

Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in DVB services delivered directly over IP protocols



Reference

RTS/JTC-DVB-263

Keywords

3GPP, audio, broadcasting, digital, DVB, IP,
MPEG, TV, video

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Individual copies of the present document can be downloaded from:

<http://www.etsi.org>

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at

<http://portal.etsi.org/tb/status/status.asp>

If you find errors in the present document, please send your comment to one of the following services:

http://portal.etsi.org/chaicor/ETSI_support.asp

Copyright Notification

No part may be reproduced except as authorized by written permission.
The copyright and the foregoing restriction extend to reproduction in all media.

© European Telecommunications Standards Institute 2010.
© European Broadcasting Union 2010.
All rights reserved.

DECT™, **PLUGTESTS™**, **UMTS™**, **TIPHON™**, the TIPHON logo and the ETSI logo are Trade Marks of ETSI registered for the benefit of its Members.

3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners.

LTE™ is a Trade Mark of ETSI currently being registered for the benefit of its Members and of the 3GPP Organizational Partners.

GSM® and the GSM logo are Trade Marks registered and owned by the GSM Association.

Contents

Intellectual Property Rights	7
Foreword.....	7
Introduction	7
1 Scope	10
2 References	10
2.1 Normative references	10
2.2 Informative references.....	11
3 Definitions and abbreviations.....	12
3.1 Definitions	12
3.2 Abbreviations	12
4 Systems layer.....	14
4.1 Transport over IP Networks/RTP Packetization Formats.....	14
4.1.1 RTP packetization of H.264/AVC	14
4.1.2 RTP packetization of SVC.....	14
4.1.3 RTP packetization of VC-1.....	14
4.1.4 RTP packetization of MPEG-4 HE AAC v2	15
4.1.5 RTP packetization of MPEG-4 HE AAC v2 in combination with MPEG Surround.....	15
4.1.6 RTP packetization of AMR-WB+	15
4.1.7 RTP packetization of AC-3.....	15
4.1.8 RTP packetization of Enhanced AC-3.....	15
4.2 File storage for download services	16
4.2.1 MP4 files.....	16
4.2.1.1 MP4 file storage of H.264/AVC	16
4.2.1.2 MP4 file storage of SVC	17
4.2.1.3 MP4 file storage of VC-1	17
4.2.1.4 MP4 file storage of MPEG-4 HE AAC v2 in combination with MPEG Surround	17
4.2.1.5 MP4 file storage of AC-3 and Enhanced AC-3.....	17
4.2.2 3GP files	17
4.2.2.1 3GP file storage of H.264/AVC	17
4.2.2.2 3GP file storage of VC-1.....	18
4.2.2.3 3GP file storage of MPEG-HE AAC v2 in combination with MPEG Surround.....	18
5 Video.....	18
5.1 H.264/AVC video.....	18
5.1.1 Profile and level	18
5.1.2 Video usability information	19
5.1.3 Frame rate	20
5.1.4 Aspect ratio	20
5.1.5 Luminance resolution	20
5.1.6 Chromaticity	20
5.1.7 Chrominance format	20
5.1.8 Random access points.....	20
5.1.8.1 Definition	20
5.1.8.2 Time Interval between RAPs	21
5.1.9 Sequence parameter sets and picture parameter sets.....	21
5.1.10 Active Format Description.....	21
5.2 SVC video	21
5.2.1 General.....	22
5.2.1.1 Profile and level	22
5.2.1.2 Backward Compatibility	23
5.2.1.3 Reference Base Pictures.....	24
5.2.2 Video usability information	24
5.2.2.1 Sequence parameter sets	24
5.2.2.2 Subset sequence parameter sets.....	24

5.2.3	Frame rate	25
5.2.4	Aspect ratio	25
5.2.5	Luminance resolution	25
5.2.6	Chromaticity	25
5.2.7	Chrominance format	25
5.2.8	Active Format Description.....	26
5.2.9	SVC Random access points	26
5.2.9.1	Definition	26
5.2.9.2	Time Interval between SVC RAPs.....	27
5.2.10	Sequence parameter sets and picture parameter sets.....	27
5.3	VC-1 video	28
5.3.1	Profile and level	28
5.3.2	Frame rate	28
5.3.3	Aspect ratio	29
5.3.4	Luminance resolution	29
5.3.5	Chromaticity	29
5.3.6	Random access points	29
5.3.7	Active Format Description.....	30
6	Audio	30
6.1	MPEG-4 AAC, HE AAC, HE AAC v2 and MPEG Surround audio	30
6.1.1	Audio mode	31
6.1.2	Profiles	31
6.1.3	Bit rate	32
6.1.4	Sampling frequency	32
6.1.5	Dynamic range control.....	32
6.1.6	Matrix downmix	32
6.2	AMR-WB+ audio	32
6.2.1	Audio mode	32
6.2.2	Sampling frequency	32
6.3	AC-3 audio	32
6.3.1	Audio mode	33
6.3.2	Bit rate	33
6.3.3	Sampling frequency	33
6.4	Enhanced AC-3 audio	33
6.4.1	Audio mode	33
6.4.2	Substreams	33
6.4.3	Bit rate	33
6.4.4	Sampling frequency	34
6.4.5	Stream mixing.....	34
Annex A (informative): Description of the implementation guidelines		35
A.1	Introduction	35
A.2	Systems.....	36
A.2.1	Protocol stack	36
A.2.2	Transport of H.264/AVC video.....	36
A.2.3	Transport of SVC video	36
A.2.4	Transport of VC-1 video	37
A.2.5	Transport of MPEG-4 HE AAC v2 audio	37
A.2.6	Transport of MPEG-4 HE AAC v2 in combination with MPEG Surround audio.....	37
A.2.7	Transport of AMR-WB+ audio	38
A.2.8	Transport of AC-3 audio	38
A.2.9	Transport of Enhanced AC-3 audio.....	38
A.2.10	Synchronization of content delivered over IP	39
A.2.11	Synchronization with content delivered over MPEG-2 TS	39
A.2.12	Service discovery	40
A.2.13	Linking to applications	40
A.2.14	Capability exchange	40
A.3	Video	40
A.3.1	H.264/AVC video.....	40

A.3.1.1	Overview	40
A.3.1.2	Network Abstraction Layer (NAL).....	41
A.3.1.3	Video Coding Layer (VCL).....	41
A.3.1.4	Explanation of H.264/AVC profiles and levels	43
A.3.1.5	Summary of key tools and parameter ranges for capability A to F IRDs	44
A.3.1.6	Other video parameters	45
A.3.2	SVC video	46
A.3.2.1	Overview	46
A.3.2.2	Network Abstraction Layer (NAL).....	46
A.3.2.3	Video Coding Layer (VCL).....	47
A.3.2.4	Explanation of SVC profiles and levels.....	49
A.3.2.5	Summary of key tools and parameter ranges for capability B to F IRDs	50
A.3.2.6	Other video parameters	51
A.3.3	VC-1 video	51
A.3.3.1	Overview	51
A.3.3.2	Explanation of VC-1 profiles and levels.....	51
A.3.3.3	Summary of key tools and parameter ranges for capability A to E IRDs	52
A.4	Audio.....	53
A.4.1	MPEG-4 AAC, HE AAC, HE AAC v2, and MPEG Surround	53
A.4.1.1	MPEG-4 AAC, HE AAC, HE AAC v2 and MPEG Surround Levels and Main Parameters for DVB	55
A.4.1.2	Methods for signalling SBR and/or PS data	56
A.4.1.3	Methods for signalling MPEG Surround data.....	57
A.4.2	Extended AMR-WB (AMR-WB+)	57
A.4.2.1	Main AMR-WB+ parameters for DVB	58
A.4.3	AC-3.....	58
A.4.4	Enhanced AC-3	60
A.5	The DVB IP datacast application	61
A.6	Future work	61
Annex B (normative):	TS 102 005 usage in DVB IP datacast.....	62
B.1	Scope	62
B.2	Introduction	62
B.3	Systems layer.....	62
B.3.1	Transport over IP networks/RTP packetization formats.....	62
B.3.1.1	Further constraints on RTP packetization of MPEG-4 HE AAC v2.....	62
B.3.1.2	Further constraints on RTP packetization of MPEG-4 HE AAC v2 in combination with MPEG Surround	63
B.3.2	File storage for download services	63
B.4	Video.....	63
B.4.1	H.264/AVC	63
B.4.1.1	Profile and level	63
B.4.1.2	Sample aspect ratio	63
B.4.1.3	Frame rate, luminance resolution, and picture aspect ratio.....	64
B.4.1.4	Chromaticity	64
B.4.1.5	Chrominance format	64
B.4.1.6	Random access points.....	64
B.4.1.7	Output latency.....	64
B.4.1.8	Active Format Description.....	64
B.4.2	VC-1	65
B.4.2.1	Profile and level	65
B.4.2.2	Bit rate	65
B.4.2.3	Sample aspect ratio	65
B.4.2.4	Frame rate, luminance resolution and picture aspect ratio.....	65
B.4.2.5	Chromaticity	65
B.4.2.6	Random Access Points.....	66
B.4.2.7	Active Format Description.....	66
B.5	Audio.....	66

B.5.1	MPEG-4 HE AAC v2 audio and MPEG-4 HE AAC v2 audio in combination with MPEG Surround.....	66
B.5.1.1	Audio mode	66
B.5.1.2	Profiles	66
B.5.1.3	Bit rate	66
B.5.1.4	Sampling frequency	66
B.5.1.5	Dynamic range control.....	66
B.5.1.6	Matrix downmix	67
B.5.2	AMR-WB+ audio	67
Annex C (informative):	Bibliography.....	68
History		69

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: "*Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards*", which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (<http://webapp.etsi.org/IPR/home.asp>).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Specification (TS) has been produced by Joint Technical Committee (JTC) Broadcast of the European Broadcasting Union (EBU), Comité Européen de Normalisation ELECTrotechnique (CENELEC) and the European Telecommunications Standards Institute (ETSI).

NOTE: The EBU/ETSI JTC Broadcast was established in 1990 to co-ordinate the drafting of standards in the specific field of broadcasting and related fields. Since 1995 the JTC Broadcast became a tripartite body by including in the Memorandum of Understanding also CENELEC, which is responsible for the standardization of radio and television receivers. The EBU is a professional association of broadcasting organizations whose work includes the co-ordination of its members' activities in the technical, legal, programme-making and programme-exchange domains. The EBU has active members in about 60 countries in the European broadcasting area; its headquarters is in Geneva.

European Broadcasting Union
CH-1218 GRAND SACONNEX (Geneva)
Switzerland
Tel: +41 22 717 21 11
Fax: +41 22 717 24 81

The Digital Video Broadcasting Project (DVB) is an industry-led consortium of broadcasters, manufacturers, network operators, software developers, regulatory bodies, content owners and others committed to designing global standards for the delivery of digital television and data services. DVB fosters market driven solutions that meet the needs and economic circumstances of broadcast industry stakeholders and consumers. DVB standards cover all aspects of digital television from transmission through interfacing, conditional access and interactivity for digital video, audio and data. The consortium came together in 1993 to provide global standardisation, interoperability and future proof specifications.

Introduction

The present document addresses the use of video and audio coding in DVB services delivered over IP protocols. It specifies the use of H.264/AVC video and SVC video as specified in ITU-T Recommendation H.264 [1] and ISO/IEC 14496-10 [2], VC-1 video as specified in SMPTE 421M [10], MPEG-4 AAC/HE AAC/HE AAC v2 audio as specified in ISO/IEC 14496-3 [3], MPEG Surround audio as specified in ISO/IEC 23003-1 [20], Extended AMR-WB (AMR-WB+) audio as specified in TS 126 290 [8] and AC-3 and Enhanced AC-3 audio as specified in TS 102 366 [13].

The present document adopts a "toolbox" approach for the general case of DVB applications delivered directly over IP. A common generic toolbox is used by all DVB services, where each DVB application can select the most appropriate tool from within that toolbox. Annex B of the present document specifies application-specific constraints on the use of the toolbox for the particular case of DVB IP Datacast services.

Clauses 4 to 6 of the present document provide the Digital Video Broadcasting (DVB) specifications for the systems, video, and audio layer, respectively. For information, some of the key features are summarized below, but clauses 4 to 6 should be consulted for all normative specifications:

Systems:

- H.264/AVC, SVC, VC-1, MPEG-4 AAC/HE AAC/HE AAC v2, MPEG Surround, AMR-WB+, AC-3 and Enhanced AC-3 encoded data is delivered over IP in RTP packets.

Video:

The following hierarchical classification of IP-IRDs is specified through Capability categorization of the video codec:

- **Capability A IP-IRDs** are capable of decoding either bitstreams conforming to H.264/AVC Baseline profile at Level 1b with `constraint_set1_flag` being equal to 1 as specified in [1] or else bitstreams conforming to VC-1 Simple Profile at level LL as specified in [10] or else both.
- **Capability B IP-IRDs** are capable of decoding either bitstreams conforming to H.264/AVC Baseline profile at Level 1.2 with `constraint_set1_flag` being equal to 1 as specified in [1] or else bitstreams conforming to H.264/AVC Scalable Baseline profile at Level 1.2 as specified in [1] or else bitstreams conforming to VC-1 Simple Profile at level ML as specified in [10] or else any combination of those. Capability B IP-IRDs that are capable of decoding bitstreams conforming to H.264/AVC Scalable Baseline profile at Level 1.2 are also capable of decoding bitstreams conforming to H.264/AVC Baseline profile at Level 1.2 with `constraint_set1_flag` being equal to 1.
- **Capability C IP-IRDs** are capable of decoding either bitstreams conforming to H.264/AVC Baseline profile at Level 2 with `constraint_set1_flag` being equal to 1 as specified in [1] or else bitstreams conforming to H.264/AVC Scalable Baseline profile at Level 2 as specified in [1] or else bitstreams conforming to VC-1 Advanced Profile at level L0 as specified in [10] or else any combination of those. Capability C IP-IRDs that are capable of decoding bitstreams conforming to H.264/AVC Scalable Baseline profile at Level 2 are also capable of decoding bitstreams conforming to H.264/AVC Baseline profile at Level 2 with `constraint_set1_flag` being equal to 1.
- **Capability DB IP-IRDs** are capable of decoding either bitstreams conforming to H.264/AVC Baseline profile at Level 3 with `constraint_set1_flag` being equal to 1 as specified in [1] or else bitstreams conforming to H.264/AVC Scalable Baseline profile at Level 3 as specified in [1] or else bitstreams conforming to VC-1 Advanced Profile at level L0 as specified in [10] or else any combination of those. Capability DB IP-IRDs that are capable of decoding bitstreams conforming to H.264/AVC Scalable Baseline profile at Level 3 are also capable of decoding bitstreams conforming to H.264/AVC Baseline profile at Level 3 with `constraint_set1_flag` being equal to 1.
- **Capability D IP-IRDs** are capable of decoding either bitstreams conforming to H.264/AVC Main profile at level 3 as specified in [1] (and optionally capable of decoding bitstreams conforming to H.264/AVC High profile at level 3 as specified in [1]) or else bitstreams conforming to H.264/AVC Scalable High profile at Level 3 as specified in [1] or else bitstreams conforming to VC-1 Advanced Profile at level L1 as specified in [10] or else any combinations of those. Capability D IP-IRDs that are capable of decoding bitstreams conforming to H.264/AVC Scalable High profile at Level 3 are also capable of decoding bitstreams conforming to H.264/AVC Main profile at Level 3 and bitstreams conforming to H.264/AVC High profile at Level 3.
- **Capability E IP-IRDs** are capable of decoding either bitstreams conforming to H.264/AVC High profile at level 4 as specified in [1] or else bitstreams conforming to H.264/AVC Scalable High profile at Level 4 as specified in [1] or else bitstreams conforming to VC-1 Advanced Profile at level L3 as specified in [10] or any combination of those. Capability E IP-IRDs that are capable of decoding bitstreams conforming to H.264/AVC Scalable High profile at Level 4 are also capable of decoding bitstreams conforming to H.264/AVC High profile at Level 4.
- **Capability F IP-IRDs** are capable of decoding either bitstreams conforming to H.264/AVC High profile at Level 4.2 as specified in [1] or else bitstreams conforming to H.264/AVC Scalable High profile at Level 4.2 as specified in [1] or else bitstreams conforming to VC-1 Advanced Profile at level L3 as specified in [10] or any combination of those. Capability F IP-IRDs that are capable of decoding bitstreams conforming to H.264/AVC Scalable High profile at Level 4.2 are also capable of decoding bitstreams conforming to H.264/AVC High profile at Level 4.2.

- IP-IRDs labelled with a particular capability Y are also capable of decoding H.264/AVC, SVC and/or VC-1 bitstreams that can be decoded by IP-IRDs labelled with a particular capability X, when X appears in the following ordered sequence at an earlier position than Y: A, B, C, DB, D, E, F. For instance, a Capability D IP-IRD that is capable of decoding bitstreams conforming to Main Profile at level 3 of H.264/AVC will additionally be capable of decoding H.264/AVC bitstreams that are also decodable by IP-IRDs with capabilities A, B, C or DB.
- It is possible that an IP-IRD may support the decoding of H.264/AVC at Capability M, SVC at Capability N (less than or equal to M) and VC-1 at Capability O where M, N and O are not the same.

Audio:

- IP-IRDs are capable of decoding bitstreams either conforming to the MPEG-4 HE AAC v2 Profile, or else bitstreams conforming to the MPEG-4 HE AAC v2 profile in combination with the MPEG Surround Baseline Profile, or else bitstreams conforming to AMR-WB+, or else bitstreams conforming to AC-3, or else bitstreams conforming to Enhanced AC-3, or any combination of the five.
- Sampling rates between 8 kHz and 48 kHz are supported by IP-IRDs.
- IP-IRDs are capable of decoding mono, parametric stereo (when the MPEG-4 HE AAC v2 Profile is used) and 2-channel stereo audio bitstreams. IP-IRDs may be capable of decoding multi-channel bitstreams.

An IP-IRD of one of the capability classes A to F above meets the minimum functionality, as specified in the present document, for decoding H.264/AVC, SVC or VC-1 video and for decoding MPEG-4 HE AAC v2, MPEG-4 HE AAC v2 in combination with MPEG Surround, AMR-WB+, AC-3 or Enhanced AC-3 audio delivered over an IP network. The specification of this minimum functionality in no way prohibits IP-IRD manufacturers from including additional features, and should not be interpreted as stipulating any form of upper limit to the performance.

Where an IP-IRD feature described in the present document is mandatory, the word "shall" is used and the text is in *italic*; all other features are optional. The specifications presented for IP-IRDs observe the following principles:

- IP-IRDs allow for future compatible extensions to the bit-stream syntax;
- all "reserved", "unspecified", and "private" bits in H.264/AVC, SVC, VC-1, MPEG-4 AAC / HE AAC / HE AAC v2, MPEG Surround, AMR-WB+, AC-3, Enhanced AC-3 and IP protocols are ignored by IP-IRDs not designed to make use of them.

The rules of operation for the encoders are features and constraints which the encoding system should adhere to in order to ensure that the transmissions can be correctly decoded. These constraints may be mandatory or optional. Where a feature or constraint is mandatory, the word "shall" is used and the text is *italic*; all other features are optional.

1 Scope

The present document specifies the use of H.264/AVC, SVC, VC-1, MPEG-4 AAC / HE AAC / HE AAC v2, MPEG Surround, AMR-WB+, AC-3 and Enhanced AC-3 for DVB conforming delivery in RTP packets over IP networks. The decoding of H.264/AVC, SVC, VC-1, MPEG-4 HE AAC v2, MPEG Surround, AMR-WB+, AC-3 and Enhanced AC-3 in IP-IRDs is specified, as well as rules of operation that encoders must apply to ensure that transmissions can be correctly decoded. These specifications may be mandatory, recommended or optional.

Annex A of the present document provides an informative description for the normative contents of the present document and the specified codecs.

Annex B of the present document defines application-specific constraints on the use of H.264/AVC, SVC, VC-1, MPEG-4 HE AAC v2, MPEG-4 HE AAC v2 in combination with MPEG Surround and AMR-WB+ for DVB IP Datacast services.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific.

- For a specific reference, subsequent revisions do not apply.
- Non-specific reference may be made only to a complete document or a part thereof and only in the following cases:
 - if it is accepted that it will be possible to use all future changes of the referenced document for the purposes of the referring document;
 - for informative references.

Referenced documents which are not found to be publicly available in the expected location might be found at <http://docbox.etsi.org/Reference>.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

2.1 Normative references

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

- [1] ITU-T Recommendation H.264: "Advanced video coding for generic audiovisual services".
- [2] ISO/IEC 14496-10:2009: "Information Technology - Coding of audio-visual objects - Part 10: Advanced Video Coding".
- [3] ISO/IEC 14496-3:2005: "Information technology - Coding of audio-visual objects - Part 3: Audio", including ISO/IEC 14496-3:2005/AMD.2:2006, ISO/IEC 14496-3:2005/AMD.5:2007 and all relevant Corrigenda.
- [4] IETF RFC 3550: "RTP: A Transport Protocol for Real-Time Applications".
- [5] IETF RFC 3640: "RTP Payload Format for Transport of MPEG-4 Elementary Streams".
- [6] IETF RFC 3984: "RTP Payload Format for H.264 Video".
- [7] ETSI TS 126 244: "Universal Mobile Telecommunications System (UMTS); Transparent end-to-end packet switched streaming service (PSS); 3GPP file format (3GP) (3GPP TS 26.244 Release 6)".

- [8] ETSI TS 126 290: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); Audio codec processing functions; Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec; Transcoding functions (3GPP TS 26.290 Release 6)".
- [9] IETF RFC 4352: "RTP Payload Format for Extended Adaptive Multi-Rate Wideband (AMR-WB+) Audio Codec".
- [10] SMPTE 421M: "VC-1 Compressed Video Bitstream Format and Decoding Process".
- [11] IETF RFC 4425: "RTP Payload Format for Video Codec 1 (VC-1)".
- [12] SMPTE RP2025-2007: "VC-1 Bitstream Storage in the ISO Base Media File".
- [13] ETSI TS 102 366: "Digital Audio Compression (AC-3, Enhanced AC-3) Standard".
- [14] IETF RFC 4184: "RTP Payload Format for AC-3 Audio".
- [15] IETF RFC 4598: "Real-time Transport Protocol (RTP) Payload Format for Enhanced AC-3 (E-AC-3) Audio".
- [16] ISO/IEC 14496-12:2005: "Information Technology - Coding of Audio-Visual Objects - Part 12: ISO base media file format".
- [17] Void.
- [18] Void.
- [19] ISO/IEC 14496-15:2004/Amd.2:2008: "Information Technology - Coding of Audio-Visual Objects - Part 15: Advanced Video Coding (AVC) file format Amd. 2: File format support for Scalable Video Coding (SVC)".
- [20] ISO/IEC 23003-1:2007: "Information technology - MPEG audio technologies - Part 1: MPEG Surround", including ISO/IEC 23003-1:2007/Cor:2008 "Information Technology - MPEG audio technologies - Part 1: MPEG Surround, TECHNICAL CORRIGENDUM 1".
- [21] IETF RFC 5691: "RTP Payload Format for Elementary Streams with MPEG Surround Multi-Channel Audio".

2.2 Informative references

The following referenced documents are not essential to the use of the ETSI deliverable but they assist the user with regard to a particular subject area. For non-specific references, the latest version of the referenced document (including any amendments) applies.

- [i.1] IETF RFC 2250: "RTP Payload Format for MPEG1/MPEG2 Video".
- [i.2] ETSI TS 101 154: "Digital Video Broadcasting (DVB); Implementation guidelines for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream".
- [i.3] ETSI TS 102 154: "Digital Video Broadcasting (DVB); Implementation guidelines for the use of Video and Audio Coding in Contribution and Primary Distribution Applications based on the MPEG-2 Transport Stream".
- [i.4] EBU Recommendation R.68: "Alignment level in digital audio production equipment and in digital audio recorders".
- [i.5] ETSI TS 126 234: "Universal Mobile Telecommunications System (UMTS); Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs (3GPP TS 26.234 Release 6)".
- [i.6] ETSI TS 126 273: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); ANSI-C code for the fixed-point Extended Adaptive Multi-Rate - Wideband (AMR-WB+) speech codec (3GPP TS 26.273 Release 6)".

- [i.7] ETSI TS 126 304: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec; Floating-point ANSI-C code (3GPP TS 26.304 Release 6)".
- [i.8] ETSI TS 126 346: "Universal Mobile Telecommunications System (UMTS); Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs (3GPP TS 26.346 Release 6)".
- [i.9] ITU-R Recommendation BT.709: "Parameter values for the HDTV standards for production and international programme exchange".
- [i.10] ISO/IEC 14496-3:2009/FDAM 1:2009: "Information technology - Generic coding of moving picture and associated audio information - Part 3: Audio, Final Draft Amendment 1 - HD-AAC Profile, MPEG Surround signalling".

