ETSI TS 102 005 V1.2.1 (2006-04)

Technical Specification

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-171

Keywords
3GPP, audio, broadcasting, digital, DVB, 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

Contents

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).

Founded in September 1993, the DVB Project is a market-led consortium of public and private sector organizations in the television industry. Its aim is to establish the framework for the introduction of MPEG-2 based digital television services. Now comprising over 200 organizations from more than 25 countries around the world, DVB fosters market-led systems, which meet the real needs, and economic circumstances, of the consumer electronics and the broadcast industry.

Introduction

The present document addresses the use of video and audio coding in DVB services delivered over IP protocols without involving an MPEG-2 Transport Stream. It specifies the use of H.264/AVC video as specified in ITU-T Recommendation H.264 and ISO/IEC 14496-10 [1], VC-1 video as specified in SMPTE 421M [19], HE AAC v2 audio as specified in ISO/IEC 14496-3 including Amendments 1 and 2 [2] and Extended AMR-WB (AMR-WB+) audio as specified in TS 126 290 [14].

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.

The present document does not address specifications for the use of video and audio coding in applications based on the MPEG-2 Transport Stream, regardless of whether or not the Transport Stream is delivered over IP. 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 [7], whilst that for contribution and primary distribution applications is given in TS 102 154 [8]. RFC 2250 [6] is used for the transport of an MPEG-2 TS in RTP packets over IP.

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, VC-1, HE AAC v2 and AMR-WB+ 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 [19] 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 VC-1 Simple Profile at level ML as specified in [19] or else both.

- **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 VC-1 Advanced Profile at level L0 as specified in [19] or else both.

- **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 VC-1 Advanced Profile at level L1 as specified in [19] or else both.

- **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 VC-1 Advanced Profile at level L3 as specified in [19] or both.

- IP-IRDs labelled with a particular capability Y are also capable of decoding H.264/AVC and/or VC-1 bitstreams 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, 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, or C.

- It is possible that an IP-IRD may support the decoding of H.264/AVC at Capability M and VC-1 at Capability N where M and N are not the same.

Audio:

- IP-IRDs are capable of decoding either bitstreams conforming to MPEG-4 Audio HE AAC v2 Profile or else bitstreams conforming to AMR-WB+ or else both.

- Sampling rates between 8 kHz and 48 kHz are supported by IP-IRDs.

- IP-IRDs support mono, parametric stereo (when MPEG-4 Audio HE AAC v2 Profile is used) and 2-channel stereo; support of multi-channel is optional.

An IP-IRD of one of the capability classes A to E above meets the minimum functionality, as specified in the present document, for decoding H.264/AVC or VC-1 video and for decoding HE AAC v2 or AMR-WB+ 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, VC-1, HE AAC v2, AMR-WB+ 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, VC-1, HE AAC v2 and AMR-WB+ for DVB conforming delivery in RTP packets over IP networks. The decoding of H.264/AVC, VC-1, HE AAC v2 and AMR-WB+ 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, VC-1, HE AAC v2 and AMR-WB+ for DVB IP Datacast services.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- 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.
- For a non-specific reference, the latest version applies.

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

[1] ITU-T Recommendation H.264: "Advanced video coding for generic audiovisual services " / ISO/IEC 14496-10 (2005): "Information Technology - Coding of audio-visual objects Part 10: Advanced Video Coding".

[2] ISO/IEC 14496-3: "Information technology - Generic coding of moving picture and associated audio information - Part 3: Audio" including ISO/IEC 14496-3 / AMD-1 (2001): "Bandwidth Extension" and ISO/IEC 14496-3 (2001) AMD-2 (2004): "Parametric Coding for High Quality Audio".

[3] IETF RFC 3550: "RTP, A Transport Protocol for Real Time Applications".

[4] IETF RFC 3640: "RTP payload for transport of generic MPEG-4 elementary streams".

[5] IETF RFC 3984: "RTP payload for transport of H.264".

[6] IETF RFC 2250: "RTP Payload Format for MPEG1/MPEG2 Video".

[7] 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 ".

[8] 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".

[9] EBU Recommendation R.68: "Alignment level in digital audio production equipment and in digital audio recorders".

[10] 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 5)".

[11] 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)".

[12] ISO/IEC 14496-14:2003, "Information Technology - Coding of Audio-Visual Objects - Part 14: MP4 file format".

[13] 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)".

[14] 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)".

[15] IETF RFC 4352: "RTP Payload Format for Extended Adaptive Multi-Rate Wideband (AMR-WB+) Audio Codec".

[16] 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)".

[17] 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)".

[18] ETSI TS 126 346: "Universal Mobile Telecommunications System (UMTS); Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs (3GPP TS 26.346 Release 6)".

[19] SMPTE 421M: "Proposed SMPTE Standard for Television: VC-1 Compressed Video Bitstream Format and Decoding Process".

[20] IETF RFC 4425: "RTP Payload Format for Video Codec 1 (VC-1)".

[21] ITU-R Recommendation BT.709: "Parameter values for the HDTV standards for production and international programme exchange".

[22] ETSI TS 102 468: "IP Datacast over DVB-H: Set of Specifications for Phase 1".

