was approved by CEN on 1994-11-25. CEN members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration.

Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the Central Secretariat or to any CEN member.

This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CEN member into its own language and notified to the Central Secretariat has the same status as the official versions.

CEN members are the national standards bodies of Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and United Kingdom.

CEN

European Committee for Standardization
Comité Européen de Normalisation
Europäisches Komitee für Normung

Central Secretariat: rue de Stassart 36, B-1050 Brussels

© 1995 All rights of reproduction and communication in any form and by any means reserved in all countries to CEN and its members.

Ref. No. EN ISO/IEC 11172-2:1995 E
 **Documents**
Document Sharing Hub

Foreword

This European Standard was taken over by CEN from the work of ISO/IEC/JTC 1, Information technology.

The Technical Board had decided to submit the final draft for Formal Vote. The result was positive.

For the time being, this European Standard exists in English and French only.

This European Standard shall be given the status of a national standard, either by publication of an identical text or by endorsement, at the latest by August 1995, and conflicting national standards shall be withdrawn at the latest by August 1995.

According to the CEN/CENELEC Internal Regulations, the following countries are bound to implement this European Standard: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and United Kingdom.

**Information technology — Coding of
moving pictures and associated audio for
digital storage media at up to about
1,5 Mbit/s —**

Part 2:
Video

*Technologies de l'information — Codage de l'image animée et du son
associé pour les supports de stockage numérique jusqu'à environ
1,5 Mbit/s —*

Partie 2: Vidéo



Contents

	Page
Foreword	v
Introduction	1
Section 1. General	
1.1 Scope	6
1.2 Normative references	6
Section 2. Technical elements	
2.1 Definitions	7
2.2 Symbols and abbreviations	18
2.3 Method of describing bitstream syntax	21
2.4 Requirements	22
Annex A (normative) 8 by 8 Inverse discrete cosine transform	44
Annex B (normative) Variable length code tables	44
Annex C (normative) Video buffering verifier	51
Annex D (informative) Guide to encoding video	52
Annex E (informative) Bibliography	106
Annex F (informative) List of patent holders	106
Figure 1 — Example of temporal picture structure	2
Figure 2 — Simplified video encoder block diagram	3
Figure 3 — Macroblock structure	4
Figure 4 — Basic video decoder block diagram	5
Figure 5 — The position of luminance and chrominance samples	23
Figure 6 — The arrangement of blocks in a macroblock	24
Figure C.1 — VBV buffer occupancy	52
Figure D.1 — Coding and decoding process	53
Figure D.2 — Dependency relationship between I, B, and P-pictures	56
Figure D.3 — Typical sequence of pictures in display order	56
Figure D.4 — Typical sequence of pictures in bitstream order	56
Figure D.5 — Macroblock structure	56
Figure D.6 — Block structure	57
Figure D.7 — Simplified decoder block diagram	57
Figure D.8 — Conversion of CCIR 601 to SIF	59
Figure D.9 — Luminance subsampling filter tap weights	59
Figure D.10 — Chrominance subsampling filter tap weights	60
Figure D.11 — Example of filtering and subsampling of a line of pels	60
Figure D.12 — Conversion from CCIR 601 into SIF	60
Figure D.13 — Source input with significant pel area shaded dark	61
Figure D.14 — Model decoder	62
Figure D.15 — Default intra quantization matrix	67
Figure D.16 — Default non-intra quantization matrix	67
Figure D.17 — Examples of groups of pictures in display order	68
Figure D.18 — Example groups of pictures in display order	69
Figure D.19 — Examples of groups of pictures and temporal references	71
Figure D.20 — Example group of pictures containing 1 500 pictures	71
Figure D.21 — Possible arrangement of slices in a 256×192 picture	73
Figure D.22 — Possible arrangement of slices in a 256×192 picture	74
Figure D.23 — Group of pictures in display order	78
Figure D.24 — Interpolation of half pel shifts	79
Figure D.25 — Integer pel and half pel displacements	80
Figure D.26 — Logarithmic search method for integer pel shifts	81
Figure D.27 — Example group of pictures in display order	82

	Page
Figure D.28 — Transformation of pels to coefficients	85
Figure D.29 — Uniform quantizer characteristics	86
Figure D.30 — Quantized coefficient block in zigzag scan order	88
Figure D.31 — Example quantized coefficients	89
Figure D.32 — Selection of macroblock types in P-pictures	94
Figure D.33 — Characteristic MC/No MC	95
Figure D.34 — Characteristic intra/non-intra	96
Figure D.35 — Dead zone quantizer characteristic	97
Figure D.36 — Selection of macroblock type in B-pictures	100
Figure D.37 — Example group of pictures	103
Figure D.38 — Sequences	104
Figure D.39 — Upsampling filter for luminance	104
Figure D.40 — Upsampling filter for chrominance	104
Figure D.41 — Simplified decoder block diagram	105
Table B.1 — Variable length codes for macroblock_address_increment	44
Table B.2a — Variable length codes for macroblock_type in intra-coded pictures (I-pictures)	45
Table B.2b — Variable length codes for macroblock_type in predictive-coded pictures (P-pictures)	45
Table B.2c — Variable length codes for macroblock_type in bidirectionally predictive-coded pictures (B-pictures)	45
Table B.2d — Variable length codes for macroblock_type in dc intra-coded pictures (D-pictures)	45
Table B.3 — Variable length codes for coded_block_pattern	46
Table B.4 — Variable length codes for motion_horizontal_forward_code, motion_vertical_forward_code, motion_horizontal_backward_code, and motion_vertical_backward_code	47
Table B.5a — Variable length codes for dct_dc_size_luminance	47
Table B.5b — Variable length codes for dct_dc_size_chrominance	48
Table B.5c — Variable length codes for dct_coeff_first and dct_coeff_next	48
Table B.5d — Variable length codes for dct_coeff_first and dct_coeff_next	49
Table B.5e — Variable length codes for dct_coeff_first and dct_coeff_next	50
Table B.5f — Encoding of run and level following an escape code either as a 14-bit fixed length code ($-127 \leq \text{level} \leq 127$) or as a 22-bit fixed length code ($-255 \leq \text{level} \leq -128$, $128 \leq \text{level} \leq 255$)	51
Table D.1 — Conversion of source formats	61
Table D.2 — Pel aspect ratio	65
Table D.3 — Picture rate	65
Table D.4 — Time code fields	69
Table D.5 — Picture types	71
Table D.6 — f_codes	72
Table D.7 — Range of motion vectors	80
Table D.8 — Grid spacings for logarithmic searches	81
Table D.9 — Differential motion code	83
Table D.10 — Modulus for motion vectors	83
Table D.11 — Macroblock type VLC for I-pictures	84
Table D.12 — Differential dc size and VLC	87
Table D.13 — Differential dc additional code	88
Table D.14 — Example run lengths and levels	89
Table D.15 — Combination codes for DCT quantized coefficients. s = 0 for positive level, s = 1 for negative level	90

	Page
Table D.16 — Zero run length codes	91
Table D.17 — Level codes for DCT quantized coefficients	91
Table D.18 — Example run lengths, values, and VLC codes	91
Table D.19 — Macroblock type VLC for P-pictures	92
Table D.20 — VLC table for coded block pattern	93
Table D.21 — Macroblock type VLC for B-pictures	99
Table D.22 — Example of the coded data elements needed to generate repeated pictures	101

Descriptors: Data processing, moving pictures, video data, video recording, data storage devices, digital storage, coded representation, coding (data conversion).

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

International Standard ISO/IEC 11172-2 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Sub-Committee SC 29, *Coded representation of audio, picture, multimedia and hypermedia information*.

ISO/IEC 11172 consists of the following parts, under the general title *Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s*:

- *Part 1: Systems*;
- *Part 2: Video*;
- *Part 3: Audio*;
- *Part 4: Compliance testing*.

Annex A, Annex B and Annex C form an integral part of this part of ISO/IEC 11172. Annex D, Annex E and Annex F are for information only.

Introduction

NOTE Readers interested in an overview of the MPEG Video layer should read this Introduction and then proceed to Annex D, before returning to clauses 1 and 2.

0.1 Purpose

This part of ISO/IEC 11172 was developed in response to the growing need for a common format for representing compressed video on various digital storage media such as CDs, DATs, Winchester disks and optical drives. This part of ISO/IEC 11172 specifies a coded representation that can be used for compressing video sequences to bitrates around 1,5 Mbit/s. The use of this part of ISO/IEC 11172 means that motion video can be manipulated as a form of computer data and can be transmitted and received over existing and future networks. The coded representation can be used with both 625-line and 525-line television and provides flexibility for use with workstation and personal computer displays.

