

I n t e r n a t i o n a l T e l e c o m m u n i c a t i o n U n i o n

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

H.263

(01/2005)

SERIES H: AUDIOVISUAL AND MULTIMEDIA SYSTEMS

Infrastructure of audiovisual services – Coding of moving
video

Video coding for low bit rate communication

ITU-T Recommendation H.263



ITU-T H-SERIES RECOMMENDATIONS
AUDIOVISUAL AND MULTIMEDIA SYSTEMS

CHARACTERISTICS OF VISUAL TELEPHONE SYSTEMS	H.100–H.199
INFRASTRUCTURE OF AUDIOVISUAL SERVICES	
General	H.200–H.219
Transmission multiplexing and synchronization	H.220–H.229
Systems aspects	H.230–H.239
Communication procedures	H.240–H.259
Coding of moving video	H.260–H.279
Related systems aspects	H.280–H.299
Systems and terminal equipment for audiovisual services	H.300–H.349
Directory services architecture for audiovisual and multimedia services	H.350–H.359
Quality of service architecture for audiovisual and multimedia services	H.360–H.369
Supplementary services for multimedia	H.450–H.499
MOBILITY AND COLLABORATION PROCEDURES	
Overview of Mobility and Collaboration, definitions, protocols and procedures	H.500–H.509
Mobility for H-Series multimedia systems and services	H.510–H.519
Mobile multimedia collaboration applications and services	H.520–H.529
Security for mobile multimedia systems and services	H.530–H.539
Security for mobile multimedia collaboration applications and services	H.540–H.549
Mobility interworking procedures	H.550–H.559
Mobile multimedia collaboration inter-working procedures	H.560–H.569
BROADBAND AND TRIPLE-PLAY MULTIMEDIA SERVICES	
Broadband multimedia services over VDSL	H.610–H.619

For further details, please refer to the list of ITU-T Recommendations.

ITU-T Recommendation H.263

Video coding for low bit rate communication

Summary

This Recommendation specifies a coded representation that can be used for compressing the moving picture component of audio-visual services at low bit rates. The basic configuration of the video source coding algorithm is based on ITU-T Rec. H.261 and is a hybrid of inter-picture prediction to utilize temporal redundancy and transform coding of the remaining signal to reduce spatial redundancy. The source coder can operate on five standardized video source formats: sub-QCIF, QCIF, CIF, 4CIF and 16CIF, and can also operate using a broad range of custom video formats.

The decoder has motion compensation capability, allowing optional incorporation of this technique in the coder. Half pixel precision is used for the motion compensation, as opposed to ITU-T Rec. H.261 where full pixel precision and a loopfilter are used. Variable length coding is used for the symbols to be transmitted.

In addition to the basic video source coding algorithm, eighteen negotiable coding options are included for improved compression performance and the support of additional capabilities. Additional supplemental information may also be included in the bitstream for enhanced display capability and for external usage.

This third edition of H.263 integrates Annexes U, V, W and X, which were approved and published separately before, with a number of corrections and clarifications:

- correction of Figure 8;
- clarification in Table 1 regarding BPPmaxKb table and Picture Padding;
- clarification in clause 5.3.2 regarding macroblock stuffing preceding a start code;
- clarification regarding interaction between H.263 Annex J and IDCT rounding error;
- clarification in Annex N regarding parsability of the GN/MBA field in BCM;
- clarification in Annex O regarding direct bi-dir predicted MBs and picture extrapolation;
- clarification on the use of B pictures with intra reference pictures in Annex O;
- clarification on the use of Annex P with Annex N;
- correction of Figure U.7 in Annex U.

Source

ITU-T Recommendation H.263 was approved on 13 January 2005 by ITU-T Study Group 16 (2005-2008) under the ITU-T Recommendation A.8 procedure.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g. interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

© ITU 2005

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

CONTENTS

	Page
1 Scope	1
2 References.....	1
2.1 Normative references.....	1
2.2 Informative references.....	1
3 Brief specification.....	2
3.1 Video input and output	2
3.2 Digital output and input.....	2
3.3 Sampling frequency.....	2
3.4 Source coding algorithm.....	2
3.5 Bit rate	5
3.6 Buffering.....	6
3.7 Symmetry of transmission.....	6
3.8 Error handling.....	6
3.9 Multipoint operation.....	6
4 Source coder	7
4.1 Source format	7
4.2 Video source coding algorithm	9
4.3 Coding control.....	13
4.4 Forced updating.....	13
4.5 Byte alignment of start codes	13
5 Syntax and semantics.....	13
5.1 Picture layer.....	22
5.2 Group of Blocks Layer	32
5.3 Macroblock layer.....	33
5.4 Block layer.....	40
6 Decoding process.....	44
6.1 Motion compensation	44
6.2 Coefficients decoding.....	46
6.3 Reconstruction of blocks	48
Annex A – Inverse transform accuracy specification.....	48
Annex B – Hypothetical Reference Decoder.....	49
Annex C – Considerations for multipoint.....	51
C.1 Freeze picture request.....	51
C.2 Fast update request	51
C.3 Freeze picture release	51
C.4 Continuous Presence Multipoint and Video Multiplexing (CPM).....	51

Annex D – Unrestricted Motion Vector mode.....	53
D.1 Motion vectors over picture boundaries.....	53
D.2 Extension of the motion vector range.....	54
Annex E – Syntax-based Arithmetic Coding mode.....	56
E.1 Introduction.....	56
E.2 Specification of SAC encoder.....	57
E.3 Specification of SAC decoder.....	58
E.4 Syntax.....	58
E.5 PSC_FIFO.....	59
E.6 Header layer symbols.....	59
E.7 Macroblock and Block layer symbols.....	60
E.8 SAC models.....	60
Annex F – Advanced Prediction mode.....	64
F.1 Introduction.....	64
F.2 Four motion vectors per macroblock.....	64
F.3 Overlapped motion compensation for luminance.....	65
Annex G – PB-frames mode.....	68
G.1 Introduction.....	68
G.2 PB-frames and INTRA blocks.....	68
G.3 Block layer.....	68
G.4 Calculation of vectors for the B-picture in a PB-frame.....	69
G.5 Prediction of a B-block in a PB-frame.....	69
Annex H – Forward error correction for coded video signal.....	71
H.1 Introduction.....	71
H.2 Error correction framing.....	71
H.3 Error correcting code.....	71
H.4 Relock time for error corrector framing.....	72
Annex I – Advanced INTRA Coding mode.....	72
I.1 Introduction.....	72
I.2 Syntax.....	73
I.3 Decoding process.....	74
Annex J – Deblocking Filter mode.....	80
J.1 Introduction.....	80
J.2 Relation to UMV and AP modes (Annexes D and F).....	81
J.3 Definition of the deblocking edge filter.....	81
Annex K – Slice Structured mode.....	85
K.1 Introduction.....	85
K.2 Structure of slice layer.....	86

Annex L – Supplemental enhancement information	88
L.1 Introduction	88
L.2 PSUPP format.....	88
L.3 Do Nothing	89
L.4 Full-Picture Freeze Request	89
L.5 Partial-Picture Freeze Request	89
L.6 Resizing Partial-Picture Freeze Request	90
L.7 Partial-Picture Freeze-Release Request.....	90
L.8 Full-Picture Snapshot Tag	90
L.9 Partial-Picture Snapshot Tag	91
L.10 Video Time Segment Start Tag	91
L.11 Video Time Segment End Tag	91
L.12 Progressive Refinement Segment Start Tag	91
L.13 Progressive Refinement Segment End Tag	91
L.14 Chroma Keying Information	92
L.15 Extended Function Type.....	94
Annex M – Improved PB-frames mode.....	94
M.1 Introduction	94
M.2 B _{PB} -macroblock prediction modes	95
M.3 Calculation of vectors for bidirectional prediction of a the B-macroblock....	95
M.4 MODB table	95
Annex N – Reference Picture Selection mode.....	96
N.1 Introduction	96
N.2 Video source coding algorithm	97
N.3 Channel for back-channel messages.....	97
N.4 Syntax.....	98
N.5 Decoder process.....	101
Annex O – Temporal, SNR, and Spatial Scalability mode.....	102
O.1 Overview	102
O.2 Transmission order of pictures	106
O.3 Picture layer syntax	107
O.4 Macroblock layer syntax	108
O.5 Motion vector decoding.....	112
O.6 Interpolation filters	112
Annex P – Reference picture resampling.....	115
P.1 Introduction	115
P.2 Syntax	118
P.3 Resampling algorithm	120
P.4 Example of implementation	123
P.5 Factor-of-4 resampling	126

Annex Q – Reduced-Resolution Update mode	130
Q.1 Introduction	130
Q.2 Decoding procedure.....	131
Q.3 Extension of referenced picture.....	133
Q.4 Reconstruction of motion vectors.....	134
Q.5 Enlarged overlapped motion compensation for luminance.....	136
Q.6 Upsampling of the reduced-resolution reconstructed prediction error.....	138
Q.7 Block boundary filter.....	141
Annex R – Independent Segment Decoding mode	143
R.1 Introduction	143
R.2 Mode operation.....	143
R.3 Constraints on usage.....	144
Annex S – Alternative INTER VLC mode	145
S.1 Introduction	145
S.2 Alternative INTER VLC for coefficients	145
S.3 Alternative INTER VLC for CBPY	146
Annex T – Modified Quantization mode	146
T.1 Introduction	146
T.2 Modified DQUANT Update.....	146
T.3 Altered quantization step size for chrominance coefficients.....	147
T.4 Modified coefficient range	148
T.5 Usage restrictions	148
Annex U – Enhanced reference picture selection mode	149
U.1 Introduction	149
U.2 Video source coding algorithm	150
U.3 Forward-channel syntax	151
U.4 Decoder process.....	169
U.5 Back-channel messages	174
Annex V – Data-partitioned slice mode.....	178
V.1 Scope	178
V.2 Structure of data partitioning.....	178
V.3 Interaction with other optional modes.....	181
Annex W – Additional supplemental enhancement information specification	185
W.1 Scope	185
W.2 References	185
W.3 Additional FTYPE Values.....	185
W.4 Recommended maximum number of PSUPP octets	185
W.5 Fixed-point IDCT	185
W.6 Picture message	195

Annex X – Profiles and levels definition	200
X.1 Scope	200
X.2 Profiles of preferred mode support.....	200
X.3 Picture formats and picture clock frequencies.....	204
X.4 Levels of performance capability	205
X.5 Generic capability definitions for use with ITU-T Rec. H.245.....	209
Appendix I – Error tracking.....	213
I.1 Introduction	213
I.2 Error tracking.....	213
Appendix II – Recommended optional enhancement.....	214

ITU-T Recommendation H.263

Video coding for low bit rate communication

1 Scope

This Recommendation specifies a coded representation that can be used for compressing the moving picture component of audio-visual services at low bit rates. The basic configuration of the video source coding algorithm is based on ITU-T Rec. H.261. Eighteen negotiable coding options are included for improved performance and increased functionality. This Recommendation contains Version 2 of ITU-T Rec. H.263, which is fully compatible with the original Recommendation, adding only optional features to the original Version 1 content of the Recommendation.

2 References

2.1 Normative references

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

- [1] ITU-R Recommendation BT.601-5 (1995), *Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios*.

Reference [1] is referenced herein to define the (Y, C_B, C_R) colour space and its 8-bit integer representation for pictures used by video codecs designed according to this Recommendation. (Reference [1] is not used to define any other aspects of this Recommendation.)

2.2 Informative references

The following additional ITU-T Recommendations are mentioned for illustration purposes in this text.

- [2] ITU-T Recommendation H.223 (2001), *Multiplexing protocol for low bit rate multimedia communication*.
- [3] ITU-T Recommendation H.242 (2004), *System for establishing communication between audiovisual terminals using digital channels up to 2 Mbit/s*.
- [4] ITU-T Recommendation H.245 (2005), *Control protocol for multimedia communication*.
- [5] ITU-T Recommendation H.261 (1993), *Video codec for audiovisual services at $p \times 64$ kbit/s*.
- [6] ITU-T Recommendation H.262 (2000) | ISO/IEC 13818-2:2000, *Information technology – Generic coding of moving pictures and associated audio information: Video*.
- [7] ITU-T Recommendation H.324 (2002), *Terminal for low bit-rate multimedia communication*.

3 Brief specification

An outline block diagram of the codec is given in Figure 1.

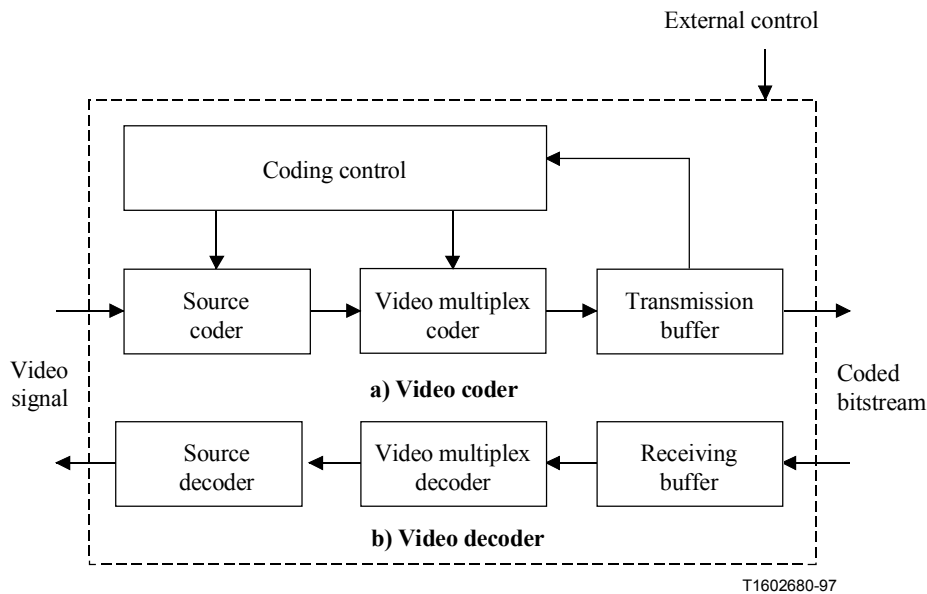


Figure 1/H.263 – Outline block diagram of the video codec

3.1 Video input and output

To permit a single Recommendation to cover use in and between regions using 625- and 525-line television standards, the standardized source formats on which the source coder operates are based on a Common Intermediate Format (CIF). It is also possible using external negotiation (for example, ITU-T Rec. H.245), to enable the use of a wide range of optional custom source formats. The standards of the input and output television signals, which may, for example, be composite or component, analogue or digital and the methods of performing any necessary conversion to and from the source coding format are not subject to recommendation.

3.2 Digital output and input

The video coder provides a self-contained digital bitstream which may be combined with other multi-facility signals (for example as defined in ITU-T Rec. H.223). The video decoder performs the reverse process.

3.3 Sampling frequency

Pictures are sampled at an integer multiple of the video line rate. This sampling clock and the digital network clock are asynchronous.

3.4 Source coding algorithm

A hybrid of inter-picture prediction to utilize temporal redundancy and transform coding of the remaining signal to reduce spatial redundancy is adopted. The decoder has motion compensation capability, allowing optional incorporation of this technique in the coder. Half pixel precision is used for the motion compensation, as opposed to ITU-T Rec. H.261 where full pixel precision and a loopfilter are used. Variable length coding is used for the symbols to be transmitted.

In addition to the core H.263 coding algorithm, eighteen negotiable coding options can be used, either together or separately (subject to certain restrictions). Additional supplemental information may also be included in the bitstream for enhanced display capability and for external usage. A forward error correction method for application to the resulting video bitstream is provided for use,

when necessary. The negotiable coding options, forward error correction, and supplemental information usage are described in the subsequent subclauses.

3.4.1 Continuous Presence Multipoint and Video Multiplex mode

In this optional mode, up to four separate video "Sub-Bitstreams" can be sent within the same video channel. This feature is designed for use in continuous presence multipoint application or other situations in which separate logical channels are not available, but the use of multiple video bitstreams is desired (see also Annex C).

3.4.2 Unrestricted Motion Vector mode

In this optional mode, motion vectors are allowed to point outside the picture. The edge pixels are used as prediction for the "non-existing" pixels. With this mode a significant gain is achieved if there is movement across the edges of the picture, especially for the smaller picture formats (see also Annex D). Additionally, this mode includes an extension of the motion vector range so that larger motion vectors can be used. This is especially useful in case of camera movement and large picture formats.

3.4.3 Syntax-based Arithmetic Coding mode

In this optional mode, arithmetic coding is used instead of variable length coding. The SNR and reconstructed pictures will be the same, but significantly fewer bits will be produced (see also Annex E).

3.4.4 Advanced Prediction mode

In this optional mode, Overlapped Block Motion Compensation (OBMC) is used for the luminance part of P-pictures (see also Annex F). Four 8×8 vectors instead of one 16×16 vector are used for some of the macroblocks in the picture. The encoder has to decide which type of vectors to use. Four vectors use more bits, but give better prediction. The use of this mode generally gives a considerable improvement. A subjective gain is achieved because OBMC results in less blocking artifacts.

3.4.5 PB-frames mode

A PB-frame consists of two pictures being coded as one unit. The name PB comes from the name of picture types in ITU-T Rec. H.262 where there are P-pictures and B-pictures. Thus a PB-frame consists of one P-picture which is predicted from the previous decoded P-picture and one B-picture which is predicted from both the previous decoded P-picture and the P-picture currently being decoded. The name B-picture was chosen because parts of B-pictures may be bidirectionally predicted from the past and future pictures. With this coding option, the picture rate can be increased considerably without substantially increasing the bit rate (see also Annex G). However, an Improved PB-frames mode is also provided (see also Annex M). The original PB-frames mode is retained herein only for purposes of compatibility with systems made prior to the adoption of the Improved PB-frames mode.

3.4.6 Forward Error Correction

A forward error correction method is specified for use when necessary for the protection of the video bitstream for use, when appropriate. The method provided for forward error correction is the same BCH code method specified also in ITU-T Rec. H.261 (see also Annex H).

3.4.7 Advanced INTRA Coding mode

In this optional mode, INTRA blocks are first predicted from neighbouring INTRA blocks prior to coding (see also Annex I). Separate Variable Length Code (VLC) tables are defined for the INTRA blocks. The technique is applied to INTRA-macroblocks within INTRA pictures and to

INTRA-macroblocks within INTER pictures. This mode significantly improves the compression performance over the INTRA coding of the core H.263 syntax.

3.4.8 Deblocking Filter mode

In this optional mode, a filter is applied across the 8×8 block edge boundaries of decoded I- and P-pictures to reduce blocking artifacts (see also Annex J). The purpose of the filter is to mitigate the occurrence of block edge artifacts in the decoded picture. The filter affects the picture that is used for the prediction of subsequent pictures and thus lies within the motion prediction loop.

3.4.9 Slice Structured mode

In this optional mode, a "slice" layer is substituted for the GOB layer of the bitstream syntax (see also Annex K). The purposes of this mode are to provide enhanced error resilience capability, to make the bitstream more amenable to use with an underlying packet transport delivery, and to minimize video delay. A slice is similar to a GOB, in that it is a layer of the syntax that lies between the picture layer and the macroblock layer. However, the use of a slice layer allows a flexible partitioning of the picture, in contrast with the fixed partitioning and fixed transmission order required by the GOB structure.

3.4.10 Supplemental enhancement information

Additional supplemental information may be included in the bitstream to signal enhanced display capability or to provide information for external usage (see also Annex L). This supplemental information can be used to signal a full-picture or partial-picture freeze or freeze-release request with or without resizing, can also be used to tag particular pictures or sequences of pictures within the video stream for external usage, and can be used to convey chroma key information for video compositing. The supplemental information may be present in the bitstream even though the decoder may not be capable of providing the enhanced capability to use it, or even to properly interpret it – simply discarding the supplemental information is allowable by decoders unless a requirement to provide the requested capability has been negotiated by external means.

3.4.11 Improved PB-frames mode

This optional mode represents an improvement compared to the PB-frames mode option (see also Annexes G and M). The main difference between the two modes is that in the Improved PB-frames mode, each B-block may be forward predicted using a separate motion vector or backward predicted with a zero vector. This significantly improves coding efficiency in situations in which downscaled P-vectors are not good candidates for B-prediction. The backward prediction is particularly useful when there is a scene cut between the previous P-frame and the PB-frame.

3.4.12 Reference Picture Selection mode

An optional mode is provided which improves the performance of real-time video communication over error-prone channel by allowing temporal prediction from pictures other than the most recently-sent reference picture (see also Annex N). This mode can be used with back-channel status messages which are sent back to the encoder to inform it about whether its bitstream is being properly received. In error-prone channel environments, this mode allows the encoder to optimize its video encoding for the conditions of the channel.

3.4.13 Temporal, SNR and Spatial Scalability mode

In this optional mode, there is support for temporal, SNR, and spatial scalability (see also Annex O). Scalability implies that a bitstream is composed of a base layer and one or more associated enhancement layers. The base layer is a separately decodable bitstream. The enhancement layers can be decoded in conjunction with the base layer to increase perceived quality by either increasing the picture rate, increasing the picture quality, or increasing the picture size. SNR scalability refers to enhancement information to increase the picture quality without increasing

picture resolution. Spatial scalability refers to enhancement information to increase the picture quality by increasing picture resolution either horizontally, vertically, or both. There is also support for temporal scalability by the use of B-pictures. A B-picture is a scalability enhancement containing pictures that can be bidirectionally predicted from two pictures in a reference layer, one temporally previous to the current picture and one temporally subsequent. B-pictures allows enhancement layer information to be used to increase perceived quality by increasing the picture rate of the displayed enhanced video sequence. This mode can be useful for heterogenous networks with varying bandwidth capacity and also in conjunction with error correction schemes.

3.4.14 Reference Picture Resampling mode

A syntax is provided for supporting an optional mode for which the reference picture used for video image prediction is processed by a resampling operation prior to its use in forming a predictor for the current input picture (see also Annex P). This allows efficient dynamic selection of an appropriate image resolution for video encoding, and can also support picture warping for use as a global motion compensator or special-effect generator.

3.4.15 Reduced-Resolution Update mode

An optional mode is provided which allows reduced-resolution updates to a reference picture having a higher resolution (see also Annex Q). This mode is expected to be used when encoding a highly active scene, and allows an encoder to increase the picture rate at which moving parts of a scene can be represented, while maintaining a higher resolution representation in more static areas of the scene.

3.4.16 Independent Segment Decoding mode

An optional mode is provided which allows a picture to be constructed without any data dependencies which cross GOB or multi-GOB video picture segments or slice boundaries (see also Annex R). This mode provides error robustness by preventing the propagation of erroneous data across the boundaries of the video picture segment areas.

3.4.17 Alternative INTER VLC mode

An optional mode is provided which improves the efficiency of INTER picture coding when significant changes are evident in the picture (see also Annex S). This efficiency improvement is obtained by allowing a VLC code originally designed for INTRA pictures to be used for some INTER picture coefficients as well.

3.4.18 Modified Quantization mode

An optional mode is provided which improves the bit-rate control ability for encoding, reduces chrominance quantization error, expands the range of representable DCT coefficients, and places certain restrictions on coefficient values (see also Annex T). This mode modifies the semantics of the differential quantization step size parameter of the bitstream by broadening the range of step size changes that can be specified. It also reduces the quantization step size used for chrominance data. The range of DCT coefficient levels is broadened to ensure that any possible coefficient value can be encoded to within the accuracy allowed by the step size. Certain restrictions are also placed on coefficients in this mode to increase error detection performance and minimize decoder complexity.

3.5 Bit rate

The transmission clock is provided externally. The video bit rate may be variable. In this Recommendation, no constraints on the video bit rate are given; constraints will be given by the terminal or the network.

3.6 Buffering

The encoder shall control its output bitstream to comply with the requirements of the hypothetical reference decoder defined in Annex B. Video data shall be provided on every valid clock cycle. This can be ensured by MCBPC stuffing (see Tables 7 and 8) or, when forward error correction is used, also by forward error correction stuffing frames (see Annex H).

The number of bits created by coding any single picture shall not exceed a maximum value specified by the parameter BPPmaxKb which is measured in units of 1024 bits. The minimum allowable value of the BPPmaxKb parameter depends on the largest picture size that has been negotiated for use in the bitstream (see Table 1). The picture size is measured as the picture width times height for the luminance (Y) component, measured in pixels. An encoder may use a larger value for BPPmaxKb than as specified in Table 1, provided the larger value is first negotiated by external means, for example ITU-T Rec. H.245.

When the Temporal, SNR, and Spatial Scalability mode (Annex O) is employed, the number of bits sent for each picture in each enhancement layer shall not exceed the maximum value specified by BPPmaxKb.

Table 1/H.263 – Minimum BPPmaxKb for different source picture formats

Y picture size in pixels (= width × height as given in picture header)	Minimum BPPmaxKb
up to 25 344 (or QCIF)	64
25 360 to 101 376 (or CIF)	256
101 392 to 405 504 (or 4CIF)	512
405 520 and above	1024

3.7 Symmetry of transmission

The codec may be used for bidirectional or unidirectional visual communication.

3.8 Error handling

Error handling should be provided by external means (for example, ITU-T Rec. H.223). If it is not provided by external means (for example, in ITU-T Rec. H.221), the optional error correction code and framing as described in Annex H can be used.

A decoder can send a command to encode one or more GOBs (or slices if Annex K is employed) of its next picture in INTRA mode with coding parameters such as to avoid buffer overflow. A decoder can also send a command to transmit only non-empty GOB headers if the Slice Structured mode (see Annex K) is not in use. The transmission method for these signals is by external means (for example, ITU-T Rec. H.245).

3.9 Multipoint operation

Features necessary to support switched multipoint operation are included in Annex C.

4 Source coder

4.1 Source format

The source coder operates on non-interlaced pictures having a source format which is defined in terms of:

- 1) the picture format, as determined by the number of pixels per line, the number of lines per picture, and the pixel aspect ratio; and
- 2) the timing between pictures, as determined by the Picture Clock Frequency (PCF). For example, the Common Intermediate Format (CIF) has 352 pixels per line, 288 lines, a pixel aspect ratio of 12:11, and a picture clock frequency of 30 000/1001 pictures per second.

The source coder operates on non-interlaced pictures occurring at a Picture Clock Frequency (PCF) of 30 000/1001 (approximately 29.97) times per second, termed the CIF PCF. It is also possible to negotiate the use of an optional custom PCF by external means. A custom PCF is given by $1\,800\,000/(\text{clock divisor} * \text{clock conversion factor})$ where clock divisor can have values of 1 through 127 and clock conversion factor can either be 1000 or 1001. The tolerance on the picture clock frequency is ± 50 ppm.

Pictures are coded as luminance and two colour difference components (Y , C_B and C_R). These components and the codes representing their sampled values are as defined in ITU-R Rec. BT.601-5.

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

These values are nominal ones and the coding algorithm functions with input values of 1 through to 254.

There are five standardized picture formats: sub-QCIF, QCIF, CIF, 4CIF and 16CIF. It is also possible to negotiate a custom picture format. For all of these picture formats, the luminance sampling structure is dx pixels per line, dy lines per picture in an orthogonal arrangement. Sampling of each of the two colour difference components is at $dx/2$ pixels per line, $dy/2$ lines per picture, orthogonal. The values of dx , dy , $dx/2$ and $dy/2$ are given in Table 2 for each of the standardized picture formats.

Table 2/H.263 – Number of pixels per line and number of lines for each of the standardized H.263 picture formats

Picture format	Number of pixels for luminance (dx)	Number of lines for luminance (dy)	Number of pixels for chrominance (dx/2)	Number of lines for chrominance (dy/2)
sub-QCIF	128	96	64	48
QCIF	176	144	88	72
CIF	352	288	176	144
4CIF	704	576	352	288
16CIF	1408	1152	704	576

For all picture formats, colour difference samples are sited such that their block boundaries coincide with luminance block boundaries as shown in Figure 2. The pixel aspect ratio is the same for each of the standardized picture formats and is the same as defined for QCIF and CIF in ITU-T Rec. H.261: $(288/3):(352/4)$, which simplifies to 12:11 in relatively prime numbers. The picture

area covered by all of the standardized picture formats except the sub-QCIF picture format has an aspect ratio of 4:3.

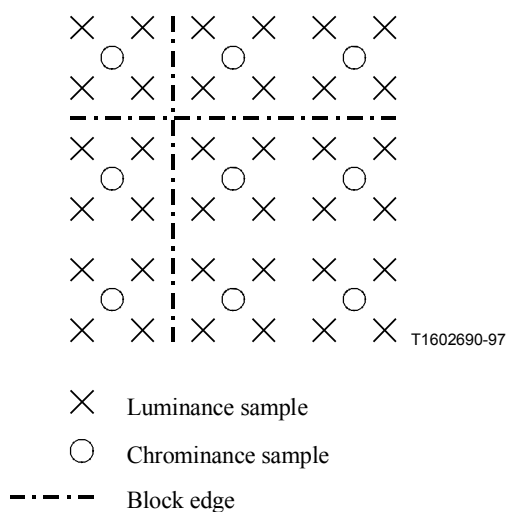


Figure 2/H.263 – Positioning of luminance and chrominance samples

Custom picture formats can have a custom pixel aspect ratio as described in Table 3, if the custom pixel aspect ratio use is first negotiated by external means. Custom picture formats can have any number of lines and any number of pixels per line, provided that the number of lines is divisible by four and is in the range [4, ... , 1152], and provided that the number of pixels per line is also divisible by four and is in the range [4, ... , 2048]. For picture formats having a width or height that is not divisible by 16, the picture is decoded in the same manner as if the width or height had the next larger size that would be divisible by 16 and then the picture is cropped at the right and the bottom to the proper width and height for display purposes only.

Table 3/H.263 – Custom pixel aspect ratios

Pixel aspect ratio	Pixel width:Pixel height
Square	1:1
CIF	12:11
525-type for 4:3 picture	10:11
CIF for 16:9 picture	16:11
525-type for 16:9 picture	40:33
Extended PAR	m:n, m and n are relatively prime

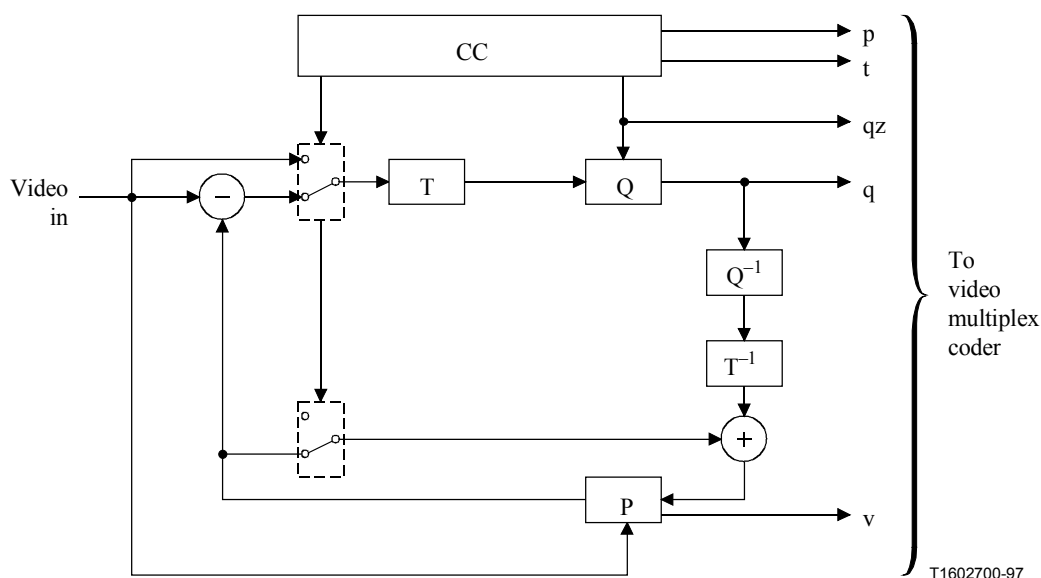
All decoders and encoders shall be able to operate using the CIF picture clock frequency. Some decoders and encoders may also support custom picture clock frequencies. All decoders shall be able to operate using the sub-QCIF picture format. All decoders shall also be able to operate using the QCIF picture format. Some decoders may also operate with CIF, 4CIF or 16CIF, or custom picture formats. Encoders shall be able to operate with one of the picture formats sub-QCIF and QCIF. The encoders determine which of these two formats are used, and are not obliged to be able to operate with both. Some encoders can also operate with CIF, 4CIF, 16CIF, or custom picture formats. Which optional formats and which picture clock frequencies can be handled by the decoder is signalled by external means, for example ITU-T Rec. H.245. For a complete overview of possible picture formats and video coding algorithms, refer to the terminal description, for example ITU-T Rec. H.324.

NOTE – For CIF, the number of pixels per line is compatible for practical purposes with sampling the active portions of the luminance and colour difference signals from 525- or 625-line sources at 6.75 and 3.375 MHz respectively. These frequencies have a simple relationship to those in ITU-R Rec. BT.601-5.

Means shall be provided to restrict the maximum picture rate of encoders by having a minimum number of non-transmitted pictures between transmitted ones. Selection of this minimum number shall be by external means (for example, ITU-T Rec. H.245). For the calculation of the minimum number of non-transmitted pictures in PB-frames mode, the P-picture and the B-picture of a PB-frames unit are taken as two separate pictures.

4.2 Video source coding algorithm

The source coder is shown in generalized form in Figure 3. The main elements are prediction, block transformation and quantization.



- T Transform
- Q Quantizer
- P Picture Memory with motion compensated variable delay
- CC Coding control
- p Flag for INTRA/INTER
- t Flag for transmitted or not
- qz Quantizer indication
- q Quantizing index for transform coefficients
- v Motion vector

Figure 3/H.263 – Source coder

4.2.1 GOBs, slices, macroblocks and blocks

Each picture is divided either into Groups Of Blocks (GOBs) or into slices.

A Group Of Blocks (GOB) comprises of up to $k * 16$ lines, where k depends on the number of lines in the picture format and depends on whether the optional Reduced-Resolution Update mode is in use (see Annex Q). The dependencies are shown in Table 4. If the number of lines is less than or equal to 400 and the optional Reduced-Resolution Update mode is not in use, then $k = 1$. If the number of lines is less than or equal to 800 and the optional Reduced-Resolution Update mode is in use or the number of lines is more than 400, then $k = 2$. If the number of lines is more than 800, then $k = 4$. When using custom picture sizes, the number of lines in the last (bottom-most) GOB may be less than $k * 16$ if the number of lines in the picture is not divisible by $k * 16$. However, every GOB in each of the standardized picture formats has $k * 16$ lines, as the number of lines in

each standardized picture format is an integer multiple of $k * 16$. Thus, for example, if the optional Reduced-Resolution mode is not in use, the number of GOBs per picture is 6 for sub-QCIF, 9 for QCIF, and 18 for CIF, 4CIF and 16CIF. The GOB numbering is done by use of vertical scan of the GOBs, starting with the upper GOB (number 0) and ending with the bottom-most GOB. An example of the arrangement of GOBs in a picture is given for the CIF picture format in Figure 4. Data for each GOB consists of a GOB header (may be empty) followed by data for macroblocks. Data for GOBs is transmitted per GOB in increasing GOB number.

Table 4/H.263 – Parameter k for GOB size definition

Number of Lines dy	Value of k when not in RRU mode	Value of k when in RRU mode
4, ... , 400	1	2
404, ... , 800	2	2
804, ... , 1152	4	4

The Slice Structured mode is described in Annex K. Slices are similar to GOBs in that they are a multi-macroblock layer of the syntax, but slices have a more flexible shape and usage than GOBs, and slices may appear in the bitstream in any order, under certain conditions.

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

Figure 4/H.263 – Arrangement of Groups of Blocks in a CIF picture

Each GOB is divided into macroblocks. The macroblock structure depends on whether the optional Reduced-Resolution Update (RRU) mode is in use (see Annex Q). Unless in RRU mode, each macroblock relates to 16 pixels by 16 lines of Y and the spatially corresponding 8 pixels by 8 lines of C_B and C_R . Further, a macroblock consists of four luminance blocks and the two spatially corresponding colour difference blocks as shown in Figure 5. Each luminance or chrominance block

thus relates to 8 pixels by 8 lines of Y, C_B or C_R. Unless in RRU mode, a GOB comprises one macroblock row for sub-QCIF, QCIF and CIF, two macroblock rows for 4CIF and four macroblock rows for 16CIF.

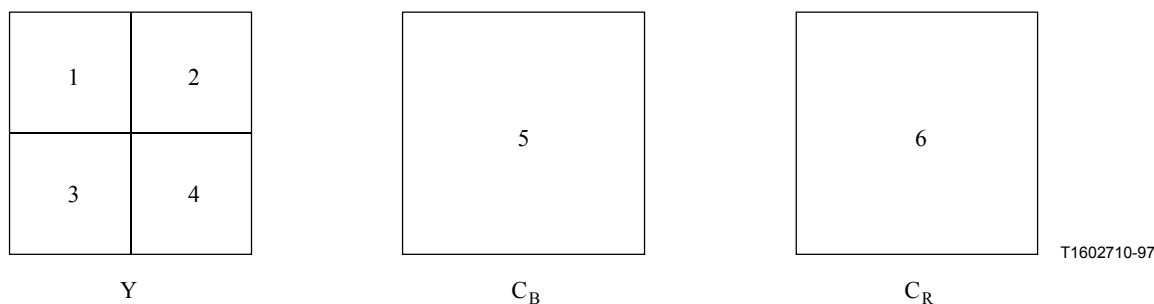


Figure 5/H.263 – Arrangement of blocks in a macroblock

When in RRU mode, a macroblock relates to 32 pixels by 32 lines of Y and the spatially corresponding 16 pixels by 16 lines of C_B and C_R, and each luminance or chrominance block relates to 16 pixels by 16 lines of Y, C_B or C_R. Furthermore, a GOB comprises one macroblock row for CIF and 4CIF, and two macroblock rows for 16CIF.

The macroblock numbering is done by use of horizontal scan of the macroblock rows from left to right, starting with the upper macroblock row and ending with the lower macroblock row. Data for the macroblocks is transmitted per macroblock in increasing macroblock number. Data for the blocks is transmitted per block in increasing block number (see Figure 5).

The criteria for choice of mode and transmitting a block are not subject to recommendation and may be varied dynamically as part of the coding control strategy. Transmitted blocks are transformed and resulting coefficients are quantized and entropy coded.

4.2.2 Prediction

The primary form of prediction is inter-picture and may be augmented by motion compensation (see 4.2.3). The coding mode in which temporal prediction is applied is called INTER; the coding mode is called INTRA if no temporal prediction is applied. The INTRA coding mode can be signalled at the picture level (INTRA for I-pictures or INTER for P-pictures) or at the macroblock level in P-pictures. In the optional PB-frames mode, B-pictures are always coded in INTER mode. The B-pictures are partly predicted bidirectionally (refer to Annex G).

In total, H.263 has seven basic picture types (of which only the first two are mandatory) which are defined primarily in terms of their prediction structure:

- 1) INTRA: A picture having no reference picture(s) for prediction (also called an I-picture);
- 2) INTER: A picture using a temporally previous reference picture (also called a P-picture);
- 3) PB: A frame representing two pictures and having a temporally previous reference picture (see Annex G);
- 4) Improved PB: A frame functionally similar but normally better than a PB-frame (see Annex M);
- 5) B: A picture having two reference pictures, one of which temporally precedes the B-picture and one of which temporally succeeds the B-picture and has the same picture size (see Annex O);
- 6) EI: A picture having a temporally simultaneous reference picture which has either the same or a smaller picture size (see Annex O); and

- 7) EP: A picture having two reference pictures, one of which temporally precedes the EP-picture and one of which is temporally simultaneous and has either the same or a smaller picture size (see Annex O).

As used herein, a "reference" or "anchor" picture is a picture that contains data which can be used by reference as a basis for the decoding of another picture. This use by reference is also known as "prediction", although it sometimes may actually indicate use in a reverse-temporal direction.

4.2.3 Motion compensation

The decoder will accept one vector per macroblock or, if the Advanced Prediction mode or the Deblocking Filter mode is used, one or four vectors per macroblock (see Annexes F and J). If the PB-frames mode is used, one additional delta vector can be transmitted per macroblock for adaptation of the motion vectors for prediction of the B-macroblock. Similarly, an Improved PB-frame macroblock (see Annex M) can include an additional forward motion vector. B-picture macroblocks (see Annex O) can be transmitted together with a forward and a backward motion vector, and EP-pictures can be transmitted with a forward motion vector.

Both horizontal and vertical components of the motion vectors have integer or half integer values. In the default prediction mode, these values are restricted to the range $[-16, 15.5]$ (this is also valid for the forward and backward motion vector components for B-pictures).

In the Unrestricted Motion Vector mode, however, the maximum range for vector components is increased. If PLUSPTYPE is absent, the range is $[-31.5, 31.5]$, with the restriction that only values that are within a range of $[-16, 15.5]$ around the predictor for each motion vector component can be reached if the predictor is in the range $[-15.5, 16]$. If PLUSPTYPE is absent and the predictor is outside $[-15.5, 16]$, all values within the range $[-31.5, 31.5]$ with the same sign as the predictor plus the zero value can be reached. If PLUSPTYPE is present, the motion vector values are less restricted (see also Annex D).

In the Reduced-Resolution Update mode, the motion vector range is enlarged to approximately double size, and each vector component is restricted to have only a half-integer or zero value. Therefore, the range of each motion vector component is $[-31.5, 30.5]$ in the default Reduced-Resolution Update mode (see Annex Q) and a larger range if the Unrestricted Motion Vector mode is also used (see also Annex D).

A positive value of the horizontal or vertical component of the motion vector signifies that the prediction is formed from pixels in the referenced picture which are spatially to the right or below the pixels being predicted.

Motion vectors are restricted such that all pixels referenced by them are within the coded picture area, except when the Unrestricted Motion Vector mode, the Advanced Prediction mode, the Deblocking Filter mode is used (see Annexes D, F and J), or in B- and EP-pictures of the Temporal, SNR, and Spatial Scalability mode (see Annex O).

4.2.4 Quantization

Unless the optional Advanced INTRA Coding mode or Modified Quantization mode is in use, the number of quantizers is 1 for the first coefficient of INTRA blocks and 31 for all other coefficients. Within a macroblock the same quantizer is used for all coefficients except the first one of INTRA blocks. The decision levels are not defined. The first coefficient of INTRA blocks is nominally the transform dc value uniformly quantized with a step size of 8. Each of the other 31 quantizers use equally spaced reconstruction levels with a central dead-zone around zero and with a step size of an even value in the range 2 to 62. For the exact formulas, refer to 6.2. For quantization using the Advanced INTRA Coding mode, refer to Annex I. For quantization using the Modified Quantization mode, refer to Annex T.

NOTE – For the smaller quantization step sizes, the full dynamic range of the transform coefficients cannot be represented unless the optional Modified Quantization mode is in use.

4.3 Coding control

Several parameters may be varied to control the rate of generation of coded video data. These include processing prior to the source coder, the quantizer, block significance criterion and temporal subsampling. The proportions of such measures in the overall control strategy are not subject to recommendation.

When invoked, temporal subsampling is performed by discarding complete pictures.

A decoder can signal its preference for a certain tradeoff between spatial and temporal resolution of the video signal. The encoder shall signal its default tradeoff at the beginning of the call and shall indicate whether it is capable of responding to decoder requests to change this tradeoff. The transmission method for these signals is by external means (for example, ITU-T Rec. H.245).

4.4 Forced updating

This function is achieved by forcing the use of the INTRA mode of the coding algorithm. The update pattern is not defined. To control the accumulation of inverse transform mismatch error, each macroblock shall be coded in INTRA mode at least once every 132 times when coefficients are transmitted for this macroblock in P-pictures. A similar requirement applies when using optional EP-pictures (see Annex O), for which each macroblock shall be coded in an INTRA or upward mode at least once every 132 times when coefficients are transmitted for the macroblock.

4.5 Byte alignment of start codes

Byte alignment of start codes is achieved by inserting a stuffing codeword consisting of less than 8 zero-bits before the start code such that the first bit of the start code is the first (most significant) bit of a byte. A start code is therefore byte aligned if the position of its most significant bit is a multiple of 8 bits from the first bit in the H.263 bitstream. All picture, slice, and EOSBS start codes shall be byte aligned, and GOB and EOS start codes may be byte aligned.

NOTE 1 – The number of bits spent for a certain picture is variable but always a multiple of 8 bits.

NOTE 2 – H.324 requires H.263 encoders to align picture start codes with the start of logical information units passed to the Adaptation Layer (AL_SDUs).

5 Syntax and semantics

The video syntax is arranged in a hierarchical structure with four primary layers. From top to bottom the layers are:

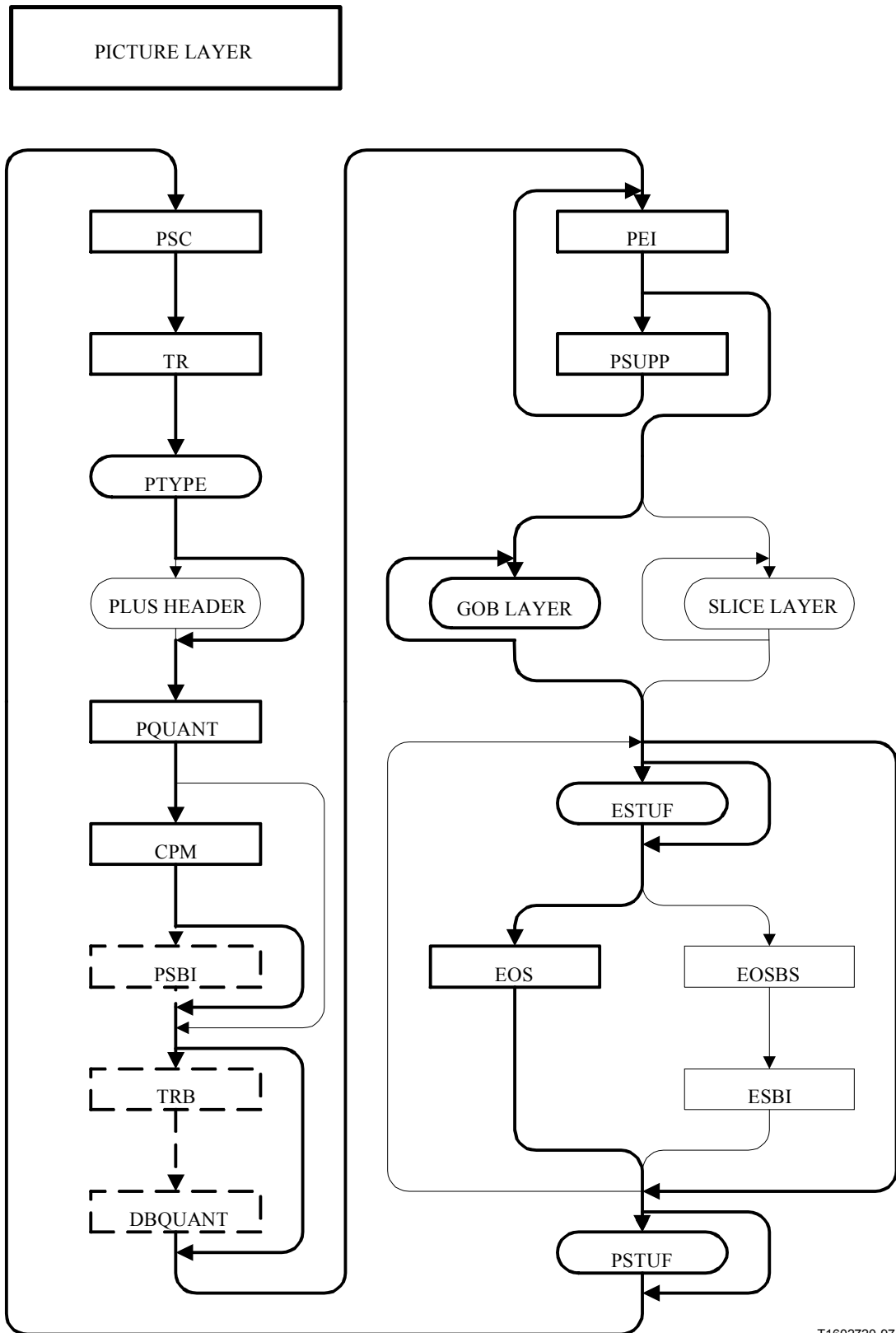
- Picture;
- Group of Blocks, or slice, or video picture segment;
- Macroblock;
- Block.

The syntax diagram is shown in Figure 6. A guide for interpretation of the diagram consists of the following:

- 1) Arrow paths show the possible flows of syntax elements. Any syntax element which has zero length is considered absent for arrow path diagramming purposes (thus, for example, there is an arrow path bypassing PSTUF despite the mandatory nature the PSTUF field since the length of the PSTUF field may be zero).
- 2) Abbreviations and semantics for each syntax element are as defined in later clauses.
- 3) The set of syntax elements and arrow paths shown using thick solid lines denotes the syntax flow of the "baseline" mode of operation, without the use of any optional enhancements. (This syntax was also present in the Version 1 of this Recommendation, and remains unchanged in any way.)

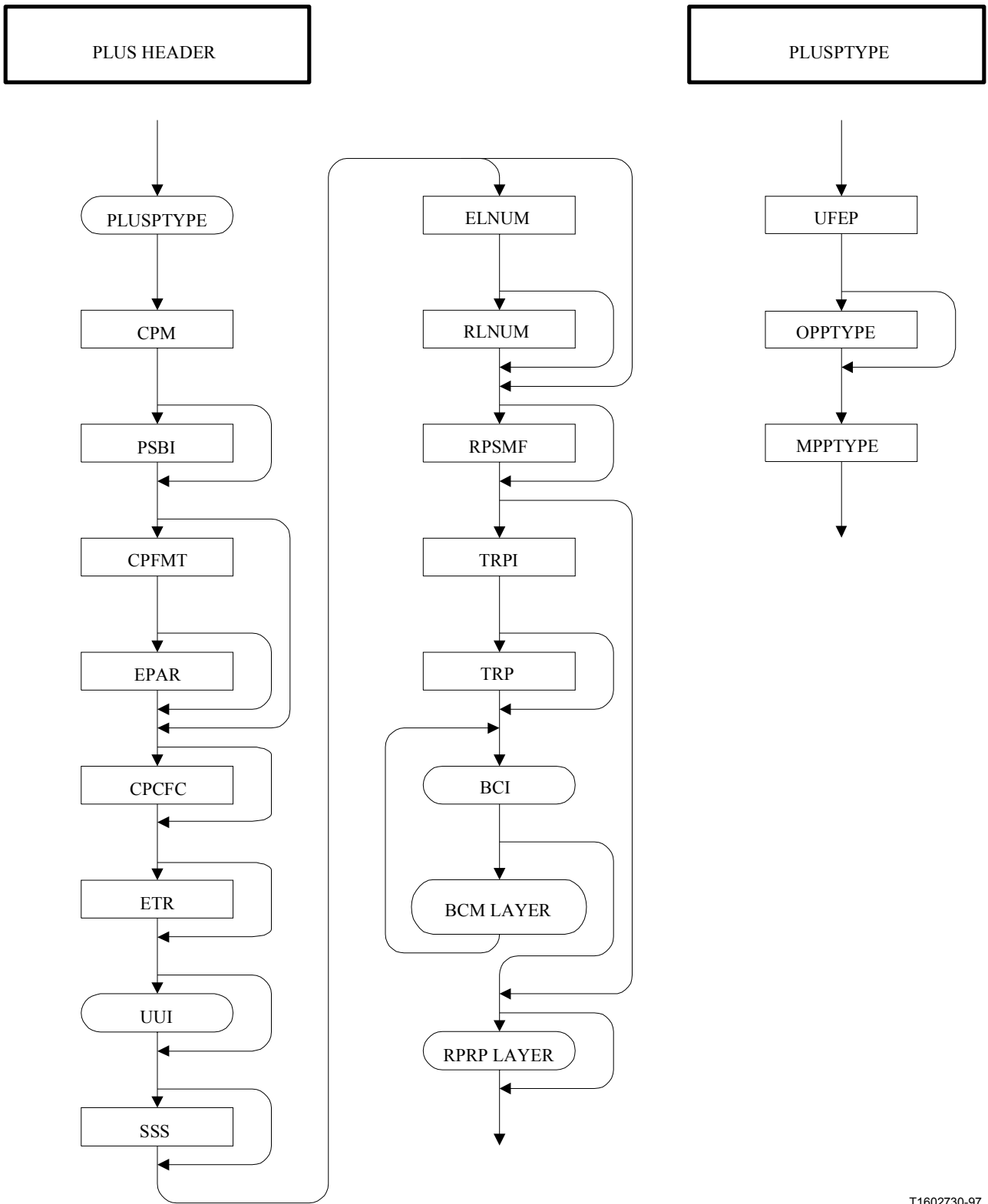
- 4) The set of syntax elements and arrow paths shown using thick dotted lines denotes the additional elements in the syntax flow of the optional enhancements which have been present in both Version 1 and Version 2 of this Recommendation. (This syntax remains unchanged in any way.)
- 5) The set of syntax elements and arrow paths shown using thin solid lines denotes the additional new elements in the syntax flow of the optional enhancements which are specific to the additional optional features added in Version 2. (This syntax was not present in Version 1.)
- 6) Syntax element fields shown with square-edged boundaries indicate fixed-length fields, and those with rounded boundaries indicate variable-length fields. One syntax element (DQUANT) is shown with both types of boundaries, because it can have either a variable or fixed length.
- 7) A fixed-length field is defined to be a field for which the length of the field is not dependent on the data in the content of the field itself. The length of this field is either always the same, or is determined by the prior data in the syntax flow.
- 8) The term "layer" is used to refer to any part of the syntax which can be understood and diagrammed as a distinct entity.

Unless specified otherwise, the most significant bit is transmitted first. This is bit 1 and is the left most bit in the code tables in this Recommendation. Unless specified otherwise, all unused or spare bits are set to "1". Spare bits shall not be used until their functions are specified by the ITU-T.



T1602720-97

Figure 6/H.263 – Syntax diagram for the video bitstream (part 1 of 7)



T1602730-97

Figure 6/H.263 – Syntax diagram for the video bitstream (part 2 of 7)

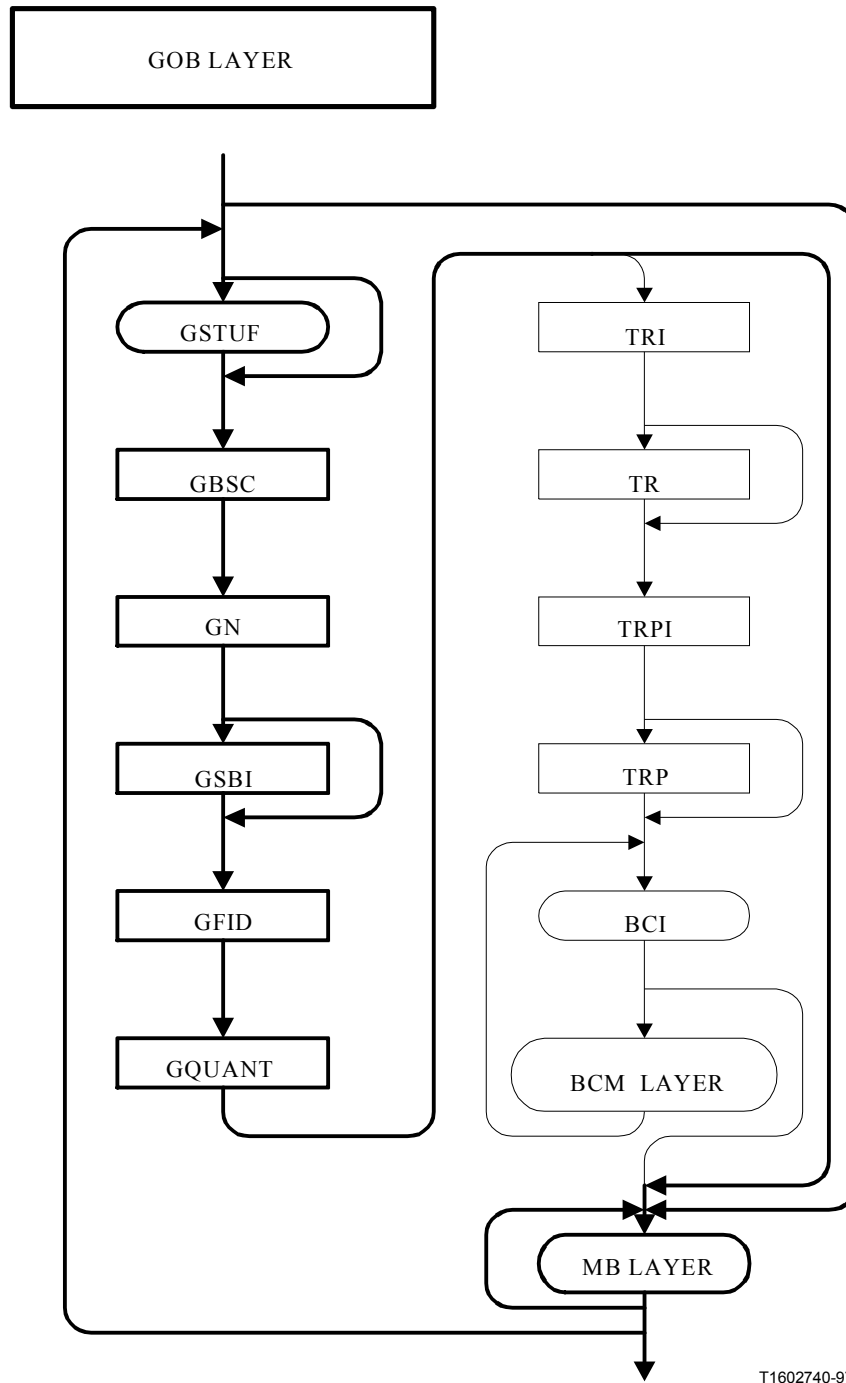


Figure 6/H.263 – Syntax diagram for the video bitstream (part 3 of 7)

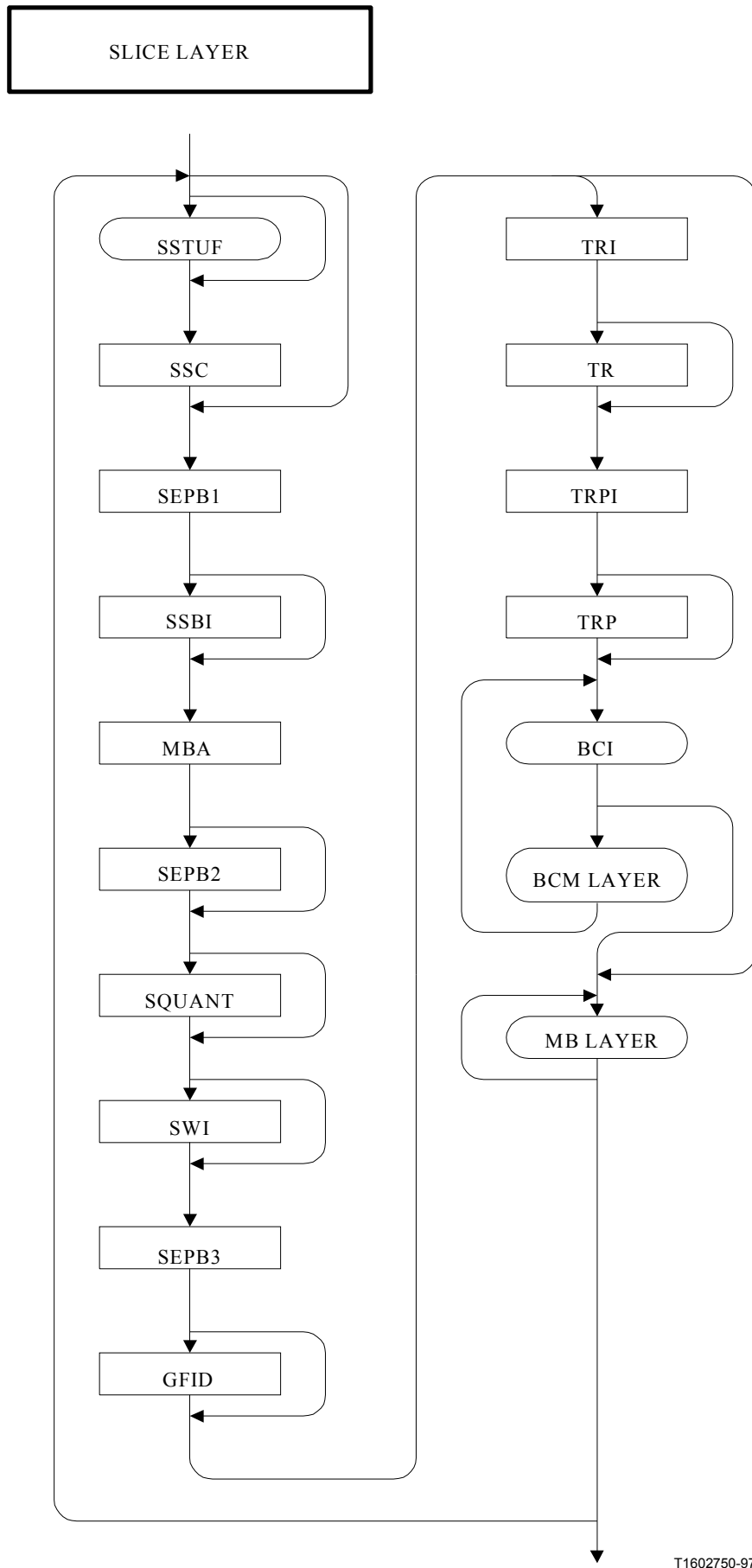


Figure 6/H.263 – Syntax diagram for the video bitstream (part 4 of 7)

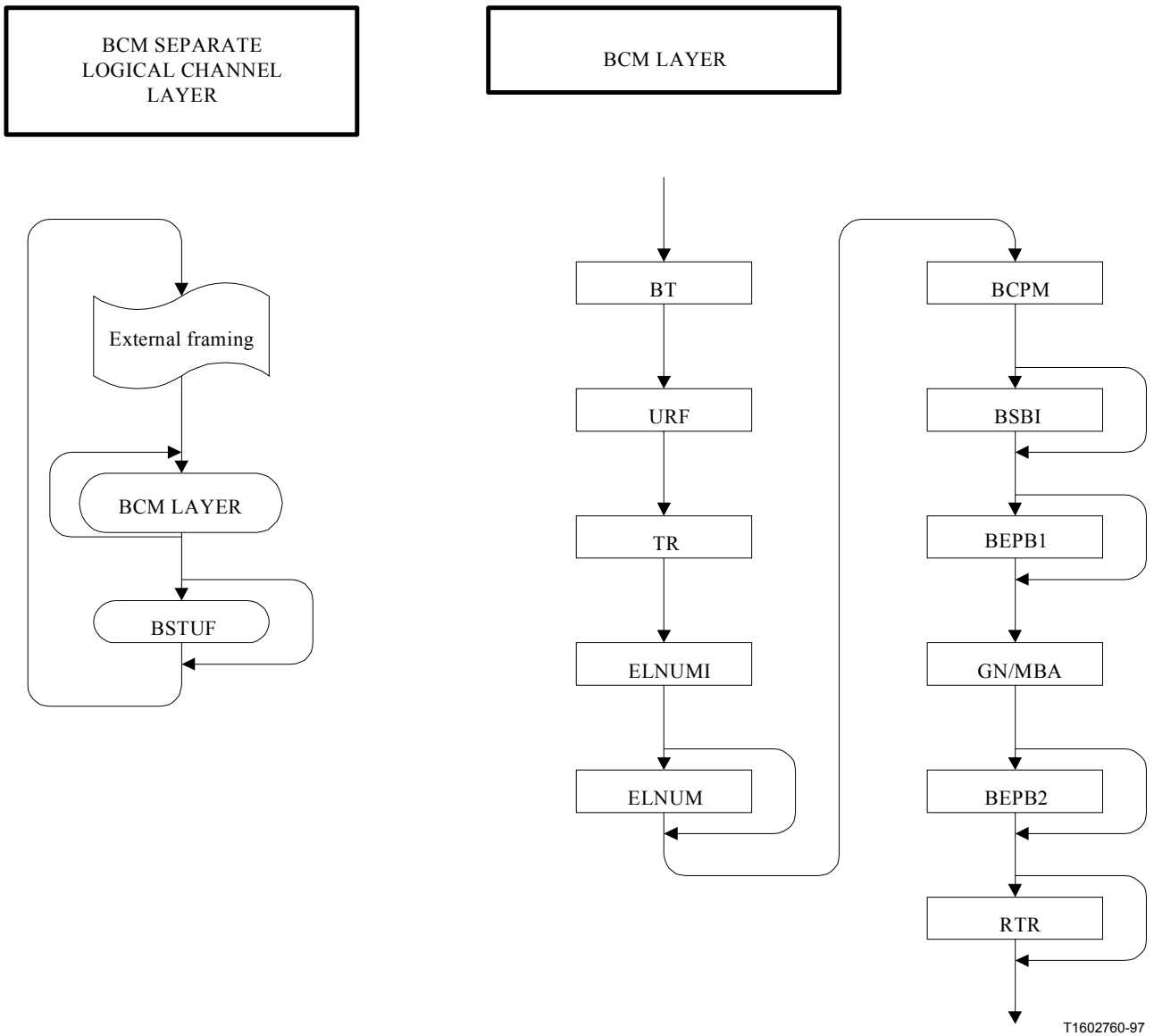
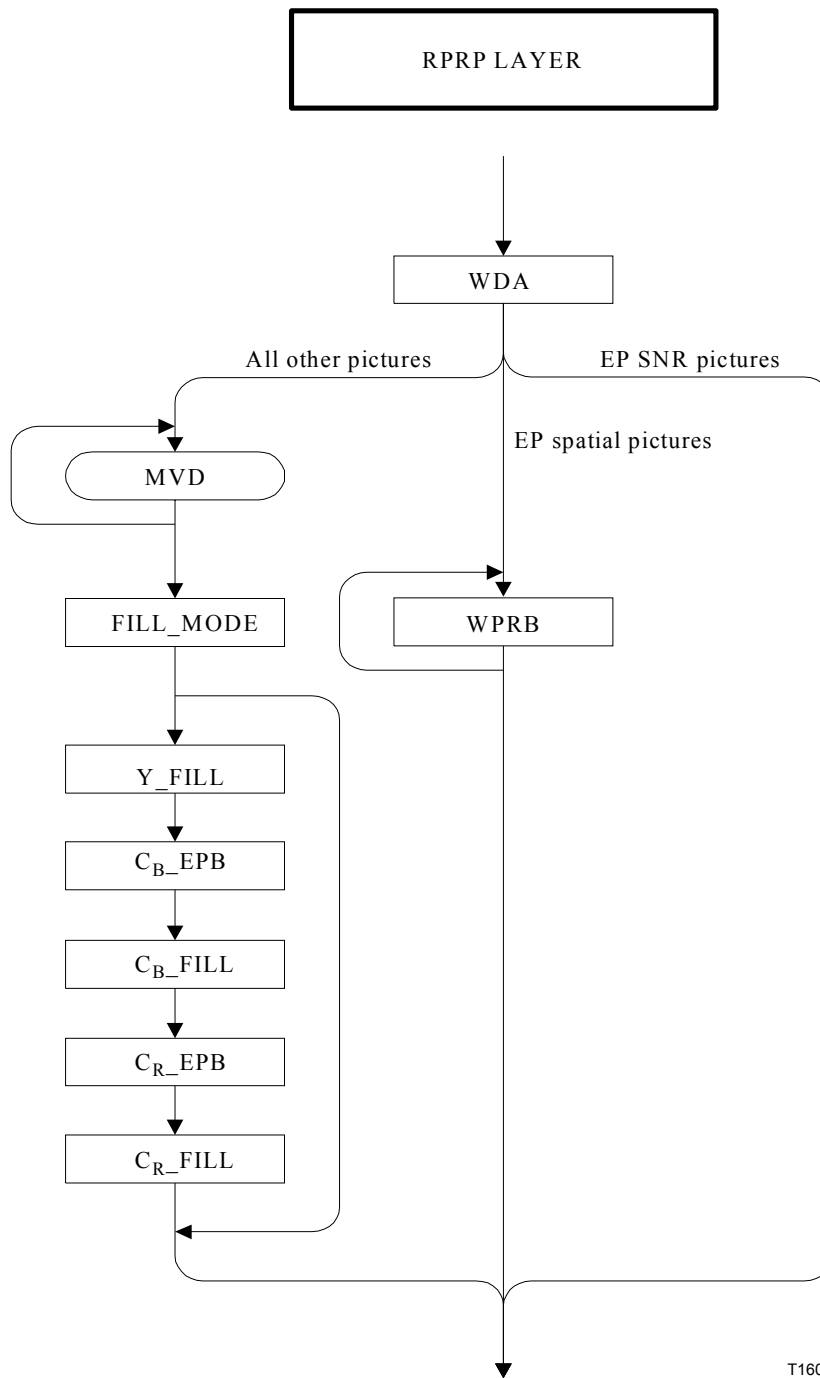
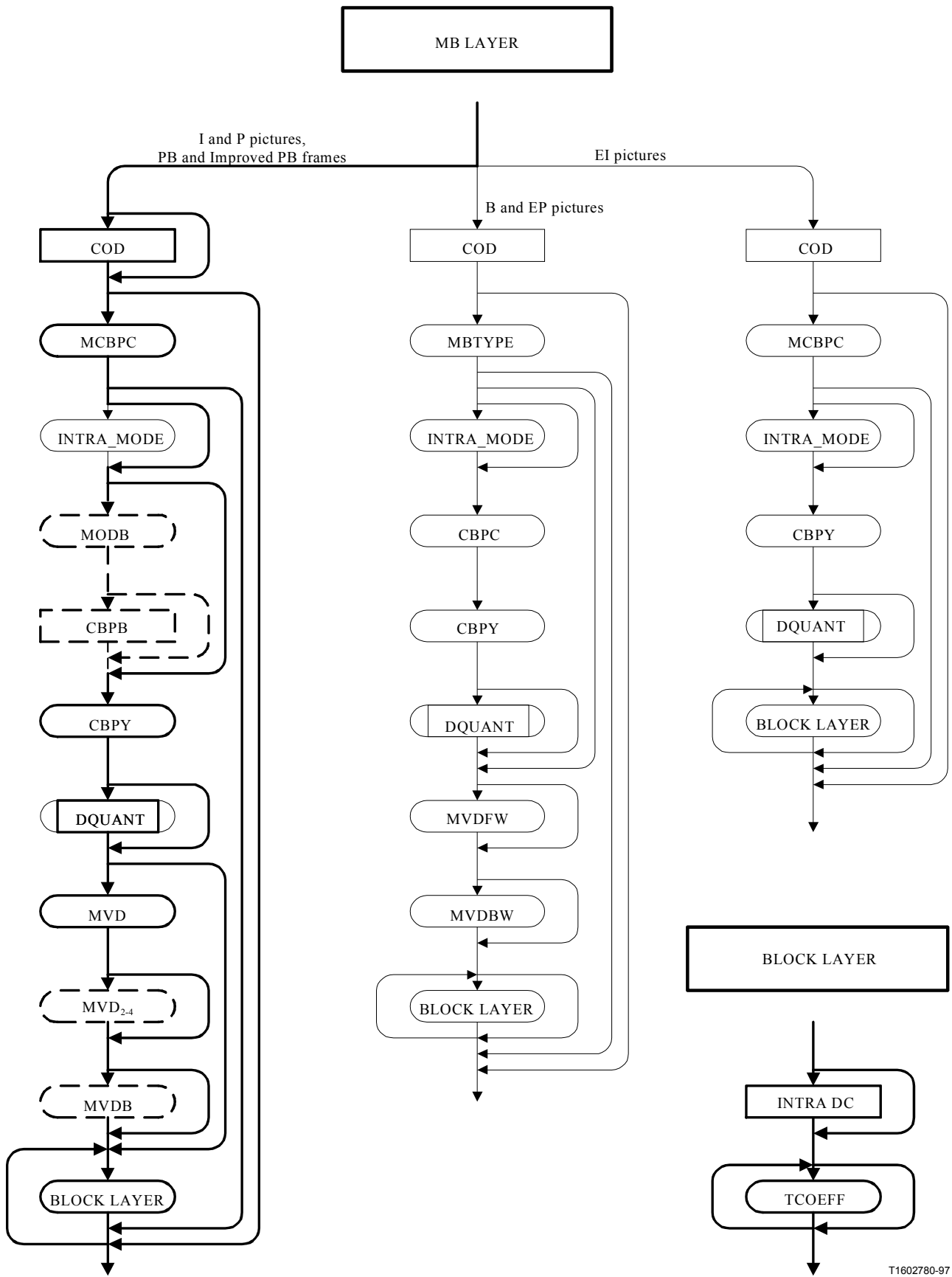


Figure 6/H.263 – Syntax diagram for the video bitstream (part 5 of 7)



T1602770-97

Figure 6/H.263 – Syntax diagram for the video bitstream (part 6 of 7)



T1602780-97

Figure 6/H.263 – Syntax diagram for the video bitstream (part 7 of 7)

5.1 Picture layer

Data for each picture consists of a picture header followed by data for Group of Blocks or for slices, eventually followed by an optional end-of-sequence code and stuffing bits. The structure is shown in Figure 7 for pictures that do not include the optional PLUSPTYPE data field. PSBI is only present if indicated by CPM. TR_B and DBQUANT are only present if PTYPE indicates use of the "PB-frame" mode (unless the PLUSPTYPE field is present and the use of DBQUANT is indicated therein).

The optional PLUSPTYPE data field is present when so indicated in bits 6-8 of PTYPE. When present, an additional set of data is included in the bitstream, which immediately follows PTYPE and precedes PQUANT. In addition, the CPM and PSBI fields are moved forward in the picture header when PLUSPTYPE is present, so that they appear immediately after PLUSPTYPE rather than being located after PQUANT. The format for the additional data following PLUSPTYPE is shown in Figure 8. All fields in this additional picture header data after PLUSPTYPE are optional, depending on whether their presence is indicated in PLUSPTYPE. When the Slice Structured mode (see Annex K) is in use, slices are substituted for GOBs in the location shown in Figure 7.

Combinations of PSUPP and PEI may not be present, and may be repeated when present. EOS and EOSBS+ESBI may not be present, while ESTUF may be present only if EOS or EOSBS is present. EOS shall not be repeated unless at least one picture start code appears between each pair of EOS codes. Picture headers for dropped pictures are not transmitted.