3 Definitions and abbreviations

3.1 Definitions

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

bitstream: coded representation of a video or audio signal

DVB IP Datacast application: application that complies with the DVB IP Datacast Umbrella Specification [22]

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

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
ACELP	lgebraic 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
BWE	BandWith Extension
CABAC	Context Adaptive Binary Arithmetic Coding
CIF	Common Interchange Format
DEMUX	DeMUltipleXer
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
HDTV	High Definition TeleVision
HE AAC	High-Efficiency Advanced Audio Coding
IP	Internet Protocol
IPDC	IP Data Casting
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 (ISO/IEC JTC 1/SC 29/WG 11)
MUX	MUltipleXer
NAL	Network Abstraction Layer
NTP	Network Time Protocol
PS	Parametric Stereo
PSS	Packet switched Streaming Service
QCIF	Quarter Common Interchange Format
QMF	Quadrature Mirror Filter
RTCP	RTP Control Protocol
RTP	Real-time Transport Protocol
RTSP	Real-Time Streaming Protocol
SBR	Spectral Band Replication
SBR	Spectral Band Replication
SR	Sender Report
TCP	Transmission Control Protocol
TCX	Transform Coded Excitation
UDP	User Datagram Protocol
VCEG	Video Coding Experts Group (ITU-T SG16 Q.6: Video Coding)
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 Packetizations Formats

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

While the general RTP specification is defined in RFC 3550 [3], 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, in clause 4.1.2 for VC-1, in clause 4.1.3 for HE AAC v2, and in clause 4.1.4 for AMR-WB+.

4.1.1 RTP Packetizations of H.264/AVC

For transport over IP, the H.264/AVC data is packetized in RTP packets using RFC 3984 [5].

Encoding: *RFC 3984 [5] shall be used for packetization into RTP.*

Decoding: *Each IP-IRD shall be able to receive RTP packets with H.264/AVC data as defined in RFC 3984 [5].*

4.1.2 RTP Packetization of VC-1

For transport over IP, the VC-1 data is packetized in RTP packets using RFC 4425 [20].

Encoding: *RFC 4425 [20] shall be used for packetization into RTP.*

Decoding: *Each IP-IRD shall be able to receive RTP packets with VC-1 data as defined in RFC 4425 [20].*

4.1.3 RTP Packetization of HE AAC v2

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

Encoding: *RFC 3640 [4] shall be used for packetization into RTP.*

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

4.1.4 RTP packetization of AMR-WB+

For transport over IP, the AMR-WB+ data is packetized in RTP packets using RFC 4352 [15].

Encoding: *RFC 4352 [15] shall be used for packetization in RTP.*

Decoding: Each IP-IRD shall support [15] to receive AMR-WB+ data contained in RTP packets.

4.2 File formats

4.2.1 MP4 file format

This clause describes usage of MP4 file format [12] in downloading services supporting this feature.

Encoding: *The MP4 file shall be created according to the MPEG-4 Part 14 [12] 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 1Mbytes. The size of "moof" boxes shall be equal to or less than 300 kbytes.

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: *Each IP-IRD shall support this feature and 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.*

4.2.2 3GPP file format

This clause describes usage of 3GPP file format [13] in downloading services supporting this feature.

Encoding: *The 3GP file shall conform to the Basic profile of the 3GPP Release 6 file format [13].*

Decoding: *Each IP-IRD shall support this feature and shall be able to parse Basic profile 3GP files according to the 3GPP Release 6 file format specification [13].*

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 VC-1 as specified in [19] or else both. 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 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

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 / ISO/IEC 14496-10 [1] 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 / ISO/IEC 14496-10 [1] 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/ ISO/IEC 14496-10 [1] 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 D H.264/AVC Bitstreams shall conform to the restrictions described in ITU-T Recommendation H.264 ISO/IEC 14496-10 [1] 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 / ISO/IEC 14496-10 [1] for Level 4 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 D IP-IRDs that support H.264/AVC shall be capable of decoding and rendering pictures using Capability A, B, C 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, D and E H.264/AVC Bitstreams. Support of profiles beyond High Profile is optional. Support of levels beyond Level 4 is optional.

NOTE: 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).

5.1.2 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: *Each IP-IRD 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.

5.1.3 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 [7] are used.

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

5.1.4 Luminance resolution

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

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

5.1.5 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 [21] is recommended.

Decoding: *Each IP-IRD 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.

5.1.6 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: *Each IP-IRD 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.

5.1.7 Random Access Points

Encoding: Where channel change times are important it is recommended that sequence and picture parameter sets are sent together with a random access point (e.g. an IDR picture) at least once every 500 ms. In applications where channel change time is an issue but coding efficiency is critical, it is recommended that random access points are encoded at least once every 2 s. For those applications where channel change time is not an issue, it is recommended that random access points are encoded at least once every 5 s. When changing 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.

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

NOTE 1: Increasing the frequency of sequence and picture parameter sets and IDR pictures will reduce channel hopping time but will reduce the efficiency of the video compression.

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

5.2 VC-1

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 [*19*]. The IP-IRD shall allow any legal structure as permitted by the specifications in [*19*] in the encoded video stream even if presently "reserved" or "unused".*

To allow full compliance to the specifications in [20] 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 Profile and level

Encoding: *Capability A VC-1 Bitstreams shall conform to the restrictions described in SMPTE 421M [*20*] for Simple Profile at level LL.*

*Capability B VC-1 Bitstreams shall conform to the restrictions described in SMPTE 421M [*20*] for Simple Profile at level ML.*

*Capability C VC-1 Bitstreams shall conform to the restrictions described in SMPTE 421M [*20*] for Advanced Profile at level L0.*

*Capability D VC-1 Bitstreams shall conform to the restrictions described in SMPTE 421M [*20*] for Advanced Profile at level L1.*

*Capability E VC-1 Bitstreams shall conform to the restrictions described in SMPTE 421M [*20*] 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.

NOTE: 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).

5.2.2 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: *Each IP-IRD 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.

5.2.3 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 [7] are used.

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

5.2.4 Luminance resolution

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

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

5.2.5 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 [21] 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: *Each IP-IRD 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.

5.2.6 Chrominance Format

Encoding: It is recommended to specify the chrominance format using the syntax element CHROMAFORMAT, if this syntax element is allowed by the applied VC-1 Profile.

Decoding: *Each IP-IRD shall be capable of decoding any allowed values of CHROMAFORMAT.* It is recommended that appropriate processing be included for the rendering of pictures.

5.2.7 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.

