ETSI TR 126 948 V17.0.0 (2022-05)

iTeh STANDARD
PREVIEW
(standards.iteh.ai)

Digital cellular telecommunications system (Phase 2+) (GSM);
Universal Mobile Telecommunications System (UMTS);
LTE;
5G;
Study on video enhancements in 3GPP multimedia services
(3GPP TR 26.948 version 17.0.0 Release 17)

ETSI TR 126 948 V17.0.0 (2022-05)
https://standards.iteh.ai/catalog/standards/sist/2b5b0e2e-db5e-42df-a365-868998e213d8/etsi-tr-126-948-v17-0-0-2022-05

Reference
RTR/TSGS-0426948vh00

Keywords
5G,GSM,LTE,UMTS

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 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

If you find a security vulnerability in the present document, please report it through our Coordinated Vulnerability Disclosure Program:
https://www.etsi.org/standards/coordinated-vulnerability-disclosure

iTeh STANDARD PREVIEW (standards.iteh.ai)
ETSI TR 126 948 V17.0.0 (2022-05)
https://standards.iteh.ai/catalog/standards/sist/73ff2ef5-db5e-42d6-a365-868998e213d8/etsi-tr-126-948-v17-0-0-2022-05

Notice of disclaimer & limitation of liability

The information provided in the present deliverable is directed solely to professionals who have the appropriate degree of experience to understand and interpret its content in accordance with generally accepted engineering or other professional standard and applicable regulations.
No recommendation as to products and services or vendors is made or should be implied.
No representation or warranty is made that this deliverable is technically accurate or sufficient or conforms to any law and/or governmental rule and/or regulation and further, no representation or warranty is made of merchantability or fitness for any particular purpose or against infringement of intellectual property rights.
In no event shall ETSI be held liable for loss of profits or any other incidental or consequential damages.

Any software contained in this deliverable is provided "AS IS" with no warranties, express or implied, including but not limited to, the warranties of merchantability, fitness for a particular purpose and non-infringement of intellectual property rights and ETSI shall not be held liable in any event for any damages whatsoever (including, without limitation, damages for loss of profits, business interruption, loss of information, or any other pecuniary loss) arising out of or related to the use of or inability to use the software.

Copyright Notification

No part may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm except as authorized by written permission of ETSI.
The content of the PDF version shall not be modified without the written authorization of ETSI.
The copyright and the foregoing restriction extend to reproduction in all media.

© ETSI 2022.
All rights reserved.

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are 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 (https://ipr.etsi.org/).

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, 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.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

iTeh STANDARD PREVIEW (standards.iteh.ai)

Legal Notice

This Technical Report (TR) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

ETSI TR 126 948 V17.0.0 (2022-05) https://standards.iteh.ai/catalog/standards/sist/3f72d5d5-db5e-42df-a565-86b998e215d8/etsi-tr-126-948-v17-0-0-2022-05

The present document may refer to technical specifications or reports using their 3GPP identities. These shall be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between 3GPP and ETSI identities can be found under http://webapp.etsi.org/key/queryform.asp.

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Contents

Intellectual Property Rights 2
Legal Notice 2
Modal verbs terminology 2
Foreword 5
1 Scope 6
2 References 6
3 Definitions and abbreviations 7
3.1 Definitions 7
3.2 Abbreviations 7
4 Overview of video codecs specified for existing 3GPP multimedia services 8
5 Overview of SHVC 9
5.0 General 9
5.1 Basic SHVC architecture 9
5.2 Systems and transport interfaces of SHVC 10
5.2.1 Introduction 10
5.2.2 Parameter Set and Slice Segment Header Extensions 11
5.2.3 Layer and Scalability Identification 11
5.2.4 Layer sets 12
5.2.5 Profile, Tier, and Level (PTL) 12
5.2.6 RPS and Reference Picture List Construction 12
5.2.7 Random Access, Layer Switching, and Bitstream Splicing 13
5.2.8 Hybrid Codec Scalability and Multiview/3D Support 13
5.2.9 Hypothetical Reference Decoder (HRD) 14
5.2.10 SEI Messages 14
5.3 A comparison of SHVC and SVC 14
5.4 SHVC decoder and encoder complexity analyses 16
5.4.1 Introduction 16
5.4.2 HEVC simulcast encoder and decoder 16
5.4.3 SHVC encoder and decoder 18
5.4.4 Upsampling filter 19
5.4.5 Inter-layer texture prediction 20
5.4.6 Inter-layer motion prediction 21
5.4.7 Conclusion 22
6 Use cases 22
6.1 Multi-stream Multiparty Video Conferencing (MMVC) 22
6.1.1 The heterogeneous-device MMVC use case 22
6.1.2 The heterogeneous-bandwidth MMVC use case 24
6.1.3 Solutions for the MMVC use cases 24
6.1.3.0 General 24
6.1.3.1 HEVC simulcast 24
6.1.3.2 SHVC 25
6.1.4 Comparison of the solutions 26
6.1.4.1 Uplink bandwidth 26
6.1.4.2 Downlink bandwidth 26
6.1.4.3 Decoding complexity 27
6.1.4.4 Encoding complexity 27
6.2 MBMS 27
6.2.1 The differentiated-service MBMS use case 27
6.2.2 Solutions for the MBMS use case 28
6.2.2.1 HEVC simulcast 28
6.2.2.2 SHVC 29
6.2.3 Comparison of the solutions 29