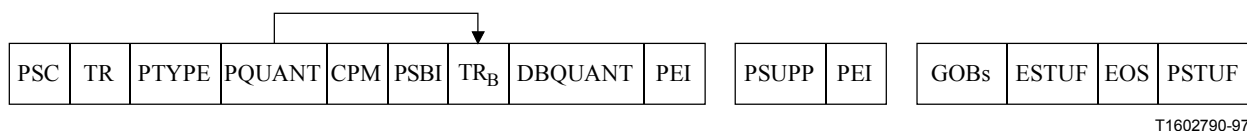


Figure 7/H.263 – Structure of picture layer (without optional PLUSPTYPE-related fields)

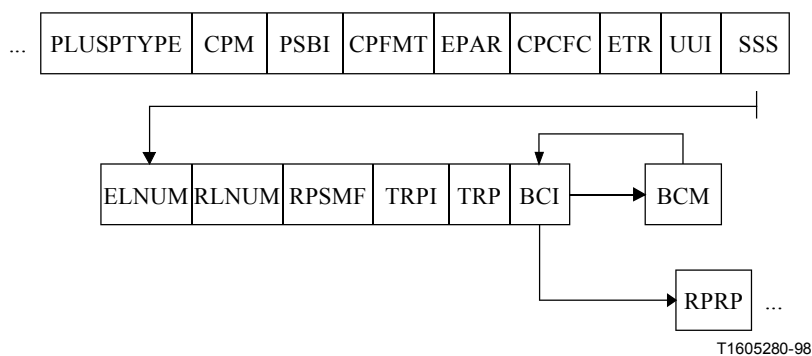


Figure 8/H.263 – Structure of optional PLUSPTYPE-related fields (located immediately after PTYPE when present)

5.1.1 Picture Start Code (PSC) (22 bits)

PSC is a word of 22 bits. Its value is 0000 0000 0000 0000 1 00000. All picture start codes shall be byte aligned. This shall be achieved by inserting PSTUF bits as necessary before the start code such that the first bit of the start code is the first (most significant) bit of a byte.

5.1.2 Temporal Reference (TR) (8 bits)

The value of TR is formed by incrementing its value in the temporally previous reference picture header by one plus the number of skipped or non-reference pictures at the picture clock frequency since the previously transmitted one. The interpretation of TR depends on the active picture clock frequency. Under the standard CIF picture clock frequency, TR is an 8-bit number which can have

256 possible values. The arithmetic is performed with only the eight LSBs. If a custom picture clock frequency is signalled in use, Extended TR in 5.1.8 and TR form a 10-bit number where TR stores the eight Least Significant Bits (LSBs) and ETR stores the two Most Significant Bits (MSBs). The arithmetic in this case is performed with the ten LSBs. In the optional PB-frames or Improved PB-frames mode, TR only addresses P-pictures; for the temporal reference for the B-picture part of PB or Improved PB frames, refer to 5.1.22.

5.1.3 Type Information (PTYPE) (Variable Length)

Information about the complete picture:

- Bit 1: Always "1", in order to avoid start code emulation.
- Bit 2: Always "0", for distinction with ITU-T Rec. H.261.
- Bit 3: Split screen indicator, "0" off, "1" on.
- Bit 4: Document camera indicator, "0" off, "1" on.
- Bit 5: Full Picture Freeze Release, "0" off, "1" on.
- Bits 6-8: Source Format, "000" forbidden, "001" sub-QCIF, "010" QCIF, "011" CIF, "100" 4CIF, "101" 16CIF, "110" reserved, "111" extended PTYPE.

If bits 6-8 are not equal to "111", which indicates an extended PTYPE (PLUSPTYPE), the following five bits are also present in PTYPE:

- Bit 9: Picture Coding Type, "0" INTRA (I-picture), "1" INTER (P-picture).
- Bit 10: Optional Unrestricted Motion Vector mode (see Annex D), "0" off, "1" on.
- Bit 11: Optional Syntax-based Arithmetic Coding mode (see Annex E), "0" off, "1" on.
- Bit 12: Optional Advanced Prediction mode (see Annex F), "0" off, "1" on.
- Bit 13: Optional PB-frames mode (see Annex G), "0" normal I- or P-picture, "1" PB-frame.

Split screen indicator is a signal that indicates that the upper and lower half of the decoded picture could be displayed side by side. This bit has no direct effect on the encoding or decoding of the picture.

Full Picture Freeze Release is a signal from an encoder which responds to a request for packet retransmission (if not acknowledged) or fast update request (see also Annex C) or picture freeze request (see also Annex L) and allows a decoder to exit from its freeze picture mode and display the decoded picture in the normal manner.

If bits 6-8 indicate a different source format than in the previous picture header, the current picture shall be an I-picture, unless an extended PTYPE is indicated in bits 6-8 and the capability to use the optional Reference Picture Resampling mode (see Annex P) has been negotiated externally (for example, ITU-T Rec. H.245).

Bits 10-13 refer to optional modes that are only used after negotiation between encoder and decoder (see also Annexes D, E, F and G, respectively). If bit 9 is set to "0", bit 13 shall be set to "0" as well.

Bits 6-8 shall not have a value of "111" which indicates the presence of an extended PTYPE (PLUSPTYPE) unless the capability has been negotiated externally (for example, ITU-T Rec. H.245) to allow the use of a custom source format or one or more of the other optional modes available only by the use of an extended PTYPE (see Annexes I through K and M through T). Whenever bit 6-8 do not have a value of "111", all of the additional modes available only by the use of an extended PTYPE shall be considered to have been set to an "off" state and shall be inferred to remain "off" unless explicitly switched on later in the bitstream.

5.1.4 Plus PTYPE (PLUSPTYPE) (Variable Length)

A codeword of 12 or 30 bits which is present only if the presence of extended PTYPE is indicated in bits 6-8 of PTYPE.

PLUSPTYPE is comprised of up to three subfields: UFEP, OPPTYPE, and MPPTYPE. OPPTYPE is present only if UFEP has a particular value.

5.1.4.1 Update Full Extended PTYPE (UFEP) (3 bits)

A fixed length codeword of 3 bits which is present only if "extended PTYPE" is indicated in PTYPE bits 6-8. When set to "000", it indicates that only those extended PTYPE fields which need to be signalled in every picture header (MPPTYPE) are included in the current picture header. When set to "001", it indicates that all extended PTYPE fields are included in the current picture header. If the picture type is INTRA or EI, this field shall be set to "001".

In addition, if PLUSPTYPE is present in each of a continuing sequence of pictures, this field shall be set to "001" at least as often as specified by a five-second or five-picture timeout period, whichever allows a larger interval of time. More specifically, the timeout period requires UFEP = "001" to appear in the PLUSPTYPE field (if PLUSPTYPE is present in every intervening picture) of the first picture header with temporal reference indicating a time interval greater than or equal to five seconds since the last occurrence of UFEP = "001", or of the fifth picture after the last occurrence of UFEP = "001" (whichever requirement allows a longer period of time as measured by temporal reference).

Encoders should set UFEP to "001" more often in error-prone environments. Values of UFEP other than "000" and "001" are reserved.

5.1.4.2 The Optional Part of PLUSPTYPE (OPPTYPE) (18 bits)

If UFEP is "001", then the following bits are present in PLUSPTYPE:

- Bits 1-3 Source Format, "000" reserved, "001" sub-QCIF, "010" QCIF, "011" CIF, "100" 4CIF, "101" 16CIF, "110" custom source format, "111" reserved;
- Bit 4 Optional Custom PCF, "0" CIF PCF, "1" custom PCF;
- Bit 5 Optional Unrestricted Motion Vector (UMV) mode (see Annex D), "0" off, "1" on;
- Bit 6 Optional Syntax-based Arithmetic Coding (SAC) mode (see Annex E), "0" off, "1" on;
- Bit 7 Optional Advanced Prediction (AP) mode (see Annex F), "0" off, "1" on;
- Bit 8 Optional Advanced INTRA Coding (AIC) mode (see Annex I), "0" off, "1" on;
- Bit 9 Optional Deblocking Filter (DF) mode (see Annex J), "0" off, "1" on;
- Bit 10 Optional Slice Structured (SS) mode (see Annex K), "0" off, "1" on;
- Bit 11 Optional Reference Picture Selection (RPS) mode (see Annex N), "0" off, "1" on;
- Bit 12 Optional Independent Segment Decoding (ISD) mode (see Annex R), "0" off, "1" on;
- Bit 13 Optional Alternative INTER VLC (AIV) mode (see Annex S), "0" off, "1" on;
- Bit 14 Optional Modified Quantization (MQ) mode (see Annex T), "0" off, "1" on;
- Bit 15 Equal to "1" to prevent start code emulation;
- Bit 16 Reserved, shall be equal to "0";
- Bit 17 Reserved, shall be equal to "0";
- Bit 18 Reserved, shall be equal to "0".

5.1.4.3 The mandatory part of PLUSPTYPE when PLUSPTYPE present (MPPTYPE) (9 bits)

Regardless of the value of UFEP, the following 9 bits are also present in PLUSPTYPE:

- Bits 1-3 Picture Type Code:
 - "000" I-picture (INTRA);
 - "001" P-picture (INTER);
 - "010" Improved PB-frame (see Annex M);
 - "011" B-picture (see Annex O);
 - "100" EI-picture (see Annex O);
 - "101" EP-picture (see Annex O);
 - "110" Reserved;
 - "111" Reserved;
- Bit 4 Optional Reference Picture Resampling (RPR) mode (see Annex P), "0" off, "1" on;
- Bit 5 Optional Reduced-Resolution Update (RRU) mode (see Annex Q), "0" off, "1" on;
- Bit 6 Rounding Type (RTYPE) (see 6.1.2);
- Bit 7 Reserved, shall be equal to "0";
- Bit 8 Reserved, shall be equal to "0";
- Bit 9 Equal to "1" to prevent start code emulation.

The encoder should control the rounding type so that P-pictures, Improved PB-frames, and EP-pictures have different values for bit 6 (Rounding Type for P-pictures) from their reference pictures for motion compensation. Bit 6 can have an arbitrary value if the reference picture is an I- or an EI-picture. Bit 6 can be set to "1" only when bits 1-3 indicate a P-picture, Improved PB-frame, or EP-picture. For other types of pictures, this bit shall always be set to "0".

5.1.4.4 The semantics of PLUSPTYPE

The mandatory part of PLUSPTYPE consists of features which are likely to be changed on a picture-by-picture basis. Primarily, these are the bits that indicate the picture type among I, P, Improved PB, B, EI, and EP. (Note that the PB-frames mode of Annex G cannot be used with PLUSPTYPE present – the Improved PB-frames mode of Annex M should be used instead.) However, they also include indications for the use of the RPR and RRU modes, as these may change from picture to picture as well.

Features that are likely to remain in use rather than being changed from one picture to another (except in obvious ways as discussed in 5.1.4.5) have been placed in the optional part of PLUSPTYPE. When UFEP is 000, the missing mode information is inferred from the type of picture and from the mode information sent in a previous PLUSPTYPE with UFEP equal to 001.

If PLUSPTYPE is present but UFEP is 000, then:

- 1) For a P-picture or Improved PB-frame (see Annex M), the pixel aspect ratio, picture width, and picture height are unchanged from those of the reference picture.
- 2) For a temporal-scalability B-picture (see Annex O) in an enhancement layer, the Reference Layer Number (RLNUM) is the same as the Enhancement Layer Number (ELNUM) if the last picture sent in the enhancement layer was an EI- or EP-picture. If the last picture sent in the enhancement layer was a B-picture, the reference layer number is the same as the reference layer number of the last B-picture. The pixel aspect ratio, picture width and

picture height are unchanged from those of the temporally subsequent reference layer picture.

Note that if temporally surrounding EI- or EP-pictures exist in the same enhancement layer as the B-picture, RLNUM (explicit or implied) shall always be equal to ELNUM. Note also that the pixel aspect ratio, picture width and picture height of a B-picture (explicit or implied) shall always be equal to those of its temporally subsequent reference layer picture.

- 3) For a SNR/Spatial scalability EP-picture (see Annex O), the pixel aspect ratio, picture width and picture height are unchanged from those of the temporally previous reference picture in the same enhancement layer.

5.1.4.5 Mode restrictions for certain picture types and mode inference rules

Certain modes do not apply to certain types of pictures. Particularly, these restrictions apply:

- 1) The following modes do not apply within I (INTRA) pictures: Unrestricted Motion Vector (see Annex D), Advanced Prediction (see Annex F), Alternative INTER VLC (see Annex S), Reference Picture Resampling (see Annex P), and Reduced-Resolution Update (see Annex Q).
- 2) The following modes do not apply within B-pictures (see Annex O): Syntax-based Arithmetic Coding (see Annex E), Deblocking Filter (see Annex J), and Advanced Prediction (see Annex F).
- 3) The following modes do not apply within EI-pictures (see Annex O): Unrestricted Motion Vector (see Annex D), Syntax-based Arithmetic Coding (see Annex E), Advanced Prediction (see Annex F), Reference Picture Resampling (see Annex P), Reduced-Resolution Update (see Annex Q), and Alternative INTER VLC (see Annex S).
- 4) The following modes do not apply within EP-pictures (see Annex O): Syntax-based Arithmetic Coding (see Annex E) and Advanced Prediction (see Annex F).

One or more of the modes listed in the above four-item list may have a mode flag having a value "1" in the optional part of PLUSPTYPE within a picture of a type that is prohibited for that mode (types I, B, EI, or EP). This condition is allowed and shall be interpreted subject to the mode inference rules which follow in the next paragraph.

Mode states are subject to the following mode inference rules:

- 1) Once a mode flag has been set to "1" in the optional part of PLUSPTYPE, the current picture and each subsequent picture in the bitstream shall be assigned a state of "on" for that mode.
- 2) An inferred state of "off" shall be assigned to any mode which does not apply within a picture having the current picture type code. However, each subsequent picture in the bitstream shall have an inferred state of "on" for that mode (unless this too results in an obvious conflict – which shall be resolved in the same way). In the case of layered scalable bitstreams (see Annex O), the mode state shall be inferred only from within the same layer of the bitstream.
- 3) The inference of state shall continue until a picture in the same layer that either contains the optional part of PLUSPTYPE or does not contain PLUSPTYPE at all is sent. If a new picture containing the optional part of PLUSPTYPE is sent, the state sent in the new message shall override the old state. If a picture is sent which does not contain PLUSPTYPE (a picture in which bits 6-8 of PTYPE is not "111"), a state of "off" shall be assigned to all modes not explicitly set to "on" in the PTYPE field, and all modes shall continue to have an inferred state of "off" until a new picture containing the optional part of PLUSPTYPE is sent.

- 4) Two modes do not require mode state inference, since the mode flags for these modes appear in the mandatory part of PLUSPTYPE. These are the Reference Picture Resampling mode (Annex P) and the Reduced-Resolution Update mode (Annex Q). The mode flag for either of these modes shall not be set unless the current picture allows the use of the mode. For example, the Reduced-Resolution Update mode bit shall not be set in an INTRA picture.

5.1.4.6 Mode interaction restrictions

Certain modes cannot be used in combination with certain other modes.

- 1) The Syntax-based Arithmetic Coding mode (see Annex E) shall not be used with the Alternative INTER VLC mode (see Annex S) or the Modified Quantization mode (see Annex T).
- 2) If PLUSPTYPE is present, the Unrestricted Motion Vector mode (see Annex D) shall not be used with the Syntax-based Arithmetic Coding mode (see Annex E).
- 3) The Independent Segment Decoding mode (see Annex R) shall not be used with the Reference Picture Resampling mode (see Annex P).
- 4) The Independent Segment Decoding mode (see Annex R) shall not be used with the Slice Structured mode without the simultaneous use of the Rectangular Slice submode of the Slice Structured mode (see Annex K).

5.1.4.7 The picture header location of CPM (1 bit) and PSBI (2 bits)

The location of the CPM and PSBI fields in the picture header depends on whether or not PLUSPTYPE is present (see 5.1.20 and 5.1.21). If PLUSPTYPE is present, then CPM follows immediately after PLUSPTYPE in the picture header. If PLUSPTYPE is not present, then CPM follows immediately after PQUANT in the picture header. PSBI always follows immediately after CPM (if CPM = "1").

5.1.5 Custom Picture Format (CPFMT) (23 bits)

A fixed length codeword of 23 bits that is present only if the use of a custom picture format is signalled in PLUSPTYPE and UFEP is '001'. When present, CPFMT consists of:

- Bits 1-4 Pixel Aspect Ratio Code: A 4-bit index to the PAR value in Table 5. For extended PAR, the exact pixel aspect ratio shall be specified in EPAR (see 5.1.6);
- Bits 5-13 Picture Width Indication: Range [0, ..., 511]; Number of pixels per line = (PWI + 1) * 4;
- Bit 14 Equal to "1" to prevent start code emulation;
- Bits 15-23 Picture Height Indication: Range [1, ..., 288]; Number of lines = PHI * 4.

Table 5/H.263 – PAR code definition

PAR code	Pixel aspect ratios
0000	Forbidden
0001	1:1 (Square)
0010	12:11 (CIF for 4:3 picture)
0011	10:11 (525-type for 4:3 picture)
0100	16:11 (CIF stretched for 16:9 picture)
0101	40:33 (525-type stretched for 16:9 picture)
0110-1110	Reserved
1111	Extended PAR

5.1.6 Extended Pixel Aspect Ratio (EPAR) (16 bits)

A fixed length codeword of 16 bits that is present only if CPFMT is present and extended PAR is indicated therein. When present, EPAR consists of:

- Bits 1-8 PAR Width: "0" is forbidden. The natural binary representation of the PAR width;
- Bits 9-16 PAR Height: "0" is forbidden. The natural binary representation of the PAR height.

The PAR Width and PAR Height shall be relatively prime.

5.1.7 Custom Picture Clock Frequency Code (CPCFC) (8 bits)

A fixed length codeword of 8 bits that is present only if PLUSPTYPE is present and UFEP is 001 and a custom picture clock frequency is signalled in PLUSPTYPE. When present, CPCFC consists of:

- Bit 1 Clock Conversion Code: "0" indicates a clock conversion factor of 1000 and "1" indicates 1001;
- Bits 2-8 Clock Divisor: "0" is forbidden. The natural binary representation of the value of the clock divisor.

The custom picture clock frequency is given by $1\ 800\ 000 / (\text{clock divisor} * \text{clock conversion factor})$ Hz.

The temporal reference counter shall count in units of the inverse of the picture clock frequency, in seconds. When the PCF changes from that specified for the previous picture, the temporal reference for the current picture is measured in terms of the prior PCF, so that the new PCF takes effect only for the temporal reference interpretation of future pictures.

5.1.8 Extended Temporal Reference (ETR) (2 bits)

A fixed length codeword of 2 bits which is present only if a custom picture clock frequency is in use (regardless of the value of UFEP). It is the two MSBs of the 10-bit number defined in 5.1.2.

5.1.9 Unlimited Unrestricted Motion Vectors Indicator (UII) (Variable length)

A variable length codeword of 1 or 2 bits that is present only if the optional Unrestricted Motion Vector mode is indicated in PLUSPTYPE and UFEP is 001. When UII is present it indicates the effective limitation of the range of the motion vectors being used.

- UII = "1" The motion vector range is limited according to Tables D.1 and D.2.
- UII = "01" The motion vector range is not limited except by the picture size.

5.1.10 Slice Structured Submode bits (SSS) (2 bits)

A fixed length codeword of 2 bits which is present only if the optional Slice Structured mode (see Annex K) is indicated in PLUSPTYPE and UFEP is 001. If the Slice Structured mode is in use but UFEP is not 001, the last values sent for SSS shall remain in effect.

- Bit 1 Rectangular Slices, "0" indicates free-running slices, "1" indicates rectangular slices;
- Bit 2 Arbitrary Slice Ordering, "0" indicates sequential order, "1" indicates arbitrary order.

5.1.11 Enhancement Layer Number (ELNUM) (4 bits)

A fixed length codeword of 4 bits which is present only if the optional Temporal, SNR, and Spatial Scalability mode is in use (regardless of the value of UFEP). The particular enhancement layer is identified by an enhancement layer number, ELNUM. Picture correspondence between layers is achieved via the temporal reference. Picture size is either indicated within each enhancement layer using the existing source format fields or is inferred by the relationship to the reference layer. The first enhancement layer above the base layer is designated as Enhancement Layer Number 2, and the base layer has number 1.

5.1.12 Reference Layer Number (RLNUM) (4 bits)

A fixed length codeword of 4 bits which is present only if the optional Temporal, SNR, and Spatial Scalability mode is in use (see Annex O) and UFEP is 001. The layer number for the pictures used as reference anchors is identified by a Reference Layer Number (RLNUM). Time correspondence between layers is achieved via the temporal reference.

Note that for B-pictures in an enhancement layer having temporally surrounding EI- or EP-pictures which are present in the same enhancement layer, RLNUM shall be equal to ELNUM (see Annex O).

5.1.13 Reference Picture Selection Mode Flags (RPSMF) (3 bits)

A fixed length codeword of 3 bits that is present only if the Reference Picture Selection mode is in use and UFEP is 001. When present, RPSMF indicates which type of back-channel messages are needed by the encoder. If the Reference Picture Selection mode is in use but RPSMF is not present, the last value of RPSMF that was sent shall remain in effect.

- 100: neither ACK nor NACK signals needed;
- 101: need ACK signals to be returned;
- 110: need NACK signals to be returned;
- 111: need both ACK and NACK signals to be returned;
- 000-011: Reserved.

5.1.14 Temporal Reference for Prediction Indication (TRPI) (1 bit)

A fixed length codeword of 1 bit that is present only if the optional Reference Picture Selection mode is in use (regardless of the value of UFEP). When present, TRPI indicates the presence of the following TRP field:

- 0: TRP field is not present;
- 1: TRP field is present.

TRPI shall be 0 whenever the picture header indicates an I- or EI-picture.

5.1.15 Temporal Reference for Prediction (TRP) (10 bits)

When present (as indicated in TRPI), TRP indicates the Temporal Reference which is used for prediction of the encoding, except for in the case of B-pictures. For B-pictures, the picture having the temporal reference TRP is used for the prediction in the forward direction. (Prediction in the reverse-temporal direction always uses the immediately temporally subsequent picture.) TRP is a ten-bit number. If a custom picture clock frequency was not in use for the reference picture, the two MSBs of TRP are zero and the LSBs contain the eight-bit TR found in the picture header of the reference picture. If a custom picture clock frequency was in use for the reference picture, TRP is a ten-bit number consisting of the concatenation of ETR and TR from the reference picture header.

When TRP is not present, the most recent temporally previous anchor picture shall be used for prediction, as when not in the Reference Picture Selection mode. TRP is valid until the next PSC, GSC, or SSC.

5.1.16 Back-Channel message Indication (BCI) (Variable length)

A variable length field of one or two bits that is present only if the optional Reference Picture Selection mode is in use. When set to "1", this signals the presence of the following optional video Back-Channel Message (BCM) field. "01" indicates the absence or the end of the video back-channel message field. Combinations of BCM and BCI may not be present, and may be repeated when present. BCI shall be set to "01" if the videomux submode of the optional Reference Picture Selection mode is not in use.

5.1.17 Back-Channel Message (BCM) (Variable length)

The Back-Channel message with syntax as specified in N.4.2, which is present only if the preceding BCI field is present and is set to "1".

5.1.18 Reference Picture Resampling Parameters (RPRP) (Variable length)

A variable length field that is present only if the optional Reference Picture Resampling mode bit is set in PLUSPTYPE. This field carries the parameters of the Reference Picture Resampling mode (see Annex P). Note that the Reference Picture Resampling mode can also be invoked implicitly by the occurrence of a picture header for an INTER coded picture having a picture size which differs from that of the previous encoded picture, in which case the RPRP field is not present and the Reference Picture Resampling mode bit is not set.

5.1.19 Quantizer Information (PQUANT) (5 bits)

A fixed length codeword of 5 bits which indicates the quantizer QUANT to be used for the picture until updated by any subsequent GQUANT or DQUANT. The codewords are the natural binary representations of the values of QUANT which, being half the step sizes, range from 1 to 31.

5.1.20 Continuous Presence Multipoint and Video Multiplex (CPM) (1 bit)

A codeword of 1 bit that signals the use of the optional Continuous Presence Multipoint and Video Multiplex mode (CPM); "0" is off, "1" is on. For the use of CPM, refer to Annex C. CPM follows immediately after PLUSPTYPE if PLUSPTYPE is present, but follows PQUANT in the picture header if PLUSPTYPE is not present.

5.1.21 Picture Sub-Bitstream Indicator (PSBI) (2 bits)

A fixed length codeword of 2 bits that is only present if Continuous Presence Multipoint and Video Multiplex mode is indicated by CPM. The codewords are the natural binary representation of the sub-bitstream number for the picture header and all following information until the next Picture or GOB start code (see also Annex C). PSBI follows immediately after CPM if CPM is "1" (the location of CPM and PSBI in the picture header depend on whether PLUSPTYPE is present).

5.1.22 Temporal Reference for B-pictures in PB-frames (TR_B) (3/5 bits)

TR_B is present if PTYPE or PLUSPTYPE indicates "PB-frame" or "Improved PB-frame" (see also Annexes G and M) and indicates the number of non-transmitted or non-reference pictures (at 29.97 Hz or the custom picture clock frequency indicated in CPCFC) since the last P- or I-picture or P-part of a PB- or Improved PB-frame and before the B-picture part of the PB- or Improved PB-frame. The codeword is the natural binary representation of the number of non-transmitted pictures plus one. It is 3 bits long for standard CIF picture clock frequency and is extended to 5 bits when a custom picture clock frequency is in use. The maximum number of non-transmitted pictures is 6 for the standard CIF picture clock frequency and 30 when a custom picture clock frequency is used.

5.1.23 Quantization information for B-pictures in PB-frames (DBQUANT) (2 bits)

DBQUANT is present if PTYPE or PLUSPTYPE indicates "PB-frame" or "Improved PB-frame" (see also Annexes G and M). In the decoding process a quantization parameter QUANT is obtained for each macroblock. With PB-frames QUANT is used for the P-block, while for the B-block a different quantization parameter BQUANT is used. QUANT ranges from 1 to 31. DBQUANT indicates the relation between QUANT and BQUANT as defined in Table 6. In this table, "/" means division by truncation. BQUANT ranges from 1 to 31; if the value for BQUANT resulting from Table 6 is greater than 31, it is clipped to 31.

Table 6/H.263 – DBQUANT codes and relation between QUANT and BQUANT

DBQUANT	BQUANT
00	$(5 \times \text{QUANT})/4$
01	$(6 \times \text{QUANT})/4$
10	$(7 \times \text{QUANT})/4$
11	$(8 \times \text{QUANT})/4$

5.1.24 Extra Insertion Information (PEI) (1 bit)

A bit which when set to "1" signals the presence of the following optional data field.

5.1.25 Supplemental Enhancement Information (PSUPP) (0/8/16 ... bits)

If PEI is set to "1", then 9 bits follow consisting of 8 bits of data (PSUPP) and then another PEI bit to indicate if a further 9 bits follow and so on. Encoders shall use PSUPP as specified in Annex L. Decoders which do not support the extended capabilities described in Annex L shall be designed to discard PSUPP if PEI is set to 1. This enables backward compatibility for the extended capabilities of Annex L so that a bitstream which makes use of the extended capabilities can also be used without alteration by decoders which do not support those capabilities.

5.1.26 Stuffing (ESTUF) (Variable length)

A codeword of variable length consisting of less than 8 zero-bits. Encoders may insert this codeword directly before an EOS codeword. Encoders shall insert this codeword as necessary to attain mandatory byte alignment directly before an EOSBS codeword. If ESTUF is present, the last bit of ESTUF shall be the last (least significant) bit of a byte, so that the start of the EOS or EOSBS codeword is byte aligned. Decoders shall be designed to discard ESTUF. See Annex C for a description of EOSBS and its use.

5.1.27 End Of Sequence (EOS) (22 bits)

A codeword of 22 bits. Its value is 0000 0000 0000 0000 1 11111. It is up to the encoder to insert this codeword or not. EOS may be byte aligned. This can be achieved by inserting ESTUF before the EOS code such that the first bit of the EOS code is the first (most significant) bit of a byte. EOS shall not be repeated unless at least one picture start code appears between each pair of EOS codes.

5.1.28 Stuffing (PSTUF) (Variable length)

A codeword of variable length consisting of less than 8 zero-bits. Encoders shall insert this codeword for byte alignment of the next PSC. The last bit of PSTUF shall be the last (least significant) bit of a byte, so that the video bitstream including PSTUF is a multiple of 8 bits from the first bit in the H.263 bitstream. Decoders shall be designed to discard PSTUF.

If for some reason the encoder stops encoding pictures for a certain time-period and resumes encoding later, PSTUF shall be transmitted before the encoder stops, to prevent that the last up to 7 bits of the previous picture are not sent until the coder resumes coding.

5.2 Group of Blocks Layer

Data for each Group of Blocks (GOB) consists of a GOB header followed by data for macroblocks. The structure is shown in Figure 9. Each GOB contains one or more rows of macroblocks. For the first GOB in each picture (with number 0), no GOB header shall be transmitted. For all other GOBs, the GOB header may be empty, depending on the encoder strategy. A decoder can signal the remote encoder to transmit only non-empty GOB headers by external means, for example ITU-T Rec. H.245. GSTUF may be present when GBSC is present. GN, GFID and GQUANT are present when GBSC is present. GSBI is present when CPM is "1" in the Picture header.

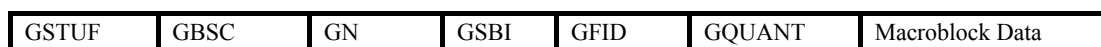


Figure 9/H.263 – Structure of GOB layer

5.2.1 Stuffing (GSTUF) (Variable length)

A codeword of variable length consisting of less than 8 zero-bits. Encoders may insert this codeword directly before a GBSC codeword. If GSTUF is present, the last bit of GSTUF shall be the last (least significant) bit of a byte, so that the start of the GBSC codeword is byte aligned. Decoders shall be designed to discard GSTUF.

5.2.2 Group of Block Start Code (GBSC) (17 bits)

A word of 17 bits. Its value is 0000 0000 0000 0000 1. GOB start codes may be byte aligned. This can be achieved by inserting GSTUF before the start code such that the first bit of the start code is the first (most significant) bit of a byte.

5.2.3 Group Number (GN) (5 bits)

A fixed length codeword of 5 bits. The bits are the binary representation of the number of the Group of Blocks. For the GOB with number 0, the GOB header including GSTUF, GBSC, GN, GSBI, GFID and GQUANT is empty; as group number 0 is used in the PSC. Group numbers 1 through 17 are used in GOB headers of the standard picture formats. Group numbers 1 through 24 are used in GOB headers of custom picture formats. Group numbers 16 through 28 are emulated in the slice header (see Annex K) when CPM = "0", and Group Numbers 25-27 and 29 are emulated in the slice header (see Annex K) when CPM = "1". Group number 31 is used in the EOS code, and group number 30 is used in the EOSBS code.

5.2.4 GOB Sub-Bitstream Indicator (GSBI) (2 bits)

A fixed length codeword of 2 bits that is only present if CPM is "1" in the picture header. The codewords are the natural binary representation of the sub-bitstream number for the GOB header and all following information until the next Picture or GOB start code (see also Annex C).

5.2.5 GOB Frame ID (GFID) (2 bits)

A fixed length codeword of 2 bits. GFID shall have the same value in every GOB (or slice) header of a given picture. Moreover, if PTYPE as indicated in a picture header is the same as for the previous transmitted picture, GFID shall have the same value as in that previous picture, provided PLUSPTYPE is not present. However, if PTYPE in a certain picture header differs from the PTYPE in the previous transmitted picture header, the value for GFID in that picture shall differ from the value in the previous picture.

If PLUSPTYPE is present, the value of GFID shall be the same as that for the previous picture (in the same layer) if the PTYPE, and PLUSPTYPE, and all of the present fields among CPFMT, EPAR, CPCFC, SSS, ELNUM, RLNUM, UUI, RPSMF, and RPRP remain in effect as for the previous picture; otherwise, GFID shall be different from that for the previous picture.

5.2.6 Quantizer Information (GQUANT) (5 bits)

A fixed length codeword of 5 bits which indicates the quantizer QUANT to be used for the remaining part of the picture until updated by any subsequent GQUANT or DQUANT. The codewords are the natural binary representations of the values of QUANT which, being half the step sizes, range from 1 to 31.

5.3 Macroblock layer

Data for each macroblock consists of a macroblock header followed by data for blocks. The structure is shown in Figure 10. COD is only present in pictures that are not of type 'INTRA', for each macroblock in these pictures. MCBPC is present when indicated by COD or when the picture is of type 'INTRA'. MODB is present for MB-type 0-4 if PTYPE indicates "PB-frame". CBPY, DQUANT, MVD and MVD₂₋₄ are present when indicated by MCBPC. CBPB and MVDB are only present if indicated by MODB. Block Data is present when indicated by MCBPC and CBPY. MVD₂₋₄ are only present when in Advanced Prediction mode (refer to Annex F) or Deblocking Filter mode (refer to Annex J). MODB, CBPB and MVDB are only present in PB-frames mode (refer to Annex G). For coding of the symbols in the Syntax-based Arithmetic Coding mode, refer to Annex E. For coding of the macroblock layer in B-, EI-, and EP-pictures, see Annex O.

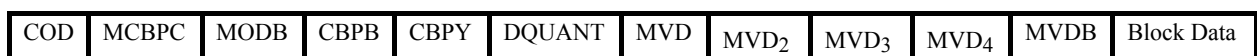


Figure 10/H.263 – Structure of macroblock layer

5.3.1 Coded macroblock indication (COD) (1 bit)

A bit which when set to "0" signals that the macroblock is coded. If set to "1", no further information is transmitted for this macroblock; in that case the decoder shall treat the macroblock as an INTER macroblock with motion vector for the whole block equal to zero and with no coefficient data. COD is only present in pictures that are not of type "INTRA", for each macroblock in these pictures.

NOTE – In Advanced Prediction mode (see Annex F), overlapped block motion compensation is also performed if COD is set to "1"; and in Deblocking Filter mode (see Annex J), the deblocking filter can also affect the values of some pixels of macroblocks having COD is set to "1".

5.3.2 Macroblock type & Coded Block Pattern for Chrominance (MCBPC) (Variable length)

MCBPC is a variable length codeword giving information about the macroblock type and the coded block pattern for chrominance. The codewords for MCBPC are given in Tables 7 and 8. MCBPC is always included in coded macroblocks.

Table 7/H.263 – VLC table for MCBPC (for I-pictures)

Index	MB type	CBPC (56)	Number of bits	Code
0	3	00	1	1
1	3	01	3	001
2	3	10	3	010
3	3	11	3	011
4	4	00	4	0001
5	4	01	6	0000 01
6	4	10	6	0000 10
7	4	11	6	0000 11
8	Stuffing	–	9	0000 0000 1

Table 8/H.263 – VLC table for MCBPC (for P-pictures)

Index	MB type	CBPC (56)	Number of bits	Code
0	0	00	1	1
1	0	01	4	0011
2	0	10	4	0010
3	0	11	6	0001 01
4	1	00	3	011
5	1	01	7	0000 111
6	1	10	7	0000 110
7	1	11	9	0000 0010 1
8	2	00	3	010
9	2	01	7	0000 101
10	2	10	7	0000 100
11	2	11	8	0000 0101
12	3	00	5	0001 1
13	3	01	8	0000 0100
14	3	10	8	0000 0011
15	3	11	7	0000 011
16	4	00	6	0001 00
17	4	01	9	0000 0010 0
18	4	10	9	0000 0001 1
19	4	11	9	0000 0001 0

Table 8/H.263 – VLC table for MCBPC (for P-pictures)

Index	MB type	CBPC (56)	Number of bits	Code
20	Stuffing	–	9	0000 0000 1
21	5	00	11	0000 0000 010
22	5	01	13	0000 0000 0110 0
23	5	10	13	0000 0000 0111 0
24	5	11	13	0000 0000 0111 1

An extra codeword is available in the tables for bit stuffing. This codeword should be discarded by decoders. If an Improved PB-frame is indicated by MPPTYPE bits 1-3 and Custom Source Format is indicated in OPPTYPE bits 1-3, then MBA shall not indicate stuffing before the first macroblock of the picture (in order to prevent start code emulation).

NOTE – Decoders should be designed to allow the macroblock type to indicate bit stuffing immediately prior to the location of a picture, GOB, or slice start code in the bitstream. However, encoders should not use macroblock-layer stuffing in this manner (for interoperability with decoders that may have been designed before the need for decoders to support this was clarified).

The macroblock type gives information about the macroblock and which data elements are present. Macroblock types and included elements are listed in Tables 9 and 10. Macroblock type 5 (indices 21-24 in Table 8) shall not be present unless an extended PTYPE (PLUSPTYPE) is present in the picture header and either the Advanced Prediction mode (see Annex F) or the Deblocking Filter mode (see Annex J) is in use, and shall not be present for the first macroblock of a picture. Also, encoders shall not allow an MCBPC code for macroblock type 5 to immediately follow seven consecutive zeros in the bitstream (as can be caused by particular INTRADC codes followed by COD = 0), in order to prevent start code emulation. Codes for macroblock type 5 can be preceded by stuffing when necessary to fulfill this requirement (for macroblocks other than the first of a picture).

Table 9/H.263 – Macroblock types and included data elements for normal pictures

Picture type	MB type	Name	COD	MCBPC	CBPY	DQUANT	MVD	MVD ₂₋₄
INTER	Not coded	–	X					
INTER	0	INTER	X	X	X		X	
INTER	1	INTER+Q	X	X	X	X	X	
INTER	2	INTER4V	X	X	X		X	X
INTER	3	INTRA	X	X	X			
INTER	4	INTRA+Q	X	X	X	X		
INTER	5	INTER4V+Q	X	X	X	X	X	X
INTER	Stuffing	–	X	X				
INTRA	3	INTRA		X	X			
INTRA	4	INTRA+Q		X	X	X		
INTRA	Stuffing	–		X				

NOTE – "X" means that the item is present in the macroblock.

Table 10/H.263 – Macroblock types and included data elements for PB-frames

Picture type	MB type	Name	COD	MCBPC	MODB	CBPY	CBPB	DQUANT	MVD	MVDB	MVD ₂₋₄
INTER	Not coded	–	X								
INTER	0	INTER	X	X	X	X	(X)		X	(X)	
INTER	1	INTER+Q	X	X	X	X	(X)	X	X	(X)	
INTER	2	INTER4V	X	X	X	X	(X)		X	(X)	X
INTER	3	INTRA	X	X	X	X	(X)		X	(X)	
INTER	4	INTRA+Q	X	X	X	X	(X)	X	X	(X)	
INTER	5	INTER4V+Q	X	X	X	X	(X)	X	X	(X)	X
INTER	Stuffing	–	X	X							

NOTE 1 – "X" means that the item is present in the macroblock.
 NOTE 2 – CBPB and MVDB are only present if indicated by MODB.
 NOTE 3 – B-blocks are always coded in INTER mode, even if the macroblock type of the PB-macroblock indicates INTRA.

The coded block pattern for chrominance signifies C_B and/or C_R blocks when at least one non-INTRADC transform coefficient is transmitted (INTRADC is the dc-coefficient for INTRA blocks, see 5.4.1), unless the optional Advanced INTRA Coding mode is in use. CBPC_N = 1 if any non-INTRADC coefficient is present for block N, else 0, for CBPC₅ and CBPC₆ in the coded block pattern. If Advanced INTRA Coding is in use, the usage is similar, but the INTRADC coefficient is indicated in the same way as the other coefficients (see Annex I). Block numbering is given in Figure 5. When MCBPC = Stuffing, the remaining part of the macroblock layer is skipped. In this case, the preceding COD = 0 is not related to any coded or not-coded macroblock and therefore the macroblock number is not incremented. For P-pictures, multiple stuffings are accomplished by multiple sets of COD = 0 and MCBPC = Stuffing. See Tables 7 and 8.

5.3.3 Macroblock mode for B-blocks (MODB) (Variable length)

MODB is present for MB-type 0-4 if PTYPE indicates "PB-frame" and is a variable length codeword indicating whether CBPB is present (indicates that B-coefficients are transmitted for this macroblock) and/or MVDB is present. In Table 11 the codewords for MODB are defined. MODB is coded differently for Improved PB-frames, as specified in Annex M.

Table 11/H.263 – VLC table for MODB

Index	CBPB	MVDB	Number of bits	Code
0			1	0
1		X	2	10
2	X	X	2	11

NOTE – "X" means that the item is present in the macroblock.

5.3.4 Coded Block Pattern for B-blocks (CBPB) (6 bits)

CBPB is only present in PB-frames mode if indicated by MODB. CBPB_N = 1 if any coefficient is present for B-block N, else 0, for each bit CBPB_N in the coded block pattern. Block numbering is given in Figure 5, the utmost left bit of CBPB corresponding with block number 1.

5.3.5 Coded Block Pattern for luminance (CBPY) (Variable length)

Variable length codeword giving a pattern number signifying those Y blocks in the macroblock for which at least one non-INTRADC transform coefficient is transmitted (INTRADC is the dc-coefficient for INTRA blocks, see 5.4.1), unless the Advanced INTRA Coding mode is in use. If

Advanced INTRA Coding is in use, INTRADC is indicated in the same manner as the other coefficients (see Annex I).

$CBPY_N = 1$ if any non-INTRADC coefficient is present for block N, else 0, for each bit $CBPY_N$ in the coded block pattern. Block numbering is given in Figure 5, the utmost left bit of CBPY corresponding with block number 1. For a certain pattern $CBPY_N$, different codewords are used for INTER and INTRA macroblocks as defined in Table 12.

Table 12/H.263 – VLC table for CBPY

Index	CBPY(INTRA) (12, 34)	CBPY(INTER) (12, 34)	Number of bits	Code
0	00 00	11 11	4	0011
1	00 01	11 10	5	0010 1
2	00 10	11 01	5	0010 0
3	00 11	11 00	4	1001
4	01 00	10 11	5	0001 1
5	01 01	10 10	4	0111
6	01 10	10 01	6	0000 10
7	01 11	10 00	4	1011
8	10 00	01 11	5	0001 0
9	10 01	01 10	6	0000 11
10	10 10	01 01	4	0101
11	10 11	01 00	4	1010
12	11 00	00 11	4	0100
13	11 01	00 10	4	1000
14	11 10	00 01	4	0110
15	11 11	00 00	2	11

5.3.6 Quantizer Information (DQUANT) (2 bits/Variable Length)

If the Modified Quantization mode is not in use, DQUANT is a two-bit code to define a change in QUANT. In Table 13 the differential values for the different codewords are given. QUANT ranges from 1 to 31; if the value for QUANT after adding the differential value is less than 1 or greater than 31, it is clipped to 1 and 31 respectively. If the Modified Quantization mode is in use, DQUANT is a variable length codeword as specified in Annex T.

Table 13/H.263 – DQUANT codes and differential values for QUANT

Index	Differential value	DQUANT
0	-1	00
1	-2	01
2	1	10
3	2	11

5.3.7 Motion Vector Data (MVD) (Variable length)

MVD is included for all INTER macroblocks (in PB-frames mode also for INTRA macroblocks) and consists of a variable length codeword for the horizontal component followed by a variable length codeword for the vertical component. Variable length codes are given in Table 14. If the Unrestricted Motion Vector mode is used and PLUSPTYPE is present, motion vectors are coded using Table D.3 instead of Table 14 (see Annex D).

Table 14/H.263 – VLC table for MVD

Index	Vector	Differences	Bit number	Codes
0	-16	16	13	0000 0000 0010 1
1	-15.5	16.5	13	0000 0000 0011 1
2	-15	17	12	0000 0000 0101
3	-14.5	17.5	12	0000 0000 0111
4	-14	18	12	0000 0000 1001
5	-13.5	18.5	12	0000 0000 1011
6	-13	19	12	0000 0000 1101
7	-12.5	19.5	12	0000 0000 1111
8	-12	20	11	0000 0001 001
9	-11.5	20.5	11	0000 0001 011
10	-11	21	11	0000 0001 101
11	-10.5	21.5	11	0000 0001 111
12	-10	22	11	0000 0010 001
13	-9.5	22.5	11	0000 0010 011
14	-9	23	11	0000 0010 101
15	-8.5	23.5	11	0000 0010 111
16	-8	24	11	0000 0011 001
17	-7.5	24.5	11	0000 0011 011
18	-7	25	11	0000 0011 101
19	-6.5	25.5	11	0000 0011 111

Table 14/H.263 – VLC table for MVD

Index	Vector	Differences	Bit number	Codes
20	-6	26	11	0000 0100 001
21	-5.5	26.5	11	0000 0100 011
22	-5	27	10	0000 0100 11
23	-4.5	27.5	10	0000 0101 01
24	-4	28	10	0000 0101 11
25	-3.5	28.5	8	0000 0111
26	-3	29	8	0000 1001
27	-2.5	29.5	8	0000 1011
28	-2	30	7	0000 111
29	-1.5	30.5	5	0001 1
30	-1	31	4	0011
31	-0.5	31.5	3	011
32	0		1	1
33	0.5	-31.5	3	010
34	1	-31	4	0010
35	1.5	-30.5	5	0001 0
36	2	-30	7	0000 110
37	2.5	-29.5	8	0000 1010
38	3	-29	8	0000 1000
39	3.5	-28.5	8	0000 0110
40	4	-28	10	0000 0101 10
41	4.5	-27.5	10	0000 0101 00
42	5	-27	10	0000 0100 10
43	5.5	-26.5	11	0000 0100 010
44	6	-26	11	0000 0100 000
45	6.5	-25.5	11	0000 0011 110
46	7	-25	11	0000 0011 100
47	7.5	-24.5	11	0000 0011 010
48	8	-24	11	0000 0011 000
49	8.5	-23.5	11	0000 0010 110
50	9	-23	11	0000 0010 100
51	9.5	-22.5	11	0000 0010 010
52	10	-22	11	0000 0010 000
53	10.5	-21.5	11	0000 0001 110
54	11	-21	11	0000 0001 100
55	11.5	-20.5	11	0000 0001 010
56	12	-20	11	0000 0001 000
57	12.5	-19.5	12	0000 0000 1110

Table 14/H.263 – VLC table for MVD

Index	Vector	Differences	Bit number	Codes
58	13	-19	12	0000 0000 1100
59	13.5	-18.5	12	0000 0000 1010
60	14	-18	12	0000 0000 1000
61	14.5	-17.5	12	0000 0000 0110
62	15	-17	12	0000 0000 0100
63	15.5	-16.5	13	0000 0000 0011 0

5.3.8 Motion Vector Data (MVD₂₋₄) (Variable length)

The three codewords MVD₂₋₄ are included if indicated by PTYPE and by MCBPC, and consist each of a variable length codeword for the horizontal component followed by a variable length codeword for the vertical component of each vector. Variable length codes are given in Table 14. MVD₂₋₄ are only present when in Advanced Prediction mode (see Annex F) or Deblocking Filter mode (see Annex J).

5.3.9 Motion Vector Data for B-macroblock (MVDB) (Variable length)

MVDB is only present in PB-frames or Improved PB-frames mode if indicated by MODB, and consists of a variable length codeword for the horizontal component followed by a variable length codeword for the vertical component of each vector. Variable length codes are given in Table 14. For the use of MVDB, refer to Annexes G and M.

5.4 Block layer

If not in PB-frames mode, a macroblock comprises four luminance blocks and one of each of the two colour difference blocks (see Figure 5). The structure of the block layer is shown in Figure 11. INTRADC is present for every block of the macroblock if MCBPC indicates macroblock type 3 or 4 (see Tables 7 and 8). TCOEF is present if indicated by MCBPC or CBPY.

In PB-frames mode, a macroblock comprises twelve blocks. First the data for the six P-blocks is transmitted as in the default H.263 mode, then the data for the six B-blocks. INTRADC is present for every P-block of the macroblock if MCBPC indicates macroblock type 3 or 4 (see Tables 7 and 8). INTRADC is not present for B-blocks. TCOEF is present for P-blocks if indicated by MCBPC or CBPY. TCOEF is present for B-blocks if indicated by CBPB.

For coding of the symbols in the Syntax-based Arithmetic Coding mode, refer to Annex E.



Figure 11/H.263 – Structure of block layer

5.4.1 DC coefficient for INTRA blocks (INTRADC) (8 bits)

A codeword of 8 bits. The code 0000 0000 is not used. The code 1000 0000 is not used, the reconstruction level of 1024 being coded as 1111 1111 (see Table 15).

Table 15/H.263 – Reconstruction levels for INTRA-mode DC coefficient

Index	FLC	Reconstruction level into inverse transform
0	0000 0001 (1)	8
1	0000 0010 (2)	16
2	0000 0011 (3)	24
...
126	0111 1111 (127)	1016
127	1111 1111 (255)	1024
128	1000 0001 (129)	1032
...
252	1111 1101 (253)	2024
253	1111 1110 (254)	2032

5.4.2 Transform Coefficient (TCOEF) (Variable length)

The most commonly occurring EVENTS are coded with the variable length codes given in Table 16. The last bit "s" denotes the sign of the level, "0" for positive and "1" for negative.

An EVENT is a combination of a last non-zero coefficient indication (LAST; "0": there are more non-zero coefficients in this block, "1": this is the last non-zero coefficient in this block), the number of successive zeros preceding the coded coefficient (RUN), and the non-zero value of the coded coefficient (LEVEL).

The remaining combinations of (LAST, RUN, LEVEL) are coded with a 22-bit word consisting of 7 bits ESCAPE, 1 bit LAST, 6 bits RUN and 8 bits LEVEL. Use of this 22-bit word for encoding the combinations listed in Table 16 is not prohibited. For the 8-bit word for LEVEL, the code 0000 0000 is forbidden, and the code 1000 0000 is forbidden unless the Modified Quantization mode is in use (see Annex T). The codes for RUN and for LEVEL are given in Table 17.

Table 16/H.263 – VLC table for TCOEF

Index	LAST	RUN	LEVEL	Bits	VLC code
0	0	0	1	3	10s
1	0	0	2	5	1111s
2	0	0	3	7	0101 01s
3	0	0	4	8	0010 111s
4	0	0	5	9	0001 1111s
5	0	0	6	10	0001 0010 1s
6	0	0	7	10	0001 0010 0s
7	0	0	8	11	0000 1000 01s
8	0	0	9	11	0000 1000 00s
9	0	0	10	12	0000 0000 111s

Table 16/H.263 – VLC table for TCOEF

Index	LAST	RUN	 LEVEL 	Bits	VLC code
10	0	0	11	12	0000 0000 110s
11	0	0	12	12	0000 0100 000s
12	0	1	1	4	110s
13	0	1	2	7	0101 00s
14	0	1	3	9	0001 1110s
15	0	1	4	11	0000 0011 11s
16	0	1	5	12	0000 0100 001s
17	0	1	6	13	0000 0101 0000s
18	0	2	1	5	1110s
19	0	2	2	9	0001 1101s
20	0	2	3	11	0000 0011 10s
21	0	2	4	13	0000 0101 0001s
22	0	3	1	6	0110 1s
23	0	3	2	10	0001 0001 1s
24	0	3	3	11	0000 0011 01s
25	0	4	1	6	0110 0s
26	0	4	2	10	0001 0001 0s
27	0	4	3	13	0000 0101 0010s
28	0	5	1	6	0101 1s
29	0	5	2	11	0000 0011 00s
30	0	5	3	13	0000 0101 0011s
31	0	6	1	7	0100 11s
32	0	6	2	11	0000 0010 11s
33	0	6	3	13	0000 0101 0100s
34	0	7	1	7	0100 10s
35	0	7	2	11	0000 0010 10s
36	0	8	1	7	0100 01s
37	0	8	2	11	0000 0010 01s
38	0	9	1	7	0100 00s
39	0	9	2	11	0000 0010 00s
40	0	10	1	8	0010 110s
41	0	10	2	13	0000 0101 0101s
42	0	11	1	8	0010 101s
43	0	12	1	8	0010 100s
44	0	13	1	9	0001 1100s
45	0	14	1	9	0001 1011s
46	0	15	1	10	0001 0000 1s
47	0	16	1	10	0001 0000 0s

Table 16/H.263 – VLC table for TCOEF

Index	LAST	RUN	 LEVEL 	Bits	VLC code
48	0	17	1	10	0000 1111 1s
49	0	18	1	10	0000 1111 0s
50	0	19	1	10	0000 1110 1s
51	0	20	1	10	0000 1110 0s
52	0	21	1	10	0000 1101 1s
53	0	22	1	10	0000 1101 0s
54	0	23	1	12	0000 0100 010s
55	0	24	1	12	0000 0100 011s
56	0	25	1	13	0000 0101 0110s
57	0	26	1	13	0000 0101 0111s
58	1	0	1	5	0111s
59	1	0	2	10	0000 1100 1s
60	1	0	3	12	0000 0000 101s
61	1	1	1	7	0011 11s
62	1	1	2	12	0000 0000 100s
63	1	2	1	7	0011 10s
64	1	3	1	7	0011 01s
65	1	4	1	7	0011 00s
66	1	5	1	8	0010 011s
67	1	6	1	8	0010 010s
68	1	7	1	8	0010 001s
69	1	8	1	8	0010 000s
70	1	9	1	9	0001 1010s
71	1	10	1	9	0001 1001s
72	1	11	1	9	0001 1000s
73	1	12	1	9	0001 0111s
74	1	13	1	9	0001 0110s
75	1	14	1	9	0001 0101s
76	1	15	1	9	0001 0100s
77	1	16	1	9	0001 0011s
78	1	17	1	10	0000 1100 0s
79	1	18	1	10	0000 1011 1s
80	1	19	1	10	0000 1011 0s
81	1	20	1	10	0000 1010 1s
82	1	21	1	10	0000 1010 0s
83	1	22	1	10	0000 1001 1s
84	1	23	1	10	0000 1001 0s
85	1	24	1	10	0000 1000 1s

Table 16/H.263 – VLC table for TCOEF

Index	LAST	RUN	LEVEL	Bits	VLC code
86	1	25	1	11	0000 0001 11s
87	1	26	1	11	0000 0001 10s
88	1	27	1	11	0000 0001 01s
89	1	28	1	11	0000 0001 00s
90	1	29	1	12	0000 0100 100s
91	1	30	1	12	0000 0100 101s
92	1	31	1	12	0000 0100 110s
93	1	32	1	12	0000 0100 111s
94	1	33	1	13	0000 0101 1000s
95	1	34	1	13	0000 0101 1001s
96	1	35	1	13	0000 0101 1010s
97	1	36	1	13	0000 0101 1011s
98	1	37	1	13	0000 0101 1100s
99	1	38	1	13	0000 0101 1101s
100	1	39	1	13	0000 0101 1110s
101	1	40	1	13	0000 0101 1111s
102	ESCAPE			7	0000 011

Table 17/H.263 – FLC table for RUNs and LEVELs

Index	Run	Code
0	0	000 000
1	1	000 001
2	2	000 010
.	.	.
.	.	.
63	63	111 111

Index	Level	Code
–	–128	see text
0	–127	1000 0001
.	.	.
125	–2	1111 1110
126	–1	1111 1111
–	0	FORBIDDEN
127	1	0000 0001
128	2	0000 0010
.	.	.
253	127	0111 1111

6 Decoding process

6.1 Motion compensation

In this clause, the motion compensation for the default H.263 prediction mode is described. For a description of motion compensation in the Unrestricted Motion Vector mode, refer to Annex D. For a description of motion compensation in the Advanced Prediction mode, refer to Annex F. For a description of motion compensation in the Reduced-Resolution Update mode, refer to Annex Q.

6.1.1 Differential motion vectors

The macroblock vector is obtained by adding predictors to the vector differences indicated by MVD (see Table 14 and Table D.3). For differential coding with four vectors per macroblock, refer to Annex F. In case of one vector per macroblock, the candidate predictors for the differential coding are taken from three surrounding macroblocks as indicated in Figure 12. The predictors are calculated separately for the horizontal and vertical components.

In the special cases at the borders of the current GOB, slice, or picture, the following decision rules are applied in increasing order:

- 1) When the corresponding macroblock was coded in INTRA mode (if not in PB-frames mode with bidirectional prediction) or was not coded (COD = 1), the candidate predictor is set to zero.
- 2) The candidate predictor MV1 is set to zero if the corresponding macroblock is outside the picture or the slice (at the left side).
- 3) Then, the candidate predictors MV2 and MV3 are set to MV1 if the corresponding macroblocks are outside the picture (at the top) or outside the GOB (at the top) if the GOB header of the current GOB is non-empty; or outside the slice when in Slice Structured mode.
- 4) Then, the candidate predictor MV3 is set to zero if the corresponding macroblock is outside the picture (at the right side).

For each component, the predictor is the median value of the three candidate predictors for this component.

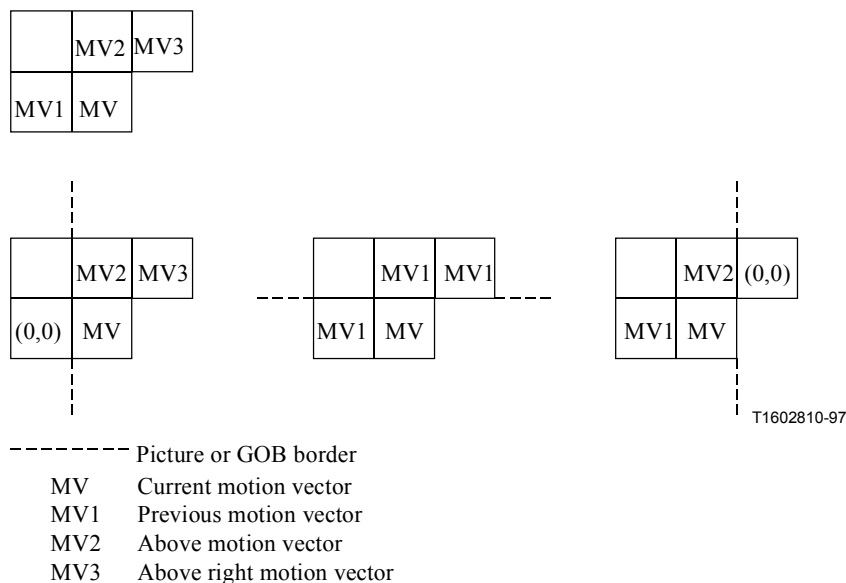


Figure 12/H.263 – Motion vector prediction

Advantage is taken of the fact that the range of motion vector component values is constrained. Each VLC word for MVD represents a pair of difference values. Only one of the pair will yield a macroblock vector component falling within the permitted range $[-16, 15.5]$. A positive value of the horizontal or vertical component of the motion vector signifies that the prediction is formed from pixels in the previous picture which are spatially to the right or below the pixels being predicted. If the unrestricted motion vector mode (see Annex D) is used, the decoding of motion vectors shall be performed as specified in D.2.

The motion vector is used for all pixels in all four luminance blocks in the macroblock. Motion vectors for both chrominance blocks are derived by dividing the component values of the macroblock vector by two, due to the lower chrominance format. The component values of the resulting quarter pixel resolution vectors are modified towards the nearest half pixel position as indicated in Table 18.

Table 18/H.263 – Modification of quarter-pixel resolution chrominance vector components

Quarter-pixel position	0	1/4	1/2	3/4	1
Resulting position	0	1/2	1/2	1/2	1

6.1.2 Interpolation for subpixel prediction

Half pixel values are found using bilinear interpolation as described in Figure 13. "/" indicates division by truncation.

The value of RCONTROL is equal to the value of the rounding type (RTYPE) bit (bit 6) in MPPTYPE (see 5.1.4.3), when the Source Format field (bits 6-8) in PTYPE indicates "extended PTYPE". Otherwise RCONTROL has an implied value of 0. Regardless of the RTYPE bit, the value of RCONTROL is set to 0 for the B-part of Improved PB frames (see Annex M).

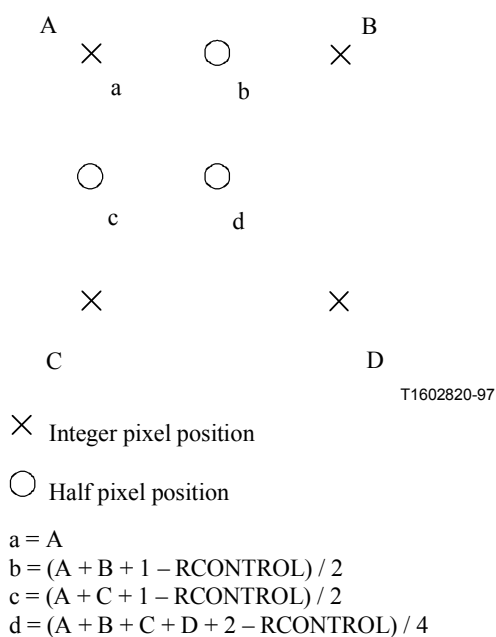


Figure 13/H.263 – Half-pixel prediction by bilinear interpolation

6.2 Coefficients decoding

6.2.1 Inverse quantization

The inverse quantization process is described in this subclause, except for when the optional Advanced INTRA Coding mode is in use (see Annex I). If LEVEL = "0", the reconstruction level REC = "0". The reconstruction level of INTRADC is given by Table 15. The reconstruction levels of all non-zero coefficients other than the INTRADC one are given by the following formulas:

$$\begin{aligned}
 |\text{REC}| &= \text{QUANT} \cdot (2 \cdot |\text{LEVEL}| + 1) && \text{if QUANT} = \text{"odd"} \\
 |\text{REC}| &= \text{QUANT} \cdot (2 \cdot |\text{LEVEL}| + 1) - 1 && \text{if QUANT} = \text{"even"}
 \end{aligned}$$

Note that this process disallows even valued numbers. This has been found to prevent accumulation of IDCT mismatch errors. After calculation of $|\text{REC}|$, the sign is added to obtain REC: $\text{REC} = \text{sign}(\text{LEVEL}) \cdot |\text{REC}|$

$\text{Sign}(\text{LEVEL})$ is given by the last bit of the TCOEF code (see Table 16) or by Table 17.

6.2.2 Clipping of reconstruction levels

After inverse quantization, the reconstruction levels of all coefficients other than the INTRADC one are clipped to the range 2048 to 2047.

6.2.3 Zigzag positioning

The quantized transform coefficients are placed into an 8×8 block according to the sequence given in Figure 14, unless the optional Advanced INTRA Coding mode is in use (see Annex I). Coefficient 1 is the dc-coefficient.

1	2	6	7	15	16	28	29
3	5	8	14	17	27	30	43
4	9	13	18	26	31	42	44
10	12	19	25	32	41	45	54
11	20	24	33	40	46	53	55
21	23	34	39	47	52	56	61
22	35	38	48	51	57	60	62
36	37	49	50	58	59	63	64

Figure 14/H.263 – Zigzag positioning of quantized transform coefficients

6.2.4 Inverse transform

After inverse quantization and zigzag of coefficients, the resulting 8×8 blocks are processed by a separable two-dimensional inverse discrete cosine transform of size 8 by 8. The output from the inverse transform ranges from -256 to $+255$ after clipping to be represented with 9 bits. The transfer function of the inverse transform is given by:

$$f(x, y) = \frac{1}{4} \sum_{u=0}^7 \sum_{v=0}^7 C(u)C(v)F(u, v) \cos \left[\pi(2x+1) \frac{u}{16} \right] \cos \left[\pi(2y+1) \frac{v}{16} \right]$$

with $u, v, x, y = 0, 1, 2, \dots, 7$

where:

x, y = spatial coordinates in the pixel domain;

u, v = coordinates in the transform domain;

$C(u) = 1/\sqrt{2}$ for $u = 0$, otherwise 1;

$C(v) = 1/\sqrt{2}$ for $v = 0$, otherwise 1.

NOTE – Within the block being transformed, $x = 0$ and $y = 0$ refer to the pixel nearest the left and top edges of the picture respectively.

The arithmetic procedures for computing the inverse transform are not defined, but should meet the error tolerance specified in Annex A.

6.3 Reconstruction of blocks

6.3.1 Summation

After motion compensation and coefficients decoding (inverse transform included), a reconstruction is formed for each luminance and chrominance block. For INTRA blocks, the reconstruction is equal to the result of the inverse transformation. For INTER blocks, the reconstruction is formed by summing the prediction and the result of the inverse transformation. The summation is performed on a pixel basis. For the summation in the Reduced-Resolution Update mode, refer to Annex Q.

6.3.2 Clipping

To prevent quantization distortion of transform coefficient amplitudes causing arithmetic overflow in the encoder and decoder loops, clipping functions are inserted. The clipper operates after the summation of prediction and reconstructed prediction error on resulting pixel values less than 0 or greater than 255, changing them to 0 and 255 respectively.

Annex A

Inverse transform accuracy specification

A.1 Generate random integer pixel data values in the range $-L$ to $+H$ according to the random number generator given below ("C" version). Arrange into 8 by 8 blocks. Data set of 10 000 blocks should each be generated for $(L = 256, H = 255)$, $(L = H = 5)$ and $(L = H = 300)$.

A.2 For each 8 by 8 block, perform a separable, orthonormal, matrix multiply, forward discrete cosine transform using at least 64-bit floating point accuracy.

$$F(u, v) = \frac{1}{4} C(u) C(v) \sum_{x=0}^7 \sum_{y=0}^7 f(x, y) \cos \left[\pi(2x+1) \frac{u}{16} \right] \cos \left[\pi(2y+1) \frac{v}{16} \right]$$

with $u, v, x, y = 0, 1, 2, \dots, 7$

where:

x, y = spatial coordinates in the pixel domain;

u, v = coordinates in the transform domain;

$C(u) = 1/\sqrt{2}$ for $u = 0$, otherwise 1;

$C(v) = 1/\sqrt{2}$ for $v = 0$, otherwise 1.

A.3 For each block, round the 64 resulting transformed coefficients to the nearest integer values. Then clip them to the range -2048 to $+2047$. This is the 12-bit input data to the inverse transform.

A.4 For each 8 by 8 block of 12-bit data produced by A.3, perform a separable, orthonormal, matrix multiply, inverse discrete cosine transform (IDCT) using at least 64-bit floating point accuracy. Round the resulting pixels to the nearest integer and clip to the range -256 to $+255$. These blocks of 8×8 pixels are the reference IDCT output data.

A.5 For each 8 by 8 block produced by A.3, apply the IDCT under test and clip the output to the range -256 to $+255$. These blocks of 8×8 pixels are the test IDCT output data.

A.6 For each of the 64 IDCT output pixels, and for each of the 10 000 block data sets generated above, measure the peak, mean and mean square error between the reference and the test data.

- A.7**
- For any pixel, the peak error should not exceed 1 in magnitude.
 - For any pixel, the mean square error should not exceed 0.06.
 - Overall, the mean square error should not exceed 0.02.
 - For any pixel, the mean error should not exceed 0.015 in magnitude.
 - Overall, the mean error should not exceed 0.0015 in magnitude.

A.8 All zeros in shall produce all zeros out.

A.9 Re-run the measurements using exactly the same data values of A.1, but change the sign on each pixel.

```
"C" program for random number generation
/* L and H shall be long, that is 32 bits */
long rand (L,H)
long    L,H;
{
    static long randx = 1;                /* long is 32 bits          */
    static double z = (double) 0x7fffffff;
    long i,j;
    double  x;                            /* double is 64 bits      */
    randx = (randx * 1103515245) + 12345;
    i = randx & 0x7fffffff;               /* keep 30 bits           */
    x = ( (double)i ) / z;                 /* range 0 to 0.99999 ... */
    x *= (L+H+1);                          /* range 0 to < L+H+1    */
    j = x;                                  /* truncate to integer    */
    return(j - L);                         /* range -L to H         */
}
```

Annex B

Hypothetical Reference Decoder

The Hypothetical Reference Decoder (HRD) is defined as follows:

B.1 The HRD and the encoder have the same clock frequency as well as the same picture clock frequency, and are operated synchronously.

B.2 The HRD receiving buffer size is $(B + \text{BPPmaxKb} * 1024 \text{ bits})$ where $(\text{BPPmaxKb} * 1024)$ is the maximum number of bits per picture that has been negotiated for use in the bitstream (see 3.6). The value of B is defined as follows:

$$B = 4 \cdot R_{\max} / PCF$$

where PCF is the effective picture clock frequency, and R_{\max} is the maximum video bit rate during the connection in bits per second. The effective picture clock frequency is the standard CIF picture clock frequency unless a custom PCF is specified in the CPCFC field of the picture header. This value for B is a minimum. An encoder may use a larger value for B, provided the larger number is first negotiated by external means, for example ITU-T Rec. H.245.

The value for R_{\max} depends on the system configuration (for example GSTN or ISDN, single or multi-link) and may be equal to the maximum bit rate supported by the physical link. Negotiation of R_{\max} is done by external means (for example, ITU-T Rec. H.245).

B.3 The HRD is initially empty.

B.4 The HRD buffer is examined at picture clock intervals (1000/PCF ms). If at least one complete coded picture is in the buffer, then all the data for the earliest picture in bitstream order is instantaneously removed (e.g., at t_{n+1} in Figure B.1). Immediately after removing the above data, the buffer occupancy must be less than B . This is a requirement on the coder output bitstream including coded picture data and MCBPC and STUF stuffing but not error correction framing bits, fill indicator (Fi), fill bits or error correction parity information described in Annex H.

For the purposes of this definition, unless the optional Temporal, SNR, and Spatial Scalability mode is in use, a complete coded picture is one normal I- or P-picture, or a PB-frame or Improved PB-frame.

When the Temporal, SNR, and Spatial Scalability mode is in use (see Annex O), each enhancement layer is given an additional HRD, for which a complete coded picture is one EI-, EP-, or B-picture. The base layer buffer shall hold the bits of the picture header as they arrive, until a sufficient amount of the picture header has arrived to determine whether the picture is a base layer or enhancement layer picture, and the number of the enhancement layer. When it can be determined that the arriving picture belongs to an enhancement layer, all bits for that picture shall be instantly transferred to the appropriate enhancement layer HRD, and any later bits which arrive continue to be placed into the enhancement layer HRD until a sufficient amount of some new picture header has arrived to determine that the bitstream should again be re-routed into another HRD buffer. The process of enhancement layer identification is instantaneous and asynchronous, independent of the picture clock interval checking times.

To meet this requirement, the number of bits for the $(n+1)$ th coded picture d_{n+1} must satisfy:

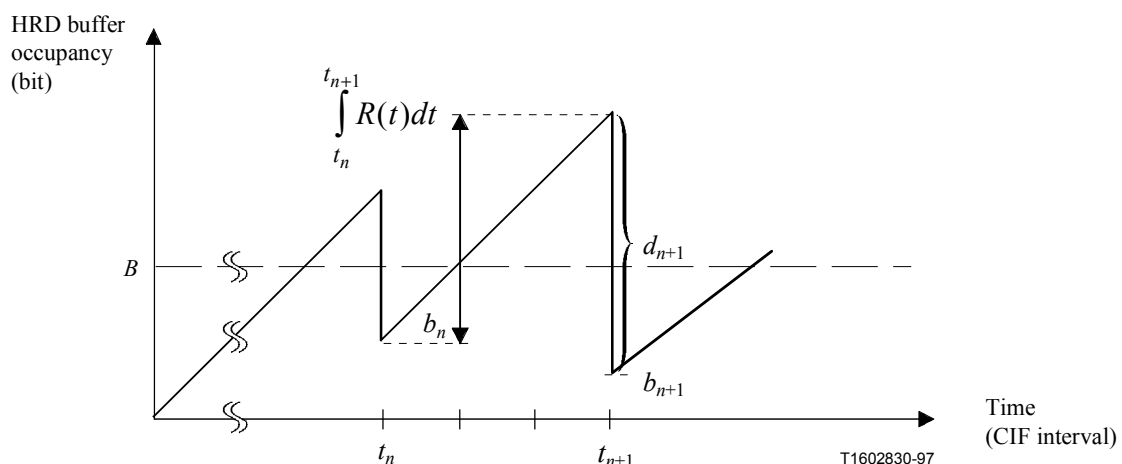
$$d_{n+1} \geq b_n + \int_{t_n}^{t_{n+1}} R(t)dt - B$$

where:

b_n is the buffer occupancy just after time t_n ;

t_n is the time the n th coded picture is removed from the HRD buffer;

$R(t)$ is the video bit rate at time t .



NOTE – Time $(t_{n+1} - t_n)$ is an integer number of CIF picture periods (1/29.97, 2/29.97, 3/29.97,...).

Figure B.1/H.263 – HRD buffer occupancy

Annex C

Considerations for multipoint

The following facilities are provided to support switched multipoint operation.

C.1 Freeze picture request

Causes the decoder to freeze its displayed picture until a freeze picture release signal is received or a time-out period of at least six seconds has expired. The transmission of this signal is by external means (for example, ITU-T Rec. H.245). Note that a similar command may also be sent using supplemental enhancement information within the picture header of the video bitstream (see L.4).

C.2 Fast update request

Causes the encoder to encode its next picture in INTRA mode with coding parameters such as to avoid buffer overflow. The transmission method for this signal is by external means (for example, ITU-T Rec. H.245).

C.3 Freeze picture release

A signal from an encoder which has responded to a fast update request and allows a decoder to exit from its freeze picture mode and display decoded pictures in the normal manner. This signal is transmitted by PTYPE (see 5.1.3) in the picture header of the first picture coded in response to the fast update request.

C.4 Continuous Presence Multipoint and Video Multiplexing (CPM)

NOTE 1 – Not used for ITU-T Rec. H.324.

In this Recommendation, a negotiable Continuous Presence Multipoint and Video Multiplexing mode is provided in which up to four independent H.263 bitstreams can be multiplexed as independent "Sub-Bitstreams" in one new video bitstream with use of the PSBI, GSBI, SSBI, and ESBI fields. Capability exchange for this mode is done by external means (for example, ITU-T Rec. H.242). Reference Picture Selection back-channel data for responding to independent Sub-Bitstreams is supported using the BCPM and BSBI fields.

When in CPM mode, the CPM field shall be set to "1" in each of the independent H.263 Sub-Bitstreams. Sub-Bitstreams are identified by a stream identifier number using the Sub-Bitstream Indicators (SBIs) in the Picture and GOB or slice and EOSBS headers of each H.263 bitstream. The SBI indicates the number of the H.263 bitstream to which that header and all following information until the next Picture or GOB header or slice header in the composed video bitstream belongs.