6 Audio

*Each IP-IRD shall be capable of decoding either audio bitstreams conforming to HE AAC v2 as specified in ISO/IEC 14496-3 including Amendments 1 and 2 [2] or else audio bitstreams conforming to Extended AMR-WB (AMR WB+) as specified in TS 126 290 [*14*] or else both.* Clause 6.1 describes the guidelines for encoding with MPEG-4 AAC, MPEG-4 HE AAC profile and MPEG HE AAC v2 profile 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. Annex B specifies application-specific constraints 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 [9].

6.1 MPEG-4 AAC profile, MPEG-4 HE AAC profile and MPEG HE AAC v2 profile

For HE AAC, the audio encoding shall conform to the requirements defined in ISO/IEC 14496-3 including Amendments 1 and 2 [2].

For HE AAC v2 the audio encoding shall conform to the requirements defined in ISO/IEC 14496-3 including ISO/IEC 14496-3 including Amendments 1 and 2 [2]

The IP-IRD design should be made under the assumption that any legal structure as permitted by ISO/IEC 14496-3 including Amendments 1 and 2 [2] may occur in the broadcast stream even if presently reserved or unused. *To allow full compliance to ISO/IEC 14496-3* [2] *and 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 1 and 2 [2].

6.1.1 Audio Mode

Encoding: *The audio shall be encoded in mono, parametric stereo or 2-channel-stereo according to the functionality defined in the HE AAC v2 Profile Level 2 or in multi-channel according to the functionality defined in the HE AAC v2 Profile Level 4, as specified in ISO/IEC 14496-3 including Amendments 1 and 2 [2].* A simulcast of a mono/parametric stereo/stereo signal together with the multi-channel signal is optional.

Decoding: *Each IP-IRD shall be capable of decoding in mono, parametric stereo or 2-channel-stereo of the functionality defined in the HE AAC v2 Profile Level 2, as specified in ISO/IEC 14496-3 including Amendments 1 and 2 [2]*. The support of multi-channel decoding in an IP-IRD is optional.

6.1.2 Profiles

Encoding: *The encoder shall use either the AAC Profile or the HE AAC Profile or the HE AAC v2 Profile.* Use of the HE AAC v2 Profile is recommended.

Decoding: *IP-IRDs shall be capable of decoding the HE AAC v2 Profile.*

6.1.3 Bit rate

Encoding: Audio may be encoded at any bit rate allowed by the applied profile and selected Level.

Decoding: *Each IP-IRD shall support any bit rate allowed by the HE AAC v2 Profile and selected Level.*

6.1.4 Sampling frequency

Encoding: Any of the audio sampling rates of the HE AAC v2 Profile Level 2 may be used for mono, parametric stereo and 2-channel stereo and of the HE AAC v2 Profile Level 4 for multichannel audio.

Decoding: *Each IP-IRD shall support each audio sampling rate permitted by the HE AAC v2 Profile Level 2 for mono, parametric stereo and 2-channel stereo and of the HE AAC v2 Profile Level 4 for multichannel audio.*

6.1.5 Dynamic range control

Encoding: The encoder may use the MPEG-4 AAC Dynamic Range Control (DRC) tool.

Decoding: *Each IP-IRD shall support the MPEG-4 AAC Dynamic Range Control (DRC) tool.*

6.1.6 Matrix downmix

Decoding: *Each IP-IRD shall support the matrix downmix as defined in MPEG-4.*

6.2 AMR-WB+

*AMR-WB+ encoding and decoding of AMR-WB+ data shall follow the guidelines described in this clause and are based on TS 126 290 [*14*].*

For AMR-WB+ the audio encoding shall conform to the requirements defined in TS 126 290 [14].

6.2.1 Audio mode

Encoding: *The audio shall be encoded in mono or stereo according to the functionality defined in the AMR-WB+ [*14*].*

Decoding: *Each IP-IRD supporting AMR-WB+ shall be capable of decoding in mono and stereo the functionality defined in the AMR-WB+, as specified in TS 126 290 [14].*

6.2.2 Sampling frequency

Encoding: Any of the audio sampling rates of the AMR-WB+ may be used for mono and stereo.

Decoding: *Each IP-IRD supporting AMR-WB+ shall support each audio sampling rate permitted by the AMR-WB+ for mono and stereo.*

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 [7] for distribution services and in TS 102 154 [8] 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 [3]. 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, as specified in ITU-T Recommendation H.264 [3] and in ISO/IEC 14496-10 [1], or else VC-1, as specified in SMPTE 421M [19]. Both 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. Both 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. For further information on the video codecs see clause A.3.

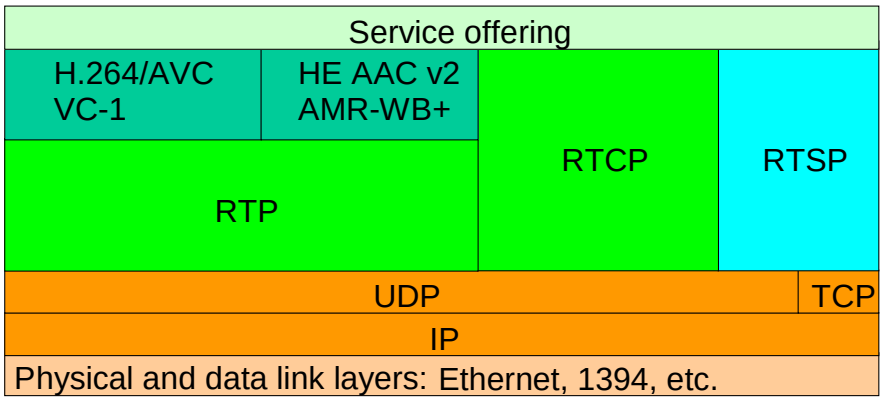
The advanced audio coding uses either MPEG-4 HE AAC v2 audio, as specified in ISO/IEC 14496-3 including Amendments 1 and 2 [2], or else Extended AMR-WB (AMR-WB+) audio as specified in TS 126 290 [14]. The MPEG-4 HE AAC v2 Profile is derived from the MPEG-2 Advanced Audio Coding (AAC), first published 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. Extended AMR-WB (AMR-WB+) has been optimized for use at low bit-rates with source material where speech predominates. 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 [22]. 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 [3]. RFC 3550 [3] 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/VC video data, RFC 3984 [5] 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 [5] specifies how to carry NAL units in RTP packets.