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ETSI TR 126 948 V17.0.0 (2022-05)
https://standards.iteh.ai/catalog/standards/sist/73ff2ef5-db5c-42df-a365-868998e213d8/etsi-tr-126-948-v17-0-0-2022-05

6.2.3.1 Transmission bandwidth 29
6.2.3.2 Decoding complexity 31
6.2.3.3 Encoding complexity 32
6.3 3GP-DASH....32
6.3.1 The 3GP-DASH use case....32
6.3.2 Solutions for the 3GP-DASH use case 32
6.3.2.1 HEVC simulcast....32
6.3.2.2 SHVC....33
6.3.3 Comparison of the solutions 34
6.3.3.0 General....34
6.3.3.1 Outgoing transmission bandwidth....34
6.3.3.1.0 General 34
6.3.3.1.1 From origin server to caches 34
6.3.3.1.2 From origin server to UEs 34
6.3.3.2 Incoming transmission bandwidth 35
6.3.3.3 Decoding complexity 36
6.3.3.4 Encoding complexity 36

7 Test cases, conditions, and results....37
7.1 Test cases and conditions 37
7.1.1 General conditions 37
7.1.2 MMVC....38
7.1.3 MBMS 39
7.1.4 3GP-DASH 40
7.2 Test results....42
7.2.1 MMVC....42
7.2.1.1 General 42
7.2.1.2 Results for aligned IRAP pictures case 42
7.2.1.3 Results for IRAP non-aligned test case 43
7.2.1.4 First set of additional results 43
7.2.1.5 Second set of additional results 44
7.2.1.6 Additional analysis for comparing SHVC and HEVC simulcast 48
7.2.1.6.1 Introduction 48
7.2.1.6.2 Uplink vs downlink transmission cost....48
7.2.1.6.3 Case by case cost/benefit analysis of SHVC vs HEVC simulcast....48
7.2.2 MBMS 52
7.2.2.0 General 52
7.2.2.1 Results for aligned IRAP pictures case 53
7.2.2.2 Results for non-aligned IRAP pictures case 54
7.2.3 3GP-DASH 55
7.2.3.0 General 55
7.2.3.1 Results for aligned IRAP pictures case 56
7.2.3.2 Results for non-aligned IRAP pictures case 57

8 Conclusions 58
8.1 Introduction 58
8.2 MMVC and IMS based telepresence....59
8.3 MBMS....59
8.4 3GP-DASH....60

Annex A: Change history 61

History 62

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ETSI TR 126 948 V17.0.0 (2022-05)
https://standards.iteh.ai/catalog/standards/sist/73ff2ef5-4b5e-42df-a365-868998e213d8/etsi-tr-126-948-v17-0-0-2022-05

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

iTeh STANDARD
PREVIEW
(standards.iteh.ai)

ETSI TR 126 948 V17.0.0 (2022-05)
https://standards.iteh.ai/catalog/standards/sist/73ff2ef5-db5e-42df-a365-868998e213d8/etsi-tr-126-948-v17-0-0-2022-05

1 Scope

The present document reports the study on video enhancements in 3GPP multimedia services. It firstly provides an overview of the video codecs and their configurations specified for existing 3GPP multimedia services, namely 3GP-DASH (TS 26.247 [1]), PSS (TS 26.234 [2]), MBMS (TS 26.346 [3]), MTSI (TS 26.114 [4], including multi-stream multiparty video conferencing), MMS (TS 26.140 [5]), and IMS Messaging and Presence (TS 26.141 [6]). Then use cases on video enhancements for existing 3GPP multimedia services are discussed, including a discussion on potential codec solutions for each of the use cases. To enable drawing conclusions, simulation conditions and simulation results for comparisons of different codecs and their configurations are provided. Performance is evaluated in typical 3GPP service environments taking into account bandwidth, quality and complexity. Based on the performance results, conclusions are made in terms of recommendations for support of enhanced video capabilities for 3GPP multimedia 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, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ETSI TR 126 948 V17.0.0 (2022-05)
https://standards.iteh.ai/catalog/standards/sist/73ff2ef5-db3e-42df-a365-868998e213d8/etsi-tr-126-948-v17.0.0-2022-05

[1] 3GPP TS 26.247: "Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)".

[2] 3GPP TS 26.234: "Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs".

[3] 3GPP TS 26.346: "Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs".

[4] 3GPP TS 26.114: "IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction".

[5] 3GPP TS 26.140: "Multimedia Messaging Service (MMS); Media formats and codecs".

[6] 3GPP TS 26.141: "IP Multimedia System (IMS) Messaging and Presence; Media formats and codecs".

[7] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[8] ITU-T Recommendation H.263 (01/2005): "Video coding for low bit rate communication".

[9] ITU-T Recommendation H.264 (V9) (02/2014): "Advanced video coding for generic audiovisual services".

[10] ITU-T Recommendation H.265 (V3) (04/2015): "High efficiency video coding".