Each Sub-Bitstream is considered as a normal H.263 bitstream and shall therefore comply with the capabilities that are exchanged by external means. The information for the different H.263 bitstreams is not transmitted in any special predefined order, an SBI can have any value independent from preceding SBIs and the picture rates for the different H.263 bitstreams may be different. The information in each individual bitstream is also completely independent from the information in the other bitstreams. For example, the GFID codewords in one Sub-Bitstream are not influenced by GFID or PTYPE codewords in other Sub-Bitstreams. Similarly, the mode state inference rules when using an extended picture type (PLUSPTYPE) in the picture header and all other aspects of the video bitstream operation shall operate independently and separately for each Sub-Bitstream.

As a matter of convention, the Sub-Bitstream with the lowest Sub-Bitstream identifier number (sent in SBI) is considered to have the highest priority in situations in which some conflict of resource

requirements may necessitate a choice of priority (unless a different priority convention is established by external means).

In order to mark the end of each Sub-Bitstream of the CPM mode, a syntax is provided as shown in Figure C.1, provided the ability to send this added syntax is first negotiated by external means (while CPM operation was defined in Version 1 of this Recommendation, the end of Sub-Bitstream syntax was added in Version 2 and thus is not considered a part of Version 1 CPM operation). The end of Sub-Bitstream syntax (ESTUF + EOSBS + ESBI) marks the end of each Sub-Bitstream, rather than the end of the entire stream as done by an EOS code.

NOTE 2 – No ability to negotiate CPM Sub-Bitstream operation as defined herein for ITU-T Rec. H.263 was adopted into any ITU-T Recommendations for terminals (such as ITU-T Rec. H.324) prior to the creation of Version 2 of this Recommendation. Thus, any external negotiation of CPM operation adopted into a future H-series terminal Recommendation shall imply support for the end of Sub-Bitstream syntax unless otherwise specified in the H-series terminal Recommendation.

There are three parts to the end of Sub-Bitstream syntax. Following mandatory byte alignment using ESTUF, an EOSBS codeword of 23 bits is sent (corresponding to a GOB header with GN = 30, which is otherwise unused in the syntax, followed by a single zero-valued bit which is reserved for future use). The EOSBS codeword is then followed by a two-bit ESBI codeword indicating which Sub-Bitstream is affected. This pair of codewords signifies that the sending of data for the associated Sub-Bitstream has ended and that any subsequent data sent for the same Sub-Bitstream shall be entirely independent from that which came before the EOSBS. In particular, the next picture for the Sub-Bitstream after the EOSBS code shall not be an INTER picture or any other picture type which can use forward temporal prediction (an I- or EI-picture is allowed, but a P-picture, PB-frame, Improved PB-frame, B-picture, or EP-picture is not).

The syntax of EOSBS and ESBI is described in the following subclauses. ESTUF is described in 5.1.26.

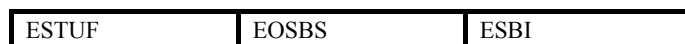


Figure C.1/H.263 – Syntax diagram for End of Sub-Bitstream Indicators

C.4.1 End Of Sub-Bitstream code (EOSBS) (23 bits)

The EOSBS code is a codeword of 23 bits. Its value is 0000 0000 0000 0000 1 11110 0. It is up to the encoder whether to insert this codeword or not. EOSBS should not be sent unless at least one picture header has previously been sent for the same Sub-Bitstream indicated in the following ESBI field, and shall not be sent unless the capability to send EOSBS has been negotiated by external means. EOSBS codes shall be byte aligned. This is achieved by inserting ESTUF before the EOSBS start code such that the first bit of the EOSBS start code is the first (most significant) bit of a byte (see 5.1.26).

The EOSBS code indicates that the sending of data for the indicated Sub-Bitstream has stopped and the Sub-Bitstream has been declared over, until re-started by issuance of another picture start code for that Sub-Bitstream. Subsequent pictures with the same Sub-Bitstream identifier number (ESBI) shall be completely independent of and shall not depend in any way on the pictures which were sent prior to the EOSBS code.

Control and other information associated with the video bitstream in general without specification of which Sub-Bitstreams these codes are meant to apply (such as a Freeze Picture Request or a Fast Update Request sent in ITU-T Rec. H.242) should be presumed to apply to all of the *active* Sub-Bitstreams only. A Sub-Bitstream is considered active if at least one picture start code has been received for the Sub-Bitstream and the last data sent which was applicable to that Sub-Bitstream was not an EOS or EOSBS + ESBI.

C.4.2 Ending Sub-Bitstream Indicator (ESBI) (2 bits)

ESBI is a fixed length codeword of two bits which follows immediately after EOSBS. It indicates the Sub-Bitstream number of the ending Sub-Bitstream. Its value is the natural two-bit binary representation of the Sub-Bitstream number.

Annex D

Unrestricted Motion Vector mode

This annex describes the optional Unrestricted Motion Vector mode of this Recommendation. The capability of this H.263 mode is signalled by external means (for example, ITU-T Rec. H.245). The use of this mode is indicated in PTYPE or PLUSPTYPE.

The range of motion vectors and the VLC table used for coding the motion vector differences for the Unrestricted Motion Vector mode depend on whether the PLUSPTYPE field of the picture header is present. When PLUSPTYPE is present, the range of motion vectors also depends on the picture size and the value of the UUI field in the picture header.

D.1 Motion vectors over picture boundaries

In the default prediction mode of this Recommendation, motion vectors are restricted such that all pixels referenced by them are within the coded picture area (see 4.2.3). In the Unrestricted Motion Vector mode, however, this restriction is removed and therefore motion vectors *are* allowed to point outside the picture. When a pixel referenced by a motion vector is outside the coded picture area, an edge pixel is used instead. This edge pixel is found by limiting the motion vector to the last full-pixel position inside the coded picture area. Limitation of the motion vector is performed on a pixel basis and separately for each component of the motion vector.

For example, if the Unrestricted Motion Vector mode is used for a QCIF picture, the referenced pixel value for the luminance component is given by the following formula:

$$R_{umv}(x, y) = R(x', y')$$

where:

x, y, x', y' = spatial coordinates in the pixel domain

$R_{umv}(x, y)$ = pixel value of the referenced picture at (x, y) when in Unrestricted Motion Vector mode

$R(x', y')$ = pixel value of the referenced picture at (x', y') when in Unrestricted Motion Vector mode

$$x' \begin{cases} = 0 & \text{if } x < 0 \\ = 175 & \text{if } x > 175 \\ = x & \text{otherwise} \end{cases}$$

$$y' \begin{cases} = 0 & \text{if } y < 0 \\ = 143 & \text{if } y > 143 \\ = y & \text{otherwise} \end{cases}$$

and the coded picture area of $R(x', y')$ is $0 \leq x' \leq 175, 0 \leq y' \leq 143$. The given boundaries are integer pixel positions; however, (x', y') can also be a half-pixel position within these boundaries.

D.1.1 Restrictions for motion vector values

If PLUSPTYPE is present in the picture header, the motion vector values are restricted such that no element of the 16×16 or (8×8) region that is selected shall have a horizontal or vertical distance more than 15 pixels outside the coded picture area. Note that this is a smaller extrapolation range than when PLUSPTYPE is not present.

NOTE 1 – When PLUSPTYPE is absent, the extrapolation range is a maximum of 31.5 pixels outside the coded picture area when the Unrestricted Motion Vector mode is used, and 16 pixels outside the coded picture area when the Advanced Prediction mode (see Annex F) is used without the Unrestricted Motion Vector mode.

NOTE 2 – When the Advanced Prediction mode (see Annex F) is in use, the motion vector for each 16×16 or (8×8) region affects a larger areas, due to overlapped block motion compensation. This can cause the effective extrapolation range to increase for the "remote" motion vectors of the Advanced Prediction mode, since the amount of overlapping (4 pixels, or 8 pixels when the Reduced-Resolution Update mode is also in use) adds to the amount of extrapolation required (even though the range of values allowed for each motion vector remains the same as when the Advanced Prediction mode is not in use).

D.2 Extension of the motion vector range

In the default prediction mode, the values for both horizontal and vertical components of the motion vectors are restricted to the range $[-16, 15.5]$ (this is also valid for the forward and backward motion vector components for B-pictures). In the Unrestricted Motion Vector mode, however, the maximum range for vector components is extended.

If the PLUSPTYPE field is not present in the picture header, the motion vector range is extended to $[-31.5, 31.5]$, with the restriction that only values that are within a range of $[-16, 15.5]$ around the predictor for each motion vector component can be reached if the predictor is in the range $[-15.5, 16]$. If the predictor is outside $[-15.5, 16]$, all values within the range $[-31.5, 31.5]$ with the same sign as the predictor plus the zero value can be reached. So, if MV_c is the motion vector component and P_c is the predictor for it, then:

$$\begin{array}{ll} -31.5 \leq MV_c \leq 0 & \text{if } -31.5 \leq P_c \leq -16 \\ -16 + P_c \leq MV_c \leq 15.5 + P_c & \text{if } -15.5 \leq P_c \leq 16 \\ 0 \leq MV_c \leq 31.5 & \text{if } 16.5 \leq P_c \leq 31.5 \end{array}$$

In the Unrestricted Motion Vector mode, the interpretation of Table 14 for MVD, MVD_{2-4} and MVDB is as follows:

- If the predictor for the motion vector component is in the range $[-15.5, 16]$, only the first column of vector differences applies.
- If the predictor for the motion vector component is outside the range $[-15.5, 16]$, the vector difference from Table 14 shall be used that results in a vector component inside the range $[-31.5, 31.5]$ with the same sign as the predictor (including zero).

The predictor for MVD and MVD_{2-4} is defined as the median value of the vector components MV1, MV2 and MV3 as defined in 6.1.1 and F.2. For MVDB, the predictor $P_c = (TR_B \times MV) / TR_D$, where MV represents a vector component for an $8 * 8$ luminance block in a P-picture (see also G.4).

If PLUSPTYPE is present, the motion vector range does not depend on the motion vector prediction value. If the UUI field is set to "1" the motion vector range depends on the picture format. For standardized picture formats up to CIF the range is $[-32, 31.5]$, for those up to 4CIF the range is $[-64, 63.5]$, and for those up to 16CIF the range is $[-128, 127.5]$, and for even larger custom picture formats the range is $[-256, 255.5]$. The horizontal and vertical motion vector ranges may be different for custom picture formats. The horizontal and vertical ranges are specified in Tables D.1 and D.2.

Table D.1/H.263 – Horizontal motion vector range when PLUSPTYPE present and UII = 1

Picture width	Horizontal motion vector range
4, ... , 352	[-32, 31.5]
356, ... , 704	[-64, 63.5]
708, ... , 1408	[-128, 127.5]
1412, ... , 2048	[-256, 255.5]

Table D.2/H.263 – Vertical motion vector range when PLUSPTYPE present and UII = 1

Picture height	Vertical motion vector range
4, ... , 288	[-32, 31.5]
292, ... , 576	[-64, 63.5]
580, ... , 1152	[-128, 127.5]

In the Reduced-Resolution Update mode, the specified range applies to the pseudo motion vectors. This implies that the resulting actual motion vector range is enlarged to approximately double size (see also Annex Q).

If UII is set to "01", the motion vectors are not limited except by their distance to the coded area border as explained in D.1.1. The same limitation applies to the actual motion vectors (not just the pseudo motion vectors) in the Reduced-Resolution Update mode.

To encode the motion vectors when PLUSPTYPE is present, Table D.3 is used to encode the difference between the motion vector and the motion vector prediction. Every entry in Table D.3 has a single value (in contrast to Table 14). The motion vector range and the use of Table D.3 to encode motion vector data apply to all picture types when PLUSPTYPE is present.

Motion vectors differences are always coded as a pair of a horizontal and a vertical component. If a pair equals (0.5, 0.5) six consecutive zeros are produced. To prevent start code emulation, this occurrence shall be followed by one bit set to "1". This corresponds to sending one additional zero motion vector component.

Table D.3/H.263 – Motion Vector Table used when PLUSPTYPE present

Absolute value of vector difference in half-pixel units	Number of bits	Codes
0	1	1
1	3	0s0
"x ₀ " + 2 (2:3)	5	0x ₀ 1s0
"x ₁ x ₀ " + 4 (4:7)	7	0x ₁ 1x ₀ 1s0
"x ₂ x ₁ x ₀ " + 8 (8:15)	9	0x ₂ 1x ₁ 1x ₀ 1s0
"x ₃ x ₂ x ₁ x ₀ " + 16 (16:31)	11	0x ₃ 1x ₂ 1x ₁ 1x ₀ 1s0
"x ₄ x ₃ x ₂ x ₁ x ₀ " + 32 (32:63)	13	0x ₄ 1x ₃ 1x ₂ 1x ₁ 1x ₀ 1s0
"x ₅ x ₄ x ₃ x ₂ x ₁ x ₀ " + 64 (64:127)	15	0x ₅ 1x ₄ 1x ₃ 1x ₂ 1x ₁ 1x ₀ 1s0
"x ₆ x ₅ x ₄ x ₃ x ₂ x ₁ x ₀ " + 128 (128:255)	17	0x ₆ 1x ₅ 1x ₄ 1x ₃ 1x ₂ 1x ₁ 1x ₀ 1s0
"x ₇ x ₆ x ₅ x ₄ x ₃ x ₂ x ₁ x ₀ " + 256 (256:511)	19	0x ₇ 1x ₆ 1x ₅ 1x ₄ 1x ₃ 1x ₂ 1x ₁ 1x ₀ 1s0
"x ₈ x ₇ x ₆ x ₅ x ₄ x ₃ x ₂ x ₁ x ₀ " + 512 (512:1023)	21	0x ₈ 1x ₇ 1x ₆ 1x ₅ 1x ₄ 1x ₃ 1x ₂ 1x ₁ 1x ₀ 1s0
"x ₉ x ₈ x ₇ x ₆ x ₅ x ₄ x ₃ x ₂ x ₁ x ₀ " + 1024 (1024:2047)	23	0x ₉ 1x ₈ 1x ₇ 1x ₆ 1x ₅ 1x ₄ 1x ₃ 1x ₂ 1x ₁ 1x ₀ 1s0
"x ₁₀ x ₉ x ₈ x ₇ x ₆ x ₅ x ₄ x ₃ x ₂ x ₁ x ₀ " + 2048 (2048:4095)	25	0x ₁₀ 1x ₉ 1x ₈ 1x ₇ 1x ₆ 1x ₅ 1x ₄ 1x ₃ 1x ₂ 1x ₁ 1x ₀ 1s0

Table D.3 is a regularly constructed reversible table. Each row represents an interval of motion vector differences in half-pixel units. The bits "...x₁x₀" denote all bits following the leading "1" in the binary representation of the absolute value of the motion vector difference. The bit "s" denotes the sign of the motion vector difference, "0" for positive and "1" for negative. The binary representation of the motion vector difference is interleaved with bits that indicate if the code continues or ends. For example, the motion vector difference -13 has sign s = 1 and binary representation 1x₂x₁x₀ = 1101. It is encoded as 0 x₂1 x₁1 x₀1 s0 = 0 11 01 11 10. The 0 in the second position of the last group of two bits indicates the end of the code.

Annex E

Syntax-based Arithmetic Coding mode

E.1 Introduction

In the Variable Length Coding/Decoding (VLC/VLD) as described in clause 5, a symbol is VLC encoded using a specific table based on the syntax of the coder. This table typically stores lengths and values of the VLC code words. The symbol is mapped to an entry of the table in a table look-up operation, and then the binary code word specified by the entry is sent out normally to a buffer for transmitting to the receiver. In VLD decoding, the received bitstream is matched entry by entry in a specific table based on the syntax of the coder. This table must be the same as the one used in the encoder for encoding the current symbol. The matched entry in the table is then mapped back to the corresponding symbol which is the end result of the VLD decoder and is then used for recovering the video pictures. This VLC/VLD process implies that each symbol must be encoded into a fixed integral number of bits. Removing this restriction of fixed integral number of bits for symbols can lead to reductions of resulting bit rates, which can be achieved by arithmetic coding.

This annex describes the optional Syntax-based Arithmetic Coding (SAC) mode of this Recommendation. In this mode, all the corresponding variable length coding/decoding operations of

this Recommendation are replaced with arithmetic coding/decoding operations. The capability of this H.263 mode is signalled by external means (for example, ITU-T Rec. H.245). The use of this mode is indicated in PTYYPE.

E.2 Specification of SAC encoder

In SAC mode, a symbol is encoded by using a specific array of integers (or a model) based on the syntax of the coder and by calling the following procedure which is specified in C language.

```

#define q1 16384
#define q2 32768
#define q3 49152
#define top 65535

static long low, high, opposite_bits, length;
void encode_a_symbol(int index, int cumul_freq[ ])
{
    length = high - low + 1;
    high = low - 1 + (length * cumul_freq[index]) / cumul_freq[0];
    low += (length * cumul_freq[index+1]) / cumul_freq[0];
    for ( ; ; ) {
        if (high < q2) {
            send out a bit "0" to PSC_FIFO;
            while (opposite_bits > 0) {
                send out a bit "1" to PSC_FIFO;
                opposite_bits--;
            }
        }
        else if (low >= q2) {
            send out a bit "1" to PSC_FIFO;
            while (opposite_bits > 0) {
                send out a bit "0" to PSC_FIFO;
                opposite_bits--;
            }
            low -= q2;
            high -= q2;
        }
        else if (low >= q1 && high < q3) {
            opposite_bits += 1;
            low -= q1;
            high -= q1;
        }
        else break;

        low *= 2;
        high = 2 * high + 1;
    }
}

```

The values of low, high and opposite_bits are initialized to 0, top and 0, respectively. PSC_FIFO is a FIFO for buffering the output bits from the arithmetic encoder. The model is specified through cumul_freq[], and the symbol is specified using its index in the model.

E.3 Specification of SAC decoder

In SAC decoder, a symbol is decoded by using a specific model based on the syntax and by calling the following procedure which is specified in C language.

```
static long          low, high, code_value, bit, length, index, cum;
int  decode_a_symbol(int cumul_freq[ ])
{
    length = high - low + 1;
    cum = (-1 + (code_value - low + 1) * cumul_freq[0]) / length;
    for (index = 1; cumul_freq[index] > cum; index++);
    high = low - 1 + (length * cumul_freq[index-1]) / cumul_freq[0];
    low += (length * cumul_freq[index]) / cumul_freq[0];
    for ( ; ; ) {
        if (high < q2);
        else if (low >= q2) {
            code_value -= q2;
            low -= q2;
            high -= q2;
        }
        else if (low >= q1 && high < q3) {
            code_value -= q1;
            low -= q1;
            high -= q1;
        }
        else break;

        low *= 2;
        high = 2 * high + 1;
        get bit from PSC_FIFO;
        code_value = 2 * code_value + bit;
    }
    return (index-1);
}
```

Again the model is specified through `cumul_freq[]`. The decoded symbol is returned through its `index` in the model. `PSC_FIFO` is a FIFO for buffering the incoming bitstream. The decoder is initialized to start decoding an arithmetic coded bitstream by calling the following procedure.

```
void decoder_reset( )
{
    code_value = 0;
    low = 0;
    high = top;
    for (int i = 1; i <= 16; i++) {
        get bit from PSC_FIFO;
        code_value = 2 * code_value + bit;
    }
}
```

E.4 Syntax

As in VLC table mode of this Recommendation, the syntax of the symbols is partitioned into four layers: Picture, Group of Blocks, Macroblock and Block. The syntax of the top three layers remain exactly the same. The syntax of the Block layer also remains quite similar, but is illustrated in Figure E.1.

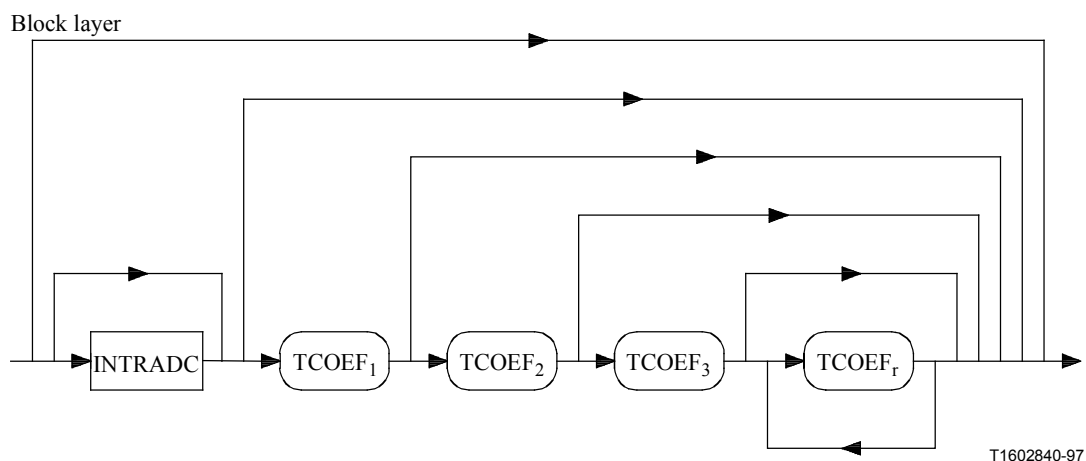


Figure E.1/H.263 – Structure of SAC Block layer

In Figure E.1, TCOEF1, TCOEF2, TCOEF3 and TCOEFr are LAST-RUN-LEVEL symbols as defined in 5.4.2, and are the possible 1st, 2nd, 3rd and rest of the symbols, respectively. TCOEF1, TCOEF2, TCOEF3 and TCOEFr are only present when one, two, three or more coefficients are present in the block layer, respectively.

E.5 PSC_FIFO

PSC_FIFO in encoder or decoder is a FIFO of size > 17 bits. In PSC_FIFO of encoder, illegal emulations of PSC and GBSC are located and are avoided by stuffing a "1" after each successive appearance of 14 "0"s (which are not part of PSC or GBSC). In PSC_FIFO of the decoder, the first "1" after each string of 14 "0"s is deleted; if instead a string of 14 "0"s is followed by a "0", it indicates that a legal PSC or GBSC is detected. The exact location of the PSC or GBSC is determined by the next "1" following the string of zeros.

E.6 Header layer symbols

The header layers of the syntax are considered to be those syntactical elements above the block and macroblock layers (see Figure 6 and the syntax specification in the text). The header levels of the basic Version 1 syntax can form three possible strings, (PSTUF)--PSC--TR--PTYPE--PQUANT--CPM--(PSBI)--(TRB-DBQUANT)--PEI--(PSUPP--PEI--...), (GSTUF)--GBSC--GN--(GSBI)--GFID--GQUANT, and (ESTUF)--(EOS)--(PSTUF). In the revised syntax of Version 2, the header levels of the syntax can have other structures (see Figure 6 and the syntax specification in the text). The header level syntax strings are directly sent to PSC_FIFO as in the normal VLC table mode of this Recommendation at encoder side, and are directly sent out from PSC_FIFO in the decoder after a legal PSC, GBSC, SSC, EOS, or EOSBS is detected.

If a header is not the first in a video session, the arithmetic encoder needs to be reset before sending the header by calling the following procedure. This procedure shall also be called at the end of a video session if (ESTUF)--EOS [or (ESTUF)--EOSBS for the Sub-Bitstream for which the last header was sent] is not sent.

```
void encoder_flush( )
{
    opposite_bits++;
    if (low < q1) {
        send out a bit "0" to PSC_FIFO;
        while (opposite_bits > 0) {
            send out a bit "1" to PSC_FIFO;
            opposite_bits- -;
        }
    }
}
```

```

    }
  }
  else {
    send out a bit "1" to PSC_FIFO;
    while (opposite_bits > 0) {
      send out a bit "0" to PSC_FIFO;
      opposite_bits--;
    }
  }
  low = 0;
  high = top;
}

```

In the decoder, after each fixed length symbol string, procedure `decoder_reset` is called.

E.7 Macroblock and Block layer symbols

Models for the macroblock and block layer symbols are included in E.8. The indices as given in the VLC tables of clause 5 are used for the indexing of the integers in the models.

The model for COD in P-pictures is named `cumf_COD`. The index for COD being "0" is 0, and is 1 for COD being "1". The model for MCBPC in P-pictures is named `cumf_MCBPC_no4MVQ` unless PLUSPTYPE is present in the picture header and either the Advanced Prediction mode (Annex F) or the Deblocking Filter mode (Annex J) is in use, in which case the model is named `cumf_MCBPC_4MVQ`. The indexes for MCBPC are defined in Table 7 for I-pictures and Table 8 for P-pictures. The model for MCBPC in I-pictures is named by `cumf_MCBPC_intra`.

The model for MODB is `cumf_MODB_G` if Annex G is used or `cumf_MODB_M` if Annex M is used. The indexes for MODB are defined in Table 11 or Table M.1, respectively. The model for CBPB_n, n = 1, 2, ..., 4, is `cumf_YCBPB`, and the model for CBPB_n, n = 5, 6, is `cumf_UVCBPB`, with index 0 for CBPB_n = 0 and index 1 for CBPB_n = 1.

The model for CBPY is `cumf_CBPY` in INTER macroblocks and `cumf_CBPY_intra` in INTRA macroblocks. The model for DQUANT is `cumf_DQUANT`. The indexing for CBPY and DQUANT is defined in Tables 12 and 13, respectively.

The model for MVD, MVD_{2,4} and MVDB is `cumf_MVD` and the model for INTRADC is `cumf_INTRADC`. The indexing is defined in Tables 14 and 15, respectively.

A non-escaped TCOEF consists of a symbol for TCOEF1/2/3/r followed by a symbol, SIGN, for the sign of the TCOEF. The models for TCOEF1, TCOEF2, TCOEF3 and TCOEF_r in INTER blocks are `cumf_TCOEF1`, `cumf_TCOEF2`, `cumf_TCOEF3`, `cumf_TCOEFr`. The models for INTRA blocks are `cumf_TCOEF1_intra`, `cumf_TCOEF2_intra`, `cumf_TCOEF3_intra`, `cumf_TCOEFr_intra`. For all TCOEFs the indexing is defined in Table 16. The model for SIGN is `cumf_SIGN`. The indexing for SIGN is 0 for positive sign and 1 for negative sign.

The models for LAST, RUN, LEVEL after ESCAPE are `cumf_LAST` (`cumf_LAST_intra`), `cumf_RUN` (`cumf_RUN_intra`), `cumf_LEVEL` (`cumf_LEVEL_intra`) for INTER (INTRA) blocks. The indexing for LAST is 0 for LAST = 0 and 1 for LAST = 1, while the indexing for RUN and LEVEL is defined in Table 17.

The model for INTRA_MODE is `cumf_INTRA_AC_DC`. The indexing is defined in Table I.1.

E.8 SAC models

```
int cumf_COD[3]={16383, 6849, 0};
```

```
int cumf_MCBPC_no4MVQ[22]={16383, 4105, 3088, 2367, 1988, 1621, 1612, 1609, 1608, 496,
353, 195, 77, 22, 17, 12, 5, 4, 3, 2, 1, 0};
```

```

int cumf_MCBPC_4MVQ[26]={16383, 6880, 6092, 5178, 4916, 3965, 3880, 3795, 3768, 1491,
1190, 889, 655, 442, 416, 390, 360, 337, 334, 331, 327, 326, 88, 57, 26, 0};
int cumf_MCBPC_intra[10]={16383, 7410, 6549, 5188, 442, 182, 181, 141, 1, 0};
int cumf_MODB_G[4]={16383, 6062, 2130, 0};
int cumf_MODB_M[7] = {16383, 6717, 4568, 2784, 1370, 655, 0};
int cumf_YCBPB[3]={16383, 6062, 0};
int cumf_UVCBPB[3]={16383, 491, 0};
int cumf_CBPY[17]={16383, 14481, 13869, 13196, 12568, 11931, 11185, 10814, 9796, 9150,
8781, 7933, 6860, 6116, 4873, 3538, 0};
int cumf_CBPY_intra[17]={16383, 13619, 13211, 12933, 12562, 12395, 11913, 11783, 11004,
10782, 10689, 9928, 9353, 8945, 8407, 7795, 0};
int cumf_DQUANT[5]={16383, 12287, 8192, 4095, 0};
int cumf_MVD[65]={16383, 16380, 16369, 16365, 16361, 16357, 16350, 16343, 16339, 16333,
16326, 16318, 16311, 16306, 16298, 16291, 16283, 16272, 16261, 16249, 16235, 16222, 16207,
16175, 16141, 16094, 16044, 15936, 15764, 15463, 14956, 13924, 11491, 4621, 2264, 1315, 854,
583, 420, 326, 273, 229, 196, 166, 148, 137, 123, 114, 101, 91, 82, 76, 66, 59, 53, 46, 36, 30, 26,
24, 18, 14, 10, 5, 0};
int cumf_INTRADC[255]={16383, 16380, 16379, 16378, 16377, 16376, 16370, 16361, 16360,
16359, 16358, 16357, 16356, 16355, 16343, 16238, 16237, 16236, 16230, 16221, 16220, 16205,
16190, 16169, 16151, 16130, 16109, 16094, 16070, 16037, 16007, 15962, 15938, 15899, 15854,
15815, 15788, 15743, 15689, 15656, 15617, 15560, 15473, 15404, 15296, 15178, 15106, 14992,
14868, 14738, 14593, 14438, 14283, 14169, 14064, 14004, 13914, 13824, 13752, 13671, 13590,
13515, 13458, 13380, 13305, 13230, 13143, 13025, 12935, 12878, 12794, 12743, 12656, 12596,
12521, 12443, 12359, 12278, 12200, 12131, 12047, 12002, 11948, 11891, 11828, 11744, 11663,
11588, 11495, 11402, 11288, 11204, 11126, 11039, 10961, 10883, 10787, 10679, 10583, 10481,
10360, 10227, 10113, 9961, 9828, 9717, 9584, 9485, 9324, 9112, 9019, 8908, 8766, 8584, 8426,
8211, 7920, 7663, 7406, 7152, 6904, 6677, 6453, 6265, 6101, 5904, 5716, 5489, 5307, 5056, 4850,
4569, 4284, 3966, 3712, 3518, 3342, 3206, 3048, 2909, 2773, 2668, 2596, 2512, 2370, 2295, 2232,
2166, 2103, 2022, 1956, 1887, 1830, 1803, 1770, 1728, 1674, 1635, 1599, 1557, 1500, 1482, 1434,
1389, 1356, 1317, 1284, 1245, 1200, 1179, 1140, 1110, 1092, 1062, 1044, 1035, 1014, 1008, 993,
981, 954, 936, 912, 894, 876, 864, 849, 828, 816, 801, 792, 777, 756, 732, 690, 660, 642, 615, 597,
576, 555, 522, 489, 459, 435, 411, 405, 396, 387, 375, 360, 354, 345, 344, 329, 314, 293, 278, 251,
236, 230, 224, 215, 214, 208, 199, 193, 184, 178, 169, 154, 127, 100, 94, 73, 37, 36, 35, 34, 33, 32,
31, 30, 29, 28, 27, 26, 20, 19, 18, 17, 16, 15, 9, 0};
int cumf_TCOEF1[104]={16383, 13455, 12458, 12079, 11885, 11800, 11738, 11700, 11681,
11661, 11651, 11645, 11641, 10572, 10403, 10361, 10346, 10339, 10335, 9554, 9445, 9427, 9419,
9006, 8968, 8964, 8643, 8627, 8624, 8369, 8354, 8352, 8200, 8192, 8191, 8039, 8036, 7920, 7917,
7800, 7793, 7730, 7727, 7674, 7613, 7564, 7513, 7484, 7466, 7439, 7411, 7389, 7373, 7369, 7359,
7348, 7321, 7302, 7294, 5013, 4819, 4789, 4096, 4073, 3373, 3064, 2674, 2357, 2177, 1975, 1798,
1618, 1517, 1421, 1303, 1194, 1087, 1027, 960, 890, 819, 758, 707, 680, 656, 613, 566, 534, 505,
475, 465, 449, 430, 395, 358, 335, 324, 303, 295, 286, 272, 233, 215, 0};
int cumf_TCOEF2[104]={16383, 13582, 12709, 12402, 12262, 12188, 12150, 12131, 12125,
12117, 12113, 12108, 12104, 10567, 10180, 10070, 10019, 9998, 9987, 9158, 9037, 9010, 9005,
8404, 8323, 8312, 7813, 7743, 7726, 7394, 7366, 7364, 7076, 7062, 7060, 6810, 6797, 6614, 6602,
6459, 6454, 6304, 6303, 6200, 6121, 6059, 6012, 5973, 5928, 5893, 5871, 5847, 5823, 5809, 5796,
5781, 5771, 5763, 5752, 4754, 4654, 4631, 3934, 3873, 3477, 3095, 2758, 2502, 2257, 2054, 1869,

```

1715, 1599, 1431, 1305, 1174, 1059, 983, 901, 839, 777, 733, 683, 658, 606, 565, 526, 488, 456, 434, 408, 380, 361, 327, 310, 296, 267, 259, 249, 239, 230, 221, 214, 0};

int cumf_TCOEF3[104]={16383, 13532, 12677, 12342, 12195, 12112, 12059, 12034, 12020, 12008, 12003, 12002, 12001, 10586, 10297, 10224, 10202, 10195, 10191, 9223, 9046, 8999, 8987, 8275, 8148, 8113, 7552, 7483, 7468, 7066, 7003, 6989, 6671, 6642, 6631, 6359, 6327, 6114, 6103, 5929, 5918, 5792, 5785, 5672, 5580, 5507, 5461, 5414, 5382, 5354, 5330, 5312, 5288, 5273, 5261, 5247, 5235, 5227, 5219, 4357, 4277, 4272, 3847, 3819, 3455, 3119, 2829, 2550, 2313, 2104, 1881, 1711, 1565, 1366, 1219, 1068, 932, 866, 799, 750, 701, 662, 605, 559, 513, 471, 432, 403, 365, 336, 312, 290, 276, 266, 254, 240, 228, 223, 216, 206, 199, 192, 189, 0};

int cumf_TCOEFr[104]={16383, 13216, 12233, 11931, 11822, 11776, 11758, 11748, 11743, 11742, 11741, 11740, 11739, 10203, 9822, 9725, 9691, 9677, 9674, 8759, 8609, 8576, 8566, 7901, 7787, 7770, 7257, 7185, 7168, 6716, 6653, 6639, 6276, 6229, 6220, 5888, 5845, 5600, 5567, 5348, 5327, 5160, 5142, 5004, 4900, 4798, 4743, 4708, 4685, 4658, 4641, 4622, 4610, 4598, 4589, 4582, 4578, 4570, 4566, 3824, 3757, 3748, 3360, 3338, 3068, 2835, 2592, 2359, 2179, 1984, 1804, 1614, 1445, 1234, 1068, 870, 739, 668, 616, 566, 532, 489, 453, 426, 385, 357, 335, 316, 297, 283, 274, 266, 259, 251, 241, 233, 226, 222, 217, 214, 211, 209, 208, 0};

int cumf_TCOEF1_intra[104]={16383, 13383, 11498, 10201, 9207, 8528, 8099, 7768, 7546, 7368, 7167, 6994, 6869, 6005, 5474, 5220, 5084, 4964, 4862, 4672, 4591, 4570, 4543, 4397, 4337, 4326, 4272, 4240, 4239, 4212, 4196, 4185, 4158, 4157, 4156, 4140, 4139, 4138, 4137, 4136, 4125, 4124, 4123, 4112, 4111, 4110, 4109, 4108, 4107, 4106, 4105, 4104, 4103, 4102, 4101, 4100, 4099, 4098, 4097, 3043, 2897, 2843, 1974, 1790, 1677, 1552, 1416, 1379, 1331, 1288, 1251, 1250, 1249, 1248, 1247, 1236, 1225, 1224, 1223, 1212, 1201, 1200, 1199, 1198, 1197, 1196, 1195, 1194, 1193, 1192, 1191, 1190, 1189, 1188, 1187, 1186, 1185, 1184, 1183, 1182, 1181, 1180, 1179, 0};

int cumf_TCOEF2_intra[104]={16383, 13242, 11417, 10134, 9254, 8507, 8012, 7556, 7273, 7062, 6924, 6839, 6741, 6108, 5851, 5785, 5719, 5687, 5655, 5028, 4917, 4864, 4845, 4416, 4159, 4074, 3903, 3871, 3870, 3765, 3752, 3751, 3659, 3606, 3580, 3541, 3540, 3514, 3495, 3494, 3493, 3474, 3473, 3441, 3440, 3439, 3438, 3425, 3424, 3423, 3422, 3421, 3420, 3401, 3400, 3399, 3398, 3397, 3396, 2530, 2419, 2360, 2241, 2228, 2017, 1687, 1576, 1478, 1320, 1281, 1242, 1229, 1197, 1178, 1152, 1133, 1114, 1101, 1088, 1087, 1086, 1085, 1072, 1071, 1070, 1069, 1068, 1067, 1066, 1065, 1064, 1063, 1062, 1061, 1060, 1059, 1058, 1057, 1056, 1055, 1054, 1053, 1052, 0};

int cumf_TCOEF3_intra[104]={16383, 12741, 10950, 10071, 9493, 9008, 8685, 8516, 8385, 8239, 8209, 8179, 8141, 6628, 5980, 5634, 5503, 5396, 5327, 4857, 4642, 4550, 4481, 4235, 4166, 4151, 3967, 3922, 3907, 3676, 3500, 3324, 3247, 3246, 3245, 3183, 3168, 3084, 3069, 3031, 3030, 3029, 3014, 3013, 2990, 2975, 2974, 2973, 2958, 2943, 2928, 2927, 2926, 2925, 2924, 2923, 2922, 2921, 2920, 2397, 2298, 2283, 1891, 1799, 1591, 1445, 1338, 1145, 1068, 1006, 791, 768, 661, 631, 630, 615, 592, 577, 576, 561, 546, 523, 508, 493, 492, 491, 476, 475, 474, 473, 472, 471, 470, 469, 468, 453, 452, 451, 450, 449, 448, 447, 446, 0};

int cumf_TCOEFr_intra[104]={16383, 12514, 10776, 9969, 9579, 9306, 9168, 9082, 9032, 9000, 8981, 8962, 8952, 7630, 7212, 7053, 6992, 6961, 6940, 6195, 5988, 5948, 5923, 5370, 5244, 5210, 4854, 4762, 4740, 4384, 4300, 4288, 4020, 3968, 3964, 3752, 3668, 3511, 3483, 3354, 3322, 3205, 3183, 3108, 3046, 2999, 2981, 2974, 2968, 2961, 2955, 2949, 2943, 2942, 2939, 2935, 2934, 2933, 2929, 2270, 2178, 2162, 1959, 1946, 1780, 1651, 1524, 1400, 1289, 1133, 1037, 942, 849, 763, 711, 591, 521, 503, 496, 474, 461, 449, 442, 436, 426, 417, 407, 394, 387, 377, 373, 370, 367, 366, 365, 364, 363, 362, 358, 355, 352, 351, 350, 0};

int cumf_SIGN[3]={16383, 8416, 0};

int cumf_LAST[3]={16383, 9469, 0};

int cumf_LAST_intra[3]={16383, 2820, 0};


```
int cumf_RUN[65]={16383, 15310, 14702, 13022, 11883, 11234, 10612, 10192, 9516, 9016, 8623,
8366, 7595, 7068, 6730, 6487, 6379, 6285, 6177, 6150, 6083, 5989, 5949, 5922, 5895, 5828, 5774,
5773, 5394, 5164, 5016, 4569, 4366, 4136, 4015, 3867, 3773, 3692, 3611, 3476, 3341, 3301, 2787,
2503, 2219, 1989, 1515, 1095, 934, 799, 691, 583, 435, 300, 246, 206, 125, 124, 97, 57, 30, 3, 2, 1,
0};
```

```
int cumf_RUN_intra[65]={16383, 10884, 8242, 7124, 5173, 4745, 4246, 3984, 3034, 2749, 2607,
2298, 966, 681, 396, 349, 302, 255, 254, 253, 206, 159, 158, 157, 156, 155, 154, 153, 106, 35, 34,
33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6,
5, 4, 3, 2, 1, 0};
```

```
int cumf_LEVEL[255]={16383, 16382, 16381, 16380, 16379, 16378, 16377, 16376, 16375, 16374,
16373, 16372, 16371, 16370, 16369, 16368, 16367, 16366, 16365, 16364, 16363, 16362, 16361,
16360, 16359, 16358, 16357, 16356, 16355, 16354, 16353, 16352, 16351, 16350, 16349, 16348,
16347, 16346, 16345, 16344, 16343, 16342, 16341, 16340, 16339, 16338, 16337, 16336, 16335,
16334, 16333, 16332, 16331, 16330, 16329, 16328, 16327, 16326, 16325, 16324, 16323, 16322,
16321, 16320, 16319, 16318, 16317, 16316, 16315, 16314, 16313, 16312, 16311, 16310, 16309,
16308, 16307, 16306, 16305, 16304, 16303, 16302, 16301, 16300, 16299, 16298, 16297, 16296,
16295, 16294, 16293, 16292, 16291, 16290, 16289, 16288, 16287, 16286, 16285, 16284, 16283,
16282, 16281, 16280, 16279, 16278, 16277, 16250, 16223, 16222, 16195, 16154, 16153, 16071,
15989, 15880, 15879, 15878, 15824, 15756, 15674, 15606, 15538, 15184, 14572, 13960, 10718,
7994, 5379, 2123, 1537, 992, 693, 611, 516, 448, 421, 380, 353, 352, 284, 257, 230, 203, 162, 161,
160, 133, 132, 105, 104, 103, 102, 101, 100, 99, 98, 97, 96, 95, 94, 93, 92, 91, 90, 89, 88, 87, 86,
85, 84, 83, 82, 81, 80, 79, 78, 77, 76, 75, 74, 73, 72, 71, 70, 69, 68, 67, 66, 65, 64, 63, 62, 61, 60,
59, 58, 57, 56, 55, 54, 53, 52, 51, 50, 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34,
33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6,
5, 4, 3, 2, 1, 0};
```

```
int cumf_LEVEL_intra[255]={16383, 16379, 16378, 16377, 16376, 16375, 16374, 16373, 16372,
16371, 16370, 16369, 16368, 16367, 16366, 16365, 16364, 16363, 16362, 16361, 16360, 16359,
16358, 16357, 16356, 16355, 16354, 16353, 16352, 16351, 16350, 16349, 16348, 16347, 16346,
16345, 16344, 16343, 16342, 16341, 16340, 16339, 16338, 16337, 16336, 16335, 16334, 16333,
16332, 16331, 16330, 16329, 16328, 16327, 16326, 16325, 16324, 16323, 16322, 16321, 16320,
16319, 16318, 16317, 16316, 16315, 16314, 16313, 16312, 16311, 16268, 16267, 16224, 16223,
16180, 16179, 16136, 16135, 16134, 16133, 16132, 16131, 16130, 16129, 16128, 16127, 16126,
16061, 16018, 16017, 16016, 16015, 16014, 15971, 15970, 15969, 15968, 15925, 15837, 15794,
15751, 15750, 15749, 15661, 15618, 15508, 15376, 15288, 15045, 14913, 14781, 14384, 13965,
13502, 13083, 12509, 12289, 12135, 11892, 11738, 11429, 11010, 10812, 10371, 9664, 9113,
8117, 8116, 8028, 6855, 5883, 4710, 4401, 4203, 3740, 3453, 3343, 3189, 2946, 2881, 2661, 2352,
2132, 1867, 1558, 1382, 1250, 1162, 1097, 1032, 967, 835, 681, 549, 439, 351, 350, 307, 306, 305,
304, 303, 302, 301, 300, 299, 298, 255, 212, 211, 210, 167, 166, 165, 164, 163, 162, 161, 160, 159,
158, 115, 114, 113, 112, 111, 68, 67, 66, 65, 64, 63, 62, 61, 60, 59, 58, 57, 56, 55, 54, 53, 52, 51,
50, 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25,
24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0};
```

```
int cumf_INTRA_AC_DC[4]={16383, 9229, 5461, 0};
```

Annex F

Advanced Prediction mode

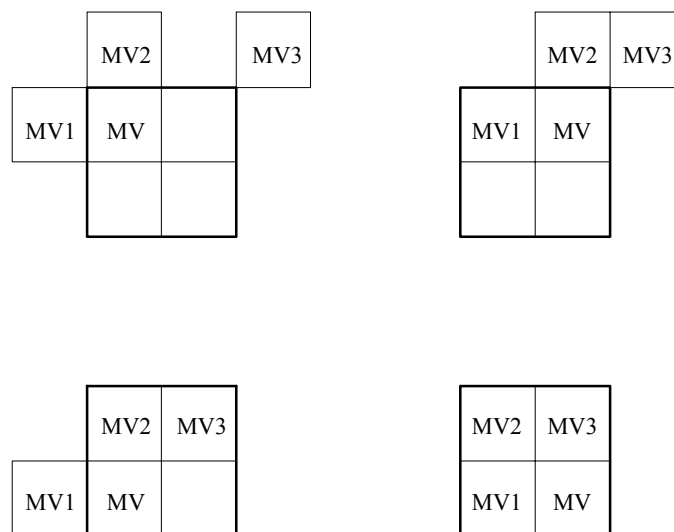
F.1 Introduction

This annex describes the optional Advanced Prediction mode of this Recommendation, including overlapped block motion compensation and the possibility of four motion vectors per macroblock. The capability of this mode is signalled by external means (for example, ITU-T Rec. H.245). The use of this mode is indicated in PTYPE. In the Advanced Prediction mode, motion vectors are allowed to cross picture boundaries as is the case in the Unrestricted Motion Vector mode (for the description of this technique, refer to D.1). The extended motion vector range feature of the Unrestricted Motion Vector mode is not automatically included in the Advanced Prediction mode, and only is active if the Unrestricted Motion Vector mode is selected. If the Advanced Prediction mode is used in combination with the PB-frames mode, overlapped motion compensation is only used for prediction of the P-pictures, not for the B-pictures.

F.2 Four motion vectors per macroblock

In this Recommendation, one motion vector per macroblock is used except when in Advanced Prediction mode or Deblocking Filter mode. In this mode, the one/four vectors decision is indicated by the MCBPC codeword for each macroblock. If only one motion vector is transmitted for a certain macroblock, this is defined as four vectors with the same value. If MCBPC indicates that four motion vectors are transmitted for the current macroblock, the information for the first motion vector is transmitted as the codeword MVD and the information for the three additional motion vectors is transmitted as the codewords MVD₂₋₄ (see also 5.3.7 and 5.3.8).

The vectors are obtained by adding predictors to the vector differences indicated by MVD and MVD₂₋₄ in a similar way as when only one motion vector per macroblock is present, according to the decision rules given in 6.1.1. Again the predictors are calculated separately for the horizontal and vertical components. However, the candidate predictors MV1, MV2 and MV3 are redefined as indicated in Figure F.1. If only one vector per macroblock is present, MV1, MV2 and MV3 are defined as for the 8 * 8 block numbered 1 in Figure 5 (this definition is given in the upper left of the four sub-figures of Figure F.1).



T1602850-97

Figure F.1/H.263 – Redefinition of the candidate predictors MV1, MV2 and MV3 for each of the luminance blocks in a macroblock

If four vectors are used, each of the motion vectors is used for all pixels in one of the four luminance blocks in the macroblock. The numbering of the motion vectors is equivalent to the numbering of the four luminance blocks as given in Figure 5. Motion vector MVD_{CHR} for both chrominance blocks is derived by calculating the sum of the four luminance vectors and dividing this sum by 8; the component values of the resulting sixteenth pixel resolution vectors are modified towards the nearest half-pixel position as indicated in Table F.1.

Table F.1/H.263 – Modification of sixteenth pixel resolution chrominance vector components

Sixteenth pixel position	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	/16
Resulting position	0	0	0	1	1	1	1	1	1	1	1	1	1	1	2	2	/2

Half-pixel values are found using bilinear interpolation as described in 6.1.2. In Advanced Prediction mode, the prediction for luminance is obtained by overlapped motion compensation as described in F.3. The prediction for chrominance is obtained by applying the motion vector MVD_{CHR} to all pixels in the two chrominance blocks (as it is done in the default prediction mode).

F.3 Overlapped motion compensation for luminance

Each pixel in an 8 * 8 luminance prediction block is a weighted sum of three prediction values, divided by 8 (with rounding). In order to obtain the three prediction values, three motion vectors are used: the motion vector of the current luminance block, and two out of four "remote" vectors:

- the motion vector of the block at the left or right side of the current luminance block;
- the motion vector of the block above or below the current luminance block.

For each pixel, the remote motion vectors of the blocks at the two nearest block borders are used. This means that for the upper half of the block the motion vector corresponding to the block above the current block is used, while for the lower half of the block the motion vector corresponding to the block below the current block is used (see Figure F.3). Similarly, for the left half of the block the motion vector corresponding to the block at the left side of the current block is used, while for the right half of the block the motion vector corresponding to the block at the right side of the current block is used (see Figure F.4).

Let (x,y) be a position in a picture measured in integer pixel units.

Let (m,n) be an integer block index in a picture, as given by:

$$m = x/8 \quad \text{and} \quad n = y/8$$

where "/" denotes division with truncation.

Let (i,j) be an integer pixel location in an 8 × 8 block, given by:

$$i = x - m \cdot 8 \quad \text{and} \quad j = y - n \cdot 8$$

resulting in:

$$(x, y) = (m \cdot 8 + i, n \cdot 8 + j)$$

Let (MV^k_x, MV^k_y) be a motion vector, which may contain full-pixel or half pixel offsets, with $k = 0, 1, \text{ or } 2$. For example, (MV^k_x, MV^k_y) can be equal to $(-7.0, 13.5)$. Here, (MV^0_x, MV^0_y) denotes the motion vector for the current block (m,n) , (MV^1_x, MV^1_y) denotes the motion vector of the block either above or below, and (MV^2_x, MV^2_y) denotes the motion vector either to the left or right of the current block (m,n) as defined above.

Then the creation of each pixel, $P(x,y)$, in an 8×8 luminance prediction block with block index (m,n) is governed by the following equation:

$$P(x,y) = (q(x,y) \cdot H_0(i,j) + r(x,y) \cdot H_1(i,j) + s(x,y) \cdot H_2(i,j) + 4) / 8$$

where $q(x,y)$, $r(x,y)$ and $s(x,y)$ are the prediction values taken from the referenced picture as defined by:

$$q(x,y) = p(x + MV^0_x, y + MV^0_y)$$

$$r(x,y) = p(x + MV^1_x, y + MV^1_y)$$

$$s(x,y) = p(x + MV^2_x, y + MV^2_y)$$

where $p(x + MV^k_x, y + MV^k_y)$ is the prediction value at position $(x + MV^k_x, y + MV^k_y)$ in the referenced picture. Note that $(x + MV^k_x, y + MV^k_y)$ may be outside the picture and can be at a full or half-pixel position. In cases using half-pixel motion vectors, $p(x + MV^k_x, y + MV^k_y)$ refers to the value obtained after the interpolation process described in 6.1.2 is applied.

The matrices $H_0(i,j)$, $H_1(i,j)$ and $H_2(i,j)$ are defined in Figures F.2, F.3 and F.4, where (i,j) denotes the column and row, respectively, of the matrix.

When neither the Slice Structured mode (see Annex K) nor the Independent Segment Decoding mode (see Annex R) are in use, remote motion vectors from other video picture segments are used in the same way as remote motion vectors inside the current GOB. If either the Slice Structured mode or the Independent Segment Decoding mode are in use, the remote motion vectors corresponding to blocks from other video picture segments are set to the motion vector of the current block, regardless of the other conditions described in the next paragraph. (See Annex R for the definition of a video picture segment.)

If one of the surrounding macroblocks was not coded, the corresponding remote motion vector is set to zero. If one of the surrounding blocks was INTRA coded, the corresponding remote motion vector is replaced by the motion vector for the current block except when in PB-frames mode. In this case (INTRA block in PB-frame mode), the INTRA block's motion vector is used (also see Annex G). If the current block is at the border of the picture and therefore a surrounding block is not present, the corresponding remote motion vector is replaced by the current motion vector. In all cases, if the current block is at the bottom of the macroblock (for block number 3 or 4, see Figure 5), the remote motion vector corresponding with an $8 * 8$ luminance block in the macroblock below the current macroblock is replaced by the motion vector for the current block.

The weighting values for the prediction are given in Figures F.2, F.3 and F.4.

4	5	5	5	5	5	5	4
5	5	5	5	5	5	5	5
5	5	6	6	6	6	5	5
5	5	6	6	6	6	5	5
5	5	6	6	6	6	5	5
5	5	6	6	6	6	5	5
5	5	5	5	5	5	5	5
4	5	5	5	5	5	5	4

Figure F.2/H.263 – Weighting values, H_0 , for prediction with motion vector of current luminance block

2	2	2	2	2	2	2	2
1	1	2	2	2	2	1	1
1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
1	1	2	2	2	2	1	1
2	2	2	2	2	2	2	2

Figure F.3/H.263 – Weighting values, H_1 , for prediction with motion vectors of the luminance blocks on top or bottom of current luminance block

2	1	1	1	1	1	1	2
2	2	1	1	1	1	2	2
2	2	1	1	1	1	2	2
2	2	1	1	1	1	2	2
2	2	1	1	1	1	2	2
2	2	1	1	1	1	2	2
2	2	1	1	1	1	2	2
2	1	1	1	1	1	1	2

Figure F.4/H.263 – Weighting values, H_2 , for prediction with motion vectors of the luminance blocks to the left or right of current luminance block

Annex G

PB-frames mode

G.1 Introduction

This annex describes the optional PB-frames mode of this Recommendation. The capability of this mode is signalled by external means (for example, ITU-T Rec. H.245). The use of this mode is indicated in PTYPE.

A PB-frame consists of two pictures being coded as one unit. The name PB comes from the name of picture types in ITU-T Rec. H.262 where there are P-pictures and B-pictures. Thus, a PB-frame consists of one P-picture which is predicted from the previous decoded P-picture and one B-picture which is predicted both from the previous decoded P-picture and the P-picture currently being decoded. The name B-picture was chosen because parts of B-pictures may be bidirectionally predicted from the past and future pictures. The prediction process is illustrated in Figure G.1.

An improved version of the PB-frames mode, termed the "Improved PB-frames mode", is described in Annex M. The PB-frames mode as described in this annex is retained herein only for purposes of compatibility with systems designed prior to the adoption of the Improved PB-frames mode. For this reason, the PB-frames mode as described in this annex cannot be used with the additional features of the syntax which require the use of PLUSPTYPE.

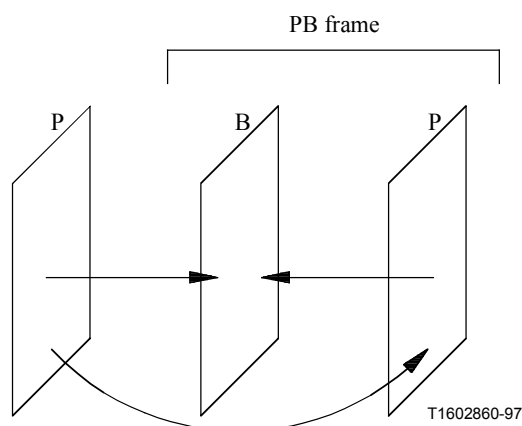


Figure G.1/H.263 – Prediction in PB-frames mode

G.2 PB-frames and INTRA blocks

When PB-frames are used, the coding mode INTRA has the following meaning (see also 5.3.2):

- The P-blocks are INTRA coded.
- The B-blocks are INTER coded with prediction as for an INTER block.

If PB-frames are used, Motion Vector Data (MVD) is included also for INTRA macroblocks in pictures for which PTYPE indicates "INTER". In this case, the vector is used for the B-blocks only. The codewords MVD_{2-4} are never used for INTRA (see also Table 10). When in both the Advanced Prediction mode and the PB-frames mode, and one of the surrounding blocks was coded in INTRA mode, the corresponding remote motion vector is not replaced by the motion vector for the current block. Instead, the remote "INTRA" motion vector is used.

G.3 Block layer

In a PB-frame, a macroblock comprises twelve blocks. First, the data for the six P-blocks is transmitted as in the default H.263 mode, then the data for the six B-blocks (see also 5.4). The

structure of the block layer is shown in Figure 11. INTRADC is present for every P-block of the macroblock if MCBPC indicates macroblock type 3 or 4 (see Tables 7 and 8). INTRADC is not present for B-blocks. TCOEF is present for P-blocks if indicated by MCBPC or CBPY; TCOEF is present for B-blocks if indicated by CBPB.

G.4 Calculation of vectors for the B-picture in a PB-frame

The vectors for the B-picture are calculated as follows (see also 6.1.1). Assume we have a vector component MV in half-pixel units to be used in the P-picture (MV represents a vector component for an 8×8 luminance block; if only one vector per macroblock is transmitted, MV has the same value for each of the four 8×8 luminance blocks). For prediction of the B-picture we need both forward and backward vector components MV_F and MV_B . These forward and backward vector components are derived from MV and eventually enhanced by a delta vector given by MV_D .

- TR_D : The increment of the temporal reference TR (or the combined extended temporal reference ETR and temporal reference TR in an Improved PB frame when a custom picture clock frequency is in use) from the last picture header (see 5.1.2). If TR_D is negative, then $TR_D = TR_D + d$ where $d = 256$ for CIF picture frequency and 1024 for any custom picture clock frequency.
- TR_B : See 5.1.2.

Assume that MV_D is the delta vector component given by $MVDB$ and corresponding with vector component MV . If $MVDB$ is not present, MV_D is set to zero. If $MVDB$ is present, the same MV_D given by $MVDB$ is used for each of the four luminance B-blocks within the macroblock.

Now, MV_F and MV_B are given in half-pixel units by the following formulae:

$$\begin{aligned}
 MV_F &= (TR_B \times MV) / TR_D + MV_D \\
 MV_B &= ((TR_B - TR_D) \times MV) / TR_D \quad \text{if } MV_D \text{ is equal to } 0 \\
 MV_B &= MV_F - MV \quad \text{if } MV_D \text{ is unequal to } 0
 \end{aligned}$$

where "/" means division by truncation. It is assumed that the scaling reflects the actual position in time of P- and B-pictures. Advantage is taken of the fact that the range of values for MV_F is constrained. Each VLC word for $MVDB$ represents a pair of difference values. Only one of the pair will yield a value for MV_F falling within the permitted range (default $[-16, 15.5]$; in Unrestricted Motion Vector mode $[-31.5, 31.5]$). The formulae for MV_F and MV_B are also used in the case of INTRA blocks where the vector data is used only for predicting B-blocks.

For chrominance blocks, MV_F is derived by calculating the sum of the four corresponding luminance MV_F vectors and dividing this sum by 8; the resulting sixteenth pixel resolution vector components are modified towards the nearest half-pixel position as indicated in Table F.1. MV_B for chrominance is derived by calculating the sum of the four corresponding luminance MV_B vectors and dividing this sum by 8; the resulting sixteenth pixel resolution vector components are modified towards the nearest half-pixel position as indicated in Table F.1.

A positive value of the horizontal or vertical component of the motion vector signifies that the prediction is formed from pixels in the referenced picture which are spatially to the right or below the pixels being predicted.

G.5 Prediction of a B-block in a PB-frame

In this clause a block means an 8×8 block. The following procedure applies for luminance as well as chrominance blocks. First, the forward and backward vectors are calculated. It is assumed that the P-macroblock (luminance and chrominance) is first decoded, reconstructed and clipped (see 6.3.2). This macroblock is called P_{REC} . Based on P_{REC} and the prediction for P_{REC} , the prediction for the B-block is calculated.

The prediction of the B-block has two modes that are used for different parts of the block:

- For pixels where the backward vector – MV_B – points inside P_{REC} , use bidirectional prediction. This is obtained as the average of the forward prediction using MV_F relative to the previous decoded picture, and the backward prediction using MV_B relative to P_{REC} . The average is calculated by dividing the sum of the two predictions by two (division by truncation).
- For all other pixels, forward prediction using MV_F relative to the previous decoded picture is used.

Figure G.2 indicates which part of a block is predicted bidirectionally (shaded part of the B-block) and which part with forward prediction only (rest of the B-block).

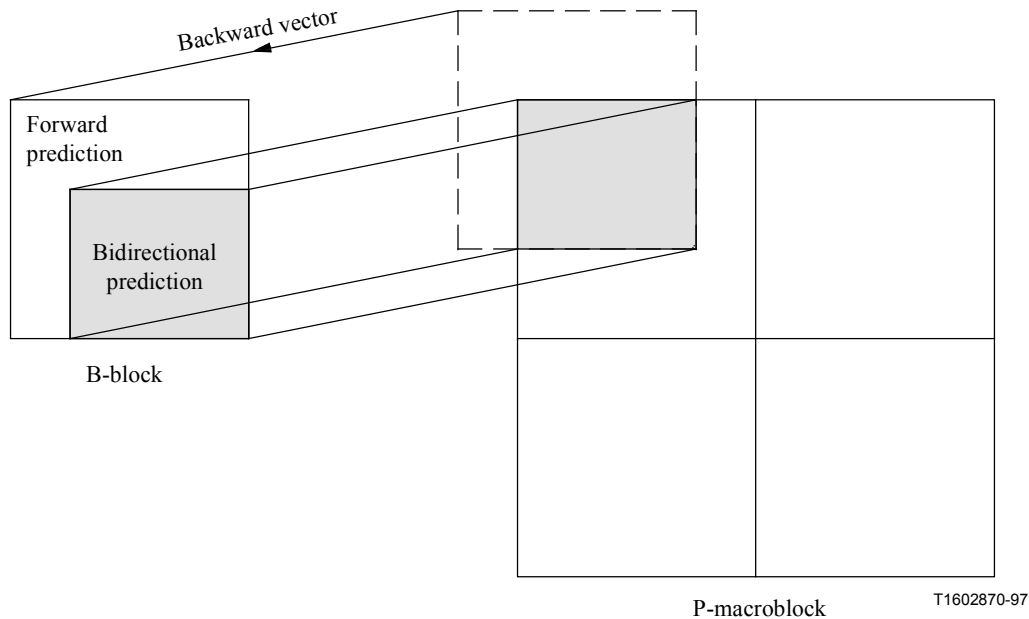


Figure G.2/H.263 – Forward and bidirectional prediction for a B-block

Bidirectional prediction is used for pixels where the backward vector – MV_B – points inside P_{REC} . These pixels are defined by the following procedures which are specified in C language.

Definitions:

- nh Horizontal position of block within a macroblock (0 or 1).
- nv Vertical position of block within a macroblock (0 or 1).
- mh(nh,nv) Horizontal vector component of block (nh,nv) in half-pixel units.
- mv(nh,nv) Vertical vector component of block (nh,nv) in half-pixel units.
- mhc Horizontal chrominance vector component.
- mvc Vertical chrominance vector component.

Procedure for luminance

```
for (nh = 0; nh <= 1; nh++) {
  for (nv = 0; nv <= 1; nv++) {
    for (i = nh * 8 + max(0, (-mh(nh,nv)+1)/2 - nh * 8);
         i <= nh * 8 + min(7, 15-(mh(nh,nv)+1)/2 - nh * 8); i++) {
      for (j = nv * 8 + max(0, (-mv(nh,nv)+1)/2 - nv * 8);
           j <= nv * 8 + min(7, 15-(mv(nh,nv)+1)/2 - nv * 8); j++) {
        predict pixel (i,j) bidirectionally
      }
    }
  }
}
```

Procedure for chrominance

```
for (i = max(0, (-mhc+1)/2); i <= min(7, 7-(mhc+1)/2); i++) {
  for (j = max(0, (-mvc+1)/2); j <= min(7, 7-(mvc+1)/2); j++) {
    predict pixel (i,j) bidirectionally;
  }
}
```

Pixels not predicted bidirectionally are predicted with forward prediction only.

Annex H

Forward error correction for coded video signal

H.1 Introduction

This annex describes an optional forward error correction method (code and framing) for transmission of H.263 encoded video data. This forward error correction may be used in situations where no forward error correction is provided by external means, for example at the multiplex or system level. It is not used for ITU-T Rec. H.324. Both the framing and the forward error correction code are the same as in ITU-T Rec. H.261.

H.2 Error correction framing

To allow the video data and error correction parity information to be identified by a decoder, an error correction framing pattern is included. This pattern consists of multiframe of eight frames, each frame comprising 1 bit framing, 1 bit fill indicator (Fi), 492 bits of coded data (or fill all 1s) and 18 bits parity (see Figure H.1). For each multiframe the frame alignment pattern formed by the framing bits of the eight individual frames is:

$$(S_1S_2S_3S_4S_5S_6S_7S_8) = (00011011)$$

The fill indicator (Fi) can be set to zero by an encoder. In this case 492 consecutive fill bits (fill all 1s) are used instead of 492 bits of coded data. This may be used for stuffing data (see 3.6).

H.3 Error correcting code

The error correction code is a BCH (511, 493) forward error correction code. Use of this by the decoder is optional. The parity is calculated against a code of 493 bits, comprising 1 bit fill indicator (Fi) and 492 bits of coded video data.

The generator polynomial is:

$$g(x) = (x^9 + x^4 + 1)(x^9 + x^6 + x^4 + x^3 + 1)$$

Example: for the input data of "01111 ... 11" (493 bits) the resulting correction parity bits are "011011010100011011" (18 bits).

H.4 Relock time for error corrector framing

Three consecutive error correction frame alignment patterns (24 bits) should be received before frame lock is deemed to have been achieved. The decoder should be designed such that frame lock will be re-established within 34 000 bits after an error corrector framing phase change.

NOTE – This assumes that the video data does not contain three correctly phased emulations of the error correction framing sequence during the relocking period.

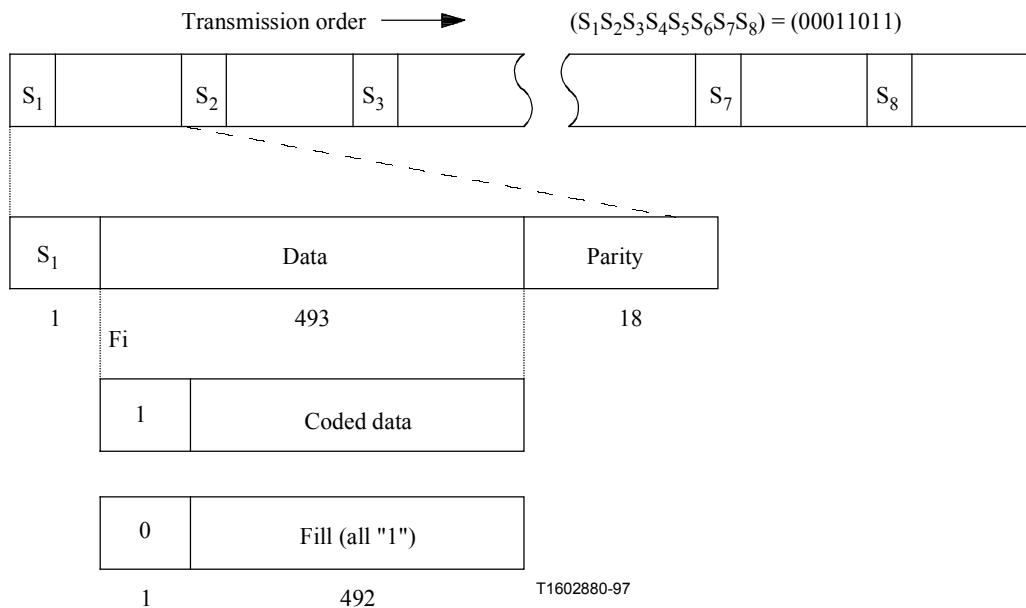


Figure H.1/H.263 – Error correcting frame

Annex I

Advanced INTRA Coding mode

This annex describes the optional Advanced INTRA Coding mode of this Recommendation. The capability of this H.263 mode is signalled by external means (for example, ITU-T Rec. H.245). The use of this mode is indicated in the PLUSPTYPE field of the picture header.

I.1 Introduction

This optional mode alters the decoding of macroblocks of type "INTRA" (macroblocks of other types are not affected). The coding efficiency of INTRA macroblocks is improved by using:

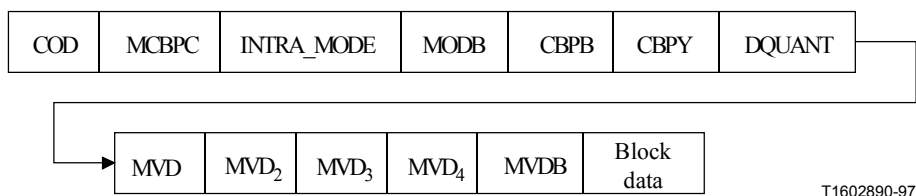
- 1) INTRA-block prediction using neighbouring INTRA blocks for the same component (Y , C_B , or C_R);
- 2) modified inverse quantization for INTRA coefficients; and
- 3) a separate VLC for INTRA coefficients.

A particular INTRA-coded block may be predicted from the block above the current block being decoded, from the block to the left of the current block being decoded, or from both. Special cases exist for situations in which the neighbouring blocks are not INTRA coded or are not in the same

video picture segment. Block prediction always uses data from the same luminance or color difference component (Y, C_B, or C_R) as the block being decoded. In prediction, DC coefficients are always predicted in some manner. The first row of AC coefficients may be predicted from those in the block above, or the first column of AC coefficients may be predicted from those in the block to the left, or only the DC coefficient may be predicted as an average from the block above and the block to the left, as signalled on a macroblock-by-macroblock basis. The remaining AC coefficients are never predicted. Inverse quantization of the INTRADC coefficient is modified to allow a varying quantization step size, unlike in the main text of this Recommendation where a fixed step size of 8 is used for INTRADC coefficients. Inverse quantization of all INTRA coefficients is performed without a "dead-zone" in the quantizer reconstruction spacing.

I.2 Syntax

When using the Advanced INTRA Coding mode, the macroblock layer syntax is altered as specified in Figure I.1. The syntax as shown in Figure I.1 is the same as that defined in 5.3 except for the insertion of an additional INTRA_MODE field for INTRA macroblocks. INTRA_MODE is present only when MCBPC indicates a macroblock of type INTRA (macroblock type 3 or 4). The prediction mode is coded using the variable length code shown in Table I.1. One prediction mode is transmitted per INTRA macroblock.



T1602890-97

Figure I.1/H.263 – Structure of macroblock layer

Table I.1/H.263 – VLC for INTRA_MODE

Index	Prediction Mode	VLC
0	0 (DC Only)	0
1	1 (Vertical DC & AC)	10
2	2 (Horizontal DC & AC)	11

I.3 Decoding process

Two scans in addition to the zigzag scan are employed. The two added scans are shown in Figure I.2-a and I.2-b, and the zigzag scan is shown in Figure 14.

1	2	3	4	11	12	13	14
5	6	9	10	18	17	16	15
7	8	20	19	27	28	29	30
21	22	25	26	31	32	33	34
23	24	35	36	43	44	45	46
37	38	41	42	47	48	49	50
39	40	51	52	57	58	59	60
53	54	55	56	61	62	63	64

a) Alternate-Horizontal scan

1	5	7	21	23	37	39	53
2	6	8	22	24	38	40	54
3	9	20	25	35	41	51	55
4	10	19	26	36	42	52	56
11	18	27	31	43	47	57	61
12	17	28	32	44	48	58	62
13	16	29	33	45	49	59	63
14	15	30	34	46	50	60	64

b) Alternate-Vertical scan (as in ITU-T Rec. H.262)

Figure I.2/H.263 – Alternate DCT scanning patterns for Advanced INTRA coding

For INTRA-coded blocks, if Prediction Mode = 0, the zigzag scan shown in Figure 14 is selected for all blocks in the macroblock; otherwise, the prediction direction is used to select a scan for the macroblock.

Prediction Mode = 1 uses the vertically adjacent block in forming a prediction. This prediction mode is designed for INTRA blocks which are dominated by stronger horizontal frequency content, so the vertically adjacent block is used to predict the horizontal frequency content of the current block, with a prediction of zero for all coefficients representing vertical AC content. Then the scanning pattern is chosen to scan the stronger horizontal frequencies prior to the vertical ones, using the Alternate-Horizontal scan.

Prediction Mode = 2 uses the horizontally adjacent block in forming a prediction. This prediction mode is designed for INTRA blocks which are dominated by stronger vertical frequency content, so the horizontally adjacent block is used to predict the vertical frequency content of the current block, with a prediction of zero for all coefficients representing horizontal AC content. Then the scanning pattern is chosen to scan the stronger vertical frequencies prior to the horizontal ones, using the Alternate-Vertical scan.

For non-INTRA blocks, the 8×8 blocks of transform coefficients are scanned with "zigzag" scanning as shown in Figure 14.

A separate VLC table is used for all INTRADC and INTRA AC coefficients. This table is specified in Table I.2. Note that the VLC codeword entries used in Table I.2 are the same as those used in the normal TCOEF table (Table 16) used when Advanced INTRA coding is not in use, but with a different interpretation of LEVEL and RUN (without altering LAST).