NOTE: Available at http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=53944.

- [i.11] ETSI TS 102 468: "Digital Video Broadcasting (DVB); IP Datacast over DVB-H: Set of Specifications for Phase 1".
- [i.12] ISO/IEC JTC 1 SC 29: "Coding of audio, picture, multimedia and hypermedia information".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

3GP file: file based on 3GPP file format [7] and its extensions and typically having a .3gp extension in its filename

bitstream: coded representation of a video or audio signal

DVB IP datacast application: application that complies with the DVB IP Datacast Umbrella Specification

IP-IRD: integrated Receiver-Decoder for DVB services delivered over IP categorized by a video decoding and rendering capability

MP4 File: file based on ISO base media file format [16] and its extensions and typically having a .mp4 extension in its filename

multi-channel audio: audio signal with more than two channels

streaming delivery session: instance of delivery of a streaming service which is characterized by a start and end time and addresses of the IP flows used for delivery of the media streams between start and end time

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

3GPP	Third Generation Partnership Project
AAC LC	Advanced Audio Coding Low Complexity
AAC	Advanced Audio Coding
AC-3	Dolby AC-3 audio coding system
ACELP	Algebraic Code Excited Linear Prediction
AMR-WB	Adaptive Multi-Rate-WideBand
AMR-WB+	Extended AMR-WB
AOT	Audio Object Type
ASO	Arbitrary Slice Ordering
AU	Access Unit
AVC	Advanced Video Coding
BWE	BandWidth Extension
CABAC	Context Adaptive Binary Arithmetic Coding

CAVLC	Content-Adaptive Variable-Length Coding
CIF	Common Interchange Format
DQId	(16*dependency_id+quality_id)
DRC	Dynamic Range Control
DVB	Digital Video Broadcasting
DVB-H	DVB-Handheld
FMO	Flexible Macroblock Ordering
GOP	Group of Picture
H.264/AVC	H.264/Advanced Video Coding, excluding SVC
HDTV	High Definition TeleVision
HE AAC	High-Efficiency Advanced Audio Coding
IDR	Instantaneous Decoding Refresh
IP	Internet Protocol
IPDC	IP Data Casting
IPI	IP Infrastructure
IRD	Integrated Receiver-Decoder
LC	Low Complexity
LF	Low Frequency
LL	Low Level
MBMS	Multimedia Broadcast/Multicast Service
ML	Medium Level
MPEG	Moving Pictures Experts Group

NOTE: See ISO/IEC JTC 1/SC 29/WG 11 [i.12].

MPS	MPEG Surround
MTU	Maximum Transmission Unit
NAL	Network Abstraction Layer
NTP	Network Time Protocol
PCM	Pulse-code modulation
PPS	Picture Parameter Set
PS	Parametric Stereo
PSS	Packet switched Streaming Service
QCIF	Quarter Common Interchange Format
QMF	Quadrature Mirror Filter
RAP	Random Access Point
RTCP	RTP Control Protocol
RTP	Real-time Transport Protocol
RTSP	Real-Time Streaming Protocol
S/PDIF	Sony/Philips Digital Interconnect Format
SAC	Spatial Audio Coding
SBR	Spectral Band Replication
SDP	Session Description Protocol
SEI	Supplemental Enhancement Information
SPS	Sequence Parameter Set
SR	Sender Report
SVC	Scalable Video Coding

NOTE: As specified in annex G of ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2].

TCP	Transmission Control Protocol
TCX	Transform Coded Excitation
TS	Transport Stream
UDP	User Datagram Protocol
VBR	Variable BitRate
VC-1	Advanced Video Coding

NOTE: According to SMPTE Standard 421M.

VCL	Video Coding Layer
VUI	Video Usability Information

4 Systems layer

The IP-IRD design should be made under the assumption that any legal structure as permitted RTP packets may occur, even if presently reserved or unused. *To allow full upward compatibility with future enhanced versions, a DVB IP-IRD shall be able to skip over data structures which are currently "reserved", or which correspond to functions not implemented by the IP-IRD. For example, an IP-IRD shall allow the presence of unknown MIME format parameters for RFC payloads, while ignoring its meaning.*

Annex B defines application-specific constraints for DVB IP Datacast services.

4.1 Transport over IP Networks/RTP Packetization Formats

When H.264/AVC, SVC, VC-1, MPEG-4 AAC / HE AAC / HE AAC v2, MPEG Surround, AMR-WB+, AC-3 and Enhanced AC-3 data is transported over IP networks, RTP, a Transport Protocol for Real-Time Applications as defined in RFC 3550 [4], shall be used. This clause specifies the transport of H.264/AVC, SVC, VC-1, MPEG-4 AAC / HE AAC / HE AAC v2, MPEG Surround, AMR-WB+, AC-3 and Enhanced AC-3 in RTP packets for delivery over IP networks and for decoding of such RTP packets in the IP-IRD.

The specification for the use of video and audio coding in broadcasting applications based on the MPEG-2 Transport Stream is given in TS 101 154 [i.2], whilst that for contribution and primary distribution applications is given in TS 102 154 [i.3]. RFC 2250 [i.1] is used for the transport of an MPEG-2 TS in RTP packets over IP.

While the general RTP specification is defined in RFC 3550 [4], RTP payload formats are codec specific and defined in separate RFCs. The specific formats of the RTP packets are specified in clause 4.1.1 for H.264/AVC, clause 4.1.2 for SVC, clause 4.1.3 for VC-1, clause 4.1.4 for MPEG-4 AAC / HE AAC / HE AAC v2, clause 4.1.5 for MPEG-4 HE AAC v2 in combination with MPEG Surround, clause 4.1.6 for AMR-WB+, clause 4.1.7 for AC-3 and clause 4.1.8 for Enhanced AC-3.

4.1.1 RTP packetization of H.264/AVC

For transport over IP, the H.264/AVC data is packetized in RTP packets using RFC 3984 [6].

Encoding: *RFC 3984 [6] shall be used for packetization into RTP.*

Decoding: *An IP-IRD that supports H.264/AVC shall be able to receive RTP packets with H.264/AVC data as defined in RFC 3984 [6].*

4.1.2 RTP packetization of SVC

For transport over IP, the SVC data is packetized in RTP packets using RFC 2250 [i.1].

Encoding: *RFC 2250 [i.1] shall be used for packetization into RTP.*

Decoding: *An IP-IRD that supports SVC shall be able to receive RTP packets with SVC data as defined in RFC 2250 [i.1].*

4.1.3 RTP packetization of VC-1

For transport over IP, the VC-1 data is packetized in RTP packets using RFC 4425 [11].

Encoding: *RFC 4425 [11] shall be used for packetization into RTP.*

Decoding: *An IP-IRD that supports VC-1 shall be able to receive RTP packets with VC-1 data as defined in RFC 4425 [11].*

4.1.4 RTP packetization of MPEG-4 HE AAC v2

For transport over IP, the MPEG-4 HE AAC v2 data is packetized in RTP packets using RFC 3640 [5].

Encoding: *RFC 3640 [5] shall be used for packetization into RTP.*

Decoding: *An IP-IRD that supports MPEG-4 HE AAC v2 shall support RFC 3640 [5] to receive MPEG-4 HE AAC v2 data contained in RTP packets.*

4.1.5 RTP packetization of MPEG-4 HE AAC v2 in combination with MPEG Surround

For transport over IP, the combination of MPEG-4 HE AAC v2 and MPEG Surround data is packetized in RTP packets using RFC 3640 [5] and RFC 5691 [21].

Encoding: *RFC 3640 [5] and RFC 5691 [21] shall be used for packetization in RTP.*

Decoding: *An IP-IRD that supports MPEG-4 HE AAC v2 in combination with MPEG Surround shall support RFC 3640 [5] and RFC 5691 [21] to receive MPEG-4 HE AAC v2 and MPEG Surround data contained in RTP packets.*

4.1.6 RTP packetization of AMR-WB+

For transport over IP, the AMR-WB+ data is packetized in RTP packets using RFC 4352 [9].

Encoding: *RFC 4352 [9] shall be used for packetization in RTP.*

Decoding: *An IP-IRD that supports AMR-WB+ shall support [9] to receive AMR-WB+ data contained in RTP packets.*

4.1.7 RTP packetization of AC-3

For transport over IP, the AC-3 data is packetized in RTP packets using RFC 4184 [14].

Encoding: *RFC 4184 [14] shall be used for packetization in RTP.*

Decoding: *An IP-IRD that supports AC-3 shall support [14] to receive AC-3 data contained in RTP packets.*

4.1.8 RTP packetization of Enhanced AC-3

For transport over IP, the Enhanced AC-3 data is packetized in RTP packets using RFC 4598 [15].

Encoding: *RFC 4598 [15] shall be used for packetization in RTP.*

Decoding: *An IP-IRD that supports Enhanced AC-3 shall support [15] to receive Enhanced AC-3 data contained in RTP packets.*

4.2 File storage for download services

4.2.1 MP4 files

This clause describes usage of MP4 files based on ISO base media file format [16] in download services supporting this feature.

Encoding: *The MP4 file shall be created according to the MPEG-4 Part 12 [16] specification with the constraints described below.*

Zero or one video track and one audio track shall be stored in the file for default presentation of contents.

The default video track (if present) shall contain Video Elementary Stream for used media format. The default audio track shall contain Audio Elementary Stream for used media format.

The default video track (if present) shall have the lowest track ID among the video tracks stored in the file. The default audio track shall have the lowest track ID among the audio tracks stored in the file.

For the default video track (if present) and the default audio track, "Track_enabled" shall be set to the value of 1 in the "flags" field of Track Header Box of the track.

The "moov" box shall be positioned after the "ftyp" box before the first "mdat". If a "moof" box is present, it shall be positioned before the corresponding "mdat" box.

Within a track, chunks shall be in decoding time order within the media data box "mdat".

Video and audio tracks shall be organized as interleaved chunks. The duration of samples stored in a chunk shall not exceed 1 second.

If the size of "moov" box becomes bigger than 1Mbytes, the file shall be fragmented by using moof header. The size of "moov" box shall be equal to or less than 1 Mbytes. The size of "moof" boxes shall be equal to or less than 300 kbit/s.

For video, random accessible samples should be stored as the first sample of each "traf". In the case of gradual decoder refresh, a random accessible sample and the corresponding recovery point should be stored in the same movie fragment. In case of audio, samples having the closest presentation time for every video random accessible sample should be stored as the first sample of each "traf". Hence, the first samples of each media in the "moof" have the approximately equal presentation times.

The sample size box ("stsz") shall be used. The compact sample size box ("stz2") shall not be used.

Only Media Data Box (mdat) is allowed to have size 1. Only the last Media Data Box (mdat) in the file is allowed to have size 0. Other boxes shall not have size 1.

Tracks other than the default video and audio tracks may be stored in the file.

Decoding: *An IP-IRD that supports this feature shall be able to render the default video track and the default audio track stored in the file as described above. The IP-IRD shall also be tolerant of additional tracks other than the default video and audio tracks stored in the file.*

4.2.1.1 MP4 file storage of H.264/AVC

H.264/AVC video bitstreams are stored in MP4 files using the AVC file format as specified in [17].

Encoding: *AVC file format [17] shall be used for storing H.264/AVC video tracks in MP4 files. In addition the restrictions defined in clause 4.2 shall apply.*

Decoding: *An IP-IRD that supports this feature shall support [17] to receive H.264/AVC data contained in MP4 files.*

4.2.1.2 MP4 file storage of SVC

SVC video bitstreams are stored in MP4 files using AVC file format support for Scalable Video Coding (SVC) as specified in [19].

Encoding: *AVC file format [19] shall be used for storing SVC video tracks in MP4 files. In addition the restrictions defined in clause 4.2 shall apply.*

Decoding: *An IP-IRD that supports this feature shall support the AVC file format [19] to receive SVC data contained in MP4 files.*

4.2.1.3 MP4 file storage of VC-1

VC-1 video bitstreams are stored in MP4 files using SMPTE RP2025 [12].

Encoding: *SMPTE RP2025 [12] shall be used for storing VC-1 video tracks in MP4 files. In addition the restrictions defined in clause 4.2 shall apply.*

Decoding: *An IP-IRD that supports this feature shall support [12] to receive VC-1 data contained in MP4 files.*

4.2.1.4 MP4 file storage of MPEG-4 HE AAC v2 in combination with MPEG Surround

MPEG-4 AAC, HE AAC or HE AAC v2 data in combination with MPEG Surround data is stored in MP4 files as specified in clause 7.2.2 of [20].

Encoding: *The MPEG Surround data shall be embedded into the HE AAC v2 data as specified in clause 7.2.3 of [20]. An additional track shall not be used for the transport or signalling of MPEG Surround data. The presence of MPEG Surround data should be signalled by setting mpsPresentFlag=1 and including the MPEG Surround configuration data in the MPEG-4 Audio AudioSpecificConfig according to [i.10].*

Decoding: *An IP-IRD that supports this feature shall support both explicit and implicit signalling of MPEG Surround data contained in MP4 files.*

4.2.1.5 MP4 file storage of AC-3 and Enhanced AC-3

AC-3 and Enhanced AC-3 audio bitstreams are stored in MP4 files using annex F of TS 102 366 [13].

Encoding: *Annex F of TS 102 366 [13] shall be used for storing AC-3 or Enhanced AC-3 audio tracks in MP4 files. In addition the restrictions defined in clause 4.2.1 shall apply.*

Decoding: *An IP-IRD that supports this feature shall support annex F of [13] to receive AC-3 and Enhanced AC-3 data contained in MP4 files.*

4.2.2 3GP files

This clause describes usage of 3GPP file format [7] in download services supporting this feature.

Encoding: *The 3GP file shall conform to the Basic profile of the 3GPP Release 6 file format [7].*

Decoding: *An IP-IRD that supports this feature shall be able to parse Basic profile 3GP files according to the 3GPP Release 6 file format specification [7].*

4.2.2.1 3GP file storage of H.264/AVC

The specifications in clause 4.2.2 shall apply.

4.2.2.2 3GP file storage of VC-1

VC-1 video bitstreams are stored in 3GP files using SMPTE RP2025 [12].

Encoding: *SMPTE RP2025 [12] shall be used for storing VC-1 video tracks in 3GP files. In addition the restrictions defined in clause 4.2.2 apply.*

Decoding: *An IP-IRD that supports this feature shall support [12] to receive VC-1 data contained in 3GP files.*

4.2.2.3 3GP file storage of MPEG-HE AAC v2 in combination with MPEG Surround

MPEG-4 AAC, HE AAC or HE AAC v2 data in combination with MPEG Surround data is stored in 3GP files in the same way as specified in clause 7.2.2 of [20].

Encoding: *The MPEG Surround data shall be embedded into the HE AAC v2 data as specified in clause 7.2.3 of [20]. An additional track shall not be used for the transport or signalling of MPEG Surround data. The presence of MPEG Surround data should be signalled by setting mpsPresentFlag=1 and including the MPEG Surround configuration data in the MPEG-4 Audio AudioSpecificConfig according to [i.10].*

Decoding: *An IP-IRD that supports this feature shall support both explicit and implicit signalling of MPEG Surround data contained in 3GP files.*

5 Video

Each IP-IRD shall be capable of decoding either video bitstreams conforming to H.264/AVC as specified in [1] or else video bitstreams conforming to SVC as specified in [1] or else video bitstreams conforming to VC-1 as specified in [10] or else any combination of them. An IP-IRD that is capable of decoding video bitstreams conforming to SVC shall also be capable of decoding video bitstreams conforming to H.264/AVC. Clause 5.1 describes the guidelines for encoding with H.264/AVC in DVB IP Network bit-streams, and for decoding this bit-stream in the IP-IRD. Clause 5.2 describes the guidelines for encoding with SVC in DVB IP Network bit-streams, and for decoding this bit-stream in the IP-IRD. Clause 5.3 describes the guidelines for encoding with VC-1 in DVB IP Network bit-streams, and for decoding this bit-stream in the IP-IRD. Annex B specifies application-specific constraints on the use of H.264/AVC and VC-1 for DVB IP Datacast services.

5.1 H.264/AVC video

This clause describes the guidelines for H.264/AVC video encoding and for decoding of H.264/AVC data in the IP-IRD.

The bitstreams resulting from H.264/AVC encoding shall conform to the corresponding profile specification in [1]. The IP-IRD shall allow any legal structure as permitted by the specifications in [1] in the encoded video stream even if presently "reserved" or "unused".

To allow full compliance to the specifications in [1] and upward compatibility with future enhanced versions, *an IP-IRD shall be able to skip over data structures which are currently "reserved", or which correspond to functions not implemented by the IP-IRD.*

5.1.1 Profile and level

Encoding: *Capability A H.264/AVC Bitstreams shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 1b of the Baseline Profile with constraint_set1_flag being equal to 1. In addition, in applications where decoders support the Main or the High Profile, the bitstream may optionally conform to these profiles.*

Capability B H.264/AVC Bitstreams shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 1.2 of the Baseline Profile with constraint_set1_flag being equal to 1. In addition, in applications where decoders support the Main or the High Profile, the bitstream may optionally conform to these profiles.

Capability C H.264/AVC Bitstreams shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 2 of the Baseline Profile with constraint_set1_flag being equal to 1. In addition, in applications where decoders support the Main or the High Profile, the bitstream may optionally conform to these profiles.

Capability DB H.264/AVC Bitstreams shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 3 of the Baseline Profile with constraint_set1_flag being equal to 1. In addition, in applications where decoders support the Main or the High Profile, the bitstream may optionally conform to these profiles.

Capability D H.264/AVC Bitstreams shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 3 of the Main Profile. In addition, in applications where decoders support the High Profile, the bitstream may optionally conform to the High Profile.

Capability E H.264/AVC Bitstreams shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 4 of the High Profile.

Capability F H.264/AVC Bitstreams shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 4.2 of the High Profile.

Decoding:

Capability A IP-IRDS that support H.264/AVC shall be capable of decoding and rendering pictures using Capability A H.264/AVC Bitstreams. Support of the Main Profile and other profiles beyond Baseline Profile with constraint_set1_flag equal to 1 is optional. Support of levels beyond Level 1b is optional.

Capability B IP-IRDS that support H.264/AVC shall be capable of decoding and rendering pictures using Capability A and B H.264/AVC Bitstreams. Support of the Main Profile and other profiles beyond Baseline Profile with constraint_set1_flag equal to 1 is optional. Support of levels beyond Level 1.2 is optional.

Capability C IP-IRDS that support H.264/AVC shall be capable of decoding and rendering pictures using Capability A, B and C H.264/AVC Bitstreams. Support of the Main Profile and other profiles beyond Baseline Profile with constraint_set1_flag equal to 1 is optional. Support of levels beyond Level 2 is optional.

Capability DB IP-IRDS that support H.264/AVC shall be capable of decoding and rendering pictures using Capability A, B, C and DB H.264/AVC Bitstreams. Support of the Main Profile and other profiles beyond Baseline Profile with constraint_set1_flag equal to 1 is optional. Support of levels beyond Level 3 is optional.

Capability D IP-IRDS that support H.264/AVC shall be capable of decoding and rendering pictures using Capability A, B, C, DB and D H.264/AVC Bitstreams. Support of the High Profile and other profiles beyond Main Profile is optional. Support of levels beyond Level 3 is optional.

Capability E IP-IRDS that support H.264/AVC shall be capable of decoding and rendering pictures using Capability A, B, C, DB, D and E H.264/AVC Bitstreams. Support of profiles beyond High Profile is optional. Support of levels beyond Level 4 is optional.

Capability F IP-IRDS that support H.264/AVC shall be capable of decoding and rendering pictures using Capability A, B, C, DB, D, E and F H.264/AVC Bitstreams. Support of profiles beyond High Profile is optional. Support of levels beyond Level 4.2 is optional.

If an IP-IRD encounters an extension which it cannot decode, it shall discard the following data until the beginning of the next NAL unit (to allow backward compatible extensions to be added in the future).

5.1.2 Video usability information

It is recommended that the IP-IRD support the use of Video Usability Information of the following syntax elements:

- Timing information (time_scale, num_units_in_tick, and fixed_frame_rate_flag).
- Picture Structure Information (pic_struct_present_flag).

- Maximum number of frames that precede any frame in the coded video sequence in decoding order and follow it in output order (num_reorder_frames).

It is recommended that encoders include these fields as appropriate.

5.1.3 Frame rate

Encoding: Each frame rate allowed by the applied H.264/AVC Profile and Level may be used. The maximum time distance between two pictures should not exceed 0,7 s.

Decoding: *An IP-IRD that supports H.264/AVC shall support each frame rate allowed by the H.264/AVC Profile and Level that is applied for decoding in the IP-IRD.* This includes variable frame rate.

5.1.4 Aspect ratio

Encoding: Each sample and picture aspect ratio allowed by the applied H.264/AVC Profile and Level may be used. It is recommended to avoid very large or very small picture aspect ratios and that those picture aspect ratios specified in [i.2] are used.

Decoding: *An IP-IRD that supports H.264/AVC shall support each sample and picture aspect ratio permitted by the applied H.264/AVC Profile and Level.*

5.1.5 Luminance resolution

Encoding: Each luminance resolution allowed by the applied H.264/AVC Profile and Level may be used.

Decoding: *An IP-IRD that supports H.264/AVC shall support each luminance resolution permitted by the applied H.264/AVC Profile and Level.*

5.1.6 Chromaticity

Encoding: It is recommended to specify the chromaticity coordinates of the colour primaries of the source using the syntax elements colour primaries, transfer characteristics, and matrix coefficients in the VUI. The use of ITU-R Recommendation BT.709 [i.9] is recommended.

Decoding: *An IP-IRD that supports H.264/AVC shall be capable of decoding any allowed values of colour primaries, transfer characteristics, and matrix coefficients.* It is recommended that appropriate processing be included for the rendering of pictures.

5.1.7 Chrominance format

Encoding: It is recommended to specify the chrominance locations using the syntax elements chroma_sample_loc_type_top_field and chroma_sample_loc_type_bottom_field in the VUI. It is recommended to use chroma sample type 0.

Decoding: *An IP-IRD that supports H.264/AVC shall be capable of decoding any allowed values of chroma_sample_loc_type_top_field and chroma_sample_loc_type_bottom_field.* It is recommended that appropriate processing be included for the rendering of pictures.

5.1.8 Random access points

5.1.8.1 Definition

A Random Access Point (RAP) shall be either:

- an IDR picture; or
- an I Picture, with an in-band recovery_point SEI message.

Where the recovery_point SEI message is present it shall:

- have the field exact_match_flag set to "1";
- have the field recovery_frame_cnt set to a value equivalent to 500 ms or less;
- only be preceded in the access unit to which it applies by:
 - access_unit_delimiter NAL, if present.
 - buffering_period SEI message, if present.

Unless the sequence parameter set and picture parameter set are provided outside the elementary stream, the random access point shall include exactly one SPS (that is active), and the PPS that is required for decoding the associated picture. Note that an I picture need not necessarily be a Random Access Point (RAP). *An I picture that is not a Random Access Point shall not contain a recovery_point SEI message.*

NOTE: The value of recovery_frame_cnt will impact on critical factors such as channel change performance.

5.1.8.2 Time Interval between RAPs

Encoding: *The Encoder shall place RAPs (along with associated sequence and picture parameter sets if these are not provided outside the elementary stream) in the video elementary stream at least once every 5 s. It is recommended that RAPs (along with associated sequence and picture parameter sets if these are not provided outside the elementary stream) occur on average at least every 2 s. Where channel change times are important it is recommended that RAPs (along with associated sequence and picture parameter sets if these are not provided outside the elementary stream) occur more frequently, such as every 500 ms.*

In systems where time-slicing is used, it is recommended that each time-slice begins with a random access point.

NOTE 1: Decreasing the time interval between RAPs may reduce channel hopping time and improve trick modes, but may reduce the efficiency of the video compression.

NOTE 2: Having a regular interval between RAPs may improve trick mode performance, but may reduce the efficiency of the video compression.

5.1.9 Sequence parameter sets and picture parameter sets

When changing syntax elements of sequence or picture parameter sets, it is recommended to use different values for seq_parameter_set_id or pic_parameter_set_id from the previous active ones, as per ISO/IEC 14496-10 [2].

5.1.10 Active Format Description

Encoding: It is recommended to specify the portion of the coded video frame intended for display by using Active Format Description and Bar Data syntax as defined in clauses B.3 and B.4 of [i.2], respectively. It is further recommended to embed the Active Format Description and Bar Data information within the video stream as defined in clause B.7 of [i.2].

Decoding: It is recommended that IP-IRDs process the pictures to be rendered considering the Active Format Description and Bar Data information embedded in the video stream.