[11] J. Boyce, Y. Yan, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits Syst. Video Technol.*, August 2015, to be published.

[12] R. Sjöberg, Y. Chen, A. Fujibayashi, M. M. Hannuksela, J. Samuelsson, T. K. Tan, Y.-K. Wang, and S. Wenger, "Overview of HEVC High-Level Syntax and Reference Picture Management," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1858–1870, Dec. 2012.

[13] 3GPP TR 26.906: "Evaluation of HEVC for 3GPP Services".

[14] G. Tech, Y. Chen, K. Müller, J.-R. Ohm, A. Vetro, and Y.-K. Wang, "Overview of the Multiview and 3D Extensions of High Efficiency Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, August 2015, to be published.

[15] ITU-R Recommendation BT.709-6 (06/2015): "Parameter values for the HDTV standards for production and international programme exchange".

[16] ITU-R Recommendation BT.2020-1 (06/2014): "Parameter values for ultra-high definition television systems for production and international programme exchange".

[17] 3GPP TR 26.923: "Study on Media Handling Aspects of IMS-based Telepresence".

[18] 3GPP TS 26.948: "Video enhancements for 3GPP Multimedia Services".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in TR 21.905 [7] apply.

3.2 Abbreviations

iTeh STANDARD PREVIEW (standards.iteh.ai)

For the purposes of the present document, the abbreviations given in TR 21.905 [7] and the following apply.

ETSI TR 126 948 V17.0.0 (2022-05)
https://standards.iteh.ai/catalog/standards/sist/73ff2ef5-db58-42df-e365-868808e213d8/etsi-tr-126-948-v17-0-0-2022-05

3D-HEVC	Three-Dimension High Efficiency Video Coding
3GPP	3rd Generation Partnership Project
API	Application Program Interface
AU	Access Unit
AVC	Advanced Video Coding
BD	Bjontegaard Delta
BL	Base Layer
BLA	Broken Link Access
BM-SC	Broadcast-Multicast - Service Centre
BP	Bitstream Partition
CABAC	Context Adaptive Binary Arithmetic Coding
CPB	Coded Picture Buffer
CRA	Clean Random Access
CTU	Coding Tree Unit
DASH	Dynamic Adaptive Streaming over HTTP
DPB	Decoded Picture Buffer
EL	Enhancement Layer
GOP	Group of Pictures
HDTV	High-Definition TeleVision
HEVC	High Efficiency Video Coding
HLS	High-Level Syntax
HRD	Hypothetical Reference Decoder
IDR	Instantaneous Decoding Refresh
IMS	IP Multimedia Subsystem
INBLD	Independent Non-Base Layer Decoding
ILP	Inter-Layer Prediction
IRAP	Intra Random Access Point
MANE	Media Aware Network Element
MBMS	Multimedia Broadcast/Multicast Service
MBMS-GW	MBMS Gateway
MMVC	Multi-stream Multiparty Video Conferencing
MMS	Multimedia Messaging Service
MRFP	Multimedia Resource Function Processor
MSB	Most Significant Bits
MTSI	Multimedia Telephony Service for IMS

MTU	Maximum Transmission Unit
MV	Motion Vector
MV-HEVC	MultiView High Efficiency Video Coding
NAL	Network Abstraction Layer
OLS	Output Layer Set
POC	Picture Order Count
PPS	Picture Parameter Set
PSNR	Peak Signal Noise Ratio
PSS	Packet-switched Streaming Service
PTL	Profile, Tier, and Level
QP	Quantization Parameter
RAP	Random Access Point
RPL	Reference Picture List
RPS	Reference Picture Set
SAO	Sample Adaptive Offset
SEI	Supplemental Enhancement Information
SPS	Sequence Parameter Set
SVC	Scalable Video Coding
SHVC	Scalable High efficiency Video Coding
TMVP	Temporal Motion Vector Prediction
UE	User Equipment
UHDTV	Ultra High-Definition TeleVision
VPS	Video Parameter Set
VUI	Video Usability Information
WPP	Wavefront Parallel Processing

iTeh STANDARD PREVIEW (standards.iteh.ai)

4 Overview of video codecs specified for existing 3GPP multimedia services

The video support in 3GPP multimedia services in Release-12 is provided in Table 1.

ETSI TR 126 948 V17.0.0 (2022-05)
https://standards.iteh.ai/catalog/standards/sist/73ff2ef5-db5e-42df-a365-868998e213d8/etsi-tr-126-948-v17-0-0-2022-05