A.2.3 Transport of VC-1 video

To transport VC-1, RFC 4425 [20] is used. Each RTP packet contains an integer number of Access Units as defined in RFC 4425 [20], 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 [19]) carried in the SDP datagrams signalling the VC-1-based session.

A.2.4 Transport of HE AAC v2 audio

To transport HE AAC v2, RFC 3640 [4] is used. RFC 3640 [4] 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 [4]. The framing structure defined in RFC 3640 [4] 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.

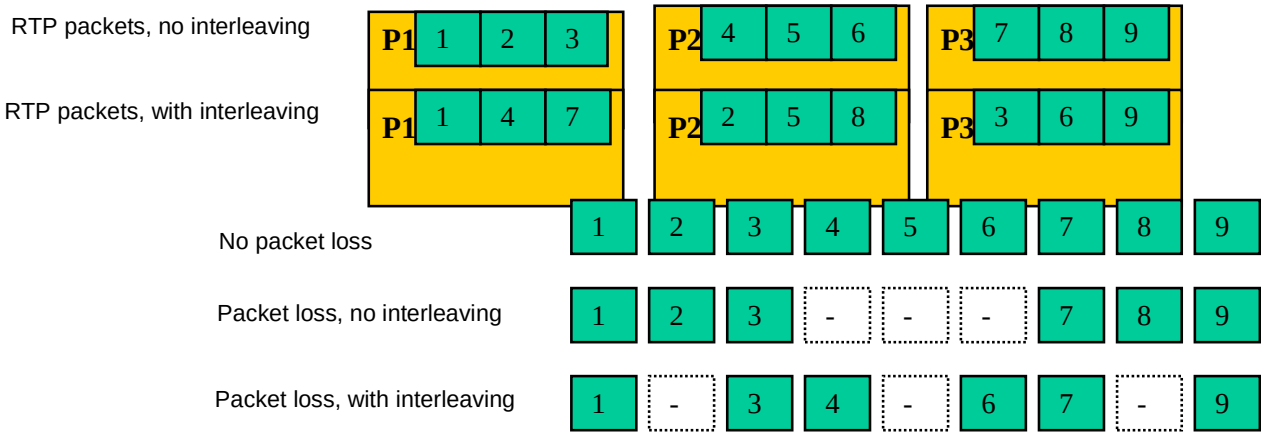


Figure A.2: Interleaving of AAC frames

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.2.5 Transport of AMR-WB+ audio

To transport AMR-WB+, RFC 4352 [15] is used. That payload is used also in both 3GPP Release TS 126 234 [11] and TS 126 346 [18] in which AMR-WB+ is the recommended codec with HE AAC v2.

The framing structure defined in [15] 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.

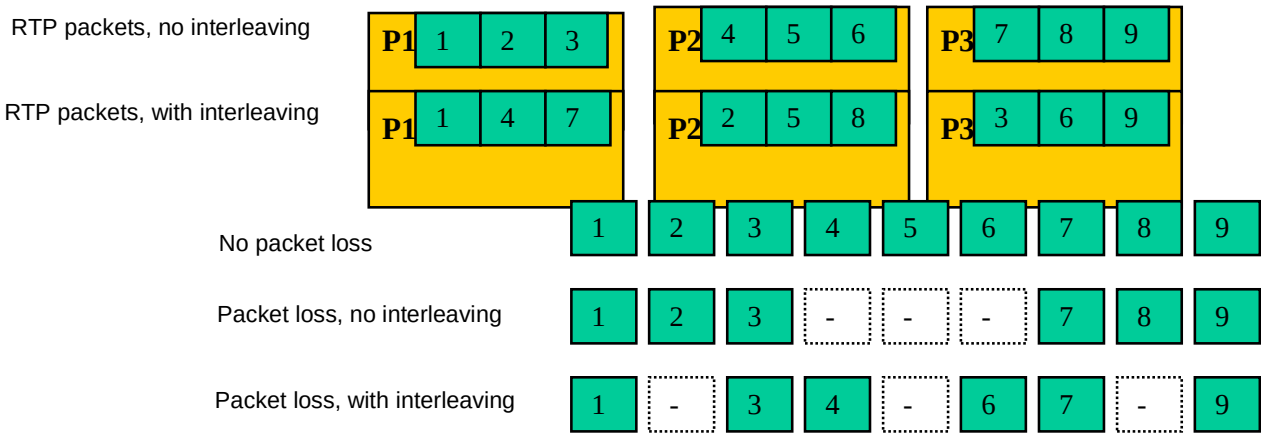


Figure A.3: Interleaving of AMR-WB+ frames

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.2.6 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 [3]. By means of figure A.3 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 [3], 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 [3]. In figure A.3, 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.

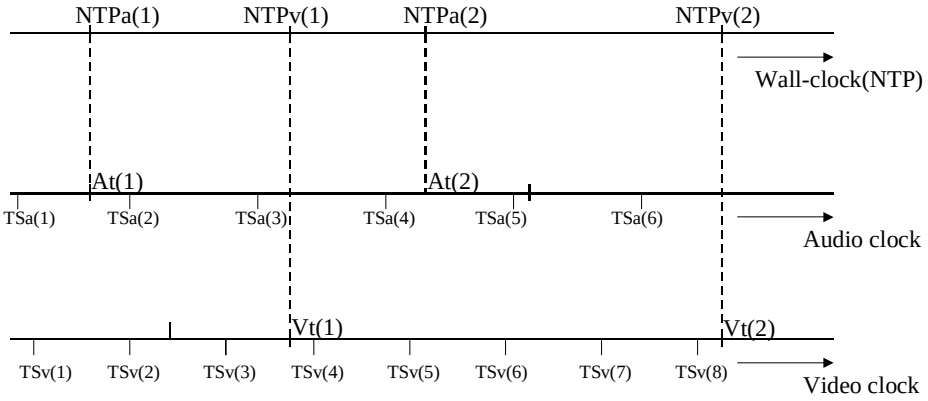


Figure A.4: RTP tools for synchronization

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 to send 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.2.7 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.2.8 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.2.9 Linking to applications

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

A.2.10 Capability exchange

By means of capability exchange protocols the sender and receiver can communicate whether the receiver has A, B, C, D or E IP-IRD capabilities for H.264/AVC 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+ decoding. 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.

