
Information technology — Coding of
audio-visual objects —

Part 10:
Advanced Video Coding

*Technologies de l'information — Codage des objets audiovisuels —
Partie 10: Codage visuel avancé*
iTeh STANDARD PREVIEW
(standards.iteh.ai)

SIST ISO/IEC 14496-10:2005

<https://standards.iteh.ai/catalog/standards/sist/3b8f0550-4e17-4862-9b5f-fe8337a55038/sist-iso-iec-14496-10-2005>

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

SIST ISO/IEC 14496-10:2005

<https://standards.iteh.ai/catalog/standards/sist/3b8f0550-4e17-4862-9b5f-fe8337a55038/sist-iso-iec-14496-10-2005>

© ISO/IEC 2004

All rights reserved. Unless otherwise specified, 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 either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword	viii
0 Introduction	ix
0.1 Prologue	ix
0.2 Purpose	ix
0.3 Applications	ix
0.4 Profiles and levels	ix
0.5 Overview of the design characteristics	x
0.6 How to read this specification	xi
1 Scope	1
2 Normative references	1
3 Definitions	1
4 Abbreviations	9
5 Conventions	9
5.1 Arithmetic operators	9
5.2 Logical operators	10
5.3 Relational operators	10
5.4 Bit-wise operators	10
5.5 Assignment operators	11
5.6 Range notation	11
5.7 Mathematical functions	11
5.8 Variables, syntax elements, and tables	12
5.9 Text description of logical operations	13
5.10 Processes	14
6 Source, coded, decoded and output data formats, scanning processes, and neighbouring relationships	14
6.1 Bitstream formats	14
6.2 Source, decoded, and output picture formats	14
6.3 Spatial subdivision of pictures and slices	16
6.4 Inverse scanning processes and derivation processes for neighbours	17
7 Syntax and semantics	28
7.1 Method of describing syntax in tabular form	28
7.2 Specification of syntax functions, categories, and descriptors	29
7.3 Syntax in tabular form	30
7.4 Semantics	47
8 Decoding process	82
8.1 NAL unit decoding process	83
8.2 Slice decoding process	84
8.3 Intra prediction process	103
8.4 Inter prediction process	115
8.5 Transform coefficient decoding process and picture construction process prior to deblocking filter process ..	137
8.6 Decoding process for P macroblocks in SP slices or SI macroblocks	144
8.7 Deblocking filter process	149
9 Parsing process	159
9.1 Parsing process for Exp-Golomb codes	159
9.2 CAVLC parsing process for transform coefficient levels	162
9.3 CABAC parsing process for slice data	169
Annex A Profiles and levels	208
A.1 Requirements on video decoder capability	208
A.2 Profiles	208
A.3 Levels	209

Annex B Byte stream format	216
B.1 Byte stream NAL unit syntax and semantics	216
B.2 Byte stream NAL unit decoding process	217
B.3 Decoder byte-alignment recovery (informative)	217
Annex C Hypothetical reference decoder	219
C.1 Operation of coded picture buffer (CPB).....	221
C.2 Operation of the decoded picture buffer (DPB).....	223
C.3 Bitstream conformance	224
C.4 Decoder conformance	226
Annex D Supplemental enhancement information	230
D.1 SEI payload syntax	231
D.2 SEI payload semantics	237
Annex E Video usability information	255
E.1 VUI syntax.....	255
E.2 VUI semantics	257
Annex F Patent Rights	267