Table 1: Video support in 3GPP multimedia services in Release-12

	H.263 [8]	H.264/AVC [9]	HEVC/H.265 [10]
DASH and PSS	Profile 0 Level 45	Constrained Baseline Profile, Level 1.3 Progressive High Profile Level 3.1 Frame-packed stereoscopic 3D video (H.264 Constrained Baseline Profile Level 1.3 or Progressive High Profile Level 3.1) Multiview stereoscopic 3D video (H.264 Stereo High Profile Level 3.1), but not for RTP based transmission	Main Profile, Main Tier, Level 3.1
MBMS		Constrained Baseline Profile, Level 1.3 Progressive High Profile Level 3.1 Frame-packed stereoscopic 3D video (H.264 Constrained Baseline Profile Level 1.3 or Progressive High Profile Level 3.1)	Main Profile, Main Tier, Level 3.1
MTSI and IMS Messaging and Presence		H.264 Constrained Baseline Profile, Level 1.2 Constrained Baseline Profile, Level 3.1	Main Profile, Main Tier, Level 3.1
MMS	Profile 0 Level 45	Constrained Baseline Profile, Level 1.3 Progressive High Profile Level 3.1 Frame-packed stereoscopic 3D video (H.264 Constrained Baseline Profile Level 1.3 or Progressive High Profile Level 3.1)	Main Profile, Main Tier, Level 3.1

iTeh STANDARD PREVIEW (standards.iteh.ai)
ETSI TR 126 948 V17.0.0 (2022-05)
https://standards.iteh.ai/catalog/standards/sist/3ff2ef5-db5e-42df-a365-868998e21360/etsi-tr-126-948-v17-0-0-2022-05

5 Overview of SHVC

5.0 General

Scalable High efficiency Video Coding (SHVC) refers to the scalable extension of H.265/HEVC, specified in Annex H of the H.265/HEVC specification [10]. This clause provides an overview of SHVC, including the basic SHVC architecture, SHVC systems and transport interfaces, a comparison of SHVC and Scalable Video Coding (SVC), the scalable extension of H.264/AVC [9], and SHVC decoder and encoder complexity analyses.

5.1 Basic SHVC architecture

Inter-layer prediction is employed in a scalable system to improve the coding efficiency of the enhancement layers. In addition to the spatial and temporal motion-compensated predictions that are available in a single-layer codec, inter-layer prediction (ILP) in SHVC uses the reconstructed video signal from a reference layer to predict the current enhancement layer. Inter layer prediction in SHVC is built upon the so called "reference index" framework. With this framework, the collocated reconstructed picture from the reference layer is treated as a long-term reference picture, and is assigned a reference index (or reference indices) in the reference picture list(s) along with other temporal reference

pictures in the current layer. Then, ILP is achieved at the block-level (Prediction Unit-level) by setting the value of the ref_idx syntax element to correspond to the inter-layer reference picture(s) in the reference picture list(s).

Figure 1 shows the SHVC codec architecture from the decoder's perspective. SHVC supports more layers, but for ease of explanation, Figure 1 only describes a two-layer scalable system consisting of the base layer (BL) and one enhancement layer (EL).

As will be discussed later in clause 8.3, one fundamental benefit of the "reference index" based SHVC architecture is that it allows the EL codec to maintain the same block level logics as a single-layer HEVC decoder. The EL codec differs from a single-layer HEVC decoder only at the high level syntax level, i.e. at or above the slice header level. Hence, the EL decoder is labeled as HEVC* in Figure 1 to reflect this. To achieve efficient inter-layer prediction, inter-layer processing is applied to the reconstructed BL pictures retrieved from the BL Decoded Picture Buffer (BL DPB); afterwards, the processed pictures are put into the EL Decoded Picture Buffer (EL DPB) and used as inter-layer reference pictures for predictive coding of the EL pictures. SHVC applies different forms of inter-layer processing depending on the types of scalability between the two layers. For example, for spatial scalability, resampling of texture and/or motion information from the reference layer is applied. By adjusting the sample bit depth during resampling, SHVC also supports bit depth scalability. For color gamut scalability, a color mapping process is applied. Further detailed discussion of inter layer processing modules supported by SHVC can be found in [11].

As shown in Figure 1, the base layer bitstream can be sent either as part of the SHVC bitstream "in-band", or obtained via "external means" in an "out-of-band" manner. In the former case when the base layer is embedded within the SHVC bitstream, the input bitstream is de-multiplexed into two separate layers. The base layer (BL) bitstream is sent to the base layer decoder and the enhancement layer (EL) bitstream is sent to the EL decoder. The BL decoder is an HEVC decoder; in the Scalable Main and Scalable Main 10 profiles currently defined in SHVC, the BL decoder conforms to either the HEVC Main or Main 10 profile. Additionally, SHVC also allows the base layer bitstream to be provided via external means, for example, through other system-level multiplexing methods. This latter function can be used to support the use case when the base layer bitstream is coded using a non-HEVC single-layer codec, for example, using H.264/AVC, MPEG-2, or even non-standardized codecs. Accordingly, this is also referred to as hybrid codec scalability. For hybrid codec scalability, the BL decoding operations are outside of the scope of the SHVC decoder. After decoding, the reconstructed BL pictures are provided to the SHVC decoder, along with some information associated with the BL pictures. The remaining SHVC decoding operations are the same as the former case with the embedded BL bitstream. The SHVC decoder applies inter-layer processing to the reconstructed BL pictures to obtain the inter-layer reference pictures for predictive coding of the EL video pictures. It is worth noting that, although BL bitstreams provided via external means are generally expected to be non-HEVC coded, an HEVC-coded BL bitstream can be provided via external means as well.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ETSI TR 126 948 V17.0.0 (2022-05)
https://standards.iteh.ai/catalog/standards/sist/73ff2ef5-db5e-42df-a365-868998e213d8/etsi-tr-126-948-v17-0-0-2022-05