This part of ISO/IEC 11172 was developed to operate principally from storage media offering a continuous transfer rate of about 1,5 Mbit/s. Nevertheless it can be used more widely than this because the approach taken is generic.

0.1.1 Coding parameters

The intention in developing this part of ISO/IEC 11172 has been to define a source coding algorithm with a large degree of flexibility that can be used in many different applications. To achieve this goal, a number of the parameters defining the characteristics of coded bitstreams and decoders are contained in the bitstream itself. This allows for example, the algorithm to be used for pictures with a variety of sizes and aspect ratios and on channels or devices operating at a wide range of bitrates.

Because of the large range of the characteristics of bitstreams that can be represented by this part of ISO/IEC 11172, a sub-set of these coding parameters known as the "Constrained Parameters" has been defined. The aim in defining the constrained parameters is to offer guidance about a widely useful range of parameters. Conforming to this set of constraints is not a requirement of this part of ISO/IEC 11172. A flag in the bitstream indicates whether or not it is a Constrained Parameters bitstream.

Summary of the Constrained Parameters:

Horizontal picture size	Less than or equal to 768 pels
Vertical picture size	Less than or equal to 576 lines
Picture area	Less than or equal to 396 macroblocks
Pel rate	Less than or equal to 396×25 macroblocks/s
Picture rate	Less than or equal to 30 Hz
Motion vector range	Less than -64 to $+63,5$ pels (using half-pel vectors) [backward_f_code and forward_f_code ≤ 4 (see Table D.7)]
Input buffer size (in VBV model)	Less than or equal to 327 680 bits
Bitrate	Less than or equal to 1 856 000 bits/s (constant bitrate)

0.2 Overview of the algorithm

The coded representation defined in this part of ISO/IEC 11172 achieves a high compression ratio while preserving good picture quality. The algorithm is not lossless as the exact pel values are not preserved during coding. The choice of the techniques is based on the need to balance a high picture quality and compression ratio with the requirement to make random access to the coded bitstream. Obtaining good picture quality at the bitrates of interest demands a very high compression ratio, which is not achievable with intraframe coding alone. The need for random access, however, is best satisfied with pure intraframe coding. This requires a careful balance between intra- and interframe coding and between recursive and non-recursive temporal redundancy reduction.

A number of techniques are used to achieve a high compression ratio. The first, which is almost independent from this part of ISO/IEC 11172, is to select an appropriate spatial resolution for the signal. The algorithm then uses block-based motion compensation to reduce the temporal redundancy. Motion compensation is used for causal prediction of the current picture from a previous picture, for non-causal prediction of the current picture from a future picture, or for interpolative prediction from past and future pictures. Motion vectors are defined for each 16-pel by 16-line region of the picture. The difference signal, the prediction error, is further compressed using the discrete cosine transform (DCT) to remove spatial correlation before it is quantized in an irreversible process that discards the less important information. Finally, the motion vectors are combined with the DCT information, and coded using variable length codes.

0.2.1 Temporal processing

Because of the conflicting requirements of random access and highly efficient compression, three main picture types are defined. Intra-coded pictures (I-Pictures) are coded without reference to other pictures. They provide access points to the coded sequence where decoding can begin, but are coded with only a moderate compression ratio. Predictive coded pictures (P-Pictures) are coded more efficiently using motion compensated prediction from a past intra or predictive coded picture and are generally used as a reference for further prediction. Bidirectionally-predictive coded pictures (B-Pictures) provide the highest degree of compression but require both past and future reference pictures for motion compensation. Bidirectionally-predictive coded pictures are never used as references for prediction. The organisation of the three picture types in a sequence is very flexible. The choice is left to the encoder and will depend on the requirements of the application. Figure 1 illustrates the relationship between the three different picture types.

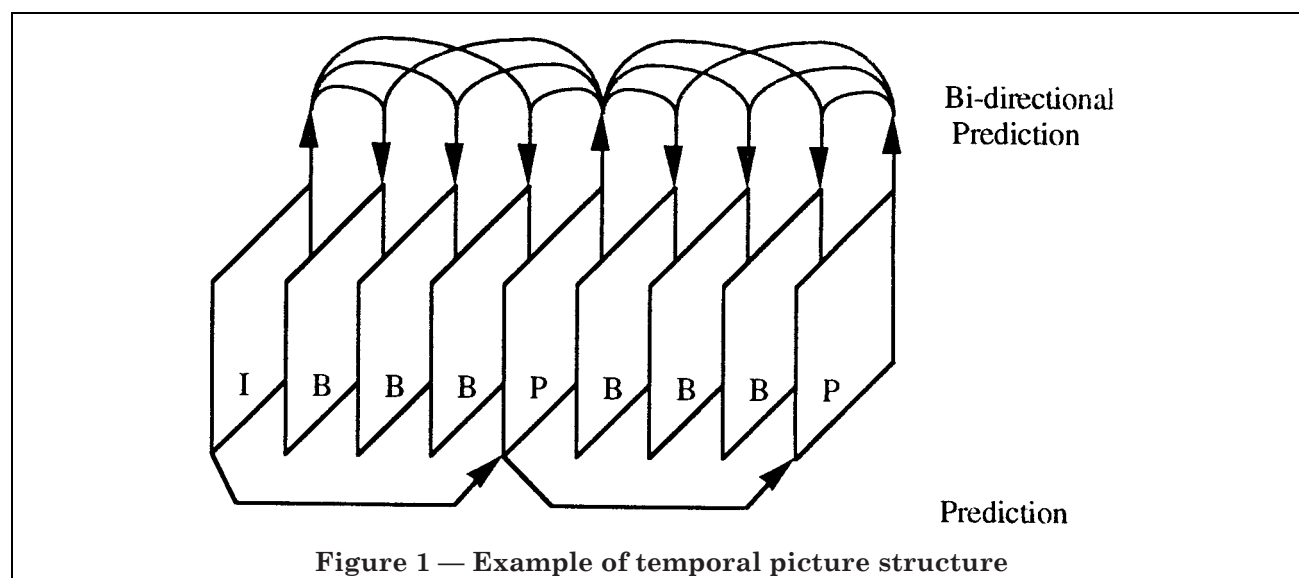


Figure 1 — Example of temporal picture structure

The fourth picture type defined in this part of ISO/IEC 11172, the D-picture, is provided to allow a simple, but limited quality, fast-forward playback mode.

0.2.2 Motion representation — macroblocks

The choice of 16 by 16 macroblocks for the motion-compensation unit is a result of the trade-off between increasing the coding efficiency provided by using motion information and the overhead needed to store it. Each macroblock can be one of a number of different types. For example, intra-coded, forward-predictive-coded, backward-predictive coded, and bidirectionally-predictive-coded macroblocks are permitted in bidirectionally-predictive coded pictures. Depending on the type of the macroblock, motion vector information and other side information are stored with the compressed prediction error signal in each macroblock. The motion vectors are encoded differentially with respect to the last coded motion vector, using variable-length codes. The maximum length of the vectors that may be represented can be programmed, on a picture-by-picture basis, so that the most demanding applications can be met without compromising the performance of the system in more normal situations.