Table I.2/H.263 – VLC for INTRA TCOEF

Index	LAST	RUN	LEVEL	Bits	VLC code
0	0	0	1	3	10s
1	0	1	1	5	1111s
2	0	3	1	7	0101 01s
3	0	5	1	8	0010 111s
4	0	7	1	9	0001 1111s

Table I.2/H.263 – VLC for INTRA TCOEF

Index	LAST	RUN	 LEVEL 	Bits	VLC code
5	0	8	1	10	0001 0010 1s
6	0	9	1	10	0001 0010 0s
7	0	10	1	11	0000 1000 01s
8	0	11	1	11	0000 1000 00s
9	0	4	3	12	0000 0000 111s
10	0	9	2	12	0000 0000 110s
11	0	13	1	12	0000 0100 000s
12	0	0	2	4	110s
13	0	1	2	7	0101 00s
14	0	1	4	9	0001 1110s
15	0	1	5	11	0000 0011 11s
16	0	1	6	12	0000 0100 001s
17	0	1	7	13	0000 0101 0000s
18	0	0	3	5	1110s
19	0	3	2	9	0001 1101s
20	0	2	3	11	0000 0011 10s
21	0	3	4	13	0000 0101 0001s
22	0	0	5	6	0110 1s
23	0	4	2	10	0001 0001 1s
24	0	3	3	11	0000 0011 01s
25	0	0	4	6	0110 0s
26	0	5	2	10	0001 0001 0s
27	0	5	3	13	0000 0101 0010s
28	0	2	1	6	0101 1s
29	0	6	2	11	0000 0011 00s
30	0	0	25	13	0000 0101 0011s
31	0	4	1	7	0100 11s
32	0	7	2	11	0000 0010 11s
33	0	0	24	13	0000 0101 0100s
34	0	0	8	7	0100 10s
35	0	8	2	11	0000 0010 10s
36	0	0	7	7	0100 01s
37	0	2	4	11	0000 0010 01s
38	0	0	6	7	0100 00s
39	0	12	1	11	0000 0010 00s
40	0	0	9	8	0010 110s
41	0	0	23	13	0000 0101 0101s
42	0	2	2	8	0010 101s

Table I.2/H.263 – VLC for INTRA TCOEF

Index	LAST	RUN	 LEVEL 	Bits	VLC code
43	0	1	3	8	0010 100s
44	0	6	1	9	0001 1100s
45	0	0	10	9	0001 1011s
46	0	0	12	10	0001 0000 1s
47	0	0	11	10	0001 0000 0s
48	0	0	18	10	0000 1111 1s
49	0	0	17	10	0000 1111 0s
50	0	0	16	10	0000 1110 1s
51	0	0	15	10	0000 1110 0s
52	0	0	14	10	0000 1101 1s
53	0	0	13	10	0000 1101 0s
54	0	0	20	12	0000 0100 010s
55	0	0	19	12	0000 0100 011s
56	0	0	22	13	0000 0101 0110s
57	0	0	21	13	0000 0101 0111s
58	1	0	1	5	0111s
59	1	14	1	10	0000 1100 1s
60	1	20	1	12	0000 0000 101s
61	1	1	1	7	0011 11s
62	1	19	1	12	0000 0000 100s
63	1	2	1	7	0011 10s
64	1	3	1	7	0011 01s
65	1	0	2	7	0011 00s
66	1	5	1	8	0010 011s
67	1	6	1	8	0010 010s
68	1	4	1	8	0010 001s
69	1	0	3	8	0010 000s
70	1	9	1	9	0001 1010s
71	1	10	1	9	0001 1001s
72	1	11	1	9	0001 1000s
73	1	12	1	9	0001 0111s
74	1	13	1	9	0001 0110s
75	1	8	1	9	0001 0101s
76	1	7	1	9	0001 0100s
77	1	0	4	9	0001 0011s
78	1	17	1	10	0000 1100 0s
79	1	18	1	10	0000 1011 1s
80	1	16	1	10	0000 1011 0s

Table I.2/H.263 – VLC for INTRA TCOEF

Index	LAST	RUN	LEVEL	Bits	VLC code
81	1	15	1	10	0000 1010 1s
82	1	2	2	10	0000 1010 0s
83	1	1	2	10	0000 1001 1s
84	1	0	6	10	0000 1001 0s
85	1	0	5	10	0000 1000 1s
86	1	4	2	11	0000 0001 11s
87	1	3	2	11	0000 0001 10s
88	1	1	3	11	0000 0001 01s
89	1	0	7	11	0000 0001 00s
90	1	2	3	12	0000 0100 100s
91	1	1	4	12	0000 0100 101s
92	1	0	9	12	0000 0100 110s
93	1	0	8	12	0000 0100 111s
94	1	21	1	13	0000 0101 1000s
95	1	22	1	13	0000 0101 1001s
96	1	23	1	13	0000 0101 1010s
97	1	7	2	13	0000 0101 1011s
98	1	6	2	13	0000 0101 1100s
99	1	5	2	13	0000 0101 1101s
100	1	3	3	13	0000 0101 1110s
101	1	0	10	13	0000 0101 1111s
102	ESCAPE			7	0000 011

Depending on the value of INTRA_MODE, either one or eight coefficients are prediction residuals that must be added to a predictor as described below. Figure I.3 shows three 8×8 blocks of final reconstructed DCT levels for the same component (Y , C_B , or C_R), labelled $RecA'(u,v)$, $RecB'(u,v)$ and $RecC'(u,v)$, where u and v are column (horizontal) and row (vertical) indices, respectively. The reconstruction process differs from the processing described in 6.2.1. INTRADC residuals are reconstructed differently by use of a variable step size, as opposed to using Table 15, and then a predictor is added to the residual values to obtain the final coefficient reconstruction value. INTRA coefficients other than INTRADC are also reconstructed differently than in 6.2.1, using a reconstruction spacing without a "dead-zone" and then in some cases adding a predictor to obtain the final coefficient reconstruction value. The block may contain both DC and AC prediction residuals.

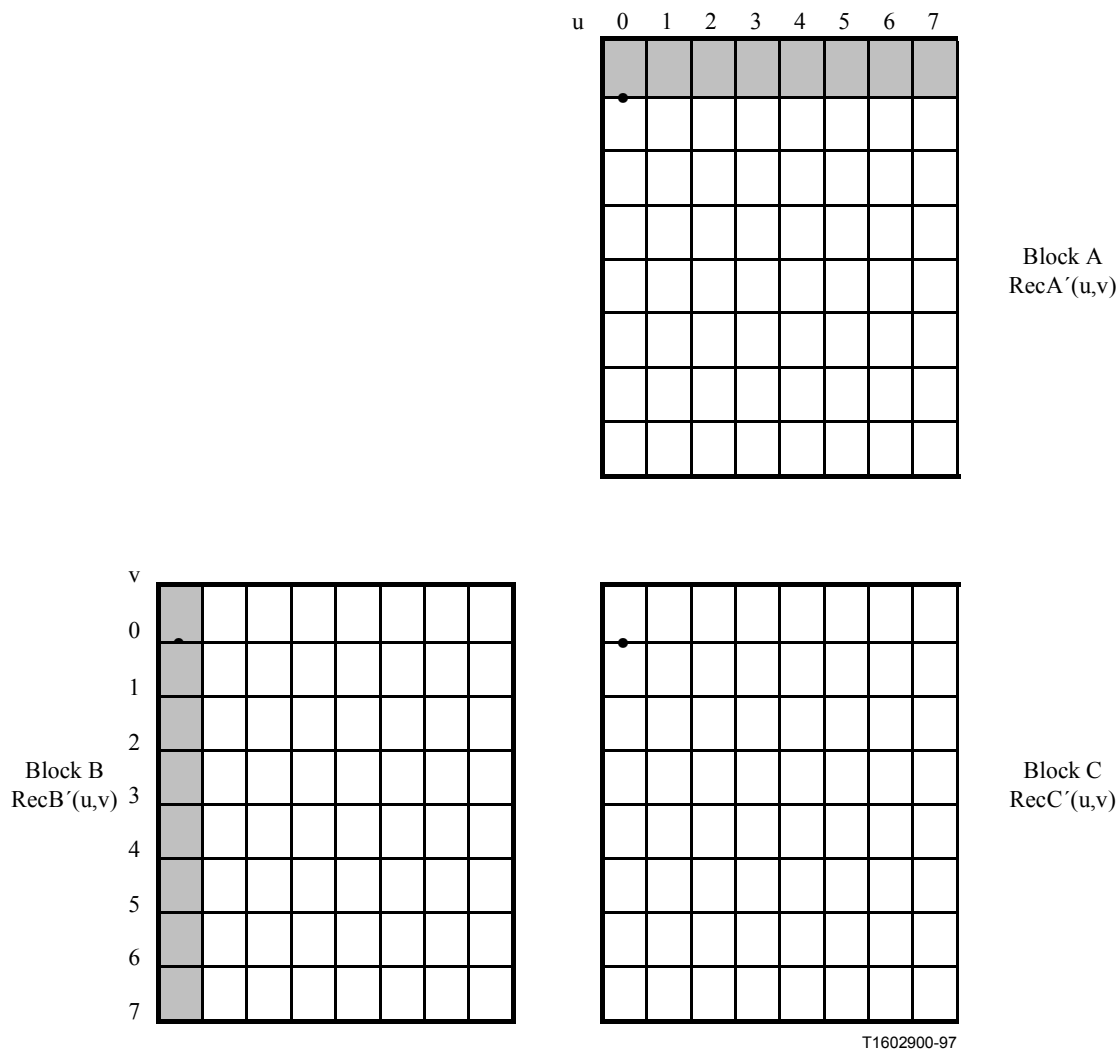


Figure I.3/H.263 – Three neighbouring blocks in the DCT domain

The definitions of MCBPC and CBPY are changed when Advanced INTRA coding is in use. When Advanced INTRA coding is in use, INTRADC transform coefficients are no longer handled as a separate case, but are instead treated in the same way as the AC coefficients in regard to MCBPC and CBPY. This means that a zero INTRADC will not be coded as a LEVEL, but will simply increase the run for the following AC coefficients.

The inverse quantization process for the B-part of an Improved PB frame (see Annex M) is not altered by the use of the Advanced INTRA coding mode.

Define $RecC(u,v)$ to be the reconstructed coefficient *residuals* of the current block. For all INTRA coefficients, the reconstructed residual value is obtained by:

$$RecC(u,v) = 2 * QUANT * LEVEL(u,v) \quad u = 0, \dots, 7, v = 0, \dots, 7$$

NOTE – LEVEL(u,v) denotes a quantity having both a magnitude and a sign in the above equation.

Define $RecC'(u,v)$ to be the *final* reconstructed coefficient values of the current block (after adjustments for prediction, oddification as described below, and clipping). The final reconstructed coefficient values $RecC'(u,v)$ are recovered by adding $RecC(u,v)$ to the appropriate prediction as signalled in the INTRA_MODE field, altering the least-significant bit if needed for oddification of the DC coefficient, and clipping.

$RecA'(u,v)$ denotes the *final* reconstructed coefficient values for the block immediately above the current block. $RecB'(u,v)$ denotes the *final* reconstructed coefficient values of the block immediately to the left of the current block.

The ability to use the reconstructed coefficient values for blocks A and B in the prediction of the coefficient values for block C depends on whether blocks A and B are in the same video picture segment as block C. A block is defined to be "in the same video picture segment" as another block only if the following conditions are fulfilled:

- 1) The relevant block is within the boundary of the picture; and
- 2) If not in Slice Structured mode (see Annex K), the relevant block is either within the same GOB or no GOB header is present for the current GOB; and
- 3) If in Slice Structured mode, the relevant block is within the same slice.

The block C to be decoded is predicted only from INTRA blocks within the same video picture segment as block C, as shown below.

If Prediction Mode = 0 is used (DC prediction only) and blocks A and B are both INTRA blocks within the same video picture segment as block C, then the DC coefficient of block C is predicted from the average (with truncation) of the DC coefficients of blocks A and B. If only one of the two blocks A and B is an INTRA block within the same video picture segment as block C, then the DC coefficient of only this one of the two blocks is used as the predictor for Prediction Mode = 0. If neither of the two blocks A and B are INTRA blocks within the same video picture segment as block C, then the prediction uses the value of 1024 as the predictor for the DC coefficient.

If Prediction Mode = 1 or 2 (vertical DC & AC or horizontal DC & AC prediction) and the referenced block (block A or block B) is not an INTRA block within the same video picture segment as block C, then the prediction uses the value of 1024 as the predictor for the DC coefficient and the value of 0 as the predictor for the AC coefficients of block C.

A process of "oddification" is applied to the DC coefficient in order to minimize the impact of IDCT mismatch errors. Certain values of coefficients can cause a round-off error mismatch between different IDCT implementations, especially certain values of the (0,0), (0,4), (4,0), and (4,4) coefficients. For example, a DC coefficient of $8k + 4$ for some integer k results in an inverse-transformed block having a constant value $k + 0.5$, for which slight errors can cause rounding in different directions on different implementations.

Define the clipAC() function to indicate clipping to the range -2048 to 2047. Define the clipDC() function to indicate clipping to the range 0 to 2047. Define the oddifyclipDC(x) function as:

```

If (x is even) {
    result = clipDC(x+1)
} else {
    result = clipDC(x)
}

```

The reconstruction for each INTRA prediction mode is then specified as follows, in which the operator "/" is defined as division by truncation:

Mode 0: DC prediction only.

```

RecC'(u,v) = clipAC( RecC(u,v) ) (u,v) ≠ (0,0), u = 0, ..., 7, v = 0, ..., 7.
If (block A and block B are both INTRA coded and are both in the same
    video picture segment as block C) {
    tempDC = RecC(0,0) + ( RecA'(0,0) + RecB'(0,0) ) / 2
} else {
    If (block A is INTRA coded and is in the same video picture segment
        as block C) {
        tempDC = RecC(0,0) + RecA'(0,0)
    }
}

```

```

    } else {
        If (block B is INTRA coded and is in the same video picture segment
            as block C) {
            tempDC = RecC(0,0) + RecB'(0,0)
        } else {
            tempDC = RecC(0,0) + 1024
        }
    }
}
RecC'(0,0) = oddifyclipDC( tempDC )

```

Mode 1: DC and AC prediction from the block above.

```

If (block A is INTRA coded and is in the same video picture segment as
    block C) {
    tempDC = RecC(0,0) + RecA'(0,0)
    RecC'(u,0) = clipAC( RecC(u,0) + RecA'(u,0) )      u = 1, ..., 7,
    RecC'(u,v) = clipAC( RecC(u,v) )                  u = 0, ..., 7, v = 1, ..., 7.
} else {
    tempDC = RecC(0,0) + 1024
    RecC'(u,v) = clipAC( RecC(u,v) )                  (u,v) ≠ (0,0), u = 0, ..., 7, v =
0, ..., 7
}
RecC'(0,0) = oddifyclipDC( tempDC )

```

Mode 2: DC and AC prediction from the block to the left.

```

If (block B is INTRA coded and is in the same video picture segment as
    block C) {
    tempDC = RecC(0,0) + RecB'(0,0)
    RecC'(0,v) = clipAC( RecC(0,v) + RecB'(0,v) )      v = 1, ..., 7,
    RecC'(u,v) = clipAC( RecC(u,v) )                  u = 1, ..., 7, v = 0, ..., 7.
} else {
    tempDC = RecC(0,0) + 1024
    RecC'(u,v) = clipAC( RecC(u,v) )                  (u,v) ≠ (0,0), u = 0, ..., 7,
                                                    v = 0, ..., 7
}
RecC'(0,0) = oddifyclipDC( tempDC )

```

Annex J

Deblocking Filter mode

J.1 Introduction

This annex describes the use of an optional block edge filter within the coding loop. The main purpose of the block edge filter is to reduce blocking artifacts. The filtering is performed on 8×8 block edges. Motion vectors may have either 8×8 or 16×16 resolution (see J.2). The processing described in this annex applies only for the P-, I-, EP-, or EI-pictures or the P-picture part of an Improved PB-frame. (Possible filtering of B-pictures or the B-picture part of an Improved PB-frame is not a matter for standardization; however, some type of filtering is recommended for improved picture quality.) The capability of this mode is signalled by external means (for example, ITU-T Rec. H.245). The use of this mode is indicated in the PLUSPTYPE field of the picture header.

NOTE – The problem of IDCT mismatch can be aggravated when using Annex J with a small quantization step size and not using the Annex W IDCT in both the encoder and decoder. In some cases, the amount of rounding error mismatch may even be amplified by the Annex J filtering process. Encoders should avoid this problem, by, for example, not using Annex J when the quantization step size is very small.

J.2 Relation to UMV and AP modes (Annexes D and F)

The use of the deblocking filter mode has similar effects on picture quality as Overlapped Block Motion Compensation (OBMC) as defined in Annex F when used alone. When both techniques are used together, a further improvement in picture quality can be obtained. The Advanced Prediction mode (see also Annex F) consists of three elements:

- 1) four motion vectors per macroblock as defined in F.2;
- 2) overlapped motion compensation for luminance as defined in F.3;
- 3) motion vectors over picture boundaries as defined in D.1.

In order for the Deblocking Filter mode to be able to provide maximum performance when complexity considerations may prevent use of the OBMC part of the Advanced Prediction mode, the Deblocking Filter mode includes the ability to use four motion vectors per macroblock and motion vectors over picture boundaries.

In summary, the three options defined in Annexes D, F and J contain the following five coding elements:

- 1) motion vectors over picture boundaries (D.1);
- 2) extension of motion vector range (D.2);
- 3) four motion vectors per macroblock (F.2);
- 4) overlapped motion compensation for luminance (F.3);
- 5) deblocking edge filter (J.3).

Table J.1 indicates which of the five elements are turned on depending on which of the three options defined in Annexes D, F and J are turned on.

Table J.1/H.263 – Feature Elements for UMV, AP, and DF modes

Unrestricted Motion Vector mode	Advanced Prediction mode	Deblocking Filter mode	Motion vectors over picture boundaries	Extension of motion vector range	Four motion vectors per macroblock	Overlapped motion compensation for luminance	Deblocking edge filter
OFF	OFF	OFF	OFF	OFF	OFF	OFF	OFF
OFF	OFF	ON	ON	OFF	ON	OFF	ON
OFF	ON	OFF	ON	OFF	ON	ON	OFF
OFF	ON	ON	ON	OFF	ON	ON	ON
ON	OFF	OFF	ON	ON	OFF	OFF	OFF
ON	OFF	ON	ON	ON	ON	OFF	ON
ON	ON	OFF	ON	ON	ON	ON	OFF
ON	ON	ON	ON	ON	ON	ON	ON

J.3 Definition of the deblocking edge filter

The filter operations are performed across 8×8 block edges at the encoder as well as on the decoder side. The reconstructed image data (the sum of the prediction and reconstructed prediction error) are clipped to the range of 0 to 255 as described in 6.3.2. Then the filtering is applied, which alters the picture that is to be stored in the picture store for future prediction. The filtering operations include an additional clipping to ensure that resulting pixel values stay in the range 0...255. No filtering is performed across a picture edge, and when the Independent Segment Decoding mode is in use, no filtering is performed across slice edges when the Slice Structured mode is in use (see Annexes K and R) or across the top boundary of GOBs having GOB headers

present when the Slice Structured mode is not in use (see Annex R). Chrominance as well as luminance data are filtered.

When the mode described in this annex is used together with the Improved PB-frames mode of Annex M, the backward prediction of the B-macroblock is based on the reconstructed P-macroblock (named P_{REC} in G.5) after the clipping operation but before the deblocking edge filter operations. The forward prediction of the B-macroblock is based on the filtered version of the previous decoded picture (the same picture data that are used for prediction of the P-macroblock).

The deblocking filter operates using a set of four (clipped) pixel values on a horizontal or vertical line of the reconstructed picture, denoted as A, B, C and D, of which A and B belong to one block called block1 and C and D belong to a neighbouring block called block2 which is to the right of or below block1. Figure J.1 shows examples for the position of these pixels.

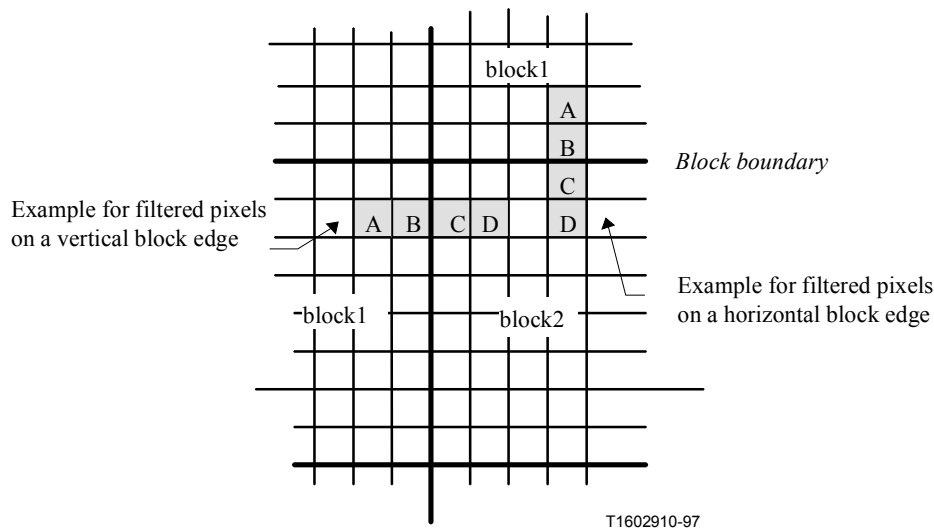


Figure J.1/H.263 – Examples of positions of filtered pixels

One or both of the following conditions must be fulfilled in order to apply the filter across a particular edge:

- Condition 1: block1 belongs to a coded macroblock ($COD==0 \parallel MB\text{-type} == INTRA$); or
- Condition 2: block2 belongs to a coded macroblock ($COD==0 \parallel MB\text{-type} == INTRA$).

If filtering is to be applied across the edge, A, B, C, and D shall be replaced by A1, B1, C1, D1 where:

$$B1 = \text{clip}(B + d1)$$

$$C1 = \text{clip}(C - d1)$$

$$A1 = A - d2$$

$$D1 = D + d2$$

$$d = (A - 4B + 4C - D) / 8$$

$$d1 = \text{UpDownRamp}(d, \text{STRENGTH})$$

$$d2 = \text{clipd1}((A - D) / 4, d1/2)$$

$$\text{UpDownRamp}(x, \text{STRENGTH}) =$$

$$\text{SIGN}(x) * (\text{MAX}(0, \text{abs}(x) - \text{MAX}(0, 2 * (\text{abs}(x) - \text{STRENGTH}))))$$

STRENGTH depends on QUANT and determines the amount of filtering. The relation between STRENGTH and QUANT is given in Table J.2.

QUANT = quantization parameter used for block2 if block2 belongs to a coded macroblock, or

QUANT = quantization parameter used for block1 if block2 does not belong to a coded macroblock (but block1 does).

Table J.2/H.263 – Relationship between QUANT and STRENGTH of filter

QUANT	STRENGTH	QUANT	STRENGTH
1	1	17	8
2	1	18	8
3	2	19	8
4	2	20	9
5	3	21	9
6	3	22	9
7	4	23	10
8	4	24	10
9	4	25	10
10	5	26	11
11	5	27	11
12	6	28	11
13	6	29	12
14	7	30	12
15	7	31	12
16	7		

The function clip(x) is defined according to 6.3.2, and the function clipd1(x, lim) clips x to the range $\pm \text{abs}(\text{lim})$. The symbol "/" denotes division by truncation toward zero.

Figure J.2 shows how the value of d1 varies as a function of d. As a result, the filter has an effect only if d is smaller than 2*STRENGTH (and different from zero). This is to prevent the filtering of strong true edges in the picture content. However, if the Reduced-Resolution Update mode is in use, STRENGTH is set to infinity, and as a result, the value of d1 is always equal to the value of d (see Q.7.2).

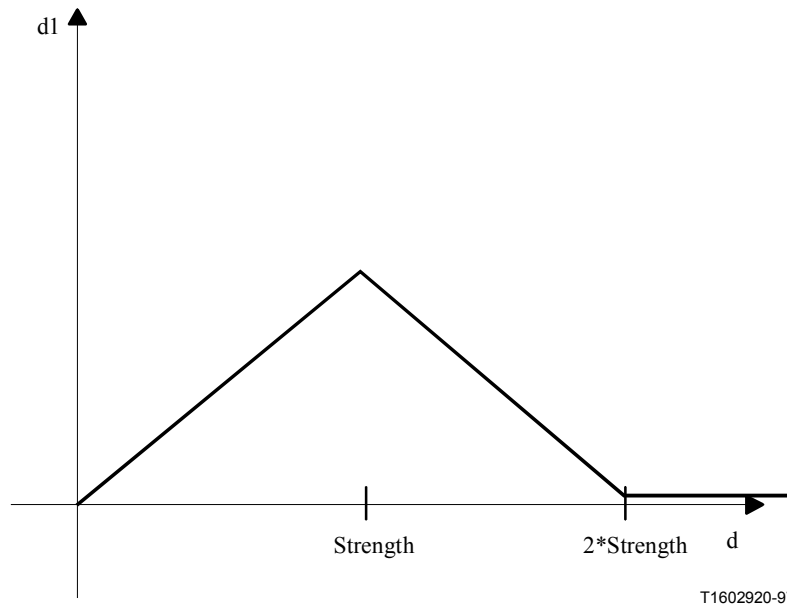


Figure J.2/H.263 – Parameter d1 as a function of parameter d for deblocking filter mode

The definition of d1 is designed to ensure that small mismatches between the encoder and decoder will remain small and will not build up over multiple pictures of a video sequence. This would be a problem, for example, with a condition that simply switches the filter on or off, because a mismatch of only ± 1 for d could then cause the filter to be switched on at the encoder side and off at the decoder side, or vice versa.

Due to rounding effects, the order of edges where filtering is performed must be specified.

Filtering across horizontal edges:

Basically this process is assumed to take place first. More precisely, the pixels $\begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix}$ that are used in

filtering across a horizontal edge shall not have been influenced by previous filtering across a vertical edge.

Filtering across vertical edges:

Before filtering across a vertical edge using pixels (A, B, C, D), all modifications of pixels (A, B, C, D) resulting from filtering across a horizontal edge shall have taken place.

Note that if one or more of the pixels (A, B, C, D) taking part in a filtering process are outside of a picture, no filtering takes place. Also, if the Independent Segment Decoding mode is in use (see Annex R) and one or more of the pixels (A, B, C, D) taking part in a filtering process are in different video picture segments (see I.3 concerning when a block is considered to be in the same video picture segment), no filtering is performed.

Annex K

Slice Structured mode

K.1 Introduction

This annex describes the optional Slice Structured mode of this Recommendation. The capability of this H.263 mode is signalled by external means (for example, ITU-T Rec. H.245). The use of this mode is indicated in the PLUSPTYPE field of the picture header. In order to facilitate optimal usage in a number of environments, this mode contains two submodes which are also capable of being signalled by external means (for example, ITU-T Rec. H.245). These two submodes are used to indicate whether or not rectangular slices will be used and/or whether the slices will be transmitted in sequential order or sent in an arbitrary order.

A slice is defined as a slice header followed by consecutive macroblocks in scanning order. An exception is for the slice that immediately follows the picture start code in the bitstream for a picture (which is not necessarily the slice starting with macroblock 0). In this case only part of the slice header is transmitted as described in K.2. The slice layer defines a video picture segment, and is used in place of the GOB layer in this optional mode. A slice video picture segment starts at a macroblock boundary in the picture and contains a number of macroblocks. Different slices within the same picture shall not overlap with each other, and every macroblock shall belong to one and only one slice.

This mode contains two submodes, which are signalled in the SSS field of the picture header:

- 1) The Rectangular Slice submode (RS): When RS is in use, the slice shall occupy a rectangular region of width specified by the SWI parameter of the slice header in units of macroblocks, and contains a number of macroblocks in scanning order within the rectangular region. When the Rectangular Slice submode is not in use, the SWI field is not present in the slice header and a slice contains a number of macroblocks in scanning order within the picture as a whole.
- 2) The Arbitrary Slice Ordering submode (ASO): When ASO is in use, the slices may appear in any order within the bitstream. When ASO is not in use, the slices must be sent in the (unique) order for which the MBA field of the slice header is strictly increasing from each slice to each subsequent slice in the picture.

Slice boundaries are treated differently from simple macroblock boundaries in order to allow slice header locations within the bitstream to act as resynchronization points for bit error and packet loss recovery and in order to allow out-of-order slice decoding within a picture. Thus, no data dependencies can cross the slice boundaries within the current picture, except for the Deblocking Filter mode which, when in use without the Independent Segment Decoding mode, filters across the boundaries of the blocks in the picture. However, motion vectors within a slice can cause data dependencies which cross the slice boundaries in the reference picture used for prediction purposes, unless the optional Independent Segment Decoding mode is in use.

The following rules are adopted to ensure that slice boundary locations can act as resynchronization points and to ensure that slices can be sent out of order without causing additional decoding delays:

- 1) The prediction of motion vector values are the same as if a GOB header were present (see 6.1.1), preventing the use of motion vectors of blocks outside the current slice for the prediction of the values of motion vectors within the slice.
- 2) The Advanced INTRA Coding mode (see Annex I) treats the slice boundary as if it were a picture boundary with respect to the prediction of INTRA block DCT coefficient values.

- 3) The assignment of remote motion vectors for use in overlapped block motion compensation within the Advanced Prediction mode also prevents the use of motion vectors of blocks outside the current slice for use as remote motion vectors (see F.3).

K.2 Structure of slice layer

The structure of the slice layer of the syntax is shown in Figure K.1 for all slices except the slice that immediately follows the picture start code in the bitstream for a picture. For the slice following the picture start code, only the emulation prevention bits (SEPB1, SEPB3, and conditionally SEPB2 as specified below), the MBA field and, when in the RS submode, also the SWI field are included.

SSTUF	SSC	SEPB1	SSBI	MBA	SEPB2	SQUANT	SWI	SEPB3	GFID	Macroblock Data
-------	-----	-------	------	-----	-------	--------	-----	-------	------	-----------------

Figure K.1/H.263 – Structure of slice layer

Refer to 5.2.5 for GFID, and to 5.3 for description of Macroblock layer.

K.2.1 Stuffing (SSTUF) (Variable length)

A codeword of variable length consisting of less than 8 bits. Encoders shall insert this codeword directly before an SSC codeword whenever needed to ensure that SSC is byte aligned. If SSTUF is present, the last bit of SSTUF shall be the last (least significant) bit of a byte, so that the start of the SSC codeword is byte aligned. Decoders shall be designed to discard SSTUF. Notice that 0 is used for stuffing within SSTUF.

K.2.2 Slice Start Code (SSC) (17 bits)

A word of 17 bits. Its value is 0000 0000 0000 0000 1. Slice start codes shall be byte aligned. This can be achieved by inserting SSTUF before the start code such that the first bit of the start code is the first (most significant) bit of a byte. The slice start code is not present for the slice which follows the picture start code.

K.2.3 Slice Emulation Prevention Bit 1 (SEPB1) (1 bit)

A single bit always having the value of "1", which is included to prevent start code emulation.

K.2.4 Slice Sub-Bitstream Indicator (SSBI) (4 bits)

A codeword of length four bits which is present only when CPM = "1" in the picture header. SSBI indicates the Sub-Bitstream number for the slice for Continuous Presence Multipoint and Video Multiplex operation, as described in Annex C. The mapping from the value of SSBI to the Sub-Bitstream number is shown in Table K.1. SSBI is not present for the slice which follows the picture start code.

Table K.1/H.263 – Values of SSBI and associated Sub-Bitstream numbers

Sub-Bitstream number	SSBI field value	GN value emulated
0	1001	25
1	1010	26
2	1011	27
3	1101	29

K.2.5 Macroblock Address (MBA) (5/6/7/9/11/12/13/14 bits)

A codeword having a length that is dependent on the current picture size and whether the Reduced-Resolution Update mode is in effect (see Annex Q). The bits are the binary representation of the macroblock number of the first macroblock in the current slice as counted from the beginning of the picture in scanning order, starting with macroblock number 0 at the upper left corner. MBA uniquely identifies which macroblock in the picture the current slice starts with. Codeword lengths for this codeword are provided in Table K.2. For custom picture sizes, the field width is given by the first entry in the table that has an equal or larger number of macroblocks, and the maximum value is the number of macroblocks in the current picture minus one. When in the Reduced-Resolution Update mode, the relevant picture size is the lower resolution update picture size rather than the picture size indicated in the picture header (see Annex Q).

Table K.2/H.263 – Specification of MBA parameter

Picture format	Default		RRU mode	
	Max value	Field width	Max value	Field width
sub-QCIF	47	6	11	5
QCIF	98	7	29	6
CIF	395	9	98	7
4CIF	1583	11	395	9
16CIF	6335	13	1583	11
2048 × 1152	9215	14	2303	12

K.2.6 Slice Emulation Prevention Bit 2 (SEPB2) (1 bit)

A single bit always having the value "1", which is included under certain conditions to prevent start code emulation. For slices other than the one following the picture start code, SEPB2 is included only if the MBA field width is greater than 11 bits and CPM = "0" in the picture header, or if the MBA field width is greater than 9 bits and CPM = "1" in the picture header. For the slice following the picture start code, SEPB2 is included only if the Rectangular Slice submode is in use.

K.2.7 Quantizer Information (SQUANT) (5 bits)

A fixed length codeword of five bits which indicates the quantizer QUANT to be used for that slice until updated by any subsequent DQUANT. The codewords are the natural binary representations of the values of QUANT which, being half the step sizes, range from 1 to 31. SQUANT is not present for the slice which follows the picture start code.

K.2.8 Slice Width Indication in Macroblocks (SWI) (3/4/5/6/7 bits)

A codeword which is present only if the Rectangular Slice submode is active, and having a length which depends on the current picture size and whether the Reduced-Resolution Update mode is active, as specified in Table K.3. For custom picture sizes, the field width is given by the next standard format size which is equal or larger in width (QCIF, CIF, ...), and the maximum value is the total number of macroblocks across the picture minus one. The last row of the table indicates the field width for picture sizes wider than 16CIF. SWI refers to the width of the current rectangular slice having its first macroblock (upper left) specified by MBA. The calculation of the actual slice width is given by:

$$\text{Actual Slice Width} = \text{SWI} + 1$$

When in the Reduced-Resolution Update mode, the relevant picture size is that of the lower resolution picture size for the update information, rather than the picture size indicated in the picture header.

Table K.3/H.263 – Specification of SWI parameter

Picture format	Default		RRU mode	
	Max value	Field width	Max value	Field width
sub-QCIF	7	4	3	3
QCIF	10	4	5	3
CIF	21	5	10	4
4CIF	43	6	21	5
16CIF	87	7	43	6
1412...2048 pixels wide	127	7	63	6

K.2.9 Slice Emulation Prevention Bit 3 (SEP3) (1 bit)

A single bit always having the value of "1" in order to prevent start code emulation.

Annex L

Supplemental enhancement information specification

L.1 Introduction

This annex describes the format of the supplemental enhancement information sent in the PSUPP field of the picture layer of this Recommendation. The capability of a decoder to provide any or all of the enhanced capabilities described in this annex may be signalled by external means (for example, ITU-T Rec. H.245). Decoders which do not provide the enhanced capabilities may simply discard any PSUPP information bits that appear in the bitstream. The presence of this supplemental enhancement information is indicated in PEI, and an additional PEI bit is inserted between every octet of PSUPP data, as described in 5.1.24 and 5.1.25.

In this annex, a distinction is made between the "decoded picture" and the "displayed picture". For purposes of this annex, the "displayed picture" is a picture having the same picture format as specified for the current picture by the picture layer of the video bitstream syntax. The "displayed picture" is constructed as described in this annex from the decoded picture, the prior displayed picture, the supplementary enhancement information described herein, and in some cases partly from a background picture which is externally controlled.

L.2 PSUPP format

The PSUPP data consists of a four-bit function type indication FTYPE, followed by a four-bit parameter data size specification DSIZE, followed by DSIZE octets of function parameter data, optionally followed by another function type indication, and so on. One function type indication value is defined as an escape code to provide for future extensibility to allow definition of more than fifteen different functions. A decoder which receives a function type indication which it does not support can discard the function parameter data for that function and then check for a

subsequent function type indication which may be supported. The defined FTYPE values are shown in Table L.1.

Table L.1/H.263 – FTYPE function type values

0	Reserved
1	Do Nothing
2	Full-Picture Freeze Request
3	Partial-Picture Freeze Request
4	Resizing Partial-Picture Freeze Request
5	Partial-Picture Freeze-Release Request
6	Full-Picture Snapshot Tag
7	Partial-Picture Snapshot Tag
8	Video Time Segment Start Tag
9	Video Time Segment End Tag
10	Progressive Refinement Segment Start Tag
11	Progressive Refinement Segment End Tag
12	Chroma Key Information
13	Reserved
14	Reserved
15	Extended Function Type

L.3 Do Nothing

No action is requested by the Do Nothing function. This function is used to prevent start code emulation. Whenever the last five or more bits of the final octet of the previous PSUPP octet are all zero and no additional PSUPP function requests are to be sent, the Do Nothing function shall be inserted into PSUPP to prevent the possibility of start code emulation. The Do Nothing function may also be sent when it is not required by rule expressed in the previous sentence. DSIZE shall be zero for the Do Nothing function.

L.4 Full-Picture Freeze Request

The full-picture freeze request function indicates that the contents of the entire prior displayed video picture shall be kept unchanged, without updating the displayed picture using the contents of the current decoded picture. The displayed picture shall then remain unchanged until the freeze picture release bit in the current PTYPE or in a subsequent PTYPE is set to 1, or until timeout occurs, whichever comes first. The request shall lapse due to timeout after five seconds or five pictures, whichever is a longer period of time. The timeout can be prevented by the issuance of another full-picture freeze request prior to or upon expiration of the timeout period (e.g., repeating the request in the header of the first picture with temporal reference indicating a time interval greater than or equal to five seconds since issuance, or in the header of the fifth picture after issuance). DSIZE shall be zero for the full-picture freeze request function.

L.5 Partial-Picture Freeze Request

The partial-picture freeze request function indicates that the contents of a specified rectangular area of the prior displayed video picture should be kept unchanged, without updating the specified area of the displayed picture using the contents of the current decoded picture. The specified area of the displayed picture shall then remain unchanged until the freeze picture release bit in the current

PTYPE or in a subsequent PTYPE is set to 1, until a partial-picture freeze-release request affecting the specified area is received, until the source format specified in a picture header differs from that of previous picture headers, or until timeout occurs, whichever comes first. Any change in the picture source format shall act as a freeze release for all active partial-picture freeze requests. The request shall lapse due to timeout after five seconds or five pictures, whichever is a longer period of time. The timeout can be prevented by the issuance of an identical partial-picture freeze request prior to or upon expiration of the timeout period (e.g., repeating the request in the header of the first picture with temporal reference indicating a time interval greater than or equal to five seconds since issuance, or in the header of the fifth picture after issuance). DSIZE shall be equal to 4 for the partial-picture freeze request. The four octets of PSUPP that follow contain the horizontal and vertical location of the upper left corner of the frozen picture rectangle, and the width and height of the rectangle, respectively, using eight bits each and expressed in units of eight pixels. For example, a 24-pixel wide and 16 pixel tall area in the upper left corner of the video display is specified by the four parameters (0, 0, 3, 2).

L.6 Resizing Partial-Picture Freeze Request

The resizing partial-picture freeze request function indicates that the contents of a specified rectangular area of the prior displayed video picture should be resized to fit into a smaller part of the displayed video picture, which should then be kept unchanged, without updating the specified area of the displayed picture using the contents of the current decoded picture. The specified area of the displayed picture shall then remain unchanged until the freeze release bit in the current PTYPE or in a subsequent PTYPE is set to 1, until a partial-picture freeze-release request affecting the specified area is received, until the source format specified in a picture header differs from that of previous picture headers, or until timeout occurs, whichever comes first. Any change in the picture source format shall act as a freeze release for all active resizing partial-picture freeze requests. The request shall lapse due to timeout after five seconds or five pictures, whichever is a longer period of time. The timeout can be prevented by the issuance of a partial-picture freeze request for the affected area of the displayed picture prior to or upon expiration of the timeout period (e.g., issuing a partial-picture freeze request in the header of the first picture with temporal reference indicating a time interval greater than or equal to five seconds since issuance, or in the header of the fifth picture after issuance). DSIZE shall be equal to 8 for the resizing partial-picture freeze request. The eight octets of PSUPP data that follow contain 32 bits used to specify the rectangular region of the affected area of the displayed picture, and then 32 bits used to specify the corresponding rectangular region of the affected area of the decoded picture. The width and height of the rectangular region in the decoded picture shall both be equal to 2^i times the width and height specified for the rectangular region in the displayed picture, where i is an integer in the range of 1 to 8. The location and size of each of these two rectangular regions is specified using the same format as such a region is specified in the partial-picture freeze request function.

L.7 Partial-Picture Freeze-Release Request

The partial-picture freeze-release request function indicates that the contents of a specified rectangular area of the displayed video picture shall be updated by the current and subsequent decoded pictures. DSIZE shall be equal to 4 for the partial-picture freeze-release request. The four octets of PSUPP data that follow specify a rectangular region of the displayed picture in the same format as such a region is specified in the partial-picture freeze request function.

L.8 Full-Picture Snapshot Tag

The full-picture snapshot tag function indicates that the current picture is labelled for external use as a still-image snapshot of the video content. DSIZE shall be equal to 4 for the full-picture snapshot tag function. The four octets of PSUPP data that follow specify a snapshot identification number for external use.

L.9 Partial-Picture Snapshot Tag

The partial-picture snapshot tag function indicates that a specified rectangular area of the current picture is labelled for external use as a still-image snapshot of the video content. DSIZE shall be equal to 8 for the partial-picture snapshot tag function. The first four octets of PSUPP data that follow specify a snapshot identification number for external use, and the remaining four octets of PSUPP data that follow specify a rectangular region of the decoded picture in the same format as such a region is specified in the partial-picture freeze request function.

L.10 Video Time Segment Start Tag

The video time segment start tag function indicates that the beginning of a specified sub-sequence of video data is labelled as a useful section of video content for external use, starting with the current picture. The tagged sub-sequence of video data shall continue until stopped by the receipt of a matching video time segment end tag function or until timeout, whichever comes first. The tagged sub-sequence shall end due to timeout after five seconds or five pictures, whichever is a longer period of time. The timeout can be prevented by the issuance of an identical video time segment start tag function prior to or upon expiration of the timeout period (e.g., repeating the video time segment start tag function in the header of the first picture with temporal reference indicating a time interval greater than or equal to five seconds since issuance, or in the header of the fifth picture after issuance). DSIZE shall be equal to 4 for the video time segment start tag function. The four octets of PSUPP data that follow specify a video time segment identification number for external use.

L.11 Video Time Segment End Tag

The video time segment end tag function indicates that the end of a specified sub-sequence of video data is labelled as a useful section of video content for external use, ending with the previous picture. DSIZE shall be equal to 4 for the video time segment start tag function. The four octets of PSUPP data that follow specify a video time segment identification number for external use.

L.12 Progressive Refinement Segment Start Tag

The progressive refinement segment start tag function indicates the beginning of a specified sub-sequence of video data which is labelled as the current picture followed by a sequence of zero or more pictures of refinement of the quality of the current picture, rather than as a representation of a continually moving scene. The tagged sub-sequence of video data shall continue until stopped by the receipt of a matching progressive refinement segment end tag function or until timeout, whichever comes first. The tagged sub-sequence shall end due to timeout after five seconds or five pictures, whichever is a longer period of time. The timeout can be prevented by the issuance of an identical progressive refinement segment start tag function prior to or upon expiration of the timeout period (e.g., repeating the progressive refinement start tag function in the header of the first picture with temporal reference indicating a time interval greater than or equal to five seconds since issuance, or in the header of the fifth picture after issuance). DSIZE shall be equal to 4 for the video time segment start tag function. The four octets of PSUPP data that follow specify a progressive refinement segment identification number for external use.

L.13 Progressive Refinement Segment End Tag

The progressive refinement segment end tag function indicates the end of a specified sub-sequence of video data which is labelled as an initial picture followed by a sequence of zero or more pictures of the refinement of the quality of the initial picture, and ending with the previous picture. DSIZE shall be equal to 4 for the video time segment start tag function. The four octets of PSUPP data that follow specify a progressive refinement identification number for external use.

L.14 Chroma Keying Information

The Chroma Keying Information Function (CKIF) indicates that the "chroma keying" technique is used to represent "transparent" and "semi-transparent" pixels in the decoded video pictures. When being presented on the display, "transparent" pixels are not displayed. Instead, a background picture which is either a prior reference picture or is an externally controlled picture is revealed. Semi-transparent pixels are displayed by blending the pixel value in the current picture with the corresponding value in the background picture. One octet is used to indicate the keying color value for each component (Y , C_B , or C_R) which is used for chroma keying. To represent pixels that are to be "semi-transparent", two threshold values, denoted as T_1 and T_2 , are used. Let α denote the transparency of a pixel; $\alpha = 255$ indicates that the pixel is opaque, and $\alpha = 0$ indicates that the pixel is transparent. For other values of α , the resulting value for a pixel should be a weighted combination of the pixel value in the current picture and the pixel value from the background picture (which is specified externally). The values of α may be used to form an image that is called an "alpha map." Thus, the resulting value for each component may be:

$$[\alpha \cdot X + (255 - \alpha) \cdot Z] / 255$$

where X is the decoded pixel component value (for Y , C_B , or C_R), and Z is the corresponding pixel component value from the background picture.

The α value can be calculated as follows. First, the distance of the pixel color from the key color value is calculated:

$$d = A_Y(X_Y - K_Y)^2 + A_B(X_B - K_B)^2 + A_R(X_R - K_R)^2$$

in which X_Y , X_B , and X_R are the Y , C_B , and C_R values of the decoded pixel color, K_Y , K_B , and K_R are the corresponding key color parameters, and A_Y , A_B , and A_R are keying flag bits which indicate which color components are used as keys. Once the distance d is calculated, the α value may be computed as specified in the following pseudo-code:

```
for each pixel
  if (d < T1) then  $\alpha = 0$ ;
  else if (d > T2) then  $\alpha = 255$ ;
  else  $\alpha = [255 \cdot (d - T_1)] / (T_2 - T_1)$ 
```

However, the precise method for performing the chroma keying operation in the decoder is not specified herein, since normative specification of the method is not needed for interoperability. The process described here is provided for illustration purposes in order to convey the intended interpretation of the data parameters.

Since the derived α value is simply a function of X_Y , X_B , and X_R , a Look-Up Table (LUT) can be built to achieve the above operation. Such an LUT has $2^{8 \times N}$ entries corresponding to all pixel values, where N is the number of color components used as keys. Each entry in the LUT would then contain the corresponding α value.

DSIZE shall be in the range of 1 to 9 (inclusive) for chroma keying information, according to the amount of data sent with the CKIF. No more than one CKIF shall be sent with a picture.

The first octet following the DSIZE octet shall contain the representation order of the current picture – streams having a lower representation order are assumed to form the background picture for streams having a higher representation order.

If DSIZE is greater than one, the next octet after the representation order octet shall be used to send six flag bits defined as:

- bit 1: A_Y : A flag bit indicating the presence of a K_Y key parameter for luminance Y values
- bit 2: A_B : A flag bit indicating the presence of a K_B key parameter for chroma C_B values
- bit 3: A_R : A flag bit indicating the presence of a K_R key parameter for chroma C_R values
- bit 4: A_1 : A flag bit indicating the presence of a T_1 threshold parameter for transparency
- bit 5: A_2 : A flag bit indicating the presence of a T_2 threshold parameter for opacity
- bit 6: RPB: A flag bit indicating the use of the reference picture as a background picture
- bit 7: reserved
- bit 8: reserved

DSIZE shall be equal to 1 or shall be equal to 2 plus the number of flag bits among A_Y , A_B , and A_R which are set to 1, plus 2 times the number of bits among A_1 and A_2 which are set to 1. If DSIZE is greater than 1, then an additional octet shall be sent to specify the value of each color component for each of the flag bits A_Y , A_B , and A_R which are set to one and two additional octets shall be sent to specify each of the flagged threshold values among T_1 and T_2 . These octets shall follow in the same order as the flag bits.

If DSIZE is equal to 1, or if all three keying color flag bits A_Y , A_B , and A_R , are zero, the keying flag bits A_Y , A_B , and A_R and keying colours K_Y , K_B , and K_R which were used for the previous keyed picture should also be used for the current picture. If no previous values have been sent for the video sequence, the default keying flag bits $A_Y = 1$, $A_B = 1$, and $A_R = 1$ and the default key colours $K_Y = 50$, $K_B = 220$, and $K_R = 100$ should be used as the previous values.

If DSIZE is equal to 1, or if both of the keying threshold flag bits A_1 and A_2 are zero, the keying threshold values T_1 and T_2 which were used for the previous keyed picture should also be used for the current picture. If no previous values have been sent for the video sequence, the default threshold values $T_1 = 48$ and $T_2 = 75$ should be used as the previous values.

In the portion where the pixels are "semi-transparent" (i.e., when $T_1 < d < T_2$), the decoded pixels typically contain the chroma key color in components where chroma keys are used. This may result in certain color artifacts. To address this problem, these pixel values may be adjusted before they are blended with the background color. Such a correction process may be applied to color components being used in the chroma key operation as indicated by the flag bits. The process is as follows:

$$X' = K + (T_2 / d)(X - K)$$

where X is the original decoded pixel component value, and X' is the corrected value.

Since the adjusted pixel values X'_Y , X'_B , and X'_R are functions of X_Y , X_B , and X_R , the color correction can be achieved by using an LUT. This LUT would have 2^{8N} entries corresponding to all pixel values, where N is the number of color components used as keys. Each entry would then contain the corresponding corrected values.

If the Reference Picture Background (RPB) flag bit is set to "1", this indicates that the temporally previous reference picture (prior to any Annex P resampling performed for the current picture) should be held as the (opaque) background of the current picture and of all subsequent chroma keyed pictures, until replaced by the arrival of another picture having an RPB flag set to "1". If the current picture has no temporally previous reference picture (i.e, if the current picture is a picture of type INTRA or EI), the picture to which the RPB flag bit refers is the picture which would normally

have been the reference picture if the current picture were of type INTER or EP as appropriate. If the RPB flag bit is set to "0", this indicates that background should remain as previously controlled (either under external control or using a reference picture which was previously stored upon receipt of a prior picture having RPB set to "1").

The use of the chroma keying which is invoked by the issuance of the chroma keying information function shall start with the current picture and shall continue until a subsequent picture of type INTRA or EI occurs or until a timeout period expires, whichever comes first. The use of chroma keying shall end due to timeout after five seconds or five pictures, whichever is a longer period of time. The timeout can be prevented by the issuance of an identical chroma keying information function prior to or upon expiration of the timeout period (e.g., repeating the chroma keying information function in the header of the first picture with temporal reference indicating a time interval greater than or equal to five seconds since issuance, or in the header of the fifth picture after issuance). The encoder shall send sufficient information with the chroma keying information function for complete resynchronization to occur with each picture of type INTRA or EI and within each timeout interval (it shall not rely on using stored or default values of the key colours or thresholds).

L.15 Extended function type

The extended function type indication is used to signal that the following PSUPP octet contains an extended function. The usage of extended functions is reserved for the ITU to have a later ability to define a larger number of backward-compatible PSUPP data functions. DSIZE shall be equal to zero for the extended function type indication. In order to allow backward compatibility of future use of the extended function type indication, decoders shall treat the second set of four bits in the octet which follows the extended function type indication as a DSIZE value indicating the number of subsequent octets of PSUPP that are to be skipped for extended function parameter data, which may be followed by additional FTYPE indications.

Annex M

Improved PB-frames mode

M.1 Introduction

This annex describes an optional Improved PB-frames mode of this Recommendation. It is therefore considered to be advantageous to use the present Improved PB-frames mode instead of the PB-frames mode defined in Annex G. The capability of this mode is signalled by external means (for example, ITU-T Rec. H.245). The use of this mode is indicated in the PLUSPTYPE field of the picture header.

Most parts of this option are similar to the PB-frame option defined in Annex G.

To avoid confusion with B-pictures as defined in Annex O, the terms B-picture, B-macroblock, and B-block will not be used in this annex. Instead, we will use the notation B_{PB} to mean the "B-part" of an Improved PB-frame. When reference is made to Annex G, B-picture and B-block shall be read as B_{PB} -picture and B_{PB} -block.

The main difference between the PB-frames mode and the Improved PB-frames mode is that in the Improved PB-frames mode, the B_{PB} -macroblock has available a forward and a backward prediction mode in addition to the bidirectional prediction mode. In this annex, MVDB (when present) refers to a forward motion vector. (Note that, in Annex G, MVDB was used to refer to an enhancement of the downsampled forward and backward vectors for bidirectional prediction, rather than a distinct forward motion vector.)

All the differences from Annex G are identified in this annex. When nothing is indicated, it means that the same procedure as described in Annex G is used.

M.2 B_{PB}-macroblock prediction modes

There are three different ways of coding a B_{PB}-macroblock. The different coding modes are signalled by the parameter MODB. The B_{PB}-macroblock coding modes are:

M.2.1 Bidirectional prediction

In the bidirectional prediction mode, prediction uses the reference pictures before and after the B_{PB}-picture (in the case of a sequence of Improved PB-frames, this means the P-picture part of the temporally previous Improved PB-frame and the P-picture part of the current Improved PB-frame). This prediction is equivalent to the prediction defined in Annex G when $MV_D = 0$. Notice that in this mode (and only in this mode), Motion Vector Data (MVD) of the PB-macroblock must be included even if the P-macroblock is INTRA coded. (Notice the difference between MVD – motion vector data – and MV_D – delta vector – defined in Annex G.)

M.2.2 Forward prediction

In the forward prediction mode, the vector data contained in MVDB are used for forward prediction from the previous reference picture (an INTRA or INTER picture, or the P-picture part of a PB or Improved PB-frame). This means that there is always only one 16×16 vector for the B_{PB}-macroblock in this prediction mode.

A simple predictor is used for coding of the forward motion vector. The rule for this predictor is that if the current macroblock is not at the far left edge of the picture or slice and the macroblock to the left has a forward motion vector, then the predictor of the forward motion vector for the current macroblock is set to the value of the forward motion vector of the block to the left; otherwise, the predictor is set to zero. The difference between the predictor and the desired motion vector is then VLC coded in the same way as vector data to be used for the P-picture (MVD).

Concerning motion vectors over picture boundaries defined in D.1, the described technique also applies for the forward B_{PB}-vector if this feature is in use (this applies for forward as well as for bidirectional prediction mode).

M.2.3 Backward prediction

In the backward prediction mode, the prediction of the B_{PB} macroblock is identical to P_{REC} (defined in G.5). No motion vector data is used for the backward prediction.

M.3 Calculation of vectors for bidirectional prediction of a the B-macroblock

In case bidirectional prediction is used, the scaled forward and backward vectors are calculated as described in Annex G when $MV_D = 0$.

M.4 MODB table

A new definition for MODB (replacing Table 11) is shown in Table M.1. It indicates the possible coding modes for a B_{PB}-block.

Table M.1/H.263 – MODB table for Improved PB-frames mode

Index	CBPB	MVDB	Number of bits	Code	Coding mode
0			1	0	Bidirectional prediction
1	x		2	10	Bidirectional prediction
2		x	3	110	Forward prediction
3	x	x	4	1110	Forward prediction
4			5	11110	Backward prediction
5	x		5	11111	Backward prediction

NOTE – The symbol "x" indicates that the associated syntax element is present.

Annex N

Reference Picture Selection mode

N.1 Introduction

This annex describes the optional Reference Picture Selection mode of this Recommendation, which operates using a modified interframe prediction method called "NEWPRED." The capability of this H.263 mode is signalled by external means (for example, ITU-T Rec. H.245). The amount of additional picture memory accommodated in the decoder may also be signalled by external means to help the memory management at the encoder. This mode can use backward channel messages sent from a decoder to an encoder to inform the encoder which part of which pictures have been correctly decoded at the decoder. The use of this mode is indicated in the PLUSPTYPE field of the picture header. This mode has two back-channel mode switches which define whether a backward channel is used and what kind of messages are returned on that backward channel from the decoder, and has another submode defined in terms of the channel for the backward channel messages.

The two back-channel mode switches of this mode determine the type of messages sent on the back-channel, specifying whether ACK (Acknowledgment messages), or NACK (Non-Acknowledgment messages) are sent. Together, the two switches define four basic methods of operation:

- 1) NEITHER, in which no back-channel data is returned from the decoder to the encoder;
- 2) ACK, in which the decoder returns only acknowledgment messages;
- 3) NACK, in which the decoder returns only non-acknowledgment messages; and
- 4) ACK+NACK, in which the decoder returns both acknowledgment and non-acknowledgment messages.

The specific type of messages to be sent as outlined above is indicated in the picture header.

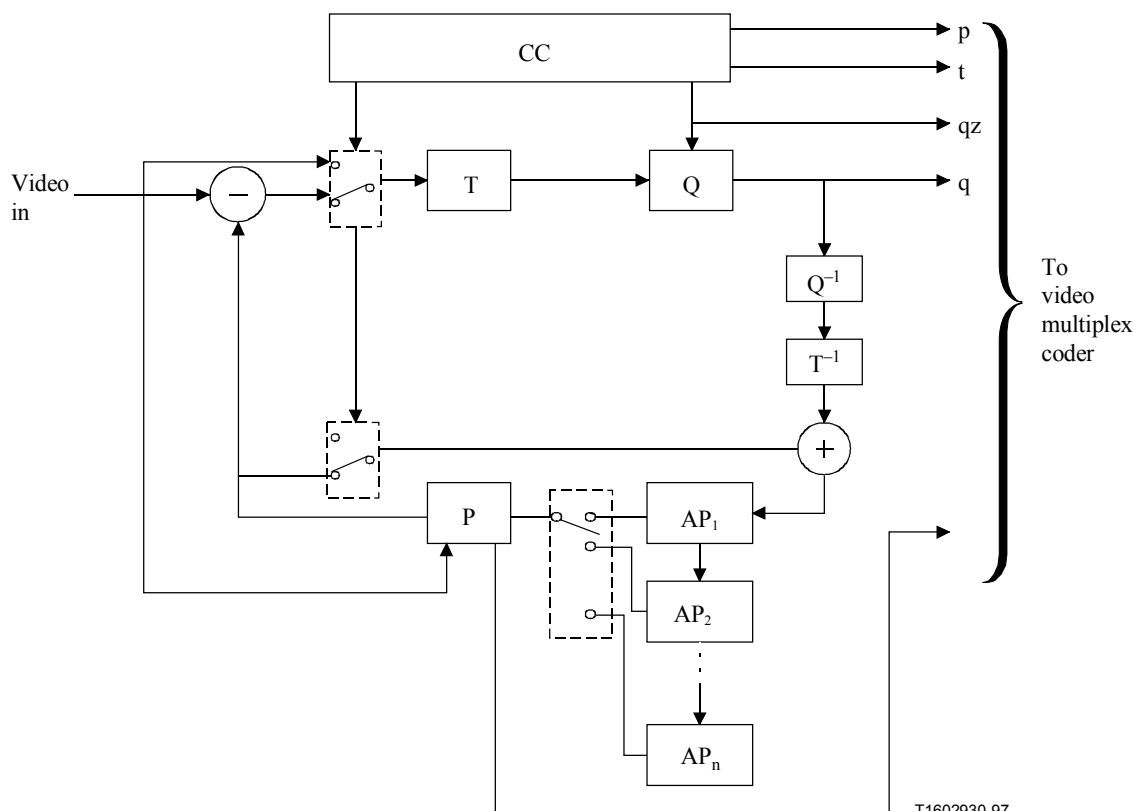
There are also two methods of operation in terms of the channel for backward channel messages:

- 1) Separate Logical Channel mode: This method of operation delivers back-channel data through a separate logical channel in the multiplex layer of the system; and
- 2) VideoMux mode: This method of operation delivers back-channel data for received video within the forward video data of a video stream of encoded data.

This annex specifies a syntax for the backward channel messages as well as for forward-channel data.

N.2 Video source coding algorithm

The source coder of this mode is shown in generalized form in Figure N.1. This figure shows a structure which uses a number of picture memories. The source coder may select one of the picture memories to suppress the temporal error propagation due to the inter-frame coding. The Independent Segment Decoding mode (see Annex R), which treats boundaries of GOBs with non-empty headers or slices as picture boundaries, can be used to avoid error propagation due to motion compensation across the boundaries of the GOBs or slices when this mode is applied to a smaller unit than a picture, such as a GOB or slice. The information to signal which picture is selected for prediction is included in the encoded bitstream. The strategy used by the encoder to select the picture to be used for prediction is out of the scope of this Recommendation.



T	Transform
Q	Quantizer
P	Picture Memory with motion compensated variable delay
AP	Additional Picture Memory
CC	Coding control
p	Flag for INTRA/INTER
t	Flag for transmitted or not
qz	Quantizer indication
q	Quantizing index for transform coefficients
v	Motion vector

Figure N.1/H.263 – Source coder for NEWPRED

N.3 Channel for back-channel messages

This mode has two methods of operation in terms of the type of channel for back-channel messages. One is the separate logical channel mode and the other is the videomux mode. The separate logical channel mode is a preferred mode and delivers the back-channel message defined in N.4.2 through the dedicated logical channel. The videomux mode is prepared for the system which cannot set up the separate extra channel for the back-channel messages due to the restriction of the number of

channels' combinations. The videomux mode delivers the back-channel messages through the same logical channel with the forward video data in the opposite direction.

N.3.1 Separate logical channel mode

The separate logical channel mode delivers back-channel messages through a dedicated logical channel opened only for the purpose of the back-channel messages. The association mechanism with the forward channel which delivers video data is provided by external means (for example, ITU-T Rec. H.245). Separate logical channel operation requires an external framing mechanism for synchronization of the messages within the back channel, as the back-channel syntax defined herein contains no synchronization flag words.

N.3.2 Videomux mode

The videomux mode delivers the back-channel messages through the same logical channel with the forward video data in the opposite direction. The syntax of the multiplexed bitstream is described in N.4.1. The back-channel messages may be inserted by using the Back-Channel message Indication (BCI) in the GOB or slice header.

N.4 Syntax

N.4.1 Forward channel

The syntax for the forward-channel data which conveys the compressed video signal is altered only in the Group of Blocks (GOB) or slice layer.

The syntax of the GOB layer is illustrated in Figure N.2. The fields of TRI, TR, TRPI, TRP, BCI, and BCM are added to Figure 9.

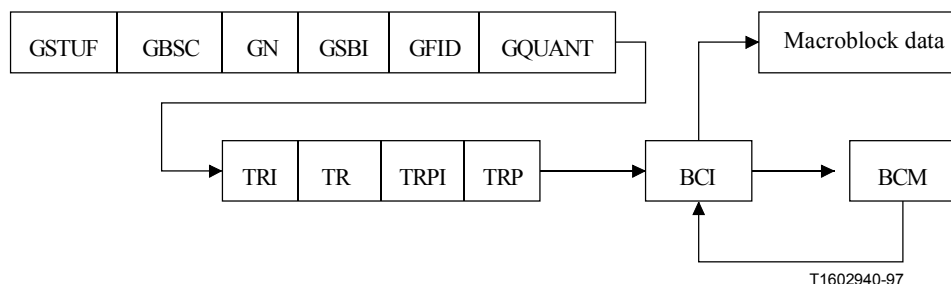


Figure N.2/H.263 – Structure of GOB layer for NEWPRED

When the optional Slice Structured mode (see Annex K) is in use, the syntax of the slice layer is modified in the same way as the GOB layer. The syntax is illustrated Figure N.3.

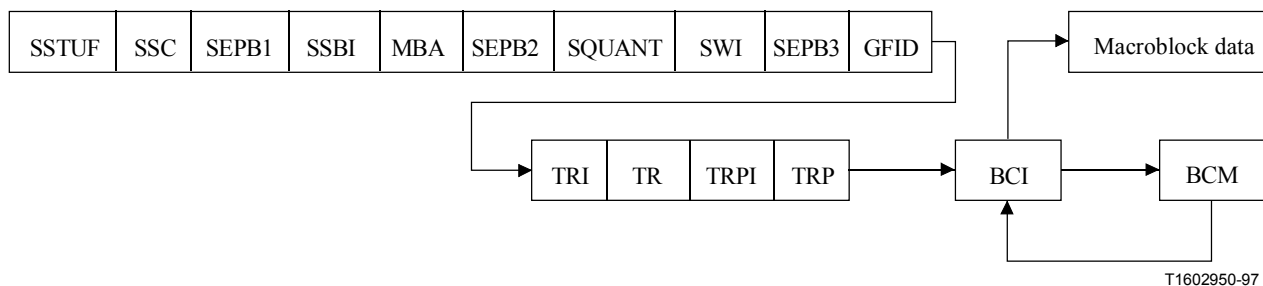


Figure N.3/H.263 – Structure of slice layer for NEWPRED

N.4.1.1 Temporal Reference Indicator (TRI) (1 bit)

TRI indicates whether or not the following TR field is present.

0: TR field is not present.

1: TR field is present.

N.4.1.2 Temporal Reference (TR) (8/10 bits)

When present, TR is an eight-bit number unless a custom picture clock frequency is in use, in which case it is a ten-bit number consisting of the concatenation of ETR and TR of the picture header.

N.4.1.3 Temporal Reference for Prediction Indicator (TRPI) (1 bit)

TRPI indicates whether or not the following TRP field is present:

0: TRP field is not present.

1: TRP field is present.

TRPI shall be equal to zero whenever the picture is an I- or EI-picture.

N.4.1.4 Temporal Reference for Prediction (TRP) (10 bits)

When present (as indicated in TRPI), TRP indicates the Temporal Reference which is used for prediction of the encoding, except in the case of B-pictures and the B-picture part of an Improved PB-frame. For B-pictures or the B-picture part of an Improved PB-frame, the picture having the temporal reference TRP is used for the prediction in the forward direction. (Prediction in the reverse-temporal direction always uses the immediately temporally subsequent picture.) TRP is a ten-bit number. If a custom picture clock frequency was not in use for the reference picture, the two MSBs of TRP are zero and the LSBs contain the eight-bit TR found in the picture header of the reference picture. If a custom picture clock frequency was in use for the reference picture, TRP is a ten-bit number consisting of the concatenation of ETR and TR from the reference picture header.

When TRP is not present, the most recent temporally previous anchor picture shall be used for prediction, as when not in the Reference Picture Selection mode. TRP is valid until the next PSC, GSC, or SSC.

N.4.1.5 Back-Channel message Indication (BCI) (Variable length)

This field contains one or two bits; when set to "1", it signals the presence of the following video Back-Channel Message (BCM) field. Otherwise, the field value is "01", which indicates the absence or the end of the video back-channel message field. Combinations of BCM and BCI may not be present, and may be repeated when present. BCI shall always be set to "01" if the videomux mode is not in use.

N.4.1.6 Back-Channel Message (BCM) (Variable length)

The Back-Channel Message having syntax as defined in N.4.2, which is present only if the preceding BCI field is set to "1".

N.4.2 Back-Channel Message (BCM) syntax

The syntax for the back-channel which conveys the acknowledgment/non-acknowledgment messages is illustrated in Figure N.4. This message is returned from a decoder to an encoder in order to tell whether a forward-channel data was correctly decoded or not.

BT	URF	TR	ELNUMI	ELNUM	BCPM	BSBI	BEPB1	GN/MBA	BEPB2	RTR	BSTUF
----	-----	----	--------	-------	------	------	-------	--------	-------	-----	-------

Figure N.4/H.263 – Structure of Back-Channel Message (BCM) syntax for NEWPRED

N.4.2.1 Back-channel message Type (BT) (2 bits)

Back-channel message type indicates if the corresponding part of the encoded message is correctly decoded or not. Which type of message is required for the encoder is indicated in the picture header of the forward channel.

00: Reserved for future use.

01: Reserved for future use.

10: NACK. This indicates the erroneous decoding of the corresponding part of the forward-channel data.

11: ACK. This indicates the correct decoding of the corresponding part of the forward-channel data.

N.4.2.2 Unreliable Flag (URF) (1 bit)

The Unreliable Flag is set to 1 when a reliable value for TR or GN/MBA is not available to the decoder. (When BT is NACK, a reliable TR may not be available at the decoder.)

0: Reliable.

1: Unreliable.

N.4.2.3 Temporal Reference (TR) (10 bits)

Temporal reference contains the TR information of the video picture segment for which the ACK/NACK is indicated in the back-channel message.

NOTE – The meaning of the term "video picture segment" as used herein is defined in Annex R. If a custom picture clock frequency was not in use for the reference picture, the two MSBs of TR are zero and the LSBs contain the eight-bit TR found in the picture header of the reference picture. If a custom picture clock frequency was in use for the reference picture, TR is a ten-bit number consisting of the concatenation of ETR and TR from the reference picture header.

N.4.2.4 Enhancement Layer Number Indication (ELNUMI) (1 bit)

The enhancement layer number indication is "0" unless the optional Temporal, SNR and Spatial Scalability mode (Annex O) is used in the forward-channel data and some enhancement layers of the forward channel are combined on one logical channel and the back-channel message refers to an enhancement layer (rather than the base layer), in which case the enhancement layer number indication shall be "1".

N.4.2.5 Enhancement Layer Number (ELNUM) (4 bits)

Enhancement layer number is present if and only if ELNUMI is "1", in which case it contains the layer number of the enhancement layer referred to in the back-channel message.

N.4.2.6 BCPM (1 bit)

BCPM is "0" unless the CPM mode is used in the forward-channel data, in which case it is "1". If BCPM is "1", this indicates that BSBI is present.

N.4.2.7 Back-channel Sub-Bitstream Indicator (BSBI) (2 bits)

A fixed length codeword of 2 bits that is only present if BCPM is "1". The BSBI is the natural binary representation of the appropriate Sub-Bitstream number in the forward-channel data for which the ACK/NACK message is indicated in the back-channel message as described in 5.2.4 and in Annex C.

N.4.2.8 Back-channel Emulation Prevention Bit 1 (BEPB1) (1 bit)

A field which is present if and only if the videomux mode is in use. This field is always set to "1" to prevent a start code emulation.

N.4.2.9 GOB Number/Macroblock Address (GN/MBA) (5/6/7/9/11/12/13/14 bits)

A GOB number or macroblock address is present in this field. If the optional Slice Structured mode (see Annex K) is not in use, this field contains the GOB number of the beginning of the video picture segment for which the NACK/ACK message is indicated in the back-channel message. If the optional Slice Structured mode is in use, this field contains the macroblock address of the beginning of the slice for which the NACK/ACK message is indicated in the back-channel message. The length of this field is the length specified elsewhere in this Recommendation for GN or MBA.

NOTE – When this field is received in the videomux mode, the use of the optional Slice Structured mode refers to the use of this mode in the video bitstream to which the BCM applies and not in the video bitstream that transports the BCM data.

N.4.2.10 Back-channel Emulation Prevention Bit 2 (BEPB2) (1 bit)

A field which is present if and only if the videomux mode is in use. This field is always set to "1" to prevent a start code emulation.

N.4.2.11 Requested Temporal Reference (RTR) (10 bits)

Requested temporal reference is present only if BT is NACK. RTR indicates the requested temporal reference of the GOB or slice associated with the NACK. Typically, it is the TR of the last correctly decoded video picture segment of the corresponding position at the decoder. If a custom picture clock frequency was not in use for the requested reference picture, the two MSBs of RTR are zero and the LSBs contain the eight-bit TR found in the picture header of the requested reference picture. If a custom picture clock frequency was in use for the requested reference picture, RTR is a ten-bit number consisting of the concatenation of ETR and TR from the requested reference picture header.

N.4.2.12 Stuffing (BSTUF) (Variable length)

This field is present if and only if the separate logical channel mode is in use and the back-channel message is the last in an external frame. BSTUF consists of a codeword of variable length consisting of zero or more bits of value "0". This field is only present at the end of an external frame.

N.5 Decoder process

The decoder of this mode may need an additional number of picture memories to store the correctly decoded video signals and their Temporal Reference (TR) information. The decoder uses the stored picture for which the TR is TRP as the reference picture for inter-frame decoding instead of the last decoded picture, if the TRP field exists in the forward-channel data. When the picture for which the TR is TRP is not available at the decoder, the decoder may send a forced INTRA update signal to the encoder by external means (for example, ITU-T Rec. H.245). Unless a different frame storage policy is negotiated by external means, correctly decoded video picture segments shall be stored into memory for use as later reference pictures on a first-in, first-out basis as shown in Figure N.1 (except for B-pictures, which are not used as reference pictures), and video picture segments which are detected as having been incorrectly decoded should not replace correctly decoded ones in this memory area.

An Acknowledgment message (ACK) and a Non-Acknowledgment message (NACK) are defined as back-channel messages. An ACK may be returned when the decoder decodes a video picture segment successfully. NACKs may be returned when the decoder fails to decode a video picture segment, and may continue to be returned until the decoder gets the expected forward channel data which includes the requested TRP or an INTRA update. Which types of message shall be sent is indicated in the RPSMF field of the picture header of the forward-channel data.

In a usage scenario known as "Video Redundancy Coding", the Reference Picture Selection mode may be used by some encoders in a manner in which more than one representation is sent for the pictured scene at the same temporal instant (usually using different reference pictures). In such a

case in which the Reference Picture Selection mode is in use and in which adjacent pictures in the bitstream have the same temporal reference, the decoder shall regard this occurrence as an indication that redundant copies have been sent of approximately the same pictured scene content, and shall decode and use the first such received picture while discarding the subsequent redundant picture(s).

Annex O

Temporal, SNR, and Spatial Scalability mode

This annex describes the optional mode of this Recommendation in support of Temporal, SNR, and Spatial Scalability. This mode may also be used in conjunction with error control schemes. The capability of this mode and the extent to which its features are supported is signalled by external means (for example, ITU-T Rec. H.245). The use of this mode is indicated in PLUSPTYPE.

O.1 Overview

Scalability allows for the decoding of a sequence at more than one quality level. This is done by using a hierarchy of pictures and enhancement pictures partitioned into one or more layers. There are three types of pictures used for scalability: B-, EI-, and EP-pictures, as explained below. Each of these has an enhancement layer number ELNUM that indicates to which layer it belongs, and a reference layer number RLNUM that indicates which layer is used for its prediction. The lowest layer is called the base layer, and has layer number 1.

Scalability is achieved by three basic methods: temporal, SNR, and spatial enhancement.

O.1.1 Temporal scalability

Temporal scalability is achieved using bidirectionally predicted pictures, or B-pictures. B-pictures allow prediction from either or both a previous and subsequent reconstructed picture in the reference layer. This property generally results in improved compression efficiency as compared to that of P-pictures. These B-pictures differ from the B-picture part of a PB- (or Improved PB-) frame (see Annexes G and M) in that they are separate entities in the bitstream: they are not syntactically intermixed with a subsequent P- (or EP-) picture.

B-pictures (and the B-part of PB- or Improved PB-frames) are not used as reference pictures for the prediction of any other pictures. This property allows for B-pictures to be discarded if necessary without adversely affecting any subsequent pictures, thus providing temporal scalability. Figure O.1 illustrates the predictive structure of P- and B-pictures.