5.2 SVC video

This clause describes the guidelines for SVC video encoding and for decoding of SVC data in the IP-IRD.

The bitstreams resulting from SVC encoding shall conform to the corresponding profile specification in [1]. The IP-IRD shall allow any legal structure as permitted by the specifications in [1] in the encoded video stream even if presently "reserved" or "unused".

To allow full compliance to the specifications in [1] and upward compatibility with future enhanced versions, *an IP-IRD shall be able to skip over data structures which are currently "reserved", or which correspond to functions not implemented by the IP-IRD.*

5.2.1 General

The restrictions for SVC Bitstreams and the capabilities for IP-IRDs are partly specified via SVC Bitstream Subsets. An SVC Bitstream Subset is a subset of an SVC Bitstream that can be obtained from the SVC Bitstream by discarding one or more access units and/or one or more VCL NAL units, starting from VCL NAL units with the largest value of DQId, and associated non-VCL NAL units in one or more access units, similar to the process specified in clause G.8.8.1 of ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2]. An SVC Bitstream Subset may be identical to the SVC Bitstream that contains the SVC Bitstream Subset. Some of the restriction for SVC Bitstreams and capabilities for IP-IRDs supporting SVC are specified by specifying restrictions for SVC Bitstream Subsets.

NOTE: As specified in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2], the variable DQId used in the following specification, which is equal to $16 * \text{dependency_id} + \text{quality_id}$, is assigned to each VCL NAL unit present in an SVC bitstream.

5.2.1.1 Profile and level

In the following specification, the term "Capability X" (where present) shall be replaced with "Capability A", "Capability B", "Capability C", "Capability DB", "Capability D", "Capability E", and "Capability F".

Encoding: *Capability B SVC Bitstream Subsets shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 1.2 of the Scalable Baseline Profile. In addition, in applications where decoders support the Scalable High Profile, the bitstream may optionally conform to this profile.*

Capability C SVC Bitstream Subsets shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 2 of the Scalable Baseline Profile. In addition, in applications where decoders support the Scalable High Profile, the bitstream may optionally conform to this profile.

Capability DB SVC Bitstream Subsets shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 3 of the Scalable Baseline Profile. In addition, in applications where decoders support the Scalable High Profile, the bitstream may optionally conform to this profile.

Capability D SVC Bitstream Subsets shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 3 of the Scalable High Profile.

Capability E SVC Bitstream Subsets shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 4 of the Scalable High Profile.

Capability F SVC Bitstream Subsets shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 4.2 of the Scalable High Profile.

Capability X SVC Bitstreams shall conform to ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] and shall contain one or more Capability X SVC Bitstream Subsets. Optionally, Capability X SVC Bitstreams may contain additional VCL NAL units and associated non-VCL NAL units that do not belong to any Capability X SVC Bitstream Subset.

Decoding: *Capability X IP-IRDs that support SVC shall be capable of decoding and rendering pictures using Capability X SVC Bitstreams. Support for SVC Bitstreams that do not contain Capability X SVC Bitstream Subsets is optional.*

Capability X IP-IRDs that support SVC shall be capable of decoding and rendering pictures that are represented by Capability X SVC Bitstream Subsets contained in a Capability X SVC Bitstream. Capability X IP-IRDs shall be capable of discarding the VCL NAL units of a Capability X SVC Bitstream that do not belong to a Capability X SVC Bitstream Subset, before decoding and rendering pictures. Support for decoding and rendering of pictures that are represented by a SVC Bitstream Subset with a conformance point beyond the conformance point of Capability X SVC Bitstream Subsets is optional.

If an IP-IRD encounters an extension which it cannot decode, it shall discard the following data until the beginning of the next NAL unit (to allow backward compatible extensions to be added in the future).

5.2.1.2 Backward Compatibility

In the following specification, Capability X IP-IRDs that support H.264/AVC and Capability X IP-IRDs that support SVC, with X being equal to B, C, DB, D, E or F, are also referred to as Capability X H.264/AVC IP-IRDs and Capability X SVC IP-IRDs, respectively.

Encoding: *When the level of the H.264/AVC Bitstream that is contained in the SVC Bitstream is less than or equal to level 1b, the H.264/AVC Bitstream shall obey the constraints of Capability A H.264/AVC Bitstreams.*

When the level of the H.264/AVC Bitstream that is contained in the SVC Bitstream is less than or equal to level 1.2, the H.264/AVC Bitstream shall obey the constraints of Capability B H.264/AVC Bitstreams.

When the level of the H.264/AVC Bitstream that is contained in the SVC Bitstream is less than or equal to level 2, the H.264/AVC Bitstream shall obey the constraints of Capability C H.264/AVC Bitstreams.

When the level of the H.264/AVC Bitstream that is contained in the SVC Bitstream is less than or equal to level 3, the H.264/AVC Bitstream shall obey the constraints of Capability D H.264/AVC Bitstreams.

When the level of the H.264/AVC Bitstream that is contained in the SVC Bitstream is less than or equal to level 4, the H.264/AVC Bitstream shall obey the constraints of Capability E H.264/AVC Bitstreams.

NOTE: Each SVC Bitstream contains an H.264/AVC Bitstream that is conforming to one or more of the non-scalable profiles specified in annex A of ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2]. The H.264/AVC Bitstream that is contained in an SVC Bitstream can be obtained by discarding NAL units as specified in Clause G.8.8.2 of ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2]. In particular, each Capability X SVC Bitstream and each Capability X SVC Bitstream Subset contains a Capability Y H.264/AVC Bitstream, with either Y being equal to X or Y appearing in the following ordered sequence at an earlier position than X: A, B, C, DB, D, E, F.

Decoding: *Capability B SVC IP-IRDs shall be capable of decoding any bitstream that a Capability B H.264/AVC IP-IRD is required to decode and resulting in the same displayed pictures as the Capability B H.264/AVC IP-IRD.*

Capability C SVC IP-IRDs shall be capable of decoding any bitstream that a Capability B or C H.264/AVC IP-IRD or a Capability B SVC IP-IRD is required to decode and resulting in the same displayed pictures as the Capability B or C H.264/AVC IP-IRD or the Capability B SVC IP-IRD.

Capability DB SVC IP-IRDs shall be capable of decoding any bitstream that a Capability B, C or DB H.264/AVC IP-IRD or a Capability B or C SVC IP-IRD is required to decode and resulting in the same displayed pictures as the Capability B, C or DB H.264/AVC IP-IRD or the Capability B or C SVC IP-IRD.

Capability D SVC IP-IRDs shall be capable of decoding any bitstream that a Capability B, C, DB or D H.264/AVC IP-IRD or a Capability B, C or DB SVC IP-IRD is required to decode and resulting in the same displayed pictures as the Capability B, C, DB or D H.264/AVC IP-IRD or the Capability B, C or DB SVC IP-IRD.

Capability E SVC IP-IRDs shall be capable of decoding any bitstream that a Capability B, C, DB, D or E H.264/AVC IP-IRD or a Capability B, C, DB or D SVC IP-IRD is required to decode and resulting in the same displayed pictures as the Capability B, C, DB, D or E H.264/AVC IP-IRD or the Capability B, C, DB or D SVC IP-IRD.

Capability F SVC IP-IRDs shall be capable of decoding any bitstream that a Capability B, C, DB, D, E or F H.264/AVC IP-IRD or a Capability B, C, DB, D or E SVC IP-IRD is required to decode and resulting in the same displayed pictures as the Capability B, C, DB, D, E or F H.264/AVC IP-IRD or the Capability B, C, DB, D or E SVC IP-IRD.

5.2.1.3 Reference Base Pictures

Encoding: *In each SVC Bitstream, the time interval between any two access units (in decoding order) that contain VCL NAL units with store_ref_base_pic_flag equal to 1 shall be greater than or equal to 100 ms.*

Decoding: *An IP-IRD that supports SVC and is labelled with a particular capability X shall support decoding and rendering pictures using Capability X SVC Bitstreams in which the time interval between any two access units (in decoding order) that contain VCL NAL units with store_ref_base_pic_flag equal to 1 is greater than or equal to 100 ms. Support for SVC Bitstreams in which the time interval between any two access units (in decoding order) that contain VCL NAL units with store_ref_base_pic_flag is less than 100 ms is optional.*

5.2.2 Video usability information

5.2.2.1 Sequence parameter sets

It is recommended that the IP-IRD supports the use of the following syntax elements in the Video Usability Information of sequence parameter sets:

- Timing information (time_scale, num_units_in_tick, and fixed_frame_rate_flag).
- Picture Structure Information (pic_struct_present_flag).
- Maximum number of frames that precede any frame in the coded video sequence in decoding order and follow it in output order (num_reorder_frames).

It is recommended that encoders include these fields in each sequence parameter set as appropriate.

5.2.2.2 Subset sequence parameter sets

It is recommended that the IP-IRD supports the use of the following syntax element in the Video Usability Information of subset sequence parameter sets:

- Maximum number of frames that precede any frame in the coded video sequence in decoding order and follow it in output order (num_reorder_frames).

It is recommended that encoders include this field in each subset sequence parameter set as appropriate.

It is recommended that the IP-IRD supports the use of the following syntax elements in the SVC Video Usability Information extension in subset sequence parameter sets, for each value *i* in the range of 0 to *vui_ext_num_entries_minus1*, inclusive, with *vui_ext_num_entries_minus1* being the corresponding field in the SVC Video Usability Information extension:

- Timing information (*vui_ext_time_scale[i]*, *vui_ext_num_units_in_tick[i]*, and *vui_ext_fixed_frame_rate_flag[i]*).
- Picture Structure Information (*vui_ext_pic_struct_present_flag[i]*).

It is recommended that encoders include these fields in each subset sequence parameter set and for each present combination of *dependency_id*, *quality_id* and *temporal_id* as appropriate.

5.2.3 Frame rate

Encoding: Each frame rate allowed by the applied SVC Profile and Level may be used. The maximum time distance between two pictures should not exceed 0,7 s.

Decoding: *An IP-IRD that supports SVC shall support each frame rate allowed by the SVC Profile and Level that is applied for decoding in the IP-IRD.* This includes variable frame rate.

5.2.4 Aspect ratio

Encoding: Each sample and picture aspect ratio allowed by the applied SVC Profile and Level may be used. It is recommended to avoid very large or very small picture aspect ratios and that those picture aspect ratios specified in [i.2] are used.

Decoding: *An IP-IRD that supports SVC shall support each sample and picture aspect ratio permitted by the applied SVC Profile and Level.*

5.2.5 Luminance resolution

Encoding: Each luminance resolution allowed by the applied SVC Profile and Level may be used.

Decoding: *An IP-IRD that supports SVC shall support each luminance resolution permitted by the applied SVC Profile and Level.*

5.2.6 Chromaticity

Encoding: It is recommended to specify the chromaticity coordinates of the colour primaries of the source using the syntax elements `colour_primaries`, `transfer_characteristics`, and `matrix_coefficients` in the VUI of each sequence parameter set and subset sequence parameter set. The use of ITU-R Recommendation BT.709 [i.9] is recommended.

Decoding: *An IP-IRD that supports SVC shall be capable of decoding any allowed values of `colour_primaries`, `transfer_characteristics`, and `matrix_coefficients`.* It is recommended that appropriate processing be included for the rendering of pictures.

5.2.7 Chrominance format

Encoding: It is recommended to specify the chrominance locations using the syntax elements `chroma_sample_loc_type_top_field` and `chroma_sample_loc_type_bottom_field` in the VUI of each sequence parameter set and subset sequence parameter set. It is recommended to use chroma sample type 0.

It is recommended that the chrominance locations specified by the syntax elements `chroma_phase_x_plus1_flag` and `chroma_phase_y_plus1` in a subset sequence parameter set be consistent with the chrominance locations specified in the VUI of the same subset sequence parameter set, as per ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2].

It is recommended that the reference layer chrominance locations specified by the syntax elements `ref_layer_chroma_phase_x_plus1_flag` and `ref_layer_chroma_phase_y_plus1` be consistent with the chrominance locations specified in the VUI of the SVC sequence parameter set that is referenced in the reference layer representation (specified by `ref_layer_dq_id`), as per ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2].

Decoding: *An IP-IRD that supports SVC shall be capable of decoding any allowed values of `chroma_sample_loc_type_top_field` and `chroma_sample_loc_type_bottom_field`.* It is recommended that appropriate processing be included for the rendering of pictures.

5.2.8 Active Format Description

- Encoding: It is recommended to specify the portion of the coded video frame intended for display by using Active Format Description and Bar Data syntax as defined in annexes B.3 and B.4 of [i.2], respectively. It is further recommended to embed the Active Format Description and Bar Data information within the video stream as defined in annex B.7 of [i.2].
- Decoding: It is recommended that IP-IRDs process the pictures to be rendered considering the Active Format Description and Bar Data information embedded in the video stream.

5.2.9 SVC Random access points

An SVC Random Access Point (RAP) is an access unit in an SVC Bitstream at which an IP-IRD can begin decoding video successfully. An SVC RAP is associated with one or more values of `dependency_id`. An SVC Random Access Point for a particular value of `dependency_id` is an access unit at which an IP-IRD can begin decoding layer pictures for the particular value of `dependency_id`. A layer picture for a particular value of `dependency_id` represents the decoded picture that is obtained when all dependency representations with `dependency_id` less than or equal to the particular value of `dependency_id` of an access unit are decoded.

5.2.9.1 Definition

An SVC RAP for a particular value of `dependency_id` is an access unit with `temporal_id` equal to 0 that contains a dependency representation with the particular value of `dependency_id` and for which either of the following conditions is true:

- the field `idr_flag` is equal to 1 for the dependency representation with the particular value of `dependency_id`; or
- all slices of the dependency representation with the particular value of `dependency_id` have `slice_type` equal to 2 or 7 and the access unit includes an in-band `recovery_point` SEI message that applies to the dependency representation with the particular value of `dependency_id`.

Where the `recovery_point` SEI message is present it shall:

- have the field `exact_match_flag` set to 1;
- have the field `recovery_frame_cnt` set to 0;
- only be preceded in the access unit by:
 - an `access_unit_delimiter` NAL, if present;
 - zero or more `buffering_period` SEI messages (which may be included in `scalable_nesting` SEI messages);
 - zero or more `recovery_point` SEI messages (which may be included in `scalable_nesting` SEI messages) that apply to dependency representations with smaller values of `dependency_id`.

Unless the SVC sequence parameter sets and picture parameter sets are provided outside the elementary stream, the SVC random access point for a particular value of `dependency_id` shall obey the following constraints:

- the SVC random access point shall include all sequence parameter sets, subset sequence parameter sets, and picture parameter sets that are referenced in the VCL NAL units of the access unit;
- the SVC random access point shall not contain any sequence parameter set that is not referenced in the VCL NAL units of the access unit;
- inside the SVC random access point, a sequence parameter set shall not be preceded by any subset sequence parameter set;
- inside the SVC random access point, a subset sequence parameter set that is referenced in VCL NAL units with a particular value of `DQId` shall not be preceded by any subset sequence parameter set that is only referenced in VCL NAL units of the random access point that are associated with a greater value of `DQId`.

An access unit in which all slices of a dependency representation with a particular value of `dependency_id` have `slice_type` equal to 2 or 7 need not necessarily be an SVC Random Access Point for the particular value of `dependency_id`. *An access unit that is not an SVC Random Access Point for a particular value of `dependency_id` shall not contain a `recovery_point SEI` message that applies to the dependency representation with the particular value of `dependency_id`.*

If an access unit represents an SVC RAP for a particular value of `dependency_id`, it shall also represent an SVC RAP for all values of `dependency_id` in the range from 0 to the particular value of `dependency_id` minus 1, inclusive.

If the maximum present value of `dependency_id` in an access unit is different from the maximum present value of `dependency_id` in the previous access unit in decoding order (when present), the access unit shall represent an SVC RAP for all values of `dependency_id` present in the access unit.

5.2.9.2 Time Interval between SVC RAPs

Encoding: *The Encoder shall place SVC RAPs for `dependency_id` equal to 0 (along with associated SVC sequence parameter sets and picture parameter sets if these are not provided outside the elementary stream) in the video elementary stream at least once every 5 s. It is recommended that SVC RAPs for `dependency_id` equal to 0 (along with associated SVC sequence parameter sets and picture parameter sets if these are not provided outside the elementary stream) occur on average at least every 2 s. Where channel change times are important it is recommended that SVC RAPs for `dependency_id` equal to 0 (along with associated SVC sequence parameter sets and picture parameter sets if these are not provided outside the elementary stream) occur more frequently, such as every 500 ms.*

For each time interval in which dependency representations with any particular value of `dependency_id` greater than 0 are present in the bitstream, the encoder shall place SVC RAPs for this particular value of `dependency_id` (along with associated SVC sequence parameter sets and picture parameter sets if these are not provided outside the elementary stream) in the video elementary stream at least once every 10 s. It is recommended that, for each time interval in which dependency representations with any particular value of `dependency_id` greater than 0 are present in the bitstream, SVC RAPs for this particular value of `dependency_id` (along with associated SVC sequence parameter sets and picture parameter sets if these are not provided outside the elementary stream) occur on average at least every 5 s.

In systems where time-slicing is used, it is recommended that each time-slice begins with an SVC RAP for `dependency_id` equal to 0.

NOTE 1: An SVC RAP for a particular value of `dependency_id` need not represent an SVC RAP for greater values of `dependency_id`.

NOTE 2: Decreasing the time interval between SVC RAPs may reduce channel hopping time and improve trick modes, but may reduce the efficiency of the video compression.

NOTE 3: Having a regular interval between SVC RAPs may improve trick mode performance, but may reduce the efficiency of the video compression.

5.2.10 Sequence parameter sets and picture parameter sets

When transmitting a new picture parameter set, it is recommended to use a value of `pic_parameter_set_id` that is different from the value of `pic_parameter_set_id` for any picture parameter set that was the previous active picture parameter set or the previous active layer picture parameter set for any value of `DQId`.

When transmitting a new sequence parameter set, it is recommended to use a value of `seq_parameter_set_id` that is different from the value of `seq_parameter_set_id` for any sequence parameter set that was the previous active SVC sequence parameter set or the previous active layer SVC sequence parameter set for `DQId` equal to 0.

When transmitting a new subset sequence parameter set, it is recommended to use a value of `seq_parameter_set_id` that is different from the value of `seq_parameter_set_id` for any subset sequence parameter set that was the previous active SVC sequence parameter set or the previous active layer SVC sequence parameter set for any value of `DQId`.

5.3 VC-1 video

This clause describes the guidelines for VC-1 video encoding and for decoding of VC-1 data in the IP-IRD.

The bitstreams resulting from VC-1 encoding shall conform to the corresponding profile specification in [10]. The IP-IRD shall allow any legal structure as permitted by the specifications in [10] in the encoded video stream even if presently "reserved" or "unused".

To allow full compliance to the specifications in [10] and upward compatibility with future enhanced versions, *an IP-IRD shall be able to skip over data structures which are currently "reserved", or which correspond to functions not implemented by the IP-IRD.*

5.3.1 Profile and level

Encoding: *Capability A VC-1 Bitstreams shall conform to the restrictions described in SMPTE 421M [10] for Simple Profile at level LL.*

Capability B VC-1 Bitstreams shall conform to the restrictions described in SMPTE 421M [10] for Simple Profile at level ML.

Capability C VC-1 Bitstreams shall conform to the restrictions described in SMPTE 421M [10] for Advanced Profile at level L0.

Capability D VC-1 Bitstreams shall conform to the restrictions described in SMPTE 421M [10] for Advanced Profile at level L1.

Capability E VC-1 Bitstreams shall conform to the restrictions described in SMPTE 421M [10] for Advanced Profile at level L3.

Decoding: *Capability A IP-IRDS that support VC-1 shall be capable of decoding and rendering pictures using Capability A VC-1 Bitstreams. Support of additional profiles and levels is optional.*

Capability B IP-IRDS that support VC-1 shall be capable of decoding and rendering pictures using Capability A and B VC-1 Bitstreams. Support of additional profiles and levels is optional.

Capability C IP-IRDS that support VC-1 shall be capable of decoding and rendering pictures using Capability A, B and C VC-1 Bitstreams. Support of additional profiles and levels is optional.

Capability D IP-IRDS that support VC-1 shall be capable of decoding and rendering pictures using Capability A, B, C and D VC-1 Bitstreams. Support of additional profiles and levels is optional.

Capability E IP-IRDS that support VC-1 shall be capable of decoding and rendering pictures using Capability A, B, C, D and E VC-1 Bitstreams. Support of additional profiles and levels is optional.

If an IP-IRD encounters an extension which it cannot decode, it shall discard the following data until the next start code prefix (to allow backward compatible extensions to be added in the future).

5.3.2 Frame rate

Encoding: Each frame rate allowed by the applied VC-1 Profile and Level may be used. The maximum time distance between two pictures should not exceed 0,7 s.

Decoding: *An IP-IRD that supports VC-1 shall support each frame rate allowed by the VC-1 Profile and Level that is applied for decoding in the IP-IRD. This includes variable frame rate.*

5.3.3 Aspect ratio

Encoding: Each sample and picture aspect ratio allowed by the applied VC-1 Profile and Level may be used. It is recommended to avoid very large or very small picture aspect ratios and that those picture aspect ratios specified in [i.2] are used.

Decoding: *An IP-IRD that supports VC-1 shall support each sample and picture aspect ratio permitted by the applied VC-1 Profile and Level.*

5.3.4 Luminance resolution

Encoding: Each luminance resolution allowed by the applied VC-1 Profile and Level may be used.

Decoding: *An IP-IRD that supports VC-1 shall support each luminance resolution permitted by the applied VC-1 Profile and Level.*

5.3.5 Chromaticity

Encoding: It is recommended to specify the chromaticity coordinates of the colour primaries of the source using the syntax elements COLOR_PRIM, TRANSFER_CHAR and MATRIX_COEF, if these syntax elements are allowed by the applied VC-1 Profile.

For Advanced Profile, the use of ITU-R Recommendation BT.709 [i.9] is recommended (video source corresponding to COLOR_PRIM, TRANSFER_CHAR and MATRIX_COEF field values equal to "1", "1", "1").

For Simple and Main Profile, the default value for the COLOR_PRIM, TRANSFER_CHAR and MATRIX_COEF field values shall be "6", "6", "6" for video sources originating from a 29.97 frame/s system and shall be "5", "5", "6" for video sources originating from a 25 frame/s system.

Decoding: *An IP-IRD that supports VC-1 shall be capable of decoding any allowed values of COLOR_PRIM, TRANSFER_CHAR and MATRIX_COEF. It is recommended that appropriate processing be included for the rendering of pictures.*

5.3.6 Random access points

Encoding: Where channel change times are important it is recommended that a Sequence Header and Entry Point Header are encoded at least once every 500 ms, if these syntax elements are allowed by the applied VC-1 Profile. In applications where channel change time is an issue but coding efficiency is critical, it is recommended that a Sequence Header and Entry Point Header are encoded at least once every 2 s, if these syntax elements are allowed by the applied VC-1 Profile. For those applications where channel change time is not an issue, it is recommended that a Sequence Header and Entry Point Header are sent at least once every 5 s, if these syntax elements are allowed by the applied VC-1 Profile.

In systems where time-slicing is used, it is recommended that each time-slice begins with a Sequence Header and Entry Point Header, if these syntax elements are allowed by the applied VC-1 Profile.

NOTE 1: Increasing the frequency of Sequence Header and Entry Point Header will reduce channel hopping time but will reduce the efficiency of the video compression.

NOTE 2: Having a regular interval between Entry Point Headers may improve trick mode performance, but may reduce the efficiency of the video compression.

5.3.7 Active Format Description

- Encoding: It is recommended to specify the portion of the coded video frame intended for display by using Active Format Description and Bar Data syntax as defined in clauses B.3 and B.4 of [i.2], respectively. It is further recommended to embed the Active Format Description and Bar Data information to the video stream as defined in clause B.8 of [i.2].
- Decoding: It is recommended that IP-IRDs process the pictures to be rendered considering the Active Format Description and Bar Data information embedded in the video stream.

6 Audio

Each IP-IRD shall be capable of decoding either audio bitstreams conforming to MPEG-4 HE AAC v2 as specified in ISO/IEC 14496-3 [3], or else audio bitstreams conforming to MPEG-4 HE AAC v2 in combination with MPEG Surround as specified in ISO/IEC 23003-1 [20], or else audio bitstreams conforming to Extended AMR-WB (AMR WB+) as specified in TS 126 290 [8], or else audio bitstreams conforming to AC-3 or Enhanced AC-3 as specified in TS 102 366 [13], or any combination of the five.