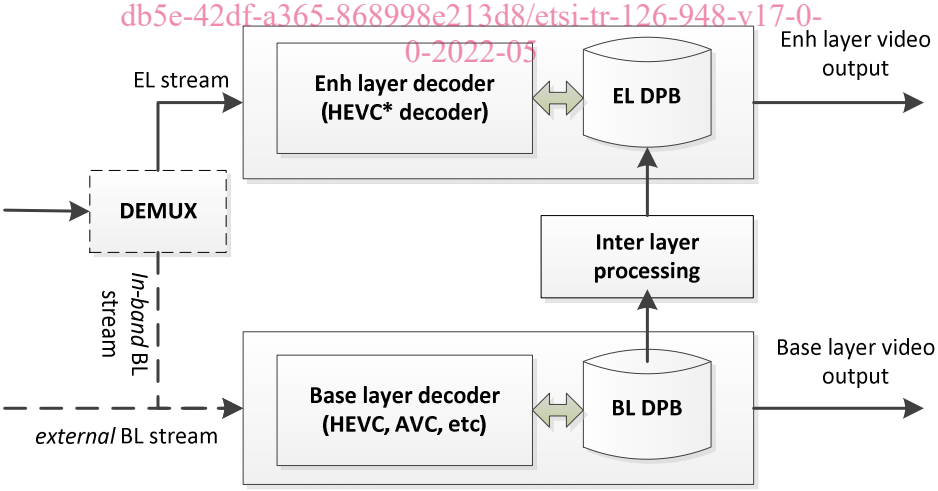

Figure 1: SHVC decoder architecture

5.2 Systems and transport interfaces of SHVC

5.2.1 Introduction

The systems and transport interfaces of a video codec, also referred to as high-level syntax (HLS), are an integral part of a video codec. An important part is the network abstraction layer (NAL), providing a (generic) interface of a video

codec to (various) networks/systems. HLS topics include (but are not limited to) bitstream structure and coded data units structures; parameter sets signalling; support of random access and stream adaptation; error resilience; coded and decoded picture buffer management and buffering model (a.k.a. hypothetical reference decoder or HRD); scalability; byte stream format; profile and level signalling; signalling of supplemental enhancement information (SEI) and video usability information (VUI); extensibility and backward compatibility.

HEVC (single-layer coding) HLS was designed with significant consideration of extensibility mechanisms. These are also referred to as hooks, which basically allow future extensions that would be backward compatible to earlier versions of the standard. Important HLS hooks in HEVC include: a) Inclusion of layer identifier (ID) in the NAL unit header, whereby the same NAL unit header syntax applies to both HEVC single-layer coding and its multi-layer extensions; b) Introduction of the video parameter set (VPS), which was introduced mainly for use with multi-layer extensions, as VPS contains cross-layer information; c) Introduction of the layer set concept and the associated signalling of multi-layer HRD parameters; d) Addition of extensibility for all types of parameter sets and slice header, which allows the same syntax structures to be used for both the base layer and enhancement layers without defining new NAL unit types and to be further extended in the future when needed.

A common HLS framework has been jointly developed for SHVC and MV-HEVC (which is largely applicable to 3D-HEVC as well). This clause focuses on the new HLS features developed for the three multi-layer HEVC extensions compared to HEVC single layer coding HLS, for which an overview can be found in [12] and TR 26.906 [13]. More details of SHVC Systems and transport interfaces be found in [11] and [14].

5.2.2 Parameter Set and Slice Segment Header Extensions

The VPS has been extended by adding the VPS extension structure to the end, which mainly includes information on: a) Scalability type and division of NAL unit header layer ID to scalability IDs; b) Layer dependency, dependency type, and independent layers; c) Layer sets and output layer sets; d) Sub-layers and inter-layer dependency of sub-layers; e) Profile, tier, and level (PTL); f) Representation format (resolution, bit depth, color format, etc.); g) Decoded picture buffer (DPB) size; h) cross-layer video usability information (VUI), which includes information on cross-layer picture type alignment, cross-layer intra random access point (IRAP) picture alignment, bit rate and picture rate of layer sets, video signal format (color primaries, transfer characteristics, etc.), usage of tiles and wavefronts and other enabled parallel processing capabilities, and additional HRD parameters.

iTeh STANDARD PREVIEW (standards.iteh.ai)

It should be noted that the VPS applies to all layers, while in the AU decoding order dimension it applies from the first AU where it is activated up to the AU when it is deactivated. Different layers (including the base layer and a non-base layer) may either share the same SPS or use different SPSs. Pictures of different layers or AUs can also share the same picture parameter set (PPS) or use different PPSs. To enable sharing between sequence parameter set (SPS) and PPS, all SPSs share the same value space of their SPS IDs, regardless of the layer ID values in their NAL unit headers; the same is true for PPSs.