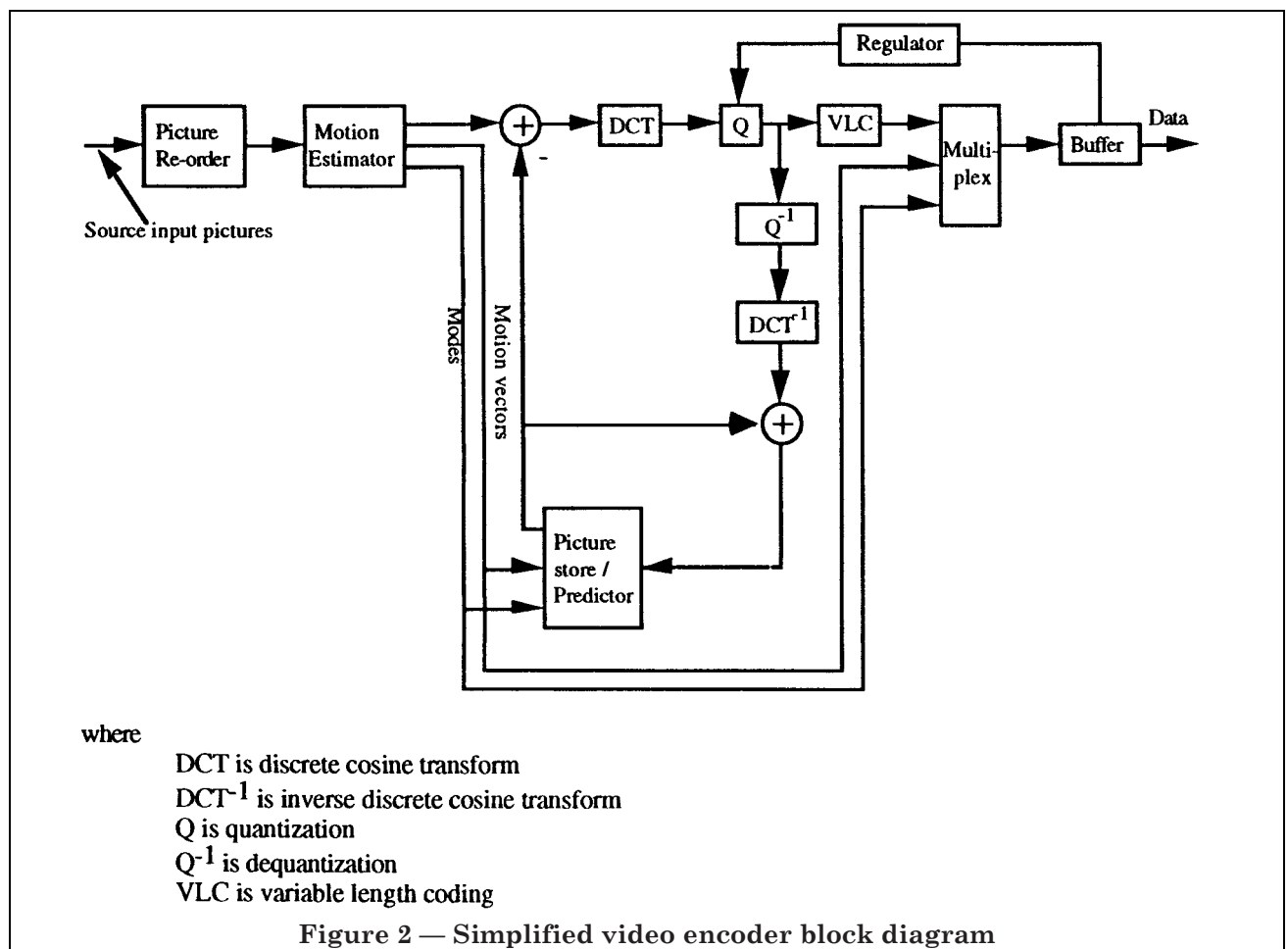
It is the responsibility of the encoder to calculate appropriate motion vectors. This part of ISO/IEC 11172 does not specify how this should be done.

0.2.3 Spatial redundancy reduction

Both original pictures and prediction error signals have high spatial redundancy. This part of ISO/IEC 11172 uses a block-based DCT method with visually weighted quantization and run-length coding. Each 8 by 8 block of the original picture for intra-coded macroblocks or of the prediction error for predictive-coded macroblocks is transformed into the DCT domain where it is scaled before being quantized. After quantization many of the coefficients are zero in value and so two-dimensional run-length and variable length coding is used to encode the remaining coefficients efficiently.

0.3 Encoding

This part of ISO/IEC 11172 does not specify an encoding process. It specifies the syntax and semantics of the bitstream and the signal processing in the decoder. As a result, many options are left open to encoders to trade-off cost and speed against picture quality and coding efficiency. This clause is a brief description of the functions that need to be performed by an encoder. Figure 2 shows the main functional blocks.

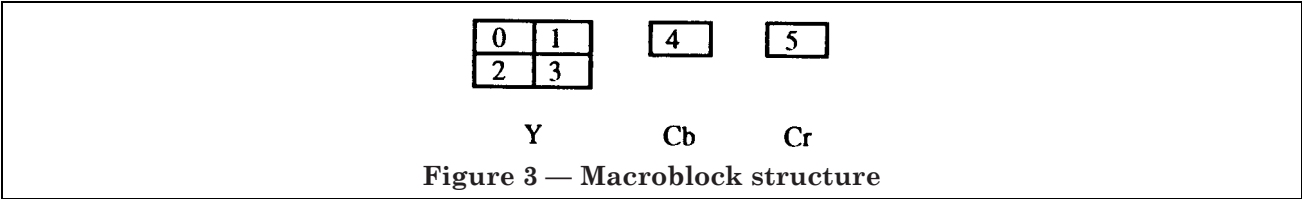


The input video signal must be digitized and represented as a luminance and two colour difference signals (Y , C_b , C_r). This may be followed by preprocessing and format conversion to select an appropriate window, resolution and input format. This part of ISO/IEC 11172 requires that the colour difference signals (C_b and C_r) are subsampled with respect to the luminance by 2 : 1 in both vertical and horizontal directions and are reformatted, if necessary, as a non-interlaced signal.

The encoder must choose which picture type to use for each picture. Having defined the picture types, the encoder estimates motion vectors for each 16 by 16 macroblock in the picture. In P-Pictures one vector is needed for each non-intra macroblock and in B-Pictures one or two vectors are needed.

If B-Pictures are used, some reordering of the picture sequence is necessary before encoding. Because B-Pictures are coded using bidirectional motion compensated prediction, they can only be decoded after the subsequent reference picture (an I or P-Picture) has been decoded. Therefore the pictures are reordered by the encoder so that the pictures arrive at the decoder in the order for decoding. The correct display order is recovered by the decoder.

The basic unit of coding within a picture is the macroblock. Within each picture, macroblocks are encoded in sequence, left to right, top to bottom. Each macroblock consists of six 8 by 8 blocks: four blocks of luminance, one block of Cb chrominance, and one block of Cr chrominance. See Figure 3. Note that the picture area covered by the four blocks of luminance is the same as the area covered by each of the chrominance blocks. This is due to subsampling of the chrominance information to match the sensitivity of the human visual system.



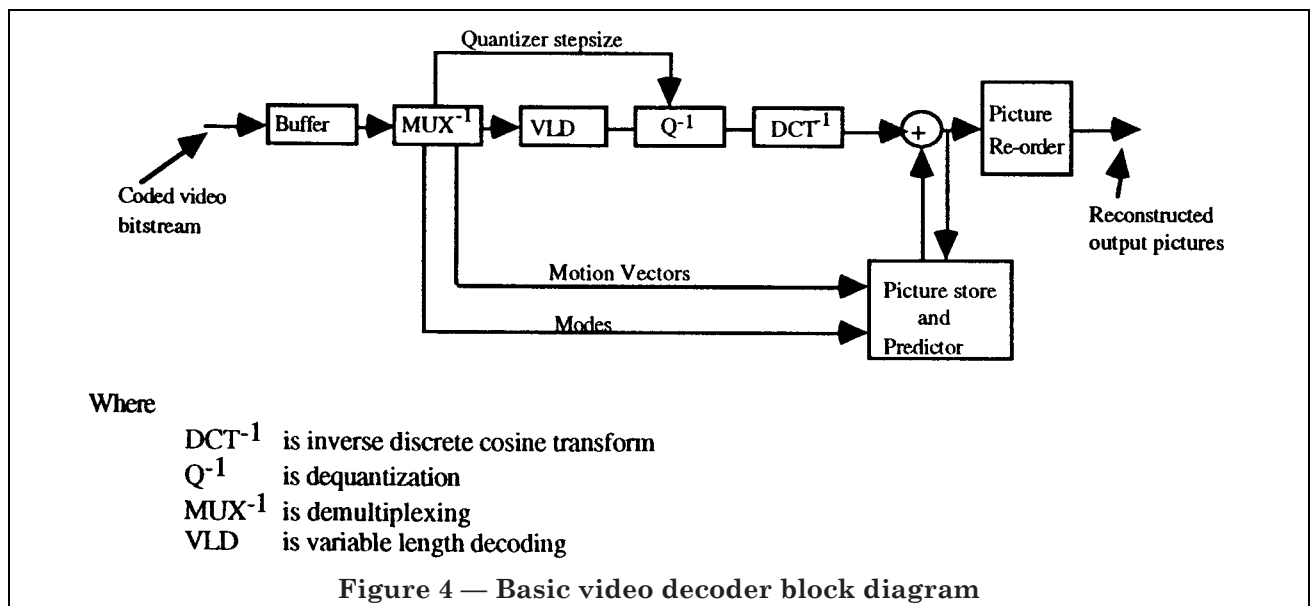
Firstly, for a given macroblock, the coding mode is chosen. It depends on the picture type, the effectiveness of motion compensated prediction in that local region, and the nature of the signal within the block. Secondly, depending on the coding mode, a motion compensated prediction of the contents of the block based on past and/or future reference pictures is formed. This prediction is subtracted from the actual data in the current macroblock to form an error signal. Thirdly, this error signal is separated into 8 by 8 blocks (4 luminance and 2 chrominance blocks in each macroblock) and a discrete cosine transform is performed on each block. Each resulting 8 by 8 block of DCT coefficients is quantized and the two-dimensional block is scanned in a zig-zag order to convert it into a one-dimensional string of quantized DCT coefficients. Fourthly, the side-information for the macroblock (mode, motion vectors etc) and the quantized coefficient data are encoded. For maximum efficiency, a number of variable length code tables are defined for the different data elements. Run-length coding is used for the quantized coefficient data.