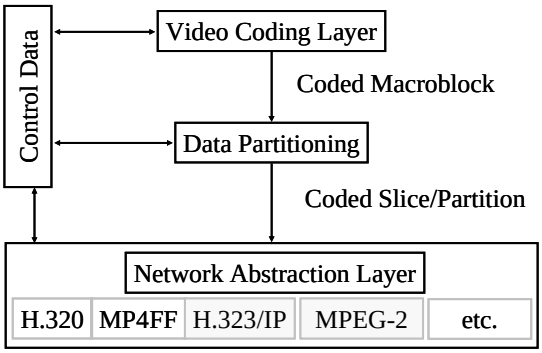


Figure A.5: Structure of H.264/AVC video encoder

A.3.1.2 Network Abstraction Layer

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 wireline 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

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.

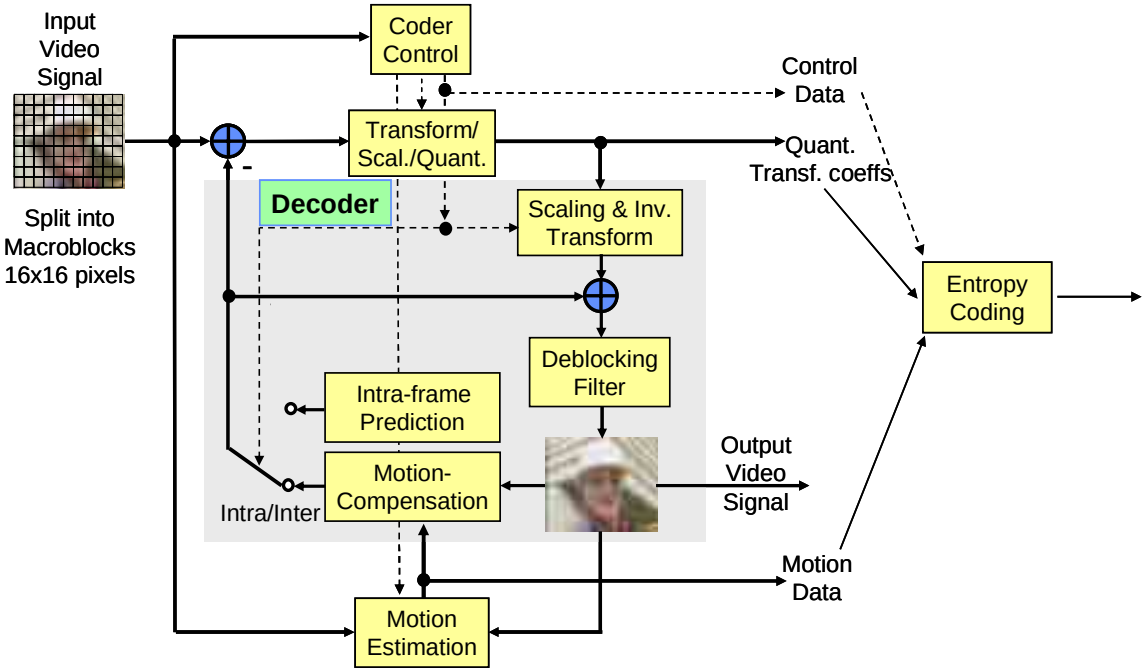


Figure A.6: Basic coding structure for H.264/AVC for a macroblock

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 must 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 [2] and by ISO/IEC as 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.