ETSI TR 126 948 V17.0.0 (2022-05)
https://standards.iteh.ai/catalog/standards/sist/73ff2e65-d65c-4f46-a255-8675098c51a8/etsi-tr-126-948-v17.0.0-2022-05

Among other smaller extensions, the slice segment header has been extended in a backward compatible manner by adding the following information: a) The discardable flag that indicates whether the picture is used for at least one of inter prediction and inter-layer prediction or neither (when neither applies the picture can be discarded without affecting the decoding of any other pictures, in the same layer or other layers); b) A flag that indicates whether an IDR picture is a bitstream splicing point (if yes, then pictures from earlier AUs would be unavailable as references for pictures of any layer starting from the current AU); c) Information on lower-layer pictures used by the current picture for inter-layer prediction; and d) POC resetting and POC most significant bits (MSB) information. The latter two sets of information are used as the basis for derivation of the inter-layer reference picture set (RPS) and for guaranteeing cross-layer POC alignment, both of which are discussed later.

5.2.3 Layer and Scalability Identification

Each layer is associated with a unique layer ID, for which the value will be increasing across pictures of different layers in decoding order within an AU. In addition, a layer is associated with scalability IDs specifying its content, which are derived from the VPS extension and denoted as view order index and auxiliary ID.

All layers of a view have the same view order index. The view order index is required to be increasing in decoding order of views. Furthermore, a view ID value is signalled for each view order index, which can be chosen without constraints, but should indicate the view's camera position (e.g. in a linear setup).

The auxiliary ID signals whether a layer is an auxiliary picture layer carrying depth, alpha or other user defined auxiliary data. By design choice, auxiliary picture layers have no normative impact on the decoding of non-auxiliary picture layers (denoted as primary picture layers).

5.2.4 Layer sets

The concept of layer sets was already introduced in HEVC version 1. A layer set is a set of independent decodable layers that contains the base layer. Layer sets are signalled in the base part of the VPS. During the development of the common multi-layer HLS, two related concepts, namely output layer sets (OLSs) and additional layer sets, were further introduced. An OLS is a layer set for which the target output layers are specified (non-target-output layers are for example those layers that are used only for inter-layer prediction but not for output/display). For example, an OLS can have two layers for output (e.g. stereoscopic viewing) but contain three layers. An HEVC single-layer decoder would only process one target output layer, i.e. the base layer, regardless of how many layers the layer set contains. This is the reason why the concept of OLS layer set was not needed in HEVC version 1.

An additional layer set is a set of independent decodable layers that does not contain the base layer. For example, if a bitstream contains two simulcast (i.e. independently coded) layers, then the non-base layer itself can be included in an additional layer set. This concept can also be used for signalling the PTL for auxiliary picture layers, which are usually coded independently from the primary picture layers. For example, a depth or alpha (i.e. transparency) auxiliary picture layer can be included in an additional layer set and indicated to conform to the Monochrome (8 bit) profile, regardless of which single-layer profile the base (primary picture) layer conforms to. Without such a design, many more profiles would need to be defined to handle all the combinations of auxiliary picture layers with single-layer profiles. To realize the benefits of this design, an independent non-base layer rewriting process was specified, which "transcodes" independent non-base layers to a bitstream that conforms to a single-layer profile.

By design choice, an additional layer set is allowed to contain more than one layer, e.g. three layers with layer ID values equal to 3, 4, and 5, where the layer with layer ID equal to 3 is an independent non-base layer. Along with this, a bitstream extraction process for additional layer sets was specified. While the extracted sub-bitstream does not contain a base layer, it is still a conforming bitstream, i.e. the multi-layer extensions of HEVC allow for a conforming multi-layer bitstream to not contain the base layer, and compliant decoding of the bitstream may not involve the base layer at all.

iTeh STANDARD PREVIEW (standards.iteh.ai) ETSI TR 126 948 V17.0.0 (2022-05) https://standards.iteh.ai/catalog/standards/sist/7311e2f5-ab5e-42df-a565-86f998e215d8/etsi-tr-126-948-v17-0-0-2022-05

5.2.5 Profile, Tier, and Level (PTL)

Compared to earlier multi-layer video coding standards, a fundamentally different approach was taken for MV-HEVC and SHVC for the specification and signalling of interoperability points (i.e. PTL in the context of HEVC and its extensions). Rather than specifying PTL for an operation point that contains a set of layers, in MV-HEVC and SHVC, PTL is specified and signalled in a layer specific manner. Consequently, a decoder that is able to decode two-layer bitstreams with 720p@30fps at the base layer and 1080p@60fps at the enhancement layer should express its capability as a list of two PTLs equivalent to {Main profile Main tier Level 3.1, Scalable Main profile Main tier Level 4.1}. A key advantage of this design is that it facilitates easy decoding of multiple layers by reusing single-layer decoders. If PTL was specified for the two layers together, then the decoder would need to be able to decode the two-layer bitstreams with both the base and enhancement layers of 1080p@60fps. In other words, over provisioning of resources would be required.

Another related innovation is the definition of the independent non-base layer decoding (INBLD) capability, which is associated with the decoding capability of one or more of the single-layer profiles. The INBLD capability, when supported, indicates the capability of a decoder to decode an independent non-base layer that is indicated in the active VPSs and SPSs to conform to a single-layer profile and is the layer with the smallest nuh_layer_id value in an additional layer set. Compared to conventional single-layer decoders, such single-layer decoders can also parse some multi-layer syntaxes such as VPS extension and handle NAL units with layer ID greater than zero. It is recommended that, when expressing the capabilities of a decoder for one or more single-layer profiles, whether the INBLD capability is supported for those profiles should also be expressed. As can be seen from its connection to additional layer sets, the INBLD capability is also part of the entire solution for auxiliary picture layers introduced to HEVC version 2.