A consequence of using different picture types and variable length coding is that the overall data rate is variable. In applications that involve a fixed-rate channel, a FIFO buffer may be used to match the encoder output to the channel. The status of this buffer may be monitored to control the number of bits generated by the encoder. Controlling the quantization process is the most direct way of controlling the bitrate. This part of ISO/IEC 11172 specifies an abstract model of the buffering system (the Video Buffering Verifier) in order to constrain the maximum variability in the number of bits that are used for a given picture. This ensures that a bitstream can be decoded with a buffer of known size.

At this stage, the coded representation of the picture has been generated. The final step in the encoder is to regenerate I-Pictures and P-Pictures by decoding the data so that they can be used as reference pictures for subsequent encoding. The quantized coefficients are dequantized and an inverse 8 by 8 DCT is performed on each block. The prediction error signal produced is then added back to the prediction signal and limited to the required range to give a decoded reference picture.

0.4 Decoding

Decoding is the inverse of the encoding operation. It is considerably simpler than encoding as there is no need to perform motion estimation and there are many fewer options. The decoding process is defined by this part of ISO/IEC 11172. The description that follows is a very brief overview of one possible way of decoding a bitstream. Other decoders with different architectures are possible. Figure 4 shows the main functional blocks.



For fixed-rate applications, the channel fills a FIFO buffer at a constant rate with the coded bitstream. The decoder reads this buffer and decodes the data elements in the bitstream according to the defined syntax. As the decoder reads the bitstream, it identifies the start of a coded picture and then the type of the picture. It decodes each macroblock in the picture in turn. The macroblock type and the motion vectors, if present, are used to construct a prediction of the current macroblock based on past and future reference pictures that have been stored in the decoder. The coefficient data are decoded and dequantized. Each 8 by 8 block of coefficient data is transformed by an inverse DCT (specified in Annex A), and the result is added to the prediction signal and limited to the defined range.

After all the macroblocks in the picture have been processed, the picture has been reconstructed. If it is an I-picture or a P-picture it is a reference picture for subsequent pictures and is stored, replacing the oldest stored reference picture. Before the pictures are displayed they may need to be re-ordered from the coded order to their natural display order. After reordering, the pictures are available, in digital form, for post-processing and display in any manner that the application chooses.

0.5 Structure of the coded video bitstream

This part of ISO/IEC 11172 specifies a syntax for a coded video bitstream. This syntax contains six layers, each of which either supports a signal processing or a system function:

Layers of the syntax	Function
Sequence layer	Random access unit: context
Group of pictures layer	Random access unit: video
Picture layer	Primary coding unit
Slice layer	Resynchronization unit
Macroblock layer	Motion compensation unit
Block layer	DCT unit

0.6 Features supported by the algorithm

Applications using compressed video on digital storage media need to be able to perform a number of operations in addition to normal forward playback of the sequence. The coded bitstream has been designed to support a number of these operations.

0.6.1 Random access

Random access is an essential feature for video on a storage medium. Random access requires that any picture can be decoded in a limited amount of time. It implies the existence of access points in the bitstream — that is segments of information that are identifiable and can be decoded without reference to other segments of data. A spacing of two random access points (Intra-Pictures) per second can be achieved without significant loss of picture quality.

0.6.2 Fast search

Depending on the storage medium, it is possible to scan the access points in a coded bitstream (with the help of an application-specific directory or other knowledge beyond the scope of this part of ISO/IEC 11172) to obtain a fast-forward and fast-reverse playback effect.

0.6.3 Reverse playback

Some applications may require the video signal to be played in reverse order. This can be achieved in a decoder by using memory to store entire groups of pictures after they have been decoded before being displayed in reverse order. An encoder can make this feature easier by reducing the length of groups of pictures.

0.6.4 Error robustness

Most digital storage media and communication channels are not error-free. Appropriate channel coding schemes should be used and are beyond the scope of this part of ISO/IEC 11172. Nevertheless the compression scheme defined in this part of ISO/IEC 11172 is robust to residual errors. The slice structure allows a decoder to recover after a data error and to resynchronize its decoding. Therefore, bit errors in the compressed data will cause errors in the decoded pictures to be limited in area. Decoders may be able to use concealment strategies to disguise these errors.

0.6.5 Editing

There is a conflict between the requirement for high coding efficiency and easy editing. The coding structure and syntax have not been designed with the primary aim of simplifying editing at any picture. Nevertheless a number of features have been included that enable editing of coded data.

Section 1. General

1.1 Scope

This part of ISO/IEC 11172 specifies the coded representation of video for digital storage media and specifies the decoding process. The representation supports normal speed forward playback, as well as special functions such as random access, fast forward playback, fast reverse playback, normal speed reverse playback, pause and still pictures. This part of ISO/IEC 11172 is compatible with standard 525- and 625-line television formats, and it provides flexibility for use with personal computer and workstation displays.

ISO/IEC 11172 is primarily applicable to digital storage media supporting a continuous transfer rate up to about 1,5 Mbit/s, such as Compact Disc, Digital Audio Tape, and magnetic hard disks. Nevertheless it can be used more widely than this because of the generic approach taken. The storage media may be directly connected to the decoder, or via communications means such as busses, LANs, or telecommunications links. This part of ISO/IEC 11172 is intended for non-interlaced video formats having approximately 288 lines of 352 pels and picture rates around 24 Hz to 30 Hz.

1.2 Normative references

The following International Standards contain provisions which, through reference in this text, constitute provisions of this part of ISO/IEC 11172. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this part of ISO/IEC 11172 are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO/IEC 11172-1:1993, *Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 1: Systems*.

ISO/IEC 11172-3:1993, *Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 3: Audio*.

CCIR Recommendation 601-2, *Encoding parameters of digital television for studios.*

CCIR Report 624-4, *Characteristics of systems for monochrome and colour television.*

CCIR Recommendation 648, *Recording of audio signals.*

CCIR Report 955-2, *Sound broadcasting by satellite for portable and mobile receivers, including Annex IV Summary description of Advanced Digital System II.*

CCITT Recommendation J.17, *Pre-emphasis used on Sound-Programme Circuits.*

IEEE Draft Standard P1180/D2:1990, *Specification for the implementation of 8×8 inverse discrete cosine transform*".

IEC publication 908:1987, *CD Digital Audio System.*

Section 2. Technical elements

2.1 Definitions

For the purposes of ISO/IEC 11172, the following definitions apply. If specific to a part, this is noted in square brackets.

2.1.1

ac coefficient [video]

any DCT coefficient for which the frequency in one or both dimensions is non-zero

2.1.2

access unit [system]

in the case of compressed audio an access unit is an audio access unit. In the case of compressed video an access unit is the coded representation of a picture

2.1.3

adaptive segmentation [audio]

a subdivision of the digital representation of an audio signal in variable segments of time

2.1.4

adaptive bit allocation [audio]

the assignment of bits to subbands in a time and frequency varying fashion according to a psychoacoustic model

2.1.5

adaptive noise allocation [audio]

the assignment of coding noise to frequency bands in a time and frequency varying fashion, according to a psychoacoustic model

2.1.6

alias [audio]

mirrored signal component resulting from sub-Nyquist sampling

2.1.7

analysis filterbank [audio]

filterbank in the encoder that transforms a broadband PCM audio signal into a set of subsampled subband samples

2.1.8

audio access unit [audio]

for Layers I and II an audio access unit is defined as the smallest part of the encoded bitstream which can be decoded by itself, where decoded means "fully reconstructed sound". For Layer III an audio access unit is part of the bitstream that is decodable with the use of previously acquired main information

2.1.9

audio buffer [audio]

a buffer in the system target decoder for storage of compressed audio data

2.1.10

audio sequence [audio]

a non-interrupted series of audio frames in which the following parameters are not changed:

- ID
- Layer
- Sampling Frequency
- For Layer I and II: Bitrate index

2.1.11

backward motion vector [video]

a motion vector that is used for motion compensation from a reference picture at a later time in display order

2.1.12

Bark [audio]

unit of critical band rate. The Bark scale is a non-linear mapping of the frequency scale over the audio range closely corresponding with the frequency selectivity of the human ear across the band

2.1.13

bidirectionally predictive-coded picture; B-picture [video]

a picture that is coded using motion compensated prediction from a past and/or future reference picture

2.1.14

bitrate

the rate at which the compressed bitstream is delivered from the storage medium to the input of a decoder

2.1.15

block companding [audio]

normalizing of the digital representation of an audio signal within a certain time period