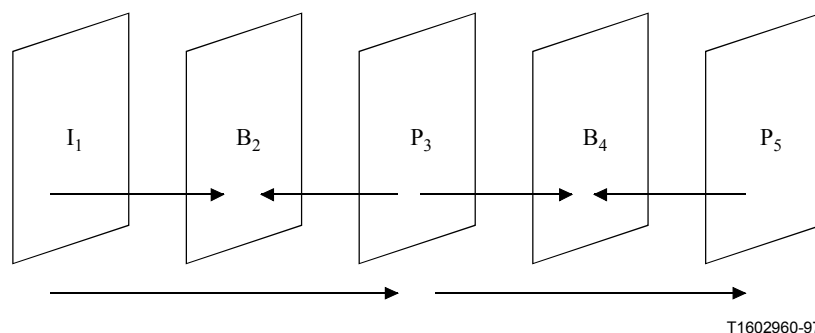


Figure O.1/H.263 – Illustration of B-picture prediction dependencies

The location of B-pictures in the bitstream is in a data-dependence order rather than in strict temporal order. (This rule is consistent with the ordering of other pictures in the bitstream, but for all picture types other than the B-picture, no such conflict arises between the data-dependence order and the temporal order.) For example, if the pictures of a video sequence were numbered 1, 2, 3, ... , then the bitstream order of the encoded pictures would be I₁, P₃, B₂, P₅, B₄, ... , where the subscript refers to the original picture number (as illustrated in Figure O.1).

There is no limit to the number of B-pictures that may be inserted between pairs of reference pictures in the reference layer (other than what is necessary to prevent temporal ambiguity from overflows of the temporal reference field in the picture header). However, a maximum number of such pictures may be signalled by external means (for example, ITU-T Rec. H.245).

The picture height, width, and pixel aspect ratio of a B-picture shall always be equal to those of its temporally subsequent reference layer picture.

Motion vectors are allowed to extend beyond the picture boundaries of B-pictures.

O.1.2 SNR scalability

The other basic method to achieve scalability is through spatial/SNR enhancement. Spatial scalability and SNR scalability are equivalent except for the use of interpolation as is described shortly. Because compression introduces artifacts and distortions, the difference between a reconstructed picture and its original in the encoder is (nearly always) a nonzero-valued picture, containing what can be called the coding error. Normally, this coding error is lost at the encoder and never recovered. With SNR scalability, these coding error pictures can also be encoded and sent to the decoder, producing an enhancement to the decoded picture. The extra data serves to increase the signal-to-noise ratio of the video picture, and hence, the term SNR scalability. Figure O.2 illustrates the data flow for SNR scalability. The vertical arrows from the lower layer illustrate that the picture in the enhancement layer is predicted from a reconstructed approximation of that picture in the reference (lower) layer.

If prediction is only formed from the lower layer, then the enhancement layer picture is referred to as an EI-picture. It is possible, however, to create a modified bidirectionally predicted picture using both a prior enhancement layer picture and a temporally simultaneous lower layer reference picture. This type of picture is referred to as an EP-picture or "Enhancement" P-picture. The prediction flow for EI- and EP-pictures is shown in Figure O.2. (Although not specifically shown in Figure O.2, an EI-picture in an enhancement layer may have a P-picture as its lower layer reference picture, and an EP-picture may have an I-picture as its lower-layer enhancement picture.)

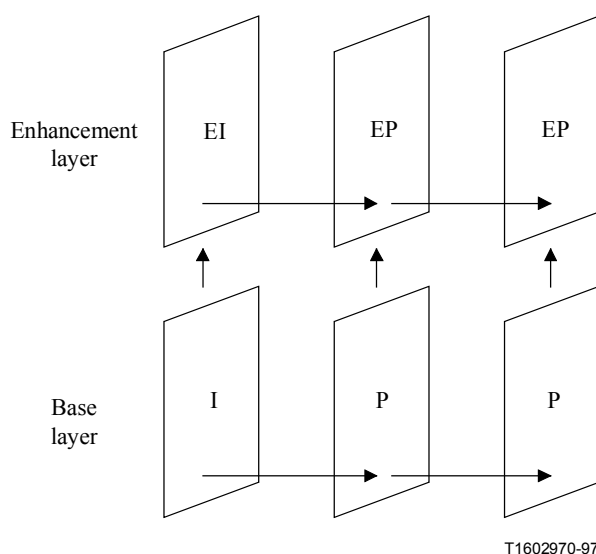


Figure O.2/H.263 – Illustration of SNR scalability

For both EI- and EP-pictures, the prediction from the reference layer uses no motion vectors. However, as with normal P-pictures, EP-pictures use motion vectors when predicting from their temporally prior reference picture in the same layer.

O.1.3 Spatial scalability

The third and final scalability method in the Temporal, SNR, and Spatial Scalability mode is spatial scalability, which is closely related to SNR scalability. The only difference is that before the picture in the reference layer is used to predict the picture in the spatial enhancement layer, it is interpolated by a factor of two either horizontally or vertically (1-D spatial scalability), or both horizontally and vertically (2-D spatial scalability). The interpolation filters for this operation are defined in O.6. For a decoder to be capable of some forms of spatial scalability, it may also need to be capable of custom picture formats. For example, if the base layer is sub-QCIF (128×96), the 2-D spatial enhancement layer picture would be 256×192 , which does not correspond to a standard picture format. Another example would be if the base layer were QCIF (176×144), with the standard pixel aspect ratio of 12:11. A 1-D horizontal spatial enhancement layer would then correspond to a picture format of 352×144 with a pixel aspect ratio of 6:11. Thus a custom picture format would have to be used for the enhancement layer in these cases. An example which does not require a custom picture format would be the use of a QCIF base layer with a CIF 2-D spatial enhancement layer. Spatial scalability is illustrated in Figure O.3.

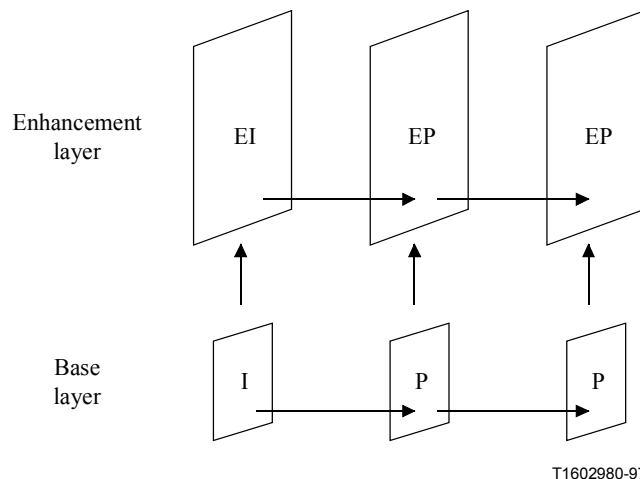


Figure O.3/H.263 – Illustration of spatial scalability

Other than requiring an upsampling process to increase the size of the reference layer picture prior to its use as a reference for the encoding process, the processing and syntax for a spatial scalability picture is functionally identical to that for an SNR scalability picture.

Since there is very little syntactical distinction between pictures using SNR scalability and pictures using spatial scalability, the pictures used for either purpose are called EI- and EP-pictures.

The picture in the base layer which is used for upward prediction in an EI- or EP-picture may be an I-picture, a P-picture, or the P-part of a PB- or Improved PB-frame (but shall not be a B-picture or the B-part of a PB- or Improved PB-frame).

O.1.4 Multilayer scalability

It is possible not only for B-pictures to be temporally inserted between pictures of types I, P, PB, and Improved PB, but also between pictures of types EI and EP, whether these consist of SNR or spatial enhancement pictures. It is also possible to have more than one SNR or spatial enhancement layer in conjunction with a base layer. Thus, a multilayer scalable bitstream can be a combination of SNR layers, spatial layers, and B-pictures. The size of a picture cannot decrease, however, with

increasing layer number. It can only stay the same or increase by a factor of two in one or both dimensions. Figure O.4 illustrates a multilayer scalable bitstream.

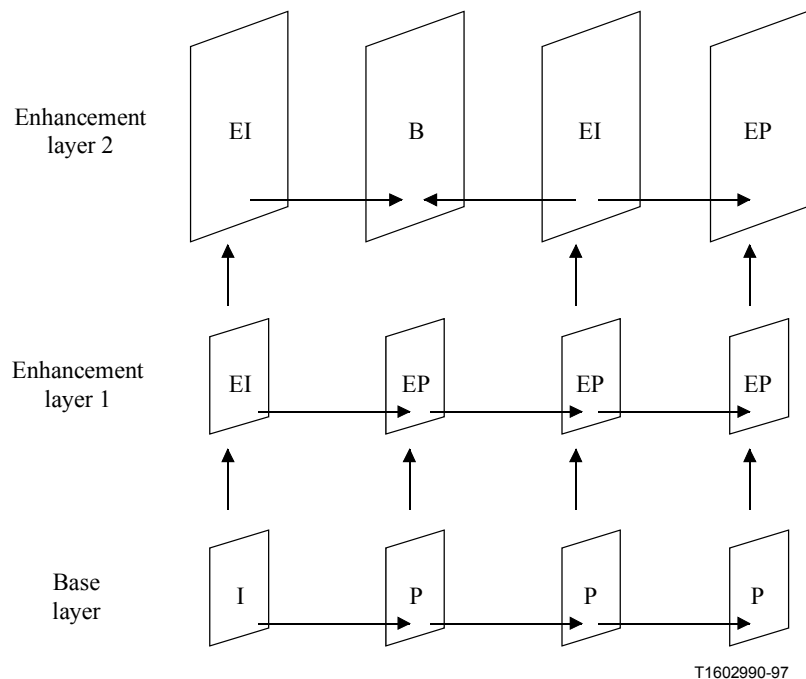


Figure O.4/H.263 – Illustration of multilayer scalability

In the case of multilayer scalability, the picture in a reference layer which is used for upward prediction in an EI- or EP-picture may be a I-, P-, EI-, or EP-picture, or may be the P-part of a PB- or Improved PB-frame in the base layer (but shall not be a B-picture or the B-part of a PB- or Improved PB-frame).

As with the two-layer case, B-pictures may occur in any layer. However, any picture in an enhancement layer which is temporally simultaneous with a B-picture in its reference layer must be a B-picture or the B-picture part of an PB- or Improved PB-frame. This is to preserve the disposable nature of B-pictures. Note, however, that B-pictures may occur in layers that have no corresponding picture in lower layers. This allows an encoder to send enhancement video with a higher picture rate than the lower layers.

The enhancement layer number and the reference layer number for each enhancement picture (B-, EI-, or EP-) are indicated in the ELNUM and RLNUM fields, respectively, of the picture header (when present). See the inference rules described in 5.1.4.4 for when these fields are not present. If a B-picture appears in an enhancement layer in which temporally surrounding SNR or spatial scalability pictures also appear, the Reference Layer Number (RLNUM) of the B-picture shall be the same as the Enhancement Layer Number (ELNUM).

The picture height, width, and pixel aspect ratio of a B-picture shall always be equal to those of its temporally subsequent reference layer picture.

O.2 Transmission order of pictures

Pictures which are dependent on other pictures shall be located in the bitstream after the pictures on which they depend.

The bitstream syntax order is specified such that for reference pictures (i.e., picture having types I, P, EI, or EP, or the P-part of PB or Improved PB), the following two rules shall be obeyed:

- 1) All reference pictures with the same temporal reference shall appear in the bitstream in increasing enhancement layer order (since each lower layer reference picture is needed to decode the next higher layer reference picture).
- 2) All temporally simultaneous reference pictures as discussed in item 1) above shall appear in the bitstream prior to any B-pictures for which any of these reference pictures is the first temporally subsequent reference picture in the reference layer of the B-picture (in order to reduce the delay of decoding all reference pictures which may be needed as references for B-pictures).

Then, the B-pictures with earlier temporal references shall follow (temporally ordered within each enhancement layer).

The bitstream location of each B-picture shall comply with the following rules:

- 1) Its bitstream location shall be after that of its first temporally subsequent reference picture in the reference layer (since the decoding of the B-picture generally depends on the prior decoding of that reference picture).
- 2) Its bitstream location shall be after that of all reference pictures that are temporally simultaneous with the first temporally subsequent reference picture in the reference layer (in order to reduce the delay of decoding all reference pictures which may be needed as references for B-pictures).
- 3) Its bitstream location shall precede the location of any additional temporally subsequent pictures other than B-pictures in its reference layer (since to allow otherwise would increase picture-storage memory requirements for the reference layer pictures).
- 4) Its bitstream location shall be after that of all EI- and EP-pictures that are temporally simultaneous with the first temporally subsequent reference picture.
- 5) Its bitstream location shall precede the location of all temporally subsequent pictures within its same enhancement layer (since to allow otherwise would introduce needless delay and increase picture-storage memory requirements for the enhancement layer).

Figure O.5 illustrates two allowable picture transmission orders given by the rules above for the layering structure shown therein (with numbers in dotted-line boxes indicating the bitstream order, separated by commas for the two alternatives).

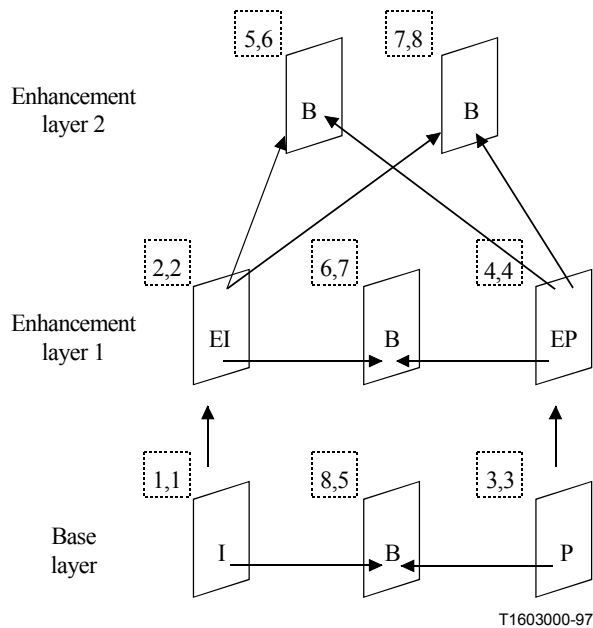


Figure O.5/H.263 – Example of picture transmission order

O.3 Picture layer syntax

The Enhancement Layer Number (ELNUM) (see 5.1.11) is always present in any enhancement (B, EI, or EP) picture and shall not be present in I- or P-pictures, or in PB- or Improved PB-frames. The Reference Layer Number (RLNUM) (see 5.1.12) is present for some enhancement pictures and is inferred for others, as described in 5.1.12.

There is exactly one base layer, and it has ELNUM and RLNUM equal to 1. The RLNUM gives the enhancement layer number of the forward and backward reference pictures for B-pictures, for the upward reference picture for EI- and EP-pictures. The reference pictures of the base layer may consist of I-, PB-, Improved PB-, and P-pictures, none of which include ELNUM or RLNUM in the picture header (their implied values are 1).

For B-pictures, RLNUM must be less or equal to ELNUM, while for EI- and EP-pictures, RLNUM must be smaller than ELNUM.

ELNUM may be different from the layer number used at the system level. Since B-pictures have no other pictures dependent on themselves, they may even be put in a separate enhancement layer by system components external to this Recommendation (for example, ITU-T Recs H.245 and H.223). Moreover, it is up to the implementor as to whether the enhancement pictures are sent in separate video channels or remain multiplexed together with the base layer pictures.

As stated in 5.1.4.5, the Deblocking Filter mode (see Annex J) does not apply within B-pictures. This is because B-pictures are not used for the prediction of any other pictures, and thus the application of a deblocking filter to these pictures is entirely a post-processing technique which is outside the scope of this Recommendation. However, the use of some type of deblocking filter for B-pictures is encouraged, and a filter such as that described in Annex J may in fact be a good design to use for that purpose.

The Temporal Reference (TR) (see 5.1.2) is defined exactly as for I- and P-pictures.

There shall be no TR_B (see 5.1.22) or DBQUANT (see 5.1.23) fields in the picture header of B-, EI- or EP-pictures.

O.4 Macroblock layer syntax

The macroblock layer syntax for B- and EP-pictures is the same, since each uses two reference pictures in a similar way. However, the interpretation varies slightly depending on the picture type. Figure O.6 indicates the B and EP syntax. The MBTYPE field indicates whether there is direct-mode prediction, forward prediction, backward/upward prediction, or bidirectional prediction. The MBTYPE is defined differently for B- and EP-pictures as described below.

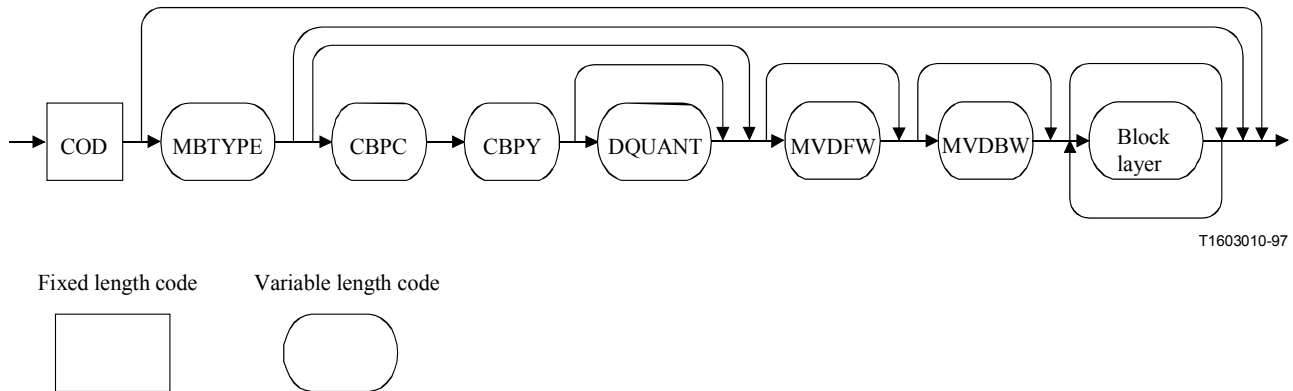


Figure O.6/H.263 – Macroblock syntax for EP- and B-pictures

The direct prediction mode is only available for B-pictures. It is a bidirectional prediction mode similar to the bidirectional mode in Improved PB-frames mode (Annex M). The only difference is that there is no restriction on which pixels can be backward predicted since the complete backward prediction image is known at the decoder. Bidirectional mode uses separate motion vectors for forward and backward prediction. In both direct and bidirectional mode, the prediction pixel values are calculated by averaging the forward and backward prediction pixels. The average is calculated by dividing the sum of the two predictions by two (division by truncation). In direct mode, when there are four motion vectors in the reference macroblock, then all four motion vectors are used as in Improved PB-frames mode (Annex M).

For B-pictures, forward prediction means prediction from a previous reference picture in the reference layer. Backward prediction means prediction from a temporally subsequent reference picture in the reference layer.

For EP-pictures, forward prediction means prediction from a previous EI- or EP-picture in the same layer, while upward prediction is used to mean prediction from the temporally simultaneous (possibly interpolated) reference picture in the reference layer. No motion vector is used for the upward prediction (which is syntactically in the same place as backward prediction for B-pictures), although one can be used for the forward prediction.

The macroblock syntax for EI-pictures is slightly different. As shown in Figure O.7, the MBTYPE and CBPC are combined into an MCBPC field. No forward prediction is used, only upward prediction from the temporally simultaneous reference picture in the reference layer. No motion vector is used.

In B and EP pictures, motion vectors over picture boundaries may be used as described in D.1 (although extension of the motion vector range as described in D.2 is only active if the Unrestricted Motion Vector mode is also in use).

Encoders shall ensure conformance to D.1.1 for all macroblocks, including those predicted with the direct prediction mode; that is, the direct prediction mode shall not be selected by the encoder unless the motion vector values inferred by the direct mode prediction process do not result in any

element of the 16×16 (or 8×8) prediction region having a horizontal or vertical distance more than 15 pixels outside the coded picture area.

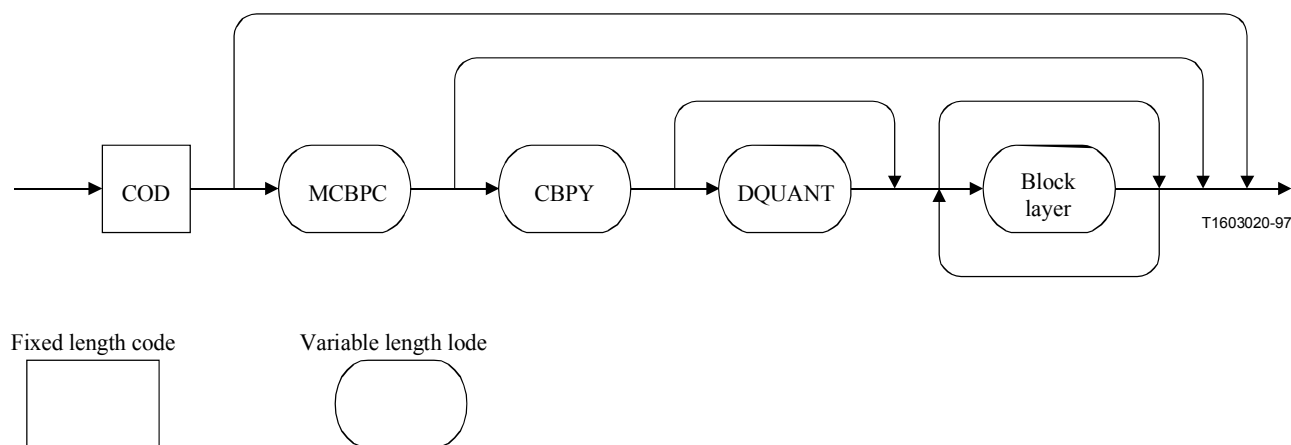


Figure O.7/H.263 – Macroblock syntax for EI-pictures

O.4.1 Coded macroblock indication (COD) (1 bit)

A bit which, when set to "0", signals that the macroblock is coded. If set to "1", no further information is transmitted for this macroblock and the macroblock is handled as "skipped" as described below.

O.4.2 MBTYPE/MCBPC (VLC)

There are different MBTYPE tables for B- and EP-pictures. For EI-pictures, there is instead a MCBPC table. Table O.1 is the MBTYPE table for B-pictures. Table O.2 is the MBTYPE table for EP-pictures. Table O.3 is the MCBPC table for EI-pictures.

For B-pictures, the "Direct (skipped)" prediction type indicates that neither MBTYPE, nor any data, is transmitted in the macroblock and that the decoder derives forward and backward motion vectors and the corresponding bidirectional prediction. This is signalled by the COD bit. The "Forward (no texture)", the "Backward (no texture)", and the "Bi-dir (no texture)" prediction types for a B-picture indicate forward, backward, and bi-directional prediction with no coefficients, and with one transmitted motion vector for forward and backward prediction, and two transmitted motion vectors for bidirectional prediction.

For EP-pictures, the "Forward (skipped)" prediction type indicates that no additional data is sent for the macroblock, so that the decoder should use forward prediction with a zero motion vector and no coefficients. The "Upward (no texture)" and "Bi-dir (no texture)" prediction types for an EP-picture indicate upward and bidirectional prediction with no coefficients, and with zero motion vector(s).

For EI-pictures, the "Upward (skipped)" prediction type indicates that no additional data is sent for the macroblock, so that the decoder should use upward prediction with a zero motion vector and no coefficients.

Table O.1/H.263 – MBTYPE VLC codes for B-pictures

Index	Prediction type	MVDFW	MVDBW	CBPC + CBPY	DQUANT	MBTYPE	Bits
–	Direct (skipped)					(COD = 1)	0
0	Direct			X		11	2
1	Direct + Q			X	X	0001	4
2	Forward (no texture)	X				100	3
3	Forward	X		X		101	3
4	Forward + Q	X		X	X	0011 0	5
5	Backward (no texture)		X			010	3
6	Backward		X	X		011	3
7	Backward + Q		X	X	X	0011 1	5
8	Bi-dir (no texture)	X	X			0010 0	5
9	Bi-dir	X	X	X		0010 1	5
10	Bi-dir + Q	X	X	X	X	0000 1	5
11	INTRA			X		0000 01	6
12	INTRA + Q			X	X	0000 001	7
13	Stuffing					0000 0000 1	9

Table O.2/H.263 – MBTYPE VLC codes for EP-pictures

Index	Prediction type	MVDFW	MVDBW	CBPC + CBPY	DQUANT	MBTYPE	Bits
–	Forward (skipped)					(COD = 1)	0
0	Forward	X		X		1	1
1	Forward + Q	X		X	X	001	3
2	Upward (no texture)					010	3
3	Upward			X		011	3
4	Upward + Q			X	X	0000 1	5
5	Bi-dir (no texture)					0001 0	5
6	Bi-dir	X		X		0001 1	5
7	Bi-dir + Q	X		X	X	0000 01	6
8	INTRA			X		0000 001	7
9	INTRA + Q			X	X	0000 0001	8
10	Stuffing					0000 0000 1	9

Table O.3/H.263 – MCBPC VLC codes for EI-pictures

Index	Prediction type	Code block pattern (56)	CBPY	DQUANT	MCBPC	Bits
–	Upward (skipped)				(COD = 1)	0
0	Upward	00	X		1	1
1	Upward	01	X		001	3
2	Upward	10	X		010	3
3	Upward	11	X		011	3
4	Upward + Q	00	X	X	0001	4
5	Upward + Q	01	X	X	0000 001	7
6	Upward + Q	10	X	X	0000 010	7
7	Upward + Q	11	X	X	0000 011	7
8	INTRA	00	X		0000 0001	8
9	INTRA	01	X		0000 1001	8
10	INTRA	10	X		0000 1010	8
11	INTRA	11	X		0000 1011	8
12	INTRA+Q	00	X	X	0000 1100	8
13	INTRA+Q	01	X	X	0000 1101	8
14	INTRA+Q	10	X	X	0000 1110	8
15	INTRA+Q	11	X	X	0000 1111	8
16	Stuffing				0000 0000 1	9

O.4.3 Coded Block Pattern for Chrominance (CBPC) (Variable length)

When present, CBPC indicates the coded block pattern for chrominance blocks as described in Table O.4. CBPC is present only for EP- and B-pictures, when its presence is indicated by MBTYPE (see Tables O.1 and O.2).

Table O.4/H.263 – CBPC VLC codes

Index	Code block pattern (56)	CBPC	Bits
0	00	0	1
1	01	10	2
2	10	111	3
3	11	110	3

O.4.4 Coded Block Pattern for Luminance (CBPY) (Variable length)

When present, CBPY indicates which blocks in the luminance portion of the macroblock are present. CBPY is present only when its presence is indicated by MBTYPE (see Tables O.1, O.2 and O.3). CBPY is coded as described in 5.3.5 and in Table 12. Upward predicted macroblocks in EI- and EP-pictures, bidirectional predicted macroblocks in EP-pictures, and INTRA macroblocks in EI-, EP-, and B-pictures use the CBPY definition for INTRA macroblocks, and other macroblock types in EI-, EP-, and B-pictures use the CBPY definition for INTER macroblocks.

O.4.5 Quantizer Information (DQUANT) (2 bits/Variable length)

DQUANT is used as in other picture macroblock types. See 5.3.6 and Annex T.

O.4.6 Motion vector data (MVDFW, MVDBW) (Variable length)

MVDFW is the motion vector data for the forward vector, if present. MVDBW is the motion vector data for the backward vector, if present (allowed only in B-pictures). The variable length codewords are given in Table 14, or in Table D.3 if the Unrestricted Motion Vector mode is used (see Annex D).

O.5 Motion vector decoding

O.5.1 Differential motion vectors

Motion vectors for forward, backward or bidirectionally predicted blocks are differentially encoded. To recover the macroblock motion vectors, a prediction is added to the motion vector differences. The predictions are formed in a similar manner to that described in 6.1.1, except that forward motion vectors are predicted only from forward motion vectors in surrounding macroblocks, and backward motion vectors are predicted only from backward motion vectors in surrounding macroblocks. The same decision rules apply for the special cases at picture, GOB or slice borders as described in 6.1.1. If a neighboring macroblock does not have a motion vector of the same type (forward or backward), the candidate predictor for that macroblock is zero for that motion vector type.

O.5.2 Motion vectors in direct mode

For macroblocks coded in direct mode, no vector differences are transmitted. Instead, the forward and backward motion vectors are directly computed from the temporally consecutive P-vector as described in G.4 with the restriction that MV_D is always zero. These derived vectors are not used for prediction of other motion vectors. If the corresponding area of the temporally subsequent reference picture is coded in INTRA mode, the forward and backward motion vectors assigned to that area for use in the prediction process for the direct mode shall have the value zero.

O.6 Interpolation filters

The method by which a picture is interpolated for 2-D spatial scalability is shown in Figures O.8 and O.9. The first figure shows the interpolation for interior pixels, while the second one shows the interpolation close to picture boundaries. This is the same technique used in Annex Q and in some cases in Annex P. The method by which a picture is interpolated for 1-D spatial scalability is shown in Figures O.10 and O.11. Figure O.10 shows the interpolation for interior pixels in the horizontal direction. The interpolation for the vertical direction is analogous. Figure O.11 shows the interpolation for pixels at picture boundaries. Again, the interpolation for the vertical direction is analogous. Again, this is the same technique used in some cases in Annex P.

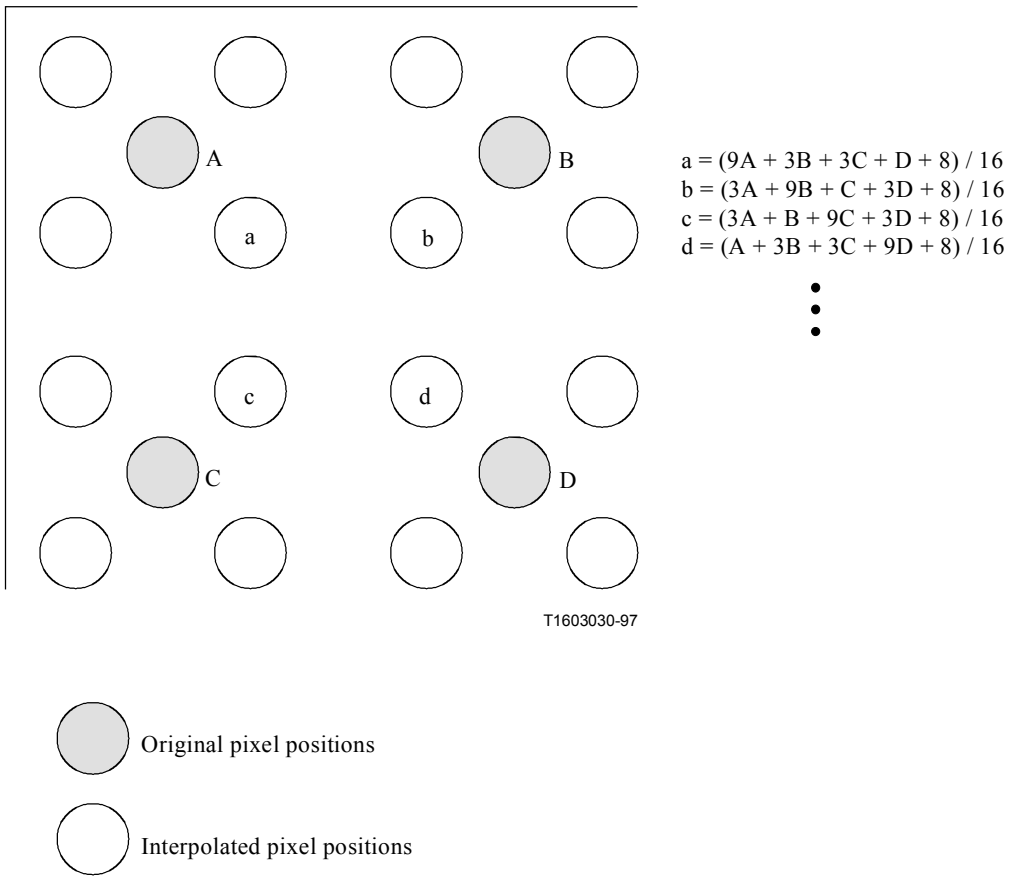


Figure O.8/H.263 – Method for interpolating pixels for 2-D scalability

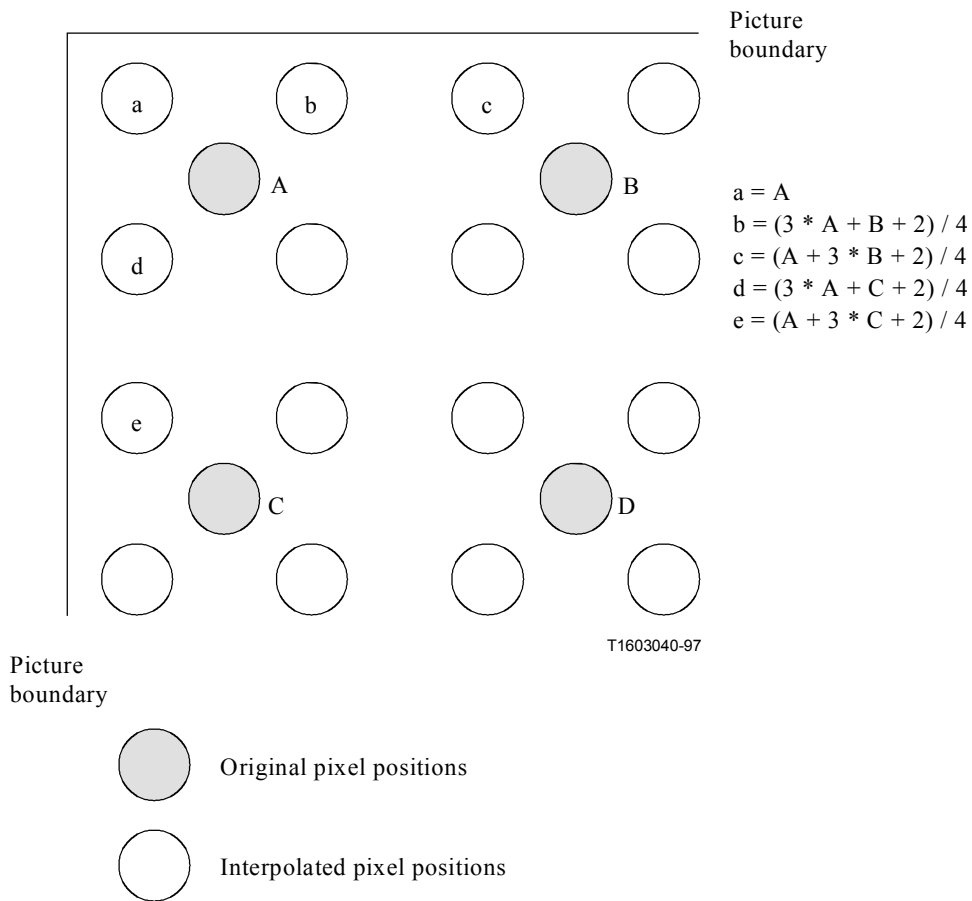


Figure O.9/H.263 – Method for 2-D interpolation at boundaries

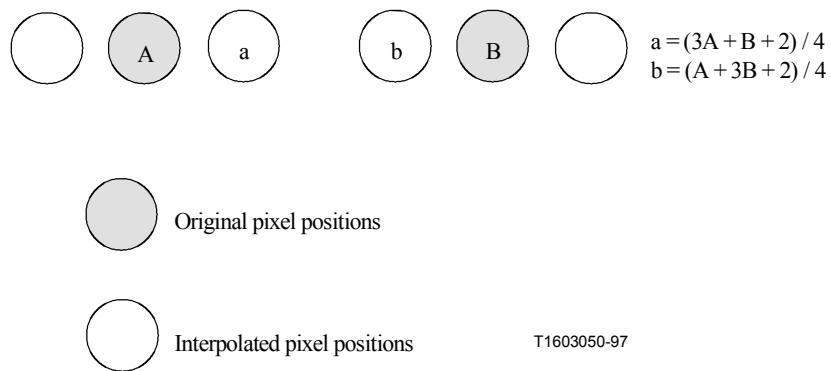


Figure O.10/H.263 – Method for interpolating pixels for 1-D scalability

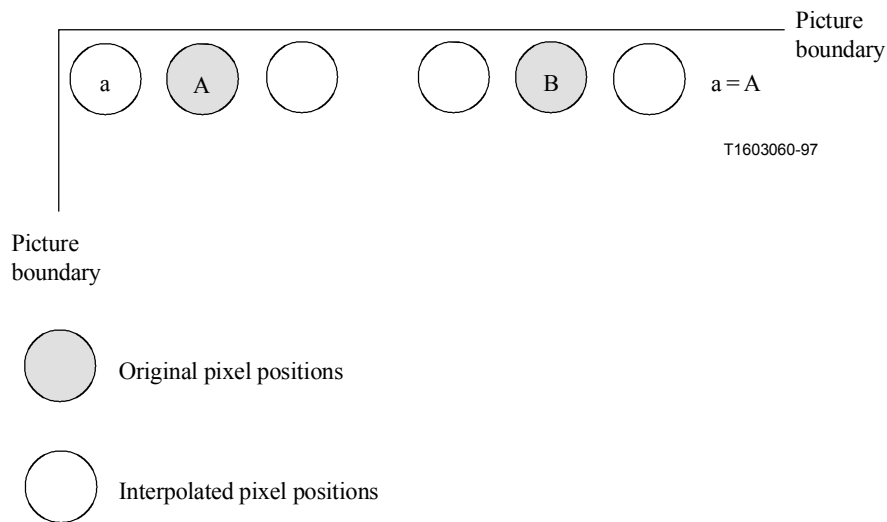


Figure O.11/H.263 – Method for 1-D interpolation at picture boundaries

Annex P

Reference picture resampling

P.1 Introduction

This annex describes the use and syntax of a resampling process which can be applied to the previous decoded reference picture in order to generate a "warped" picture for use in predicting the current picture. This resampling syntax can specify the relationship of the current picture to a prior picture having a different source format, and can also specify a "global motion" warping alteration of the shape, size, and location of the prior picture with respect to the current picture. In particular, the Reference Picture Resampling mode can be used to adaptively alter the resolution of pictures during encoding. A fast algorithm is used to generate bilinear interpolation coefficients. The capability to use this mode and the extent to which its features are supported is negotiated externally (for example, ITU-T Rec. H.245). This mode may be used in restricted scenarios that may be defined during capability negotiation (e.g., to support only factor-of-four picture resizing, to support only half-pixel resolution picture warping, or to support arbitrary image size resizing and displacements).

NOTE – The default image transformation between pictures of differing resolutions is defined to maintain the spatial alignment of the edges of the picture area and of the relative locations of the luminance and chrominance samples. This may have implications on the design of any resampling operations used for generating pictures of various resolutions at the encoder and for displaying pictures of various resolutions after decoding (particularly regarding shifts in spatial location caused by phase shifts induced in resampling). Also, since this mode can be used for dynamic adaptive picture resolution changes, operation with this mode may benefit from external negotiation to display the decoded picture at a higher resolution than its encoded picture size, in order to allow switching between encoded picture sizes without resizing the picture display.

If the Reference Picture Resampling bit in the PLUSPTYPE field is not set, but PLUSPTYPE is present and the picture is an INTER-, B-, or EP-picture or an Improved PB-frame, and the picture size differs from that of the temporally previous encoded picture, this condition invokes Reference Picture Resampling with warping parameters (see P.2.2) set equal to zero, fill mode (see P.2.3) set to *clip*, and displacement accuracy (see P.2.1) set to $\frac{1}{16}$ -pixel accuracy. This causes the resampling process to easily act as a predictively-encoded change in picture resolution. In simple factor-of-four

resolution change cases such as transitions between CIF and 4CIF, the resampling process reduces to the same simple filter used for Spatial Scalability (Annex O) or for Reduced-Resolution Update (Annex Q), except for the application of rounding control.

If the picture is an EP-picture and the Reference Picture Resampling bit is set in the PLUSPTYPE field of the picture header of the reference layer, this bit shall also be set in the picture header of the EP-picture in the enhancement layer.

If a B-picture uses the Reference Picture Resampling mode, the resampling process shall be applied to the temporally previous anchor picture and not to the temporally subsequent one. The temporally previous anchor picture to which the resampling process is applied shall be the decoded picture (i.e., prior to any resampling applied by Reference Picture Resampling if this mode is also invoked for the subsequent reference picture). The temporally subsequent anchor picture shall have the same picture size as the B-picture.

If the Reference Picture Resampling mode is invoked for an Improved PB-frame, one set of warping parameters is sent and the resampled reference picture is used as the reference for both the B- and P-parts of the Improved PB-frame.

The Reference Picture Resampling mode shall not be invoked when the Reference Picture Selection mode (see Annex N) is in use unless the values of TRPI and TRP in all picture, GOB, and slice headers of the current picture specify the use of the same reference picture – in which case the indicated reference picture is the picture that determines whether the reference picture resampling process is to be invoked implicitly and is the picture to which the resampling process shall be applied.

The Reference Picture Resampling is defined in terms of the displacement of the four corners of the current picture area. For the luminance field of the current picture with horizontal size H and vertical size V , four conceptual motion vectors \underline{v}^{00} , \underline{v}^{H0} , \underline{v}^{0V} , and \underline{v}^{HV} , are defined for the upper left, upper right, lower left, and lower right corners of the picture, respectively. These vectors describe how to move the corners of the current picture to map them onto the corresponding corners of the previous decoded picture, as shown in Figure P.1. The units of these vectors are the same as those in the reference picture grid. To generate a vector $\underline{v}(x, y)$ at some real-valued location (x, y) in the interior of the current picture, an approximation to bilinear interpolation is used, i.e., as in:

$$\underline{v}(x, y) = \left(1 - \frac{y}{V}\right) \left[\left(1 - \frac{x}{H}\right) \underline{v}^{00} + \left(\frac{x}{H}\right) \underline{v}^{H0} \right] + \left(\frac{y}{V}\right) \left[\left(1 - \frac{x}{H}\right) \underline{v}^{0V} + \left(\frac{x}{H}\right) \underline{v}^{HV} \right]$$

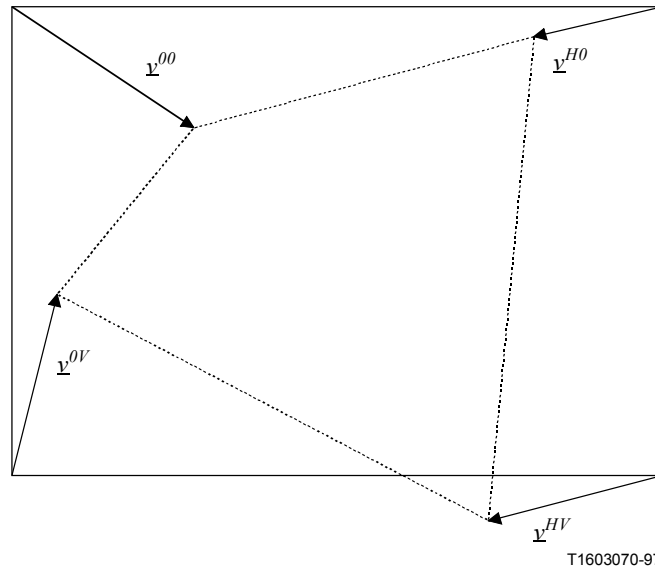


Figure P.1/H.263 – Example of conceptual motion vectors used for warping

The horizontal size H and vertical size V of the current picture, and the horizontal size H_R and vertical size V_R of the reference picture are those indicated by the picture header, regardless of whether these values are divisible by 16 or not. If the picture width or height is not divisible by 16, the additional area shall be generated by adding pixels to the resampled picture using the same fill mode as used in the resampling process.

For the simplicity of description, the resampling vectors \underline{r}^0 , \underline{r}^x , \underline{r}^y and \underline{r}^{xy} are defined as:

$$\begin{aligned}\underline{r}^0 &= \underline{v}^{00} \\ \underline{r}^x &= \underline{v}^{H0} - \underline{v}^{00} \\ \underline{r}^y &= \underline{v}^{0V} - \underline{v}^{00} \\ \underline{r}^{xy} &= \underline{v}^{00} - \underline{v}^{H0} - \underline{v}^{0V} + \underline{v}^{HV}\end{aligned}$$

Using this definition, the equation for bilinear interpolation is rewritten as:

$$\underline{v}(x, y) = \underline{r}^0 + \left(\frac{x}{H}\right)\underline{r}^x + \left(\frac{y}{V}\right)\underline{r}^y + \left(\frac{x}{H}\right)\left(\frac{y}{V}\right)\underline{r}^{xy}$$

For warping purposes, it is assumed that the coordinates of the upper left corner of the picture area are $(x, y) = (0, 0)$ and that each pixel has unit height and width, so that the centres of the pixels lie at the points $(x, y) = \left(i_L + \frac{1}{2}, j_L + \frac{1}{2}\right)$ for $i_L = 0, \dots, H - 1$ and $j_L = 0, \dots, V - 1$, where the L subscript indicates that i_L and j_L pertain to the luminance field. (As the pixel aspect ratio is usually constant or the conversion of the aspect ratio is performed by this resampling, there is no need to consider the actual pixel aspect ratio for these purposes.) Using this convention, the x - and y -displacements at the locations of interest in the luminance field of the reference picture are:

$$v_x(i_L, j_L) = \frac{1}{HV} \left[HVr_x^0 + \left(i_L + \frac{1}{2}\right)Vr_x^x + \left(j_L + \frac{1}{2}\right)Hr_x^y + \left(i_L + \frac{1}{2}\right)\left(j_L + \frac{1}{2}\right)r_x^{xy} \right]$$

$$v_y(i_L, j_L) = \frac{1}{HV} \left[HVr_y^0 + \left(i_L + \frac{1}{2}\right) Vr_y^x + \left(j_L + \frac{1}{2}\right) Hr_y^y + \left(i_L + \frac{1}{2}\right) \left(j_L + \frac{1}{2}\right) r_y^{xy} \right]$$

Because all positions and phases must be computed relative to the centre of the upper-left corner pixel, which has the coordinate $(x, y) = \left(\frac{1}{2}, \frac{1}{2}\right)$, the quantities of primary interest are:

$$\begin{aligned} x_R(i_L, j_L) - \frac{1}{2} &= \left(i_L + \frac{1}{2}\right) + v_x(i_L, j_L) - \frac{1}{2} \\ &= \frac{1}{HV} \left[HVr_x^0 + \left(i_L + \frac{1}{2}\right) (HV + Vr_x^x) + \left(j_L + \frac{1}{2}\right) Hr_x^y + \left(i_L + \frac{1}{2}\right) \left(j_L + \frac{1}{2}\right) r_x^{xy} \right] - \frac{1}{2} \\ y_R(i_L, j_L) - \frac{1}{2} &= \left(i_L + \frac{1}{2}\right) + v_y(i_L, j_L) - \frac{1}{2} \\ &= \frac{1}{HV} \left[HVr_y^0 + \left(j_L + \frac{1}{2}\right) Vr_y^x + \left(j_L + \frac{1}{2}\right) (HV + Hr_y^y) + \left(i_L + \frac{1}{2}\right) \left(j_L + \frac{1}{2}\right) r_y^{xy} \right] - \frac{1}{2} \end{aligned}$$

Once a location has been determined in the prior decoded reference picture using an approximation of these equations, bilinear interpolation as specified later in this annex shall be used to generate a value for the resampled pixel.

Each resampling vector can be decomposed into two components, with the first component describing the geometrical warping and the second component accounting for any difference in size between the predicted picture (horizontal size H and vertical size V) and the reference picture (horizontal size H_R and vertical size V_R). This decomposition is as follows:

$$\begin{aligned} \underline{v}^{00} &= \underline{v}_{warp}^{00} + \underline{v}_{size}^{00} = \underline{v}_{warp}^{00} + (0,0) \\ \underline{v}^{H0} &= \underline{v}_{warp}^{H0} + \underline{v}_{size}^{H0} = \underline{v}_{warp}^{H0} + (H_R - H, 0) \\ \underline{v}^{0V} &= \underline{v}_{warp}^{0V} + \underline{v}_{size}^{0V} = \underline{v}_{warp}^{0V} + (0, V_R - V) \\ \underline{v}^{HV} &= \underline{v}_{warp}^{HV} + \underline{v}_{size}^{HV} = \underline{v}_{warp}^{HV} + (H_R - H, V_R - V) \end{aligned}$$

P.2 Syntax

Whenever the reference picture resampling bit is set in the PLUSTPYPE field of the picture header, the RPRP field of the picture header includes parameters which control the Reference Picture Resampling process. This includes a two-bit Warping Displacement Accuracy (WDA) field, may include eight warping parameters or one-bit warping parameter refinements, and includes a fill mode, as described in this subclause.

P.2.1 Warping Displacement Accuracy (WDA) (2 bits)

A two-bit warping displacement accuracy field WDA appears first in the RPRP field of the bitstream, and indicates the accuracy of the displacements for each pixel. A value of "10" indicates that the x - and y -displacements for each pixel are quantized to half-pixel accuracy. A value of "11" indicates that the displacements are quantized to $\frac{1}{16}$ -pixel accuracy. The use of other values is reserved.

P.2.2 Warping parameters (Variable length)

When the Reference Picture Resampling parameters are sent for an INTER- or B-picture, or an Improved PB-frame, eight warping parameters are included in the picture header using the variable length code (VLC) shown in Table D.3. For an EP-picture using SNR scalability, the warping parameters of the lower layer are used and no warping parameters are transmitted. If the Reference Picture Resampling bit is set in the PLUSPTYPE field of the picture header for an EP-picture using spatial scalability, the warping parameters of the lower layer are refined to the accuracy needed for the current layer by multiplying the warping parameter for each up-sampled dimension (warping parameters with subscript x and/or y) of the lower layer by two and adding the value of one additional bit that is sent in place of the associated warping parameter to define the least significant bit of the warping parameter.

The eight integer warping parameters (or their one-bit refinements) are sent in the following order:

$$w_x^0, w_y^0, w_x^x, w_y^x, w_x^y, w_y^y, w_x^{xy}, \text{ and } w_y^{xy}$$

When not a one-bit refinement, these parameters are sent in a similar manner as motion vector differences in the Unrestricted Motion Vector mode when PLUSPTYPE is present, using Table D.3 with no restriction on the range of the parameters (i.e., a range of -4095 to $+4095$). As when encoding motion vector difference pairs, an emulation prevention bit shall be added as needed after each pair of warping parameters is sent, such that if the all-zero codeword (the value $+1$ in half-pixel units) of Table D.3 is used for both warping parameters in the pair (w_x^0 and w_y^0 , w_x^x and w_y^y , or w_x^{xy} and w_y^{xy}) the pair of codewords is followed by a single bit equal to 1 to prevent start code emulation.

These eight warping parameters are interpreted as picture corner displacements relative to the displacements that would be induced by removing the resizing component of the resampling vectors. The warping parameters are scaled to represent half-pixel offsets in the luminance field of the current picture, and the range for the values of these parameters is -4095 to $+4095$. The warping parameters are defined using the resampling vectors by the relations:

$$\begin{aligned} w_x^0 &= 2r_x^0 & w_y^0 &= 2r_y^0 \\ w_x^x &= 2(r_x^x - (H_R - H)) & w_y^x &= 2r_y^x \\ w_x^y &= 2r_x^y & w_y^y &= 2(r_y^y - (V_R - V)) \\ w_x^{xy} &= 2r_x^{xy} & w_y^{xy} &= 2r_y^{xy} \end{aligned}$$

P.2.3 Fill Mode (FILL_MODE) (2 bits)

For an INTER- or B-picture or an Improved PB-frame, immediately following the VLC-coded warping parameters in the picture header are two bits which define the fill-mode action to be taken for the values of pixels for which the calculated location in the reference picture lies outside of the reference picture area. The meaning of these two bits is shown in Table P.1 and their location is shown in Figure P.2. For an EP-picture, the fill-mode action is the same as that for the reference layer, and the two fill-mode bits are not sent.

Table P.1/H.263 – Fill-mode bits/action

Fill-mode bits	Fill action
0 0	<i>color</i>
0 1	<i>black</i>
1 0	<i>gray</i>
1 1	<i>clip</i>

If the fill mode is *clip*, the coordinates of locations in the prior reference picture are independently limited as in the Unrestricted Motion Vector mode so that pixel values outside of the prior reference picture area are estimated by extrapolation from the values of pixels at the image border. If the fill mode is *black*, luminance samples outside of the prior reference picture area are assigned a value of $Y = 16$ and chrominance samples are assigned a value of $C_B = C_R = 128$. If the fill mode is *gray*, luminance and chrominance values are assigned a value of $Y = C_B = C_R = 128$. If the fill mode is *color*, then additional fields are sent to specify a fill color, as described in the next subclause.

P.2.4 Fill Color Specification (Y_FILL, C_B_EPB, C_B_FILL, C_R_EPB, C_R_FILL) (26 bits)

If the fill mode is *color* and the picture is not an EP-picture, then the fill-mode bits are followed in the bitstream by three eight-bit integers, Y_fill , C_B_fill , and C_R_fill , which specify a fill color precisely. Between these three eight-bit integers are two emulation prevention bits (C_B_EPB and C_R_EPB) each of which is equal to 1. The format of this color specification, which is present only when the fill mode is *color*, is shown in Figure P.2. Each eight-bit integer field is sent using its natural representation. For an EP-picture, the fill-mode action (and fill color) is the same as that for the reference layer, and the fill color specification is not sent.

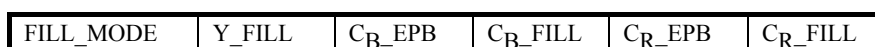


Figure P.2/H.263 – Format of fill mode and fill color specification data

P.3 Resampling algorithm

The method described in this clause shall be mathematically equal in result to that used to generate the samples of the resampled reference picture. Using the integer warping parameters $w_x^0, w_y^0, w_x^x, w_y^x, w_x^y, w_y^y, w_x^{xy}$, and w_y^{xy} integer parameters

$u_x^{00}, u_y^{00}, u_x^{H0}, u_y^{H0}, u_x^{0V}, u_y^{0V}, u_x^{HV}$, and u_y^{HV} which denote the x - and y -displacements at the

corners of the luminance field in $\frac{1}{32}$ -pixel accuracy (the actual displacements are obtained by dividing these values by 32) are defined as:

$$\begin{aligned}
 u_x^{00} &= 16w_x^0 & u_y^{00} &= 16w_y^0 \\
 u_x^{H0} &= 16(w_x^0 + w_x^x + 2(H_R - H)) & u_y^{H0} &= 16(w_y^0 + w_y^x) \\
 u_x^{0V} &= 16(w_x^0 + w_x^y) & u_y^{0V} &= 16(w_y^0 + w_y^y + 2(V_R - V)) \\
 u_x^{HV} &= 16(w_x^0 + w_x^x + w_x^y + w_x^{xy} + 2(H_R - H)) & u_y^{HV} &= 16(w_y^0 + w_y^x + w_y^y + w_y^{xy} + 2(V_R - V))
 \end{aligned}$$

Next, H' and V' , which denote the horizontal and vertical size of the *virtual frame*, are defined as the smallest integers that satisfy the following condition:

$$H' \geq H, V' \geq V, H' = 2^m, V' = 2^n, m \text{ and } n \text{ are positive integers}$$

By applying bilinear extrapolation to the corner vectors of the luminance field, the integer parameters $u_x^{LT}, u_y^{LT}, u_x^{RT}, u_y^{RT}, u_x^{LB}, u_y^{LB}, u_x^{RB}$, and u_y^{RB} which denote the x - and y -displacements of the luminance field at the *virtual points* $(x, y) = (0, 0), (H', 0), (0, V'),$ and (H', V') in $\frac{1}{32}$ -pixel accuracy (the actual displacements are obtained by dividing these values by 32) are defined as:

$$\begin{aligned}
 u_x^{LT} &= u_x^{00} & u_y^{LT} &= u_y^{00} \\
 u_x^{RT} &= \left((H - H')u_x^{00} + H'u_x^{H0} \right) // H & u_y^{RT} &= \left((H - H')u_y^{00} + H'u_y^{H0} \right) // H \\
 u_x^{LB} &= \left((V - V')u_x^{00} + V'u_x^{0V} \right) // V & u_y^{LB} &= \left((V - V')u_y^{00} + V'u_y^{0V} \right) // V \\
 u_x^{RB} &= \left((V - V') \left((H - H')u_x^{00} + H'u_x^{H0} \right) + V' \left((H - H')u_x^{0V} + H'u_x^{HV} \right) \right) // (HV) \\
 u_y^{RB} &= \left((V - V') \left((H - H')u_y^{00} + H'u_y^{H0} \right) + V' \left((H - H')u_y^{0V} + H'u_y^{HV} \right) \right) // (HV)
 \end{aligned}$$

where " //" denotes integer division that rounds the quotient to the nearest integer, and rounds half integer values away from 0.

In the remainder of this annex, it is assumed that the centres of the pixels lie at the points $(x, y) = \left(i + \frac{1}{2}, j + \frac{1}{2} \right)$ for both the luminance and chrominance fields. Integer parameters i, j are defined as:

- $i = 0, \dots, H - 1$ and $j = 0, \dots, V - 1$ for luminance; and
- $i = 0, \dots, \frac{H}{2} - 1$ and $j = 0, \dots, \frac{V}{2} - 1$ for chrominance.

This implies that different coordinate systems are used for luminance and chrominance as shown in Figure P.3. Using the coordinate system for chrominance, the integer parameters $u_x^{LT}, u_y^{LT}, u_x^{RT}, u_y^{RT}, u_x^{LB}, u_y^{LB}, u_x^{RB}$, and u_y^{RB} defined above can also be regarded as the x - and y -displacements of the chrominance field at virtual points $(x, y) = (0, 0), (H'/2, 0), (0, V'/2),$ and $(H'/2, V'/2)$ in $\frac{1}{64}$ -pixel accuracy (the actual displacements are obtained by dividing these values by 64). Using these parameters and an additional parameter S , which is defined as 2 for luminance and as 1 for chrominance, the resampling algorithm for the luminance and chrominance fields are defined using common equations.

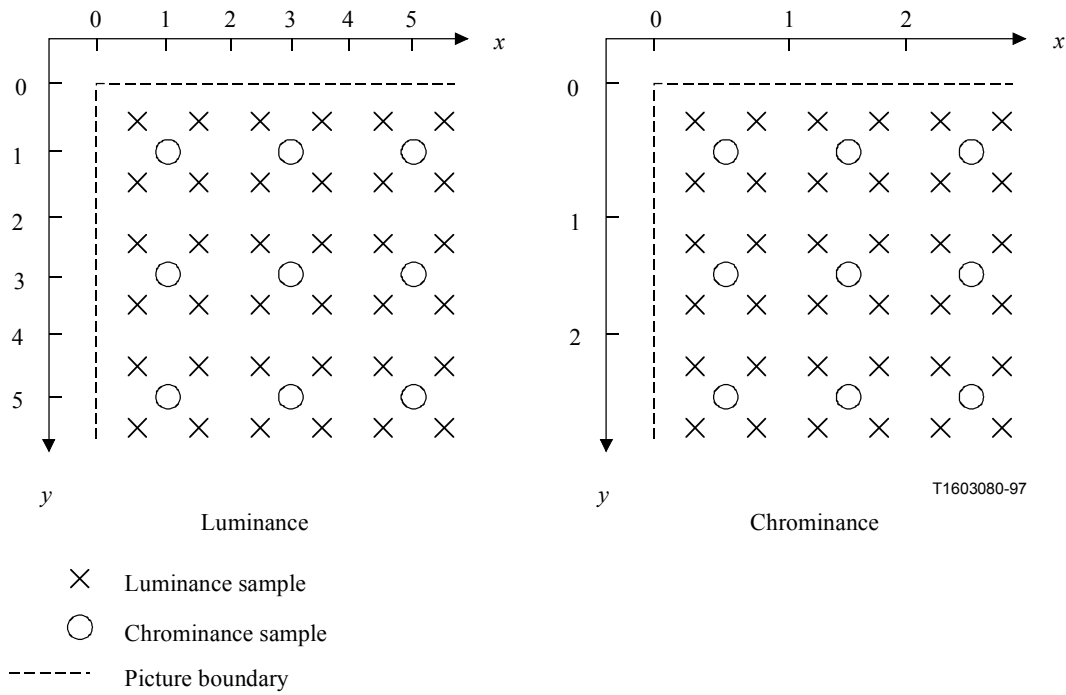


Figure P.3/H.263 – Coordinate systems for luminance and chrominance fields

The integer parameters $u_x^L(j)$, $u_y^L(j)$, $u_x^R(j)$, and $u_y^R(j)$ which denote the x - and y -displacements of the picture field at $(x, y) = (0, j + \frac{1}{2})$ and $(SH'/2, j + \frac{1}{2})$ in $\frac{S}{64}$ -pixel accuracy (the actual displacements are obtained by dividing these values by $64/S$) are defined using one dimensional linear interpolation as:

$$u_x^L(j) = ((SV' - 2j - 1)u_x^{LT} + (2j + 1)u_x^{LB}) // (SV') \quad u_y^L(j) = ((SV' - 2j - 1)u_y^{LT} + (2j + 1)u_y^{LB}) // (SV')$$

$$u_x^R(j) = ((SV' - 2j - 1)u_x^{RT} + (2j + 1)u_x^{RB}) // (SV') \quad u_y^R(j) = ((SV' - 2j - 1)u_y^{RT} + (2j + 1)u_y^{RB}) // (SV')$$

where "///" denotes integer division that rounds the quotient to the nearest integer, and rounds half integer values away from 0.

Finally, the parameters that specify the transformed position in the reference picture become:

$$I_R(i, j) = Pi + ((SH' - 2i - 1)u_x^L(j) + (2i + 1)u_x^R(j) + 32H'/P) /// (64H'/P)$$

$$J_R(i, j) = Pj + ((SH' - 2i - 1)u_y^L(j) + (2i + 1)u_y^R(j) + 32H'/P) /// (64H'/P)$$

$$i_R(i, j) = I_R(i, j) /// P \quad j_R(i, j) = J_R(i, j) /// P$$

$$\varnothing_x = I_R(i, j) - (I_R(i, j) /// P)P \quad \varnothing_y = J_R(i, j) - (J_R(i, j) /// P)P$$

where:

- "///" integer division with truncation towards the negative infinity;
- "/" integer division (in this case resulting in no loss of accuracy);

P accuracy of x - and y -displacements ($P = 2$ when WDA = "10" and $P = 16$ when WDA = "11" or is absent, see P.2.1 for the definition of WDA);

$\left(\frac{I_R(i,j)}{P} + \frac{1}{2}, \frac{J_R(i,j)}{P} + \frac{1}{2}\right)$ (x, y) location of the transformed position (both $I_R(i, j)$ and $J_R(i, j)$ are integers);

$\left(i_R(i,j) + \frac{1}{2}, j_R(i,j) + \frac{1}{2}\right)$ (x, y) location of the sampling point near the transformed position (both $i_R(i, j)$ and $j_R(i, j)$ are integers);

$(\emptyset_x, \emptyset_y)$ bilinear interpolation coefficients of the transformed position (both \emptyset_x and \emptyset_y are integers).

The computation of this equation can be simplified by replacing the divisions by shift operations, since $64H'/P = 2^{m+2}$ when $P = 16$ and $64H'/P = 2^{m+5}$ when $P = 2$.

Using these parameters, the sample value, $E_P(i, j)$, of the pixel located at $(x, y) = \left(i + \frac{1}{2}, j + \frac{1}{2}\right)$ in the resampled picture is obtained using bilinear interpolation by:

$$E_P(i, j) = \left((P - \emptyset_y) \left((P - \emptyset_x) E_R(i_R, j_R) + \emptyset_x E_R(i_R + 1, j_R) \right) + \emptyset_y \left((P - \emptyset_x) E_R(i_R, j_R + 1) + \emptyset_x E_R(i_R + 1, j_R + 1) \right) + P^2 / 2 - 1 + RCRPR \right) / P^2$$

where "/" denotes division by truncation. i_r and j_r are simplified notations for $i_R(i, j)$ and $j_R(i, j)$, and $E_R(i_R, j_R)$ denotes the sample value of the pixel located at $(x, y) = \left(i_R + \frac{1}{2}, j_R + \frac{1}{2}\right)$ in the reference picture after extrapolation using the proper fill mode if necessary. The value of parameter RCRPR is defined as follows:

- For a B-picture (or the B-part of an Improved PB-frame) which has a P-picture as its temporally subsequent anchor picture, RCRPR is equal to the rounding type (RTYPE) bit in MPPTYPE (see 5.1.4.3) of this temporally subsequent P-picture. This implies that for Improved PB-frames, RCRPR has the same value for the P-part and the B-part.
- For other types of pictures, RCRPR is equal to the RTYPE bit of the current picture.

P.4 Example of implementation

In this clause, an implementation example of the algorithm described in the previous clause is provided as pseudo-code.

P.4.1 Displacements of virtual points

When a large picture is coded, a straightforward implementation of the equation for obtaining parameters u_x^{RB} and u_y^{RB} shown in P.3 may necessitate the usage of variables requiring more than 32 bits for their binary representation. For systems which cannot easily use 64-bit integer or floating point registers, an example of an algorithm which does not require variables with more than 32 bits for the calculation of u_x^{RB} and u_y^{RB} is shown below.

Since H, V, H' , and V' are divisible by 4, the definition of u_x^{RB} can be rewritten as:

$$u_x^{RB} = \left((V_Q - V_Q') \left((H_Q - H_Q') u_x^{00} + H_Q' u_x^{H0} \right) + V_Q' \left((H_Q - H_Q') u_x^{0V} + H_Q' u_x^{HV} \right) \right) // A$$

where $H_Q = H/4$, $V_Q = V/4$, $H_Q' = H'/4$, $V_Q' = V'/4$, $A = H_Q V_Q$, and $"/"$ denotes integer division that rounds the quotient to the nearest integer, and rounds half integer values away from 0. Next, parameters T_T and T_B are defined as:

$$T_T = (H_Q - H_Q') u_x^{00} + H_Q' u_x^{H0}$$

$$T_B = (H_Q - H_Q') u_x^{0V} + H_Q' u_x^{HV}$$

for simplicity of description. Using operator $"/"$ which denotes integer division with truncation towards the negative infinity, and operator $"\%"$ which is defined as $a \% b = a - (a // b) b$, the value of u_x^{RB} can be obtained via the following pseudo-code:

```

q = (V_Q - V_Q') * (T_T // A) + V_Q' * (T_B // A) + ((V_Q - V_Q') * (T_T % A) + V_Q'
  * (T_B % A)) // A;
r = ((V_Q - V_Q') * (T_T % A) + V_Q' * (T_B % A)) % A;
if (q < 0)
  u_x^{RB} = q + (r + (A - 1) / 2) / A;
else
  u_x^{RB} = q + (r + A / 2) / A;

```

The value of u_y^{RB} can also be calculated using this algorithm.

P.4.2 Resampling algorithm

To simplify the description of the algorithm, a function `prior_sample` is defined. Its purpose is to generate a pixel value for any integer location (i_p, j_p) relative to the prior reference picture sampling grid:

```

clip(x_min, x, x_max) {
  if (x < x_min) {
    return x_min;
  } else if (x > x_max) {
    return x_max;
  } else {
    return x;
  }
}
prior_sample (i_p, j_p) {
  if (FILL_MODE = clip) {
    i_c = clip (0, i_p, S*H_R / 2 - 1);
    j_c = clip (0, j_p, S*V_R / 2 - 1);
    return prior_ref[i_c, j_c];
  } else {
    if ((i_p < 0) OR (i_p > S*H_R / 2 - 1) OR (j_p < 0) OR (j_p > S*V_R / 2 - 1) {
      return fill_value;
    } else {
      return prior_ref[i_p, j_p];
    }
  }
}

```

In the pseudo-code, `prior_ref[i, j]` indicates the sample in column i and row j in the temporally previous reference picture.

Next, a filter function that implements the bilinear interpolation described in P.3 is defined. It is assumed that all arguments of the following function are integers and that the bilinear interpolation coefficients, \emptyset_x and \emptyset_y , are quantized in the range 0, ..., $P-1$ (inclusive).

```
filter( $x_0, y_0, \emptyset_x, \emptyset_y$ ) {
    return  $[(P-\emptyset_y) \cdot ((P-\emptyset_x) \cdot \text{prior\_sample}(x_0, y_0) + \emptyset_x \cdot \text{prior\_sample}(x_0+1, y_0)) +$ 
         $\emptyset_y \cdot ((P-\emptyset_x) \cdot \text{prior\_sample}(x_0, y_0+1) + \emptyset_x \cdot \text{prior\_sample}(x_0+1, y_0+1))] +$ 
         $P^2 / 2 - 1 + RCRPRJ / P^2;$ 
}
```

Finally, the method for warping the reference picture to generate a prediction for the current picture can be specified in terms of these functions. The pixels of the prediction picture can be generated in a raster scan order. It is assumed that the values $u_x^L(j)$, $u_y^L(j)$, $u_x^R(j)$, and $u_y^R(j)$ are already calculated and loaded into variables u_x^L , u_y^L , u_x^R , and u_y^R . Defining parameter D as $D = 64H'/P$ and noting that $H' = 2^m$, the sample values of the pixels in the j th line of the resampled field (the topmost line is defined as the 0th line) is obtained by the following pseudo-code:

```
 $a_x^i = D * P + 2 * (u_x^R - u_x^L);$ 
 $a_y^i = 2 * (u_y^R - u_y^L);$ 
 $a_x = u_x^L * S * 2^m + (u_x^R - u_x^L); + D / 2;$ 
 $a_y = j * D * P + u_y^L * S * 2^m + (u_y^R - u_y^L); + D / 2;$ 
for ( $i = 0; i < S * H / 2; i++$ ) {
     $I_R = a_x /// D;$ 
     $J_R = a_y /// D;$ 
     $i_R = I_R /// P;$ 
     $j_R = J_R /// P;$ 
     $\emptyset_x = I_R - (i_R * P);$ 
     $\emptyset_y = J_R - (j_R * P);$ 
     $\text{new\_ref}[i, j] = \text{filter}(i_R, j_R, \emptyset_x, \emptyset_y);$ 
     $a_x += a_x^i;$ 
     $a_y += a_y^i;$ 
}
```

where all the variables used in this code are integer variables and $\text{new_ref}[i, j]$ indicates the sample generated for column i and row j in the resampled reference picture. According to the definition of the parameters, all the divisions in this code can be replaced by binary shift operations. For example, when $P = 16$:

```
 $I_R = a_x /// D;$ 
 $J_R = a_y /// D;$ 
 $i_R = I_R /// P;$ 
 $j_R = J_R /// P;$ 
 $\emptyset_x = I_R - (i_R * P);$ 
 $\emptyset_y = J_R - (j_R * P);$ 
```

can be rewritten, assuming that a_x , a_y , I_R , and J_R are binary-coded integer variables in two's complement representation, as:

$$\begin{aligned}I_R &= a_x \gg (m+2); \\J_R &= a_y \gg (m+2); \\i_R &= I_R \gg 4; \\j_R &= J_R \gg 4; \\\emptyset_x &= I_R \& 15; \\\emptyset_y &= J_R \& 15;\end{aligned}$$

where " $\gg N_{shift}$ " denotes a right arithmetic binary shift by N_{shift} bits (N_{shift} is a positive integer), and "&" denotes a bit-wise AND operation.

P.5 Factor-of-4 resampling

Factor-of-4 resampling, which converts both the horizontal and vertical size of the picture by a factor of 2 or $\frac{1}{2}$, is a special case of the resampling algorithm described in P.3. The simplified description of the resampling algorithm for this special case is provided in this clause.

The value of the parameter RCRPR used in Figures P.4 to P.6 is determined by the Rounding Type (RTYPE) bit in MPPTYPE (see 5.1.4.3) as described in P.3. Additionally, "/" in the figures indicates division by truncation.

P.5.1 Factor-of-4 upsampling

The pixel value interpolation method used in factor-of-4 upsampling for internal pixels are shown in Figure P.4. By assuming the existence of pixels outside the picture according to the selected fill mode (see P.2.3 and P.2.4), the same interpolation method is applied for boundary pixels. The interpolation method for boundary pixels when *clip* is selected as the fill mode is shown in Figure P.5. Since precise factor-of-4 upsampling requires x- and y-displacements with at least $\frac{1}{4}$ -pixel accuracy, the Warping Displacement Accuracy (WDA) field specified in P.2.1 must be set to "11" or the resampling must be implicitly invoked in order to use this upsampling method.