5.2.6 RPS and Reference Picture List Construction

In addition to the five RPS lists (RefPicSetStCurrBefore, RefPicSetStCurrAfter, RefPicSetStFoll, RefPicSetLtCurr, and RefPicSetLtFoll) defined in HEVC version 1, two more RPS lists, RefPicSetInterLayer0 and RefPicSetInterLayer1 (denoted as RpsIL0 and RpsIL1, respectively), were introduced to contain inter-layer reference pictures. Given a current picture, those inter-layer reference pictures are included into two sets depending on whether they have view ID values greater or smaller than the current picture. If the base view has greater view ID than the current picture, then those with greater view IDs are included into RpsIL0 and those with smaller view IDs into RpsIL1, and vice versa. The derivation of RpsIL0 and RpsIL1 is based on VPS extension signalling (of layer dependency and inter-layer dependency of sub-layers) as well as slice header signalling (of lower-layer pictures used by the current picture for inter-layer prediction).

When constructing the initial reference picture list 0 (i.e. RefPicListTemp0), pictures in RpsIL0 are inserted immediately after pictures in RefPicSetStCurrBefore, and pictures in RpsIL1 are inserted last, after pictures in RefPicSetLtCurr. When constructing the initial reference picture list 1 (i.e. RefPicListTemp1), pictures in RpsIL1 are inserted immediately after pictures in RefPicSetStCurrAfter, and pictures in RpsIL0 are inserted last, after pictures in RefPicSetLtCurr. Otherwise the reference picture list construction process stays the same as for HEVC single-layer coding.

5.2.7 Random Access, Layer Switching, and Bitstream Splicing

Compared to AVC, HEVC provides more flexible and convenient random access and splicing operations, by allowing conforming bitstreams to start with a clean random access (CRA) or broken link access (BLA) picture. In addition, MV-HEVC and SHVC support non-cross-layer aligned IRAP pictures of any type (IDR, CRA, or BLA), and a conforming bitstream can start with any type of IRAP access unit, including an IRAP AU where the base layer picture is an IRAP picture while (some of) the enhancement layer pictures are non-IRAP pictures. This allows easy splicing of multi-layer bitstreams at any type of IRAP AU and random accessing from such AU. Non-cross-layer aligned IRAP pictures also allow for flexible layer switching.

To support non-cross-layer aligned IRAP pictures, the multi-layer POC design needs to ensure cross-layer POC alignment within any AU. Cross-layer POC alignment is needed to ensure that the in-layer RPS derivation and the output order of pictures of target output layers are correct.

The multi-layer HEVC design allows extremely flexible layering structures. Basically, a picture of any layer may be absent at any AU. For example, the highest layer ID value can vary from AU to AU, which was disallowed in SVC and MVC. Such flexibilities imposed a great challenge on the multi-layer POC design. In addition, although a bitstream after layer or sub-layer switching is not required to be conforming, the design should still enable a conforming decoding behaviour to work with layer and sub-layer switching, including cascaded switching behaviour. This is achieved by a POC resetting approach.

iTeh STANDARD PREVIEW (standards.iteh.ai)

The basic idea of POC resetting is to reset the POC value when decoding a non-IRAP picture (as determined by the POC derivation process in HEVC version 1), such that the final POC values of pictures of all layers of the AU are identical. In addition, to ensure that POC values of pictures in earlier AUs are also cross-layer aligned and that POC delta values of pictures within each layer remain proportional to the associated presentation time delta values, POC values of pictures in earlier AUs are reduced by a specified amount.

ETSI TR 126 948 V17.0.0 (2022-05)
https://standards.iteh.ai/catalog/standards/sist/7a6f2a7b-dc5e-4a1b-a505-a0c7a2e21508/etsi-tr-126-948-v17.0.0-2022-05

To work with all possible layering structures as well as some picture loss situations, the POC resetting period is specified based on the POC resetting period ID that is optionally signalled in the slice header. Each non-IRAP picture that belongs to an AU that contains at least one IRAP picture will be the start of a POC resetting period in the layer containing the non-IRAP picture. In that AU, each picture would be the start of a POC resetting period in the layer containing the picture. POC resetting and decreasing of POC values of same-layer pictures are applied only for the first picture within each POC resetting period, such that these operations would not be performed more than necessary; otherwise POC values would be messed up.

5.2.8 Hybrid Codec Scalability and Multiview/3D Support

The HEVC multi-layer extensions support the base layer being coded by other codecs, e.g. AVC. A simple approach was taken for this functionality by specifying the base layer being provided by an external means, i.e. not specified by the standard. Basically, except for information on the representation format and whether the base layer is a target output layer as signalled in the VPS extension, no other information about the base layer is included in the bitstream (as input to the enhancement-layer decoder). Decoder implementations can implement an application program interface (API) to accept the sample values of the decoded base layer picture for each AU plus some other minimum amount of information required for decoding the enhancement layer pictures, including whether it is an IRAP picture and, if yes, the IRAP picture type (IDR, CRA, or BLA). The base layer pictures and this latter information may be provided by the base layer decoder through the API, however this is not part of the enhancement-layer decoder specified by HEVC version 2. The output of base layer pictures is the responsibility of the base layer decoder, and output synchronization between a base layer picture and an enhancement layer picture in the same AU, when needed, is externally controlled (e.g. by using presentation timestamps). The association of a base layer decoded picture to an AU is also the responsibility of external means.