2.1.16

block [video]

all 8-row by 8-column orthogonal block of pels

2.1.17

bound [audio]

the lowest subband in which intensity stereo coding is used

2.1.18

byte aligned

a bit in a coded bitstream is byte-aligned if its position is a multiple of 8-bits from the first bit in the stream

2.1.19

byte

sequence of 8-bits

2.1.20

channel

a digital medium that stores or transports an ISO/IEC 11172 stream

2.1.21

channel [audio]

the left and right channels of a stereo signal

2.1.22

chrominance (component) [video]

a matrix, block or single pel representing one of the two colour deference signals related to the primary colours in the manner defined in CCIR Rec 601. The symbols used for the colour difference signals are Cr and Cb

2.1.23**coded audio bitstream [audio]**

a coded representation of an audio signal as specified in ISO/IEC 11172-3

2.1.24**coded video bitstream [video]**

a coded representation of a series of one or more pictures as specified in this part of ISO/IEC 11172

2.1.25**coded order [video]**

the order in which the pictures are stored and decoded. This order is not necessarily the same as the display order

2.1.26**coded representation**

a data element as represented in its encoded form

2.1.27**coding parameters [video]**

the set of user-definable parameters that characterize a coded video bitstream. Bitstreams are characterised by coding parameters. Decoders are characterised by the bitstreams that they are capable of decoding

2.1.28**component [video]**

a matrix, block or single pel from one of the three matrices (luminance and two chrominance) that make up a picture

2.1.29**compression**

reduction in the number of bits used to represent an item of data

2.1.30**constant bitrate coded video [video]**

a compressed video bitstream with a constant average bitrate

2.1.31**constant bitrate**

operation where the bitrate is constant from start to finish of the compressed bitstream

2.1.32**constrained parameters [video]**

the values of the set of coding parameters defined in **2.4.3.2**

2.1.33**constrained system parameter stream (CSPS) [system]**

an ISO/IEC 11172 multiplexed stream for which the constraints defined in **2.4.6** of ISO/IEC 11172-1 apply

2.1.34**CRC**

cyclic redundancy code

2.1.35**critical band rate [audio]**

psychoacoustic function of frequency. At a given audible frequency it is proportional to the number of critical bands below that frequency. The units of the critical band rate scale are Barks

2.1.36**critical band [audio]**

psychoacoustic measure in the spectral domain which corresponds to the frequency selectivity of the human ear. This selectivity is expressed in Bark

2.1.37

data element

an item of data as represented before encoding and after decoding

2.1.38

de-coefficient [video]

the DCT coefficient for which the frequency is zero in both dimensions

2.1.39

dc-coded picture; D-picture [video]

a picture that is coded using only information from itself. Of the DCT coefficients in the coded representation, only the dc-coefficients are present

2.1.40

DCT coefficient

the amplitude of a specific cosine basis function

2.1.41

decoded stream

the decoded reconstruction of a compressed bitstream

2.1.42

decoder input buffer [video]

the first-in first-out (FIFO) buffer specified in the video buffering verifier

2.1.43

decoder input rate [video]

the data rate specified in the video buffering verifier and encoded in the coded video bitstream

2.1.44

decoder

an embodiment of a decoding process

2.1.45

decoding (process)

the process defined in ISO/IEC 11172 that reads an input coded bitstream and produces decoded pictures or audio samples

2.1.46

decoding time-stamp; DTS [system]

a field that may be present in a packet header that indicates the time that an access unit is decoded in the system target decoder

2.1.47

de-emphasis [audio]

filtering applied to an audio signal after storage or transmission to undo a linear distortion due to emphasis

2.1.48

dequantization [video]

the process of rescaling the quantized DCT coefficients after their representation in the bitstream has been decoded and before they are presented to the inverse DCT

2.1.49

digital storage media; DSM

a digital storage or transmission device or system

2.1.50

discrete cosine transform; DCT [video]

either the forward discrete cosine transform or the inverse discrete cosine transform. The DCT is an invertible, discrete orthogonal transformation. The inverse DCT is defined in Annex A

2.1.51**display order [video]**

the order in which the decoded pictures should be displayed. Normally this is the same order in which they were presented at the input of the encoder

2.1.52**dual channel mode [audio]**

a mode, where two audio channels with independent programme contents (e.g. bilingual) are encoded within one bitstream. The coding process is the same as for the stereo mode

2.1.53**editing**

the process by which one or more compressed bitstreams are manipulated to produce a new compressed bitstream. Conforming edited bitstreams must meet the requirements defined in this part of ISO/IEC 11172

2.1.54**elementary stream [system]**

a generic term for one of the coded video, coded audio or other coded bitstreams

2.1.55**emphasis [audio]**

filtering applied to an audio signal before storage or transmission to improve the signal-to-noise ratio at high frequencies

2.1.56**encoder**

an embodiment of an encoding process

2.1.57**encoding (process)**

a process, not specified in ISO/IEC 11172, that reads a stream of input pictures or audio samples and produces a valid coded bitstream as defined in ISO/IEC 11172

2.1.58**entropy coding**

variable length lossless coding of the digital representation of a signal to reduce redundancy

2.1.59**fast forward playback [video]**

the process of displaying a sequence, or parts of a sequence, of pictures in display-order faster than real-time

2.1.60**FFT**

fast fourier transformation. A fast algorithm for performing a discrete Fourier transform (an orthogonal transform)

2.1.61**filterbank [audio]**

a set of band-pass filters covering the entire audio frequency range

2.1.62**fixed segmentation [audio]**

a subdivision of the digital representation of an audio signal into fixed segments of time

2.1.63**forbidden**

the term “forbidden” when used in the clauses defining the coded bitstream indicates that the value shall never be used. This is usually to avoid emulation of start codes

2.1.64

forced updating [video]

the process by which macroblocks are intra-coded from time-to-time to ensure that mismatch errors between the inverse DCT processes in encoders and decoders cannot build up excessively

2.1.65

forward motion vector [video]

a motion vector that is used for motion compensation from a reference picture at an earlier time in display order

2.1.66

frame [audio]

a part of the audio signal that corresponds to audio PCM samples from an Audio Access Unit

2.1.67

free format [audio]

any bitrate other than the defined bitrates that is less than the maximum valid bitrate for each layer

2.1.68

future reference picture [video]

the future reference picture is the reference picture that occurs at a later time than the current picture in display order

2.1.69

granules [Layer II] [audio]

the set of 3 consecutive subband samples from all 32 subbands that are considered together before quantization. They correspond to 96 PCM samples

2.1.70

granules [Layer III] [audio]

576 frequency lines that carry their own side information

2.1.71

group of pictures [video]

a series of one or more coded pictures intended to assist random access. The group of pictures is one of the layers in the coding syntax defined in this part of ISO/IEC 11172

2.1.72

hann window [audio]

a time function applied sample-by-sample to a block of audio samples before Fourier transformation

2.1.73

huffman coding

a specific method for entropy coding

2.1.74

hybrid filterbank [audio]

a serial combination of subband filterbank and MDCT

2.1.75

IMDCT [audio]

inverse modified discrete cosine transform

2.1.76

intensity stereo [audio]

a method of exploiting stereo irrelevance or redundancy in stereophonic audio programmes based on retaining at high frequencies only the energy envelope of the right and left channels

2.1.77**interlace [video]**

the property of conventional television pictures where alternating lines of the picture represent different instances in time

2.1.78**intra coding [video]**

coding of a macroblock or picture that uses information only from that macroblock or picture

2.1.79**intra-coded picture; I-picture [video]**

a picture coded using information only from itself

2.1.80**ISO/IEC 11172 (multiplexed) stream [system]**

a bitstream composed of zero or more elementary streams combined in the manner defined in ISO/IEC 11172-1

2.1.81**joint stereo coding [audio]**

any method that exploits stereophonic irrelevance or stereophonic redundancy

2.1.82**joint stereo mode [audio]**

a mode of the audio coding algorithm using joint stereo coding

2.1.83**layer [audio]**

one of the levels in the coding hierarchy of the audio system defined in ISO/IEC 11172-3

2.1.84**layer [video and systems]**

one of the levels in the data hierarchy of the video and system specifications defined in ISO/IEC 11172-1 and this part of ISO/IEC 11172

2.1.85**luminance (component) [video]**

a matrix, block or single pel representing a monochrome representation of the signal and related to the primary colours in the manner defined in CCIR Rec 601. The symbol used for luminance is Y

2.1.86**macroblock [video]**

the four 8 by 8 blocks of luminance data and the two corresponding 8 by 8 blocks of chrominance data coming from a 16 by 16 section of the luminance component of the picture. Macroblock is sometimes used to refer to the pel data and sometimes to the coded representation of the pel values and other data elements defined in the macroblock layer of the syntax defined in this part of ISO/IEC 11172. The usage is clear from the context