Clause 6.1 describes the guidelines for encoding with MPEG-4 AAC, MPEG-4 HE AAC and MPEG-4 HE AAC v2, and MPEG-4 AAC, MPEG-4 HE AAC and MPEG-4 HE AAC v2 in combination with MPEG Surround, and for decoding this bit-stream in the IP-IRD. Clause 6.2 describes the guidelines for encoding with AMR-WB+ and for decoding this bit-stream in the IP-IRD. Clause 6.3 describes the guidelines for encoding with AC-3 and for decoding this bit-stream in the IP-IRD. Clause 6.4 describes the guidelines for encoding with Enhanced AC-3 and for decoding this bit-stream in the IP-IRD. Annex B specifies application-specific constraints on the use of MPEG-4 HE AAC v2, MPEG-4 HE AAC v2 in combination with MPEG Surround and AMR-WB+ for DVB IP Datacast services.

The recommended level for reference tones for transmission is 18 dB below clipping level, in accordance with EBU Recommendation R.68 [i.4].

6.1 MPEG-4 AAC, HE AAC, HE AAC v2 and MPEG Surround audio

For MPEG-4 AAC and HE AAC, the audio encoding shall conform to the requirements defined in ISO/IEC 14496-3 [3].

For MPEG-4 HE AAC v2 the audio encoding shall conform to the requirements defined in ISO/IEC 14496-3 [3] including Amendment 2 [3].

For MPEG-4 AAC, HE AAC or HE AAC v2 in combination with MPEG Surround, the audio encoding shall conform to the requirements defined in ISO/IEC 14496-3 including Amendments 2 and 5 [3] and ISO/IEC 23003-1 [20].

The IP-IRD design should be made under the assumption that any legal structure as permitted by ISO/IEC 14496-3 including Amendments 2 and 5 [3] and ISO/IEC 23003-1 [20] may occur in the broadcast stream even if presently reserved or unused. *To allow full compliance to ISO/IEC 14496-3 [3] and ISO/IEC 23003-1 [20], and to ensure upward compatibility with future enhanced versions, a DVB IP-IRD shall be able to skip over data structures which are currently "reserved", or which correspond to functions not implemented by the IP-IRD. For example, an IP-IRD which is not designed to make use of the extension payload shall skip over that portion of the bit-stream.*

The following clauses are based on ISO/IEC 14496-3 including Amendments 2 and 5 [3] and ISO/IEC 23003-1 [20].

6.1.1 Audio mode

Encoding: *For MPEG-4 AAC, the audio shall be encoded in mono or 2-channel stereo according to the functionality defined in the MPEG-4 AAC Profile Level 2 or in multi-channel according to the functionality defined in the MPEG-4 AAC Profile Level 4.*

For MPEG-4 HE AAC, the audio shall be encoded in mono or 2-channel stereo according to the functionality defined in the MPEG-4 HE AAC Profile Level 2 or in multi-channel according to the functionality defined in the MPEG-4 HE AAC Profile Level 4.

For MPEG-4 HE AAC v2, the audio shall be encoded in mono, parametric stereo or 2-channel stereo according to the functionality defined in the MPEG-4 HE AAC v2 Profile Level 2.

For MPEG-4 AAC in combination with MPEG Surround, the audio shall be encoded according to the functionality defined in both the MPEG-4 AAC Profile Level 2 and the MPEG Surround Baseline Profile Level 4, or according to the functionality defined in both the MPEG-4 AAC Profile Level 4 and the MPEG Surround Baseline Profile Level 5

For MPEG-4 HE AAC in combination with MPEG Surround, the audio shall be encoded according to the functionality defined in both the MPEG-4 HE AAC Profile Level 2 and the MPEG Surround Baseline Profile Level 4, or according to the functionality defined in both the MPEG-4 HE AAC Profile Level 4 and the MPEG Surround Baseline Profile Level 5

For MPEG-4 HE AAC v2 in combination with MPEG Surround, the audio shall be encoded according to the functionality defined in both the MPEG-4 HE AAC v2 Profile Level 2 and the MPEG Surround Baseline Profile Level 4.

A simulcast of a mono/parametric stereo/stereo signal together with the multi-channel signal is optional.

Decoding: *An IP-IRD that supports MPEG-4 HE AAC v2 shall be capable of decoding in mono, parametric stereo or 2-channel-stereo according to the functionality defined in the MPEG-4 HE AAC v2 Profile Level 2.]. An IP-IRD that supports HE AAC v2 may be capable of decoding multi channel according to the functionality defined in the MPEG-4 HE AAC Profile Level 4.*

An IP-IRD that supports MPEG-4 HE AAC v2 audio, is capable of decoding MPEG Surround and is capable of providing up to 2.0 channels of output shall be capable of providing decoder output according to MPEG Surround Baseline Profile Level 1.

An IP-IRD that supports MPEG-4 HE AAC v2 audio, is capable of decoding MPEG Surround and is capable of providing more than two and up to 7.1 channels of output shall be capable of providing decoder output according to MPEG Surround Baseline Profile Level 3.

An IP-IRD that supports MPEG-4 HE AAC audio up to Level 3, is capable of decoding MPEG Surround and is capable of providing 7.1 channels or more of output shall be capable of providing decoder output according to MPEG Surround Baseline Profile Level 4.

An IP-IRD that supports MPEG-4 HE AAC audio at Level 4, is capable of decoding MPEG Surround and is capable of providing 7.1 channels or more of output shall be capable of providing decoder output according to MPEG Surround Baseline Profile Level 5.

The support of multi-channel decoding in an IP-IRD is optional.

6.1.2 Profiles

Encoding: *The encoder shall use either the MPEG-4 AAC Profile or the MPEG-4 HE AAC Profile or the MPEG-4 HE AAC v2 Profile. Use of the MPEG-4 HE AAC v2 Profile is recommended. For audio encoded using MPEG Surround, the MPEG Surround Baseline Profile shall be used.*

Decoding: *An IP-IRD that supports MPEG-4 HE AAC v2 shall be capable of decoding the MPEG-4 HE AAC v2 Profile. An IP-IRD that supports MPEG Surround shall be capable of decoding according to the MPEG Surround Baseline Profile.*

6.1.3 Bit rate

- Encoding: Audio may be encoded at any bit rate allowed by the applied profile and selected Level.
- Decoding: *An IP-IRD that supports MPEG-4 HE AAC v2 shall support any bit rate allowed by the MPEG-4 HE AAC v2 Profile and selected Level.*

6.1.4 Sampling frequency

- Encoding: Any of the audio sampling frequencies of the MPEG-4 HE AAC v2 Profile Level 2 may be used for mono, parametric stereo and 2-channel stereo and of the MPEG-4 HE AAC Profile Level 4 for multichannel audio. For audio encoded using MPEG Surround, the sampling frequency of the MPEG Surround data shall be equal to the sampling frequency of the core audio stream.
- Decoding: *An IP-IRD that supports MPEG-4 HE AAC v2 shall support each audio sampling rate permitted by the MPEG-4 HE AAC v2 Profile Level 2 for mono, parametric stereo and 2-channel stereo and of the MPEG-4 HE AAC Profile Level 4 for multichannel audio. An IP-IRD that supports MPEG Surround shall support each audio sampling rate permitted by the supported MPEG Surround Baseline Profile Level.*

6.1.5 Dynamic range control

- Encoding: The encoder may use the MPEG-4 AAC Dynamic Range Control (DRC) tool.
- Decoding: *An IP-IRD that supports MPEG-4 HE AAC v2 shall support the MPEG-4 AAC Dynamic Range Control (DRC) tool.*

6.1.6 Matrix downmix

- Decoding: *An IP-IRD that supports MPEG-4 HE AAC v2 shall support the matrix downmix as defined in MPEG-4.*

6.2 AMR-WB+ audio

AMR-WB+ encoding and decoding of AMR-WB+ data shall follow the guidelines described in this clause and are based on TS 126 290 [8].

For AMR-WB+ the audio encoding shall conform to the requirements defined in TS 126 290 [8].

6.2.1 Audio mode

- Encoding: *The audio shall be encoded in mono or stereo according to the functionality defined in the AMR-WB+ [8].*
- Decoding: *An IP-IRD that supports AMR-WB+ shall be capable of decoding in mono and stereo the functionality defined in the AMR-WB+, as specified in TS 126 290 [8].*

6.2.2 Sampling frequency

- Encoding: Any of the audio sampling rates of the AMR-WB+ may be used for mono and stereo.
- Decoding: *An IP-IRD that supports AMR-WB+ shall support each audio sampling rate permitted by the AMR-WB+ for mono and stereo.*

6.3 AC-3 audio

The encoding and decoding of an AC-3 elementary stream shall conform to the requirements defined in TS 102 366 [13] excluding annex E. Annex E specifies the Enhanced AC-3 bitstream syntax.

6.3.1 Audio mode

Encoding: *The audio shall be encoded in mono, 2-channel-stereo or multi-channel, as specified in TS 102 366 [13] excluding annex E.*

Decoding: *An IP-IRD that supports AC-3 shall be capable of decoding to mono, or 2-channel-stereo PCM, as specified in TS 102 366, [13] excluding annex E. Support for decoding to multi-channel PCM in an IP-IRD is optional.*

6.3.2 Bit rate

Encoding: *Audio may be encoded at any bit rate listed in TS 102 366 [13], excluding annex E.*

Decoding: *An IP-IRD that supports AC-3 shall support all bit rates listed in TS 102 366 [13], excluding annex E.*

6.3.3 Sampling frequency

Encoding: *Audio may be encoded at any sample rate listed in TS 102 366 [13], excluding annex E.*

Decoding: *An IP-IRD that supports AC-3 shall support all sample rates listed in TS 102 366 [13], excluding annex E.*

6.4 Enhanced AC-3 audio

The encoding and decoding of an Enhanced AC-3 elementary stream shall conform to the requirements defined in TS 102 366 [13] including annex E.

6.4.1 Audio mode

Encoding: *The audio shall be encoded in mono, 2-channel-stereo or multi-channel, as specified in TS 102 366 [13].*

Decoding: *An IP-IRD that supports Enhanced AC-3 shall be capable of decoding to mono, or 2-channel-stereo PCM, as specified in TS 102 366, [13]. Support for decoding to multi-channel PCM in an IP-IRD is optional.*

6.4.2 Substreams

Encoding *The Enhanced AC-3 elementary stream shall contain no more than three independent substreams in addition to the independent substream containing the main audio programme. The main audio programme shall only be delivered in independent substream 0 and dependent substreams associated with independent substream 0. All substreams within an Enhanced AC-3 bitstream shall be encoded with the same number of audio blocks per syncframe.*

Decoding *An IP-IRD that supports Enhanced AC-3 shall be able to accept Enhanced AC-3 elementary streams that contain more than one substream. IP-IRDs shall be capable of decoding independent substream 0.*

6.4.3 Bit rate

Encoding: *Audio may be encoded at any bit rate up to and including 3 024 kbit/s.*

Decoding: *An IP-IRD that supports Enhanced AC-3 shall support a maximum bit rate of 3 024 kbit/s.*

6.4.4 Sampling frequency

- Encoding: Audio may be encoded at a sample rate of 32 kHz, 44,1 kHz or 48 kHz. *All substreams present in an Enhanced AC-3 bitstream shall be encoded at the same sample rate.*
- Decoding: *An IP-IRD that supports Enhanced AC-3 shall support sample rates of 32 kHz, 44,1 kHz and 48 kHz.*

6.4.5 Stream mixing

In some applications, the audio decoder may be capable of simultaneously decoding two different programme elements, carried in two separate Enhanced AC-3 elementary streams, or in separate independent substreams within a single Enhanced AC-3 elementary stream, and then combining the programme elements into a complete programme.

- Encoding: *The elementary stream or independent substream that carries the associated audio services to be mixed with the main programme audio shall not contain more audio channels than the main audio programme.*

The elementary stream or independent substream carrying the associated audio service shall contain mixing metadata, as defined in TS 102 366 [13], for use by the decoder to control the mixing process.

To match the default user volume adjustment setting in the decoder, the pgmscl field in the associated programme elementary stream or independent substream shall be set to a positive value of 12 dB.

A minimum functionality mixer is described in clause E.4 of TS 102 366 [13]. Elementary streams or independent substreams intended to be combined together for reproduction according to this mixing process shall meet the following constraints:

The elementary stream or independent substream that carries the associated audio services to be mixed with the main programme audio shall contain no more than two audio channels;

- Decoding: *If audio access units from two audio services which are to be simultaneously decoded do not have identical RTP timestamp values indicated in their corresponding RTP headers (indicating that the audio encoding was not frame synchronous) then the audio frames (access units) of the main audio service shall be presented to the audio decoder for decoding and presentation at the time indicated by the RTP timestamp. An associated service, which is being simultaneously decoded, shall have its audio frames (access units), which are in closest time alignment (as indicated by the RTP timestamp) to those of the main service being decoded, presented to the audio decoder for simultaneous decoding.*
- IP-IRDs shall set the default user volume adjustment of the associated programme level to minus 12 dB.*

Annex A (informative): Description of the implementation guidelines

A.1 Introduction

The present document defines how advanced audio and video compression algorithms may be used for all DVB services delivered directly over IP protocols without the use of an intermediate MPEG-2 Transport Stream. An example of this type of DVB service is DVB-H, using multi-protocol encapsulation. The corresponding guidelines for audio-visual coding for DVB services which use an MPEG-2 Transport Stream are given in TS 101 154 [i.2] for distribution services and in TS 102 154 [i.3] for contribution services. Examples of Transport Stream based DVB service are the familiar DVB-S, DVB-C and DVB-T transmissions.

The "systems layer" of the present document addresses issues related to transport and synchronization of advanced audio and video. The systems layer is based on the use of RTP, a generic Transport Protocol for Real-Time Applications as defined in RFC 3550 [4]. Use of RTP requires the definition of payload formats that are specific for each content format, and so the system layer specifies which RTP payload formats to use for transport of advanced audio and video, as well as applicable constraints for that. Further information on the systems layer is given in clause A.2.

The advanced video coding uses either H.264/AVC or SVC, as specified in ITU-T Recommendation H.264 [1] and in ISO/IEC 14496-10 [2], or else VC-1, as specified in SMPTE 421M [10]. These algorithms use an architecture based on a motion-compensated block transform, like the older MPEG-1 and MPEG-2 algorithms. However, unlike the earlier algorithms, they have smaller, dynamically selected block sizes to allow the encoder to represent both large and small moving objects more efficiently. They also support greater precision in the representation of motion vectors and use more sophisticated variable-length coding to represent the coded information more efficiently. The algorithms include loop filtering to help reduce the visibility of blocking artefacts that may appear when the encoder is highly stressed by extremely critical source material. SVC is a scalable extension of H.264/AVC that enables scalability at a bitstream level and allows the transmission of different temporal and/or spatial resolutions and/or different fidelities in a single bitstream. For further information on the video codecs see clause A.3.

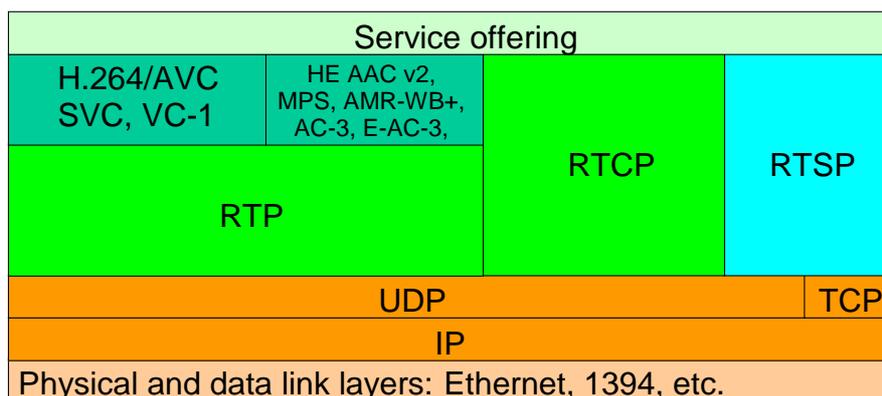
The advanced audio coding uses either MPEG-4 AAC, HE AAC or HE AAC v2 audio, as specified in ISO/IEC 14496-3 [3], or else MPEG-4 AAC, HE AAC or HE AAC v2 as specified in ISO/IEC 14496-3 [3] in combination with MPEG Surround as specified in ISO/IEC 23003-1 [20], or else Extended AMR-WB (AMR-WB+) audio as specified in TS 126 290 [8], or else AC-3 or Enhanced AC-3 audio as specified in TS 102 366 [13]. The MPEG-4 HE AAC v2 Profile is derived from the MPEG-2 Advanced Audio Coding (AAC), first published in 1997. MPEG-4 AAC is closely based on MPEG-2 AAC but includes some further enhancements such as perceptual noise substitution to give better performance at low bit rates. The MPEG-4 HE AAC Profile adds spectral band replication, to allow more efficient representation of high-frequency information by using the lower harmonic as a reference. The MPEG-4 HE AAC v2 Profile adds the parametric stereo tool to the MPEG-4 HE AAC Profile, to allow a more efficient representation of the stereo image at low bit rates. The combination of MPEG-4 AAC, HE AAC or HE AAC v2 with MPEG Surround enables mono or stereo backward compatible multi-channel coding, to allow an efficient representation of surround sound at low bit rates. Extended AMR-WB (AMR-WB+) has been optimized for use at low bit rates with source material where speech predominates. AC-3 is an audio coding format designed to encode multiple channels of audio into a low bit rate format. Dolby Digital, which is a branded version of AC-3, encodes up to 5.1 channels of audio. Enhanced AC-3 is a development of AC-3 that improves low data rate performance and supports a more flexible bitstream syntax to support new audio services. Dolby Digital Plus, which is a branded version of Enhanced AC-3, encodes up to 7.1 channels of audio, and enables multiple audio services to be carried within a single bit stream. For further information on the audio codecs see clause A.4.

A wide range of potential applications are covered by the present document, ranging from HDTV services to low-resolution services delivered to small portable receivers. A particular example of the latter type of service is the DVB IP Datacast application. A common generic toolbox is used by all DVB services, where each DVB application can select the most appropriate tool from within that toolbox. Annex B of the specification defines application-specific constraints on the use of the toolbox for the particular case of DVB IP Datacast services. For further information on the DVB IP Datacast application and the background to the constraints that have been defined, see clause A.5.

A.2 Systems

A.2.1 Protocol stack

For delivery of DVB Services over IP-based networks a protocol stack is defined in a suite of DVB specifications. The systems part the present document addresses only the part of the protocol stack that is related to the transport and synchronization of audio and video. This part of the DVB-IP protocol stack is given in figure A.1. For completeness, RTCP and RTSP are also included, as they are relevant for RTP usage, though there are no specific guidelines for RTCP and RTSP defined in the present document.



NOTE: Specifications for RTCP and RTSP usage are beyond the scope of the present document.

Figure A.1: The part of the DVB-IP protocol stack relevant for the transport of advanced audio and video

The transport of audio and video data is based on RTP, a generic Transport Protocol for Real-Time Applications as defined in RFC 3550 [4]. RFC 3550 [4] specifies the elements of the RTP transport protocol that are independent of the data that is transported, while separate RFCs define how to use RTP for transport of specific data such as coded audio and video.

A.2.2 Transport of H.264/AVC video

To transport H.264/AVC video data, RFC 3984 [6] is used. The H.264/AVC specification [1] distinguishes conceptually between a Video Coding Layer (VCL), and a Network Abstraction Layer (NAL). The VCL contains the video features of the codec (transform, quantization, motion compensation, loop filter, etc.). The NAL layer formats the VCL data into Network Abstraction Layer units (NAL units) suitable for transport across the applied network or storage medium. A NAL unit consists of a one-byte header and the payload; the header indicates the type of the NAL unit and other information, such as the (potential) presence of bit errors or syntax violations in the NAL unit payload, and information regarding the relative importance of the NAL unit for the decoding process. RFC 3984 [6] specifies how to carry NAL units in RTP packets.

A.2.3 Transport of SVC video

To transport SVC video data, RFC 2250 [i.1] is used. The H.264/AVC specification, which includes SVC, [1] distinguishes conceptually between a Video Coding Layer (VCL), and a Network Abstraction Layer (NAL). The VCL contains the video features of the codec (transform, quantization, motion compensation, loop filter, etc.). The NAL layer formats the VCL data into Network Abstraction Layer units (NAL units) suitable for transport across the applied network or storage medium. A NAL unit consists of a one-byte or four-byte header and the payload; the header indicates the type of the NAL unit and other information, such as the (potential) presence of bit errors or syntax violations in the NAL unit payload, and information regarding the relative importance of the NAL unit for the decoding process. The four-byte header that is only used for prefix NAL units and coded slice NAL units of enhancement layers contains additional scalability information. RFC 2250 [i.1] specifies how to carry NAL units in RTP packets.

A.2.4 Transport of VC-1 video

To transport VC-1, RFC 4425 [11] is used. Each RTP packet contains an integer number of Access Units as defined in RFC 4425 [11], which are byte-aligned. Each Access Unit (AU) starts with the AU header, followed by a variable length payload. The AU payload normally contains data belonging to exactly one VC-1 frame. However, the data may be split between multiple AUs if it would otherwise cause the RTP packet to exceed the Maximum Transmission Unit (MTU) size, to avoid IP-level fragmentation.

In the VC-1 Advanced Profile, the sequence layer header contains the parameters required to initialize the VC-1 decoder. These parameters apply to all entry-point segments until the next occurrence of a sequence layer header in the coded bit stream. Neither a sequence layer header nor an entry-point segment header is defined for the VC-1 Simple and Main Profiles. For these profiles, the decoder initialization parameters are conveyed as Decoder Initialization Metadata structures (see annex J of SMPTE 421M [10]) carried in the SDP datagrams signalling the VC-1-based session.

A.2.5 Transport of MPEG-4 HE AAC v2 audio

To transport MPEG-4 AAC, HE AAC or HE AAC v2, RFC 3640 [5] is used. RFC 3640 [5] supports both implicit signalling as well as explicit signalling by means of conveying the AudioSpecificConfig() as the required MIME parameter "config", as defined in RFC 3640 [5].

The framing structure defined in RFC 3640 [5] does support carriage of multiple AAC frames in one RTP packet with optional interleaving to improve error resiliency in packet loss. For example, if each RTP packet carries three AAC frames, then with interleaving the RTP packets may carry the AAC frames as given in figure A.2.

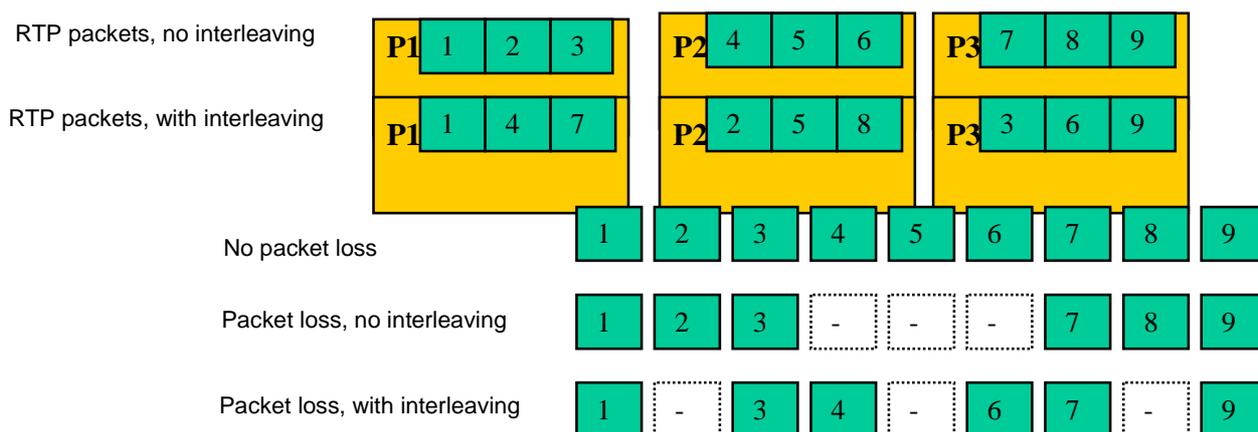


Figure A.2: Interleaving of AAC frames

Without interleaving, then RTP packet P1 carries the AAC frames 1, 2 and 3, while packet P2 and P3 carry the frames 4, 5 and 6 and the frames 7, 8 and 9, respectively. When P2 gets lost, then AAC frames 4, 5 and 6 get lost, and hence the decoder needs to reconstruct three missing AAC frames that are contiguous. In this example, interleaving is applied so that P1 carries 1, 4 and 7, P2 carries 2, 5 and 8, and P3 carries 3, 6 and 9. When P2 gets lost in this case, again three frames get lost, but due to the interleaving, the frames that are immediately adjacent to each lost frame are received and can be used by the decoder to reconstruct the lost frames, thereby exploiting the typical temporal redundancy between adjacent frames to improve the perceptual performance of the receiver.

A.2.6 Transport of MPEG-4 HE AAC v2 in combination with MPEG Surround audio

To transport MPEG-4 AAC or MPEG-4 HE AAC or MPEG-4 HE AAC v2 in combination with MPEG Surround, RFC 3640 [5] and RFC 5691 [21] are used. RFC 3640 [5] supports both implicit signalling as well as explicit signalling of the HE AAC v2 configuration by means of conveying a AudioSpecificConfig() as the required MIME parameter "config", as defined in RFC 3640 [5]. In addition, RFC 5691 [21] supports explicit signalling of the MPEG Surround configuration by means of conveying an AudioSpecificConfig() as the required optional MIME parameter "MPS-config".

Alternatively, MPEG Surround data can be transported in separate RTP packets according to RFC 5691 [21]. RFC 5691 [21] also supports interleaving as described above and defines constraints to ensure that the MPEG-4 AAC or MPEG-4 HE AAC or MPEG-4 HE AAC v2 and MPEG Surround streams are synchronized. RFC 5691 [21] requires explicit signalling of the MPEG Surround configuration by means of conveying the AudioSpecificConfig() as the required MIME parameter "config".

A.2.7 Transport of AMR-WB+ audio

To transport AMR-WB+, RFC 4352 [9] is used. That payload is used also in both 3GPP Release TS 126 234 [i.5] and TS 126 346 [i.8] in which AMR-WB+ is the recommended codec with MPEG-4 HE AAC v2.

The framing structure defined in [9] does support carriage of multiple AMR-WB+ frames in one RTP packet with optional interleaving to improve error resiliency in packet loss. The overhead due to payload starts from three bytes per RTP-packet. The use of interleaving increases the overhead per packet slightly; in minimum 4 bits for each frame in the payload (rounded upwards to full bytes in case of odd number of frames).

For example, if each RTP packet carries three AMR-WB+ frames, then with interleaving the AMR-WB+ packets may carry the AMR-WB+ frames as given in figure A.3.

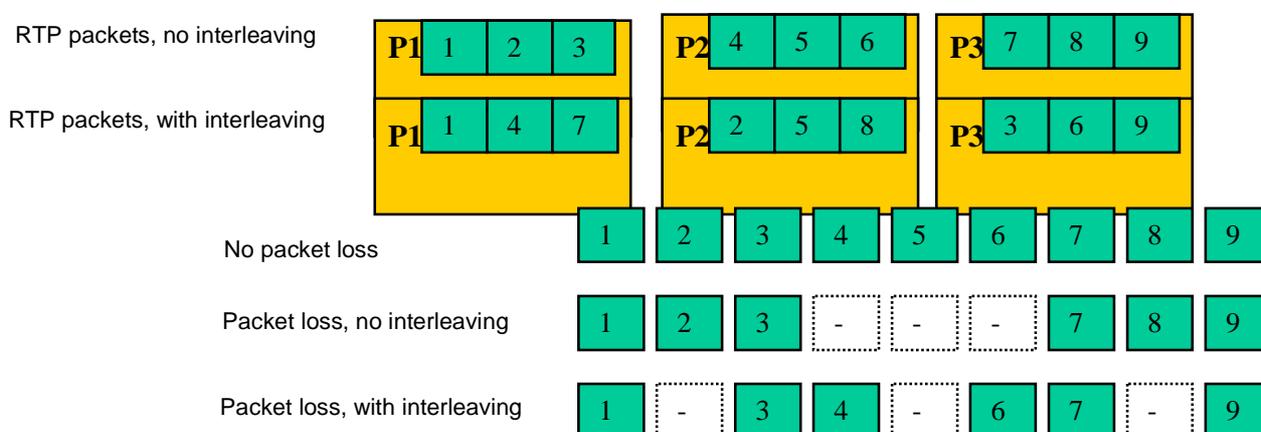


Figure A.3: Interleaving of AMR-WB+ frames

Without interleaving, then RTP packet P1 carries the AMR-WB+ frames 1, 2 and 3, while packet P2 and P3 carry the frames 4, 5 and 6 and the frames 7, 8 and 9, respectively. When P2 gets lost, then AMR-WB+ frames 4, 5 and 6 get lost, and hence the decoder needs to reconstruct three missing AMR-WB+ frames that are contiguous. In this example, interleaving is applied so that P1 carries 1, 4 and 7, P2 carries 2, 5 and 8, and P3 carries 3, 6 and 9. When P2 gets lost in this case, again three frames get lost, but due to the interleaving, the frames that are immediately adjacent to each lost frame are received and can be used by the decoder to reconstruct the lost frames, thereby exploiting the typical temporal redundancy between adjacent frames to improve the perceptual performance of the receiver.

A.2.8 Transport of AC-3 audio

To transport AC-3 audio, RFC 4184 [14] is used. The framing structure defined in RFC 4184 [14] supports carriage of multiple AC-3 frames in one RTP packet. It also supports fragmentation of AC-3 frames in cases where the frame exceeds the Maximum Transmission Unit (MTU) of the network. Fragmentation may take into account the partial frame decoding capabilities of AC-3 to achieve higher resilience to packet loss by setting the fragmentation boundary at the "5/8ths point" of the frame.

A.2.9 Transport of Enhanced AC-3 audio

To transport Enhanced AC-3 audio, RFC 4598 [15] is used. The framing structure defined in RFC 4598 [15] supports carriage of multiple Enhanced AC-3 frames in one RTP packet. Recommendations for concatenation decisions which reduce the impact of packet loss by taking into account the configuration of multiple channels and programs are provided. It also supports fragmentation of Enhanced AC-3 frames in cases where the frame exceeds the MTU of the network.

A.2.10 Synchronization of content delivered over IP

RTP also provides tools for synchronization. For that purpose, an RTP time stamp is present in the RTP header; the RTP time stamps are used to determine the presentation time of the audio and video access units. The method to synchronize content transported in RTP packets is described RFC 3550 [4]. By means of figure A.4 a simplified summary is given below:

- a) RTP time stamps convey the sampling instant of access units at the encoder. The RTP time stamp is expressed in units of a clock, which is required to increase monotonically and linearly. The frequency of this clock is specified for each payload format, either explicitly or by default. Often, but not necessarily, this clock is the sampling clock. In figure A.4, $TSa(i)$ and $TSv(j)$ are RTP time stamps that are used to present the access units at the correct timing at the receiver; this requires that the receiver reconstructs the video clock and audio clock with the same mutual offset in time as at the sender.
- b) When transporting RTP packets, the RTCP Control Protocol, also defined in RFC 3550 [4], is used for purposes such as monitoring and control. RTCP data is carried in RTCP packets. There are several RTCP packet types, one of which is the Sender Report (SR) RTCP packet type. Each RTCP SR packet contains an RTP time stamp and an NTP time stamp; both time stamps correspond to the same instant in time. However, the RTP time stamp is expressed in the same units as RTP time stamps in data packets, while the NTP time stamp is expressed in "wallclock" time; see clause 4 of RFC 3550 [4]. In figure A.4, $NTPa(k)$ and $NTPv(n)$ are the NTP time stamps of the audio and video RTCP packets. $At(k)$ and $Vt(n)$ are the values of the audio and video clock at the same instant in time as $NTPa(k)$ and $NTPv(n)$, respectively. Each $SR(k)$ for audio provides $NTPa(k)$ as NTP time stamp and $At(k)$ as RTP time stamp. Similarly, each $SR(n)$ for video provides $NTPv(n)$ as the NTP time stamps and $Vt(n)$ as RTP time stamp.

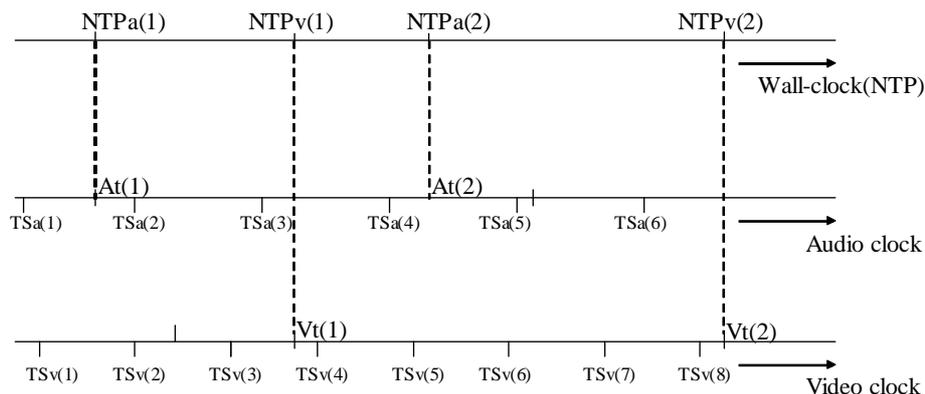


Figure A.4: RTP tools for synchronization

- c) Synchronized playback of streams is only possible if the streams use the same wall-clock to encode NTP values in SR packets. If the same wall-clock is used, receivers can achieve synchronization by using the correspondence between RTP and NTP time stamps. To synchronize an audio and a video stream, one needs to receive an RTCP SR packet relating to the audio stream, and an RTCP SR packet relating to the video stream. These SR packets provide a pair of NTP timestamps and their corresponding RTP timestamps that is used to align the media. For example, in figure A.4, $[NTPv(k) - NTPa(n)]$ represents the offset in time between $Vt(k)$ and $At(n)$, expressed in wallclock time.
- d) The time between sending subsequent RTCP SR packets may vary; the default RTCP timing rules suggest sending an RTCP SR packet every 5 s. This means that upon entering a streaming session there may be an initial delay - on average a 2,5 s duration if the default RTCP timing rules are used - when the receiver does not yet have the necessary information to perform inter-stream synchronization.

A.2.11 Synchronization with content delivered over MPEG-2 TS

Applications may require synchronization of audiovisual content delivered over IP with content delivered over an MPEG-2 TS. For example, a broadcaster may wish to provide audio in another language as part of a broadcast program, but using transport over IP instead of transporting this additional audio stream over the same MPEG-2 TS as the broadcast program.

Synchronization of a stream delivered over IP with a broadcast program requires that the receiver knows the timing relationship between the RTP time stamps of the stream that is delivered over IP and the MPEG-2 time stamps of the broadcast program. It is beyond the scope of the present document how to convey such timing relationship.

A.2.12 Service discovery

For discovery of DVB services over IP it is referred to the IPI specification for low and mid level (PSI/SI equivalent) functionality and to the GBS specification for higher level (SI/metadata related, except structures and containers) functionality.

A.2.13 Linking to applications

Audio and video delivered over IP can be presented in an MHP application by means of including appropriate URLs.

A.2.14 Capability exchange

By means of capability exchange protocols the sender and receiver can communicate whether the receiver has A, B, C, DB, D, E, or F IP-IRD capabilities for H.264/AVC or, except for capability A, for SVC decoding. In addition, it can also be communicated whether the receiver has multi-channel or only mono/stereo capabilities for HE AAC v.2 decoding or whether the receiver supports AMR-WB+, AC-3 or Enhanced AC-3 decoding, and whether decoding of multiple Enhanced AC-3 substreams is supported. For capability exchange protocols it is referred to the IPI specification.

A.3 Video

A.3.1 H.264/AVC video

A.3.1.1 Overview

The part of the H.264/AVC standard referenced in the present document specifies the coding of video (in 4:2:0 chroma format) that contains either progressive or interlaced frames, which may be mixed together in the same sequence. Generally, a frame of video contains two interleaved fields, the top and the bottom field. The two fields of an interlaced frame, which are separated in time by a field period (half the time of a frame period), may be coded separately as two fields or together as a frame. A progressive frame should always be coded as a single frame; however, it can still be considered to consist of two fields at the same instant of time. H.264/AVC covers a Video Coding Layer (VCL), which is designed to efficiently represent the video content, and a Network Abstraction Layer (NAL), which formats the VCL representation of the video and provides header information in a manner appropriate for conveyance by a variety of transport layers or storage media. The structure of H.264/AVC video encoder is shown in figure A.5.

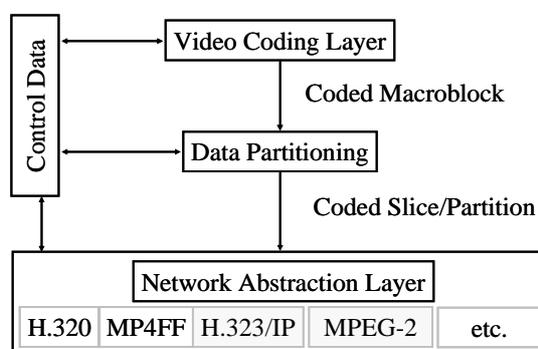


Figure A.5: Structure of H.264/AVC video encoder

A.3.1.2 Network Abstraction Layer (NAL)

The Video Coding Layer (VCL), which is described below, is specified to efficiently represent the content of the video data. The Network Abstraction Layer (NAL) is specified to format that data and provide header information in a manner appropriate for conveyance by the transport layers or storage media. All data are contained in NAL units, each of which contains an integer number of bytes. A NAL unit specifies a generic format for use in both packet-oriented and bitstream systems. The format of NAL units for both packet-oriented transport and bitstream is identical except that each NAL unit can be preceded by a start code prefix in a bitstream-oriented transport layer. The NAL facilitates the ability to map H.264/AVC VCL data to transport layers such as:

- RTP/IP for any kind of real-time wire-line and wireless Internet services (conversational and streaming);
- File formats, e.g. ISO "MP4" for storage and MMS;
- H.32X for wire-line and wireless conversational services;
- MPEG-2 systems for broadcasting services, etc.

The full degree of customization of the video content to fit the needs of each particular application was outside the scope of the H.264/AVC standardization effort, but the design of the NAL anticipates a variety of such mappings.

One key concept of the NAL is parameter sets. A parameter set is supposed to contain information that is expected to rarely change over time. There are two types of parameter sets:

- sequence parameter sets, which apply to a series of consecutive coded video pictures; and
- picture parameter sets, which apply to the decoding of one or more individual pictures.

The sequence and picture parameter set mechanism decouples the transmission of infrequently changing information from the transmission of coded representations of the values of the samples in the video pictures. Each VCL NAL unit contains an identifier that refers to the content of the relevant picture parameter set, and each picture parameter set contains an identifier that refers to the content of the relevant sequence parameter set. In this manner, a small amount of data (the identifier) can be used to refer to a larger amount of information (the parameter set) without repeating that information within each VCL NAL unit.

A.3.1.3 Video Coding Layer (VCL)

The video coding layer of H.264/AVC is similar in spirit to other standards such as MPEG-2 Video. It consists of a hybrid of temporal and spatial prediction in conjunction with transform coding. Figure A.6 shows a block diagram of the video coding layer for a macroblock, which consists of a 16x16 luma block and two 8x8 chroma blocks.

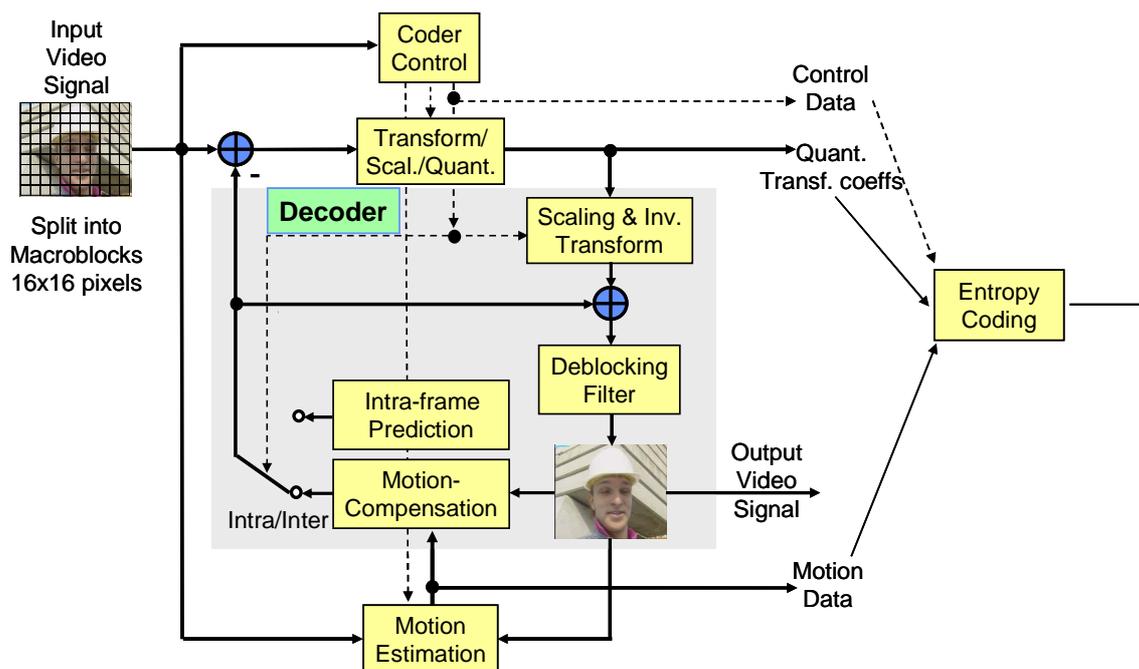


Figure A.6: Basic coding structure for H.264/AVC for a macroblock

In summary, the picture is split into macroblocks. The first picture of a sequence or a random access point is typically coded in Intra, i.e. without using other information than the information contained in the picture itself. Each sample of a luma or chroma block of a macroblock in such an Intra frame is predicted using spatially neighbouring samples of previously coded blocks. The encoding process is to choose which and how neighbouring samples are used for Intra prediction which is simultaneously conducted at encoder and decoder using the transmitted Intra prediction side information.

For all remaining pictures of a sequence or between random access points, typically Inter coding is utilized. Inter coding employs prediction (motion compensation) from other previously decoded pictures. The encoding process for Inter prediction (motion estimation) consists of choosing motion data comprising the reference picture and a spatial displacement that is applied to all samples of the macroblock. The motion data which are transmitted as side information are used by encoder and decoder to simultaneously provide the inter prediction signal.

The residual of the prediction (either Intra or Inter) which is the difference between the original and the predicted macroblock is transformed. The transform coefficients are scaled and quantized. The quantized transform coefficients are entropy coded and transmitted together with the side information for either Intra-frame or Inter-frame prediction.

The encoder contains the decoder to conduct prediction for the next blocks or next picture. Therefore, the quantized transform coefficients are inverse scaled and inverse transformed in the same way as at the decoder side resulting in the decoded prediction residual. The decoded prediction residual is added to the prediction. The result of that addition is fed into a deblocking filter which provides the decoded video as its output.

The new features of H.264/AVC compared to MPEG-2 Video are listed as follows: variable block-size motion compensation with small block sizes from 16x16 luma samples down to 4x4 luma samples per block, quarter-sample-accurate motion compensation, motion vectors pointing over picture boundaries, multiple reference picture motion compensation, decoupling of referencing order from display order, decoupling of picture representation methods from picture referencing capability, weighted prediction, improved "skipped" and "direct" motion inference, directional spatial prediction for intra coding, in-the-loop deblocking filtering, 4x4 block-size transform, hierarchical block transform, short word-length/exact-match inverse transform, context-adaptive binary arithmetic entropy coding, flexible slice size, FMO, ASO, redundant pictures, data partitioning, SP/SI synchronization/switching pictures.

A.3.1.4 Explanation of H.264/AVC profiles and levels

Profiles and levels specify conformance points. These conformance points are designed to facilitate interoperability between various applications of the standard that have similar functional requirements. A *profile* specifies a set of coding tools or algorithms that can be used in generating a conforming bit-stream, whereas a *level* places constraints on certain key parameters of the bitstream. All decoders conforming to a specific profile should support all features in that profile. Encoders are not required to make use of any particular set of features supported in a profile but have to provide conforming bitstreams, i.e. bitstreams that can be decoded by conforming decoders.

The first version of H.264/AVC was published in May 2003 by ITU-T as Recommendation H.264 [1] and by ISO/IEC 14496-10 [2]. Three Profiles define sub-sets of the syntax and semantics:

- Baseline Profile.
- Extended Profile.
- Main Profile.

The Fidelity Range Extensions Amendment of H.264/AVC, agreed in July 2004, added some additional tools and defined four new Profiles (of which only the first is relevant for the present document):

- High Profile.
- High 10 Profile.
- High 4:2:2 Profile.
- High 4:4:4 Profile.

The relationship between High Profile and the original three Profiles, in terms of the major tools from the toolbox that may be used, is illustrated by figure A.7.

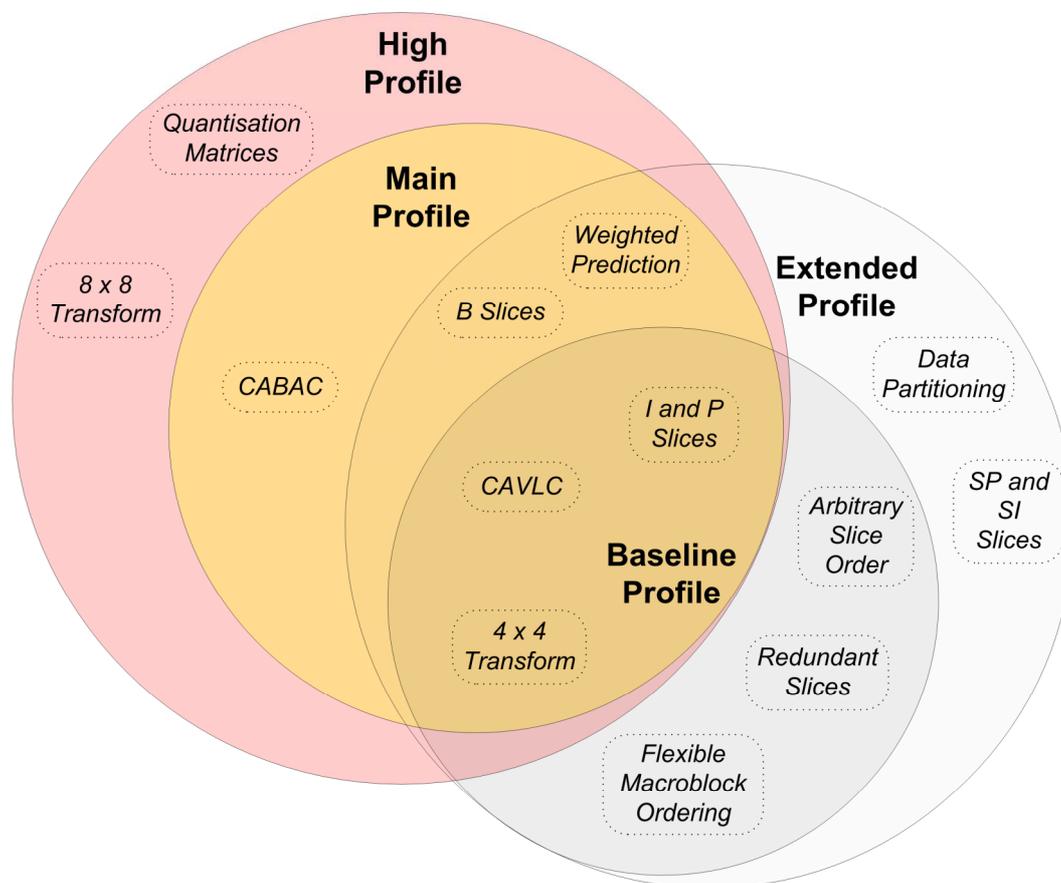


Figure A.7: Relationship between high profile and the three original profiles

The present document only uses Baseline, Main, and High Profile. These contain the following features:

Baseline Profile:

The Baseline Profile contains the following restricted set of coding features:

- I and P Slices: Intra coding of macroblocks through the use of I slices; P slices add the option of Inter coding using one temporal prediction signal.
- 4x4 Transform: The prediction residual is transformed and quantized using 4x4 blocks.
- CAVLC: The symbols of the coder (e.g. quantized transform coefficients, intra predictors, motion vectors) are entropy-coded using a variable length code.
- FMO: This feature of Baseline allowing arbitrary sampling of the Macroblocks within a slice is not used in the present document. The main reason is to achieve decodability by Main or High profile decoders, which is signalled by `constrained_set1_flag` being equal to 1.
- ASO: This feature of Baseline allowing arbitrary order of slices within a picture is not used in the present document. The main reason is to achieve decodability by Main or High profile decoders, which is signalled by `constrained_set1_flag` being equal to 1.
- Redundant Slices: This feature of Baseline allowing transmission of a redundant slice that approximates the primary slice is not used in the present document. The main reason is to achieve decodability by Main or High profile decoders, which is signalled by `constrained_set1_flag` being equal to 1.

Main Profile:

Except for FMO, ASO, and Redundant Slices, Main Profile contains all features of Baseline Profile and the following additional ones:

- B Slices: Enhanced Inter coding using up to two temporal prediction signals that are superimposed for the predicted block.
- Weighted Prediction: Allowing the temporal prediction signal in P and B slices to be weighted by a factor.
- CABAC: An alternative entropy coding to CAVLC providing higher coding efficiency at higher complexity, which is based on context-adaptive binary arithmetic coding.

High Profile:

High Profile contains all features of Main Profile and the following additional ones:

- 8x8 Transform: In addition to the 4x4 Transform, the encoder can choose to code the prediction residual using a, 8x8 Transform.
- Quantization Matrix: The encoder can choose to apply weights to the transform coefficients, which provides a weighted fidelity of reproduction for these.

A.3.1.5 Summary of key tools and parameter ranges for capability A to F IRDs

Table A.1 summarizes the assignment of profiles and levels to the seven IP-IRDs that are specified in the present document.

Table A.1

Capability	Mandatory profile	Optional profile	Additional constraint on mandatory profile	Level	Max frame size (macro-blocks)	Example video formats	Max VCL bit rate (kbit/s)
A	Baseline	Main or High	constraint_set1_flag = 1	1b	99	176 x 144, 15Hz	128
B	Baseline	Main or High	constraint_set1_flag = 1	1.2	396	352 x 288, 15Hz QCIF = 176 x 144, 30Hz	384
C	Baseline	Main or High	constraint_set1_flag = 1	2	396	CIF = 352 x 288, 30Hz	2 000
DB	Baseline	Main or High	constraint_set1_flag = 1	3	1 620	625 SD = 720 x 576, 25Hz 525 SD = 720 x 480, 30Hz	10 000
D	Main	High	None	3	1 620	625 SD = 720 x 576, 25Hz 525 SD = 720 x 480, 30Hz	10 000
E	High	-	None	4	8 192	1 080i HD = 1 920 x 1 080, 25/30Hz 720p HD = 1 280 x 720, 50/60Hz	25 000
F	High	-	None	4.2	8 704	1080p HD = 1 920 x 1 080, 50/60Hz	62 500

The following should be noted.

IP-IRDs with Capability A, B, C, and DB specify the Baseline profile with the additional constraint that `constraint_set1_flag` should be set equal to 1 making these bitstreams also decodable by Main or High profile decoders. The reason for this additional constraint is that our investigations have shown that the features that are contained in Baseline but are not contained in Main profile (FMO, ASO, and redundant pictures) and are disabled by setting `constraint_set1_flag` equal to 1 do not provide any benefit at the packet error rates envisioned to be typical for the applications in which the present document will be used. IP-IRDs with capability D should be conforming to Main profile without any additional constraints. IP-IRDs with capability E or F should be conforming to High profile without any additional constraints.

Because of the additional constraint and the requirements in H.264/AVC, IP-IRDs labelled with a particular capability Y are capable of decoding and rendering pictures that can be decoded by IP-IRDs labelled with a particular capability X with X appearing in the following ordered sequence at an earlier position than Y: A, B, C, DB, D, E, F.. For instance, Capability D IP-IRDs are capable of decoding bitstreams conforming to Main Profile at level 3 of H.264/AVC and below. Additionally, Capability D IP-IRDs are capable of decoding bitstreams that are also decodable by IP-IRDs with capabilities A, B, C, or DB.

In addition to the mandatory requirements on IP-IRDs and Bitstreams, the optional use of the following Bitstreams is allowed given that the IP-IRD is capable of decoding it. For Capability A, B, C, and DB Bitstreams, encoders may optionally generate Main or High Profile bitstreams. For Capability D Bitstreams, encoders may optionally generate High Profile bitstreams.

Each level specifies a maximum number of macroblocks per second that can be processed by a corresponding decoder (not explicitly listed in table A.1). Additionally, the maximum number of macroblocks per frame is restricted as well. For example, for the Capability D IP-IRD, the maximum number of macroblocks per frame is given as 1 620 corresponding to a 625 SD picture (level 3 of H.264/AVC). Together with the maximum number of macroblocks per second that can be processed which are given as 40 500, the maximum frame rate is given as 25 frames per second. Please note that this also permits the processing of 525 SD pictures at 30 frames per second.

A.3.1.6 Other video parameters

The present document is supposed to cover a large variety of applications. Therefore, we do not specify parameters such as frame rate, aspect ratio, chromaticity, chroma, and random access points as restrictively as they are specified in TS 101 154 [i.2].

For parameters such as frame rate and aspect ratio, the constraints as specified in H.264/AVC are sufficient and need no further adjustment. It is only recommended to avoid extreme values.

For parameters such as chromaticity and chroma, it is recommended to utilize the parameters that are specified in the VUI of H.264/AVC which is part of the sequence parameter set.

Random access points are provided through so-called Instantaneous Decoding Refresh (IDR) pictures. In our recommendations, we distinguish broadcast and other applications. For broadcast applications it is recommended that sequence and picture parameter sets are sent together with a random access point (e.g. an IDR picture) to be encoded at least once every 500 ms. For multicast or streaming applications a maximum interval of 5 s between random access points should not be exceeded.

A.3.2 SVC video

A.3.2.1 Overview

The basic SVC design can be classified as layered video codec. In each layer, the basic concepts of motion-compensated prediction and intra prediction are employed as in H.264/AVC. The redundancy between different layers is exploited by additional inter-layer prediction concepts that include prediction mechanisms for macroblock mode information, motion parameters, and texture data (intra and residual data). Each SVC bit stream includes a sub-stream, which is compliant to one or more non-scalable profile of H.264/AVC.

Similar to the underlying H.264/AVC standard, the SVC design includes a Video Coding Layer (VCL) and a Network Abstraction Layer (NAL). While the VCL represents the coded source content, the NAL formats the VCL representation and provides header information appropriate for conveyance by transport layers and storage media.

Scalability is provided at the bit stream level. A bit stream with reduced spatio and/or temporal resolution and/or fidelity can be obtained by discarding NAL units from a scalable bit stream. The NAL units that are required for decoding of a specific spatio-temporal resolution and bit rate are identified by syntax elements inside the NAL unit header or by a preceding so-called prefix NAL unit.

A.3.2.2 Network Abstraction Layer (NAL)

As for H.264/AVC, the Network Abstraction Layer (NAL) is specified to format that data and provide header information in a manner appropriate for conveyance by the transport layers or storage media. All data are contained in NAL units, each of which contains an integer number of bytes. A NAL unit specifies a generic format for use in both packet-oriented and bitstream systems. The format of NAL units for both packet-oriented transport and bitstream is identical except that each NAL unit can be preceded by a start code prefix in a bitstream-oriented transport layer. The NAL facilitates the ability to map SVC VCL data to transport layers.

The full degree of customization of the video content to fit the needs of each particular application was outside the scope of the H.264/AVC and SVC standardization effort, but the design of the NAL anticipates a variety of such mappings.

One key concept of the NAL for H.264/AVC and SVC is parameter sets. A parameter set is supposed to contain information that is expected to rarely change over time. In SVC, there are three types of parameter sets:

- sequence parameter sets, which apply to the base layer of a series of consecutive coded video pictures;
- subset sequence parameter sets, which apply to one or more enhancement layers of a series of consecutive coded video pictures;
- picture parameter sets, which apply to the decoding of one or more layer representations of individual pictures.

The parameter set mechanism decouples the transmission of infrequently changing information from the transmission of coded representations of the values of the samples in the video pictures. Each VCL NAL unit contains an identifier that refers to the content of the relevant picture parameter set, and each picture parameter set contains an identifier that refers to the content of the relevant sequence parameter set (for base layer VCL NAL units) or subset sequence parameter set (for enhancement layer VCL NAL units). In this manner, a small amount of data (the identifier) can be used to refer to a larger amount of information (the parameter set) without repeating that information within each VCL NAL unit.

In contrast to the non-scalable profiles of H.264/AVC, the NAL concept for SVC was extended to provide a mechanism for easy bitstream manipulation and association of NAL units to scalable layers. The one-byte NAL unit header of H.264/AVC is extended by additional three bytes for the SVC NAL unit types. This extended header includes the parameters required for identifying the scalable layer (corresponding to a particular temporal resolution, spatial resolution, and fidelity) to which a VCL NAL units belongs as well as additional information assisting bit stream adaptations. Each SVC bit stream includes a sub-stream, which is compliant to a non-scalable profile of H.264/AVC. Standard H.264/AVC NAL units (non-SVC NAL units) do not include the extended SVC NAL unit header. However, these data are not only useful for bit stream adaptations, but some of them are also required for the SVC decoding process. In order to attach this SVC related information to non-SVC NAL units, so-called prefix NAL units are introduced. These NAL units directly precede all non-SVC VCL NAL units in an SVC bit stream and contain the SVC NAL unit header extension.

A.3.2.3 Video Coding Layer (VCL)

A simplified block diagram of the SVC coding structure is illustrated below. Each representation of the video source with a particular spatial resolution and fidelity that is included in an SVC bit stream is referred to as a layer (shaded area in figure A.8) and is characterized by a layer identifier. In each access unit, the layers are encoded in increasing order of their layer identifiers. For the coding of a layer, already transmitted data of another layer with a smaller layer identifier can be employed as described in the following paragraphs. The layer to predict from can be selected on an access unit basis and is referred to as the reference layer. The layer with a layer identifier equal to 0, which may only be present in some access units, is coded in conformance with one of the non-scalable H.264/AVC profiles and is referred to as base layer. The layers that employ data of other layers for coding are referred to as enhancement layers. An enhancement layer is called spatial enhancement layer when the spatial resolution changes relative to its reference layer; and it is called fidelity enhancement layer when the spatial resolution is identical to that of its reference layer.

The number of layers present in a SVC bit stream is dependent on the needs of an application. SVC supports up to 128 layers in a bit stream. With the currently specified profiles, the maximum number of enhancement layers in a bit stream is limited to 47 and at most 2 of those can represent spatial enhancement layers.

Similarly to H.264/AVC, the input pictures of each spatial or fidelity layer are split into macroblocks and slices. A macroblock represents a square area of 16x16 luma samples and 8x8 samples of the two chroma components. The macroblocks are organized in slices, which can be parsed independently. For the purpose of intra prediction, motion-compensated prediction, and transform coding, a macroblock can be split into smaller partitions or blocks.

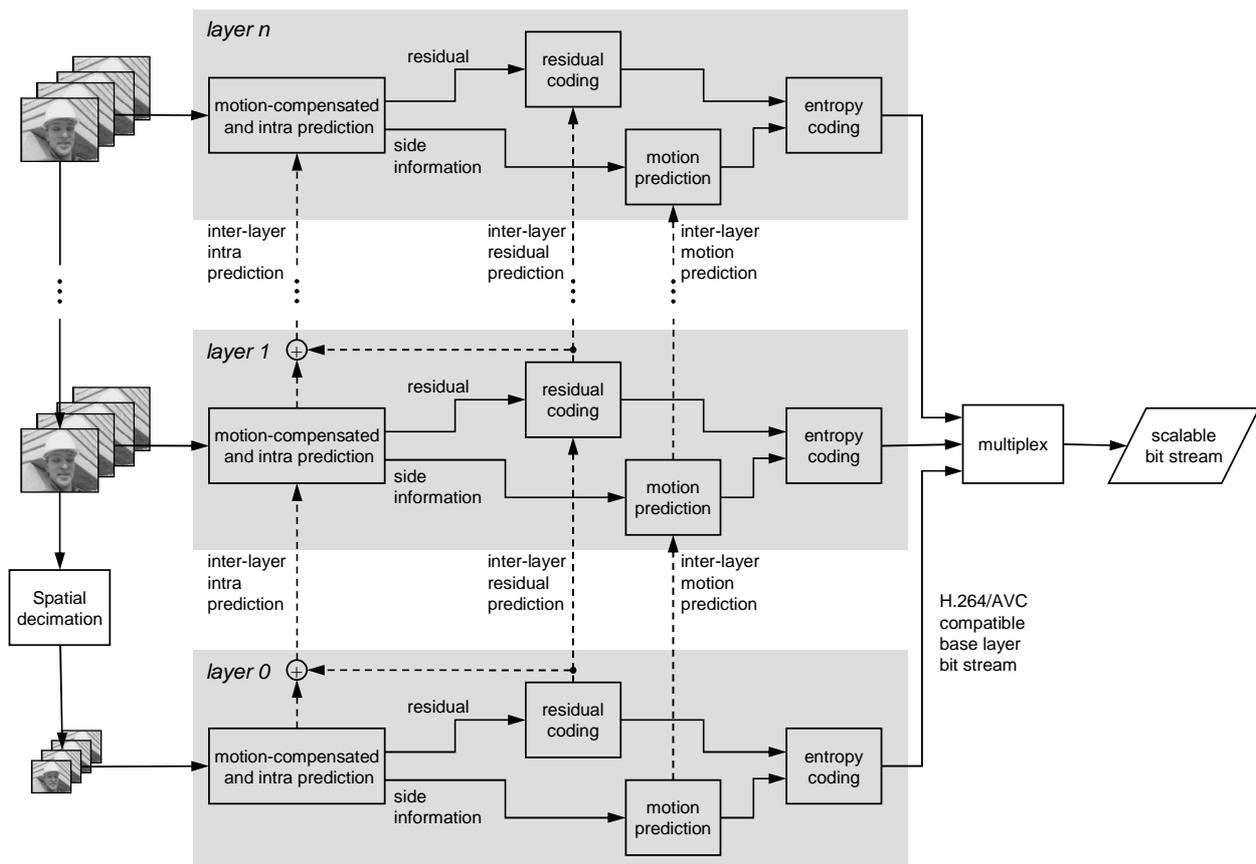


Figure A.8: Basic coding structure for H.264/AVC for a macroblock

Inside each layer, the SVC design basically follows the design of the underlying H.264/AVC standard for single-layer coding. The samples of each macroblock are either predicted by intra-picture or inter-picture prediction. With intra-picture prediction, each sample of a block is predicted using spatially neighbouring samples of previously coded blocks in the same picture. With inter-picture prediction, the prediction signal of a partition is built by a spatially displaced region of a previously coded picture of the same layer. The residual representing the difference between the original and the prediction signal for a block is transformed using a decorrelating transform. The transform coefficients are scaled and quantized. The quantized transform coefficients are entropy coded together with other information including the macroblock coding type, the quantization step size, and the intra prediction modes or the motion information consisting of identifiers specifying the employed reference pictures and corresponding displacement (or motion) vectors. The motion vector components are differentially coded using motion vectors of neighbouring blocks as predictors. The decoded representation of the residual is obtained by inverse scaling and inverse transformation of the quantized transform coefficients. The obtained decoded residual is then added to the prediction signal, and the result is additionally processed by a deblocking filter before output and potential storage as a reference picture for inter-picture coding of following pictures.

In addition to these basic coding tools of H.264/AVC, SVC provides so-called inter-layer prediction methods, which allow an exploitation of the statistical dependencies between different layers for improving the coding efficiency (reducing the bit rate) of enhancement layers. All inter-layer prediction tools can be chosen on a macroblock or sub-macroblock basis allowing an encoder to select the coding mode that gives the highest coding efficiency. For SVC enhancement layers, an additional macroblock coding mode is provided, in which the macroblock prediction signal is completely inferred from co-located blocks in the reference layer without transmitting any additional side information. When the co-located reference layer blocks are intra-coded, the prediction signal is built by the (for spatial scalable coding) potentially up-sampled reconstructed intra signal of the reference layer - a prediction method also referred to as inter-layer intra prediction. Otherwise, the enhancement layer macroblock is inter-picture predicted as described above, but the macroblock partitioning - specifying the decomposition into smaller block with different motion parameters - and the associated motion parameters are completely derived from the co-located blocks in the reference layer. This concept is also referred to as inter-layer motion prediction. For the conventional inter-coded macroblock types of H.264/AVC, the (scaled) motion vector of the reference layer blocks can also be used as replacement for usual spatial motion vector predictor. A further inter-layer prediction tool referred to as inter-layer residual prediction targets a reduction of the bit rate required for transmitting the residual signal of inter-coded macroblocks. With the usage of residual prediction, the (up-sampled) residual of the co-located reference layer blocks is subtracted from the enhancement layer residual (difference between the original and the inter-picture prediction signal) and only the resulting difference, which often has a smaller energy than the original residual signal, is encoded using transform coding as described above. For fidelity enhancement layers, the inter-layer intra and residual prediction are performed in the transform coefficient domain in order to avoid multiple inverse transform operations at the decoder side.

As an important feature of the SVC design, each spatial and fidelity enhancement layer can be decoded with a single motion compensation loop. For the employed reference layers, only the intra-coded macroblocks and residual blocks that are used for inter-layer prediction need to be reconstructed and the motion vectors need to be decoded. The computationally complex operations of motion-compensated prediction and deblocking only need to be performed for the target layer to be displayed.

Temporal scalability can be achieved by partitioning the access units into a temporal base and one or more temporal enhancement layers and restricting the encoding structure in a way that for each access unit of a specific temporal layer, only access units of the same or a coarser temporal layer are employed for inter-picture prediction.

A.3.2.4 Explanation of SVC profiles and levels

As for H.264/AVC, profiles and levels specify conformance points. These conformance points are designed to facilitate interoperability between various applications of the standard that have similar functional requirements. A *profile* specifies a set of coding tools or algorithms that can be used in generating a conforming bit-stream, whereas a *level* places constraints on certain key parameters of the bitstream. All decoders conforming to a specific profile should support all features in that profile. Encoders are not required to make use of any particular set of features supported in a profile but have to provide conforming bitstreams, i.e. bitstreams that can be decoded by conforming decoders.

ITU-T Recommendation H.264 [1] and ISO/IEC 14496-10 [2] define three SVC profiles:

- Scalable Baseline Profile.
- Scalable High Profile.
- Scalable High Intra Profile.

The present document only uses the Scalable Baseline and Scalable High Profile. These contain the following features:

Scalable Baseline Profile:

The Scalable Baseline Profile contains the following set of coding features:

- Base layer bitstream is conforming to the H.264/AVC Baseline profile with `constraint_set1_flag` equal to 1.
- Enhancement layers support all features of the Baseline profile with `constraint_set1_flag` equal to 1 and additionally the following features:
 - Inter-layer prediction.
 - B slices.
 - Multiple slice groups (FMO) with `slice_group_map_type` equal to 2.

- Weighted prediction.
 - Arbitrary slice order (ASO).
 - Redundant slices.
 - Weighting of transform coefficients (quantization matrices).
 - 8x8 transform: only for levels greater than 2.1.
 - CABAC: only for levels greater than 2.1.
- The support for spatial scalable coding is restricted to resolution ratios of 1.5 and 2 between successive spatial layers in both horizontal and vertical direction and to macroblock-aligned cropping.

Scalable High Profile:

The Scalable High Profile contains the following set of coding features:

- Base layer bitstream is conforming to the H.264/AVC High profile.
- Enhancement layers support all features of the High profile and additionally support inter-layer prediction.
- Spatial scalable coding is supported with arbitrary resolution ratios.

A.3.2.5 Summary of key tools and parameter ranges for capability B to F IRDs

Table A.2 summarizes the assignment of profiles and levels to the six IP-IRDs that are specified in the present document.

Table A.2

Capability	Mandatory profile	Optional profile	Level	Max frame size (macro-blocks)	Example video formats	Max VCL bit rate (kbit/s)
B	Scalable Baseline	Scalable High	1.2	396	352 x 288, 15Hz QCIF = 176 x 144, 30Hz	480
C	Scalable Baseline	Scalable High	2	396	CIF = 352 x 288, 30Hz	2 500
DB	Scalable Baseline	Scalable High	3	1 620	625 SD = 720 x 576, 25Hz 525 SD = 720 x 480, 30Hz	12 500
D	Scalable High	-	3	1 620	625 SD = 720 x 576, 25Hz 525 SD = 720 x 480, 30Hz	12 500
E	Scalable High	-	4	8 192	1 080i HD = 1 920 x 1 080, 25/30Hz 720p HD = 1 280 x 720, 50/60Hz	25 000
F	Scalable High	-	4.2	8 704	1 080p HD = 1 920 x 1 080, 50/60Hz	62 500

The following should be noted.

IP-IRDs with Capability B, C, and DB should be conforming to the Scalable Baseline profile. IP-IRDs with capability D, E or F should be conforming to the Scalable High profile.

IP-IRDs labelled with a particular capability Y are capable of decoding and rendering pictures that can be decoded by IP-IRDs labelled with a particular capability X with X appearing in the following ordered sequence at an earlier position than Y: A, B, C, DB, D, E, F. For instance, Capability D IP-IRDs are capable of decoding bitstreams conforming to Scalable High Profile at level 3 and below. Additionally, Capability D IP-IRDs are capable of decoding bitstreams that are also decodable by IP-IRDs with capabilities B, C, or DB.

IP-IRDs supporting SVC and labelled with a particular capability X are additionally capable of decoding H.264/AVC bitstreams that are decodable by IP-IRDs with the capability X that support H.264/AVC.

In addition to the mandatory requirements on IP-IRDs and Bitstreams, the optional use of the following Bitstreams is allowed given that the IP-IRD is capable of decoding it. For Capability A, B, C, and DB Bitstreams, encoders may optionally generate Scalable High Profile bitstreams.

Each level specifies a maximum number of macroblocks per second that can be processed by a corresponding decoder (not explicitly listed in table A.2). Additionally, the maximum number of macroblocks per frame is restricted as well. For example, for the Capability D IP-IRD, the maximum number of macroblocks per frame is given as 1 620 corresponding to a 625 SD picture (level 3). Together with the maximum number of macroblocks per second that can be processed which are given as 40 500, the maximum frame rate is given as 25 frames per second. Please note that this also permits the processing of 525 SD pictures at 30 frames per second.

A.3.2.6 Other video parameters

The present document is supposed to cover a large variety of applications. Therefore, we do not specify parameters such as frame rate, aspect ratio, chromaticity, chroma, and random access points as restrictively as they are specified in TS 101 154 [i.2].

For parameters such as frame rate and aspect ratio, the constraints as specified in H.264/AVC (which includes SVC) are sufficient and need no further adjustment. It is only recommended to avoid extreme values.

For parameters such as chromaticity and chroma, it is recommended to utilize the parameters that are specified in the VUI of the H.264/AVC specification which is part of the sequence parameter set and subset sequence parameter set.

Random access points are provided through so-called Instantaneous Decoding Refresh (IDR) pictures. In our recommendations, we distinguish broadcast and other applications. For broadcast applications it is recommended that sequence and picture parameter sets are sent together with a random access point (e.g. an IDR picture) to be encoded at least once every 500 ms for at least the base layer. For multicast or streaming applications a maximum interval of 5 s between random access points in the base layer should not be exceeded.

A.3.3 VC-1 video

A.3.3.1 Overview

The VC-1 bit stream is defined as a hierarchy of layers. This is conceptually similar to the notion of a protocol stack of networking protocols. The outermost layer is called the sequence layer. The other layers are entry-point, picture, slice, macroblock and block. In the Simple and Main profiles, a sequence in the sequence layer consists of a series of one or more coded pictures. In the Advanced profile, a sequence consists of one or more entry-point segments, where each entry-point segment consists of a series of one or more pictures, and where the first picture in each entry-point segment provides random access.

In the VC-1 Advanced Profile, the sequence layer header contains the parameters required to initialize the VC-1 decoder. These parameters apply to all entry-point segments until the next occurrence of a sequence layer header in the coded bit stream. For Simple and Main Profiles, the decoder initialization parameters are conveyed as Decoder Initialization Metadata structures (see annex J of SMPTE 421M [10]) carried in the SDP datagrams signalling the VC-1-based session, rather than via a sequence layer header and an entry-point segment header. Therefore, all IP IRDs supporting VC-1 should be capable of extracting this data from the SDP datagrams.

A.3.3.2 Explanation of VC-1 profiles and levels

As with MPEG-2 and H.264/AVC, Profiles and Levels are used to specify conformance points for VC-1. A profile defines a sub-set of the VC-1 standard which include a specific set of coding tools and syntax. A level is a defined set of constraints on the values which can be taken by key parameters (such as bit rate or video resolution) within a particular profile. A decoder claiming conformance to a specific profile should support all features in that profile. Encoders are not required to make use of any particular set of features supported in a profile but have to provide conforming bitstreams, i.e. bitstreams that can be decoded by conforming decoders.

Three profiles have been specified: Simple, Main and Advanced. For each profile a number of levels have been defined: two levels with Simple Profile, three levels with Main Profile and five levels with Advanced Profile. Note that VC-1 levels have been defined to be specific to particular profiles; this is in contrast with MPEG-2 and H.264/AVC where levels are largely independent of profiles.

Table A.3 summarizes the coding tools that are included in each profile.

Table A.3

Feature	Simple Profile	Main Profile	Advanced Profile
Baseline intra frame compression	✓	✓	✓
Variable-sized transform	✓	✓	✓
16-bit transform	✓	✓	✓
Overlapped transform	✓	✓	✓
4 motion vector per macroblock	✓	✓	✓
1/4 pixel luminance motion compensation	✓	✓	✓
1/4 pixel chrominance motion compensation		✓	✓
Start codes		✓	✓
Extended motion vectors		✓	✓
Loop filter		✓	✓
Dynamic resolution change		✓	✓
Adaptive macroblock quantization		✓	✓
B frames		✓	✓
Intensity compensation		✓	✓
Range adjustment		✓	✓
Field and frame coding modes			✓
GOP Layer			✓
Display metadata			✓

The Advanced Profile bitstream includes a number of fields which provide information useful to the post-decode display process. This information, collectively known as "display metadata" is output by the decoding process. Its use in the display process is optional, but recommended.

A.3.3.3 Summary of key tools and parameter ranges for capability A to E IRDs

Five combinations of profile and level have been defined in the present document as VC-1 IP-IRDs with Capability A to E. The combinations of VC-1 profile and level for each of the five Capabilities have been chosen to facilitate the design of an IP-IRD that has the computational resource required to support both H.264/AVC and VC-1 at the same Capability. However, the differences between the two standards mean that this alignment cannot be guaranteed.

Table A.4 summarizes the assignment of profiles and levels to the five IP-IRDs that are specified in the present document.

Table A.4

Capability	Profile	Level	Max frame size (macroblocks)	Example Video Formats	Max bit rate (kbit/s)
A	Simple	LL	99	176 x 144, 15 Hz	96
B	Simple	ML	396	352 x 288, 15 Hz 320 x 240, 24 Hz QCIF = 176 x 144, 30 Hz	384
C	Advanced	L0	396	CIF = 352 x 288, 30 Hz	2 000
D	Advanced	L1	1,620	625 SD = 720 x 576, 25 Hz 525 SD = 720 x 480, 30 Hz	10,000
E	Advanced	L3	8,192	1 080i HD = 1 920 x 1 080, 25/30 Hz 720p HD = 1 280 x 720, 50/60 Hz	45,000

Note that IP-IRDs labelled with a particular capability Y are capable of decoding and rendering pictures that can be decoded by IP-IRDs labelled with a particular capability X with X being an earlier letter than Y in the alphabet. For instance, Capability D IP-IRDs are capable of decoding bitstreams conforming to Advanced Profile at L1 of VC-1 and below. Additionally, Capability D IP-IRDs are capable of decoding bitstreams that are also decodable by IP-IRDs with capabilities A, B, or C.

A.4 Audio

A.4.1 MPEG-4 AAC, HE AAC, HE AAC v2, and MPEG Surround

The principle problem of traditional perceptual audio codecs at low bit rates is, that they would need more bits to encode the whole spectrum accurately than available. The results are either coding artefacts or the transmission of a reduced bandwidth audio signal. To resolve this problem, MPEG decided to add a bandwidth extension technology as a new tool to the MPEG-4 audio toolbox. With SBR the higher frequency components of the audio signal are reconstructed at the decoder based on transposition and additional helper information. This method allows an accurate reproduction of the higher frequency components with a much higher coding efficiency compared to a traditional perceptual audio codec. Within MPEG the resulting audio codec is called MPEG-4 High Efficiency AAC (HE AAC) and is the combination of the MPEG-4 Audio Object Types AAC-Low Complexity (LC) and Spectral Band Replication (SBR). It is not a replacement for AAC, but rather a superset which extends the reach of high-quality MPEG-4 Audio to much lower bitrates. HE AAC decoders will decode both, plain AAC and the enhanced AAC plus SBR. The result is a backward compatible extension of the standard.

The basic idea behind SBR is the observation that usually a strong correlation between the characteristics of the high frequency range of a signal (further referred to as "highband") and the characteristics of the low frequency range (further referred to as "lowband") of the same signal is present. Thus, a good approximation of the representation of the original input signal highband can be achieved by a transposition from the lowband to the highband. In addition to the transposition, the reconstruction of the highband incorporates shaping of the spectral envelope. This process is controlled by transmission of the highband spectral envelope of the original input signal. Additional guidance information for the transposing process is sent from the encoder, which controls means, such as inverse filtering, noise and sine addition. This transmitted side information is further referred to as SBR data.

In June 2004 MPEG extended its toolbox with the Audio Object Type Parametric Stereo (PS), which enables stereo coding at very low bitrates. The principle behind the PS tool is to transmit a mono signal coded in HE AAC format together with a description of the stereo image. The PS tool is used at bit rates in the low bit rate range. The resulting MPEG profile is called MPEG-4 HE AAC v2. Figure A.9 shows the different MPEG tools used in the MPEG-4 HE AAC v2 profile. An MPEG-4 HE AAC v2 decoder will decode all three profiles; MPEG-4 AAC, HE AAC and HE AAC v2.

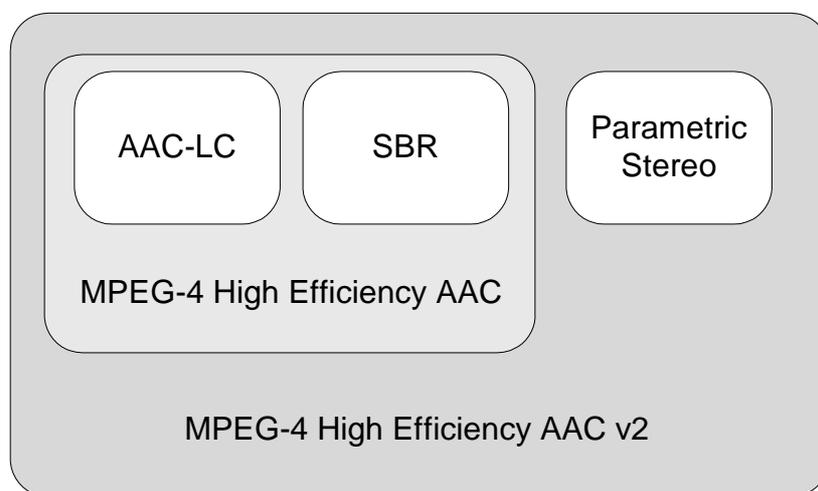


Figure A.9: MPEG tools used in the MPEG-4 HE AAC v2 profile

Figure A.10 shows a block diagram of a MPEG-4 HE AAC v2 Encoder. At the lowest bitrates the PS tool is used. At higher bitrates, normal stereo operation is performed. The PS encoding tool estimates the parameters characterizing the perceived stereo image of the input signal. These parameters are embedded in the SBR data. If the PS tool is used, a stereo to mono downmix of the input signal is applied, which is then fed into the aacPlus encoder operating in mono. SBR data is embedded into the AAC bitstream by means of the `extension_payload()` element. Two types of SBR extension data can be signalled through the `extension_type` field of the `extension_payload()`. For compatibility reasons with existing AAC only decoders, two different methods for signalling the existence of an SBR payload can be selected. Both methods are described below.

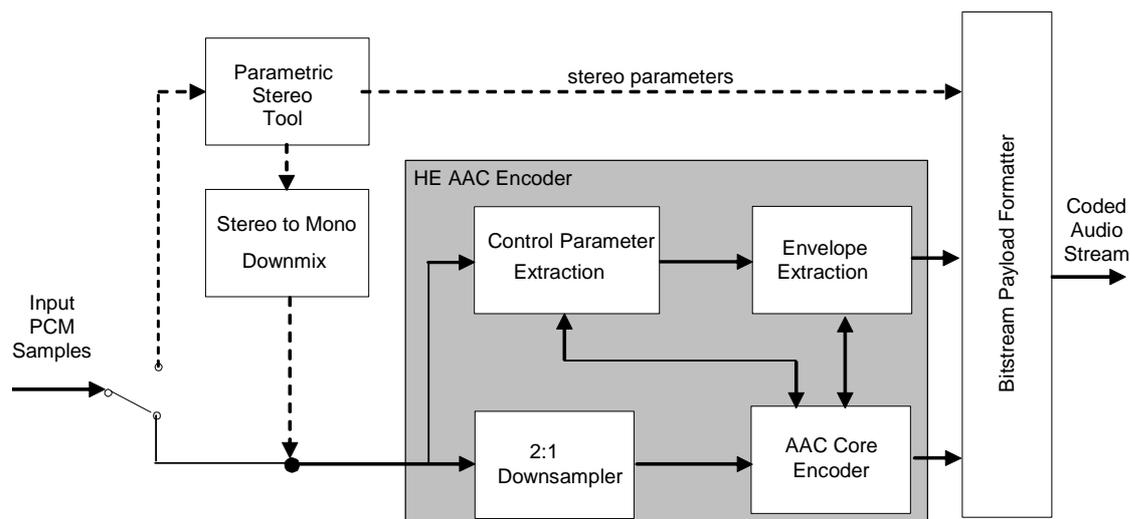


Figure A.10: MPEG-4 HE AAC v2 encoder

The MPEG-4 HE AAC v2 decoder is depicted in figure A.11. The coded audio stream is fed into a demultiplexing unit prior to the AAC decoder and the SBR decoder. The AAC decoder reproduces the lower frequency part of the audio spectrum. The time domain output signal from the underlying AAC decoder at the sampling rate $f_{s_{AAC}}$ is first fed into a 32 channel Quadrature Mirror Filter (QMF) analysis filterbank. Secondly, the high frequency generator module recreates the highband by patching QMF subbands from the existing low band to the high band. Furthermore, inverse filtering is applied on a per QMF subband basis, based on the control data obtained from the bitstream. The envelope adjuster modifies the spectral envelope of the regenerated highband, and adds additional components such as noise and sinusoids, all according to the control data in the bitstream. In case of a stream using Parametric Stereo, the mono output signal from the underlying HE AAC decoder is converted into a stereo signal. This processing is carried out in the QMF domain and is controlled by the Parametric Stereo parameters embedded in the SBR data. Finally a 64 channel QMF synthesis filterbank is applied to retain a time-domain output signal at twice the sampling rate, i.e. $f_{s_{out}} = f_{s_{SBR}} = 2 \times f_{s_{AAC}}$.

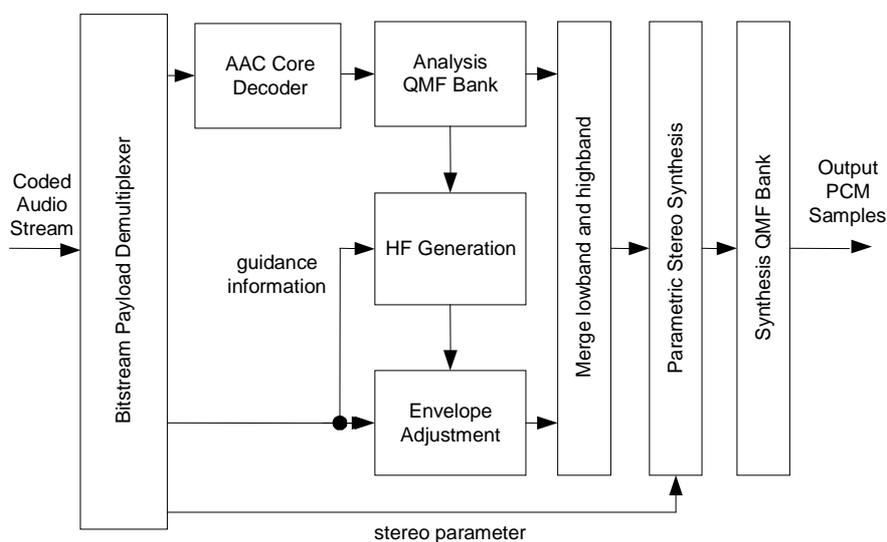


Figure A.11: MPEG-4 HE AAC v2 decoder

In January 2007 MPEG finalized the standardization of MPEG Surround (Spatial Audio Coding, (SAC)) [20]. MPEG Surround is capable of re-creating N channels based on $M < N$ transmitted channels and additional control data. In the preferred modes of operating the spatial audio coding system, the M channels can either be a single mono channel or a stereo channel pair. The control data represents a significant lower data rate than the data rate required for transmitting all N channels, making the coding very efficient while at the same time ensuring compatibility with M channel devices.

The MPEG Surround standard incorporates a number of tools enabling features that allow for broad application of the standard. A key feature is the ability to scale the spatial image quality gradually from very low spatial overhead towards transparency. Another key feature is that the decoder input can be made compatible to existing matrixed surround technologies.

As an example, for 5.1 multi-channel audio, the MPEG Surround encoder creates a stereo (or mono) downmix signal and spatial information describing the full 5.1 material in a highly efficient parameterized format. The spatial information is transmitted alongside the downmix. Figure A.12 shows the principle of MPEG Surround.

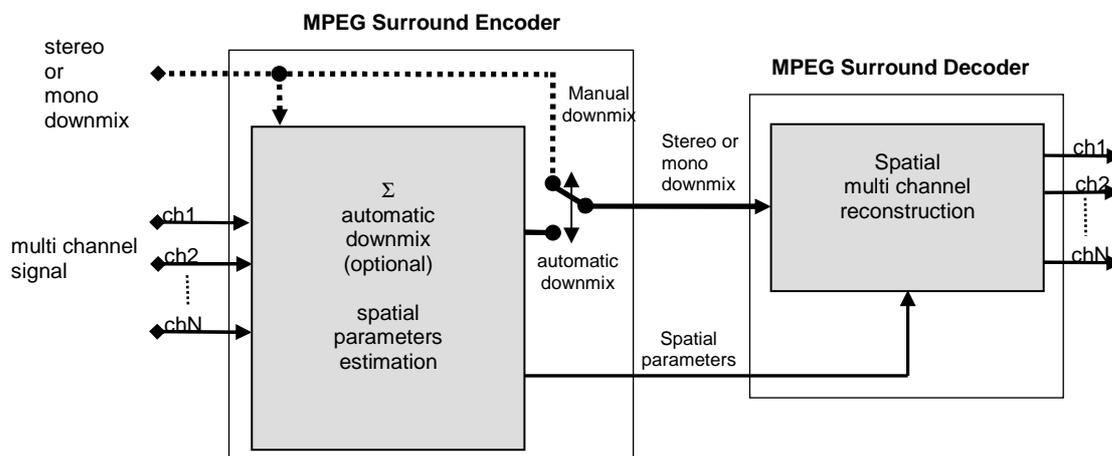


Figure A.12: Principles of MPEG Surround; the downmix is coded using MPEG-4 AAC, HE AAC or HE AAC v2

By using MPEG Surround, existing services can easily be upgraded to provide surround sound in a backward compatible fashion. While a stereo decoder in an existing legacy consumer device ignores the MPEG Surround data and plays back the stereo signal without any quality degradation, an MPEG Surround-enabled decoder will deliver high quality multi-channel audio.

The MPEG Surround decoder can operate in modes that render the multi-channel signal to multi-channel output, stereo output or operate in a two-channel headphone mode to produce a virtual surround output signal.

A.4.1.1 MPEG-4 AAC, HE AAC, HE AAC v2 and MPEG Surround Levels and Main Parameters for DVB

MPEG-4 provides a huge toolset for the coding of audio objects. In order to allow effective implementations of the standard, subsets of this toolset have been identified that can be used for specific applications. The function of these subsets, called "Profiles," is to limit the toolset a conforming decoder should implement. For each of these Profiles, one or more Levels have been specified, thus restricting the computational complexity.

The MPEG-4 HE AAC v2 Profile is a superset of the MPEG-4 AAC Profile. Besides the Audio Object Type (AOT) AAC LC (which is present in the AAC Profile), it includes the AOT SBR and the AOT PS. Levels are introduced within these Profiles in such a way that a decoder supporting the MPEG-4 HE AAC v2 Profile at a given level can decode an MPEG-4 AAC Profile and an HE AAC Profile stream at the same or lower level.

Table A.5: Levels within the MPEG-4 HE AAC v2 Profile

Level	Max. channels/object	Max. AAC sampling rate, SBR not present (kHz)	Max. AAC sampling rate, SBR present (kHz)	Max. SBR sampling rate, (kHz) (in/out)
1	NA	NA	NA	NA
2	2	48	24	24/48 (see note 1)
3	2	48	48 (see note 3)	48/48 (see note 2)
4	5	48	24/48 (see note 4)	48/48 (see note 2)
5	5	96	48	48/96

NOTE 1: A level 2 HE AAC v2 Profile decoder implements the baseline version of the parametric stereo tool. Higher level decoders are not being limited to the baseline version of the parametric stereo tool.

NOTE 2: For Level 3 and Level 4 decoders, it is mandatory to operate SBR in a downsampled mode if the sampling rate of the AAC core is higher than 24 kHz. Hence, if SBR operates on a 48 kHz AAC signal, the internal sampling rate of SBR will be 96 kHz, however, the output signal will be downsampled by SBR to 48 kHz.

NOTE 3: If Parametric Stereo data is present the maximum AAC sampling rate is 24 kHz, if Parametric stereo data is not present the maximum AAC sampling rate is 48 kHz.

NOTE 4: For one or two channels the maximum AAC sampling rate, with SBR present, is 48 kHz. For more than two channels the maximum AAC sampling rate, with SBR present, is 24 kHz.

For DVB, level 2 for mono and stereo as well as level 4 for multichannel audio signals are supported. The Low Frequency Enhancement channel of a 5.1 audio signal is included in the level 4 definition of the number of channels.

The MPEG Surround Baseline Profile defines six levels using the MPEG Surround AOT as given in table A.6.

Table A.6: Levels of the Baseline MPEG Surround Profile

Level	Tree configurations	Max. number output channels	Max. sampling rate [kHz]	Max. bandwidth residual coding [QMF bands]
1	515, 525, 727 (see note 1), (see note 4)	2.0	48	0 (see note 2)
2	515, 525, 727 (see note 4)	5.1	48	0 (see note 2)
3	515, 525, 727 (see note 4)	5.1	48	64 (see note 3)
4	515, 525, 727 (see note 4)	7.1	48	64 (see note 3)
5	515, 525, 757, 727 (see note 4)	7.1	48	64 (see note 3)
6	515, 525, 757, 727, plus arbitrary tree extension	32 incl. LFE	96	64 (see note 3)

NOTE 1: This level provides a 2-channel stereo output.

NOTE 2: Residual coding data, if present in the bitstream, is not utilized, hence the residual decoding tool is not required.

NOTE 3: A low power decoder utilizes only residual coding data for the first 8 QMF bands, corresponding to approximately 2.7 kHz bandwidth.

NOTE 4: Arbitrary tree extension data, if present, is not utilized.

For DVB applications, levels 1, 3 or 4 (using mono and stereo core coding) as well as the level 5 (using multichannel core coding) are supported.

A.4.1.2 Methods for signalling SBR and/or PS data

When the SBR and/or PS tools are used, several methods are available to signal the presence of the SBR and/or PS data [3]. Within the context of DVB services over IP it is recommended that backward compatible explicit signalling is used. In this case the respective extension Audio Object Type is signalled at the end of the AudioSpecificConfig().

A.4.1.3 Methods for signalling MPEG Surround data

When MPEG Surround is used, and the MPEG-4 AAC / HE AAC / HE AAC v2 and MPEG Surround payloads are transported together within a single RTP stream (see clause A.2.5), the MPEG Surround information is implicitly signalled by the presence of MPS fill elements in the MPEG-4 AAC/HE AAC/HE AAC v2 payload. It is recommended that signalling of MPEG Surround information is also enabled at the RTP level by conveying the MIME parameter MPS-profile-level-id as described in RFC 5691 [21].

When the MPEG Surround payload is transported in a separate RTP stream from the core payload (see clause A.2.5), the MPEG Surround information is signalled by conveying the MIME parameter profile-level-id for the MPEG Surround RTP stream as described in RFC 5691 [21].

A.4.2 Extended AMR-WB (AMR-WB+)

The AMR-WB+ audio codec can encode mono and stereo, up to 48 kbit/s for stereo. It supports also downmixing to mono at a decoder. The AMR-WB+ codec has been fully specified in TS 126 290 [8] including error concealment. The source code for both encoder and decoder has been fully specified in TS 126 304 [i.7] and in TS 126 273 [i.6]. The transport has been specified in RFC 4352 [9].

Figure A.13 presents the AMR-WB+ encoder structure. The input signal is separated in two bands. The first band is the Low-Frequency (LF) signal, which is critically sampled at $F_s/2$. The second band is the high-frequency (HF) signal, which is also downsampled to obtain a critically sampled signal. The LF and HF signals are then encoded using two different approaches: the LF signal is encoded and decoded using the "core" encoder/decoder, based on switched ACELP and Transform Coded eXcitation (TCX). In ACELP mode, the standard AMR-WB codec is used. The HF signal is encoded with relatively few bits using a BandWidth Extension (BWE) method.

The parameters transmitted from encoder to decoder are the mode selection bits, the LF parameters and the HF parameters. The codec operates in superframes of 1 024-samples. The parameters for each of them are decomposed into four packets of identical size.

When the input signal is stereo, the left and right channels are combined into mono signal for ACELP/TCX encoding, whereas the stereo encoding receives both input channels.

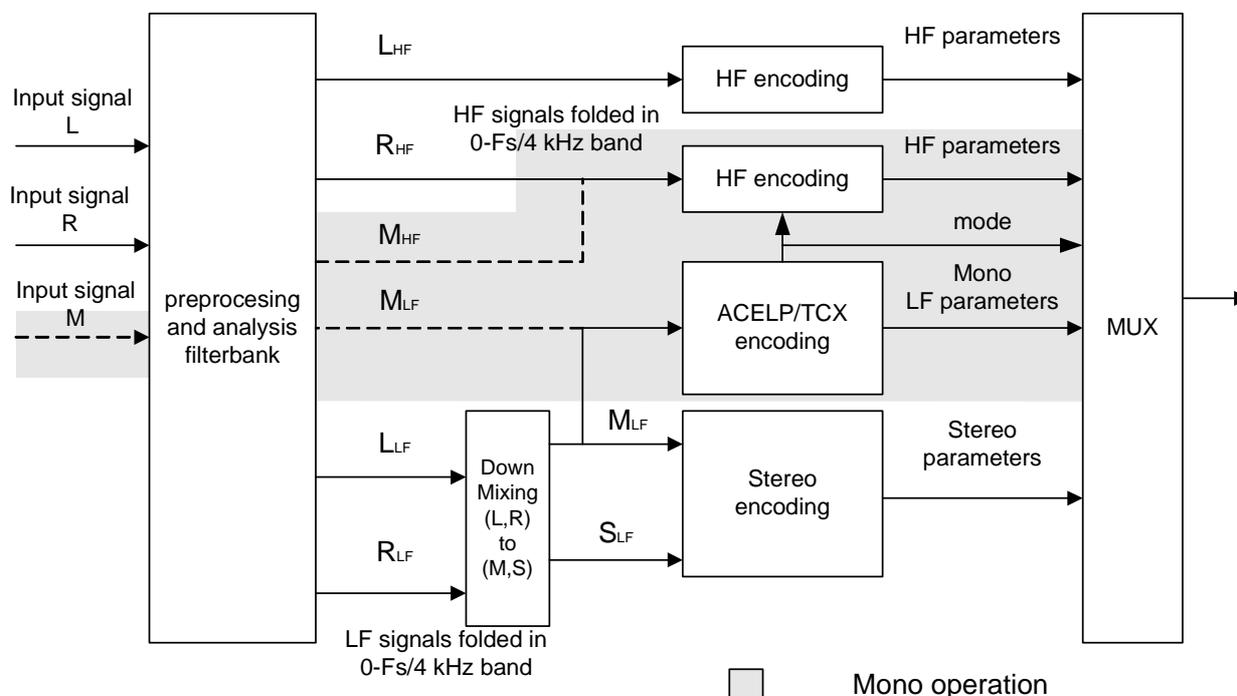


Figure A.13: High-level structure of AMR-WB+ encoder

Figure A.14 presents the AMR-WB+ decoder structure. The LF and HF bands are decoded separately after which they are combined in a synthesis filterbank. If the output is restricted to mono only, the stereo parameters are omitted and the decoder operates in mono mode.

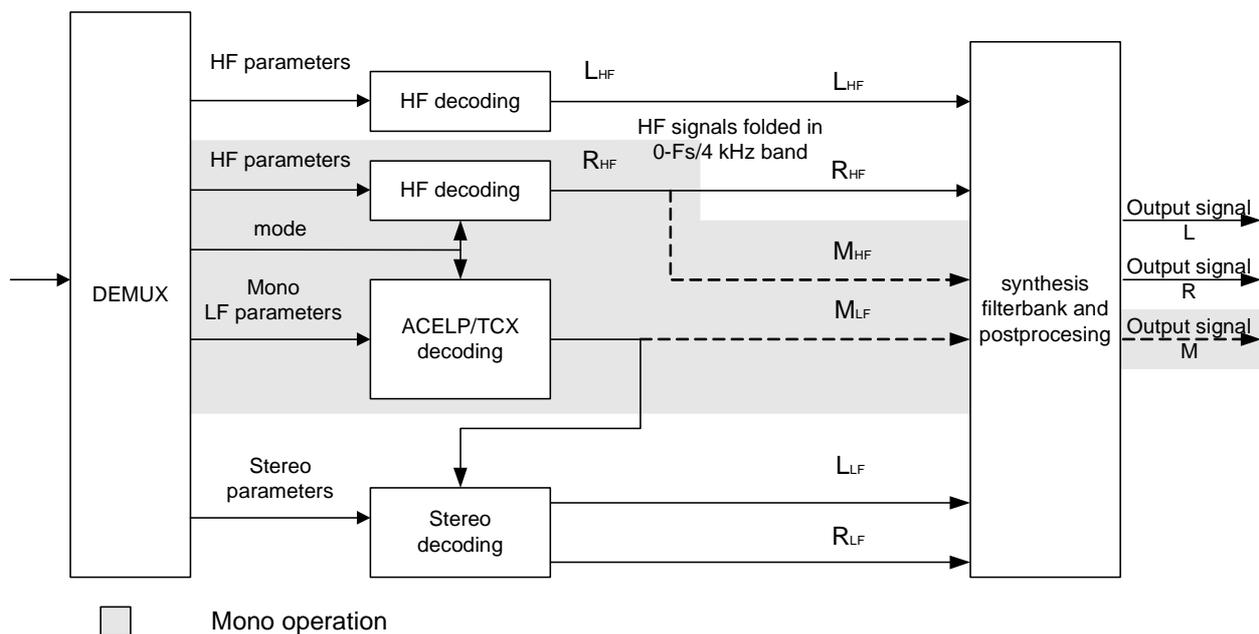


Figure A.14: High-level structure of AMR-WB+ decoder

A.4.2.1 Main AMR-WB+ parameters for DVB

The AMR-WB+ codec has been designed for mobile applications. Therefore no additional restrictions are required for IPDC over DVB-H or other DVB applications.

A.4.3 AC-3

The AC-3 digital compression algorithm can encode from 1 to 5.1 channels of source audio from a PCM representation into a serial bit stream at data rates ranging from 32 kbit/s to 640 kbit/s. The 0,1 channel refers to a fractional bandwidth channel intended to convey only low frequency signals.

The AC-3 algorithm achieves high coding gain by coarsely quantizing a frequency domain representation of the audio signal. A block diagram of this process is shown in figure A.15. The first step in the encoding process is to transform the representation of audio from a sequence of PCM time samples into a sequence of blocks of frequency coefficients. This is done in the analysis filter bank. Overlapping blocks of 512 time samples are multiplied by a time window and transformed into the frequency domain. Due to the overlapping blocks, each PCM input sample is represented in two sequential transformed blocks. The frequency domain representation may then be decimated by a factor of two so that each block contains 256 frequency coefficients. The individual frequency coefficients are represented in binary exponential notation as a binary exponent and a mantissa. The set of exponents is encoded into a coarse representation of the signal spectrum which is referred to as the spectral envelope. This spectral envelope is used by the core bit allocation routine which determines how many bits to use to encode each individual mantissa. The spectral envelope and the coarsely quantized mantissas for 6 audio blocks (1 536 audio samples per channel) are formatted into an AC-3 frame. The AC-3 bit stream is a sequence of AC-3 frames.

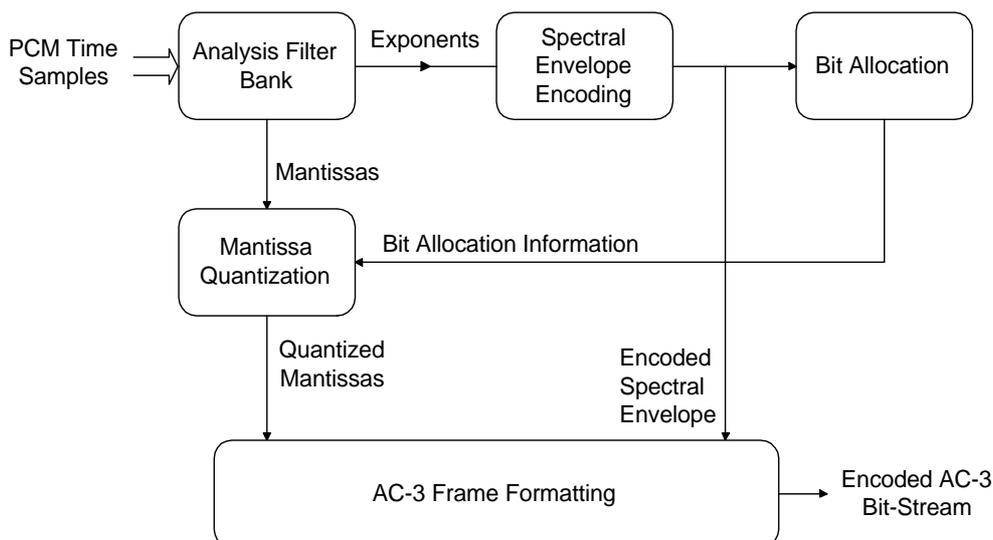


Figure A.15: The AC-3 encoder

The actual AC-3 encoder is more complex than indicated in figure A.15. The following functions not shown above are also included:

- 1) A frame header is attached which contains information (bit rate, sample rate, number of encoded channels, etc.) required to synchronize to and decode the encoded bit stream.
- 2) Error detection codes are inserted in order to allow the decoder to verify that a received frame of data is error free.
- 3) The analysis filterbank spectral resolution may be dynamically altered so as to better match the time/frequency characteristic of each audio block.
- 4) The spectral envelope may be encoded with variable time/frequency resolution.
- 5) A more complex bit allocation may be performed, and parameters of the core bit allocation routine modified so as to produce a more optimum bit allocation.
- 6) The channels may be coupled together at high frequencies in order to achieve higher coding gain for operation at lower bit rates.
- 7) In the two-channel mode, a rematrixing process may be selectively performed in order to provide additional coding gain, and to allow improved results to be obtained in the event that the two-channel signal is decoded with a matrix surround decoder.

The decoding process is basically the inverse of the encoding process. The decoder, shown in figure A.16, should synchronize to the encoded bit stream, check for errors, and de-format the various types of data such as the encoded spectral envelope and the quantized mantissas. The bit allocation routine is run and the results used to unpack and de-quantize the mantissas. The spectral envelope is decoded to produce the exponents. The exponents and mantissas are transformed back into the time domain to produce the decoded PCM time samples.

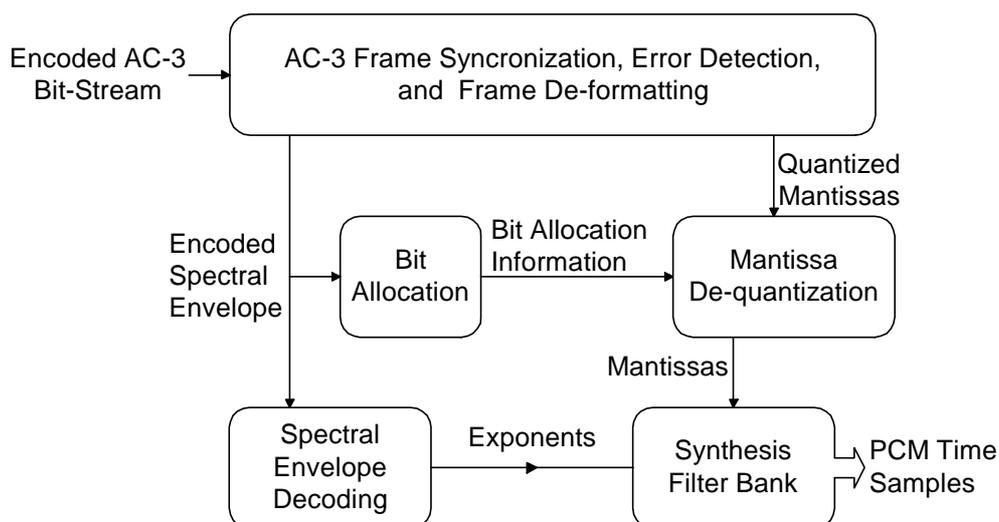


Figure A.16: The AC-3 decoder

The actual AC-3 decoder is more complex than indicated in figure A.16. The following decoder operations not shown above are included:

- 1) Error concealment or muting may be applied in case a data error is detected.
- 2) Channels which have had their high-frequency content coupled together should be de-coupled.
- 3) Dematrixing should be applied (in the 2-channel mode) whenever the channels have been rematrixed.
- 4) The synthesis filterbank resolution should be dynamically altered in the same manner as the encoder analysis filter bank had been during the encoding process.

A.4.4 Enhanced AC-3

Enhanced AC-3 is an evolution of the AC-3 coding system. The addition of a number of low data rate coding tools enables use of Enhanced AC-3 at a lower bit rate than AC-3 for high quality, and use at much lower bit rates than AC-3 for medium quality. A greatly expanded and more flexible bitstream syntax enables a number of advanced features, including expanded data rate flexibility and support for Variable BitRate (VBR) coding. A bitstream structure based on substreams allows delivery of programs containing more than 5.1 channels of audio to support next-generation content formats, supporting channel configuration standards developed for D-Cinema and support for multiple audio programs carried within a single bit-stream, suitable for deployment of services such as Hearing Impaired/Visual Impaired. To control the combination of audio programs carried in separate substreams or bitstreams, Enhanced AC-3 includes comprehensive mixing metadata, enabling a content creator to control the mixing of two audio streams in an IP-IRD. To ensure compatibility of the most complex bitstream configuration with even the simplest Enhanced AC-3 decoder, the bitstream structure is hierarchical - decoders will accept any Enhanced AC-3 bitstream and will extract only the portions that are supported by that decoder without requiring additional processing. To address the need to connect IP-IRDs that include Enhanced AC-3 to the millions of home theatre systems that feature legacy AC-3 decoders via S/PDIF, it is possible to perform a modest complexity conversion of an Enhanced AC-3 bitstream to an AC-3 stream for S/PDIF compatibility.

Enhanced AC-3 includes the following coding tools that improve coding efficiency when compared to AC-3:

- Spectral Extension: recreates a signal's high frequency amplitude spectrum from side data transmitted in the bit stream. This tool offers improvements in reproduction of high frequency signal content at low data rates.
- Transient Pre-Noise Processing: synthesizes a clause of PCM data just prior to a transient. This feature improves low data rate performance for transient signals.
- Adaptive Hybrid Transform Processing: improves coding efficiency and quality by increasing the length of the transform. This feature improves low data rate performance for signals with primarily tonal content.

- Enhanced Coupling: improves on traditional coupling techniques by allowing the technique to be used at lower frequencies than conventional coupling, thus increasing coder efficiency.

A.5 The DVB IP datacast application

Annex B of the present document defines application-specific constraints on the use of the toolbox for the particular case of DVB IP Datacast applications. These applications are mainly focused on handheld devices with severe limitations on computational resources and battery. Hence, the allowed values of parameters such as the picture size are limited. In addition, the desire to harmonize the such applications with 3GPP specifications has led to a strong recommendation that each IP-IRD that is to be used for DVB IP Datacast applications is capable of decoding video bitstreams conforming to H.264/AVC [1].

A.6 Future work

In common with TS 101 154 [i.2] and TS 102 154 [i.3], the present document is a living document, subject to periodic revision. The intention is to develop revisions in a largely backwards compatible manner, so that no changes to the mandatory functionality of a previously defined IP-IRD are made between one edition and the next.

One specific issue is the possibility of extending the video specification to include even higher resolution content, such as 1 080 p at 50 Hz and 60 Hz frame rate, also for VC-1. If this is done, it is likely that VC-1 Advanced Profile at Level L4 would be chosen.

Annex B (normative): TS 102 005 usage in DVB IP datacast

B.1 Scope

This annex describes the usage of TS 102 005 in TS 102 468 [i.11], specifying additional constraints that apply to the specifications in clauses 1 to 6 of the present document.

B.2 Introduction

This annex contains the technical specifications that address the requirements for DVB IP Datacast applications. These are mainly focused on handheld devices with severe limitations on computational resources and battery. Hence, the allowed values of parameters such as the picture size are limited. Nevertheless, IP-IRDs permitting larger spatial video resolutions may also be used in DVB IP Datacast applications.

Conversely, it is not mandatory for IP datacast services which do not conform to TS 102 468 [i.11] to follow the additional constraints specified in this annex.

B.3 Systems layer

This clause specifies constraints on the RTP payload formats, 3GPP file format, and "MP4" file format that are to be used for DVB IP Datacast applications.

B.3.1 Transport over IP networks/RTP packetization formats

The specifications in clause 4.1, including its constituent clauses shall apply subject to the following further constraint on clause 4.1.1 for the RTP Packetization of H.264/AVC for DVB IP Datacast applications and on clause 4.1.4 for the RTP Packetization of MPEG-4 HE AAC v2 and HE AAC v2 in combination with MPEG Surround for DVB IP Datacast applications. Further constraints on RTP packetization of H.264/AVC

Encoding: *The Single NAL Unit Mode or the Non-Interleaved Mode of RFC 3984 [6] shall be used for the packetization of H.264/AVC data into RTP.*

Decoding: *Each IP-IRD supporting H.264/AVC shall be able to receive Single NAL Unit Mode and Non-Interleaved Mode RTP packets with H.264/AVC data as defined in RFC 3984 [6].*

B.3.1.1 Further constraints on RTP packetization of MPEG-4 HE AAC v2

Encoding: *The interleaving mode of RFC 3640 [5] shall not be used for the packetization of MPEG-4 HE AAC v2 data into RTP.*

Decoding: *An IP IRD supporting MPEG-4 HE AAC v2 shall be able to decode non-interleaved access units of RFC 3640 [5].*

B.3.1.2 Further constraints on RTP packetization of MPEG-4 HE AAC v2 in combination with MPEG Surround

Encoding: *The interleaving mode of RFC 3640 [5] shall not be used for the packetization of MPEG-4 HE AAC v2 data in combination with MPEG Surround data into RTP. The MPEG Surround data shall be embedded within the core MPEG-4 AAC, HE AAC or HE AAC v2 audio bit-stream, and this bit-stream shall be transported in a single RTP stream according to RFC 5691 [21].*

Decoding: *An IP IRD supporting MPEG-4 HE AAC v2 in combination with MPEG Surround shall be able to decode non-interleaved access units of RFC 3640 [5].*

B.3.2 File storage for download services

The specifications in clause 4.2 including its constituent clauses, shall apply.

B.4 Video

This clause specifies constraints on the video encoding, decoding and rendering for DVB IP Datacast applications.

It is strongly recommended that each IP-IRD that is to be used for DVB IP Datacast applications is capable of decoding video bitstreams conforming to H.264/AVC as specified in [1]. IP-IRDs that are used for DVB IP Datacast applications may be capable of decoding video bitstreams conforming to VC-1 as specified in [10]. *Encoded video bitstreams for DVB IP Datacast applications shall conform to either H.264/AVC, or VC-1.*

Clause B.4.1 defines the constraints for encoding and decoding with H.264/AVC and clause B.4.2 defines the constraints for encoding and decoding with VC-1.

B.4.1 H.264/AVC

B.4.1.1 Profile and level

Encoding: *For all Capability Bitstreams except Capability C Bitstreams the specifications in clause 5.1.1 shall apply.*

Capability C Bitstreams RTP packetized for real-time delivery shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 1.3 of the Baseline Profile with constraint_set1_flag being equal to 1.

Capability C Bitstreams encapsulated in 3GPP file format or in "MP4" file format shall conform to the restrictions described in ITU-T Recommendation H.264 [1]/ISO/IEC 14496-10 [2] for Level 2 of the Baseline Profile with constraint_set1_flag being equal to 1.

Decoding: *For all Capability IP-IRDs, the specifications in clause 5.1.1 shall apply in terms of the signalling of Profile and Level. However, it should be noted that IP-IRDs used for DVB IP Datacast applications are only required to be capable of decoding and rendering pictures from bitstreams that are subject to the additional constraints in terms of Sample Aspect Ratio, Frame Rate, Luminance Resolution and Picture Aspect Ratio that are specified in clauses B.4.1.2 and B.4.1.3.*

B.4.1.2 Sample aspect ratio

Encoding: *Square (1:1) sample aspect ratio shall be used.*

Decoding: *Each IP-IRD supporting H.264/AVC shall support decoding and rendering pictures with square (1:1) sample aspect ratio.*

B.4.1.3 Frame rate, luminance resolution, and picture aspect ratio

The specifications on frame rate in clause 5.1.3, picture aspect ratio in clause 5.1.4, and luminance resolution in clause 5.1.5 are further constrained as follows.

Encoding: *One of the picture sizes listed in table B.1 shall be used for the indicated capability class. The video frame rate shall not exceed the maximum frame rate specified for the picture size in the indicated capability class. The picture size shall not change during a streaming delivery session.*

Decoding: *Each IP-IRD supporting H.264/AVC shall support decoding and rendering video encoded using the picture sizes and video frame rates indicated in table B.1. Additionally, lower frame rates and variable frame rates shall be supported.*

Table B.1: H.264/AVC pictures sizes for DVB IP datacast applications

Capability class	Horizontal resolution (samples)	Vertical resolution (samples)	Maximum frame rate (f/s)	Display Aspect ratio
A	176	144	15	1.22:1
A	128	96	30	4:3 (1.33:1)
A	144	80	30	16:9 (1.80:1)
B	176	144	30	1.22:1
B	320	240	15	4:3 (1.33:1)
B	320	176	15	16:9 (1.82:1)
C	320	240	30	4:3 (1.33:1)
C	320	176	30	16:9 (1.82:1)
C	400	224	30	16:9 (1.79:1)

B.4.1.4 Chromaticity

The specifications in clause 5.1.6 shall apply.

B.4.1.5 Chrominance format

The specifications in clause 5.1.7 shall apply.

B.4.1.6 Random access points

Encoding: *A Random Access Point shall be an IDR picture. Unless the sequence parameter set and picture parameter set are provided outside the elementary stream, the random access point shall include exactly one SPS (that is active), and the PPS that is required for decoding the associated picture.*

B.4.1.7 Output latency

Encoding: *Each H.264/AVC sequence parameter set shall contain a vui_parameters syntax structure including the num_reorder_frames syntax element (indicating maximum number of frames that precede any frame in the coded video sequence in decoding order and follow it in output order during the streaming delivery session).*

NOTE: For fixed frame applications the num_reorder_frames can be used to compute the maximum decoding to output latency in the sequence.

B.4.1.8 Active Format Description

The specifications in clause 5.1.10 shall apply.

B.4.2 VC-1

B.4.2.1 Profile and level

The specifications in clause 5.3.1 shall apply in terms of the signalling of Profile and Level. However, it should be noted that IP-IRDs used for DVB IP Datacast applications are only required to be capable of decoding and rendering pictures from bitstreams that are subject to the additional constraints in terms of bit rate, sample aspect ratio, frame rate, luminance resolution and picture aspect ratio that are specified in clauses B.4.2.2, B.4.2.3 and B.4.2.4.

B.4.2.2 Bit rate

The specifications in clause 5.3.1 are constrained as follows:

Encoding: *The maximum bit rate of a Capability C Bitstream shall not exceed 768 kbit/s.*

Decoding: *Each IP-IRD supporting VC-1 shall support any bit rate allowed by the indicated VC-1 Profile and Level, subject to a maximum of 768 kbit/s for a Capability C Bitstream.*

B.4.2.3 Sample aspect ratio

Encoding: *Square (1:1) sample aspect ratio shall be used.*

Decoding: *Each IP-IRD supporting VC-1 shall support decoding and rendering pictures with square (1:1) sample aspect ratio.*

B.4.2.4 Frame rate, luminance resolution and picture aspect ratio

The specifications on frame rate in clause 5.3.2, picture aspect ratio in clause 5.3.3, and luminance resolution in clause 5.3.4 are further constrained as follows:

Encoding: *One of the picture sizes listed in table B.2 shall be used for the indicated capability class. The video frame rate shall not exceed the maximum frame rate specified for the picture size in the indicated capability class. The picture size shall not change during a streaming delivery session.*

Decoding: *Each IP-IRD supporting VC-1 shall support decoding and rendering video encoded using the picture sizes and video frame rates indicated in table B.2. Additionally, lower frame rates and variable frame rates shall be supported.*

Table B.2: VC-1 Pictures sizes for DVB IP Datacast applications

Capability Class	Horizontal resolution (samples)	Vertical resolution (samples)	Maximum frame rate (f/s)	Display Aspect ratio
A	176	144	15	1.22:1
A	128	96	30	4:3 (1.33:1)
A	144	80	30	16:9 (1.80:1)
B	176	144	30	1.22:1
B	320	240	15	4:3 (1.33:1)
B	320	176	15	16:9 (1.82:1)
C	320	240	30	4:3 (1.33:1)
C	320	176	30	16:9 (1.82:1)
C	400	224	30	16:9 (1.79:1)

B.4.2.5 Chromaticity

The specifications in clause 5.3.5 shall apply.

B.4.2.6 Random Access Points

The specifications in clause 5.3.6 shall apply.

B.4.2.7 Active Format Description

The specifications in clause 5.3.7 shall apply.

B.5 Audio

This clause specifies constraints on the audio encoding and decoding for DVB IP Datacast applications.

Each IP-IRD that is to be used for DVB IP Datacast applications shall be capable of decoding audio bitstreams conforming to the MPEG-4 HE AAC v2 profile as specified in ISO/IEC 14496-3 [3]. In addition, IP-IRDs that are used for DVB IP Datacast applications may be capable of decoding audio bitstreams conforming to MPEG-4 HE AAC v2 as specified in ISO/IEC 14496-3 [3] in combination with MPEG Surround as specified in ISO/IEC 23003-1 [20] or audio bitstreams conforming to AMR-WB+ as specified in TS 126 290 [8]. Encoded audio bitstreams for DVB IP Datacast applications shall conform to either MPEG-4 AAC, HE AAC or HE AAC v2, the combination of MPEG-4 AAC, HE AAC or HE AAC v2 with MPEG Surround, or AMR-WB+.

Clause B.5.1 defines the constraints for encoding and decoding with MPEG-4 HE AAC v2, whilst clause B.5.2 defines the constraints for encoding and decoding with AMR-WB+.

B.5.1 MPEG-4 HE AAC v2 audio and MPEG-4 HE AAC v2 audio in combination with MPEG Surround

B.5.1.1 Audio mode

The specifications in clause 6.1.1 shall apply.

B.5.1.2 Profiles

The specifications in clause 6.1.2 shall apply.

B.5.1.3 Bit rate

The specifications in clause 6.1.3 are constrained as follows:

Encoding: *The maximum bit rate of the encoded audio shall not exceed 192 kbit/s for a stereo pair. For Capability A and B bitstreams containing video, the maximum audio bit rate shall not exceed 128 kbit/s for a stereo pair. The maximum bit rate of the encoded audio shall not exceed 320 kbit/s for multi-channel audio.*

Decoding: *Each IP-IRD supporting MPEG-4 HE AAC v2 shall support any bit rate allowed by the MPEG-4 HE AAC v2 Profile and selected Level, subject to a maximum of 192 kbit/s for a stereo pair.*

B.5.1.4 Sampling frequency

The specifications in clause 6.1.4 shall apply.

B.5.1.5 Dynamic range control

The specifications in clause 6.1.5 shall apply.

B.5.1.6 Matrix downmix

The specifications in clause 6.1.6 are constrained as follows:

Decoding: The support of matrix downmix as defined in MPEG-4 is optional for each IP-IRD.

B.5.2 AMR-WB+ audio

AMR-WB+ encoding and decoding of AMR-WB+ data in the IP Datacast IP-IRD shall follow the guidelines described in clauses 6.2.1 and 6.2.2.

Annex C (informative): Bibliography

- ISO/IEC 14496-14:2003: "Information Technology - Coding of Audio-Visual Objects - Part 14: MP4 file format".
- ISO/IEC 14496-15:2004: "Information technology - Coding of audio-visual objects - Part 15: Advanced Video Coding (AVC) file format".
- IETF Internet-Draft Submission draft-ietf-avt-rtp-svc-20: "RTP Payload Format for SVC Video".

NOTE: Available at <https://datatracker.ietf.org/drafts/draft-ietf-avt-rtp-svc/>.

History

Document history		
V1.1.1	March 2005	Publication
V1.2.1	April 2006	Publication
V1.3.1	July 2007	Publication
V1.4.1	March 2010	Publication