5.2.9 Hypothetical Reference Decoder (HRD)

The main new developments of HRD of the common HLS compared to HEVC version 1 include the following three aspects relevant for MV- and 3D-HEVC. Firstly, the bitstream conformance tests specified in HEVC version 1 are classified into two sets and a third set is additionally specified. The first set of tests is for testing the conformance of the entire bitstream and its temporal subsets. The second set of bitstream conformance tests is for testing the conformance of the layer sets specified by the active VPS and their temporal subsets. For the first and second sets of tests, only the base layer pictures are decoded and other pictures are ignored by the decoder. The third set of tests is for testing the conformance of the OLSs specified in the VPS extension and their temporal subsets.

The second aspect is the introduction of bitstream partition (BP) specific coded picture buffer (CPB) operations, wherein each BP contains one or more layers, and CPB parameters for each BP can be signalled and applied. These parameters can be utilized by transport systems that transmit different sets of layers in different physical or logical channels; one extreme example is one channel for each layer. The layer specific CPB parameters are also a basis for defining the semantics of layer specific PTL. The third aspect is the layer specific DPB management operations, where each layer exclusively uses its own sub-DPB. To ensure the design works with (cascaded) layer switching behavior, sharing of a particular memory unit across layers is disallowed.

5.2.10 SEI Messages

SEI messages in HEVC version 1 have been adapted to be applicable in the multi-layer contexts, in a backward compatible fashion, some of them with significant semantics changes. In addition, some new SEI messages are specified that apply to all multi-layer HEVC extensions.

The layers not present SEI message can indicate which layers are dropped. The inter-layer constrained tile sets SEI message, which can indicate cross-layer region of interest coding based on tiles. The BP nesting SEI message and the BP initial arrival time SEI message can be used to signal BP buffering parameters for CPB operations. The sub-bitstream property SEI message provides the bit rate information for a sub-bitstream created by discarding those pictures in the layers that do not belong to the output layers of the OLSs specified by the active VPS and that do not affect the decoding of the output layers. The alpha channel information SEI message provides information about alpha channel sample values and post-processing applied to the decoded alpha plane auxiliary pictures, and one or more associated primary pictures. Other new SEI messages that apply to all multi-layer HEVC extensions include the temporal motion vector prediction constraints SEI message and the frame-field information SEI message.

iTeh STANDARD PREVIEW (standards.iteh.ai) ETSI TR 126 948 V17.0.0 (2022-05) https://standards.iteh.ai/catalog/standards/sist/73ff2ef5-db5a-42df-a365-868998e212d8/etsi-tr-126-948-v17-0-0-2022-05

5.3 A comparison of SHVC and SVC

The scalable extension to H.264/AVC, commonly referred to as SVC, is the most recent scalable video coding standard preceding SHVC. In this clause a brief comparison between SHVC and SVC is provided.

The first difference between SHVC and SVC is the indication (i.e. signalling) of inter layer prediction. The previous SVC standard uses a base_mode_flag signalled at the macroblock level to enable inter-layer prediction, and makes further block-level operation changes depending on the value of base_mode_flag. In comparison, SHVC uses the "reference index" based framework, as discussed above. Because the ref_idx syntax element already exists in the single-layer HEVC standard at the Prediction Unit-level, referencing an inter-layer reference picture can be carried out in a transparent manner at the block-level. In other words, all block-level logic, including parsing and interpretation of the syntax elements, decoding and reconstruction, loop filtering, and other related processes, of the EL codec can be kept unchanged from those of a single-layer HEVC codec. Any necessary changes to the EL decoder, denoted as HEVC* in Figure 1, are only at slice header-level and above, that is, at the high level syntax level. By keeping the detailed block-level operations compatible with a single-layer codec, SHVC can be implemented by reusing/repurposing most parts of an existing HEVC implementation; thus, the implementation cost of SHVC can be reduced significantly.

Secondly, to achieve inter-layer prediction, the only BL information that the EL needs to access is the reconstructed pictures from the BL DPB, which includes the reconstructed texture samples, and typically also the BL motion information. As the BL DPB needs to be provided as an open interface in a single-layer codec implementation, such scalable codec architecture requires no change at all to the BL codec, and allows the BL codec to essentially operate as a black box. In contrast, in addition to using the BL reconstructed texture to predict the EL, SVC also applies cross-layer syntax prediction and cross-layer residual prediction, the implementation of which requires the BL codec to be redesigned to provide much more information than what it would generally need to provide as a single-layer codec. As such, implementation of the SHVC codec is much simpler than that of SVC. Further, operating the BL codec as a "black box" also allows SHVC to support hybrid codec scalability discussed earlier, thus providing expanded backward compatibility support.