2.1.87**mapping [audio]**

conversion of an audio signal from time to frequency domain by subband filtering and/or by MDCT

2.1.88**masking [audio]**

a property of the human auditory system by which an audio signal cannot be perceived in the presence of another audio signal

2.1.89**masking threshold [audio]**

a function in frequency and time below which an audio signal cannot be perceived by the human auditory system

2.1.90

MDCT [audio]

modified discrete cosine transform

2.1.91

motion compensation [video]

the use of motion vectors to improve the efficiency of the prediction of pel values. The prediction uses motion vectors to provide offsets into the past and/or future reference pictures containing previously decoded pel values that are used to form the prediction error signal

2.1.92

motion estimation [video]

the process of estimating motion vectors during the encoding process

2.1.93

motion vector [video]

a two-dimensional vector used for motion compensation that provides an offset from the coordinate position in the current picture to the coordinates in a reference picture

2.1.94

MS stereo [audio]

a method of exploiting stereo irrelevance or redundancy in stereophonic audio programmes based on coding the sum and difference signal instead of the left and right channels

2.1.95

non-intra coding [video]

coding of a macroblock or picture that uses information both from itself and from macroblocks and pictures occurring at other times

2.1.96

non-tonal component [audio]

a noise-like component of an audio signal

2.1.97

nyquist sampling

sampling at or above twice the maximum bandwidth of a signal

2.1.98

pack [system]

a pack consists of a pack header followed by one or more packets. It is a layer in the system coding syntax described in ISO/IEC 11172-1

2.1.99

packet data [system]

contiguous bytes of data from an elementary stream present in a packet

2.1.100

packet header [system]

the data structure used to convey information about the elementary stream data contained in the packet data

2.1.101

packet [system]

a packet consists of a header followed by a number of contiguous bytes from an elementary data stream. It is a layer in the system coding syntax described in ISO/IEC 11172-1

2.1.102

padding [audio]

a method to adjust the average length in time of an audio frame to the duration of the corresponding PCM samples, by conditionally adding a slot to the audio frame

2.1.103**past reference picture [video]**

the past reference picture is the reference picture that occurs at an earlier time than the current picture in display order

2.1.104**pel aspect ratio [video]**

the ratio of the nominal vertical height of pel on the display to its nominal horizontal width

2.1.105**pel [video]**

picture element

2.1.106**picture period [video]**

the reciprocal of the picture rate

2.1.107**picture rate [video]**

the nominal rate at which pictures should be output from the decoding process

2.1.108**picture [video]**

source, coded or reconstructed image data. A source or reconstructed picture consists of three rectangular matrices of 8-bit numbers representing the luminance and two chrominance signals. The Picture layer is one of the layers in the coding syntax defined in this part of ISO/IEC 11172. Note that the term “picture” is always used in ISO/IEC 11172 in preference to the terms field or frame

2.1.109**polyphase filterbank [audio]**

a set of equal bandwidth filters with special phase interrelationships, allowing for an efficient implementation of the filterbank

2.1.110**prediction [video]**

the use of a predictor to provide an estimate of the pel value or data element currently being decoded

2.1.111**predictive-coded picture; P-picture [video]**

a picture that is coded using motion compensated prediction from the past reference picture

2.1.112**prediction error [video]**

the difference between the actual value of a pel or data element and its predictor

2.1.113**predictor [video]**

a linear combination of previously decoded pel values or data elements

2.1.114**presentation time-stamp; PTS [system]**

a field that may be present in a packet header that indicates the time that a presentation unit is presented in the system target decoder

2.1.115**presentation unit; PU [system]**

a decoded audio access unit or a decoded picture

2.1.116**psychoacoustic model [audio]**

a mathematical model of the masking behaviour of the human auditory system

2.1.117

quantization matrix [video]

a set of sixty-four 8-bit values used by the dequantizer

2.1.118

quantized DCT coefficients [video]

DCT coefficients before dequantization. A variable length coded representation of quantized DCT coefficients is stored as part of the compressed video bitstream

2.1.119

quantizer scalefactor [video]

a data element represented in the bitstream and used by the decoding process to scale the dequantization

2.1.120

random access

the process of beginning to read and decode the coded bitstream at an arbitrary point

2.1.121

reference picture [video]

reference pictures are the nearest adjacent I- or P-pictures to the current picture in display order

2.1.122

reorder buffer [video]

a buffer in the system target decoder for storage of a reconstructed I-picture or a reconstructed P-picture

2.1.123

requantization [audio]

decoding of coded subband samples in order to recover the original quantized values

2.1.124

reserved

the term “reserved” when used in the clauses defining the coded bitstream indicates that the value may be used in the future for ISO/IEC defined extensions

2.1.125

reverse playback [video]

the process of displaying the picture sequence in the reverse of display order

2.1.126

scalefactor band [audio]

a set of frequency lines in Layer III which are scaled by one scalefactor

2.1.127

scalefactor index [audio]

a numerical code for a scalefactor

2.1.128

scalefactor [audio]

factor by which a set of values is scaled before quantization

2.1.129

sequence header [video]

a block of data in the coded bitstream containing the coded representation of a number of data elements

2.1.130

side information

information in the bitstream necessary for controlling the decoder

2.1.131

skipped macroblock [video]

a macroblock for which no data are stored

2.1.132**slice [video]**

a series of macroblocks. It is one of the layers of the coding syntax defined in this part of ISO/IEC 11172

2.1.133**slot [audio]**

a slot is an elementary part in the bitstream. In Layer I a slot equals four bytes, in Layers II and III one byte

2.1.134**source stream**

a single non-multiplexed stream of samples before compression coding

2.1.135**spreading function [audio]**

a function that describes the frequency spread of masking

2.1.136**start codes [system and video]**

32-bit codes embedded in that coded bitstream that are unique. They are used for several purposes including identifying some of the layers in the coding syntax

2.1.137**STD input buffer [system]**

a first-in first-out buffer at the input of the system target decoder for storage of compressed data from elementary streams before decoding

2.1.138**stereo mode [audio]**

mode, where two audio channels which form a stereo pair (left and right) are encoded within one bitstream. The coding process is the same as for the dual channel mode

2.1.139**stuffing (bits); stuffing (bytes)**

code-words that may be inserted into the compressed bitstream that are discarded in the decoding process. Their purpose is to increase the bitrate of the stream

2.1.140**subband [audio]**

subdivision of the audio frequency band

2.1.141**subband filterbank [audio]**

a set of band filters covering the entire audio frequency range. In ISO/IEC 11172-3 the subband filterbank is a polyphase filterbank

2.1.142**subband samples [audio]**

the subband filterbank within the audio encoder creates a filtered and subsampled representation of the input audio stream. The filtered samples are called subband samples, From 384 time-consecutive input audio samples, 12 time-consecutive subband samples are generated within each of the 32 subbands

2.1.143**syncword [audio]**

a 12-bit code embedded in the audio bitstream that identifies the start of a frame

2.1.144**synthesis filterbank [audio]**

filterbank in the decoder that reconstructs a PCM audio signal from subband samples

2.1.145

system header [system]

the system header is a data structure defined in ISO/IEC 11172-1 that carries information summarising the system characteristics of the ISO/IEC 11172 multiplexed stream

2.1.146

system target decoder; STD [system]

a hypothetical reference model of a decoding process used to describe the semantics of an ISO/IEC 11172 multiplexed bitstream

2.1.147

time-stamp [system]

a term that indicates the time of an event

2.1.148

triplet [audio]

a set of 3 consecutive subband samples from one subband. A triplet from each of the 32 subbands forms a granule

2.1.149

tonal component [audio]

a sinusoid-like component of an audio signal

2.1.150

variable bitrate

operation where the bitrate varies with time during the decoding of a compressed bitstream

2.1.151

variable length coding; VLC

a reversible procedure for coding that assigns shorter code-words to frequent events and longer code-words to less frequent events

2.1.152

video buffering verifier; VBV [video]

a hypothetical decoder that is conceptually connected to the output of the encoder. Its purpose is to provide a constraint on the variability of the data rate that an encoder or editing process may produce

2.1.153

video sequence [video]

a series of one or more groups of pictures. It is one of the layers of the coding syntax defined in this part of ISO/IEC 11172

2.1.154

zig-zag scanning order [video]

a specific sequential ordering of the DCT coefficients from (approximately) the lowest spatial frequency to the highest

2.2 Symbols and abbreviations

The mathematical operators used to describe this International Standard are similar to those used in the C programming language. However, integer division with truncation and rounding are specifically defined. The bitwise operators are defined assuming twos-complement representation of integers. Numbering and counting loops generally begin from zero.