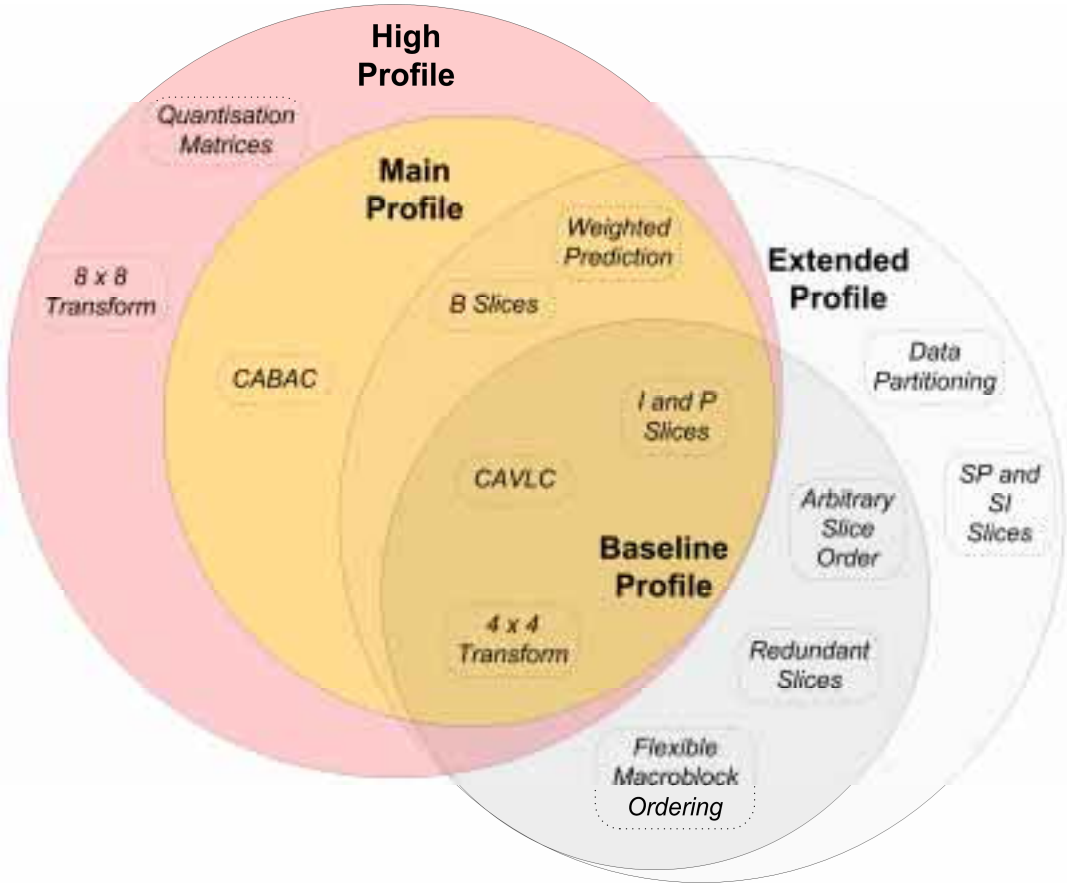


Figure A.7: Relationship between High Profile and the three Original Profiles

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 slices that approximates the primary slice is not used in the present document. The main reason is to achieve decodability by Main of 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 addional 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 E IRDs

Table A.1 summarizes the assignment of profiles and levels to the five 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 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
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 1080, 25/30Hz 720p HD = 1 280 x 720, 50/60Hz	20 000

The following should be noted.

IP-IRDs with Capability A, B, and C specify the Baseline profile with the additional constraint that constraint_set1_flag must 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 must be conforming to Main profile without any additional constraints. IP-IRDs with capability E must be conforming to Main 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 being an earlier letter than Y in the alphabet. 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, or C.

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, and C 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 the table). 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 [7].

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 VC-1 video

A.3.2.1 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 [19]) 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 must be capable of extracting this data from the SDP datagrams.

A.3.2.2 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 must 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.2 summarizes the coding tools that are included in each profile.

Table a.2

Feature	Simple Profile	Main Profile	Advanced Profile
Baseline intra frame compression	✓	✓	✓
Variable-sized transform	✓	✓	✓
16-bit transform	✓	✓	✓
Overlapped transform	✓	✓	✓
4 motion vector per macroblock	✓	✓	✓
¼ pixel luminance motion compensation	✓	✓	✓
¼ 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.3.2.3 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.3 summarizes the assignment of profiles and levels to the five IP-IRDs that are specified in the present document.

Table A.3

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 320x240, 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 = 1280 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 High Efficiency AAC v2 (HE AAC v2)

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.7 shows the different MPEG tools used in the MPEG-4 HE AAC v2 profile. A HE AAC v2 decoder will decode all three profile, AAC-LC, HE AAC and HE AAC v2.

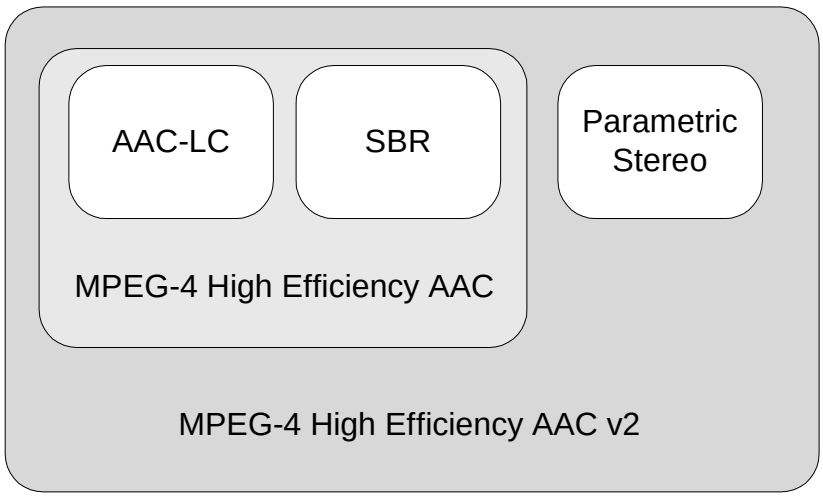


Figure A.8: MPEG Tools used in the HE AAC v2 Profile

Figure A.9 shows a block diagram of a 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.

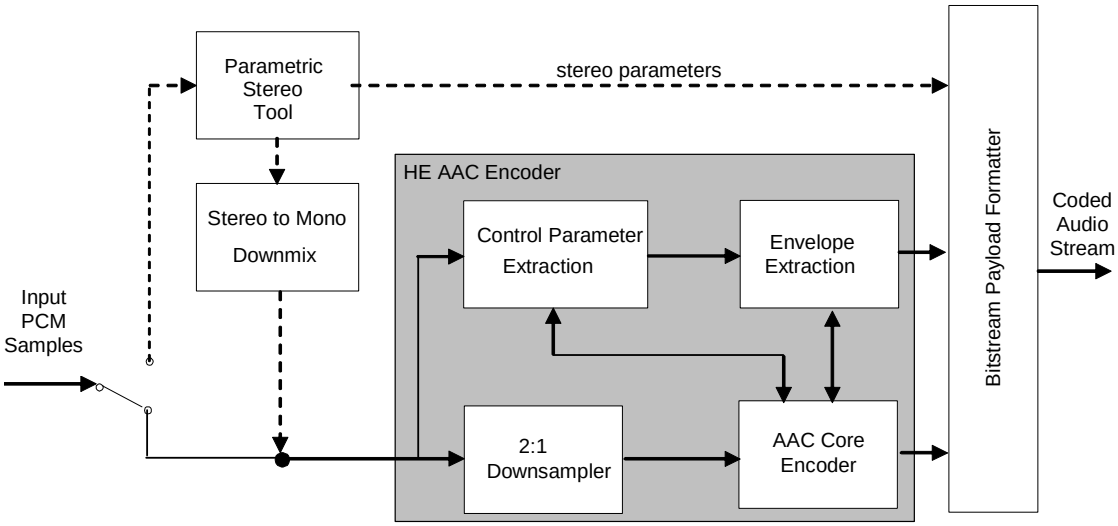


Figure A.9: HE AAC v2 Encoder

The HE AAC v2 decoder is depicted in figure A.9. 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 fs_{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. $fs_{out} = fs_{SBR} = 2 \times fs_{AAC}$.