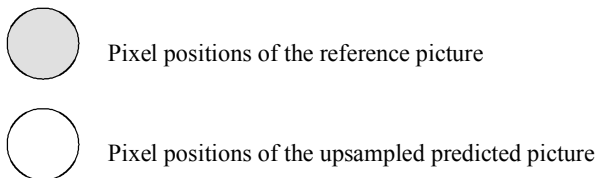
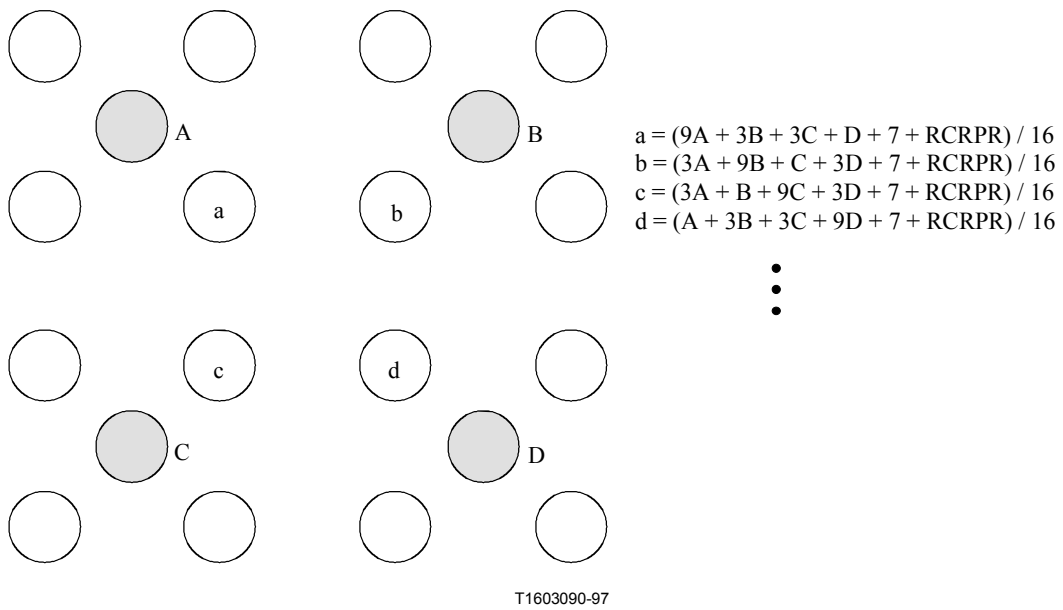


Figure P.4/H.263 – Factor-of-4 upsampling for pixels inside the picture

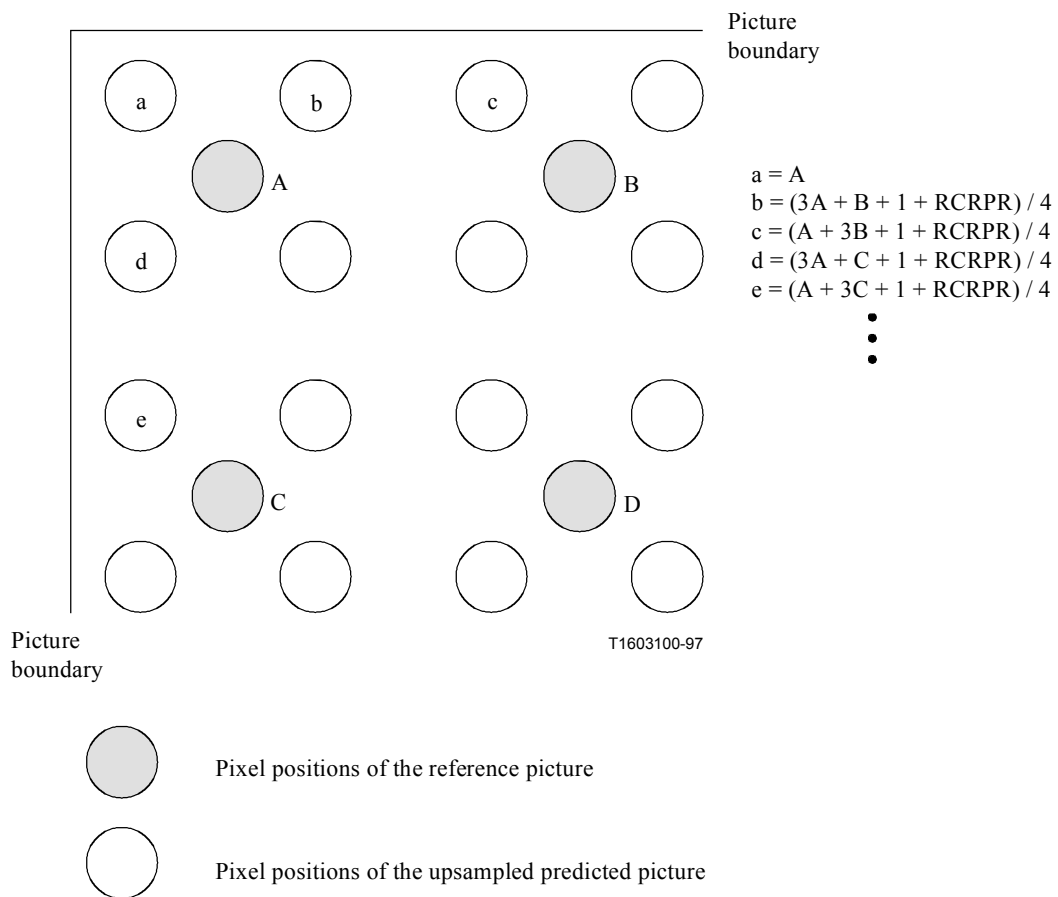


Figure P.5/H.263 – Factor-of-4 upsampling for pixels at the picture boundary (fill mode = *clip*)

P.5.2 Factor-of-4 downsampling

The pixel value interpolation method for factor-of-4 downsampling is shown in Figure P.6. Since x- and y-displacements with $\frac{1}{2}$ -pixel accuracy is sufficient for precise factor-of-4 downsampling, both "10" and "11" are allowed as the value of the displacement accuracy (WDA) field (if present) specified in P.2.1.

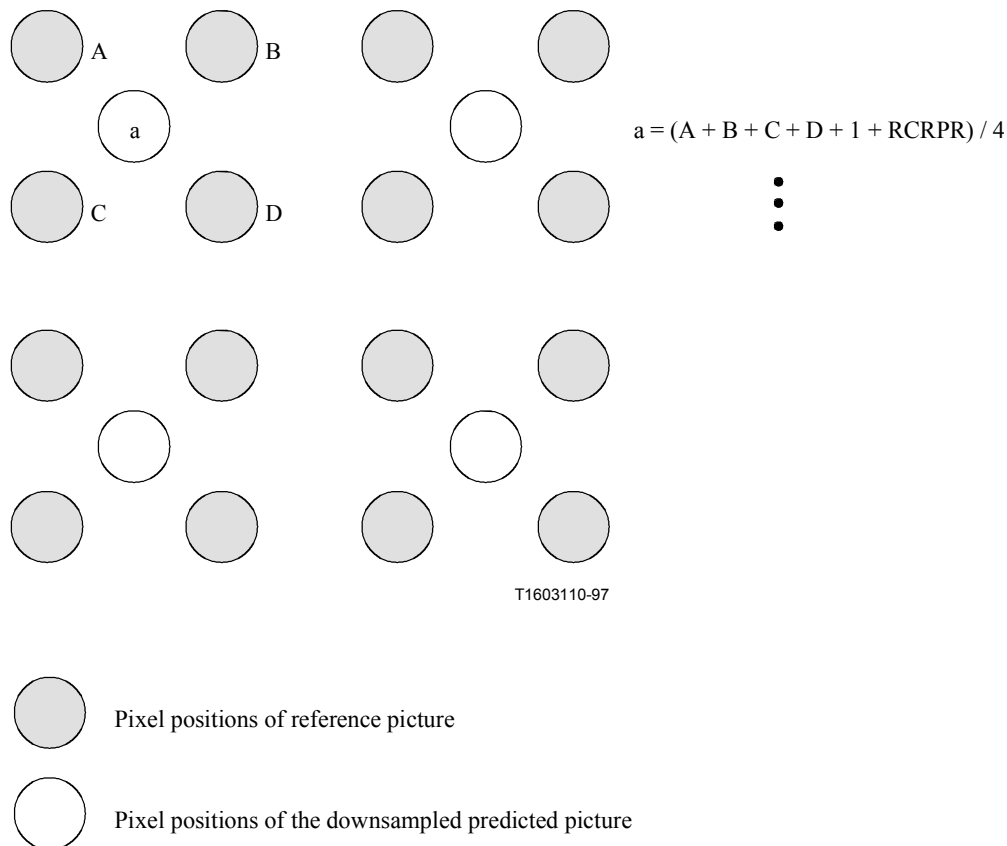


Figure P.6/H.263 – Factor-of-4 downsampling

Annex Q

Reduced-Resolution Update mode

Q.1 Introduction

This annex describes an optional Reduced-Resolution Update mode of this Recommendation. The capability of this mode is signalled by external means (for example, ITU-T Rec. H.245). The use of this mode is indicated in the PLUSPTYPE field of the picture header.

The Reduced-Resolution Update mode is expected to be used when encoding a highly active scene, and provides the opportunity to increase the coding picture rate while maintaining sufficient subjective quality. This mode allows the encoder to send update information for a picture that is encoded at a reduced resolution, while preserving the detail in a higher resolution reference image to create a final image at the higher resolution.

The syntax of the bitstream in this mode is identical to the syntax for coding without the mode, but the semantics, or interpretation of the bitstream is somewhat different. In this mode, the portion of the picture covered by a macroblock is twice as wide and twice as high. Thus, there is approximately one-quarter the number of macroblocks as there would be without this mode. Motion vector data also refers to blocks of twice the normal height and width, or 32×32 and 16×16 instead of the normal 16×16 and 8×8 . On the other hand, the DCT or texture data should be thought of as describing 8×8 blocks on a reduced-resolution version of the picture. To produce the final picture, the texture data is decoded at a reduced resolution and then up-sampled to the full resolution of the picture. After upsampling, the full resolution texture image is added to the (already full resolution) motion compensated image to create the image for display and further reference.

In this mode, a picture which has the horizontal size H and vertical size V as indicated in the picture header is created as a final image for display.

In this mode, a referenced picture used for prediction and created for further decoding has a horizontal size H_R and a vertical size V_R which are the same as in the default mode as defined in 4.1. That is, the H_R and V_R are:

$$H_R = ((H + 15) / 16) * 16$$

$$V_R = ((V + 15) / 16) * 16$$

where H and V are the horizontal size and the vertical size as indicated in the picture header and "/" is defined as division by truncation.

Then in this annex, the texture is coded at a reduced resolution with height and width of H_C and V_C where:

$$H_C = ((H_R + 31) / 32) * 32$$

$$V_C = ((V_R + 31) / 32) * 32$$

and "/" is defined as division by truncation.

If H_C and H_R , or V_C and V_R are not identical to each other, such as in the QCIF format, an extension of the referenced picture is performed, and the picture is decoded in the same manner as if the width and height are H_C and V_C . Then the resulting picture which is tiled by $32 * 32$ macroblocks is cropped at the right and the bottom to the width H_R and height V_R , and this cropped picture is stored as a reference picture for further decoding.

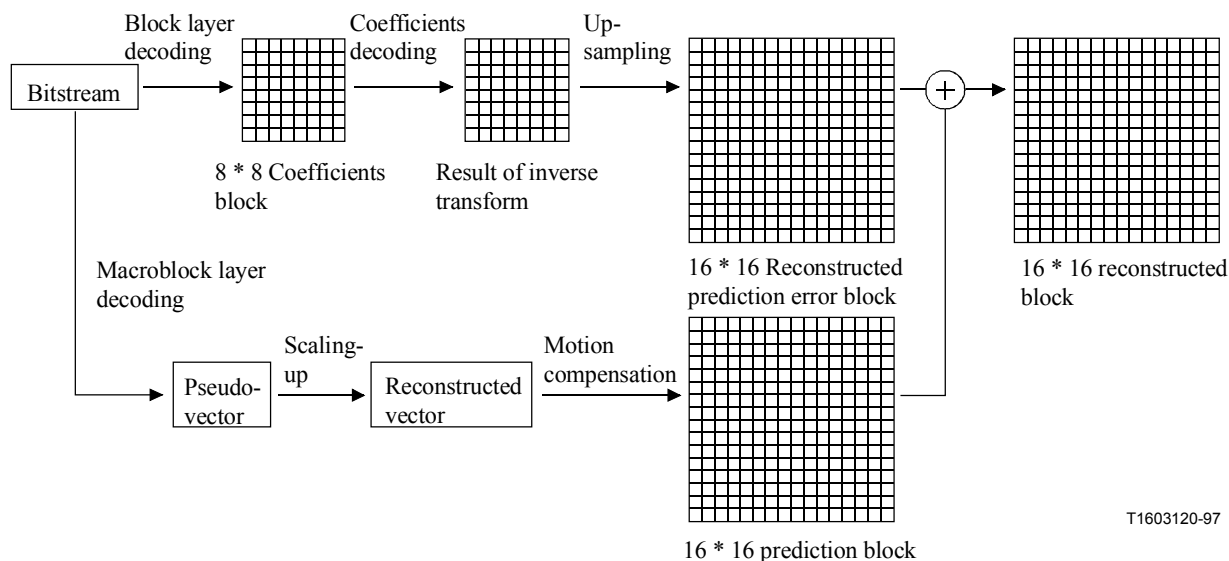
If both H and V are the same as those of a resulting picture having the width H_C and the height V_C , this resulting picture is used for display. Otherwise, this resulting picture is further cropped to the size $H * V$, and the cropped picture here is used for display purpose only.

If the Temporal, SNR, and Spatial Scalability mode (Annex O) or Reference Picture Resampling mode (Annex P) is also used with this option, the source format of the current picture might be different from that of the reference picture. In this case, the resampling of the reference picture shall be performed before decoding.

NOTE – This mode can be used with the Reference Picture Selection mode (see Annex N) without modification, since the reference picture (after possible resampling by the Reference Picture Resampling mode) has the same size as indicated in the current picture header when this option is used.

Q.2 Decoding procedure

Figure Q.1 shows the block diagram of the block decoding in the Reduced-Resolution Update mode.



T1603120-97

Figure Q.1/H.263 – Block diagram of block decoding in Reduced-Resolution Update mode

The decoding procedure is given in the following subclauses:

Q.2.1 Reference preparation

In some cases, the available reference picture has different size from H_C and V_C . Then, the reference picture shall be converted before decoding procedure according to Q.2.1.1 or Q.2.1.2.

Q.2.1.1 Reference Picture Resampling

If the Temporal, SNR, and Spatial Scalability mode (Annex O) or Reference Picture Resampling mode (Annex P) is also used with this option, the source format of the current picture might be different from that of the reference picture. In this case, the resampling of the reference picture according to each annex shall be performed first.

Q.2.1.2 Extension of Reference Picture

If H_R or V_R is not divisible by 32, such as with the QCIF format, the reference picture is extended. The detailed procedure for this extension is defined in Q.3.

Q.2.2 Macroblock layer decoding

Decoding can be thought of as operating on "enlarged" blocks of size 32×32 in luminance and 16×16 in chrominance. The texture and motion data for each enlarged block is decoded to create a 32×32 motion block and a 32×32 texture block, as described in Q.2.2.1 and Q.2.2.2 respectively. These motion and texture blocks are then added as described in Q.2.2.3.

Q.2.2.1 Motion compensation

First, each component of the macroblock motion vector (or four macroblock motion vectors) is formed from MVD (and possibly MVD_{2-4}). If in Improved PB-Frames mode, MV_F and MV_B for B-Picture is also formed from MVDB. The detailed procedure for this motion vector formation is defined in Q.4. If the current picture mode is a B-picture or EP-picture, the motion vector for forward and backward are also obtained according to Q.4.

The motion vector for the two chrominance blocks of the macroblock is obtained from the macroblock motion vector according to 6.1.1. If either the Advanced Prediction mode or the Deblocking Filter mode is in use and thus four motion vectors are defined for the macroblock, the motion vector for both chrominance blocks is obtained from the four motion vectors according to F.2. If in the Improved PB-frames mode, the creation of the chrominance vector is specified in Annex M. If a B-picture or EP-picture is used, the creation of the chrominance vector is specified in Annex O.

Then, a prediction is formed from the motion vector for an INTER macroblock. Four 16×16 luminance prediction blocks are obtained from the macroblock motion vector, and two 16×16 chrominance prediction blocks are obtained from the chrominance motion vector. For interpolation for subpixel prediction, refer to 6.1.2. If Advanced Prediction mode is also used, an enlarged overlapped motion compensation is performed to obtain four 16×16 luminance prediction blocks using enlarged weighting matrices, for which a detailed procedure is defined in Q.5. If the current picture mode is an Improved PB-frame, B-picture, or EP-picture, the prediction is obtained according to the other relevant annexes, except that the size of the predicted blocks is 16×16 instead of 8×8 .

Q.2.2.2 Texture decoding

First, the bitstream of the block layer is decoded according to 5.4. Then, coefficients are decoded and the 8×8 reduced-resolution reconstructed prediction error blocks are obtained as the result of inverse transform according to 6.2.

Then the 16×16 reconstructed prediction error blocks are obtained by upsampling the 8×8 reduced-resolution reconstructed prediction error blocks. For the creation of the edge pixels in each 16×16 reconstructed prediction error block, only the pixels which belong to the corresponding block are used. The detailed procedure is defined in Q.6.

Q.2.2.3 Reconstruction of block

For each luminance and chrominance block, a summation of the prediction and prediction error is performed. The procedure is identical to 6.3.1, except that the size of blocks is 16×16 instead of 8×8 . Then the clipping is performed according to the 6.3.2.

Then, a block boundary filter is applied to the boundary pixels of the 16×16 reconstructed blocks. The detailed procedure is described in Q.7.

Q.2.3 Picture store

If both H_R and V_R are divisible by 32, such as with the CIF format, the resulting picture reconstructed as described in Q.2.2 is stored as a reference picture as it is for further decoding. Otherwise, such as the QCIF format, the reconstructed picture which is just covered with

32×32 macroblocks is cropped at the right and the bottom to the width H_R and height V_R , and this cropped picture is stored as a reference picture for further decoding.

Q.2.4 Display

If both H and V are the same as H_C and V_C , the resulting picture in Q.2.2 is used for display purposes as it is. Otherwise, this resulting picture is further cropped to the size $H \times V$, and the cropped picture is used for display purpose only.

Q.3 Extension of referenced picture

If H_R or V_R is not divisible by 32, such as with the QCIF format, the extension of the referenced picture is performed before decoding macroblock/block layer. The width and the height of the extended reference picture for luminance are the next larger size that would be divisible by 32, and those for chrominance are the next larger size that would be divisible by 16.

NOTE – The width and height of the referenced picture in the default mode are always extended to be divisible by 16 even if the picture format has a width or height that is not divisible by 16, because the picture shall be decoded as if the width or height had the next larger size that would be divisible by 16. See 4.1.

If neither Unrestricted Motion Vector mode, Advanced Prediction mode nor Deblocking Filter mode is used with this option, the extended pixels can be arbitrary values, because the extended pixels will be never used as a reference pixels of the decoded picture to be reconstructed and displayed.

If either Unrestricted Motion Vector mode, Advanced Prediction mode or Deblocking Filter mode is also used with this option, the extension of the referenced picture is performed by duplicating the edge pixel of the referenced picture, in order to ensure the decoding when motion vectors point outside the right and bottom edge of the picture.

For example, if the Reduced-Resolution Update mode is used for a QCIF, the width of the referenced picture is 176 and the height is 144, which are not divisible by 32. In order to cover a QCIF picture with 32×32 -sized macroblocks, the number of macroblock row should be 6, and the number of macroblock column should be 5. Therefore, the width of the extended referenced picture is 192 and the height is 160.

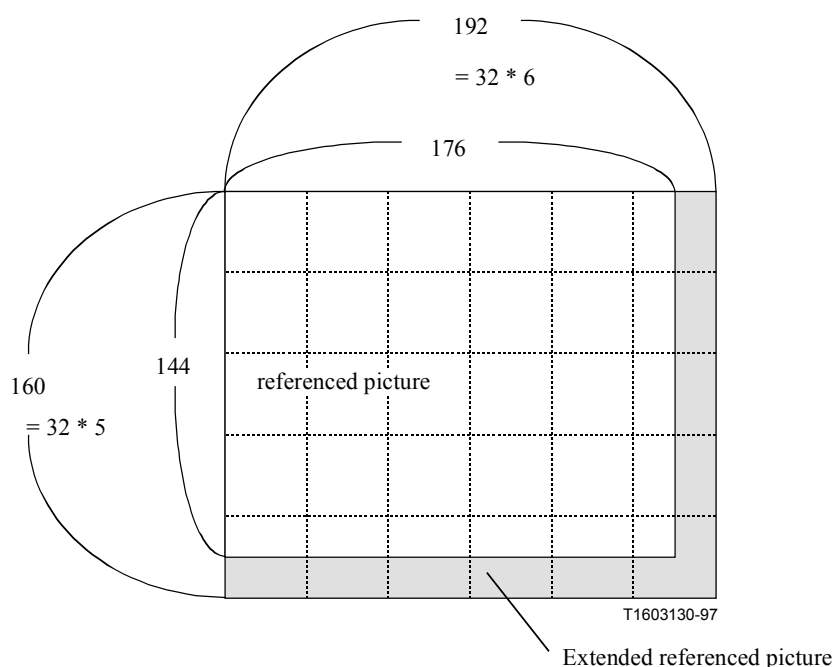


Figure Q.2/H.263 – Extension of referenced picture for QCIF picture size

The extension of the referenced picture in QCIF is illustrated in Figure Q.2. The extended referenced picture for luminance is given by the following formula:

$$R_{RRU}(x, y) = R(x', y')$$

where:

x, y = spatial coordinates of the extended referenced picture in the pixel domain

x', y' = spatial coordinates of the referenced picture in the pixel domain

$R_{RRU}(x, y)$ = pixel value of the extended referenced picture at (x, y)

$R(x', y')$ = pixel value of the referenced picture at (x', y')

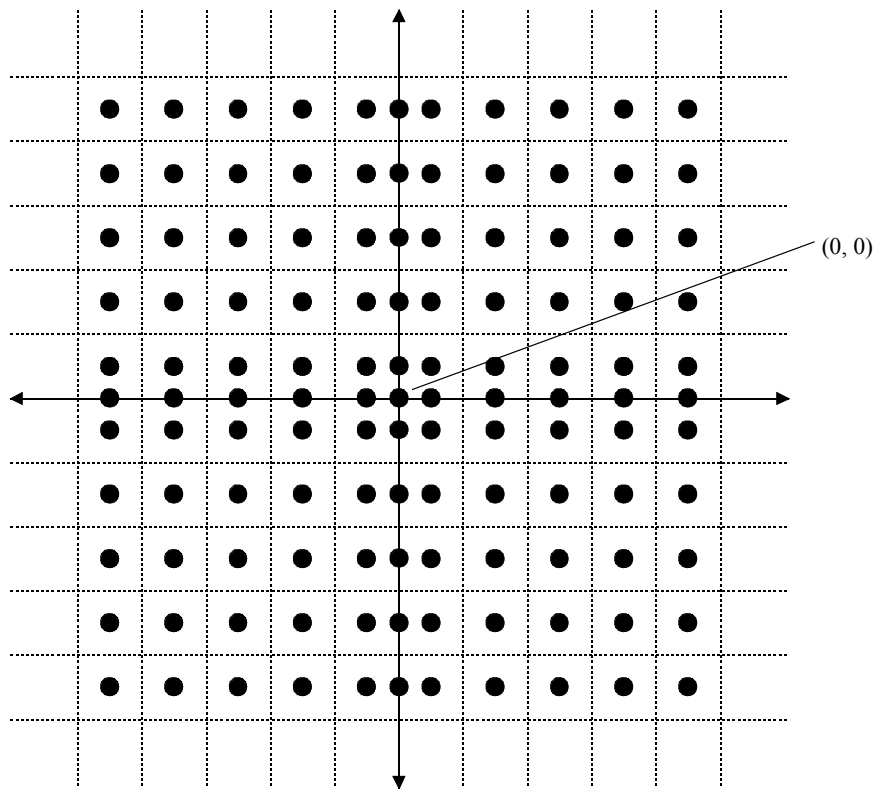
$$x' = \begin{cases} 175 & \text{if } x > 175 \text{ and } x < 192 \\ x & \text{otherwise} \end{cases}$$

$$y' = \begin{cases} 143 & \text{if } y > 143 \text{ and } y < 160 \\ y & \text{otherwise} \end{cases}$$

The referenced pictures for chrominance is also extended in the same manner.

Q.4 Reconstruction of motion vectors

In Reduced-Resolution Update mode, the motion vector range is enlarged to approximately double size in both the horizontal and vertical directions. In order to realize an enlarged range using the VLC for MVD defined in Table 14, each vector component is restricted to be only half-pel or zero value. Therefore the range of each motion vector component is $[-31.5, 30.5]$ in the default Reduced-Resolution Update mode. If the Unrestricted Motion Vector mode is used, the vector range $[-\text{limit}, \text{limit}]$ as defined in D.2 applies to the pseudo-motion vectors and translates into $[-(2 * \text{limit} - 0.5), 2 * \text{limit} - 1.5]$ for the motion vectors. For CIF this means that the pseudo-motion vector range is $[-32, 31.5]$ and the motion vector range is $[-63.5, 62.5]$. If the UUI field is set to "01", the motion vectors are unlimited. However, the motion vectors (not just the pseudo-motion vectors) are always limited to point not more than 15 pixels outside the coded area as described in D.1.1. Figure Q.3 illustrates the possible positions of macroblock motion vector or four motion vector predictions around the $(0, 0)$ vector value. The dashed lines indicate the integer coordinates.



T1603140-97

Figure Q.3/H.263 – Reconstruction of motion vector

For the macroblock using differential motion vectors in a B-picture, the motion vector for forward and backward prediction are obtained independently. In the Reduced-Resolution Update mode, the motion vector component MV_C for a luminance block is reconstructed from MVD and $MVD_{2.4}$ as follows:

- 1) Pseudo-prediction vector component **pseudo- P_C** is created from the prediction vector component P_C .

$$\text{pseudo-}P_C = 0 \quad \text{if } P_C = 0$$

$$\text{pseudo-}P_C = \text{sign}(P_C) * (|P_C| + 0.5) / 2.0 \quad \text{if } P_C \neq 0$$

"/" indicates floating-point division (without loss of accuracy). The prediction vector component P_C is defined as the median value of the vector components MV_1 , MV_2 and MV_3 as defined in 6.1.1 and F.2.

- 2) Pseudo-macroblock vector component **pseudo- MV_C** is obtained by adding the motion vector differences MVD (and $MVD_{2.4}$) from Table 14 to the pseudo- P_C .

In the default Reduced-Resolution Update mode, the value of **pseudo- MV_C** is restricted to the range $[-16, 15.5]$. Only one of the pair will yield a pseudo- MV_C falling within the permitted range. The procedure is performed in a similar way as defined in 6.1.1.

If the Unrestricted Motion Vector mode is also used with the Reduced-Resolution Update mode, **pseudo- MV_C** is obtained by adding the motion vector differences MVD (and $MVD_{2.4}$) from Table D.3.

If four motion vectors are present, the procedure is performed in a similar way as defined in F.2.

3) Motion vector component MV_C is obtained from pseudo- MV_C in the following formula:

$$MV_C = 0 \quad \text{if pseudo-}MV_C = 0$$

$$MV_C = \text{sign}(\text{pseudo-}MV_C) * (2.0 * |\text{pseudo-}MV_C| - 0.5) \quad \text{if pseudo-}MV_C \neq 0$$

As a result, each vector component is restricted to have a half-integer or zero value, and the range of each motion vector component is enlarged to approximately twice the pseudo-motion vector range.

4) If the current picture mode is an Improved PB-frame, or when the MBTYPE indicates the direct mode in a B-picture, the motion vector components MV_F and/or MV_B for forward and backward prediction are created.

First, pseudo-motion vector components **pseudo- MV_F** and/or **pseudo- MV_B** are calculated based on the rules for the prediction modes defined in Annexes O or M.

In the case of the Bidirectional prediction in the Improved PB-frames mode (see M.2.1) or when the MBTYPE indicates the direct mode in a B-picture (see O.5.2), pseudo- MV_F and pseudo- MV_B are calculated from **pseudo- MV_D** and **pseudo- MV_C** assuming that pseudo- MV_D is zero and pseudo- MV_C is MV , as defined in Annexes G and M.

In the case of forward prediction in the Improved PB-frames mode (see M.2.2), **pseudo- MV_{DB}** is obtained by decoding the variable length code MV_{DB} according to Table 12. Then, pseudo- MV_F is obtained by adding the pseudo- MV_{DB} to the **pseudo-Predictor**. In order to form the pseudo-Predictor, the predictor obtained according to the procedure defined in M.2.2 is converted to the pseudo-Predictor vector according to the formula defined in item 1 of this clause.

In the case of backward prediction in the Improved PB-frames mode (see M.2.3), pseudo- MV_B is set to zero.

Then, the motion vectors MV_F and/or MV_B for forward and backward prediction are obtained from pseudo- MV_F and/or pseudo- MV_B according to the formula defined in item 3 of this clause.

Q.5 Enlarged overlapped motion compensation for luminance

If Advanced Prediction mode is also used with Reduced-Resolution Update mode, enlarged matrices of weighting values are used to perform the overlapped motion compensation. Except that the size of each block and weighting matrices is 16×16 , the procedure of the creation of each prediction block is identical to the description in F.3.

The enlarged matrices of weighting values for the 16×16 luminance prediction are given in Figures Q.4, Q.5 and Q.6.

4	4	5	5	5	5	5	5	5	5	5	5	5	5	4	4
4	4	5	5	5	5	5	5	5	5	5	5	5	5	4	4
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	5	5	5	6	6	6	6	6	6	6	6	5	5	5	5
5	5	5	5	6	6	6	6	6	6	6	6	5	5	5	5
5	5	5	5	6	6	6	6	6	6	6	6	5	5	5	5
5	5	5	5	6	6	6	6	6	6	6	6	5	5	5	5
5	5	5	5	6	6	6	6	6	6	6	6	5	5	5	5
5	5	5	5	6	6	6	6	6	6	6	6	5	5	5	5
5	5	5	5	6	6	6	6	6	6	6	6	5	5	5	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
4	4	5	5	5	5	5	5	5	5	5	5	5	5	4	4
4	4	5	5	5	5	5	5	5	5	5	5	5	5	4	4

Figure Q.4/H.263 – Weighting values, H_0 , for prediction with motion vector of current 16×16 luminance block

2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1
1	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1
1	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

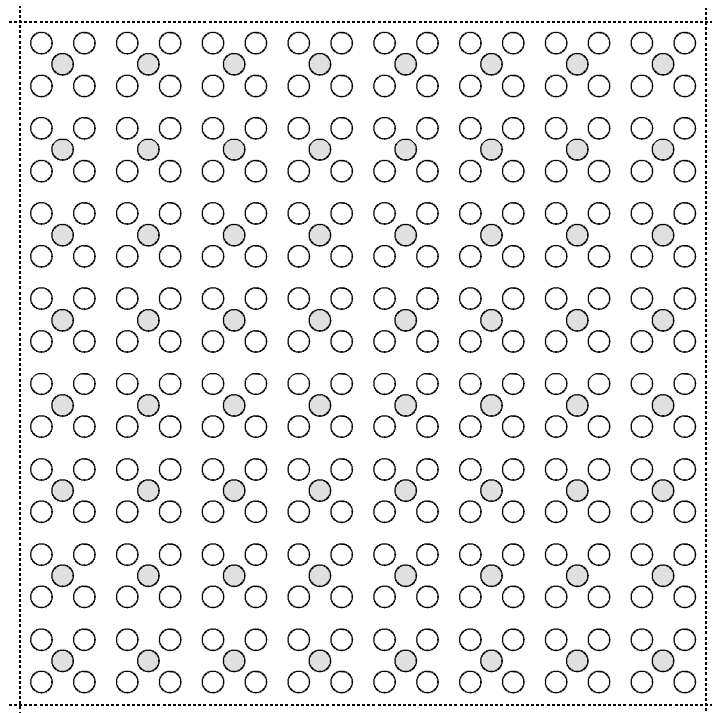
Figure Q.5/H.263 – Weighting values, H_1 , for prediction with motion vector of current 16×16 luminance blocks on top or bottom of current 16×16 luminance block

2	2	1	1	1	1	1	1	1	1	1	1	1	1	2	2
2	2	1	1	1	1	1	1	1	1	1	1	1	1	2	2
2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2
2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2
2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2
2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2
2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2
2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2
2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2
2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2
2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2
2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2
2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2
2	2	1	1	1	1	1	1	1	1	1	1	1	1	2	2
2	2	1	1	1	1	1	1	1	1	1	1	1	1	2	2

Figure Q.6/H.263 – Weighting values, H_2 , for prediction with motion vector of 16×16 luminance blocks to the left or right of current 16×16 luminance block

Q.6 Upsampling of the reduced-resolution reconstructed prediction error

The 16×16 reconstructed prediction error block is obtained by upsampling the 8×8 reduced-resolution reconstructed prediction error block. In order to realize a simple implementation, filtering is closed within a block which enables to perform an individual upsampling on block basis. Figure Q.7 shows the positioning of samples. The upsampling procedure for the luminance and chrominance pixels which are inside the 16×16 reconstructed prediction error blocks is defined in Q.6.1. For the creation of the luminance and chrominance pixels which are at the boundary of 16×16 reconstructed prediction error block, the procedure is defined in Q.6.2. Chrominance blocks as well as luminance blocks are up-sampled. The symbol "/" in Figures Q.8 and Q.9 indicates division by truncation.



T1603150-97

- Position of samples in 8 * 8 reduced-resolution reconstructed prediction error block
- Position of samples in 16 * 16 reconstructed prediction error block
- Block edge

Figure Q.7/H.263 – Positioning of samples in 8 × 8 reduced-resolution reconstructed prediction error block and 16 × 16 reconstructed prediction error block

Q.6.1 Upsampling procedure for the pixels inside a 16 × 16 reconstructed prediction error block

The creation of reconstructed prediction error for pixels inside block is described in Figure Q.8. "/" indicates division by truncation.

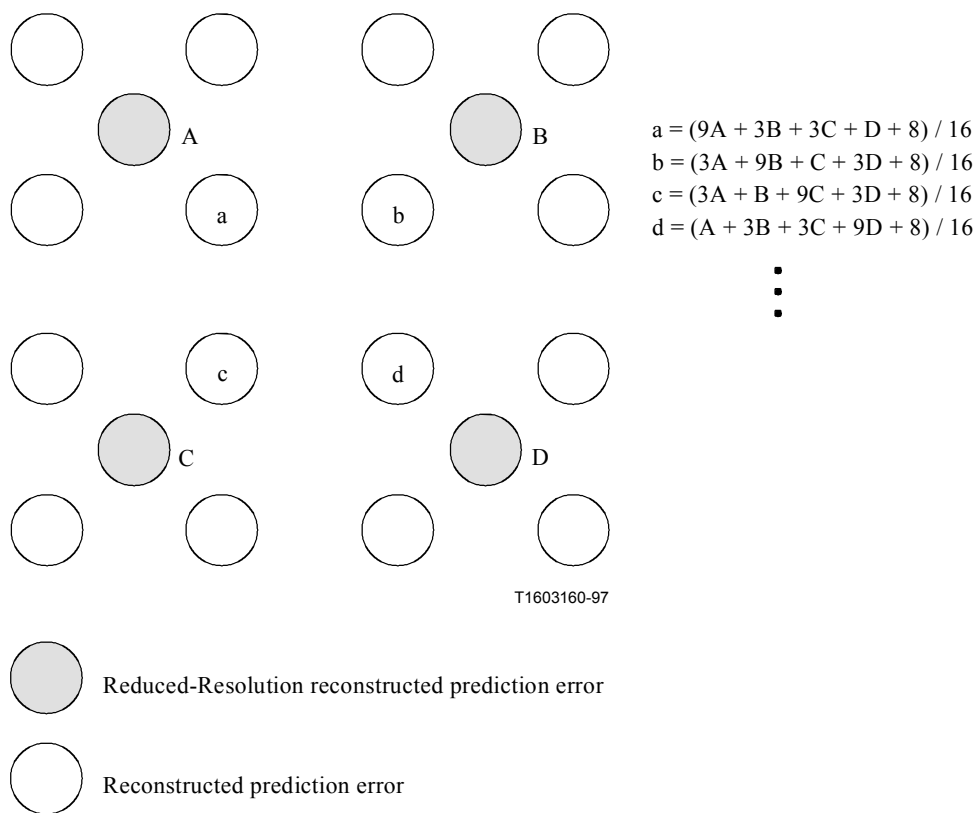


Figure Q.8/H.263 – Creation of reconstructed prediction error for pixels inside block

Q.6.2 Upsampling procedure for the pixels at the boundary of 16×16 reconstructed prediction error block

The creation of reconstructed prediction error for pixels of a 16×16 block is shown in Figure Q.9.

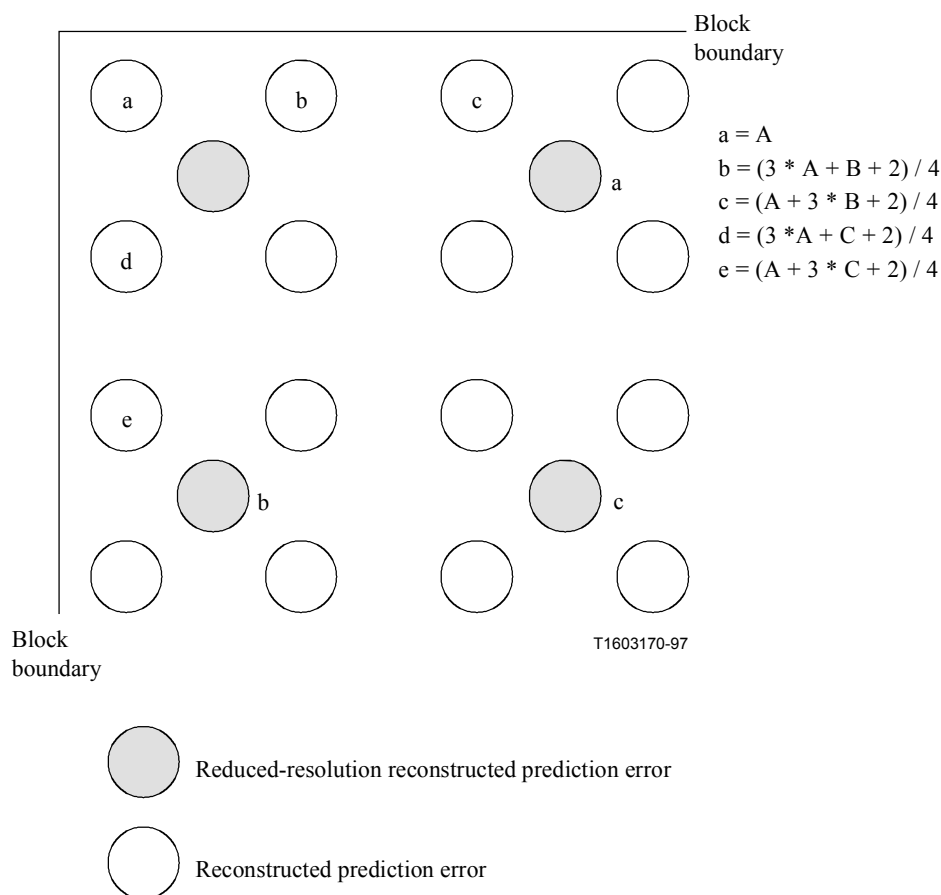


Figure Q.9/H.263 – Creation of reconstructed prediction error for pixels at the block boundary

Q.7 Block boundary filter

The filter operations are performed along the edges of the 16×16 reconstructed blocks at the encoder as well as on the decoder side. There are two alternative of filtering, depending on whether Deblocking Filter mode is used or not.

The default filtering in Reduced-Resolution Update mode is performed according to Q.7.1.

If Deblocking Filter mode is also used with Reduced-Resolution Update mode, the filtering is performed according to Q.7.2.

In both cases, filtering is performed on the complete reconstructed image data before storing the data in the picture store for future prediction. No filtering is performed across picture edges, slice edges in Slice Structured mode (see Annex K), or GOB boundaries having GOB headers present in Independent Segment Decoding mode (see Annex R). Chrominance as well as luminance data is filtered.

Q.7.1 Definition of the default block boundary filter

In the Reduced-Resolution Update mode, the default filtering is performed according to this subclause.

If A and B are two pixel values on a line – horizontal or vertical – of the reconstructed picture, and A belongs to one 16×16 block called block1 whereas B belongs to a neighboring 16×16 block called block2 which is to the right or below of block1. Figure Q.10 shows examples for the position of these pixels.

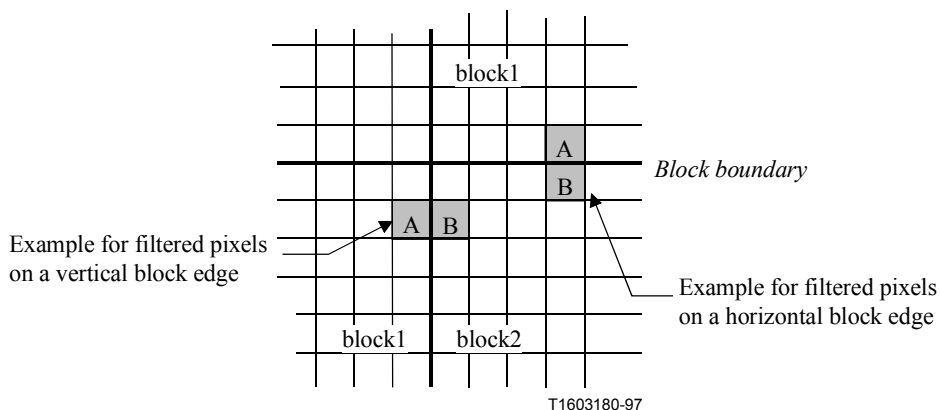


Figure Q.10/H.263 – Default block boundary filter

One of the following conditions must be fulfilled in order to turn the filter on for a particular edge:

- block1 belongs to a coded macroblock ($COD==0 \parallel MB\text{-type} == INTRA$); **or**
- block2 belongs to a coded macroblock ($COD==0 \parallel MB\text{-type} == INTRA$).

A shall be replaced by A1 and B shall be replaced by B1. "/" indicates division by truncation.

$$A1 = (3 * A + B + 2) / 4$$

$$B1 = (A + 3 * B + 2) / 4$$

The order of edges where filtering is performed is identical to the description provided in J.3.

Q.7.2 Definition of the block boundary filter when Deblocking Filter mode is used

If the Deblocking Filter mode (see Annex J) is used with the Reduced-Resolution Update mode, the filtering which is defined in Annex J with one modification is performed on the boundary pixels of 16×16 luminance and chrominance blocks, in place of the filtering described in Q.7.1. The one modification of the filtering in Annex J is that the parameter STRENGTH is given the value of positive infinity. This implies that the function $UpDownRamp(x, STRENGTH)$ defined in J.3 becomes a linear function of x.

As a result, the procedure of the deblocking filter described in J.3 is redefined in the following manner:

$$B1 = clip(B + d1)$$

$$C1 = clip(C - d1)$$

$$A1 = A - d2$$

$$D1 = D + d2$$

$$d1 = (A - 4B + 4C - D) / 8$$

$$d2 = clipd1((A - D) / 4, d1/2)$$

Annex R

Independent Segment Decoding mode

R.1 Introduction

This annex describes the optional Independent Segment Decoding mode of this Recommendation, which allows a picture to be decoded without the presence of any data dependencies across slice boundaries or across GOB boundaries having non-empty GOB headers. The use of this mode is indicated in the PLUSPTYPE field of the picture header. The capability to use this optional mode is negotiated by external means (for example, ITU-T Rec. H.245).

When the use of this mode is indicated, the video picture segment boundaries (as defined by the boundaries of the slices or the upper boundaries of the GOBs for which GOB headers are sent, or the boundaries of the picture, whichever bounds a region in the smallest way) are treated as picture boundaries when decoding, including the treatment of motion vectors which cross those boundaries (which result in boundary extrapolation when the Unrestricted Motion Vector mode, the Advanced Prediction mode, the Deblocking Filter mode, or the Temporal, SNR, and Spatial Scalability mode are in use, and which are prohibited when none of those optional modes are in use).

R.2 Mode operation

A video picture segment is defined by the following.

If the Slice Structured mode (see Annex K) is not in use, then one GOB or a plural number of consecutive GOBs forms one video picture segment. The location of the top of each video picture segment is indicated by the presence of a non-empty GOB header for which the border of the video picture segment lies just above the macroblocks in the GOB for which a header is present, or the top of the picture, whichever is lower. The location of the bottom of each video picture segment is defined by the top of the next video picture segment, or the bottom of the picture, whichever is uppermost.

If the Slice Structured mode (see Annex K) is in use, then each slice forms one video picture segment.

In the Independent Segment Decoding mode, each video picture segment is decoded with complete independence from all other video picture segments, and is also independent of all data outside the same video picture segment location in the reference picture(s). This includes:

- 1) no use of motion vectors outside of the current video picture segment for motion vector prediction (as in 6.1.1);
- 2) no use of motion vectors outside of the current video picture segment as remote motion vectors for overlapped block motion compensation when the Advanced Prediction mode is in use (see F.3);
- 3) no deblocking filter operation across video picture segment boundaries (see J.3);
- 4) no use of motion vectors which reference data outside the current video picture segment unless the Unrestricted Motion Vector mode (see Annex D), the Advanced Prediction mode (see Annex F), the Deblocking Filter mode (see Annex J) or the Temporal, SNR, and Spatial Scalability mode (see Annex O) are in use; in which case the borders of the current video picture segment in the prior picture are extrapolated as described in Annex D to form predictions of the pixels which reference the out-of-bounds region;
- 5) no bilinear interpolation across the boundaries of the $\frac{1}{4}$ -size or $\frac{1}{2}$ -size region corresponding to the current video picture segment for upward prediction in spatial scalability EI- and EP-pictures (as defined in Annex O);

- 6) when the Reduced-Resolution Update mode (see Annex Q) is in use, no block boundary filter operation across video picture segment boundaries;
- 7) no use of the Reference Picture Resampling mode with the Independent Segment Decoding mode.

R.3 Constraints on usage

Certain restrictions are placed on the use of other aspects of the video coding syntax when the Independent Segment Decoding mode is in use. These restrictions are made to prevent two pathological cases which otherwise would make operation of the Independent Segment Decoding mode difficult.

R.3.1 Constraint on segment shapes

In the use of the Slice Structured mode (Annex K) without the use of the Rectangular Slice submode (see K.1), there can arise cases in which the shape of a video picture segment may be non-convex (having "inside corners", or even comprising two distinct and separated regions of the picture).

Therefore, the Independent Segment Decoding mode shall not be used with the Slice Structured mode without the simultaneous use of the Rectangular Slice submode of the Slice Structured mode (see Annex K). This constraint is mandated to prevent the need for difficult special-case treatment in order to determine how and when to perform extrapolation of each video picture segment.

R.3.2 Constraint on changes of segment shapes

If the shape of the video picture segments were allowed to change in any way from picture to picture in the bitstream, there could arise cases in which the bitstream would be difficult to decode. This is because in such cases the bitstream content is not sufficient to determine the shape of each video picture segment prior to the possible appearance of motion vectors in the bitstream which require knowledge of the video picture segment shape for proper interpretation.

Therefore, when the Independent Segment Decoding mode is in use, the video picture segmentation for all pictures and frames allowing temporal prediction (i.e., all P-, B-, and EP-pictures and all Improved PB-frames) shall be the same as that used in its temporal reference picture. Also, when the Independent Segment Decoding mode is in use, the video picture segmentation for all EI-pictures shall either be the same or shall differ only by sub-dividing the video picture segmentation used in its reference picture. Also, the Independent Segment Decoding mode shall not be used in any picture or frame which uses reference pictures (all picture types except INTRA) unless the Independent Segment Decoding mode is also used in all of the reference picture(s) for the current picture. As a result of this constraint, the shape of the video picture segments in the Independent Segment Decoding mode shall never change from picture to picture except as changed in I- and EI-pictures (and the manner in which EI-pictures can change segmentation is itself also somewhat constrained).

Annex S

Alternative INTER VLC mode

S.1 Introduction

This annex describes an optional Alternative INTER VLC mode of this Recommendation, which improves the efficiency of inter-picture coding when significant changes are evident in the picture. This efficiency improvement is obtained by allowing some VLC codes originally designed for INTRA pictures to be used for some INTER picture coefficients and CBPY data as well. The use of this mode is indicated in the PLUSPTYPE field of the picture header. The capability to use this optional mode is negotiated by external means (for example, ITU-T Rec. H.245). The mode contains two syntax alterations, one for the encoding of INTER coefficients and another for the encoding of the INTER CBPY values.

S.2 Alternative INTER VLC for coefficients

The concept behind the design of the INTRA VLC table of Annex I is to use the same codewords as in the original INTER VLC but with a different interpretation of LEVEL and RUN. The INTRA VLC is better suited in cases where there are many and/or large-valued coefficients.

The INTRA VLC is constructed so that codewords have the same value for LAST (0 or 1) in both the INTER and INTRA tables. The INTRA table is therefore produced by "reshuffling" the meaning of the codewords with the same value of LAST. Furthermore, for events with large |LEVEL| the INTRA table uses a codeword which in the INTER table has a large RUN. In INTER blocks having a large number of large-magnitude coefficients, it can sometimes be more efficient to use the INTRA table than the INTER table, and in some such cases the choice of the VLC table can be apparent to the decoder since decoding using the INTER table would result in RUN values so large as to indicate the presence of more than 64 coefficients for a block. Under these circumstances, the INTRA table can be used to improve the efficiency of INTER coding.

S.2.1 Encoder action

The encoder may use the INTRA VLC table for coding an INTER block whenever the decoder can detect its use – in other words, whenever decoding using the INTER VLC table would cause coefficients outside the 64 coefficients of a block to be addressed.

The encoder would normally choose to use the INTRA VLC table for coding an INTER block only when the above condition is satisfied and also when the INTRA VLC usage results in fewer bits than the INTER VLC for the same coefficient values. This will often be the case when there are many large coefficients, due to the way the INTRA VLC was produced (since the ordinary INTER VLC table contains long run-lengths for the same codewords in which the INTRA VLC table contains large coefficient amplitudes).

S.2.2 Decoder action

The decoding process is as follows:

- 1) The decoder first receives all coefficient codes of a block.
- 2) The codewords are then interpreted assuming that INTER VLC is used. If the addressing of coefficients stays inside the 64 coefficients of a block, the VLC decoding is finished.
- 3) If coefficients outside the block are addressed, the codewords shall be interpreted according to the INTRA VLC.

S.3 Alternative INTER VLC for CBPY

The INTER CBPY codewords (Table 12) are designed with the assumption that there are more Y blocks with all zero coefficients than there are with at least one non-zero coefficient. When both C_B and C_R blocks have at least one non-zero coefficient, i.e., $CBPC_5 = CBPC_6 = 1$, this assumption no longer holds. For this reason, when the Alternative INTER VLC mode is in use, the CBPY codewords as defined in Table 12 for INTRA macroblocks shall also be used for INTER macroblocks whenever $CBPC_5 = CBPC_6 = 1$.

Annex T

Modified Quantization mode

T.1 Introduction

This annex describes an optional Modified Quantization mode of this Recommendation, which modifies quantizer operation. The use of this mode is indicated in the PLUSPTYPE field of the picture header. The capability to use this optional mode is negotiated by external means (for example, ITU-T Rec. H.245).

This mode includes four key features:

- 1) The bit-rate control ability for encoding is improved by altering the syntax for the DQUANT field.
- 2) Chrominance fidelity is improved by specifying a smaller step size for chrominance than that for luminance data.
- 3) The range of representable coefficient values is extended to allow the representation of any possible true coefficient value to within the accuracy allowed by the quantization step size.
- 4) The range of quantized coefficient levels is restricted to those which can reasonably occur, to improve the detectability of errors and minimize decoding complexity.

T.2 Modified DQUANT Update

This mode modifies the semantics of the DQUANT field. With this mode, it is possible to use DQUANT either to modify QUANT by plus or minus a small amount, or to signal any specific new value for QUANT. The size of the small amount of modification depends on the current value of QUANT. By use of this mode, more flexible control of the quantizer step size can be specified in the DQUANT field.

The codeword for DQUANT in this mode is no longer a two-bit fixed length field. It is a variable length field which can either be two bits or six bits in length. Whether it is two or six bits depends on the first bit of the code. The description below is therefore split into two sections, depending on the first bit.

T.2.1 Small-step QUANT alteration

When the first bit of the DQUANT field is 1, only one additional bit is sent in DQUANT. The single additional bit is used to modify QUANT by a differential value. The change in the value of QUANT is dependent on the second bit of DQUANT and on the prior value of QUANT, as shown in Table T.1.

Example: If the previous value of QUANT is 29 and DQUANT is signalled with the codeword "11", then the differential value is +2, and thus the resulting new QUANT value is 31.

Table T.1/H.263 – Semantics of small-step QUANT alteration

Prior QUANT	Change of QUANT	
	DQUANT = 10	DQUANT = 11
1	+2	+1
2-10	-1	+1
11-20	-2	+2
21-28	-3	+3
29	-3	+2
30	-3	+1
31	-3	-5

T.2.2 Arbitrary QUANT selection

When the first bit of the DQUANT field is 0, five additional bits are sent in DQUANT. The following five bits represent a new QUANT as defined in 5.1.19.

Example: Regardless of the current value of QUANT, if DQUANT is signalled with the code word '001111', then the new value of QUANT is 15.

T.3 Altered quantization step size for chrominance coefficients

When the modified quantization mode is in use, the quantization parameter of the chrominance coefficients is different from the quantization parameter of the luminance. The luminance quantization parameter is signalled in the bitstream. It is called QUANT. When this mode is in use, a different quantization parameter termed QUANT_C is used for the inverse quantization of chrominance coefficients. The relation between QUANT and QUANT_C is given in Table T.2. If the Deblocking Filter mode (see Annex J) is in use, QUANT_C shall also be used for the application of the deblocking filter to the chrominance data. Whenever QUANT is discussed herein in any other context, it shall mean the quantization step size of the luminance.

Table T.2/H.263 – Relationship between QUANT and QUANT_C

Range of QUANT	Value of QUANT_C
1-6	QUANT_C = QUANT
7-9	QUANT_C = QUANT - 1
10-11	9
12-13	10
14-15	11
16-18	12
19-21	13
22-26	14
27-31	15

T.4 Modified coefficient range

When the Modified Quantization mode is in use, quantized DCT coefficients having quantization level magnitudes greater than 127 can be represented in the bitstream. This has two advantages:

- 1) Encoder performance is improved by allowing the true full range of possible coefficient values to be represented.
- 2) Encoder complexity is reduced by eliminating the need to increase the quantization step size upon encountering certain large coefficient values which would otherwise be unrepresentable.

It is possible that the correct value of a DCT coefficient prior to quantization in the encoder will have a magnitude as high as 2040. Thus, a range of -127 to $+127$ for LEVEL is insufficient to cover the entire range of possible coefficient values whenever the quantization parameter QUANT or QUANT_C is less than 8. The expanded coefficient range broadens the range of LEVEL to allow any true coefficient value to be more properly encoded.

When the Modified Quantization mode is in use, the meaning of the LEVEL field following an ESCAPE code (0000 011, as per 5.4.2) is altered. In this mode, rather than being forbidden, the bit sequence 1000 0000 is used to represent an EXTENDED-ESCAPE code. An AC coefficient of magnitude greater than 127 is represented by sending an EXTENDED-ESCAPE code, immediately followed by a fixed-length EXTENDED-LEVEL field of 11 bits. An extended coefficient value is encoded into the EXTENDED-LEVEL field by taking the least significant 11 bits of the two's-complement binary representation of LEVEL and cyclically rotating them to the right by 5-bit positions. This rotation is necessary to prevent start code emulation. The cyclic rotation is illustrated in Figure T.1.

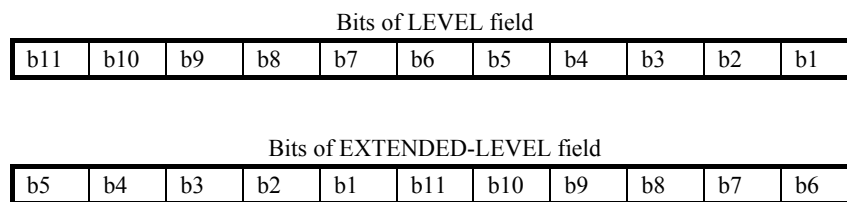


Figure T.1/H.263 – Cyclic rotation of coefficient representation

T.5 Usage restrictions

When the Modified Quantization mode is in use, certain restrictions are placed on the encoded coefficient values. This has several benefits:

- 1) The detectability of bit errors is improved by prohibiting certain unreasonable coefficient values, thus allowing these values to be recognized as bit errors by the decoder; and
- 2) Decoder complexity is reduced by reducing the wordlength necessary for inverse quantization prior to clipping;
- 3) Start-code emulation is prevented for coefficients encoded using the EXTENDED-ESCAPE mechanism described in T.4.

These restrictions are as follows. When the Modified Quantization mode is in use:

- 1) For any coefficient, the reconstruction level magnitude $|REC|$ produced by the inverse quantization process described in 6.2.1 using the current value of QUANT or QUANT_C as appropriate, and the encoded value of LEVEL, shall be less than 4096. This additional restriction applies to all coefficients, regardless of whether the coefficient is sent using the EXTENDED-ESCAPE mechanism or not.

- 2) The bitstream shall not use either the normal ESCAPE code or the EXTENDED-ESCAPE code to encode a combination of LAST, RUN, and LEVEL for which there exists a codeword entry in the applicable VLC table, which is either Table 16 (see 5.4.2) or Table I.2 (see I.3).
- 3) The EXTENDED-ESCAPE code shall be used only when the quantization parameter for the coefficient (QUANT or QUANT_C) is less than eight (8).
- 4) The EXTENDED-ESCAPE code shall be used only when it is followed by an EXTENDED-LEVEL field representing a value of LEVEL which is outside of the range -127 to +127.

Annex U

Enhanced reference picture selection mode

U.1 Introduction

This annex describes the optional Enhanced Reference Picture Selection (ERPS) mode of this Recommendation. The capability to use this optional mode is negotiated by external means (for example, ITU-T Rec. H.245). The amount of picture memory accommodated in the decoder for ERPS operation should also be signalled by external means. The use of this mode shall be indicated by setting the formerly reserved bit 16 of the optional part of the PLUSPTYPE (OPPTYPE) to "1". The mode provides benefits for both error resilience and coding efficiency by using a memory buffer of reference pictures.

A sub-mode of the ERPS mode is specified for Sub-Picture Removal. The purpose of Sub-Picture Removal is to reduce the amount of memory required to store multiple reference pictures. The memory reduction is accomplished by specifying the partitioning of each reference picture into smaller rectangular units called sub-pictures. The encoder can then indicate to the decoder that specific sub-picture areas of specific reference pictures will not be used as a reference for the prediction of subsequent pictures, thus allowing the memory allocated in the decoder for storing these areas to be used to store data from other reference pictures. The support for this sub-mode and the allowed fragmentation of the picture memory into minimum picture units (MPUs) for Sub-Picture Removal as defined herein is also negotiated by external means (for example, ITU-T Rec. H.245).

A sub-mode of the ERPS mode is specified for enabling two-picture backward prediction in B pictures. This sub-mode can enhance performance by providing encoders for B pictures not only with an ability to use multiple references for forward prediction, but also to use more than one reference picture for backward prediction. The support for this sub-mode is negotiated by external means (for example, ITU-T Rec. H.245).

For error resilience, the ERPS mode can use backward channel messages, which are signalled by external means (for example, ITU-T Rec. H.245) sent from a decoder to an encoder to inform the encoder which pictures or parts of pictures have been incorrectly decoded. The ERPS mode provides enhanced performance compared to the Reference Picture Selection (RPS) mode defined in Annex N. It shall not be used simultaneously with the RPS mode. (It can be used in such a way as to provide essentially the same functionality as the RPS mode.)

For coding efficiency, motion compensation can be extended to prediction from multiple pictures. The extension of motion compensation to multi-picture prediction is achieved by extending each motion vector by a picture reference parameter that is used to address a macroblock or block prediction region for motion compensation in any of the multiple reference pictures. The picture

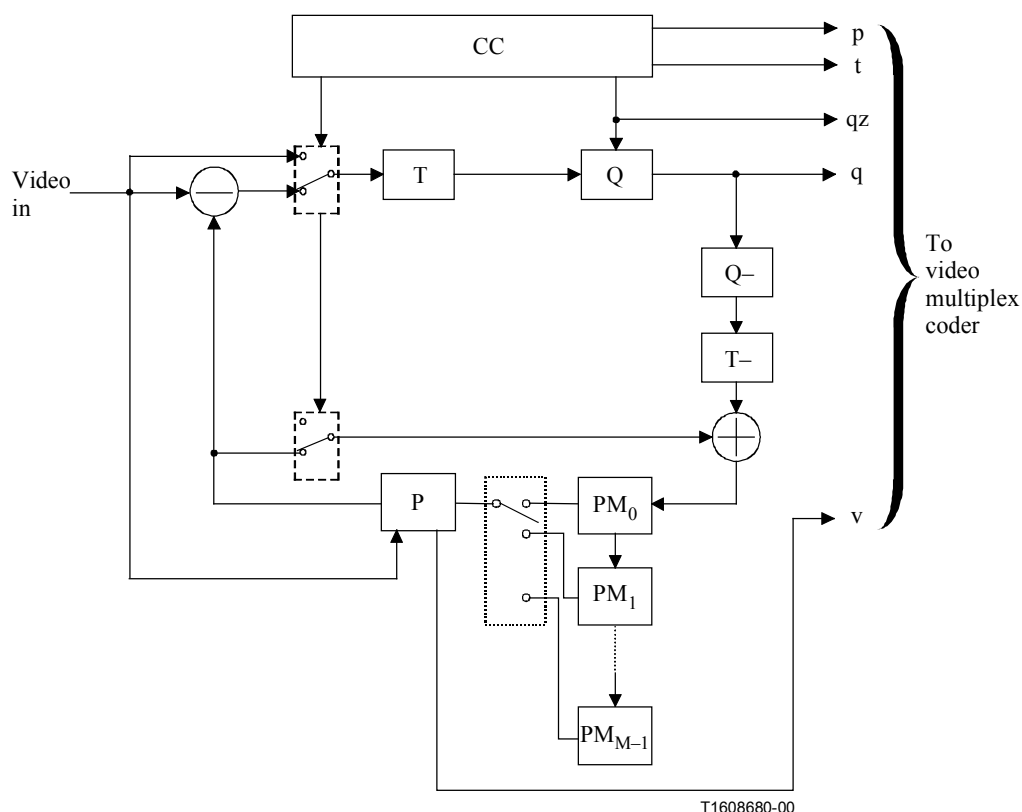
reference parameter is a variable length code specifying a relative buffer index. The reference pictures are assembled in a buffering scheme that is controlled by the encoder.

The ERPS mode shall not be used with the Syntax-based Arithmetic Coding mode (see Annex E) or the Data Partitioned Slice mode (see Annex V).

Once activated, the ERPS mode shall not be inactivated in subsequent pictures in the bitstream unless the initial inactivation occurs in an I or EI picture and any reactivation is also in an I or EI picture and is accompanied by a buffer reset (RESET equal to "1"). If inactivated, the entire contents of the ERPS multi-picture buffer shall be set to "unused" status.

U.2 Video source coding algorithm

The source coder of this mode is shown in generalized form in Figure U.1. This figure shows a structure that uses a number of picture memories.



T	Transform
Q	Quantizer
P	Picture Memory with motion compensated variable delay
PM	Picture Memory
CC	Coding control
p	Flag for INTRA/INTER
t	Flag for transmitted or not
qz	Quantizer indication
q	Quantizing index for transform coefficients
v	Motion vector

Figure U.1/H.263 – Source coder for Enhanced Reference Picture Selection mode

The video source coding algorithm can be extended to multi-picture motion compensation. Enhanced coding efficiency may be achieved by allowing reference picture selection on the macroblock level. A picture buffering scheme with relative indexing is employed for efficient addressing of pictures in the multi-picture buffer. The multi-picture buffer control may work in two distinct types of operation.

In the first of these two types of operation, a "Sliding Window" over time can be accommodated by the buffer control unit. In such a buffering scheme using M picture memories $PM_0 \dots PM_{M-1}$, the most recent preceding (up to M) decoded and reconstructed pictures are stored in the picture memories and can be used as references for decoding. If the number of pictures maximally accommodated by the multi-picture buffer corresponds to M , the motion estimation when coding a picture m , if $0 \leq m \leq M - 1$, can utilize m pictures. When coding a picture $m \geq M$, the maximum number of pictures M can be used. Alternatively, a second "Adaptive Memory Control" type of operation can be used for a more flexible and specific control of the picture memories than with the simple "Sliding Window" scheme.

The operation of the ERPS mode results in the assignment of "unused" status to some pictures or sub-picture areas of pictures that have been sent to the decoder. Once some picture or area of a picture has been assigned to "unused" status, the bitstream shall not contain any data that causes a reference to any "unused" area for the prediction of subsequent pictures. By managing the assignment of "unused" status to previous pictures, the encoder shall ensure that sufficient memory is available in the decoder to store all data needed for the representation of subsequent pictures. The overall buffer size and structure is conveyed to the decoder in the bitstream, and the encoder shall control the buffer such that the specified total capacity is not exceeded by stored picture data that has not been assigned to "unused" status.

The source coder may select one or several of the picture memories to suppress temporal error propagation caused by inter-picture coding. The Independent Segment Decoding mode (see Annex R), which treats boundaries of GOBs with non-empty headers or slices as picture boundaries, can be used to avoid spatial error propagation due to motion compensation across the boundaries of the GOBs or slices when this mode is applied to a smaller unit than a picture, such as a GOB or slice. The information to signal which picture is selected for prediction is included in the encoded bitstream.

The strategy used by the encoder to select the picture or pictures to be used for prediction is out of the scope of this Recommendation.

U.3 Forward-channel syntax

The syntax is altered in the picture, Group of Blocks (GOB), and slice layers. When indicated by a parameter MRPA being equal to "1", the syntax is also altered in the macroblock layer. In the picture, GOB, and slice layers, an Enhanced Reference Picture Selection layer (ERPS layer) is inserted. In the macroblock layer, picture reference parameters are inserted under certain conditions to enable multi-picture motion compensation.

U.3.1 Syntax of the picture, GOB, and slice layer

The Enhanced Reference Picture Selection syntax for the PLUS header (otherwise as shown in Figure 8) is shown in Figure U.2. The fields of RPSMF, PN, and the ERPS layer are inserted into the PLUS header. The fields of TRPI, TRP, BCI, and BCM are not present (since they are only needed for the RPS mode of Annex N, which is not allowed when the ERPS mode is active).

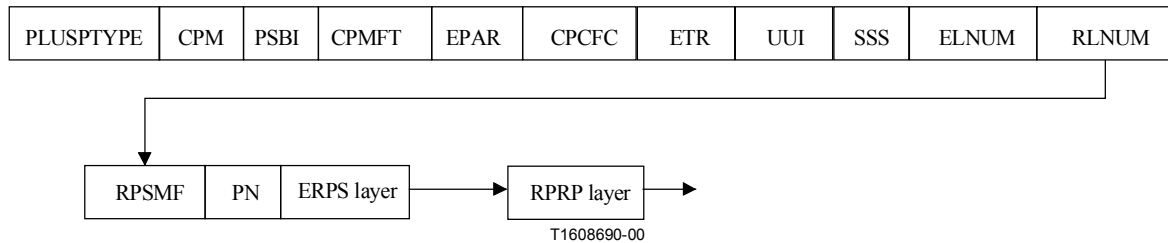


Figure U.2/H.263 – Structure of PLUS header for the ERPS mode

The syntax for the GOB layer is shown in Figure U.3. The fields of PNI, PN, NOERPSL, and the ERPS layer are added to the syntax (otherwise defined as in Figure 9).

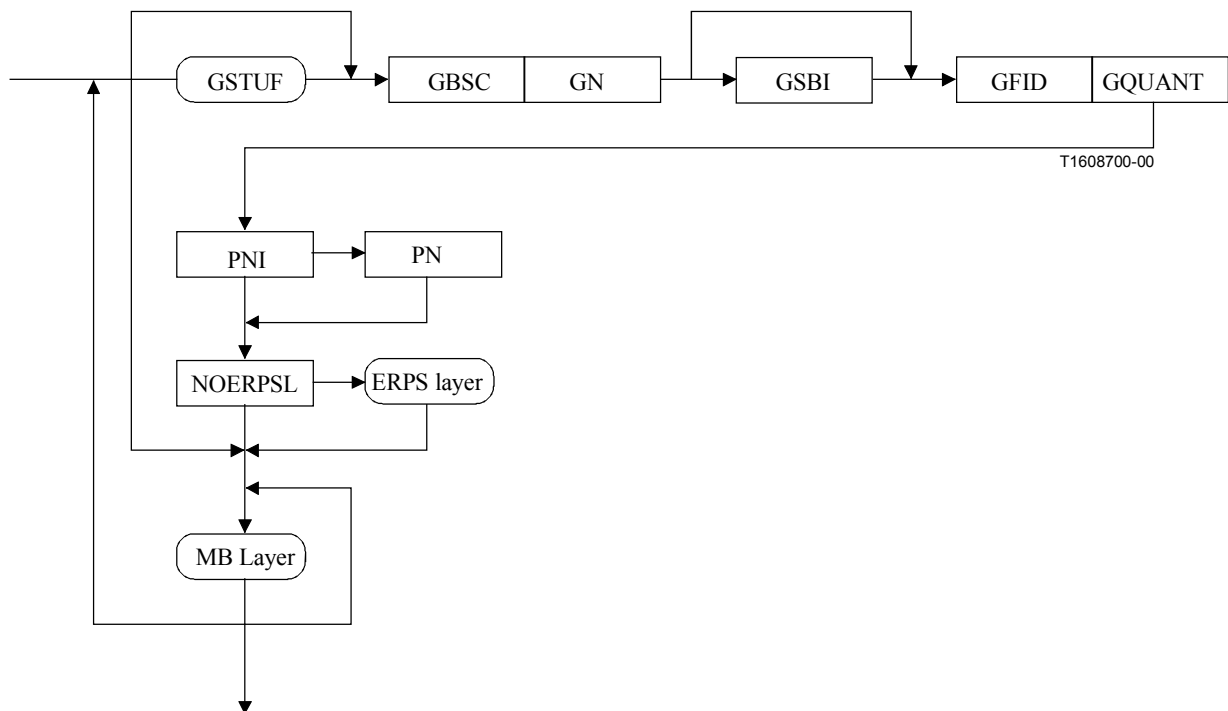


Figure U.3/H.263 – Structure of GOB layer for the ERPS mode

When the optional Slice Structured mode (see Annex K) is in use, the syntax of the slice layer is modified in the same way as the GOB layer. The syntax for the slice layer is shown in Figure U.4. The slice that immediately follows the picture start code in the bitstream also includes all of the added fields PNI, PN, NOERPSL, and the ERPS layer.

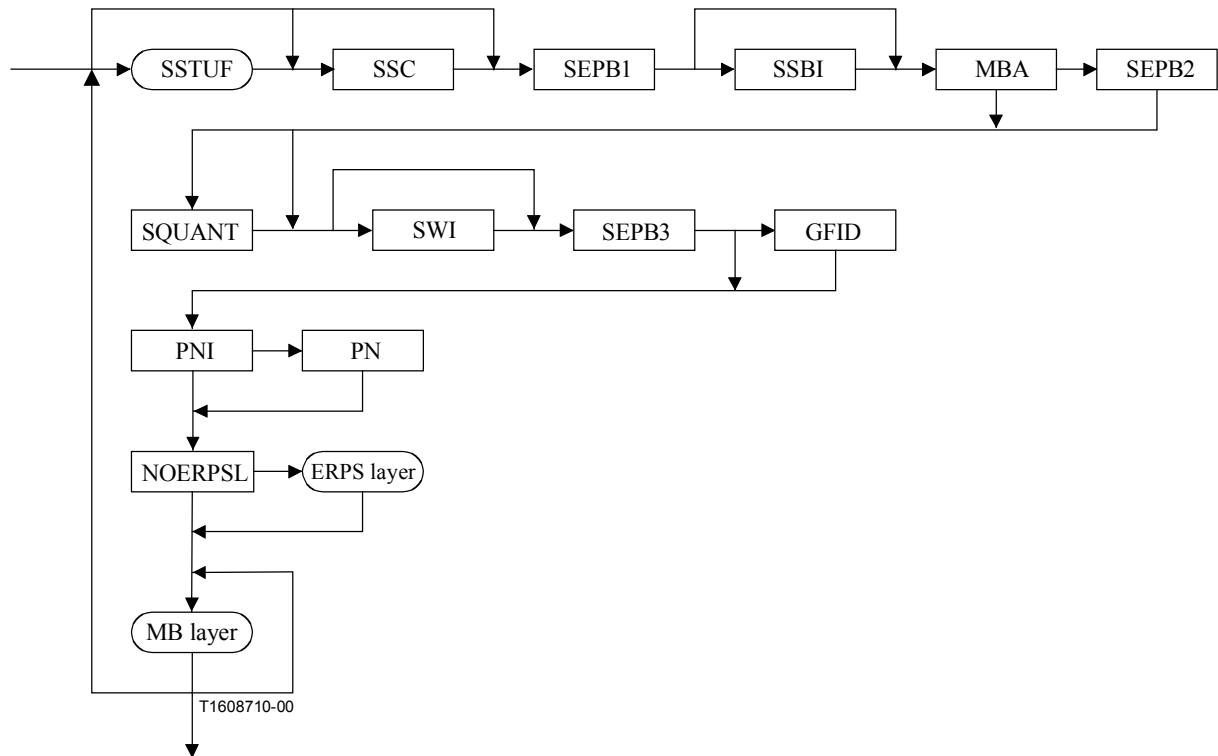


Figure U.4/H.263 – Structure of slice layer for the ERPS mode

The ERPS layer is shown in Figure U.5.

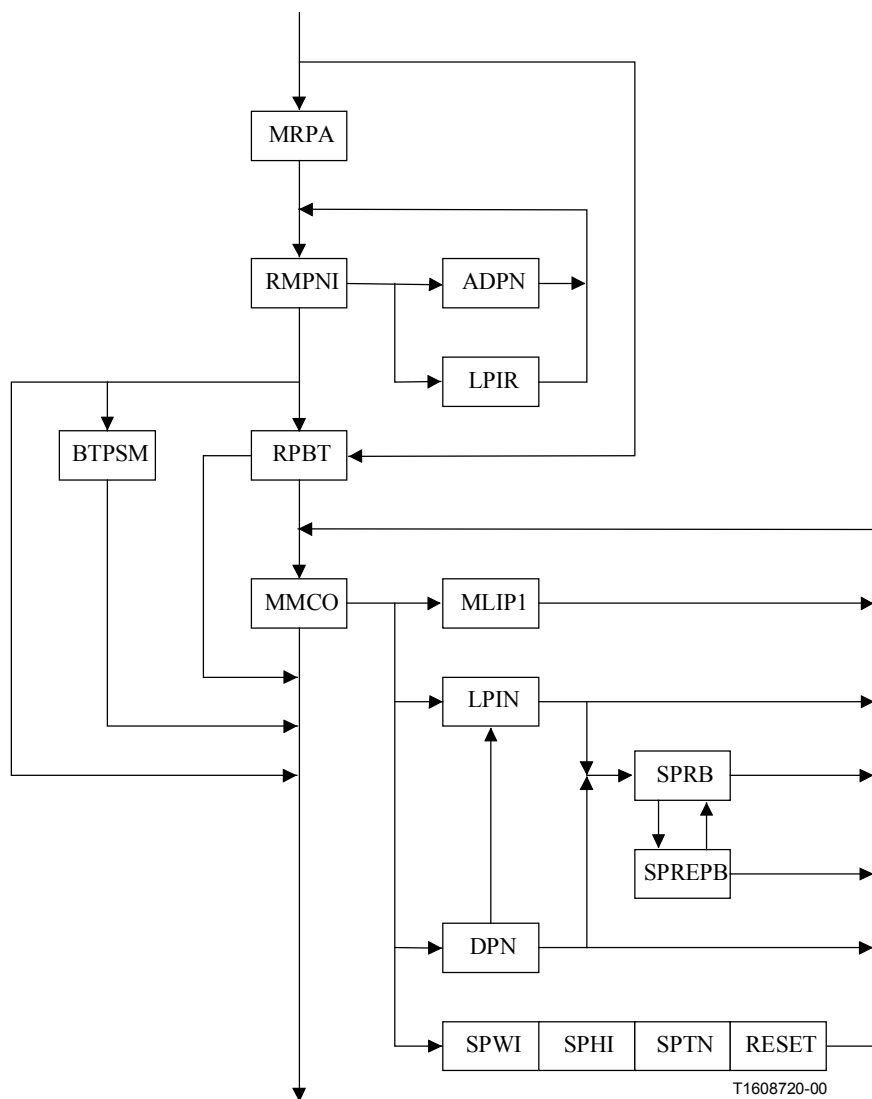


Figure U.5/H.263 – Structure of the ERPS layer

Variable length codes for the ADPN, LPIR, MLIP1, DPN, LPIN, SPTN, PR, PR₀, PR₂, PR₃, PR₄, PRB and PRFW fields are given in Table U.1.