LIST OF FIGURES

Figure 6-1 – Nominal vertical and horizontal locations of 4:2:0 luma and chroma samples in a frame	15
Figure 6-2 – Nominal vertical and horizontal sampling locations of samples top and bottom fields	16
Figure 6-3 – A picture with 11 by 9 macroblocks that is partitioned into two slices	17
Figure 6-4 – Partitioning of the decoded frame into macroblock pairs	17
Figure 6-5 – Macroblock partitions, sub-macroblock partitions, macroblock partition scans, and sub-macroblock partition scans.	18
Figure 6-6 – Scan for 4x4 luma blocks.....	19
Figure 6-7 – Neighbouring macroblocks for a given macroblock	20
Figure 6-8 – Neighbouring macroblocks for a given macroblock in MBAFF frames.....	21
Figure 6-9 – Determination of the neighbouring macroblock, blocks, and partitions (informative)	22
Figure 7-1 – The structure of an access unit not containing any NAL units with nal_unit_type equal to 0, 7, 8, or in the range of 12 to 31, inclusive	53
Figure 8-1 – Intra_4x4 prediction mode directions (informative)	104
Figure 8-2 –Example for temporal direct-mode motion vector inference (informative)	125
Figure 8-3 – Directional segmentation prediction (informative).....	126
Figure 8-4 – Integer samples (shaded blocks with upper-case letters) and fractional sample positions (un-shaded blocks with lower-case letters) for quarter sample luma interpolation.	131
Figure 8-5 – Fractional sample position dependent variables in chroma interpolation and surrounding integer position samples A, B, C, and D.	133
Figure 8-6 – Assignment of the indices of dcY to luma4x4BlkIdx.	138
Figure 8-7 – Assignment of the indices of dcC to chroma4x4BlkIdx.	139
Figure 8-8 – a) Zig-zag scan. b) Field scan	140
Figure 8-9 – Boundaries in a macroblock to be filtered (luma boundaries shown with solid lines and chroma boundaries shown with dashed lines).....	150
Figure 8-10 – Convention for describing samples across a 4x4 block horizontal or vertical boundary	153
Figure 9-1 – Illustration of CABAC parsing process for a syntax element SE (informative)	170
Figure 9-2 – Overview of the arithmetic decoding process for a single bin (informative).....	197

Figure 9-3 – Flowchart for decoding a decision.....	198
Figure 9-4 – Flowchart of renormalization	200
Figure 9-5 – Flowchart of bypass decoding process	201
Figure 9-6 – Flowchart of decoding a decision before termination	202
Figure 9-7 – Flowchart for encoding a decision.....	203
Figure 9-8 – Flowchart of renormalization in the encoder.....	204
Figure 9-9 – Flowchart of PutBit(B).....	204
Figure 9-10 – Flowchart of encoding bypass	205
Figure 9-11 – Flowchart of encoding a decision before termination	206
Figure 9-12 – Flowchart of flushing at termination	206
Figure C-1 – Structure of byte streams and NAL unit streams for HRD conformance checks.....	219
Figure C-2 – HRD buffer model	220
Figure E-1 – Location of chroma samples for top and bottom fields as a function of chroma_sample_loc_type_top_field and chroma_sample_loc_type_bottom_field	262

LIST OF TABLES

Table 6-1 – ChromaFormatFactor values.....	15
Table 6-2 – Specification of input and output assignments for subclauses 6.4.7.1 to 6.4.7.5	21
Table 6-3 – Specification of mbAddrN.....	25
Table 6-4 – Specification of mbAddrN and yM	27
Table 7-1 – NAL unit type codes.....	48
Table 7-2 – Meaning of primary_pic_type.....	59
Table 7-3 – Name association to slice_type.....	62
Table 7-4 – reordering_of_pic_nums_idc operations for reordering of reference picture lists.....	68
Table 7-5 – Interpretation of adaptive_ref_pic_marking_mode_flag	69
Table 7-6 – Memory management control operation (memory_management_control_operation) values	70
Table 7-7 – Allowed collective macroblock types for slice_type	72
Table 7-8 – Macroblock types for I slices.....	73
Table 7-9 – Macroblock type with value 0 for SI slices	74
Table 7-10 – Macroblock type values 0 to 4 for P and SP slices	75
Table 7-11 – Macroblock type values 0 to 22 for B slices.....	76
Table 7-12 – Specification of CodedBlockPatternChroma values.....	78
Table 7-13 – Relationship between intra_chroma_pred_mode and spatial prediction modes	78
Table 7-14 – Sub-macroblock types in P macroblocks	79
Table 7-15 – Sub-macroblock types in B macroblocks	80
Table 8-1 – Refined slice group map type	88
Table 8-2 – Specification of Intra4x4PredMode[luma4x4BlkIdx] and associated names	104
Table 8-3 – Specification of Intra16x16PredMode and associated names.....	109
Table 8-4 – Specification of Intra chroma prediction modes and associated names.....	111
Table 8-5 – Specification of the variable colPic	119