2.2.1 Arithmetic operators

+	Addition.
−	Subtraction (as a binary operator) or negation (as a unary operator).
++	Increment.
--	Decrement.
*	Multiplication.
^	Power.
/	Integer division with truncation of the result toward zero. For example, $7/4$ and $-7/-4$ are truncated to 1 and $-7/4$ and $7/-4$ are truncated to -1 .
//	Integer division with rounding to the nearest integer. Half-integer values are rounded away from zero unless otherwise specified. For example $3//2$ is rounded to 2, and $-3//2$ is rounded to -2 .
DIV	Integer division with truncation of the result towards $-\infty$.
	Absolute value. $ x = x$ when $x > 0$ $ x = 0$ when $x == 0$ $ x = -x$ when $x < 0$
%	Modulus operator. Defined only for positive numbers.
Sign()	Sign(x) = 1 $x > 0$ 0 $x == 0$ -1 $x < 0$
NINT()	Nearest integer operator. Returns the nearest integer value to the real-valued argument. Half-integer values are rounded away from zero.
sin	Sine.
cos	Cosine.
exp	Exponential.
$\sqrt{\quad}$	Square root.
$\log_{10}$	Logarithm to base ten.
$\log_e$	Logarithm to base e.
$\log_2$	Logarithm to base 2.

2.2.2 Logical operators

	Logical OR.
&&	Logical AND.
!	Logical NOT.

2.2.3 Relational operators

>	Greater than.
>=	Greater than or equal to.
<	Less than.
<=	Less than or equal to.
==	Equal to.
!=	Not equal to.
max [...]	the maximum value in the argument list.
min [...]	the minimum value in the argument list.

2.2.4 Bitwise operators

A two's complement number representation is assumed where the bitwise operators are used.

&	AND.
	OR.
>>	Shift right with sign extension.
<<	Shift left with zero fill.

2.2.5 Assignment

=	Assignment operator.
---	----------------------

2.2.6 Mnemonics

The following mnemonics are defined to describe the different data types used in the coded bit-stream.

bslbf	Bit string, left bit first, where "left" is the order in which bit strings are written in ISO/IEC 11172. Bit strings are written as a string of 1s and 0s within single quote marks, e.g. '1000 0001'. Blanks within a bit string are for ease of reading and have no significance.
ch	Channel. If ch has the value 0, the left channel of a stereo signal or the first of two independent signals is indicated. (Audio)
nch	Number of channels; equal to 1 for single_channel mode, 2 in other modes. (Audio)
gr	Granule of 3 * 32 subband samples in audio Layer II, 18 * 32 sub-band samples in audio Layer III. (Audio)
main_data	The main_data portion of the bitstream contains the scalefactors, Huffman encoded data, and ancillary information. (Audio)
main_data_beg	The location in the bitstream of the beginning of the main_data for the frame. The location is equal to the ending location of the previous frame's main_data plus one bit. It is calculated from the main_data_end value of the previous frame. (Audio)
part2_length	The number of main_data bits used for scalefactors. (Audio)
rpchof	Remainder polynomial coefficients, highest order first. (Audio)
sb	Subband. (Audio)
sblimit	The number of the lowest sub-band for which no bits are allocated. (Audio)
scfsi	Scalefactor selection information. (Audio)
switch_point_l	Number of scalefactor band (long block scalefactor band) from which point on window switching is used. (Audio)
switch_point_s	Number of scalefactor band (short block scalefactor band) from which point on window switching is used. (Audio)
uimbsf	Unsigned integer, most significant bit first.
vlclbf	Variable length code, left bit first, where "left" refers to the order in which the VLC codes are written.
window	Number of the actual time slot in case of block_type == 2, 0 ≤ window ≤ 2. (Audio)

The byte order of multi-byte words is most significant byte first.

2.2.7 Constants

π	3,14159265358...
e	2,71828182845...

2.3 Method of describing bitstream syntax

The bitstream retrieved by the decoder is described in 2.4.2. Each data item in the bitstream is in bold type. It is described by its name, its length in bits, and a mnemonic for its type and order of transmission.

The action caused by a decoded data element in a bitstream depends on the value of that data element and on data elements previously decoded. The decoding of the data elements and definition of the state variables used in their decoding are described in 2.4.3. The following constructs are used to express the conditions when data elements are present, and are in normal type:

Note this syntax uses the “C”-code convention that a variable or expression evaluating to a non-zero value is equivalent to a condition that is true.

while (condition) {	If the condition is true, then the group of data elements occurs next in the data stream. This repeats until the condition is not true.
data_element	
. . .	
}	
do {	
data_element	The data element always occurs at least once.
. . .	
} while (condition)	The data element is repeated until the condition is not true.
if (condition) {	If the condition is true, then the first group of data elements occurs next in the data stream.
data_element	
. . .	
}	
else {	If the condition is not true, then the second group of data elements occurs next in the data stream.
data_element	
. . .	
}	
for (expr1; expr2; expr3) {	expr1 is an expression specifying the initialization of the loop. Normally it specifies the initial state of the counter, expr2 is a condition specifying a test made before each iteration of the loop. The loop terminates when the condition is not true. expr3 is an expression that is performed at the end of each iteration of the loop, normally it increments a counter.
data_element	
. . .	
}	

Note that the most common usage of this construct is as follows:

for (i = 0; i < n; i++) {	The group of data elements occurs n times. Conditional constructs within the group of data elements may depend on the value of the loop control variable i, which is set to zero for the first occurrence, incremented to one for the second occurrence, and so forth.
data_element	
. . .	
}	

As noted, the group of data elements may contain nested conditional constructs. For compactness, the {} may be omitted when only one data element follows.

data_element []	data_element [] is an array of data. The number of data elements is indicated by the context.
data_element [n]	data_element [n] is the n + 1th element of an array of data.
data_element [m][n]	data_element [m][n] is the m + 1, n + 1th element of a two-dimensional array of data.
data_element [l][m][n]	data_element [l][m][n] is the l + 1, m + 1, n + 1th element of a three-dimensional array of data.
data_element [m..n]	is the inclusive range of bits between bit m and bit n in the data_element.

While the syntax is expressed in procedural terms, it should not be assumed that 2.4.3 implements a satisfactory decoding procedure. In particular, it defines a correct and error-free input bitstream. Actual decoders must include a means to look for start codes in order to begin decoding correctly, and to identify errors, erasures or insertions while decoding. The methods to identify these situations, and the actions to be taken, are not standardized.

Definition of bytealigned function

The function bytealigned () returns 1 if the current position is on a byte boundary, that is the next bit in the bitstream is the first bit in a byte. Otherwise it returns 0.

Definition of nextbits function

The function nextbits () permits comparison of a bit string with the next bits to be decoded in the bitstream.

Definition of next_start_code function

The next_start_code function removes any zero bit and zero byte stuffing and locates the next start code.

Syntax	No. of bits	Mnemonic
next_start_code() { while (!bytealigned()) zero_bit while (nextbits() != '0000 0000 0000 0000 0000 0001') zero_byte }	1 8	"0" "00000000"

This function checks whether the current position is bytealigned. If it is not, zero stuffing bits are present. After that any number of zero bytes may be present before the start-code. Therefore start-codes are always bytealigned and may be preceded by any number of zero stuffing bits.

2.4 Requirements

2.4.1 Coding structure and parameters

Video sequence

A coded video sequence commences with a sequence header and is followed by one or more groups of pictures and is ended by a sequence_end_code. Immediately before each of the groups of pictures there may be a sequence header. Within each sequence, pictures shall be decodable continuously.

In each of these repeated sequence headers all of the data elements with the permitted exception of those defining the quantization matrices (load_intra_quantizer_matrix, load_non_intra_quantizer_matrix and optionally intra_quantizer_matrix and non_intra_quantizer_matrix) shall have the same values as in the first sequence header. The quantization matrices may be redefined each time that a sequence header occurs in the bitstream. Thus the data elements load_intra_quantizer_matrix, load_non_intra_quantizer_matrix and optionally intra_quantizer_matrix and non_intra_quantizer_matrix may have any (non-forbidden) values.

Repeating the sequence header allows the data elements of the initial sequence header to be repeated in order that random access into the video sequence is possible. In addition the quantization matrices may be changed inside the video sequence as required.

Sequence header

A video sequence header commences with a sequence_header_code and is followed by a series of data elements.

Group of pictures

A group of pictures is a series of one or more coded pictures intended to assist random access into the sequence. In the stored bitstream, the first coded picture in a group of pictures is an I-Picture. The order of the pictures in the coded stream is the order in which the decoder processes them in normal playback. In particular, adjacent B-Pictures in the coded stream are in display order. The last coded picture, in display order, of a group of pictures is either an I-Picture or a P-Picture.