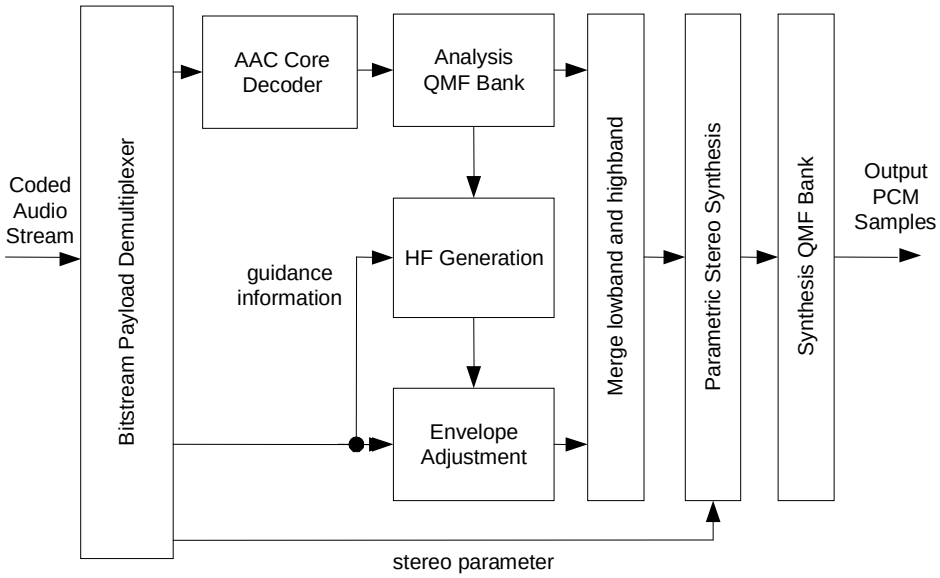


Figure A.10: HE AAC v2 Decoder

A.4.1.1 HE AAC v2 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 must implement. For each of these Profiles, one or more Levels have been specified, thus restricting the computational complexity.

The HE AAC v2 Profile is introduced as a superset of the 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 HE AAC v2 Profile at a given level can decode an AAC Profile and an HE AAC Profile stream at the same or lower level.

Table A.4:Levels within the 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 be 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 24kHz, if Parametric stereo data is not present the maximum AAC sampling rate is 48kHz.

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 the level 2 for mono and stereo as well as the level 4 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.

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

In case of usage of SBR and/or PS several ways how to signal the presence of SBR and/or PS data are possible [2]. Within the context of DVB services over IP it is recommended to use backward compatible explicit signalling. Here the respective extension Audio Object Type is signalled at the end of the AudioSpecificConfig().

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 [14] including error concealment. The source code for both encoder and decoder has been fully specified in TS 126 304 [17] TS 126 273 [16]. The transport has been specified in RFC 4352 [15].

Figure A.11 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 Fs/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.12 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.