Table 8-6 – Specification of PicCodingStruct(X).....	119
Table 8-7 – Specification of mbAddrCol, yM, and vertMvScale.....	120
Table 8-8 – Assignment of prediction utilization flags.....	122
Table 8-9 – Derivation of the vertical component of the chroma vector in field coding mode.....	128
Table 8-10 – Differential full-sample luma locations.....	131
Table 8-11 – Assignment of the luma prediction sample predPartLX _L [x _L , y _L].....	133
Table 8-12 – Specification of mapping of idx to c _{ij} for zig-zag and field scan.....	140
Table 8-13 – Specification of QP _C as a function of qP ₁	140
Table 8-14 – Derivation of indexA and indexB from offset dependent threshold variables α and β	156
Table 8-15 – Value of filter clipping variable t _{C0} as a function of indexA and bS.....	157
Table 9-1 – Bit strings with “prefix” and “suffix” bits and assignment to codeNum ranges (informative).....	159
Table 9-2 – Exp-Golomb bit strings and codeNum in explicit form and used as ue(v) (informative).....	160
Table 9-3 – Assignment of syntax element to codeNum for signed Exp-Golomb coded syntax elements se(v).....	160
Table 9-4 – Assignment of codeNum to values of coded_block_pattern for macroblock prediction modes.....	161
Table 9-5 – coeff_token mapping to TotalCoeff(coeff_token) and TrailingOnes(coeff_token).....	164
Table 9-6 – Codeword table for level_prefix.....	166
Table 9-7 – total_zeros tables for 4x4 blocks with TotalCoeff(coeff_token) 1 to 7.....	167
Table 9-8 – total_zeros tables for 4x4 blocks with TotalCoeff(coeff_token) 8 to 15.....	168
Table 9-9 – total_zeros tables for chroma DC 2x2 blocks.....	168
Table 9-10 – Tables for run_before.....	168
Table 9-11 – Association of ctxIdx and syntax elements for each slice type in the initialisation process.....	171
Table 9-12 – Values of variables m and n for ctxIdx from 0 to 10.....	172
Table 9-13 – Values of variables m and n for ctxIdx from 11 to 23.....	172
Table 9-14 – Values of variables m and n for ctxIdx from 24 to 39.....	172
Table 9-15 – Values of variables m and n for ctxIdx from 40 to 53.....	173
Table 9-16 – Values of variables m and n for ctxIdx from 54 to 59.....	173
Table 9-17 – Values of variables m and n for ctxIdx from 60 to 69.....	173
Table 9-18 – Values of variables m and n for ctxIdx from 70 to 104.....	174
Table 9-19 – Values of variables m and n for ctxIdx from 105 to 165.....	175
Table 9-20 – Values of variables m and n for ctxIdx from 166 to 226.....	176
Table 9-21 – Values of variables m and n for ctxIdx from 227 to 275.....	177
Table 9-22 – Values of variables m and n for ctxIdx from 277 to 337.....	178
Table 9-23 – Values of variables m and n for ctxIdx from 338 to 398.....	179
Table 9-24 – Syntax elements and associated types of binarization, maxBinIdxCtx, and ctxIdxOffset.....	181
Table 9-25 – Bin string of the unary binarization (informative).....	182
Table 9-26 – Binarization for macroblock types in I slices.....	184
Table 9-27 – Binarization for macroblock types in P, SP, and B slices.....	185
Table 9-28 – Binarization for sub-macroblock types in P, SP, and B slices.....	186
Table 9-29 – Assignment of ctxIdxInc to binIdx for all ctxIdxOffset values except those related to the syntax elements coded_block_flag, significant_coeff_flag, last_significant_coeff_flag, and coeff_abs_level_minus1.....	188

Table 9-30 – Assignment of ctxIdxBlockCatOffset to ctxBlockCat for syntax elements coded_block_flag, significant_coeff_flag, last_significant_coeff_flag, and coeff_abs_level_minus1	189
Table 9-31 – Specification of ctxIdxInc for specific values of ctxIdxOffset and binIdx	195
Table 9-32 – Specification of ctxBlockCat for the different blocks.....	196
Table 9-33 – Specification of rangeTabLPS depending on pStateIdx and qCodIRangeIdx	199
Table 9-34 – State transition table	200
Table A-1 – Level limits	211
Table A-2 – Baseline profile level limits	212
Table A-3 – Main profile level limits	213
Table A-4 – Extended profile level limits.....	213
Table A-5 – Maximum frame rates (frames per second) for some example frame sizes	214
Table D-1 – Interpretation of pic_struct.....	239
Table D-2 – Mapping of ct_type to source picture scan	240
Table D-3 – Definition of counting_type values.....	240
Table D-4 – scene_transition_type values.	247
Table E-1 – Meaning of sample aspect ratio indicator.....	257
Table E-2 – Meaning of video_format.....	258
Table E-3 – Colour primaries.....	259
Table E-4 – Transfer characteristics	260
Table E-5 – Matrix coefficients	261
Table E-6 – Divisor for computation of $\Delta_{f,dpb}(n)$	263

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. 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.

ISO/IEC 14496-10 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This second edition cancels and replaces the first edition (ISO/IEC 14496-10:2003) which has been technically revised.

This part of ISO/IEC 14496 is technically aligned with ITU-T Rec. H.264 but is not published as identical text.

ISO/IEC 14496 consists of the following parts, under the general title *Information technology — Coding of audio-visual objects*:

- Part 1: *Systems*
- Part 2: *Visual*
- Part 3: *Audio*
- Part 4: *Conformance testing*
- Part 5: *Reference software*
- Part 6: *Delivery Multimedia Integration Framework (DMIF)*
- Part 7: *Optimized reference software for coding of audio-visual objects*
- Part 8: *Carriage of ISO/IEC 14496 contents over IP networks*
- Part 9: *Reference hardware description*
- Part 10: *Advanced Video Coding*
- Part 11: *Scene description and application engine*
- Part 12: *ISO base media file format*
- Part 13: *Intellectual Property Management and Protection (IPMP) extensions*
- Part 14: *MP4 file format*
- Part 15: *Advanced Video Coding (AVC) file format*
- Part 16: *Animation Framework eXtension (AFX)*
- Part 17: *Streaming text format*
- Part 18: *Font compression and streaming*
- Part 19: *Synthesized texture stream*

ITeH STANDARD PREVIEW
(standards.iteh.ai)
SIST ISO/IEC 14496-10:2005
<https://standards.iteh.ai/catalog/standards/sist/3b8f0550-4e17-4862-9b5f-fe8337a55038/sist-iso-iec-14496-10-2005>

0 Introduction

This clause does not form an integral part of this Recommendation | International Standard.

0.1 Prologue

This subclause does not form an integral part of this Recommendation | International Standard.

As the costs for both processing power and memory have reduced, network support for coded video data has diversified, and advances in video coding technology have progressed, the need has arisen for an industry standard for compressed video representation with substantially increased coding efficiency and enhanced robustness to network environments. Toward these ends the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) formed a Joint Video Team (JVT) in 2001 for development of a new Recommendation | International Standard.

0.2 Purpose

This subclause does not form an integral part of this Recommendation | International Standard.

This Recommendation | International Standard was developed in response to the growing need for higher compression of moving pictures for various applications such as videoconferencing, digital storage media, television broadcasting, internet streaming, and communication. It is also designed to enable the use of the coded video representation in a flexible manner for a wide variety of network environments. The use of this Recommendation | International Standard allows motion video to be manipulated as a form of computer data and to be stored on various storage media, transmitted and received over existing and future networks and distributed on existing and future broadcasting channels.

0.3 Applications

This subclause does not form an integral part of this Recommendation | International Standard.

This Recommendation | International Standard is designed to cover a broad range of applications for video content including but not limited to the following:

CATV	Cable TV on optical networks, copper, etc.
DBS	Direct broadcast satellite video services
DSL	Digital subscriber line video services
DTTB	Digital terrestrial television broadcasting
ISM	Interactive storage media (optical disks, etc.)
MMM	Multimedia mailing
MSPN	Multimedia services over packet networks
RTC	Real-time conversational services (videoconferencing, videophone, etc.)
RVS	Remote video surveillance
SSM	Serial storage media (digital VTR, etc.)

0.4 Profiles and levels

This subclause does not form an integral part of this Recommendation | International Standard.

This Recommendation | International Standard is designed to be generic in the sense that it serves a wide range of applications, bit rates, resolutions, qualities, and services. Applications should cover, among other things, digital storage media, television broadcasting and real-time communications. In the course of creating this Specification, various requirements from typical applications have been considered, necessary algorithmic elements have been developed, and these have been integrated into a single syntax. Hence, this Specification will facilitate video data interchange among different applications.

Considering the practicality of implementing the full syntax of this Specification, however, a limited number of subsets of the syntax are also stipulated by means of "profiles" and "levels". These and other related terms are formally defined in clause 3.

A "profile" is a subset of the entire bitstream syntax that is specified by this Recommendation | International Standard. Within the bounds imposed by the syntax of a given profile it is still possible to require a very large variation in the performance of encoders and decoders depending upon the values taken by syntax elements in the bitstream such as the specified size of the decoded pictures. In many applications, it is currently neither practical nor economic to implement a decoder capable of dealing with all hypothetical uses of the syntax within a particular profile.

In order to deal with this problem, "levels" are specified within each profile. A level is a specified set of constraints imposed on values of the syntax elements in the bitstream. These constraints may be simple limits on values. Alternatively they may take the form of constraints on arithmetic combinations of values (e.g. picture width multiplied by picture height multiplied by number of pictures decoded per second).

Coded video content conforming to this Recommendation | International Standard uses a common syntax. In order to achieve a subset of the complete syntax, flags, parameters, and other syntax elements are included in the bitstream that signal the presence or absence of syntactic elements that occur later in the bitstream.

0.5 Overview of the design characteristics

This subclause does not form an integral part of this Recommendation | International Standard.

The coded representation specified in the syntax is designed to enable a high compression capability for a desired image quality. The algorithm is not lossless, as the exact source sample values are typically not preserved through the encoding and decoding processes. A number of techniques may be used to achieve highly efficient compression. Encoding algorithms (not specified in this Recommendation | International Standard) may select between inter and intra coding for block-shaped regions of each picture. Inter coding uses motion vectors for block-based inter prediction to exploit temporal statistical dependencies between different pictures. Intra coding uses various spatial prediction modes to exploit spatial statistical dependencies in the source signal for a single picture. Motion vectors and intra prediction modes may be specified for a variety of block sizes in the picture. The prediction residual is then further compressed using a transform to remove spatial correlation inside the transform block before it is quantised, producing an irreversible process that typically discards less important visual information while forming a close approximation to the source samples. Finally, the motion vectors or intra prediction modes are combined with the quantised transform coefficient information and encoded using either variable length codes or arithmetic coding.

0.5.1 Predictive coding

This subclause does not form an integral part of this Recommendation | International Standard.

Because of the conflicting requirements of random access and highly efficient compression, two main coding types are specified. Intra coding is done without reference to other pictures. Intra coding may provide access points to the coded sequence where decoding can begin and continue correctly, but typically also shows only moderate compression efficiency. Inter coding (predictive or bi-predictive) is more efficient using inter prediction of each block of sample values from some previously decoded picture selected by the encoder. In contrast to some other video coding standards, pictures coded using bi-predictive inter prediction may also be used as references for inter coding of other pictures.

The application of the three coding types to pictures in a sequence is flexible, and the order of the decoding process is generally not the same as the order of the source picture capture process in the encoder or the output order from the decoder for display. The choice is left to the encoder and will depend on the requirements of the application. The decoding order is specified such that the decoding of pictures that use inter-picture prediction follows later in decoding order than other pictures that are referenced in the decoding process.

0.5.2 Coding of progressive and interlaced video

This subclause does not form an integral part of this Recommendation | International Standard.

This Recommendation | International Standard specifies a syntax and decoding process for video that originated in either progressive-scan or interlaced-scan form, which may be mixed together in the same sequence. The two fields of an interlaced frame are separated in capture time while the two fields of a progressive frame share the same capture time. Each field may be coded separately or the two fields may be coded together as a frame. Progressive frames are typically coded as a frame. For interlaced video, the encoder can choose between frame coding and field coding. Frame coding or field coding can be adaptively selected on a picture-by-picture basis and also on a more localized basis within a coded

frame. Frame coding is typically preferred when the video scene contains significant detail with limited motion. Field coding typically works better when there is fast picture-to-picture motion.

0.5.3 Picture partitioning into macroblocks and smaller partitions

This subclause does not form an integral part of this Recommendation | International Standard.

As in previous video coding Recommendations and International Standards, a macroblock, consisting of a 16x16 block of luma samples and two corresponding blocks of chroma samples, is used as the basic processing unit of the video decoding process.

A macroblock can be further partitioned for inter prediction. The selection of the size of inter prediction partitions is a result of a trade-off between the coding gain provided by using motion compensation with smaller blocks and the quantity of data needed to represent the data for motion compensation. In this Recommendation | International Standard the inter prediction process can form segmentations for motion representation as small as 4x4 luma samples in size, using motion vector accuracy of one-quarter of the luma sample grid spacing displacement. The process for inter prediction of a sample block can also involve the selection of the picture to be used as the reference picture from a number of stored previously-decoded pictures. Motion vectors are encoded differentially with respect to predicted values formed from nearby encoded motion vectors.

Typically, the encoder calculates appropriate motion vectors and other data elements represented in the video data stream. This motion estimation process in the encoder and the selection of whether to use inter prediction for the representation of each region of the video content is not specified in this Recommendation | International Standard.

0.5.4 Spatial redundancy reduction

This subclause does not form an integral part of this Recommendation | International Standard.

Both source pictures and prediction residuals have high spatial redundancy. This Recommendation | International Standard is based on the use of a block-based transform method for spatial redundancy removal. After inter prediction from previously-decoded samples in other pictures or spatial-based prediction from previously-decoded samples within the current picture, the resulting prediction residual is split into 4x4 blocks. These are converted into the transform domain where they are quantised. After quantisation many of the transform coefficients are zero or have low amplitude and can thus be represented with a small amount of encoded data. The processes of transformation and quantisation in the encoder are not specified in this Recommendation | International Standard.

0.6 How to read this specification

This subclause does not form an integral part of this Recommendation | International Standard.

It is suggested that the reader starts with clause 1 (Scope) and moves on to clause 3 (Definitions). Clause 6 should be read for the geometrical relationship of the source, input, and output of the decoder. Clause 7 (Syntax and semantics) specifies the order to parse syntax elements from the bitstream. See subclauses 7.1-7.3 for syntactical order and see subclause 7.4 for semantics; i.e., the scope, restrictions, and conditions that are imposed on the syntax elements. The actual parsing for most syntax elements is specified in clause 9 (Parsing process). Finally, clause 8 (Decoding process) specifies how the syntax elements are mapped into decoded samples. Throughout reading this specification, the reader should refer to clauses 2 (Normative references), 4 (Abbreviations), and 5 (Conventions) as needed. Annexes A through E also form an integral part of this Recommendation | International Standard.

Annex A defines three profiles (Baseline, Main, and Extended), each being tailored to certain application domains, and defines the so-called levels of the profiles. Annex B specifies syntax and semantics of a byte stream format for delivery of coded video as an ordered stream of bytes. Annex C specifies the hypothetical reference decoder and its use to check bitstream and decoder conformance. Annex D specifies syntax and semantics for supplemental enhancement information message payloads. Finally, Annex E specifies syntax and semantics of the video usability information parameters of the sequence parameter set.

Throughout this specification, statements appearing with the preamble "NOTE -" are informative and are not an integral part of this Recommendation | International Standard.

iTeh STANDARD PREVIEW **(standards.iteh.ai)**

SIST ISO/IEC 14496-10:2005

<https://standards.iteh.ai/catalog/standards/sist/3b8f0550-4e17-4862-9b5f-fe8337a55038/sist-iso-iec-14496-10-2005>

Information technology — Coding of audio-visual objects —

Part 10:

Advanced Video Coding

1 Scope

This document specifies ITU-T Recommendation H.264 | ISO/IEC International Standard ISO/IEC 14496-10 video coding.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

- ITU-T Recommendation T.35 (2000), *Procedure for the allocation of ITU-T defined codes for non-standard facilities*
- ISO/IEC 11578:1996, *Information technology — Open Systems Interconnection — Remote Procedure Call (RPC)*
- ISO/CIE 10527:1991, *CIE standard colorimetric observers*

<https://standards.iteh.ai/catalog/standards/sist/3b8f0550-4e17-4862-9b5f-fe8337a55038/sist-iso-iec-14496-10-2005>

3 Definitions

For the purposes of this Recommendation | International Standard, the following definitions apply.

- 3.1 access unit:** A set of *NAL units* always containing exactly one *primary coded picture*. In addition to the *primary coded picture*, an access unit may also contain one or more *redundant coded pictures* or other *NAL units* not containing *slices* or *slice data partitions* of a *coded picture*. The decoding of an access unit always results in a *decoded picture*.
- 3.2 AC transform coefficient:** Any *transform coefficient* for which the *frequency index* in one or both dimensions is non-zero.
- 3.3 adaptive binary arithmetic decoding process:** An entropy *decoding process* that derives the values of *bins* from a *bitstream* produced by an *adaptive binary arithmetic encoding process*.
- 3.4 adaptive binary arithmetic encoding process:** An entropy *encoding process*, not normatively specified in this Recommendation | International Standard, that codes a sequence of *bins* and produces a *bitstream* that can be decoded using the *adaptive binary arithmetic decoding process*.
- 3.5 arbitrary slice order:** A *decoding order* of *slices* in which the *macroblock address* of the first *macroblock* of some *slice* of a *picture* may be smaller than the *macroblock address* of the first *macroblock* of some other preceding *slice* of the same *coded picture*.
- 3.6 B slice:** A *slice* that may be decoded using *intra prediction* from decoded samples within the same *slice* or *inter prediction* from previously-decoded *reference pictures*, using at most two *motion vectors* and *reference indices* to *predict* the sample values of each *block*.
- 3.7 bin:** One bit of a *bin string*.
- 3.8 binarization:** A set of *bin strings* for all possible values of a *syntax element*.

- 3.9 binarization process:** A unique mapping process of all possible values of a *syntax element* onto a set of *bin strings*.
- 3.10 bin string:** A string of *bins*. A bin string is an intermediate binary representation of values of *syntax elements* from the *binarization* of the *syntax element*.
- 3.11 bi-predictive slice:** See *B slice*.
- 3.12 bitstream:** A sequence of bits that forms the representation of *coded pictures* and associated data forming one or more *coded video sequences*. Bitstream is a collective term used to refer either to a *NAL unit stream* or a *byte stream*.
- 3.13 block:** An MxN (M-column by N-row) array of samples, or an MxN array of *transform coefficients*.
- 3.14 bottom field:** One of two *fields* that comprise a *frame*. Each row of a *bottom field* is spatially located immediately below a corresponding row of a *top field*.
- 3.15 bottom macroblock (of a macroblock pair):** The *macroblock* within a *macroblock pair* that contains the samples in the bottom row of samples for the *macroblock pair*. For a *field macroblock pair*, the bottom macroblock represents the samples from the region of the *bottom field* of the *frame* that lie within the spatial region of the *macroblock pair*. For a *frame macroblock pair*, the bottom macroblock represents the samples of the *frame* that lie within the bottom half of the spatial region of the *macroblock pair*.
- 3.16 broken link:** A location in a *bitstream* at which it is indicated that some subsequent *pictures* in *decoding order* may contain serious visual artefacts due to unspecified operations performed in the generation of the *bitstream*.
- 3.17 byte:** A sequence of 8 bits, written and read with the most significant bit on the left and the least significant bit on the right. When represented in a sequence of data bits, the most significant bit of a byte is first.
- 3.18 byte-aligned:** A position in a *bitstream* is byte-aligned when the position is an integer multiple of 8 bits from the position of the first bit in the *bitstream*. A bit or *byte* or *syntax element* is said to be byte-aligned when the position at which it appears in a *bitstream* is byte-aligned.
- 3.19 byte stream:** An encapsulation of a *NAL unit stream* containing *start code prefixes* and *NAL units* as specified in Annex B.
- 3.19.1 can:** A term used to refer to behaviour that is allowed, but not necessarily required.
- 3.20 category:** A number associated with each *syntax element*. The category is used to specify the allocation of *syntax elements* to *NAL units* for *slice data partitioning*. It may also be used in a manner determined by the application to refer to classes of *syntax elements* in a manner not specified in this Recommendation | International Standard.
- 3.21 chroma:** An adjective specifying that a sample array or single sample is representing one of the two colour difference signals related to the primary colours. The symbols used for a chroma array or sample are Cb and Cr.
- NOTE - The term chroma is used rather than the term chrominance in order to avoid the implication of the use of linear light transfer characteristics that is often associated with the term chrominance.
- 3.22 coded field:** A *coded representation* of a *field*.
- 3.23 coded frame:** A *coded representation* of a *frame*.
- 3.24 coded picture:** A *coded representation* of a *picture*. A coded picture may be either a *coded field* or a *coded frame*. Coded picture is a collective term referring to a *primary coded picture* or a *redundant coded picture*, but not to both together.
- 3.25 coded picture buffer (CPB):** A first-in first-out buffer containing *access units* in *decoding order* specified in the *hypothetical reference decoder* in Annex C.
- 3.26 coded representation:** A data element as represented in its coded form.
- 3.27 coded video sequence:** A sequence of *access units* that consists, in decoding order, of an *IDR access unit* followed by zero or more non-IDR *access units* including all subsequent *access units* up to but not including any subsequent *IDR access unit*.
- 3.28 component:** An array or single sample from one of the three arrays (*luma* and two *chroma*) that make up a *field* or *frame*.