The following is an example of groups of pictures taken from the beginning of a video sequence. In this example the first group of pictures contains seven pictures and subsequent groups of pictures contain nine pictures. There are two B-pictures between successive P-pictures and also two B-pictures between successive I- and P-pictures. Picture '1I' is used to form a prediction for picture '4P'. Pictures '4P' and '1I' are both used to form predictions for pictures '2B' and '3B'. Therefore the order of pictures in the coded sequence shall be '1I', '4P', '2B', '3B'. However, the decoder should display them in the order '1I', '2B', '3B', '4P'.

At the encoder input,

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
I	B	B	P	B	B	P	B	B	I	B	B	P	B	B	P	B	B	I	B	B	P	B	B	P

At the encoder output, in the stored bitstream, and at the decoder input,

1	4	2	3	7	5	6	10	8	9	13	11	12	16	14	15	19	17	18	22	20	21	25	23	24
I	P	B	B	P	B	B	I	B	B	P	B	B	P	B	B	I	B	B	P	B	B	P	B	B

where the double vertical bars mark the group of pictures boundaries. Note that in this example, the first group of pictures is two pictures shorter than subsequent groups of pictures, since at the beginning of video coding there are no B-pictures preceding the first I-Picture. However, in general, in display order, there may be B-Pictures preceding the first I-Picture in the group of pictures, even for the first group of pictures to be decoded.

At the decoder output,

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

A group of pictures may be of any length. A group of pictures shall contain one or more I-Pictures. Applications requiring random access, fast-forward playback, or fast and normal reverse playback may use relatively short groups of pictures. Groups of pictures may also be started at scene cuts or other cases where motion compensation is ineffective.

The number of consecutive B-Pictures is variable. Neither B- nor P-Pictures need be present.

A video sequence of groups of pictures that is read by the decoder may be different from the one at the encoder output due to editing.

Picture

A source or reconstructed picture consists of three rectangular matrices of eight-bit numbers; a luminance matrix (Y), and two chrominance matrices (Cb and Cr). The Y-matrix shall have an even number of rows and columns, and the Cb and Cr matrices shall be one half the size of the Y-matrix in both horizontal and vertical dimensions.

The Y, Cb and Cr components are related to the primary (analogue) Red, Green and Blue Signals (E'_R , E'_G and E'_B) as described in CCIR Recommendation 601. These primary signals are gamma pre-corrected. The assumed value of gamma is not defined in this part of ISO/IEC 11172 but may typically be in the region approximately 2,2 to approximately 2,8. Applications which require accurate colour reproduction may choose to specify the value of gamma more accurately, but this is outside the scope of this part of ISO/IEC 11172.

The luminance and chrominance samples are positioned as shown in Figure 5, where “x” marks the position of the luminance (Y) samples and “0” marks the position of the chrominance (Cb and Cr) samples:

x		x		x		x		x		x
	0				0				0	
x		x		x		x		x		x
—	—	—	—	—	—	—	—	—	—	—
x		x		x		x		x		x
	0				0				0	
x		x		x		x		x		x
—	—	—	—	—	—	—	—	—	—	—
x		x		x		x		x		x
	0				0				0	
x		x		x		x		x		x

Figure 5 — The position of luminance and chrominance samples

There are four types of coded picture that use different coding methods.

An **Intra-coded picture (I-picture)** is coded using information only from itself.

A **Predictive-coded picture (P-picture)** is a picture which is coded using motion compensated prediction from a past I-Picture or P-Picture.

A **Bidirectionally predictive-coded picture (B-picture)** is a picture which is coded using motion compensated prediction from a past and/or future I-Picture or P-Picture.

A **dc coded (D) picture** is coded using information only from itself. Of the DCT coefficients only the dc ones are present. The D-Pictures shall not be in a sequence containing any other picture types.

Slice

A **slice** is a series of an arbitrary number of macroblocks with the order of macroblocks starting from the upper-left of the picture and proceeding by raster-scan order from left to right and top to bottom. The first and last macroblocks of a slice shall not be skipped macroblocks (see 2.4.4.4). Every slice shall contain at least one macroblock. Slices shall not overlap and there shall be no gaps between slices. The position of slices may change from picture to picture. The first slice shall start with the first macroblock in the picture and the last slice shall end with the last macroblock in the picture.

Macroblock

A **macroblock** contains a 16-pel by 16-line section of luminance component and the spatially corresponding 8-pel by 8-line section of each chrominance component. A macroblock has 4 luminance blocks and 2 chrominance blocks. The term “macroblock” can refer to source or reconstructed data or to scaled, quantized coefficients. The order of blocks in a macroblock is top-left, top-right, bottom-left, bottom-right blocks for Y, followed by Cb and Cr. Figure 6 shows the arrangement of these blocks. A skipped macroblock is one for which no information is stored (see 2.4.4.4).

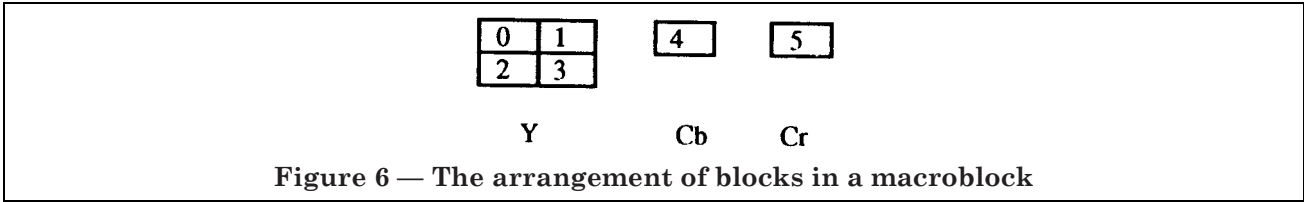


Figure 6 — The arrangement of blocks in a macroblock

Block

A **block** is an orthogonal 8-pel by 8-line section of a luminance or chrominance component. The term “block” can refer either to source and reconstructed data or to the corresponding coded data elements.

Reserved, Forbidden and Marker bit

The terms “reserved” and “forbidden” are used in the description of some values of several fields in the coded bitstream.

The term “reserved” indicates that the value may be used in the future for ISO/IEC-defined extensions.

The term “forbidden” indicates a value that shall never be used (usually in order to avoid emulation of start codes).

The term “marker_bit” indicates a one bit field in which the value zero is forbidden. These marker bits are introduced at several points in the syntax to avoid start-code emulation.

2.4.2 Specification of the coded video bitstream syntax

2.4.2.1 Start codes

Start codes are reserved bit patterns that do not otherwise occur in the video stream. All start codes are bytealigned.

Name	Hexadecimal value
picture_start_code	00000100
slice_start_codes (including slice_vertical_positions)	00000101 through 000001AF
reserved	000001B0
reserved	000001B1
user_data_start_code	000001B2
sequence_header_code	000001B3
sequence_error_code	000001B4
extension_start_code	000001B5
reserved	000001B6
sequence_end_code	000001B7
group_start_code	000001B8
system start codes (see note)	000001B9 through 000001FF
NOTE System start codes are defined in ISO/IEC 11172-1.	

The use of the start codes is defined in the following syntax description with the exception of the sequence_error_code. The sequence_error_code has been allocated for use by the digital storage media interface to indicate where uncorrectable errors have been detected.

2.4.2.2 Video sequence layer

Syntax	No. of bits	Mnemonic
<pre> video_sequence() { next_start_code() do { sequence_header() do { group_of_pictures() } while (nextbits() == group_start_code) } while (nextbits() == sequence_header_code) sequence_end_code } </pre>	32	bslbf

2.4.2.3 Sequence header

Syntax	No. of bits	Mnemonic
sequence_header() {		
sequence_header_code	32	bslbf
horizontal_size	12	uimsbf
vertical_size	12	uimsbf
pel_aspect_ratio	4	uimsbf
picture_rate	4	uimsbf
bit_rate	18	uimsbf
marker_bit	1	"1"
vbv_buffer_size	10	uimsbf
constrained_parameters_flag	1	
load_intra_quantizer_matrix	1	
if (load_intra_quantizer_matrix)		
intra_quantizer_matrix []	8 * 64	uimsbf
load_non_intra_quantizer_matrix	1	
if (load_non_intra_quantizer_matrix)		
non_intra_quantizer_matrix []	8 * 64	uimsbf
next_start_code()		
if (nextbits() == extension_start_code) {		
extension_start_code	32	bslbf
while (nextbits () != '0000 0000 0000 0000 0000 0001') {		
sequence_extension_data	8	
}		
next_start_code()		
}		
if (nextbits() == user_data_start_code) {		
user_data_start_code	32	bslbf
while (nextbits() != '0000 0000 0000 0000 0000 0001') {		
user_data	8	
}		
next_start_code()		
}		
}		