Table U.1/H.263 – Variable length codes for ADPN, LPIR, MLIP1, DPN, LPIN, SPTN, PR, PR₀, PR₂, PR₃, PR₄, PRB, and PRFW

Absolute position	Number of bits	Codes
0	1	1
"x ₀ " + 1 (1:2)	3	0x ₀ 0
"x ₁ x ₀ " + 3 (3:6)	5	0x ₁ 1x ₀ 0
"x ₂ x ₁ x ₀ " + 7 (7:14)	7	0x ₂ 1x ₁ 1x ₀ 0
"x ₃ x ₂ x ₁ x ₀ " + 15 (15:30)	9	0x ₃ 1x ₂ 1x ₁ 1x ₀ 0
"x ₄ x ₃ x ₂ x ₁ x ₀ " + 31 (31:62)	11	0x ₄ 1x ₃ 1x ₂ 1x ₁ 1x ₀ 0
"x ₅ x ₄ x ₃ x ₂ x ₁ x ₀ " + 63 (63:126)	13	0x ₅ 1x ₄ 1x ₃ 1x ₂ 1x ₁ 1x ₀ 0
"x ₆ x ₅ x ₄ x ₃ x ₂ x ₁ x ₀ " + 127 (127:254)	15	0x ₆ 1x ₅ 1x ₄ 1x ₃ 1x ₂ 1x ₁ 1x ₀ 0
"x ₇ x ₆ x ₅ x ₄ x ₃ x ₂ x ₁ x ₀ " + 255 (255:510)	17	0x ₇ 1x ₆ 1x ₅ 1x ₄ 1x ₃ 1x ₂ 1x ₁ 1x ₀ 0
"x ₈ x ₇ x ₆ x ₅ x ₄ x ₃ x ₂ x ₁ x ₀ " + 511 (511:1022)	19	0x ₈ 1x ₇ 1x ₆ 1x ₅ 1x ₄ 1x ₃ 1x ₂ 1x ₁ 1x ₀ 0
"x ₉ x ₈ x ₇ x ₆ x ₅ x ₄ x ₃ x ₂ x ₁ x ₀ " + 1023 (1023:2046)	21	0x ₉ 1x ₈ 1x ₇ 1x ₆ 1x ₅ 1x ₄ 1x ₃ 1x ₂ 1x ₁ 1x ₀ 0
"x ₁₀ x ₉ x ₈ x ₇ x ₆ x ₅ x ₄ x ₃ x ₂ x ₁ x ₀ " + 2047 (2047:4094)	23	0x ₁₀ 1x ₉ 1x ₈ 1x ₇ 1x ₆ 1x ₅ 1x ₄ 1x ₃ 1x ₂ 1x ₁ 1x ₀ 0

U.3.1.1 Reference Picture Selection Mode Flags (RPSMF) (3 bits)

RPSMF is a 3-bit fixed-length codeword that is present in the PLUS header whenever the ERPS mode is in use (regardless of the value of UFEP). RPSMF shall not be present in the GOB or slice layer. When present, RPSMF indicates which type of back-channel messages are needed by the encoder. The values of RPSMF shall be as defined in 5.1.13.

U.3.1.2 Picture Number Indicator (PNI) (1 bit)

PNI is a single-bit fixed-length codeword that is always present at the GOB or slice layer when the ERPS mode is in use, and is not present in the PLUS header. When present, PNI indicates whether or not the following PN field is also present.

"0": PN field is not present.

"1": PN field is present.

U.3.1.3 Picture Number (PN) (10 bits)

PN is a 10-bit fixed-length codeword that is always present in the PLUS header when the ERPS mode is in use, and is present at the GOB or slice layer only when indicated by PNI.

PN shall be incremented by 1 for each coded and transmitted picture, in a 10-bit modulo operation, relative to the PN of the previous stored picture. The term "stored picture" is defined in U.3.1.5.7. For EI and EP pictures, PN shall be incremented from the value in the last stored EI or EP picture within the same scalability enhancement layer. For B pictures, PN shall be incremented from the value in the most temporally-recent stored non-B picture in the reference layer of the B picture which precedes the B picture in bitstream order (a picture which is temporally subsequent to the B picture). B pictures are not stored in the multi-picture buffer, as they are not used as references for subsequent pictures. Thus, a picture immediately following a B picture in the reference layer of the B picture or another B picture which immediately follows a B picture shall have the same PN as the

B picture. Similarly, if a non-B picture is present in the bitstream which is not stored, the picture following this non-B picture (in the same enhancement layer, in the case of Annex O operation) shall have the same PN as the non-stored non-B picture.

In a usage scenario known as "Video Redundancy Coding", the ERPS mode may be used by some encoders in a manner in which more than one representation is sent for the pictured scene at the same temporal instant (usually using different reference pictures). In such a case in which the ERPS mode is in use and in which adjacent pictures in the bitstream have the same temporal reference and the same picture number, the decoder shall regard this occurrence as an indication that redundant copies have been sent of approximately the same pictured scene content, and shall decode and use the first such received picture while discarding the subsequent redundant picture(s).

The PN serves as a unique ID for each picture stored in the picture buffer within 1024 coded and stored pictures. Therefore, a picture cannot be kept in the buffer after more than 1023 subsequent coded and stored pictures (in the same enhancement layer, in the case of Annex O operation) unless it has been assigned a long-term picture index as specified below. The encoder shall ensure that the bitstream shall not specify retaining any short-term picture after more than 1023 subsequent stored pictures. A decoder which encounters a picture number on a current stored picture having a value equal to the picture number of some other short-term picture in the multi-picture buffer should treat this condition as an error.

U.3.1.4 No Enhanced Reference Picture Selection Layer (NOERPSL) (1 bit)

NOERPSL is a single-bit fixed-length codeword that is present at the GOB or slice level whenever the ERPS mode is in use. It is not present in the PLUS header. The values of NOERPSL shall be as follows:

"0": The ERPS layer is sent.

"1": The ERPS layer is not sent.

If NOERPSL is "1", all ERPS settings and re-mappings in effect for the picture shall be applied also for the relevant video picture segment that follows the GOB or slice layer data. ERPS layer information sent at the GOB or slice level governs the decoding process for the video picture segment preceded by the GOB or slice level data, and does not affect the decoding process of any other video picture segment. (See Annex R for the definition of a video picture segment.)

U.3.1.5 Enhanced Reference Picture Selection layer (ERPS) (variable length)

The ERPS layer is always present at the picture level when the ERPS mode is in use, and is present at the GOB or slice level if NOERPSL is "0". It specifies the buffer indexing used to decode the current picture or video picture segment, and manages the contents of the picture buffer.

U.3.1.5.1 Multiple Reference Pictures Active (MRPA) (1 bit)

MRPA is a single-bit fixed-length codeword that is present only if the picture coding type indicates a P picture, an EP picture, an Improved PB frame, or B picture. MRPA is the first element in the ERPS layer if present. MRPA specifies whether the number of active reference pictures for forward-prediction or backward-prediction decoding of the current picture or video picture segment may be larger than one. The value of MRPA shall be as follows:

"1": More than one reference picture may be used for forward or backward motion compensation.

"0": Only one reference picture is used for forward or backward motion compensation. In this case, the extensions of the macroblock layer syntax in U.3.2 do not apply.

MRPA may be changed from video picture segment to video picture segment, so that different video picture segments may address different numbers of reference pictures.

MRPA shall be "0" in any picture which invokes the Reference Picture Resampling mode (see Annex P), and the same picture shall be indicated as the forward reference picture to be used at both the picture and GOB or slice levels for any such current picture. If the current picture is a B picture, the backward reference picture shall have the same size as the current picture, and any reference picture resampling process shall be applied only to the forward reference picture. Reference picture resampling shall be invoked only if the multi-picture buffer contains sufficient "unused" capacity to store the resampled forward reference picture, but after the resampled reference picture is used for the decoding of the current picture, the resampled forward reference picture shall not be stored in the multi-picture buffer.

U.3.1.5.2 Re-Mapping of Picture Numbers Indicator (RMPNI) (variable length)

RMPNI is a variable-length codeword that is present in the ERPS layer if the picture is a P, EP, Improved PB, or B picture. RMPNI indicates whether any default picture indices are to be re-mapped for motion compensation of the current picture or video picture segment – and how the re-mapping of the relative indices into the multi-picture buffer is to be specified if indicated. RMPNI is transmitted using Table U.2. If RMPNI indicates the presence of an ADPN or LPIR field, an additional RMPNI field immediately follows the ADPN or LPIR field.

Table U.2/H.263 – RMPNI operations for re-mapping of reference pictures

Value	Re-mapping specified
"1"	ADPN field is present and corresponds to a negative difference to add to a picture number prediction value
"010"	ADPN field is present and corresponds to a positive difference to add to a picture number prediction value
"011"	LPIR field is present and specifies the long-term index for a reference picture
"001"	End loop for re-mapping of picture relative indexing default order

A picture reference parameter is a relative index into the ordered set of pictures. The RMPNI, ADPN, and LPIR fields allow the order of that relative indexing into the multi-picture buffer to be temporarily altered from the default index order for the decoding of a particular picture or video picture segment. The default index order is for the short-term pictures (i.e., pictures which have not been given a long-term index) to precede the long-term pictures in the reference indexing order. Within the set of short-term pictures, the default order is for the pictures to be ordered starting with the most recent buffered reference picture and proceeding through to the oldest reference picture (i.e., in decreasing order of picture number in the absence of wrapping of the ten-bit picture number field). Within the set of long-term pictures, the default order is for the pictures to be ordered starting with the picture with the smallest long-term index and proceeding up to the picture with long-term index equal to the most recent value of MLIP1 – 1.

For example, if the buffer contains three short-term pictures with short-term picture numbers 300, 302, and 303 (which were transmitted in increasing picture-number order) and two long-term pictures with long-term picture indices 0 and 3, the default index order is:

- default relative index 0 refers to the short-term picture with picture number 303;
- default relative index 1 refers to the short-term picture with picture number 302;
- default relative index 2 refers to the short-term picture with picture number 300;
- default relative index 3 refers to the long-term picture with long-term picture index 0; and
- default relative index 4 refers to the long-term picture with long-term picture index 3.

The first ADPN or LPIR field that is received (if any) moves a specified picture out of the default order to the relative index of zero. The second such field moves a specified picture to the relative

index of one, etc. The set of remaining pictures not moved to the front of the relative indexing order in this manner shall retain their default order amongst themselves and shall follow the pictures that have been moved to the front of the buffer in relative indexing order.

If MRPA is "0", no more than one ADPN or LPIR field shall be present in the same ERPS layer unless the current picture is a B picture. If the current picture is a B picture and MRPA is "0", no more than two ADPN or LPIR fields shall be present in the same ERPS layer.

Any re-mapping of picture numbers specified for some picture shall not affect the decoding process for any other picture. Any re-mapping of picture numbers specified for some video picture segment shall not affect the decoding process for any other video picture segment. A re-mapping of picture numbers specified for a picture shall only affect the decoding process for any video picture segment within that picture in two ways:

- If NOERPSL is "1" at the GOB or slice level, then the re-mapping specified at the picture level is also used for the associated video picture segment.
- If the picture is a B picture, the re-mapping specified at the picture level shall specify the calculation of the value of TR_B and TR_D for direct bidirectional prediction.

An RMPNI "end loop" indication is the last element of the ERPS layer for a B picture if MRPA is "0". In a B picture with MRPA equal to "1", an RMPNI "end loop" indication is followed by BTPSM. In a P or EP picture or Improved PB frame, an RMPNI "end loop" indication is followed by RPBT.

Within one ERPS layer, RMPNI shall not specify the placement of any individual reference picture into more than one re-mapped position in relative index order.

U.3.1.5.3 Absolute Difference of Picture Numbers (ADPN) (variable length)

ADPN is a variable-length codeword that is present only if indicated by RMPNI. ADPN follows RMPNI when present. ADPN is transmitted using Table U.1, where the index into the table corresponds to $ADPN - 1$. ADPN represents the absolute difference between the picture number of the currently re-mapped picture and the prediction value for that picture number. If no previous ADPN fields have been sent within the current ERPS layer, the prediction value shall be the picture number of the current picture. If some previous ADPN field has been sent, the prediction value shall be the picture number of the last picture that was re-mapped using ADPN.

If the picture number prediction is denoted PNP, and the picture number in question is denoted PNQ, the decoder shall determine PNQ from PNP and ADPN in a manner mathematically equivalent to the following:

```
if (RMPNI == "1") { // a negative difference
    if (PNP - ADPN < 0)
        PNQ = PNP - ADPN + 1024;
    else
        PNQ = PNP - ADPN;
} else { // a positive difference
    if (PNP + ADPN > 1023)
        PNQ = PNP + ADPN - 1024;
    else
        PNQ = PNP + ADPN;
}
```

The encoder shall control RMPNI and ADPN such that the decoded value of ADPN shall not be greater than or equal to 1024.

As an example implementation, the encoder may use the following process to determine values of ADPN and RMPNI to specify a re-mapped picture number in question, PNQ:

```
DELTA = PNQ - PNP;  
if (DELTA < 0) {  
    if (DELTA < -511)  
        MDELTA = DELTA + 1024;  
    else  
        MDELTA = DELTA;  
}else{  
    if (DELTA > 512)  
        MDELTA = DELTA - 1024;  
    else  
        MDELTA = DELTA;  
}
```

```
ADPN = abs(MDELTA);
```

where abs() indicates an absolute value operation. Note that the index into Table U.1 corresponds to the value of ADPN – 1, rather than the value of ADPN itself.

RMPNI would then be determined by the sign of MDELTA.

U.3.1.5.4 Long-term Picture Index for Re-Mapping (LPIR) (variable length)

LPIR is a variable-length codeword that is present only if indicated by RMPNI. LPIR follows RMPNI when present. LPIR is transmitted using Table U.1. It represents the long-term picture index to be re-mapped. The prediction value used by any subsequent ADPN re-mappings is not affected by LPIR.

U.3.1.5.5 B-picture Two-Picture Prediction Sub-Mode (BTPSM) (1 bit)

BTPSM is a single-bit fixed-length codeword that is present only in a B picture (see Annex O) and only when MRPA is "1". It follows an RMPNI "end loop" indication and is the last element of the ERPS layer for the B picture when present. It indicates whether the two-picture backward-prediction sub-mode is in use for the picture as follows:

"0": Single-picture backward prediction.

"1": Two-picture backward prediction.

BTPSM has an implied value of "0" if not present (when MRPA is "0").

The set of pictures available for use as forward-prediction references is the set of pictures in the multi-picture buffer other than the set of backward reference pictures. The set of backward reference pictures is determined by the value of BTPSM. If single-picture backward prediction is specified by BTPSM, the first picture in (possibly re-mapped) relative index order is the only backward reference picture. If two-picture backward prediction is specified by BTPSM, the first two pictures in (possibly re-mapped) relative index order are the two backward reference pictures. The relative index for forward prediction then becomes a relative index into the set of forward reference pictures.

The contents of the multi-picture buffer are not affected by the presence of a B picture. The B picture is not stored in the multi-picture buffer and is not used as a reference for the coding of subsequent pictures.

U.3.1.5.6 Reference Picture Buffering Type (RPBT) (1 bit)

RPBT is a single-bit fixed-length codeword that specifies the buffering type of the currently decoded picture. It follows an RMPNI "end loop" indication when the picture is not an I, EI, or B picture. It is the first element of the ERPS layer if the picture is an I or EI picture. It is not present if the picture is a B picture. The values for RPBT are defined as follows:

"1": Sliding Window.

"0": Adaptive Memory Control.

In the "Sliding Window" buffering type, the current decoded picture shall be added to the buffer with index 0, and any marking of pictures as "unused" in the buffer is performed automatically in a first-in-first-out fashion among the set of short-term pictures. In this case, if the buffer has sufficient "unused" capacity to store the current picture, no additional pictures shall be marked as "unused" in the buffer. If the buffer does not have sufficient "unused" capacity to store the current picture, the picture (or pictures as necessary to free the needed amount of memory in the case of sub-picture removal) with the largest default index (or indices as necessary in the case of sub-picture removal) among the short-term pictures in the buffer shall be marked as "unused". In the "Sliding Window" buffering type, no additional information is transmitted to control the buffer contents.

In the "Adaptive Memory Control" buffering type, the encoder explicitly specifies any addition to the buffer or marking of data as "unused" in the buffer, and may also assign long-term indices to short-term pictures. The current picture and other pictures may be explicitly marked as "unused" in the buffer, as specified by the encoder. This buffering type requires further information that is controlled by memory management control operation (MMCO) parameters.

RPBT, if present in GOB or slice layers, shall be the same as in the picture layer. Any MMCO command present in GOB or slice layers shall convey the same operation as some MMCO command in the picture layer.

If the picture is a B picture, RPBT shall not be present and the decoded picture shall not be stored in the multi-picture buffer. This ensures that a B picture shall not affect the contents of the multi-picture buffer.

Similarly, the B-picture part of an Improved PB frame shall not be stored in the buffer. All control fields associated with controlling the storage of an Improved PB frame shall be considered to be associated with controlling the storage of only the P-picture part of the Improved PB frame.

U.3.1.5.7 Memory Management Control Operation (MMCO) (variable length)

MMCO is a variable-length codeword that is present only when RPBT indicates "Adaptive Memory Control", and may occur multiple times if present. It specifies a control operation to be applied to manage the multi-picture buffer memory. The MMCO parameter is followed by data necessary for the operation specified by the value of MMCO, and then an additional MMCO parameter follows – until the MMCO value indicates the end of the list of such operations. MMCO commands do not affect the buffer contents or the decoding process for the decoding of the current picture – rather, they specify the necessary buffer status for the decoding of subsequent pictures in the bitstream. The values and control operations associated with MMCO are defined in Table U.3.

Table U.3/H.263 – Memory Management Control Operation (MMCO) values

Value	Memory Management Control Operation	Associated data fields following
"1"	End MMCO loop	None (end of ERPS layer)
"011"	Mark a short-term picture as "Unused"	DPN
"0100"	Mark a long-term picture as "Unused"	LPIN
"0101"	Assign a long-term index to a picture	DPN and LPIN
"00100"	Mark short-term sub-picture areas as "Unused"	DPN and SPRB
"00101"	Mark long-term sub-picture areas as "Unused"	LPIN and SPRB
"00110"	Specify the maximum long-term picture index	MLIP1
"00111"	Specify the buffer size and structure	SPWI, SPHI, SPTN, and RESET

All memory management control operations specified using MMCO shall be specified in the picture layer. Some or all of the same operations as are specified at the picture layer may also be specified at the GOB or slice layer (with the same associated data). MMCO shall not specify memory operations at the GOB or slice layer that are not also specified with the same associated data at the picture layer.

A buffer size and structure specification MMCO command shall be the first MMCO command if present. No more than one buffer size and structure specification MMCO command shall be present in a given ERPS layer. A buffer size and structure specification MMCO command with RESET equal to "1" shall be present in the first picture in which the ERPS mode is activated in any series of ERPS mode pictures in the bitstream. A buffer size and structure specification MMCO command with RESET equal to "1" shall precede any use of MMCO to indicate marking sub-picture areas of any short-term or long-term pictures as "unused". The sub-picture width and height specified in a buffer size and structure specification MMCO command shall not differ from the value of these parameters in a prior buffer size and structure specification MMCO command unless the current picture is an I or EI picture with RESET equal to "1". The picture height and width shall not change within the bitstream except within a picture containing a buffer size and structure specification MMCO command with RESET equal to "1" (or within a picture in which the ERPS mode is not in use).

If a B picture using single-picture backward prediction is present in the bitstream, exactly one temporally subsequent non-B picture in the reference layer of the B picture shall precede the B picture in bitstream order, as specified in O.2. No memory management control operations shall be present within any ERPS layer of this immediately temporally subsequent non-B picture within the reference layer of the B picture which mark any part of that immediately temporally succeeding non-B picture as "unused", since that reference layer picture is needed for display until after the decoding of the B picture.

The transmission order constraints specified in O.2 are adjusted as necessary for B pictures using two-picture backward prediction. If a B picture using two-picture backward prediction is present in the bitstream, exactly two temporally subsequent non-B pictures in the reference layer of the B picture shall precede the B picture in bitstream order. The other restrictions on the transmission order of the B picture in the bitstream specified in O.2 shall apply, but as adjusted for the use of two temporally subsequent reference layer pictures. No memory management control operations shall be present within any ERPS layer of these two immediately temporally subsequent two non-B pictures within the reference layer of the B picture which mark any part of these two non-B pictures as "unused", since these reference layer pictures are needed for display until after the decoding of the B picture.

A "stored picture" is defined as a non-B picture which does not contain an MMCO command in its ERPS layer which marks that picture as "unused". If the current picture is not a stored picture, its ERPS layer shall not contain any of the following types of MMCO commands:

- an MMCO command to specify the buffer size and structure with RESET equal to "1";
- any MMCO command which marks any other picture as "unused" that has not also been marked as "unused" in the ERPS layer of a prior stored picture;
- any MMCO command which assigns a long-term index to a picture that has not also been assigned the same long-term index in the ERPS layer of a prior stored picture; or
- any MMCO command which marks sub-picture areas of any picture as "unused" that have not also been marked as "unused" in the ERPS layer of a prior stored picture.

U.3.1.5.8 Difference of Picture Numbers (DPN) (variable length)

DPN is present when indicated by MMCO. DPN follows MMCO if present. DPN is transmitted using codewords in Table U.1 and is used to calculate the PN of a picture for a memory control operation. It is used in order to assign a long-term index to a picture, mark a short-term picture as "unused", or mark sub-picture areas of a short-term picture as "unused". If the current decoded picture number is PNC and the decoded value from Table U.1 is DPN, an operation mathematically equivalent to the following equations shall be used for calculation of PNQ, the specified picture number in question:

```
if (PNC - DPN < 0)
    PNQ = PNC - DPN + 1024;
else
    PNQ = PNC - DPN;
```

Similarly, the encoder may compute the DPN value to encode using the following relation:

```
if (PNC - PNQ < 0)
    DPN = PNC - PNQ + 1024;
else
    DPN = PNC - PNQ;
```

For example, if the decoded value of DPN is zero and MMCO indicates marking a short-term picture as "unused", the current decoded picture shall be marked as "unused".

U.3.1.5.9 Long-term Picture Index (LPIN) (variable length)

LPIN is present when indicated by MMCO. LPIN is transmitted using codewords in Table U.1 and specifies the long-term picture index of a picture. It follows DPN if the operation is to assign a long-term index to a picture. It follows MMCO if the operation is to mark a long-term picture as "unused" or to mark sub-picture areas of a long-term picture as "unused".

U.3.1.5.10 Sub-Picture Removal Bit-map (SPRB) (fixed length)

SPRB is a fixed-length codeword that contains one bit for each sub-picture and is present when indicated by MMCO. The number of bits of SPRB data is determined by the most recent values of SPWI and SPHI. SPRB is used to indicate which sub-picture areas of a buffered picture are to be marked as "unused". SPRB follows DPN if the operation is to mark sub-picture areas of a short-term picture as "unused", and follows LPIN if the operation is to mark sub-picture areas of a long-term picture as "unused".

Sub-pictures are numbered in raster scan order starting from the upper-left corner of the picture. For example, consider a case in which a reference picture, specified by DPN, is partitioned into six sub-pictures. Let "s₁ s₂ s₃ s₄ s₅ s₆" represent six bits of SPRB data. If bit s_i is "1", then the decoder should mark the ith sub-picture in the indicated reference picture as "unused". For example, if the SPRB is "000110", then the fourth and fifth sub-pictures areas are marked as "unused".

To prevent start code emulation, all necessary SPREPB emulation prevention bits shall be inserted within or following the SPRB data as specified in U.3.1.5.11.

If SPRB is present and the specified picture has been previously been affected by a prior SPRB bit-map, the bit-map specified by SPRB shall contain a "1" for any sub-picture area that contained a "1" in the previous SPRB bit-map. Every SPRB bit-map shall contain at least one bit having the value "0" and at least one bit having the value "1".

U.3.1.5.11 Sub-Picture Removal Emulation Prevention Bit (SPREPB) (one bit)

SPREPB is a single-bit fixed-length codeword having the value "1" which shall be inserted immediately after any string of 8 consecutive zero bits of SPRB data.

U.3.1.5.12 Maximum Long-term Picture Index Plus 1 (MLIP1) (variable length)

MLIP1 is a variable-length codeword that is present if indicated by MMCO. MLIP1 follows MMCO if present. MLIP1 is transmitted using codewords in Table U.1. If present, MLIP1 is used to determine the maximum index allowed for long-term reference pictures (until receipt of another value of MLIP1). The decoder shall initially assume MLIP1 is "0" until some other value has been received. Upon receiving an MLIP1 parameter, the decoder shall consider all long-term pictures having indices greater than the decoded value of MLIP1 – 1 as "unused" for referencing by the decoding process for subsequent pictures. For all other pictures in the multi-picture buffer, no change of status shall be indicated by MLIP1.

U.3.1.5.13 Sub-Picture Width Indication (SPWI) (7 bits)

SPWI is a fixed-length codeword of 7 bits that is present if indicated by MMCO. SPWI follows MMCO when indicated. SPWI specifies the width of a sub-picture in units of 16 luminance samples, such that the indicated sub-picture width is $16 \cdot (\text{SPWI} + 1)$ luminance samples. The current picture has a width in sub-picture units of $\text{ceil}(\text{ceil}(\text{pw}/16)/(\text{SPWI} + 1))$ sub-pictures, where pw is the width of the picture and "/" indicates floating-point division. For positive numbers, the ceiling function, $\text{ceil}(x)$, equals x if x is an integer and otherwise $\text{ceil}(x)$ equals one plus the integer part of x . If a minimum picture unit (MPU) size defining the minimum width and height of a sub-picture has been negotiated by external means (for example, ITU-T Rec. H.245), the sub-picture width specified by SPWI shall be an integer multiple of the width of the MPU; otherwise, the sub-picture width specified by SPWI shall be such that SPWI is equal to $\text{ceil}(\text{pw}/16) - 1$.

U.3.1.5.14 Sub-Picture Height Indication (SPHI) (7 bits)

SPHI is a fixed-length codeword of 7 bits that is present if SPWI is present (as indicated by MMCO). SPHI follows SPWI if present. SPHI specifies the height of a sub-picture in units of 16 luminance samples, such that the indicated sub-picture height is $16 \cdot \text{SPHI}$. The allowed range of values of SPHI is from 1 to 72. The current picture has a height of $\text{ceil}(\text{ceil}(\text{ph}/16)/\text{SPHI})$ sub-pictures, where ph is the height of the picture and "/" indicates floating-point division. If a minimum picture unit (MPU) size defining the minimum width and height of a sub-picture has been negotiated by external means (for example, ITU-T Rec. H.245), the sub-picture height specified by SPHI shall be an integer multiple of the height of the MPU; otherwise, the sub-picture height specified by SPHI shall be such that SPHI is equal to $\text{ceil}(\text{ph}/16)$.

U.3.1.5.15 Sub-Picture Total Number (SPTN) (variable length)

SPTN is a variable-length codeword that is present if SPWI and SPHI are present (as indicated by MMCO). SPTN follows SPHI if present. SPTN is coded using Table U.1, where the index into Table U.1 corresponds to the decoded value of SPTN – 1. The decoded value of SPTN is the total operational size capacity of the multi-picture buffer in units of sub-pictures as specified by SPWI and SPHI. The memory capacity needed for the decoding of current pictures is not included in SPTN – only the memory capacity needed for storing the reference pictures to use for the prediction of other pictures. When sub-picture removal is not in use (i.e., when SPWI and SPHI have whole-picture dimensions), the maximum number of active short-term reference pictures (for example, for sliding window operation) is thus given by SPTN minus the number of pictures that have been assigned to long-term indices and have not been subsequently marked as "unused".

U.3.1.5.16 Buffer Reset Indicator (RESET) (1 bit)

RESET is a single-bit fixed-length codeword that is present if SPWI, SPHI, and SPTN are present (as indicated by MMCO). RESET follows SPTN if present. The values of RESET shall be as follows:

- "0": The buffer contents are not reset.
- "1": The buffer contents are reset.

If RESET is "1", all pictures in the multi-picture buffer (but not the current picture unless specified separately) shall be marked "unused" (including both short-term and long-term pictures).

U.3.2 Macroblock layer syntax

U.3.2.1 P-picture and Improved PB frames macroblock syntax

The macroblock layer syntax is modified if the ERPS layer is present for P pictures and Improved PB frames when the number of selected forward reference pictures may be greater than one, as indicated by MRPA. The field MRPA is signalled in the ERPS layer. The macroblock layer syntax is shown in Figure U.6 when MRPA is "1". Otherwise, the macroblock syntax format in a P picture or Improved PB frame is not altered from that shown in Figure 10.

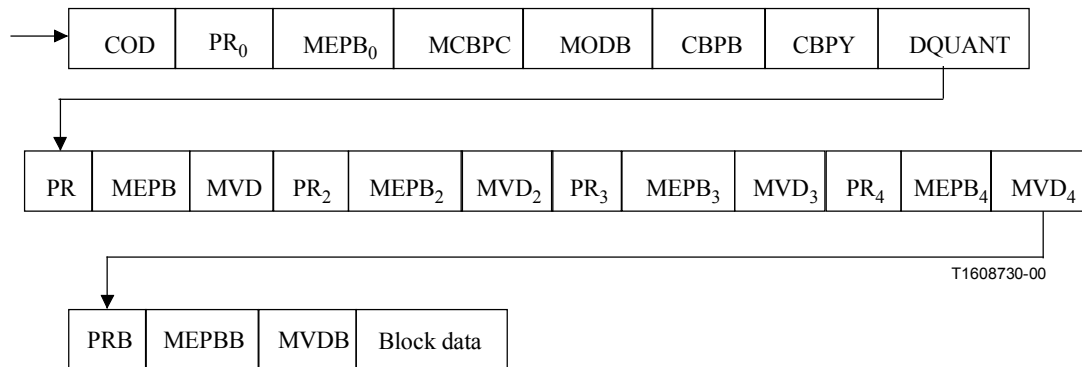


Figure U.6/H.263 – Structure of P-picture and Improved PB frame macroblock layer for the ERPS mode

U.3.2.1.1 Interpretation of COD

If the COD bit is "1", no further information is transmitted for the macroblock. In that case, the decoder shall treat the macroblock as an INTER macroblock with the motion vector for the entire macroblock equal to zero, picture reference parameter equal to zero, and with no coefficient data. If the COD bit is "0", indicating that the macroblock is coded, the syntax of the macroblock layer is depicted in Figure U.6 with the fields PR₀, PR, PR₂, PR₃, PR₄, and PRB being included in the syntax. PR₀, PR, PR₂, PR₃, PR₄, and PRB each consist of a variable length codeword as given in Table U.1.

U.3.2.1.2 Picture Reference Parameter 0 (PR₀) (variable length)

PR₀ is a variable length codeword as specified in Table U.1. It is present whenever COD is "0". If PR₀ has a decoded value of zero (codeword "1"), it indicates that further information will follow for the macroblock. If decoded as non-zero, it indicates the coding of the macroblock using only a picture reference parameter.

If the field PR₀ does not have a decoded value of zero (codeword "1"), no further information is transmitted for this macroblock. In that case the decoder shall treat the macroblock as an INTER macroblock with the motion vector for the whole block equal to zero, the picture reference parameter equal to PR₀, and with no coefficient data.

If the field PR₀ has a decoded value of zero (codeword "1"), the macroblock is coded. The meaning and usage of the fields MCBPC, CBPB, CBPY, and DQUANT remains unaltered. The field PR is included together with the field MVD for all INTER macroblocks (and in Improved PB frames mode also for INTRA macroblocks). The use of MODB in Improved PB frames is described in U.3.2.1.4.

U.3.2.1.3 Macroblock Emulation Prevention Bit 0 (MEPB₀) (1 bit)

MEPB₀ is a single-bit fixed-length codeword having the value "1" that follows PR₀ if, and only if, PR₀ is present and has a decoded value of "1" (codeword "000"), and either of the following two conditions are satisfied:

- 1) the slice structured mode (see Annex K) is in use; or
- 2) the COD for the current macroblock immediately follows after another macroblock which also has COD = "0" and PR₀ = "1" (codeword "000"), and the PR₀ of the previous macroblock is not followed by an MEPB₀ bit.

The purpose of MEPB₀ is to prevent start-code emulation and, in the slice structured mode, to aid in determining the number of macroblocks in a slice.

U.3.2.1.4 Macroblock Picture Reference parameters (PR, PR₂₋₄, and PRB) (variable length)

PR is the primary picture reference parameter. PR is present whenever MVD is present. The three codewords PR₂₋₄ are included together with MVD₂₋₄ if indicated by PTYPE and if MCBPC specifies an INTER4V or INTER4V+Q macroblock (a macroblock of type 2 or 5 in Tables 8 and 9). PR₂₋₄ and MVD₂₋₄ are only present when in Advanced Prediction mode (see Annex F) or Deblocking Filter mode (see Annex J). PRB is only present in an Improved PB frame when MODB indicates that MVDB is present. PR, PR₂₋₄, and PRB each specify a picture reference relative index into the multi-picture buffer.

PR is used as the picture reference parameter for motion compensation of the entire macroblock if the macroblock is not an INTER4V or INTER4V+Q macroblock. If the macroblock is an INTER4V or INTER4V+Q macroblock, PR is used for motion compensated prediction of the first of the four 8 × 8 luminance blocks in the macroblock and for the two chrominance blocks of the macroblock (with the motion compensation process otherwise as specified in 6.1). PR₂₋₄ are used for motion compensation of the remaining three 8 × 8 blocks of luminance data in the macroblock. If MODB indicates that MVDB is present, PRB is the picture reference parameter for forward prediction of the B part of the Improved PB frame.

In Improved PB frames when MODB indicates B_{PB} bidirectional prediction, the values of TR_D and TR_B shall be computed as the temporal reference increments based on the temporal reference data of the current picture and that of the most recent previous reference picture, regardless of whether or not the most recent previous reference picture has been re-mapped to a difference relative index order, marked as "unused", or assigned to a long-term index. The picture used as the forward reference picture for B_{PB} bidirectional prediction in Improved PB frames shall be the picture specified by PR.

U.3.2.1.5 Macroblock Emulation Prevention Bits (MEPB, MEPB₂₋₄ and MEPBB) (1 bit each)

MEPB, MEPB₂₋₄, and MEPBB are each a single bit having the value "1" if present. Each shall be present if, and only if, the Unrestricted Motion Vector mode (see Annex D) is not in use and the associated PR, PR₂₋₄, or PRB field is present and has the decoded value "1" (codeword "000"). Their purpose is to prevent start-code emulation.

U.3.2.2 B-picture and EP-picture macroblock syntax

The macroblock layer syntax for B and EP pictures (see Annex O) is modified in a similar fashion as in P pictures. The COD bit, if equal to "1", indicates a skipped macroblock as defined in Annex O, using a picture reference parameter of zero for the forward (skipped) prediction in an EP picture and for the forward part of direct (skipped) bidirectional prediction in a B picture and using the first backward prediction picture for the backward part of direct (skipped) bidirectional

prediction in a B picture (in the case of two-picture backward prediction, as when BSBBW is present and equal to "0"). If COD is "0", a PR_0 parameter is inserted into the syntax and is used in a similar manner as described in U.3.2.1.2. If PR_0 is present and does not have a decoded value of zero (codeword "1"), it indicates that the macroblock is to be predicted with forward INTER prediction using a zero-valued motion vector and a picture reference parameter of PR_0 . If PR_0 has a decoded value of zero, MBTYPE follows and specifies the macroblock type. The format of the CBPC, CBPY, and DQUANT fields is unchanged. The MVDFW and MVDBW fields are encoded in the same manner as when the ERPS mode is not in use, but are each used in conjunction with a picture reference, and possibly an emulation prevention bit.

For a B picture, the backward reference pictures in the multi-picture buffer are defined as follows:

- in the case of single-picture backward prediction, there is only one backward reference picture, which is the first picture in (possibly re-mapped) relative index order; and
- in the case of two-picture backward prediction, there are two backward reference pictures, which are the first two pictures in (possibly re-mapped) relative index order.

The forward reference pictures in the multi-picture buffer are defined as the pictures in the multi-picture buffer other than the backward reference pictures. The relative indexing for forward prediction is a relative index into the forward reference picture set, and the relative indexing for backward prediction is a relative index into the backward reference picture set.

For example, if the buffer contains three short-term pictures with short-term picture numbers 300, 302, and 303 (which were transmitted in increasing picture-number order) and two long-term pictures with long-term picture indices 0 and 3, the default index order in the case of two-picture backward prediction is:

- default backward relative index 0 refers to the short-term picture with picture number 303;
- default backward relative index 1 refers to the short-term picture with picture number 302;
- default forward relative index 0 refers to the short-term picture with picture number 300;
- default forward relative index 1 refers to the long-term picture with long-term picture index 0; and
- default forward relative index 2 refers to the long-term picture with long-term picture index 3;

and in the case of single-picture backward prediction:

- the single default backward reference picture is the short-term picture with picture number 303;
- default forward relative index 0 refers to the short-term picture with picture number 302;
- default forward relative index 1 refers to the short-term picture with picture number 300;
- default forward relative index 2 refers to the long-term picture with long-term picture index 0; and
- default forward relative index 3 refers to the long-term picture with long-term picture index 3;

and if these pictures have been re-mapped to a new relative indexing order of short-term picture 302, followed by short-term picture 303, followed by long-term picture 0, followed by short-term picture 300, followed by long-term picture 3, the new relative index order in the case of two-picture backward prediction is:

- re-mapped backward relative index 0 refers to the short-term picture with picture number 302;
- re-mapped backward relative index 1 refers to the short-term picture with picture number 303;

- re-mapped forward relative index 0 refers to the long-term picture with long-term picture index 0;
- re-mapped forward relative index 1 refers to the short-term picture with picture number 300; and
- re-mapped forward relative index 2 refers to the long-term picture with long-term picture index 3;

and in the case of single-picture backward prediction:

- the single re-mapped backward reference picture is the short-term picture with picture number 302;
- re-mapped forward relative index 0 refers to the short-term picture with picture number 303;
- re-mapped forward relative index 1 refers to the long-term picture with long-term picture index 0;
- re-mapped forward relative index 2 refers to the short-term picture with picture number 300; and
- re-mapped forward relative index 3 refers to the long-term picture with long-term picture index 3.

The TR_D used for direct bidirectional prediction in a B picture shall be computed as the temporal reference increment between the first forward reference picture in (possibly re-mapped) relative index order and the first backward reference picture in (possibly re-mapped) relative index order (i.e., if two-picture backward prediction is in use, this would be the picture referenced when BSBBW is "0" as described in U.3.2.2.3). The TR_B used for direct bidirectional prediction in a B picture shall be computed as the temporal reference increment between the B picture and the first forward reference picture in (possibly re-mapped) relative index order. The relative index order used in the computation of TR_D and TR_B shall be that specified by the ERPS layer at the picture level of the B picture syntax (i.e., re-mappings at the GOB or slice level shall not affect the values of TR_D and TR_B). (See Figure U.7.)

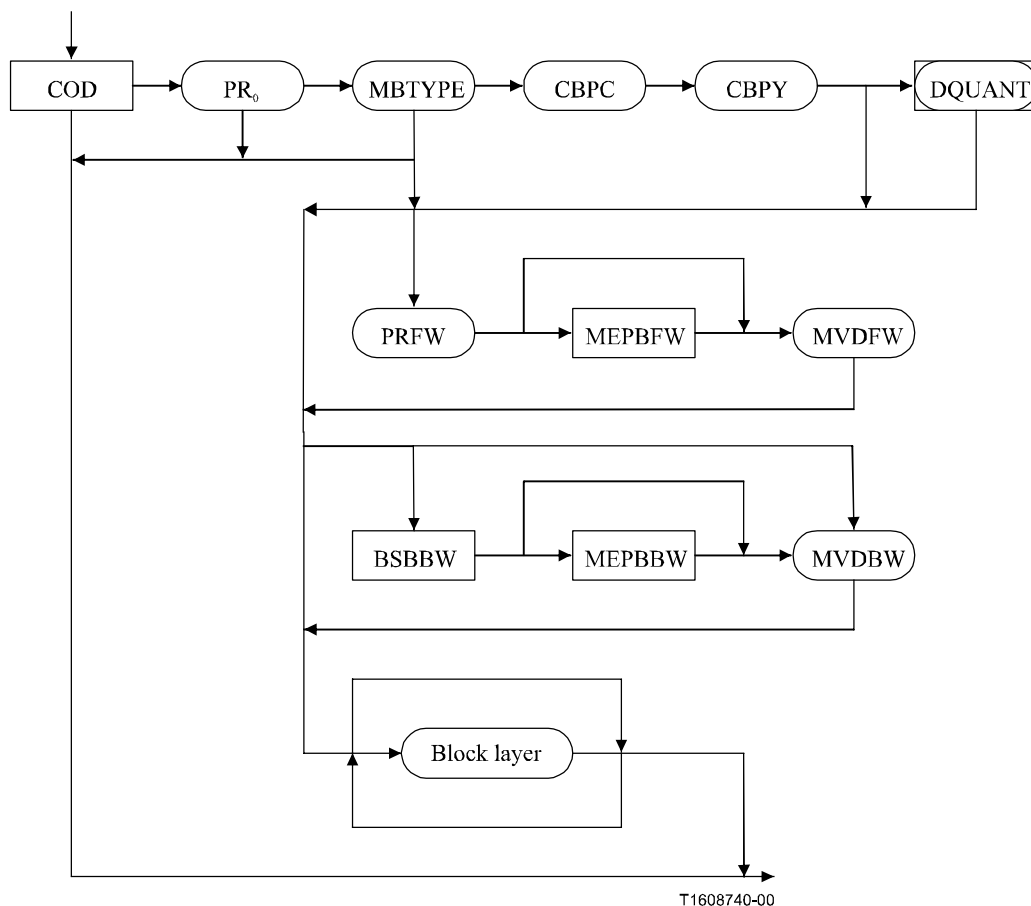


Figure U.7/H.263 – Structure of EP- and B-picture macroblock layer for the ERPS mode

U.3.2.2.1 Picture Reference for Forward Prediction (PRFW) (variable length)

PRFW is a variable-length picture reference parameter that is present whenever forward motion vector data is present, and is encoded using Table U.1. PRFW is a relative index into the set of forward reference pictures.

U.3.2.2.2 Emulation Prevention Bit for Forward Prediction (MEPBFW) (1 bit)

MEPBFW is a single-bit fixed-length codeword having the value "1" which shall be inserted after PRFW if and only if PRFW is present and has a decoded value of "1" (codeword "000") and the unrestricted motion vector mode (see Annex D) is not in use.

U.3.2.2.3 B-picture Selection Bit for Backward Prediction (BSBBW) (1 bit)

BSBBW is a single-bit fixed-length codeword that is present only for B pictures when MVDBW is present and only when two-picture backward prediction is specified for the B-picture operation. The meaning of this bit shall be defined as:

- "0": Prediction from the first backward reference picture in relative index order (in default order, this would be the most recent short-term reference picture if that picture has not been assigned a long-term index or marked as "unused").
- "1": Prediction from the second backward reference picture in relative index order (in default order, this would be the second-most recent short-term reference picture if neither of the last two reference pictures has been assigned a long-term index or marked as "unused").

U.3.2.2.4 Emulation Prevention Bit for Backward Prediction (MEPBBW) (1 bit)

MEPBBW is a single-bit fixed-length codeword having the value "1" that is present only under the following conditions:

- BSBBW is present and equal to "0"; and
- the unrestricted motion vector mode (see Annex D) is not in use; and
- BSBBW is preceded by five bits having the value "00000".

U.4 Decoder process

The decoder for the ERPS mode stores the reference pictures for inter-picture decoding in a multi-picture buffer. The decoder may need additional memory capacity to store the multiple decoded pictures (relative to the memory capacity needed without support of the ERPS mode). The decoder replicates the multi-picture buffer of the encoder according to the reference picture buffering type and any memory management control operations specified in the bitstream. The buffering scheme may also be operated when partially erroneous pictures are decoded.

Each transmitted and stored picture is assigned a Picture Number (PN) which is stored with the picture in the multi-picture buffer. PN represents a sequential picture counting identifier for stored pictures. PN is constrained, using modulo 1024 arithmetic operation. For the first transmitted picture, PN should be "0". For each and every other transmitted and stored picture, PN shall be increased by 1 (within a given scalability layer, if Annex O is in use). If the difference (modulo 1024) of the PNs of two consecutively received and stored pictures is not 1, the decoder should infer a loss of pictures or corruption of data. In such a case, a back-channel message indicating the loss of pictures may be sent to the encoder.

Besides the PN, each picture stored in the multi-picture buffer has an associated index, called the default index. When a picture is first added to the multi-picture buffer, it is given default index 0 – unless it is assigned to a long-term index. The indices of pictures in the multi-picture buffer are modified when pictures are added to or removed from the multi-picture buffer.

The pictures stored in the multi-picture buffers can also be divided into two categories: long-term pictures and short-term pictures. A long-term picture can stay in the multi-picture buffer for a long time (more than 1023 coded and stored picture intervals). The current picture is initially considered a short-term picture. Any short-term picture can be changed to a long-term picture by assigning it a long-term index according to information in the bitstream. The PN is the unique ID for all short-term pictures in the multi-picture buffer. When a short-term picture is changed to a long-term picture, it is also assigned a long-term picture index (LPIN). A long-term picture index is assigned to a picture by associating its PN to an LPIN. Once a long-term picture index has been assigned to a picture, the only potential subsequent use of the long-term picture's PN within the bitstream shall be in a repetition of the long-term index assignment. The PNs of the long-term pictures are unique within 1024 transmitted and stored pictures. Therefore, the PN of a long-term picture cannot be used for assignment of a long-term index after 1023 transmitted subsequent stored pictures. LPIN becomes the unique ID for the life of a long-term picture.

PN (for a short-term picture) or LPIN (for a long-term picture) can be used to re-map the pictures into re-mapped indices for efficient reference picture addressing.

U.4.1 Decoder process for short-/long-term picture management

The decoder may have both long-term pictures and short-term pictures in its multi-picture buffer. The MLIP1 field is used to indicate the maximum long-term picture index allowed in the buffer. If no prior value of MLIP1 has been sent, no long-term pictures shall be in use, i.e., MLIP1 shall initially have an implied value of "0" upon invocation of the ERPS mode. Upon receiving an MLIP1 parameter, a new MLIP1 shall take effect until another value of MLIP1 is received. Upon receiving a new MLIP1 parameter in the bitstream, all long-term pictures with associated long-term indices greater than or equal to MLIP1 shall be considered marked "unused". The frequency of transmitting MLIP1 is out of the scope of this Recommendation. However, the encoder should send an MLIP1 parameter upon receiving an error message, such as an INTRA request message.

A short-term picture can be changed to a long-term picture by using an MMCO command with an associated DPN and LPIN. The short-term picture number is derived from DPN and the long-term picture index is LPIN. Upon receiving such an MMCO command, the decoder shall change the short-term picture with PN indicated by DPN to a long-term picture and shall assign it to the long-term index indicated by LPIN. If a long-term picture with the same long-term index already exists in the buffer, the previously-existing long-term picture shall be marked "unused". An encoder shall not assign a long-term index greater than $MLIP1 - 1$ to any picture. If LPIN is greater than $MLIP1 - 1$, this condition should be treated by the decoder as an error. For error resilience, the encoder may send the same long-term index assignment operation or MLIP1 specification message repeatedly. If the picture specified in a long-term assignment operation is already associated with the required LPIN, no action shall be taken by the decoder. An encoder shall not assign the same picture to more than one long-term index value. If the picture specified in a long-term index assignment operation is already associated with a different long-term index, this condition should be treated as an error. An encoder shall only change a short-term picture to a long-term picture within 1024 transmitted consecutive stored pictures. In other words, a short-term picture shall not stay in the short-term buffer after more than 1023 subsequent stored pictures have been transmitted. An encoder shall not assign a long-term index to a short-term picture that has been marked as "unused" by the decoding process prior to the first such assignment message in the bitstream. An encoder shall not assign a long-term index to a picture number that has not been sent.

U.4.2 Decoder process for reference picture buffer mapping

The decoder employs indices when referencing a picture for motion compensation on the macroblock layer using the fields PR_0 , PR, PR_2 , PR_3 , PR_4 , PRB, PRFW, and BSBBW. In pictures other than B pictures, these indices are the default relative indices of pictures in the multi-picture buffer when the fields ADPN and LPIR are not present in the current picture, GOB, or slice layer as applicable, and are re-mapped indices when these fields are present. In B pictures, the first one or two pictures (depending on BTPSM) in relative index order are used for backward prediction, and the forward picture reference parameters specify a relative index into the remaining pictures for use in forward prediction.

The indices of pictures in the multi-picture buffer can be re-mapped onto newly specified indices by transmitting the RMPNI, ADPN, and LPIR fields. RMPNI indicates whether ADPN or LPIR is present. If ADPN is present, RMPNI specifies the sign of the difference to be added to a picture number prediction value. The ADPN value corresponds to the absolute difference between the PN of the picture to be re-mapped and a prediction of that PN value. The first transmitted ADPN is computed as the absolute difference between the PN of the current picture and the PN of the picture to be re-mapped. The next transmitted ADPN field represents the difference between the PN of the previous picture that was re-mapped using ADPN and that of another picture to be re-mapped. The process continues until all necessary re-mapping is complete. The presence of re-mappings specified using LPIR does not affect the prediction value for subsequent re-mappings using ADPN. If RMPNI indicates the presence of an LPIR field, the re-mapped picture corresponds to a long-term picture with a long-term index of LPIR. If any pictures are not re-mapped to a specific order by

RMPNI, these remaining pictures shall follow after any pictures having a re-mapped order in the indexing scheme, following the default order amongst these non-re-mapped pictures.

If the decoder detects a missing picture, it may invoke some concealment process, and may insert an error-concealed picture into the multi-picture buffer. Missing pictures can be identified if one or several picture numbers are missing or if a picture not stored in the multi-picture buffer is indicated in a transmitted ADPN or LPIR. Concealment may be conducted by copying the closest temporally preceding picture that is available in the multi-picture buffer into the position of the missing picture. The temporal order of the short-term pictures in the multi-picture buffer can be inferred from their default index order and PN fields. In addition or instead, the decoder may send a forced INTRA update signal to the encoder by external means (for example, ITU-T Rec. H.245), or the decoder may use external means or back-channel messages (for example, ITU-T Rec. H.245) to indicate the loss of pictures to the encoder. A concealed picture may be inserted into the multi-picture buffer when using the "Sliding Window" buffering type. If a missing picture is detected when decoding a GOB or Slice layer, the concealment may be applied to the picture as if the missing picture had been detected at the picture layer.

U.4.3 Decoder process for sub-picture removal

Sub-picture removal may be used to reduce the amount of memory required to save multiple reference pictures. In sub-picture removal, each reference picture is partitioned into smaller equal-sized sub-pictures. The memory reduction is accomplished by marking undesired sub-pictures as "unused". The strategy used by the encoder to decide which of the sub-pictures to mark as "unused" is outside the scope of this Recommendation. The encoder signals to the decoder the size of the sub-pictures and which of the sub-pictures to mark as "unused" using MMCO commands in the enhanced reference picture selection (ERPS) layer. The encoder shall not send information in the bitstream that causes any samples in reference pictures or sub-pictures that it has caused to be marked as "unused" to be indicated for use in the prediction of subsequent pictures.

The sub-picture removal capability is negotiated by external means (for example, ITU-T Rec. H.245). In addition, the decoder signals, also by external means, the minimum partition unit (MPU) which is described in terms of a minimum width and height (in units of 16 luminance samples) of a sub-picture and the total amount of memory it has available for its multi-picture buffer. Memory management is facilitated by the partition rules described below.

Each reference picture is partitioned into rectangular sub-pictures of equal size. The encoder specifies the sub-picture size which shall be an integer multiple of the MPU. The width and height of the sub-picture shall be integer multiples of the minimum width and height negotiated externally as the MPU. The upper-left-hand corner of the first sub-picture is coincident with the upper-left-hand corner of the reference picture. Consequently, the entire partition may be described by specifying the width and height of a sub-picture. If the picture size is not an integer multiple of the sub-picture size, some sub-pictures may extend beyond the right and bottom boundaries of the reference picture. When a sub-picture that extends past the reference picture boundary is saved, a convenient memory management strategy is to set aside enough memory to save the entire sub-picture, rather than just the memory necessary to save the portion of the reference picture that lies within that sub-picture. This is the convention which shall be followed in any calculation of buffer spare capacity for the purpose of determining buffer fullness (e.g., in order to determine whether to automatically mark buffered pictures as "unused" in "sliding window" operation). A decoder designed such that each sub-picture occupies the same amount of memory will prevent the possibility of memory fragmentation.

An example method designed to access referenced picture samples when sub-picture removal is in use is described briefly as follows. One important element in any reference picture access technique is a mechanism to identify where the samples in each sub-picture are stored in memory. If there are R reference pictures and each picture is partitioned into S sub-pictures, then there are a total of $K = R \cdot S$ sub-pictures. For example, the sub-picture in the upper-left hand corner of the first

reference picture number can be considered sub-picture number 0, and the sub-picture to the right of it can be considered sub-picture number 1 and so on in raster scan order progressing from reference picture 1 to R until all K sub-pictures have a label. The total buffer capacity is SPTN sub-picture memory buffers, and SPTN is typically less than K. A K-element array can be defined, subPicMem[K], such that $t = \text{subPicMem}[k]$ corresponds to the sub-picture memory area that contains the samples in the kth sub-picture. For example, a case can be considered in which $R = 5$ reference pictures each have $S = 12$ sub-pictures. Then the samples for the 6th sub-picture in reference picture 3 would be found in sub-picture memory area $t = \text{subPicMem}[k]$ where $k = 3 \cdot S + 6 = 42$.

For example, when referencing samples for motion-compensated prediction of one block of luminance or chrominance data when the Advanced Prediction and Reduced Resolution Update optional modes are not in use, it is necessary to acquire $n \times m$ samples, where n and m may take values of 8 or 9 to accommodate half-integer motion compensation. Since the samples in one block may lie in up to four different sub-pictures, four separate cases must be considered. In all cases, the first step is to find the location in memory that contains the upper-left hand sample (U) of the block to be referenced. The sub-picture containing U can be identified by dividing the horizontal or vertical location of U by the sub-picture width or height. If U lies in sub-picture k, then that sample will be located in the subPicMem[k] sub-picture memory area. Next, if both the sample $m - 1$ samples to the right of U (i.e., the upper-right-hand corner of the block) and the sample $n - 1$ samples down from U (i.e., the lower-left-hand corner of the block) lie in sub-picture k, this can be considered case number one. If the sample $n - 1$ samples down from U lies within k, but the sample $m - 1$ samples to the right of U does not, this can be considered case two. If the sample $m - 1$ samples to the right of U lies within k, but the sample $n - 1$ down does not, this can be considered case three. Otherwise, when both the sample $m - 1$ samples to the right of U and the one $n - 1$ samples down lie outside of sub-picture k, this can be considered case four.

In case number one, all samples in the reference block are contained within the kth sub-picture. In this case, all relevant $n \times m$ samples may be found in sub-picture memory area subPicMem[k] and it is a simple task to access them. In case two, the samples that lie in the kth sub-picture can be obtained from sub-picture memory area subPicMem[k] and the remaining samples can be obtained from subPicMem[k_r] where k_r is the sub-picture to the right of k. In case three, the samples that lie in the kth sub-picture can be obtained from memory area subPicMem[k], and the remaining samples can be obtained from subPicMem[k_d] where k_d is the sub-picture below k. In case four, the samples that lie in the kth sub-picture can be obtained from sub-picture memory area subPicMem[k] and the remaining samples can be obtained from memory areas subPicMem[k_r], subPicMem[k_d] and subPicMem[k_{rd}] where k_r and k_d are defined above and k_{rd} is the sub-picture to the right and below k.

U.4.4 Decoder process for multi-picture motion compensation

Multi-picture motion compensation is applied if the MRPA field indicates the use of more than one reference picture. For multi-picture motion compensation, the decoder chooses a reference picture as indicated using the fields PR₀, PR, PR₂, PR₃, PR₄, PRB, PRFW and BSBBW on the macroblock layer. Once the reference picture is specified, the decoding process for motion compensation proceeds as described in 6.1.

In case four motion vectors per macroblock are used and the MRPA field indicates the use of more than one reference picture, the picture reference index for both chrominance blocks is that associated with the first of the four motion vectors (with the motion compensation process otherwise as specified by 6.1).

U.4.5 Decoder process for reference picture buffering

The buffering of the currently decoded picture can be specified using the reference picture buffering type (RPBT) for non-B pictures. The buffering may follow a first-in, first-out ("Sliding Window") mode. Alternatively, the buffering may follow a customized adaptive buffering ("Adaptive Memory Control") operation that is specified by the encoder in the forward channel. B pictures do not affect buffer contents.

The "Sliding Window" buffering type operates as follows. First, the decoder determine whether the picture can be stored into "unused" buffer capacity. If there is insufficient "unused" buffer capacity, the short-term picture with the largest default index (i.e., the oldest short-term picture in the buffer) shall be marked as "unused". This process is repeated if necessary (in the case of sub-picture removal) until sufficient memory capacity is freed to hold the current decoded picture. The current picture is stored in the buffer and assigned a default relative buffer index of zero. The default relative index of all other short-term pictures is incremented by one. The default relative index of all long-term pictures is incremented by one minus the number of short-term pictures removed.

In the "Adaptive Memory Control" buffering type, specified pictures or sub-picture areas may be removed from the multi-picture buffer explicitly. The currently decoded picture, which is initially considered a short-term picture, may be inserted into the buffer with default relative index 0, may be assigned to a long-term index, or may be marked as "unused" by the encoder. Other short-term pictures may also be assigned to long-term indices. The buffering process shall operate in a manner functionally equivalent to the following: First, the current picture is added to the multi-picture buffer with default relative index 0, and the default relative indices of all other pictures is incremented by one. Then, the MMCO commands are processed:

- If MMCO indicates a reset of the buffer contents by using RESET equal to "1", all pictures in the buffer are marked as "unused" except the current picture (which will be the picture with default relative index 0 since a buffer reset must be the first MMCO command as required by U.3.1.5.7).
- If MMCO indicates a maximum long-term index using MLIP1, all long-term pictures having long-term indices greater than or equal to MLIP1 are marked as "unused" and the default relative index order of the remaining pictures are not affected.
- If MMCO indicates that a picture is to be marked as "unused" in the multi-picture buffer and if that picture has not already been marked as "unused", the specified picture is marked as "unused" in the multi-picture buffer and the default relative index of all subsequent pictures in default order is decremented by one.
- If MMCO indicates that sub-picture areas of some picture are to be marked as "unused" in the multi-picture buffer, the specified sub-picture areas are marked as "unused" and the default relative index order of the pictures is not affected. As required by U.3.1.5.10, not all sub-picture areas of any given picture will be marked "unused" by a sub-picture removal MMCO command (instead, the encoder should send an MMCO command marking the picture as a whole as "unused").
- If MMCO indicates the assignment of a long-term index to a specified short-term picture and if the specified long-term index has not already been assigned to the specified short-term picture, the specified short-term picture is marked in the buffer as a long-term picture with the specified long-term index. If another picture is already present in the buffer with the same long-term index as the specified long-term index, the other picture is marked as "unused". All short-term pictures that were subsequent to the specified short-term picture in default relative index order and all long-term pictures having a long-term index less than the specified long-term index have their associated default relative indices decremented by one. The specified picture is assigned to a default relative index of one plus the highest of the decremented default relative indices, or zero if there are no such decremented indices.

The resulting buffered quantity of pictures or sub-picture regions not marked as "unused" shall not exceed the buffer capacity indicated by the most recent value of SPTN. If the decoder detects this condition, it should be treated as an error.

U.5 Back-channel messages

An out-of-band channel, which need not necessarily be reliable, can be used to convey back-channel messages. The syntax of this out-of-band channel (which could be a separate logical channel, for example using ITU-T Rec. H.223 or ITU-T Rec. H.225.0) should be the one defined herein. The "videomux" operation of back-channel messages as defined in Annex N is not supported in the ERPS mode.

U.5.1 BCM separate logical channel layer

The BCM layer as specified in U.5.2 should be carried by a BCM separate logical channel layer as shown in Figure U.8.

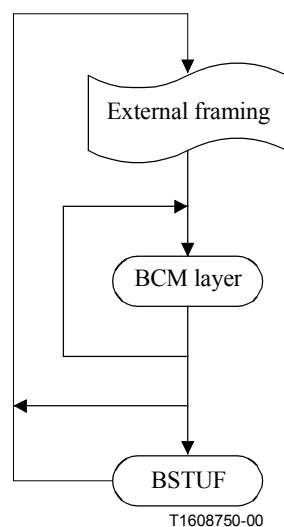


Figure U.8/H.263 – Structure of BCM separate logical channel layer for ERPS mode

U.5.1.1 External framing

External framing of back-channel messages should be provided as shown in Figure U.8. The external framing is used to determine the starting point for the back-channel messages and the amount of back-channel message data to follow.

U.5.1.2 Back-channel Stuffing (BSTUF) (variable length)

BSTUF is a variable-length codeword that may be present only after the last back-channel message in an external frame. BSTUF consists of a codeword of variable length consisting of one or more bits of value "0".

U.5.2 Back-channel message layer syntax

The syntax for the back-channel message (BCM) layer defined herein shall be as shown in Figure U.9.

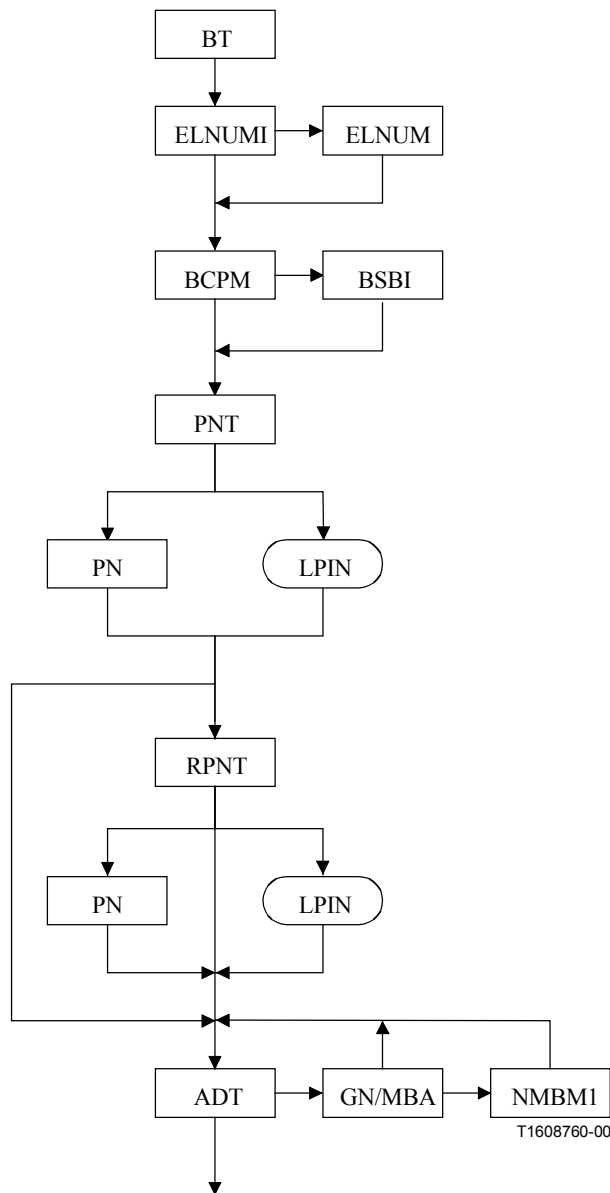


Figure U.9/H.263 – Structure of Back-Channel Message (BCM) layer for ERPS mode

U.5.2.1 Back-channel message Type (BT) (2 bits)

BT is a 2-bit fixed-length codeword which indicates the type of back-channel message. BT is the first codeword present in each back-channel message. Which type or types of message are requested by the encoder is indicated in the RPSMF field of the forward-channel syntax. The values of BT shall be defined as:

- "00": Reserved for future use.
- "01": Reserved for future use.
- "10": NACK. This indicates the loss or erroneous decoding of the corresponding part of the forward channel data.
- "11": ACK. This indicates the correct decoding of the corresponding part of the forward channel data.

U.5.2.2 Enhancement Layer Number Indication (ELNUMI) (1 bit)

ELNUMI is a single-bit fixed-length codeword that follows BT in the back-channel message. ELNUMI shall be "0" unless the optional Temporal, SNR, and Spatial Scalability mode (see Annex O) is used in the forward channel and some enhancement layers of the forward channel are combined in one logical channel and the back-channel message refers to an enhancement layer (rather than the base layer), in which case ELNUMI shall be "1".

U.5.2.3 Enhancement Layer Number (ELNUM) (4 bits)

ELNUM is a 4-bit fixed-length codeword that is present only if ELNUMI is "1". It follows ELNUMI if present. When present, ELNUM contains the layer number of the enhancement layer referred to in the back-channel message.

U.5.2.4 Back-channel CPM Indicator (BCPM) (1 bit)

BCPM is a single-bit fixed-length codeword that follows ELNUMI or ELNUM in the back-channel message. BCPM shall be "0" unless the CPM mode (see 5.2.4 and Annex C) is used in the forward channel data, in which case BCPM shall be "1". If BCPM is "1", this indicates that BSBI is present.

U.5.2.5 Back-channel Sub-Bitstream Indicator (BSBI) (2 bits)

BSBI is a 2-bit fixed-length codeword that follows BCPM when present. BSBI is present only if BCPM is "1". BSBI is the natural binary representation of the Sub-Bitstream number in the forward channel data to which the back-channel message refers (see 5.2.4 and Annex C).

U.5.2.6 Picture Number Type (PNT) (1 bit)

PNT is a single-bit fixed-length codeword that is always present and follows BCPM or BSBI in the back-channel message. The values of PNT shall be defined as:

"0": The message concerns a picture specified by a short-term picture number (PN).

"1": The message concerns a picture specified by a long-term picture index (LPIN).

PNT is followed by PN or LPIN, depending on the value of PNT. PN and LPIN shall be represented as specified for use in forward channel data in U.3.1.3 and U.3.1.5.9, respectively.

U.5.2.7 Requested Picture Number Type (RPNT) (2 bits)

RPNT is a 2-bit fixed-length codeword that is present only if BT indicates a NACK message. It follows PN or LPIN when present. It determines how to identify a picture in the multi-picture buffer which may be used as a reference for the coding of subsequent pictures. The values of RPNT shall be defined as:

"00": No valid pictures in buffer – Buffer should be reset by an I or EI picture with RESET equal to "1".

"01": No particular picture is identified to be used as a reference.

"10": A picture which may be used as a reference is identified by a short-term picture number (PN).

"11": A picture which may be used as a reference is identified by a long-term picture index (LPIN).

If RPNT is "10" or "11", RPNT is followed by PN or LPIN, depending on the value of RPNT. PN and LPIN shall be represented as specified for use in forward channel data in U.3.1.3 and U.3.1.5.9, respectively. Typically the PN or LPIN specified using RPNT identifies the last correctly decoded spatially-corresponding picture area for the picture or region identified in the back-channel message.

U.5.2.8 Additional Data Type (ADT) (2 bits)

ADT is a 2-bit fixed-length codeword that is present after PN, LPIN, or RPNT, as determined by PNT (in an ACK message) or RPNT (in a NACK message). It may occur multiple times if present. It specifies the type of additional data used to identify a region of the picture of concern to which the back-channel message applies. The values of ADT shall be defined as:

"00": End of additional data.

"01": A region is identified by only a GN/MBA field.

"10": A region is identified as a raster-scan area within a picture by GN/MBA and NMBM1.

"11": A region is identified as a raster-scan area within a rectangular slice by GN/MBA and NMBM1.

If ADT is "00", no more data follows in the back-channel message. If ADT is "01", ADT is followed by GN/MBA and then by another ADT. If ADT is "10" or "11", ADT is followed by GN/MBA and NMBM1 and then by another ADT.

If ADT is "10", the region is identified as a region starting at a particular spatial location specified by GN/MBA and containing a specified number of macroblocks in raster-scan order within the picture. If ADT is "11", the region is identified as a region starting at a particular spatial location specified by GN/MBA and containing a specified number of macroblocks in raster-scan order within a rectangular slice. If ADT is present only once and is "00", the region identified is the picture as a whole. If ADT is present more than once, the value "00" is used only to end the loop rather than to identify a region.

U.5.2.9 GOB Number/Macroblock Address (GN/MBA) (5/6/7/9/11/12/13/14 bits)

GN/MBA is a fixed-length codeword which specifies a GOB number or macroblock address. GN/MBA follows ADT when present. GN/MBA is present when indicated by ADT. If the optional Slice Structured mode (see Annex K) is not in use, GN/MBA contains the GOB number of the beginning of an area to which the back-channel message refers. If the optional Slice Structured mode is in use, GN/MBA contains the macroblock address of the beginning of the area to which the back-channel message refers. The length of this field shall be as specified elsewhere in this Recommendation for GN or MBA.

U.5.2.10 Number of Macroblocks Minus 1 (NMBM1) (5/6/7/9/11/12/13/14 bits)

NMBM1 is a fixed-length codeword which specifies a number of macroblocks. NMBM1 is present when indicated by ADT. It follows GN/MBA when present. It contains the natural representation of the number of specified macroblocks minus 1. The length of this field shall be the length defined for a macroblock address in K.2.5 and in Table K.2.

Annex V

Data-partitioned slice mode

V.1 Scope

This annex describes the optional data-partitioned slice (DPS) mode of H.263. The capability of this mode is signalled by external means (for example, ITU-T Rec. H.245). The use of this mode shall be indicated by setting the formerly-reserved bit 17 of the optional part of the PLUSPTYPE (OPPTYPE) to "1". This mode uses the header structure defined in Annex K.

Data partitioning provides robustness in error prone environments. This is accomplished using a rearrangement of the H.263 syntax to enable early detection of and recovery from errors that have been introduced during transmission.

V.2 Structure of data partitioning

When data partitioning is used, the data is arranged as a video picture segment, as defined in R.2. The MBs in the segment are rearranged so that the header information for all the MBs in the segment are transmitted together, followed by the MVs for all the MBs in the segment, and then by the DCT coefficients for all the MBs in the segment. The segment header uses the same syntax as described in K.2. The header, MV, and DCT partitions are separated by markers, allowing for resynchronization at the end of the partition in which an error occurred. Each segment shall contain the data for an integer number of MBs. When this mode is in use, the syntax shown in Figure V.1 shall be used.

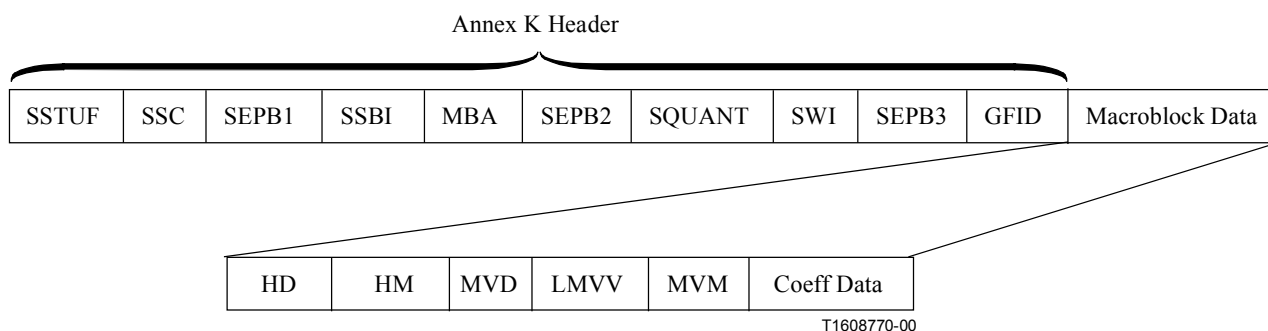


Figure V.1/H.263 – Data partitioning syntax

Note that when this annex is not active, the MV and DCT data are transmitted in an interleaved fashion for all the MBs in a video picture segment, in which case an error normally results in the loss of all information for the remaining MBs in the packet.

V.2.1 Header Data (HD) (Variable length)

The Header Data field contains the COD and MCBPC information for all the MBs in the packets, plus the MODB data in case of PB-frames or Improved PB-frames. A reversible variable length code (RVLC) is used to combine the COD and the MCBPC for all the MBs in the packet. This code is shown in Tables V.1 through V.5. If Annex O is in use, the COD is only combined with the MB TYPE to form the RVLC for B and EP pictures using Tables V.3 and V.4, and the CBPC is coded with codewords in Table O.4. If COD = 0 and Annex G or Annex M is in use, the codeword for the COD+MCBPC shall be immediately followed by the reversible variable-length encoded data corresponding to the MODB field of the macroblock. Table V.6 shall be used for PB-frames, Table V.7 shall be used for Improved PB-frames.

V.2.2 Header Marker (HM) (9 bits)

A codeword of 9 bits. Its value is 1010 0010 1. The HM terminates the header partition. When reversed decoding is used by a decoder, the decoder searches for this marker. This value cannot occur naturally in the HD field.

V.2.3 Motion vector data layer (Variable length)

V.2.3.1 Motion vector difference coding

For the motion vectors, the RVLC codewords shown in Table D.3 are used to encode the difference between the motion vector and the motion vector prediction. Note that this annex only uses the entropy coding from Annex D, but not its other aspects unless Annex D is also in use.