- 3.29 complementary field pair:** A collective term for a *complementary reference field pair* or a *complementary non-reference field pair*.
- 3.30 complementary non-reference field pair:** Two *non-reference fields* that are in consecutive *access units* in *decoding order* as two *coded fields* of opposite parity where the first *field* is not already a paired *field*.
- 3.31 complementary reference field pair:** Two *reference fields* that are in consecutive *access units* in *decoding order* as two *coded fields* and share the same value of the *frame_num syntax element*, where the second *field* in *decoding order* is not an *IDR picture* and does not include a *memory_management_control_operation syntax element* equal to 5.
- 3.32 context variable:** A variable specified for the *adaptive binary arithmetic decoding process* of a *bin* by an equation containing recently decoded *bins*.
- 3.33 DC transform coefficient:** A *transform coefficient* for which the *frequency index* is zero in all dimensions.
- 3.34 decoded picture:** A *decoded picture* is derived by decoding a *coded picture*. A *decoded picture* is either a *decoded frame*, or a *decoded field*. A *decoded field* is either a *decoded top field* or a *decoded bottom field*.
- 3.35 decoded picture buffer (DPB):** A buffer holding *decoded pictures* for reference, output reordering, or output delay specified for the *hypothetical reference decoder* in Annex C.
- 3.36 decoder:** An embodiment of a *decoding process*.
- 3.37 decoding order:** The order in which *syntax elements* are processed by the *decoding process*.
- 3.38 decoding process:** The process specified in this Recommendation | International Standard that reads a *bitstream* and derives *decoded pictures* from it.
- 3.39 direct prediction:** An *inter prediction* for a *block* for which no *motion vector* is decoded. Two *direct prediction* modes are specified that are referred to as *spatial direct prediction* and *temporal prediction* mode.
- 3.40 decoder under test (DUT):** A *decoder* that is tested for conformance to this Recommendation | International Standard by operating the *hypothetical stream scheduler* to deliver a conforming *bitstream* to the *decoder* and to the *hypothetical reference decoder* and comparing the values and timing of the output of the two *decoders*.
- 3.41 emulation prevention byte:** A *byte* equal to 0x03 that may be present within a *NAL unit*. The presence of emulation prevention bytes ensures that no sequence of consecutive *byte-aligned bytes* in the *NAL unit* contains a *start code prefix*.
- 3.42 encoder:** An embodiment of an *encoding process*.
- 3.43 encoding process:** A process, not specified in this Recommendation | International Standard, that produces a *bitstream* conforming to this Recommendation | International Standard.
- 3.44 field:** An assembly of alternate rows of a *frame*. A *frame* is composed of two *fields*, a *top field* and a *bottom field*.
- 3.45 field macroblock:** A *macroblock* containing samples from a single *field*. All *macroblocks* of a *coded field* are *field macroblocks*. When *macroblock-adaptive frame/field decoding* is in use, some *macroblocks* of a *coded frame* may be *field macroblocks*.
- 3.46 field macroblock pair:** A *macroblock pair* decoded as two *field macroblocks*.
- 3.47 field scan:** A specific sequential ordering of *transform coefficients* that differs from the *zig-zag scan* by scanning columns more rapidly than rows. *Field scan* is used for *transform coefficients* in *field macroblocks*.
- 3.48 flag:** A variable that can take one of the two possible values 0 and 1.
- 3.49 frame:** A *frame* contains an array of *luma* samples and two corresponding arrays of *chroma* samples. A *frame* consists of two *fields*, a *top field* and a *bottom field*.
- 3.50 frame macroblock:** A *macroblock* representing samples from the two *fields* of a *coded frame*. When *macroblock-adaptive frame/field decoding* is not in use, all *macroblocks* of a *coded frame* are *frame macroblocks*. When *macroblock-adaptive frame/field decoding* is in use, some *macroblocks* of a *coded frame* may be *frame macroblocks*.
- 3.51 frame macroblock pair:** A *macroblock pair* decoded as two *frame macroblocks*.