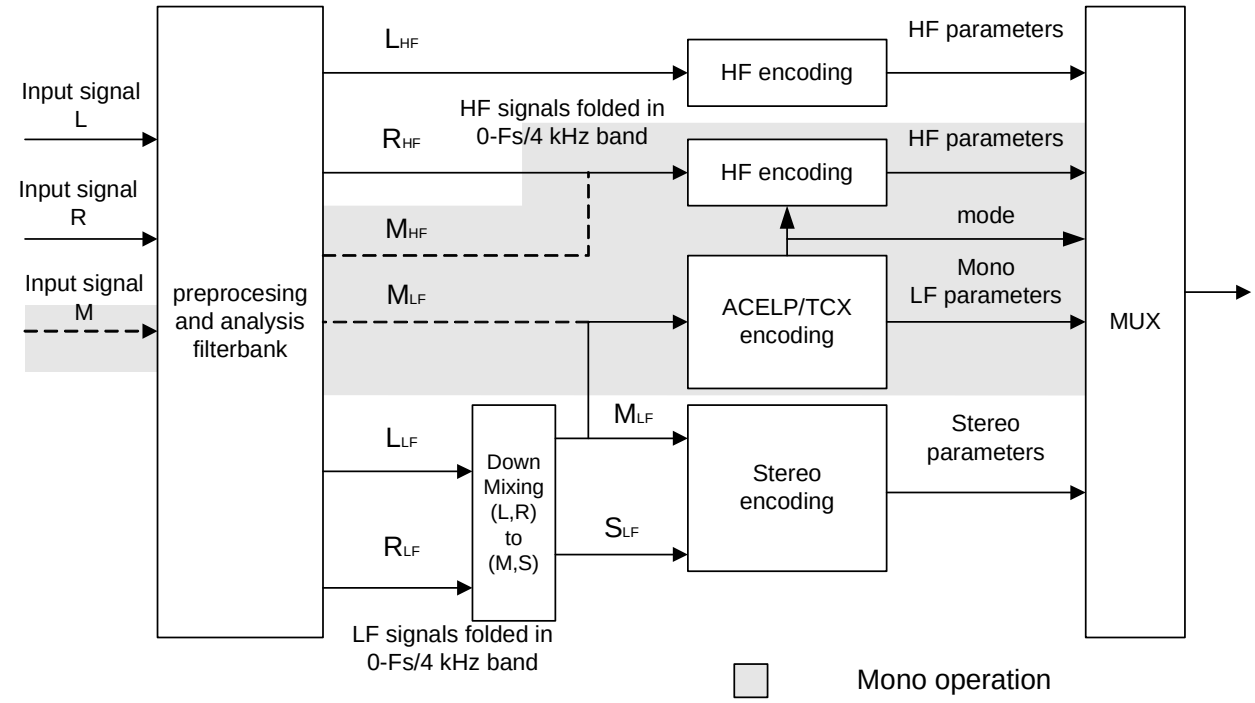


Figure A.11: High-level structure of AMR-WB+ encoder

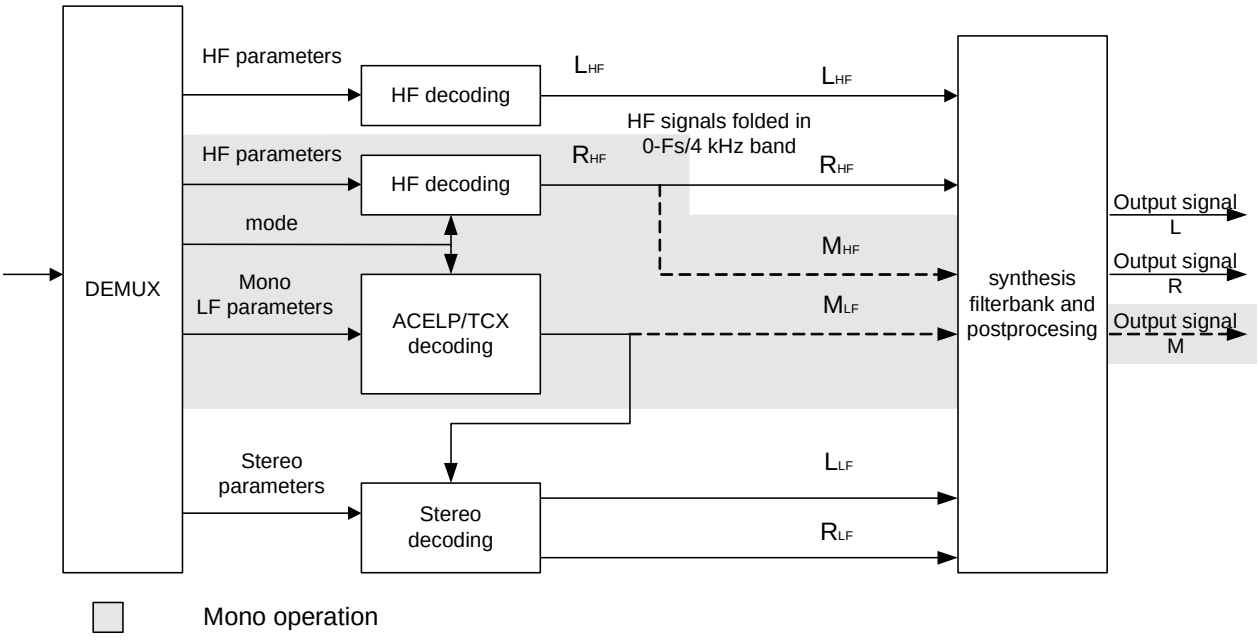


Figure A.12: High-level structure of AMR-WB+ decoder

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.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 [22]. 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 [7] and TS 102 154 [8], 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 1080 p at 50 Hz and 60 Hz frame rate. If this is done, it is likely that H.264/AVC High Profile at Level 4.2 and 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 [22] through specifying additional constraints that apply to the specifications in clauses 1-6 of the present document.

B.2 Introduction

This annex contains the technical specifications that address the requirements for DVB IP Datacast applications [22]. 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 [22] 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.

Encoding: *The Single NAL Unit Mode or the Non-Interleaved Mode of RFC 3984 [5] shall be used for the packetization of H.264/AVC data into RTP.*

Decoding: *Each IP-IRD 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 [5].*

B.3.2 File formats

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 [19]. *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, whilst 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 | ISO/IEC 14496-10 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 / ISO/IEC 14496-10 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 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 5.1.2, picture aspect ratio in clause 5.1.3, and luminance resolution in clause 5.1.4 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 shall support decoding and rendering video encoded using the picture sizes and video frame rates indicated in the 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 5.1.5 shall apply.

B.4.1.5 Chrominance Format

The specifications in clause 5.1.6 shall apply.

B.4.1.6 Random Access Points

The specifications in clause 5.1.7 shall apply.

B.4.2 VC-1

B.4.2.1 Profile and level

The specifications in clause 5.2.1 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 5.2.1 are constrained as follows:

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

Decoding: *Each IP-IRD shall support any bit rate allowed by the indicated VC-1 Profile and Level, subject to a maximum of 768kbit/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 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 5.2.2, picture aspect ratio in clause 5.2.3, and luminance resolution in clause 5.2.4 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 shall support decoding and rendering video encoded using the picture sizes and video frame rates indicated in the 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 5.2.5 shall apply.

B.4.2.6 Chrominance Format

The specifications in clause 5.2.6 shall apply.

B.4.2.7 Random Access Points

The specifications in clause 5.2.7 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 HE AAC v2 profile as specified in ISO/IEC 14496-3 including Amendments 1 and 2 [2]. In addition, IP-IRDs that are used for DVB IP Datacast applications may be capable of decoding audio bitstreams conforming to AMR-WB+ as specified in TS 126 290 [14]. *Encoded audio bitstreams for DVB IP Datacast applications shall conform to either HE AAC v2 or AMR-WB+*.

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

B.5.1 HE AAC v2

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 bitrate 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 shall support any bit rate allowed by the HE AAC v2 Profile and selected Level, subject to a maximum of 192kbit/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+

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.

History

Document history		
V1.1.1	March 2005	Publication
V1.2.1	April 2006	Publication