V.2.3.2 Prediction of motion vector values

The first motion vector in the packet is coded using a predictor value of 0 for both horizontal and vertical components, and the MVs for the subsequent coded MBs are coded predictively using the MV difference (MVD). This differs from the method otherwise used for coding the MVs in which the MVs following a skipped or INTRA MB are coded using a predictor value of 0 for both horizontal and vertical components.

Forward Direction: $MV_i = MV_{i-1} + MVD_i = MV_{i-1} + (MV_i - MV_{i-1})$

Backward Direction: $MV_{i-1} = MV_i - MVD_i = MV_i - (MV_i - MV_{i-1})$.

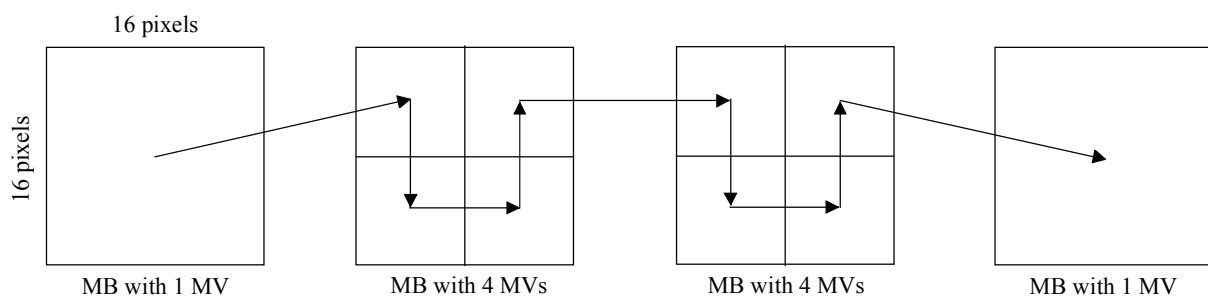
(MV_i and MVD_i are the i th MV and MV Difference in the packet respectively).

The motion vector information for the last motion vector in the packet is coded in this manner and is also coded again in the LMVV field as described below in V.2.4. This allows the decoder to independently decode the sequence of MVs using two different prediction paths:

- 1) in the forward direction, starting from the beginning of the motion data of the packet; and
- 2) in the backward direction, from the end of the motion data in a packet.

This provides robustness for better error detection and concealment.

NOTE 1 – When the DPS mode is not in use, motion vectors are predictively coded, with the prediction of the current motion vector being the median value of 3 motion vectors of neighboring locations as described in 6.1.1. Because packets in this annex are formed in a way such that the number of MBs coded in each packet is variable, using the median predictive coding method (which involves motion vectors on different rows of the frame) would prevent reversible decoding of the motion vectors in a slice. When the DPS mode is in use, a single prediction thread is formed for the MVs in the whole packet. This is shown in Figure V.2.



T1608780-00

Figure V.2/H.263 – Single thread motion vector prediction

In case of B pictures or EP pictures (Annex O), MVDFW and MVDBW may be present as indicated by the MBTYPE codeword in Tables V.3 and V.4. MVDFW is predictively encoded

using the same single prediction thread as described above and MVDBW (when present in B pictures) shall be encoded as specified in O.4.6. MVDFW and MVDBW shall be coded with the codewords from Table D.3.

In case of PB-frames (Annex G) and Improved PB-frames (Annex M), the MVDB data shall be encoded as specified in corresponding annexes and shall be coded using the codewords from Table D.3.

NOTE 2 – If the backward decoding mode is engaged in a B frame (Annex O) or in Improved PB-frames (Annex M), MVDB and MVDBW should be discarded by the decoder as the Motion Vector data for the backward prediction may not be recovered properly across the packet boundaries.

V.2.3.3 Start-code emulation prevention in motion vector difference coding

The MVD start-code-emulation avoidance method is changed from the method described in D.2, in order to facilitate independent parsing in the backward direction. The MV partition shall be scanned from left to right and a MVD = 0 (codeword "1") shall be inserted after any two MVDs that are both equal to 1 (codeword "000"). If a third MVD = 1 codeword follows these two MVD = 1 codewords in the original bitstream (before insertion), it shall be considered the first MVD = 1 codeword detected in the remaining codewords in the MV partition. It shall not be considered a second MVD = 1 codeword, and shall not have a MVD = 0 codeword inserted after it. This differs from Annex D, in which the bit is only inserted when two consecutive MVD = 1 (codeword "000") form a pair (i.e., when the first MVD is the horizontal component, and the second is the vertical component). If Annex D and Annex V are both in use, this Annex V method of start-code-emulation avoidance method shall be used instead of the method described in D.2.

V.2.4 Last Motion Vector Value (LMVV) (Variable length)

The LMVV field contains the last MV in the packet. It is coded using a predictor value of 0 for both the horizontal and vertical components. If there are no motion vectors or only one motion vector in the packet, LMVV shall not be present. (This use of a fixed zero-valued predictor enables the use of reversible decoding.)

V.2.5 Motion Vector Marker (MVM) (10 bits)

A codeword of 10 bits having the value "0000 0000 01". The MVM terminates the motion vector partition. When reverse decoding is used in a decoder, the decoder searches for this marker. The Motion Vector Marker (MVM) shall not be included in the packet if the packet does not contain Motion Vector Data (if all the macroblocks in the packet are intra-coded or with CODs equal to 1).

V.2.6 Coefficient Data Layer (Variable length)

The DCT data layer contains INTRA_MODE (if present), CBPB (if present), CBPC (if present), CBPY, DQUANT (if present), and DCT coefficients coded as specified in I.2, 5.3.4, O.4.3, 5.3.5, 5.3.6, and 5.4.2, respectively. The syntax diagram of DCT Data is illustrated in Figure V.3. The presence of CBPC is indicated in Tables V.3 and V.4.

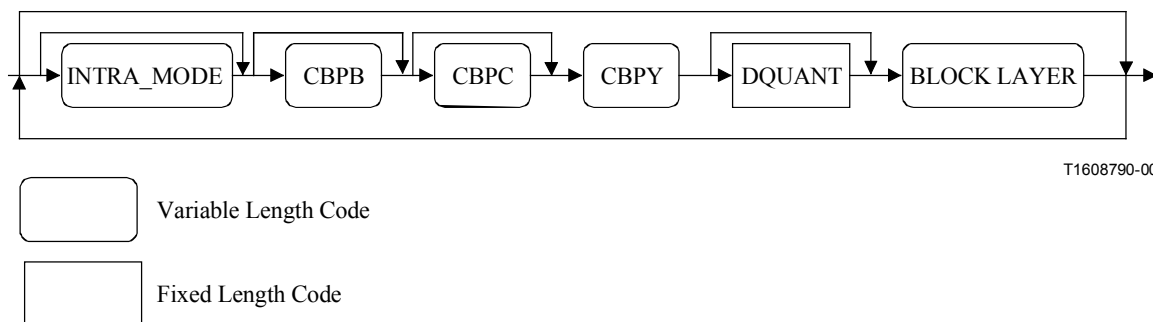


Figure V.3/H.263 – Coefficient Data syntax

V.3 Interaction with other optional modes

The DPS mode acts effectively as a sub-mode of the Slice Structured mode of Annex K, and uses its outer picture and slice header structures. The SS mode shall therefore be indicated as being in use whenever the DPS mode is in use. Both of the other sub-modes of the Slice Structured mode (the Arbitrary Slice Ordering and Rectangular Slice sub-modes) may be used in conjunction with the DPS mode.

The Syntax-Based Arithmetic Coding mode of Annex E shall not be used with this annex, as it does not allow for reversible decoding.

Annex H Forward Error Correction should not be used with this annex, as it can result in the bitstream being disrupted in undesirable places. However, the use of Annex H with the DPS mode is not forbidden, as the FEC defined in Annex H is required in some existing standard system designs.

The Temporal, SNR, and Spatial Scalability (TSSS) mode of Annex O may be used in conjunction with the DPS mode. When the TSSS and DPS modes are used together, the codewords provided in Tables V.3, V.4 and V.5 shall be used instead of those defined in Annex O.

Annex U shall not be used with this annex.

Table V.1/H.263 – COD + MCBPC RVLC table for INTRA MBs

MB type	CBPC (56)	Codeword (for combined COD + MCBPC)	Number of bits
3 (INTRA)	00	1	1
3	01	010	3
3	10	0110	4
3	11	01110	5
4 (INTRA + Q)	00	00100	5
4	01	011110	6
4	10	001100	6
4	11	0111110	7
stuffing		0011100	7

Table V.2/H.263 – COD + MCBPC RVLC table for INTER MBs

MB type	CBPC (56)	Codeword (for combined COD+MCBPC)	Number of bits
skipped		1	1
0 (INTER)	00	010	3
0	10	00100	5
0	01	011110	6
0	11	0011100	7
1 (INTER + Q)	00	01110	5
1	10	00011000	8
1	01	011111110	9
1	11	0111111110	11
2 (INTER4V)	00	0110	4
2	10	01111110	8
2	01	00111100	8
2	11	000010000	9
3 (INTRA)	00	001100	6
3	11	0001000	7
3	10	001111100	9
3	01	000111000	9
4 (INTRA + Q)	00	0111110	7
4	11	0011111100	10
4	10	0001111000	10
4	01	0000110000	10
5 (INTER4V + Q)	00	00111111100	11
5	01	00011111000	11
5	10	00001110000	11
5	11	00000100000	11
stuffing		0111111110	10

Table V.3/H.263 – MBTYPE RVLC codes for B MBs

Index	Prediction type	MVDFW	MVDBW	CBPC + CBPY	DQUANT	MBTYPE	Bits
–	Direct (skipped)					1 (COD=1)	1
0	Direct			X		010	3
1	Direct + Q			X	X	001100	6
2	Forward (no texture)	X				00100	5
3	Forward	X		X		011110	6
4	Forward + Q	X		X	X	01111110	8
5	Backward (no texture)		X			0110	4
6	Backward		X	X		01110	5
7	Backward + Q		X	X	X	00111100	8
8	Bi-Dir (no texture)	X	X			0011100	7
9	Bi-Dir	X	X	X		0001000	7
10	Bi-Dir + Q	X	X	X	X	0111110	7
11	INTRA			X		00011000	8
12	INTRA + Q			X	X	011111110	9
13	Stuffing					001111100	9

Table V.4/H.263 – MBTYPE RVLC table for EP MBs

Index	Prediction type	MVDFW	MVDBW	CBPC + CBPY	DQUANT	MBTYPE	Bits
–	Forward (skipped)					1 (COD=1)	1
0	Forward	X		X		010	3
1	Forward + Q	X		X	X	0110	4
2	Upward (no texture)					01110	5
3	Upward			X		00100	5
4	Upward + Q			X	X	011110	6
5	Bi-Dir (no texture)					001100	6
6	Bi-Dir	X		X		0111110	7
7	Bi-Dir + Q	X		X	X	0011100	7
8	INTRA			X		0001000	7
9	INTRA + Q			X	X	01111110	8
10	Stuffing					00111100	8

Table V.5/H.263 – COD + MCBPC RVLC table for EI MBs

Prediction type	QCBP (56)	Codeword (for combined COD+MCBPC)	Number of bits
Upward (skipped)		1	1
0 (Upward)	00	010	3
0	01	0110	4
0	10	01110	5
0	11	00100	5
1 (Upward + Q)	00	011110	6
1	01	001100	6
1	10	0111110	7
1	11	0011100	7
2 (INTRA)	00	0001000	7
2	01	01111110	8
2	10	00111100	8
2	11	00011000	8
3 (INTRA + Q)	00	011111110	9
3	01	001111100	9
3	10	000111000	9
3	11	000010000	9
Stuffing		0111111110	10

Table V.6/H.263 – RVLC table for MODB

Index	CBPB	MVDB	Number of bits	Code
0			3	010
1		X	4	0110
2	X	X	5	01110

NOTE – "X" means that the item is present in the macroblock.

Table V.7/H.263 – RVLC table for MODB for Improved PB-frames mode

Index	CBPB	MVDB	Number of bits	Code	Coding mode
0			3	010	Bidirectional prediction
1	X		4	0110	Bidirectional prediction
2		X	5	01110	Forward prediction
3	X	X	5	00100	Forward prediction
4			6	011110	Backward prediction
5	X		6	001100	Backward prediction

NOTE – The symbol "X" in the table above indicates that the associated syntax element is present.

Annex W

Additional supplemental enhancement information specification

W.1 Scope

This annex describes the format of the additional supplemental enhancement information sent in the PSUPP field of the picture layer of H.263, which adds to the functionality defined in Annex L. The capability of a decoder to provide any or all of the capabilities described in this annex may be signalled by external means (for example, ITU-T Rec. H.245). Decoders which do not provide the additional capabilities may simply discard any of the newly defined PSUPP information bits that appear in the bitstream. The presence of this supplemental enhancement information is indicated by the presence of both the PEI bit, and by the following PSUPP octet whose FTYPE field has one of the two newly defined values. The basic interpretation of PEI, PSUPP, FTYPE, and DSIZE is identical to Annex L and to clauses 5.1.24 and 5.1.25.

W.2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

- ISO/IEC 10646:2003, *Information technology – Universal Multiple-Octet Coded Character Set (UCS)*.
- IETF RFC 2396 (1998), *Uniform Resource Identifiers (URI): Generic Syntax*.

W.3 Additional FTYPE values

Two values that were reserved in Annex L, Table L.1 are defined as in Table W.1.

Table W.1/H.263 – FTYPE function type values

13	Fixed-Point IDCT
14	Picture Message

W.4 Recommended maximum number of PSUPP octets

When using any of the aforementioned FTYPE functions defined in this annex, the total number of PSUPP octets per picture should, in relation to the coded picture size, be kept reasonably small, and should not exceed 256 octets regardless of the coded picture size.

NOTE – Some data transmission protocols used for conveyance of the video bitstream may provide for external repetition of picture header contents for error resilience purposes, and may place limits on the amount of such data that can be repeated from a picture header (e.g., 504 bits in the IETF RFC 2429 packetization format). The inclusion of a large number of PSUPP octets may result in the lack of such an external protocol to provide for full repetition of the picture header contents.

W.5 Fixed-point IDCT

The fixed-point IDCT function indicates that a particular IDCT approximation is used in construction of the bitstream. DSIZE shall be equal to 1 for the fixed-point IDCT function. The

octet of PSUPP data that follows specifies the particular IDCT implementation. A value of 0 indicates the reference IDCT 0 as described in W.5.3; values of 1 through 255 are reserved.

W.5.1 Decoder operation

The capability of a decoder to perform a particular fixed-point IDCT may be signalled to the encoder by external means (for example, ITU-T Rec. H.245). When receiving an encoded bitstream with the fixed-point IDCT indication, a decoder shall use the particular fixed-point IDCT if it is capable of doing so.

W.5.2 Removal of forced updating

Annex A specifies the accuracy requirements for the inverse discrete cosine transform (IDCT), allowing numerous compliant implementations. To control accumulation of errors due to mismatched IDCTs at the encoder and decoder, clause 4.4, Forced Updating, requires that macroblocks be coded in INTRA mode at least once every 132 times when coefficients are transmitted.

If the fixed-point IDCT function type is indicated in the bitstream, then the forced updating requirement is removed, and the frequency of INTRA coding is unregulated. An encoder should continue to use forced updating, however, unless it has ascertained through external means that the decoder is capable of the particular fixed-point IDCT specified herein; otherwise there may be mismatch.

W.5.3 Reference IDCT 0

The reference IDCT 0 is any implementation that, for every input block, produces identical output values as the C source program listed below.

NOTE – This fixed-point IDCT is compliant with Annex A, but is not compliant with the extended range of values requirement in Annex A of ITU-T Rec. H.262 | ISO/IEC 13818-2.

```
/*
 *
 *          FIXED-POINT IDCT
 *
 * Fixed-point fast, separable idct
 * Storage precision: 16 bits signed
 * Internal calculation precision: 32 bits signed
 * Input range: 12 bits signed, stored in 16 bits
 * Output range: [-256, +255]
 * All operations are signed
 *
 */
#include <stdlib.h>
#include <stdio.h>

/*
 * Typedefs
 */

typedef short int REGISTER; /* 16 bits signed */
typedef long int LONG; /* 32 bits signed */

/*
 * Global constants
 */
```

```

const REGISTER cpo8   = 0x539f;    /* 32768*cos(pi/8)*1/sqrt(2) */
const REGISTER spo8   = 0x4546;    /* 32768*sin(pi/8)*sqrt(2)   */
const REGISTER cpo16  = 0x7d8a;    /* 32768*cos(pi/16)         */
const REGISTER spo16  = 0x18f9;    /* 32768*sin(pi/16)         */
const REGISTER c3po16 = 0x6a6e;    /* 32768*cos(3*pi/16)       */
const REGISTER s3po16 = 0x471d;    /* 32768*sin(3*pi/16)       */
const REGISTER OoR2   = 0x5a82;    /* 32768*1/sqrt(2)         */

/*
 * Function declarations
 */

void Transpose(REGISTER block[64]);
void HalfSwap(REGISTER block[64]);
void Swap(REGISTER block[64]);
void Scale(REGISTER block[64], signed char sh);
void Round(REGISTER block[64], signed char sh,
           const REGISTER min, const REGISTER max);
REGISTER Multiply(const REGISTER a, REGISTER x, signed char sh);
void Rotate(REGISTER *x, REGISTER *y,
            signed char sha, signed char shb,
            const REGISTER a, const REGISTER b,
            int inv);
void Butterfly(REGISTER column[8], char pass);
void IDCT(REGISTER block[64]);

/*
 * Transpose():
 *   Transpose a block
 * Input:
 *   REGISTER block[64]
 * Output:
 *   block
 * Return value:
 *   none
 */
void Transpose(REGISTER block[64])
{
    int i, j;
    REGISTER temp;

    for (i=0; i<8; i++) {
        for (j=0; j<i; j++) {
            temp = block[8*i+j];
            block[8*i+j] = block[8*j+i];
            block[8*j+i] = temp;
        }
    }
    return;
}

/*
 * HalfSwap():
 *   One-dimensional swap
 * Input:
 *   REGISTER block[64]
 * Output:
 *   block
 * Return value:
 *   none
 */
void HalfSwap(REGISTER block[64])
{

```

```

int i;
REGISTER temp;

for (i=0; i<8; i++) {
    temp = block[8+i];
    block[8+i] = block[32+i];
    block[32+i] = temp;
    temp = block[24+i];
    block[24+i] = block[48+i];
    block[48+i] = temp;
    temp = block[40+i];
    block[40+i] = block[56+i];
    block[56+i] = temp;
}
return;
}

/*
 * Swap():
 *     Swap and transpose a block
 * Input:
 *     REGISTER block[64]
 * Output:
 *     block
 * Return value:
 *     none
 */
void Swap(REGISTER block[64])
{
    HalfSwap(block);
    Transpose(block);
    HalfSwap(block);
}

/*
 * Scale():
 *     Scale a block
 * Input:
 *     REGISTER block[64]
 *     signed char sh
 * Output:
 *     block
 * Return value:
 *     none
 */
void Scale(REGISTER block[64], signed char sh)
{
    int i;

    if (sh>0) {
        for (i=0; i<64; i++)
            block[i] >>= sh;
    }
    else {
        for (i=0; i<64; i++)
            block[i] <<= -sh;
    }
}

/*
 * Round():
 *     Performs the final rounding of an 8x8 block
 * Input:
 *     REGISTER block[64]

```

```

*      signed char sh
*      const REGISTER min
*      const REGISTER max
* Output:
*      block
* Return value:
*      none
*/
void Round(REGISTER block[64], signed char sh,
           const REGISTER min, const REGISTER max)
{
    int i;

    for (i=0; i<64; i++) {
        if (block[i] < 0x00007FFF - (1<<(sh-1)))
            block[i] += (1<<(sh-1));
        else
            block[i] = 0x00007FFF;
        block[i] >>= sh;
        block[i] = (block[i]<min) ? min : ((block[i]>max) ? max : block[i]);
    }
    return;
}

/*
* Multiply():
*      Multiply by a constant with shift
* Input:
*      const REGISTER a
*      REGISTER x
*      signed char sh
* Output:
*      none
* Return value:
*      REGISTER, the result of the multiply
*/
REGISTER Multiply(const REGISTER a, REGISTER x, signed char sh)
{
    LONG tmp;
    REGISTER reg_out;

    /* multiply */
    tmp = (LONG)a * (LONG)x;

    /* shift */
    if (sh > 0)
        tmp >>= sh;
    else
        tmp <<= -sh;

    /* rounding and saturating */
    if (tmp < 0x7FFFFFFF - 0x00007FFF)
        tmp = tmp + 0x00007FFF;
    else
        tmp = 0x7FFFFFFF;

    reg_out = (REGISTER)(tmp >>16);

    return(reg_out);
}

/*
* Rotate():
*      Perform rotate operation on two registers

```

```

* Input:
*   REGISTER *x           pointer to the 1st register
*   REGISTER *y           pointer to the 2nd register
*   signed char sha       shift associated with factor a
*   signed char shb       shift associated with factor b
*   const REGISTER a      factor a
*   const REGISTER b      factor b
*   int inv               1 for inverse dct, 0 for forward dct
* Output:
*   *x, *y
* Return value:
*   none
*/
void Rotate(REGISTER *x, REGISTER *y,
            signed char sha, signed char shb,
            const REGISTER a, const REGISTER b,
            int inv)
{
    LONG tmp1xa, tmp1ya, tmp1xb, tmp1yb;
    LONG tmp11, tmp12;

    /*
     * intermediate calculation
     */

    tmp1xa = (LONG)(*x) * (LONG)a;
    if (sha > 0)
        tmp1xa >>= sha;
    else
        tmp1xa <<= -sha;

    tmp1ya = (LONG)(*y) * (LONG)a;
    if (sha > 0)
        tmp1ya >>= sha;
    else
        tmp1ya <<= -sha;

    tmp1xb = (LONG)(*x) * (LONG)b;
    if (shb > 0)
        tmp1xb >>= shb;
    else
        tmp1xb <<= -shb;

    tmp1yb = (LONG)(*y) * (LONG)b;
    if (shb > 0)
        tmp1yb >>= shb;
    else
        tmp1yb <<= -shb;

    /*
     * rounding and rotation
     */

    if (inv) {
        tmp1xa += 0x00007FFF;
        tmp1xb += 0x00007FFF;

        tmp11 = tmp1xb - tmp1ya;
        tmp12 = tmp1xa + tmp1yb;
    }
    else {
        tmp1ya += 0x00007FFF;
        tmp1yb += 0x00007FFF;
    }
}

```



```

    tmp11 = tmp1xb + tmp1ya;
    tmp12 = -tmp1xa + tmp1yb;
}

/*
 * final rounding
 */

*x = (REGISTER) (tmp11 >>16);
*y = (REGISTER) (tmp12 >>16);

return;
}

/*
 * Butterfly():
 *     Perform 1D IDCT on a column
 * Input:
 *     REGISTER column[8]
 *     char pass
 * Output:
 *     column
 * Return value:
 *     none
 */
void Butterfly(REGISTER column[8], char pass)
{
    int i;
    REGISTER shadow_column[8];

    /*
     * For readability, we use a shadow column
     * that contains the state of column at the
     * preceding stage of the butterfly.
     */

    /*
     * Initialization
     */

    for (i=0; i<8; i++)
        shadow_column[i] = column[i];

    /*
     * First Phase
     */

    Rotate(column+2, column+6, pass-2, pass-1, cpo8, spo8, 1);
    Rotate(column+1, column+7, pass-1, pass-1, cpo16, spo16, 1);
    Rotate(column+3, column+5, pass-1, pass-1, c3po16, s3po16, 1);

    if (pass) {
        int a, tmp=column[4], b=column[0];
        a = b+tmp;
        b = b-tmp;
        column[0] = (a - ((tmp<0) ? 1 : 0)) >> 1;
        column[4] = (b - ((tmp<0) ? 1 : 0)) >> 1;
    }
    else {
        column[0] = shadow_column[0] + shadow_column[4];
        column[4] = shadow_column[0] - shadow_column[4];
    }
}

```

```

for (i=0; i<8; i++)
    shadow_column[i] = column[i];

/*
 * Second Phase
 */

column[1] = shadow_column[1] - shadow_column[3];
column[3] = shadow_column[1] + shadow_column[3];

column[7] = shadow_column[7] - shadow_column[5];
column[5] = shadow_column[7] + shadow_column[5];

column[0] = shadow_column[0] + shadow_column[6];
column[6] = shadow_column[0] - shadow_column[6];

column[4] = shadow_column[4] + shadow_column[2];
column[2] = shadow_column[4] - shadow_column[2];

for (i=0; i<8; i++)
    shadow_column[i] = column[i];

/*
 * Third Phase
 */

column[7] = shadow_column[7] - shadow_column[3];
column[3] = shadow_column[7] + shadow_column[3];

column[1] = Multiply(OoR2, shadow_column[1], -2);
column[5] = Multiply(OoR2, shadow_column[5], -2);

for (i=0; i<8; i++)
    shadow_column[i] = column[i];

/*
 * Fourth Phase
 */

column[4] = shadow_column[4] + shadow_column[3];
column[3] = shadow_column[4] - shadow_column[3];

column[2] = shadow_column[2] + shadow_column[7];
column[7] = shadow_column[2] - shadow_column[7];

column[0] = shadow_column[0] + shadow_column[5];
column[5] = shadow_column[0] - shadow_column[5];

column[6] = shadow_column[6] + shadow_column[1];
column[1] = shadow_column[6] - shadow_column[1];

return;
}

/*
 * IDCT():
 * Perform 2D IDCT on a block
 * Input:
 * REGISTER block[64]
 * Output:
 * block
 * Return value:
 * none
 */

```

```

void IDCT(REGISTER block[64])
{
    int i;

    Scale(block, -4);

    for (i=0; i<8; i++)
        Butterfly(block+8*i, 0);

    Transpose(block);

    for (i=0; i<8; i++)
        Butterfly(block+8*i, 1);

    Round(block, 6, -256, 255);

    Swap(block);
}

```

For informative purposes, a related forward discrete cosine transform (FDCT) implementation is shown below. This fixed-point FDCT does not form an integral part of this Recommendation.

```

/*****
 *
 *                      FIXED-POINT FDCT
 *
 * Fixed-point fast, separable fdct
 * Storage precision: 16 bits signed
 * Internal calculation precision: 32 bits signed
 * Input range: 9 bits signed, stored in 16 bits
 * Output range: [-2048, +2047]
 * All operations are signed
 *
 *****/

/*
 * Function declarations
 */

void FButterfly(REGISTER column[8]);
void FDCT(REGISTER block[64]);

/*
 * FButterfly():
 *     Perform 1D FDCT on a column
 * Input:
 *     REGISTER column[8]
 * Output:
 *     column
 * Return value:
 *     none
 */
void FButterfly(REGISTER column[8])
{
    int i;
    REGISTER shadow_column[8];

    /*
     * For readability, we use a shadow column
     * that contains the state of column at the
     * preceding stage of the butterfly.
     */
}

```

```

/*
 * Initialization
 */

for (i=0; i<8; i++)
    shadow_column[i] = column[i];

/*
 * First Phase
 */

for (i=0; i<4; i++) {
    column[i]    = shadow_column[i] + shadow_column[7-i];
    column[7-i] = shadow_column[i] - shadow_column[7-i];
}

for (i=0; i<8; i++)
    shadow_column[i] = column[i];

/*
 * Second Phase
 */

column[0] = shadow_column[0] + shadow_column[3];
column[3] = shadow_column[0] - shadow_column[3];

column[1] = shadow_column[1] + shadow_column[2];
column[2] = shadow_column[1] - shadow_column[2];

column[4] = Multiply(OoR2, shadow_column[4], -2);
column[7] = Multiply(OoR2, shadow_column[7], -2);

column[6] = shadow_column[6] - shadow_column[5];
column[5] = shadow_column[6] + shadow_column[5];

for (i=0; i<8; i++)
    shadow_column[i] = column[i];

/*
 * Third Phase
 */

column[0] = shadow_column[0] + shadow_column[1];
column[1] = shadow_column[0] - shadow_column[1];

column[6] = shadow_column[6] - shadow_column[4];
column[4] = shadow_column[6] + shadow_column[4];

column[7] = shadow_column[7] - shadow_column[5];
column[5] = shadow_column[7] + shadow_column[5];

for (i=0; i<8; i++)
    shadow_column[i] = column[i];

/*
 * Fourth Phase
 */

Rotate(column+2, column+3, -2, -1, cpo8, spo8, 0);
Rotate(column+4, column+5, -1, -1, cpo16, spo16, 0);
Rotate(column+6, column+7, -1, -1, c3po16, s3po16, 0);

return;
}

```

```

/*
 * FDCT() :
 *     Perform 2D FDCT on a block
 * Input:
 *     REGISTER block[64]
 * Output:
 *     block
 * Return value:
 *     none
 */
void FDCT(REGISTER block[64])
{
    int i;

    for (i=0; i<8; i++)
        FButterfly(block+8*i);

    Transpose(block);

    for (i=0; i<8; i++)
        FButterfly(block+8*i);

    Round(block, 3, -2048, 2047);

    Swap(block);
}

```

W.6 Picture message

The picture message function indicates the presence of one or more octets representing message data. The first octet of the message data is a message header with the following structure, as shown in Figure W.1.

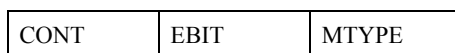


Figure W.1/H.263 – Structure of first message octet

DSIZE shall be equal to the number of octets in the message data corresponding to a picture message function, including the first octet shown in Figure W.1.

Decoders shall parse picture message data as required by basic PSUPP syntax, but decoder response to picture messages is otherwise undefined.

W.6.1 Continuation (CONT) (1 bit)

If equal to "1", CONT indicates that the message data associated with this picture message function is part of the same logical message as the message data associated with the next picture message function. If equal to "0", CONT indicates that the message data associated with this picture message function terminates the current logical message. CONT may be used, for example, to represent logical messages that span more than 14 octets.

W.6.2 End Bit Position or Track Number (EBIT) (3 bits)

For non-text picture messages, EBIT specifies the number of least significant bits that shall be ignored in the last message octet. In non-text picture messages, if CONT is "1", or if there is only one message octet (i.e., the octet in Figure W.1), EBIT shall equal "0". The number of valid message bits for a non-text picture message function excluding the CONT/EBIT/MTYPE bits is

equal to $(DSIZE - 1) \times 8 - EBITS$. The number of valid message bits for a logical message may be greater due to continuation.

For picture message types containing text information, EBIT shall contain a text track number. The precise meaning of the text track number is not specified herein, but should indicate a particular type (e.g., language) for the text. Track number zero should be considered the default track.

W.6.3 Message Type (MTYPE) (4 bits)

MTYPE indicates the type of message. The defined types are shown in Table W.2.

Table W.2/H.263 – MTYPE message type values

0	Arbitrary Binary Data
1	Arbitrary Text
2	Copyright Text
3	Caption Text
4	Video Description Text
5	Uniform Resource Identifier Text
6	Current Picture Header Repetition
7	Previous Picture Header Repetition
8	Next Picture Header Repetition, Reliable TR
9	Next Picture Header Repetition, Unreliable TR
10	Top Interlaced Field Indication
11	Bottom Interlaced Field Indication
12	Picture Number
13	Spare Reference Pictures
14..15	Reserved

W.6.3.1 Arbitrary binary data

Arbitrary binary data is used to convey any non-ISO/IEC 10646 UTF-8 coded binary message. The interpretation of contents of the arbitrary binary data are outside the scope of this Recommendation, but should begin with some identifying pattern (e.g., a four-octet identifier code) to aid in distinguishing one type of such data from others.

W.6.3.2 Arbitrary text

Arbitrary text is used to convey a generic ISO/IEC 10646 UTF-8 coded text message. More specific text messages such as copyright information should be represented with other message types (e.g., copyright text) as appropriate.

W.6.3.3 Copyright text

Copyright text shall be used only to convey intellectual property information regarding the source or the encoded representation in the bitstream. The copyright message shall be coded according to ISO/IEC 10646 UTF-8.

W.6.3.4 Caption text

Caption text shall be used only to convey caption information associated with the current and subsequent pictures of the bitstream. The caption message shall be coded according to ISO/IEC 10646 UTF-8. The caption text shall be inserted in the bitstream as if it were to be displayed in a separate text area where new text is appended at the end of previous text and earlier text scrolled

away from the point of insertion. The Form Feed (hexadecimal "0x000C") control code shall be used to indicate clearing of the visible text area. The End of Medium (hexadecimal "0x0019") control code shall be used to indicate "caption off" status. However, this Recommendation puts no restriction on how caption text is actually displayed and stored.

W.6.3.5 Video description text

Video description text shall be used only to convey descriptive information associated with the information contents of the current bitstream. The video description shall be coded according to ISO/IEC 10646 UTF-8. The video description text shall be inserted in the bitstream as if it were to be displayed in a separate text area where new text is appended at the end of previous text and earlier text scrolled away from the point of insertion. The Form Feed (hexadecimal "0x000C") control code shall be used to indicate clearing of the visible text area. The End of Medium (hexadecimal "0x0019") control code shall be used to indicate "description off" status. However, this Recommendation puts no restriction on how video description text is actually displayed and stored.

W.6.3.6 Uniform Resource Identifier (URI) text

The message consists of a uniform resource identifier (URI), as defined in IETF RFC 2396. The URI shall be coded according to ISO/IEC 10646 UTF-8.

W.6.3.7 Current picture header repetition

The picture header from the current picture is repeated in this message. The repeated bits exclude any supplemental enhancement information (PEI/PSUPP). All other bits up to the GOB or Slice layer should be included, subject to the limitations of W.4.

W.6.3.8 Previous picture header repetition

The picture header from the previously transmitted picture is repeated in this message. The repeated bits exclude the first two bytes of picture start code (PSC) and any supplemental enhancement information (PEI/PSUPP). All other bits up to the GOB or Slice layer should be included, subject to the limitations of W.4.

W.6.3.9 Next picture header repetition, reliable TR

The picture header from the next picture to be transmitted is repeated in this message. The repeated bits exclude the first two bytes of picture start code (PSC) and any supplemental enhancement information (PEI/PSUPP). All other bits up to the GOB or Slice layer should be included, subject to the limitations of W.4.

W.6.3.10 Next picture header repetition, unreliable TR

The picture header from the next picture to be transmitted is repeated in this message. The repeated bits exclude the first three bytes of picture header and any supplemental enhancement information (PEI/PSUPP). All other bits up to the GOB or Slice layer should be included, subject to the limitations of W.4. Any TR or ETR bits in the repeated picture header are not necessarily the same as the corresponding bits in the next picture header.

W.6.3.11 Interlaced field indications

In the case of interlaced field indications, the message consists of an indication of interlaced field coding. This indication does not affect the decoding process. However, it indicates that the current picture was not actually scanned as a progressive-scan picture. In other words, it indicates that the current coded picture contains only half of the lines of the full resolution source picture. DSIZE shall be 1, CONT shall be 0, and EBIT shall be 0 for interlaced field indications. In the case of interlaced field coding, each increment of the temporal reference denotes the time between the sampling of alternate half-picture fields of a picture, rather than the time between two complete

pictures. In the case of a top interlaced field indication, the current picture contains the first (i.e., top), third, fifth, etc. lines of the complete picture. In the case of a bottom interlaced field indication, the current picture contains the second, fourth, sixth, etc. lines of the complete picture. When sending interlaced field indications, an encoder shall conform to the following conventions:

- 1) The encoder shall use a picture clock frequency (custom picture clock frequency, if necessary) such that each new field of the original source video corresponds to an increment of 1 in the temporal reference.
- 2) The encoder shall use a picture size (custom picture size, if necessary) such that the picture dimensions correspond to those of a single field.
- 3) The encoder shall use a pixel aspect ratio (custom pixel aspect ratio, if necessary) such that the full-height picture aspect ratio corresponds to the picture aspect ratio derived from the pixel aspect ratio of the single field represented by the current encoded picture.

Interlaced field scanning was introduced originally as an analog video compression technique. Although progressive picture scanning is generally regarded as superior for digital compression and display, the use of interlaced field scanning has persisted in many camera and display designs. Interlaced field coding (which can be implemented with lower delay than either interlaced full-picture coding or progressive-scan picture coding at half the interlaced field rate) is therefore supported by the indications herein.

An encoder shall not send interlaced field indications unless the capability of the decoder to receive and properly process such field-based pictures has been established by external means (for example, ITU-T Rec. H.245). Failure to establish such a decoder capability may produce a visually annoying small-amplitude vertical shaking behaviour in the decoded picture received and displayed by a decoder.

For example, an encoder may use interlaced field coding with application of the Reference Picture Selection mode (specified in Annex N) or the Enhanced Reference Picture Selection mode (specified in Annex U) to allow the addressing of more than one prior field. For "525/60" interlaced field coding for a 4:3 picture aspect ratio with 704 coded luminance samples per line and 240 coded luminance lines per field, the encoder shall use a custom picture size having a picture width of 704 and a picture height of 240, a custom pixel aspect ratio of 5:11, and a custom picture clock frequency specified with a clock conversion code "1" and a clock divisor of 30. For "625/50" interlaced field coding for a 4:3 picture aspect ratio with 704 coded luminance samples per line and 288 coded luminance lines per field, the encoder shall use a custom picture size having a picture width of 704 and a picture height of 288, a custom pixel aspect ratio of 6:11, and a custom picture clock frequency specified with a clock conversion code "0" and a clock divisor of 36.

The vertical sampling positions of the chrominance samples in interlaced field coding of a top field picture are specified as shifted up by 1/4 luminance sample height relative to the field sampling grid in order for these samples to align vertically to the usual position relative to the full-picture sampling grid. The vertical sampling positions of the chrominance samples in interlaced field coding of a bottom field picture are specified as shifted down by 1/4 luminance sample height relative to the field sampling grid in order for these samples to align vertically to the usual position relative to the full-picture sampling grid. The horizontal sampling positions of the chrominance samples are specified as unaffected by the application of interlaced field coding. The vertical sampling positions are shown with their corresponding temporal sampling positions in Figure W.2.

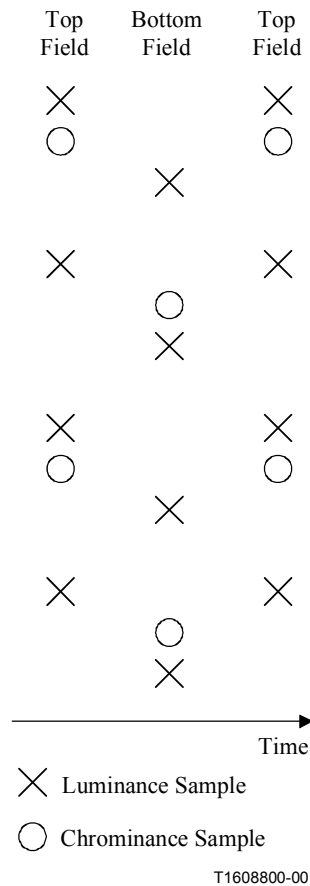


Figure W.2/H.263 – Vertical and temporal alignment of chrominance samples for interlaced field coding

W.6.3.12 Picture number

This message shall not be used if Annex U is in use. The message contains two data bytes that carry a 10-bit Picture Number. Consequently, DSIZE shall be 3, CONT shall be 0, and EBIT shall be 6. Picture Number shall be incremented by 1 for each coded and transmitted I or P picture or PB or Improved PB frame, in a 10-bit modulo operation. For EI and EP pictures, Picture Number shall be incremented for each EI or EP picture within the same scalability enhancement layer. For B pictures, Picture Number shall be incremented relative to the value in the most recent non-B picture in the reference layer of the B picture which precedes the B picture in bitstream order (a picture which is temporally subsequent to the B picture). If adjacent pictures in the same enhancement layer have the same temporal reference, and if the reference picture selection mode (see Annex N) is in use, the decoder shall regard this occurrence as an indication that redundant copies have been sent of approximately the same pictured scene content, and all of these pictures shall share the same Picture Number. If the difference (modulo 1024) of the Picture Numbers of two consecutively received non-B pictures in the same enhancement layer is not 1, and if the pictures do not represent approximately the same pictured scene content as described above, the decoder should infer a loss of pictures or corruption of data.

W.6.3.13 Spare reference pictures

Encoders can use this message to instruct decoders which pictures resemble the current motion compensation reference picture so well that one of them can be used as a spare reference picture if the actual reference picture is lost during transmission. If a decoder lacks an actual reference picture but can access a spare reference picture, it should not request for an INTRA picture update. It is up to encoders to choose the spare reference pictures, if any. The message data bytes contain the

Picture Numbers of the spare reference pictures in preference order (the most preferred appearing first). Picture Numbers refer to the values that are transmitted according to Annex U or W.6.3.12. This message can be used for P, B, PB, Improved PB, and EP picture types. However, if Annex N or Annex U is in use and if the picture is associated with multiple reference pictures, this message shall not be used. For EP pictures, the message shall be used only for forward prediction, whereas upward prediction is always done from the temporally corresponding reference layer picture. For B, PB, and Improved PB picture types, it specifies a picture for use as a forward motion prediction reference. This message shall not be used if the picture is an I or EI picture.

Annex X

Profiles and levels definition

X.1 Scope

With the variety of optional modes available in this Recommendation, it is crucial that several preferred mode combinations for operation be defined, so that option-enhanced terminals will have a high probability of connecting to each other using some syntax better than the "baseline". This annex contains a list of preferred feature combinations, which are structured into "profiles" of support. It also defines some groupings of maximum performance parameters as "levels" of support for these profiles. The primary objectives of this annex are:

- 1) to provide a simple means of describing or negotiating the capabilities of a decoder (by specifying profile and level parameters);
- 2) to encourage common enhancement features to be supported in decoders for achieving maximal interoperability; and
- 3) to describe feature sets chosen as particularly appropriate for addressing certain key applications.

The profiles and levels are defined in the following clauses and in Tables X.1 and X.2. The minimum picture interval as specified in Table X.2 is the minimum difference in time between the decoding of consecutive pictures in the bitstream. Support of any level other than level 45 implies support of all lower levels. Support of level 45 implies support of level 10.

X.2 Profiles of preferred mode support

The profiles of support are defined by the set of features supported in the decoder for each profile. Decoder support for a given profile implies support for all valid subset combinations of the constituent modes of that profile. This requirement exists so that the limitations placed upon an encoder's choice of mode combinations are minimized. This is in keeping with the primary objective of this annex, which is to describe which optional modes should be supported at the decoder to address key applications, rather than to enforce a particular small set of mode combinations upon the encoder.

X.2.1 The Baseline Profile (Profile 0)

The Baseline Profile, designated as Profile 0, is defined herein to provide a profile designation for the minimal "baseline" capability of this Recommendation. "Baseline" refers to the syntax of this Recommendation with no optional modes of operation. This profile of support is composed of only the baseline design.

X.2.2 H.320 Coding Efficiency Version 2 Backward-Compatibility Profile (Profile 1)

The H.320 Coding Efficiency Version 2 Backward Compatibility Profile, designated as Profile 1, is defined herein to provide compatibility with a feature set adopted into the H.242 capability exchange mechanism for use by H.320 circuit-switched terminal systems. It provides basic enhanced coding efficiency and simple enhanced functionality within the feature set available in the second version of this Recommendation (which did not include Annexes U, V and W). This profile of support is composed of the baseline design plus the following modes:

- 1) **Advanced INTRA Coding (Annex I)** – Use of this mode improves the coding efficiency for INTRA macroblocks (whether within INTRA pictures or predictively-coded pictures). The additional computational requirements of this mode are minimal at both the encoder and decoder (as low as a maximum of 8 additions/subtractions per 8×8 block in the decoding process plus the use of a different but very similar VLC table in order to obtain a significant improvement in coding efficiency). For these reasons, Advanced INTRA Coding is included in this basic package of support.
- 2) **Deblocking Filter (Annex J)** – Because of the significant subjective quality improvement that may be realized with a deblocking filter, these filters are widely in use as a method of post-processing in video communication terminals. Annex J represents the preferred mode of operation for a deblocking filter because it places the filter within the coding loop. This placement eases the implementation of the filter (by reducing the required memory) and somewhat improves the coding performance over a post-processing implementation. As with the Advanced Prediction mode, this mode also includes the four-motion-vector-per-macroblock feature and picture boundary extrapolation for motion compensation, both of which can further improve coding efficiency. The computational requirements of the deblocking filter are several hundred operations per coded macroblock, but memory accesses and computational dependencies are uncomplicated. This last point is what makes the Deblocking Filter preferable to Advanced Prediction for some implementations. Also, the benefits of Advanced Prediction are not as substantial when the Deblocking Filter is used as well. Thus, the Deblocking Filter is included in this basic package of support.
- 3) **Full-Picture Freeze Supplemental Enhancement Information (Annex L, clause L.4)** – The full-picture freeze is very simple to implement, requiring only that the decoder be able to stop the transfer of new pictures from its output buffer to the video display. This capability is useful for preventing the display of low-fidelity pictures while the encoder is building up a higher fidelity picture.
- 4) **Modified Quantization (Annex T)** – This mode includes an extended DCT coefficient range, modified DQUANT syntax, and a modified step size for chrominance. The first two features allow for more flexibility at the encoder and may actually decrease the encoder's computational load (by eliminating the need re-encode macroblocks when coefficient level saturation occurs). The third feature noticeably improves chrominance fidelity, typically with little added bit-rate cost and with virtually no increase in computation. At the decoder, the only significant added computational burden is the ability to parse some new bitstream symbols.

X.2.3 Version 1 Backward-Compatibility Profile (Profile 2)

The Version 1 Backward-Compatibility Profile, designated as Profile 2, is defined herein to provide enhanced coding efficiency performance within the feature set available in the first version of ITU-T Rec. H.263 (which did not include Supplemental Enhancement Information or any of the optional features which use PLUSPTYPE). This profile of support is composed of the baseline design plus the following single mode:

- 1) **Advanced Prediction (Annex F)** – From a coding efficiency standpoint, this mode is the most important of the modes available in the first version (Version 1) of this Recommendation. It includes overlapped block motion compensation, the four-motion-vector-per-macroblock feature, and it allows for motion vectors to point outside of the picture boundaries. The use of Advanced Prediction results in significant improvements in both subjective and objective performance. It does, however, require an appreciable increase in computation and introduces complicating data dependencies in the order of processing at the decoder. However, since implementations of this Recommendation that were designed prior to the adoption of the other modes in this list might have implemented Advanced Prediction by itself, Advanced Prediction-only operation is recommended for maximal quality with backward compatibility to version 1 decoders.

X.2.4 Version 2 Interactive and Streaming Wireless Profile (Profile 3)

The Version 2 Interactive and Streaming Wireless Profile, designated as Profile 3, is defined herein to provide enhanced coding efficiency performance and enhanced error resilience for delivery to wireless devices within the feature set available in the second version of this Recommendation (which did not include Annexes U, V, and W). This profile of support is composed of the baseline design plus the following modes:

- 1) **Advanced INTRA Coding (Annex I)** – See X.2.2 item 1.
- 2) **Deblocking Filter (Annex J)** – See X.2.2 item 2.
- 3) **Slice Structured Mode (Annex K)** – The Slice Structured mode is included here due to its enhanced ability to provide resynchronization points within the video bitstream for recovery from erroneous or lost data. Support for the Arbitrary Slice Ordering (ASO) and Rectangular Slice (RS) submodes of the Slice Structured mode are not included in this profile, in order to limit the complexity requirements of the decoder. The additional computational burden imposed by the Slice Structured mode is minimal, limited primarily to bitstream generation and parsing.
- 4) **Modified Quantization (Annex T)** – See X.2.2 item 4.

X.2.5 Version 3 Interactive and Streaming Wireless Profile (Profile 4)

The Version 3 Interactive and Streaming Wireless Profile, designated as Profile 4, is defined herein to provide enhanced coding efficiency performance and enhanced error resilience for delivery to wireless devices, while taking advantage of the enhanced features of the third version of this Recommendation. This profile of support is composed of the baseline design, plus the following additional features as follows:

- 1) **Profile 3** – This feature set provides several enhancements useful for support of wireless video transmission.
- 2) **Data Partitioned Slice Mode (Annex V)** – This feature enhances error resilience performance by separating motion vector data from DCT coefficient data within slices, and protects the motion vector information (the most important part of the detailed macroblock data) by using reversible variable-length coding. Support of the Arbitrary Slice Ordering (ASO) and Rectangular Slice (RS) submodes are not included in this profile, in order to limit the complexity requirements of the decoder.
- 3) **Previous Picture Header Repetition Supplemental Enhancement Information (Annex W, clause W.6.3.8)** – This feature allows the decoder to receive and recover the header information from a previous picture in case of data loss or corruption.

X.2.6 Conversational High Compression Profile (Profile 5)

The Conversational High Compression Profile, designated as Profile 5, is defined herein to provide enhanced coding efficiency performance without adding the delay associated with the use of B pictures and without adding error resilience features. This profile of support is composed of the baseline design, plus the following additional features as follows:

- 1) **Profile 1** – This feature set provides several enhancements useful for enhanced coding efficiency.
- 2) **Profile 2** – This profile adds the Advanced Prediction mode (Annex F), which provides a further enhancement of coding efficiency performance and backward-compatibility with implementations of the first version of this Recommendation.
- 3) **Unrestricted Motion Vectors with UUI = "1" (Annex D)** – Annex D has two primary features:
 - a) picture boundary extrapolation; and
 - b) longer motion vector support.

The first of these features is already supported by the inclusion of Annex J in Profile 1. The longer motion vector support can provide a significant improvement in coding efficiency, especially for large picture sizes, rapid motion, camera movement, and low picture rates. When used with PLUSPTYPE present, this mode also allows for longer motion vector differences, which can significantly simplify encoder operation. The longer motion vectors do present a potential problem for the decoder in terms of memory access, but picture-size-dependent limits on the maximum motion vector size prevent this problem from becoming an appreciable obstacle to implementation.

- 4) **Enhanced Reference Picture Selection (Annex U)** – This mode adds a significant gain in compression efficiency performance due to the ability to use multiple prior pictures as reference data for macroblock-level prediction of the subsequent pictures. The Sub-Picture Removal submode (Annex U, clause U.4.3) of the Enhanced Reference Picture Selection mode is not included in Profile 5.

X.2.7 Conversational Internet Profile (Profile 6)

The Conversational Internet Profile, designated as Profile 6, is defined herein to provide enhanced coding efficiency performance without adding the delay associated with the use of B pictures, but adding some error resilience suitable for use on Internet Protocol (IP) networks (which use packet-based data protocols with relatively large packets and which exhibit data losses rather than data corruption). This profile of support is composed of the baseline design, plus the following additional features as follows:

- 1) **Profile 5** – This feature set provides several enhancements useful for enhanced coding efficiency.
- 2) **Slice Structured mode (Annex K) with Arbitrary Slice Ordering (ASO) submode** – The Slice Structured mode is included here due to its enhanced ability to provide resynchronization points within the video bitstream for recovery from lost data packets. The Arbitrary Slice Ordering (ASO) submode of the Slice Structured mode is also included in order to allow for interleaved packetization for motion-compensated error concealment and for out-of-sequence data reception. Support for the Rectangular Slice (RS) submode of the Slice Structured mode is not included in this profile, in order to limit the complexity requirements of the decoder. The additional computational burden imposed by the Slice Structured mode is minimal, limited primarily to bitstream generation and parsing.

X.2.8 Conversational Interlace Profile (Profile 7)

The Conversational Interlace Profile, designated as Profile 7, is defined herein to provide enhanced coding efficiency performance for low-delay applications, plus support of interlaced video sources. This profile of support is composed of the baseline design, plus the following additional features as follows:

- 1) **Profile 5** – This feature set provides several enhancements useful for enhancing coding efficiency without adding delay.
- 2) **Interlaced Field Indications For 240-line and 288-line Pictures (Annex W, clause W.6.3.11)** – This feature allows video to be sent in an interlaced source picture format for compatibility with existing camera designs.

X.2.9 High Latency Profile (Profile 8)

The High Latency Profile, designated as Profile 8, is defined herein to provide enhanced coding efficiency performance for applications without critical delay constraints. This profile of support is composed of the baseline design, plus the following additional features as follows:

- 1) **Profile 6** – This feature set provides several enhancements useful for enhanced coding efficiency and robustness to data losses.
- 2) **Reference Picture Resampling (Implicit Factor-of-4 Mode Only) (Annex P, clause P.5)** – The implicit factor-of-4 mode of Reference Picture Resampling allows for automatic reference picture resampling only when the size of the new frame is changed, as indicated in the picture header. No bitstream overhead is required for this mode of operation. Predictive dynamic resolution changes allow an encoder to make intelligent trade-offs between temporal and spatial resolution. Furthermore, this simplest mode of operation for Annex P (factor-of-4 upsampling or downsampling only) adds only a modest amount of computational complexity to both the encoder or decoder, since the factor-of-4 case uses a simple fixed FIR filter (requiring roughly 4 operations per pixel, at most).
- 3) **B Pictures (Temporal Scalability, Annex O, clause O.1.1)** – This feature consists of B pictures, which are pictures allowing bidirectional temporal prediction. The addition of B pictures enhances coding efficiency performance, but at some cost in added processing power and encoding and decoding delay. The two-picture backward prediction submode for B pictures in Enhanced Reference Picture Selection mode (Annex U, clause U.3.1.5.5) is not supported in Profile 8.

X.3 Picture formats and picture clock frequencies

To ensure a high quality level of interoperability, encoders and decoders supporting a large standard picture format (QCIF, CIF, 4CIF, 16CIF) should support all smaller standard picture formats. This is a requirement of all decoders conforming to the profiles and levels defined in this annex. (As specified elsewhere in this Recommendation, decoders shall support sub-QCIF and QCIF, and encoders shall support sub-QCIF or QCIF.) For example, a decoder conforming to a profile and level defined in this annex which is capable of decoding 4CIF pictures shall also support the decoding of CIF pictures.

Decoders should be capable of operation with a smaller picture format at maximum picture rates no lower than the maximum picture rate for which it is capable of operation with a larger standard picture format. This is a requirement of all decoders conforming to the profiles and levels defined in this annex. For example, a decoder conforming to a profile and level defined in this annex which is capable of decoding 4CIF pictures at 25 pictures per second shall also be able to decode CIF, QCIF and SQCIF pictures at least at 25 pictures per second.

Encoders and decoders supporting custom picture formats and/or custom picture clock frequencies are recommended to follow the rules defined in this paragraph. These rules are requirements of all decoders conforming to the profiles and levels defined in this annex:

- 1) A decoder for any profile and level defined herein that supports a maximum picture format shall support all standard picture formats smaller or equal in both height and width than those of the maximum supported picture format. For example, a decoder supporting a custom picture format of 720×288 shall support CIF, QCIF and sub-QCIF picture decoding.
- 2) A decoder for any profile and level defined herein that supports custom picture formats shall support all standard or custom picture formats having both height and width smaller than or equal to those of the maximum supported picture format.
- 3) A decoder for any profile and level defined herein that supports a minimum picture interval with the standard picture clock frequency of $(30\ 000)/1001$ units per second shall support the same or smaller minimum picture interval for all supported picture formats having both height and width smaller than or equal to those of the maximum picture format at which the minimum picture interval is specified.
- 4) A decoder for any profile and level defined herein that supports a minimum picture interval and supports custom picture clock frequencies shall support the use of any picture clock frequency with the same or larger picture interval for all supported picture formats having both height and width smaller than or equal to those of the maximum picture format at which the minimum picture interval is specified.

X.4 Levels of performance capability

Eight levels of performance capability are defined for decoder implementation. The Hypothetical Reference Decoder has the minimal size specified in Table X.1 for all levels of Profiles 0 through 4. In Profiles 5 through 8 the Hypothetical Reference Decoder has an increased size and Enhanced Reference Picture Selection is supported with multiple reference pictures. Table X.2 defines the detailed performance parameters of each of these levels:

- 1) **Level 10** – Support of QCIF and sub-QCIF resolution decoding, capable of operation with a bit rate up to 64 000 bits per second with a picture decoding rate up to $(15\ 000)/1001$ pictures per second.
- 2) **Level 20** – Support of CIF, QCIF and sub-QCIF resolution decoding, capable of operation with a bit rate up to $2 \cdot (64\ 000) = 128\ 000$ bits per second with a picture decoding rate up to $(15\ 000)/1001$ pictures per second for CIF pictures and $(30\ 000)/1001$ pictures per second for QCIF and sub-QCIF pictures.
- 3) **Level 30** – Support of CIF, QCIF and sub-QCIF resolution decoding, capable of operation with a bit rate up to $6 \cdot (64\ 000) = 384\ 000$ bits per second with a picture decoding rate up to $(30\ 000)/1001$ pictures per second.
- 4) **Level 40** – Support of CIF, QCIF and sub-QCIF resolution decoding, capable of operation with a bit rate up to $32 \cdot (64\ 000) = 2\ 048\ 000$ bits per second with a picture decoding rate up to $(30\ 000)/1001$ pictures per second.
- 4.5) **Level 45** – Support of QCIF and sub-QCIF resolution decoding, capable of operation with a bit rate up to $2 \cdot (64\ 000) = 128\ 000$ bits per second with a picture decoding rate up to $(15\ 000)/1001$ pictures per second. Additionally, in profiles other than profiles 0 and 2, support of custom picture formats of size QCIF and smaller.
- 5) **Level 50** – Support of custom and standard picture formats of size CIF and smaller, capable of operation with a bit rate up to $64 \cdot (64\ 000) = 4\ 096\ 000$ bits per second with a picture decoding rate up to 50 pictures per second for CIF or smaller picture formats and up to $(60\ 000)/1001$ pictures per second for 352×240 and smaller picture formats.

- 6) **Level 60** – Support of custom and standard picture formats of size 720×288 and smaller, capable of operation with a bit rate up to $128 \cdot (64\ 000) = 8\ 192\ 000$ bits per second with a picture decoding rate up to 50 pictures per second for 720×288 or smaller picture formats and up to $(60\ 000)/1001$ pictures per second for 720×240 and smaller picture formats.
- 7) **Level 70** – Support of custom and standard picture formats of size 720×576 and smaller, capable of operation with a bit rate up to $256 \cdot (64\ 000) = 16\ 384\ 000$ bits per second with a picture decoding rate up to 50 pictures per second for 720×576 or smaller picture formats and up to $(60\ 000)/1001$ pictures per second for 720×480 and smaller picture formats.

The bit rate at which a particular profile and level are used in a system shall never exceed that specified in this annex. However, particular systems may include other means to signal further limits on the bit rate. Other aspects of profile and level capabilities may also be subject to additional capability restrictions when used in particular systems, but the capabilities required for decoding any bitstream for a particular profile and level defined herein shall never exceed those specified in this annex.

Table X.1/H.263 – Summary of profiles

Annex/clause below for profile listed at the right	0	1	2	3	4	5	6	7	8
5.1.5: Custom Picture Format (CPFMT)	L	L	L	L	L	L	L	L	L
5.1.7: Custom Picture Clock Frequency Code (CPCFC)	L	L	L	L	L	L	L	L	L
C: Continuous Presence Multipoint and Video Mux									
D.1: Motion vectors over picture boundaries		X	X	X	X	X	X	X	X
D.2 with UUI = '1' or UUI not present: Extension of the motion vector range						X	X	X	X
D.2 with UUI = '01': Unlimited extension of the motion vector range									
E: Syntax-based Arithmetic Coding									
F.2: Four motion vectors per macroblock		X	X	X	X	X	X	X	X
F.3: Overlapped block motion compensation			X			X	X	X	X
G: PB-Frames									
H: Forward Error Correction (use may be imposed at system level as in ITU-T Rec. H.320)									
I: Advanced Intra Coding		X		X	X	X	X	X	X
J: Deblocking Filter		X		X	X	X	X	X	X
K without submodes: Slice Structured Coding – Without submodes				X	X		X		X
K with ASO: Slice Structured Coding – With Arbitrary Slice Ordering submode							X		X
K with RS: Slice Structured Coding – With Rectangular Slice submode									
L.4: Supplemental Enhancement Full picture freeze		X				X	X	X	X
L: Supplemental Enhancement – Other SEI features									
M: Improved PB-Frames									
N: Reference Picture Selection (and submodes)									
O.1.1 Temporal (B pictures): Temporal, SNR, and Spatial Scalability – B pictures for Temporal Scalability									X

Table X.1/H.263 – Summary of profiles

Annex/clause below for profile listed at the right	0	1	2	3	4	5	6	7	8
O SNR and Spatial: <i>Temporal, SNR, and Spatial Scalability – EI and EP pictures for SNR and Spatial Scalability</i>									
P.5: <i>Reference Picture Resampling – Implicit Factor of Four</i>									X
P: <i>Reference Picture Resampling – More General Resampling</i>									
Q: <i>Reduced Resolution Update</i>									
R: <i>Independent Segment Decoding</i>									
S: <i>Alternative Inter VLC</i>									
T: <i>Modified Quantization</i>		X		X	X	X	X	X	X
U without submodes: <i>Enhanced Reference Picture Selection – Without submodes</i>						X	X	X	X
U with SPR: <i>Enhanced Reference Picture Selection – With Sub-Picture Removal submode</i>									
U with BTPSM: <i>Enhanced Reference Picture Selection – With B-Picture Two-Picture submode</i>									
V: <i>Data Partitioned Slices</i>					X				
W.6.3.8: <i>Additional SEI Specification – Prior Picture Header Repetition</i>					X				
W.6.3.11: <i>Additional SEI Specification – Interlaced Field Indications</i>								X	
W: <i>Additional SEI Specification – Other SEI features</i>									
"X" indicates that support of a feature is part of a profile.									
"L" indicates that the inclusion of a feature depends on the level within the profile.									

Table X.2/H.263 – Levels of operation

Parameter below for level listed at the right	10	20	30	40	45	50	60	70
Max picture format	QCIF (176 × 144)	CIF (352 × 288)	CIF (352 × 288)	CIF (352 × 288)	QCIF (176 × 144) support of CPFMT in profiles other than 0 and 2	CIF (352 × 288) support of CPFMT	CPFMT: 720 × 288 support of CPFMT	CPFMT: 720 × 576 support of CPFMT
Min picture interval	2002/(30 000) s	2002/(30 000) s for CIF 1001/(30 000) s for QCIF and sub-QCIF	1001/(30 000) s	1001/(30 000) s	2002/(30 000) s support of CPCFC in profiles other than 0 and 2	1/50 s at CIF or lower 1001/(60 000) s at 352 × 240 or smaller support of CPCFC	1/50 s at 720 × 288 or lower 1001/(60 000) s at 720 × 240 or smaller support of CPCFC	1/50 s at 720 × 576 or lower 1001/(60 000) s at 720 × 480 or smaller support of CPCFC
Max bit rate in 64 000 bits/s units	1	2	6	32	2	64	128	256
Max HRD B in 16 384 bit units	1: Prof. 5-8	2: Prof. 5-8	6: Prof. 5-8	32: Prof. 5-8	2: Prof. 5-8	64: Prof. 5-8	64: Prof. 5-8	256: Prof. 5-8
Max BPPmaxKb in 1024 bit units	128: Prof. 5-8	512: Prof. 5-8	512: Prof. 5-8	512: Prof. 5-8	128: Prof. 5-8	512: Prof. 5-8	1024: Prof. 5-8	1024: Prof. 5-8
Max ERPS reference pictures (Annex U)	5: Prof. 5-7 10: Prof. 8	5: Prof. 5-7 10: Prof. 8 multiplied by 2 for QCIF or sub-QCIF in Prof. 5-8	5: Prof. 5-7 10: Prof. 8 multiplied by 2 for QCIF or sub-QCIF in Prof. 5-8	5: Prof. 5-7 10: Prof. 8 multiplied by 2 for QCIF or sub-QCIF in Prof. 5-8	5: Prof. 5-7 10: Prof. 8	5: Prof. 5-7 10: Prof. 8 multiplied by 2 for QCIF or smaller in Prof. 5-8	5: Prof. 5-7 10: Prof. 8 multiplied by 2 for CIF or smaller, and by 4 for QCIF or smaller in Prof. 5-8	5: Prof. 5-7 10: Prof. 8 multiplied by 2 for CIF or smaller, and by 4 for QCIF or smaller in Prof. 5-8
NOTE 1 – In profiles for which a maximum number of reference picture buffers is not specified in Table X.2, no support for multiple reference picture buffering is required.								
NOTE 2 – In profiles for which a maximum BPPmaxKb and HRD B are not specified in Table X.2, the minimum value specified in Table X.1 applies for the specified maximum bit rate and resolution.								

X.5 Generic capability definitions for use with ITU-T Rec. H.245

Table X.3 defines a capability identifier for establishing H.263 capabilities for use in systems that use ITU-T Rec. H.245 for capability determination. These parameters shall only be included as **genericVideoCapability** within the **VideoCapability** structure and as **genericVideoMode** within the **VideoMode** structure of ITU-T Rec. H.245. Tables X.4 to X.14 define the associated capability parameters.

When included in Logical Channel Signalling or Mode Request, exactly one parameter with Parameter identifier value in the range zero to eight shall be present; that is, only one profile shall be specified.

Table X.3/H.263 – Capability identifier for H.263 capability

Capability name:	H.263
Capability class:	Video codec
Capability identifier type:	Standard
Capability identifier value:	itu-t (0) recommendation (0) h (8) 263 generic-capabilities (1) 0
MaxBitRate:	The maxBitRate field shall always be included.
NonCollapsingRaw:	This field shall not be included.
Transport:	This field shall not be included.

Table X.4/H.263 – The Baseline Profile (Profile 0) capability

Parameter name:	baselineProfile
Parameter description:	This is a collapsing GenericParameter. baselineProfile indicates the maximum level of support for the Baseline Profile when present in Capability Exchange, the maximum level to be transmitted when present in Logical Channel Signalling, and the desired level when present in Mode Request.
Parameter identifier value:	0
Parameter status:	Mandatory
Parameter type:	unsignedMin
Supersedes:	–

**Table X.5/H.263 – H.320 Coding Efficiency Version 2
Backward-Compatibility Profile (Profile 1) capability**

Parameter name:	h320Profile
Parameter description:	This is a collapsing GenericParameter. h320Profile indicates the maximum level of support for the H.320 Coding Efficiency Version 2 Backward-Compatibility Profile when present in Capability Exchange, the maximum level to be transmitted when present in Logical Channel Signalling, and the desired level when present in Mode Request.
Parameter identifier value:	1
Parameter status:	Optional
Parameter type:	unsignedMin
Supersedes:	–

Table X.6/H.263 – Version 1 Backward-Compatibility Profile (Profile 2) capability

Parameter name:	backwardCompatibleProfile
Parameter description:	This is a collapsing GenericParameter. backwardCompatibleProfile indicates the maximum level of support for the Version 1 Backward-Compatibility Profile when present in Capability Exchange, the maximum level to be transmitted when present in Logical Channel Signalling, and the desired level when present in Mode Request.
Parameter identifier value:	2
Parameter status:	Optional
Parameter type:	unsignedMin
Supersedes:	–

**Table X.7/H.263 – Version 2 Interactive and Streaming
Wireless Profile (Profile 3) capability**

Parameter name:	v2WirelessProfile
Parameter description:	This is a collapsing GenericParameter. v2WirelessProfile indicates the maximum level of support for the Version 2 Interactive and Streaming Wireless Profile when present in Capability Exchange, the maximum level to be transmitted when present in Logical Channel Signalling, and the desired level when present in Mode Request.
Parameter identifier value:	3
Parameter status:	Optional
Parameter type:	unsignedMin
Supersedes:	–

**Table X.8/H.263 – Version 3 Interactive and Streaming
Wireless Profile (Profile 4) capability**

Parameter name:	v3WirelessProfile
Parameter description:	This is a collapsing GenericParameter. v3WirelessProfile indicates the maximum level of support for the Version 3 Interactive and Streaming Wireless Profile when present in Capability Exchange, the maximum level to be transmitted when present in Logical Channel Signalling, and the desired level when present in Mode Request.
Parameter identifier value:	4
Parameter status:	Optional
Parameter type:	unsignedMin
Supersedes:	–

Table X.9/H.263 – Conversational High Compression Profile (Profile 5) capability

Parameter name:	conversationalProfile
Parameter description:	This is a collapsing GenericParameter. conversationalProfile indicates the maximum level of support for the Conversational High Compression Profile when present in Capability Exchange, the maximum level to be transmitted when present in Logical Channel Signalling, and the desired level when present in Mode Request.
Parameter identifier value:	5
Parameter status:	Optional
Parameter type:	unsignedMin
Supersedes:	–

Table X.10/H.263 – Conversational Internet Profile (Profile 6) capability

Parameter name:	conversationalInternetProfile
Parameter description:	This is a collapsing GenericParameter. conversationalInternetProfile indicates the maximum level of support for the Conversational Internet Profile when present in Capability Exchange, the maximum level to be transmitted when present in Logical Channel Signalling, and the desired level when present in Mode Request.
Parameter identifier value:	6
Parameter status:	Optional
Parameter type:	unsignedMin
Supersedes:	–

Table X.11/H.263 – Conversational Plus Interlace Profile (Profile 7) capability

Parameter name:	conversationalInterlaceProfile
Parameter description:	This is a collapsing GenericParameter. conversationalInterlaceProfile indicates the maximum level of support for the Conversational Plus Interlace Profile when present in Capability Exchange, the maximum level to be transmitted when present in Logical Channel Signalling, and the desired level when present in Mode Request.
Parameter identifier value:	7
Parameter status:	Optional
Parameter type:	unsignedMin
Supersedes:	–

Table X.12/H.263 – High Latency Profile (Profile 8) capability

Parameter name:	highLatencyProfile
Parameter description:	This is a collapsing GenericParameter. highLatencyProfile indicates the maximum level of support for the High Latency Profile when present in Capability Exchange, the maximum level to be transmitted when present in Logical Channel Signalling, and the desired level when present in Mode Request.
Parameter identifier value:	8
Parameter status:	Optional
Parameter type:	unsignedMin
Supersedes:	–

Table X.13/H.263 – Temporal Spatial Trade-Off capability

Parameter name:	temporalSpatialTradeOffCapability
Parameter description:	This is a collapsing GenericParameter. The presence of this parameter indicates that the encoder is able to vary its trade-off between temporal and spatial resolution as commanded by the remote terminal. It has no meaning when part of a receive capability.
Parameter identifier value:	9
Parameter status:	Optional
Parameter type:	logical
Supersedes:	–

Table X.14/H.263 – Video Bad Macroblocks capability

Parameter name:	videoBadMBsCap
Parameter description:	This is a collapsing GenericParameter. The presence of this parameter indicates the capability of an encoder to receive or a decoder to transmit the videoBadMBs command. When part of a transmit capability, it indicates the ability of the encoder to process videoBadMBs commands and to take appropriate corrective action toward recovery of video quality. When part of a receive capability, it indicates the ability of the decoder to send appropriate videoBadMBs indications.
Parameter identifier value:	10
Parameter status:	Optional
Parameter type:	logical
Supersedes:	–

Appendix I

Error tracking

I.1 Introduction

This appendix describes a method to recover efficiently after transmission errors if erroneous MBs are reported via a feedback channel to the encoder. The capability of sending and processing feedback information is signalled via external means (for example, by ITU-T Rec. H.245). Furthermore, format and content of the feedback message are defined externally (for example, by ITU-T Rec. H.245).

I.2 Error tracking

Because INTRA coding stops temporal error propagation, it should be used for macroblocks which are severely affected by transmission errors. This requires that the location and extent of image artefacts can be made available to the encoder. The following algorithm provides an estimated error distribution based on feedback information received by the encoder. It considers spatial error propagation caused by motion-compensated prediction as well as the delay until the reception of the feedback message. The algorithm illustrates one possible approach to evaluate feedback messages for spatio-temporal error tracking. Other algorithms are possible.

Assume N macroblocks within each frame enumerated $mb = 1 \dots N$ from top-left to bottom-right. Let $\{n_{err}, mb_{first}, mb_{last}\}$ be the feedback message to the encoder, where $mb_{first} \leq mb \leq mb_{last}$ indicates a set of erroneous macroblocks in frame n_{err} .

To evaluate the feedback message, the encoder must continuously record information during the encoding of each frame. First, the initial error $E_0(mb, n)$ that would be introduced by the loss of macroblock mb in frame n needs to be stored. Assuming a simple error concealment where erroneous macroblocks are treated as not coded, $E_0(mb, n)$ is computed as the Summed Absolute Difference (SAD) of macroblock mb in frame n and $n - 1$. Second, the number of pixels transferred from macroblock mb_{source} in frame $n - 1$ to macroblock mb_{dest} in frame n is stored in dependencies $d(mb_{source}, mb_{dest}, n)$. These dependencies are derived from the motion vectors.

Assume that a feedback message arrives before frame n_{next} is encoded, such that $n_{\text{next}} > n_{\text{err}}$. Then, the estimated error $E(\text{mb}, n_{\text{err}})$ in macroblock mb and frame n_{err} is initialized as:

$$E(\text{mb}, n_{\text{err}}) = \begin{cases} E_0(\text{mb}, n_{\text{err}}) & \text{for } \text{mb}_{\text{first}} \leq \text{mb} \leq \text{mb}_{\text{last}} \\ 0 & \text{else} \end{cases}$$

For subsequent frames n , with $n_{\text{err}} < n < n_{\text{next}}$, the error may be estimated as:

$$E(\text{mb}, n) = \sum_{i=1}^N E(i, n-1) \frac{d(i, \text{mb}, n)}{256}$$

where a uniformly distributed error in each macroblock is assumed after each iteration.

The estimated error $E(\text{mb}, n_{\text{next}} - 1)$ is incorporated into the mode decision of the next frame. For example, macroblock mb is coded in INTRA mode, if $E(\text{mb}, n_{\text{next}} - 1)$ exceeds a threshold.

In practice, error tracking information will only be stored for the latest M frames. Then, if $n_{\text{err}} < n_{\text{next}} - M$, no error-tracking information is available and the encoder must take special action. For example, the next frame may be coded in INTRA mode. However, other procedures are possible and may be more effective.

Appendix II

Recommended optional enhancement

The content of the H.263 Appendix II approved in February 1998 has become obsolete after the approval of H.263 Annex X. Since this appendix is used as a reference by many users of ITU-T texts, this appendix is provided mainly as a reference for users unaware of the contents of H.263 Annex X.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems