ITUPublications

International Telecommunication Union
Radiocommunication Sector

Report ITU-R BT.2538-1

(10/2026)

BT Series: Broadcasting service (television)

Use cases of versatile video coding multilayer profiles for broadcasting services

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from https://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Reports

(Also available online at https://www.itu.int/publ/R-REP/en)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	**Broadcasting service (television)**
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radio-wave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
TF	Time signals and frequency standards emissions

***Note**: This ITU-R Report was approved in English by the Study Group under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2026

© ITU 2026

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

REPORT ITU-R BT.2538-1

Use cases of versatile video coding multilayer profiles for broadcasting services*

(Question ITU-R 12-4/6)

(2024-2026)

TABLE OF CONTENTS

		Page
1	Introduction	2
2	Multilayer coding capability of the VVC standard	2
2.1	Multilayer profiles	2
2.2	Multilayer coding technologies	3
3	Use cases of multilayer profiles	4
3.1	Use Case 1: Optimization of frequency resource utilization	4
3.2	Use Case 2: Live content overlaid on offline-encoded video	5
3.3	Use Case 3: Premium content distribution linked to the Internet	7
3.4	Use Case 4: Sub-content delivery based on viewers' demands	9
4	Signalling of video layer information in VVC multilayer bitstreams	11
5	Signalling of video layer information in multiplexing layer	14
6	Concept of association between video and audio presentations	15

* This Report should be brought to the attention of ITU-R Working Party 4B.

List of abbreviations and acronyms

BL	Base layer
DPB	Decoded picture buffer
EL	Enhancement layer
HDTV	High-definition television
HEVC	High efficiency video coding
OTT	Over-the-top
RPR	Reference picture resampling
UHDTV	Ultra-high-definition television
VoD	Video on demand
VVC	Versatile video coding

1 Introduction

Versatile video coding (VVC), specified in Recommendation ITU-T H.266 | International Standard ISO/IEC 23090-3, specifies a video coding technology with a compression capability that is substantially beyond that of the prior generations of such standards. The coding technology is highly versatile for use in a broadened range of applications. Some key application areas of this standard include ultra-high-definition video and immersive media applications.

Versatile video coding also supports multilayer coding, by which different types of scalabilities can be realized such as spatial scalability, quality scalability, and content scalability. These scalabilities can be usefully used to provide new types of broadcasting services including “personalization” and “integration with broadband”.

This Report describes various use cases of the multilayer coding in digital broadcasting based on the multilayer profiles supported by VVC.

2 Multilayer coding capability of the VVC standard

2.1 Multilayer profiles

The VVC standard contains a total of 15 profiles. Among them, the general profiles specified in the first version of VVC are the following six profiles:

– Main 10 profile
– Main 10 Still Picture profile
– Main 10 4:4:4 profile
– Main 10 4:4:4 Still Picture profile
– Multilayer Main 10 profile
– Multilayer Main 10 4:4:4 profile.

The Main 10 profile is considered as a basic profile for traditional digital broadcastings. Profiles with a multilayer (scalable) coding capability are Multilayer Main 10 profile and Multilayer Main 10 4:4:4 profile. “10” means that the bit depth of the video signal is 10-bit, and 4:4:4 indicates sampling format of chrominance signals. The sampling format of the Multilayer Main 10 profile is 4:2:0, which is used in legacy broadcast standards, and is omitted from the profile name.

2.2 Multilayer coding technologies

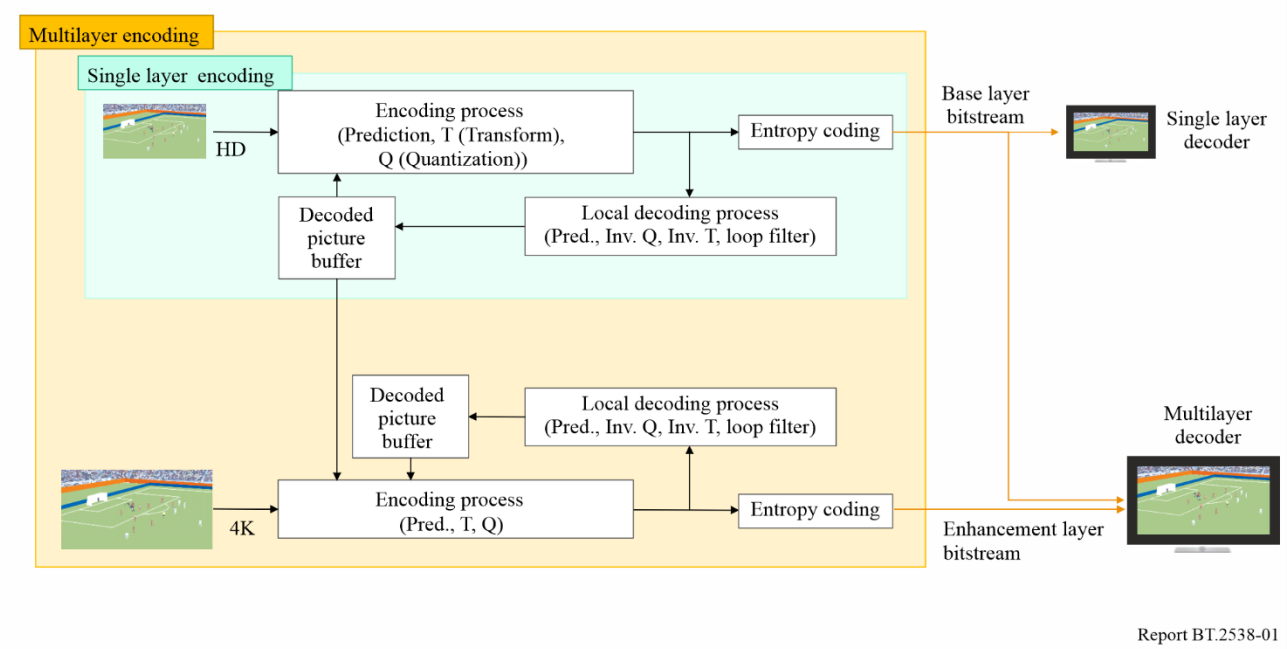
Multilayer coding collectively encodes multiple variations of the content. There are two types of variations: signal variants (e.g. UHDTV video and HDTV video) and service variants (e.g. stereo video). Figure 1 illustrates the mechanism of VVC multilayer encoding. VVC multilayer encoding comprises a base layer (BL) and enhancement layer (EL) encoding processes. BL encoding is a single-layer encoding process. Input video signal of BL is predicted by exploiting intra-frame correlation (intra prediction) or inter-frame correlation (inter prediction). Transform and quantization processes are then applied to the residual signal obtained by the prediction process. Then, quantized coefficients and mode information including prediction information are entropy coded and the resulting bitstream is output as a BL bitstream. A local decoding process is applied to the quantized coefficients and the reconstructed locally-decoded video signal is stored in a decoded picture buffer (DPB) as a reference picture for the prediction of subsequent pictures to be encoded.

The EL encoding process also consists of similar encoding and local decoding processes. However, in the prediction process of EL, the reference pictures stored in DPB of BL can also be used as reference pictures for the EL prediction process. This means that inter-layer prediction can be performed in the EL encoding process in addition to intra-layer prediction. When referencing the decoded picture of BL for predicting the input video signal of EL, conventional inter prediction does not work efficiently since the video size of the input video signal might be different between BL and EL. For this reason, when referencing DPB of BL for inter-prediction, a rescaling process is performed based on the ratio of BL and EL video size.

Encoding using the correlation between BL and EL enables reduction of the total bit rate compared to encoding those variations individually. Meanwhile, a bitstream coded in a conventional scheme with the Main 10 profile only requires a decoding process for one layer, while decoding processes for both the BL and EL bitstreams are required for decoding a bitstream coded with multilayer profiles. Since synchronization of the decoding process for each layer is necessary, additional receiver resources might be required.

FIGURE 1

Mechanism of VVC multilayer encoding



Since the rescaling process during the inter-layer prediction can be performed within the reference picture resampling (RPR) process included in the toolsets of the Main 10 profile, multilayer profiles of the VVC standard can be extended from the Main 10 profile without additional coding tools, different from the conventional standards such as the HEVC standard.

In the conventional standard, the inter-layer prediction method between BL and EL in scalable coding is different from the prediction method within the layer, so that the implementation complexity increases. Therefore, products conforming to the multilayer profiles were limited. In the VVC standard, the inter-layer prediction method between layers is unified with the prediction method within the layer. This unification brings reduction of the implementation complexity compared to the previous standard. It is expected that the barriers to introduction of the VVC standard to television receivers will be lowered. On the other hand, it is also anticipated that the market for decoders conforming to multilayer profiles will come later than that for Main 10 profile decoders. If the market size is small, the cost of receivers compatible with multilayer profiles may increase.

3 Use cases of multilayer profiles

The following use cases are identified as potential applications using multilayer profiles for broadcast and streaming services:

TABLE 1

Use cases for multilayer profiles

	Use case	Description
1	Optimization of frequency resource utilization	Quality improvement can be achieved by introducing multilayer coding into the coding of videos for both mobile receivers and TV sets.
2	Overlaying live content on top of offline-encoded video	Realizing transmission of offline-encoded video and live superimposition of breaking news on the offline-encoded video.
3	Premium content distribution via the Internet	Delivery of the content with higher resolution and better quality via Internet.
4	Sub-content delivery based on viewer needs	Selectable sub-content according to the viewers' demand such as sign-language video and picture-in-picture video.

3.1 Use case 1: Optimization of frequency resource utilization

3.1.1 Issues

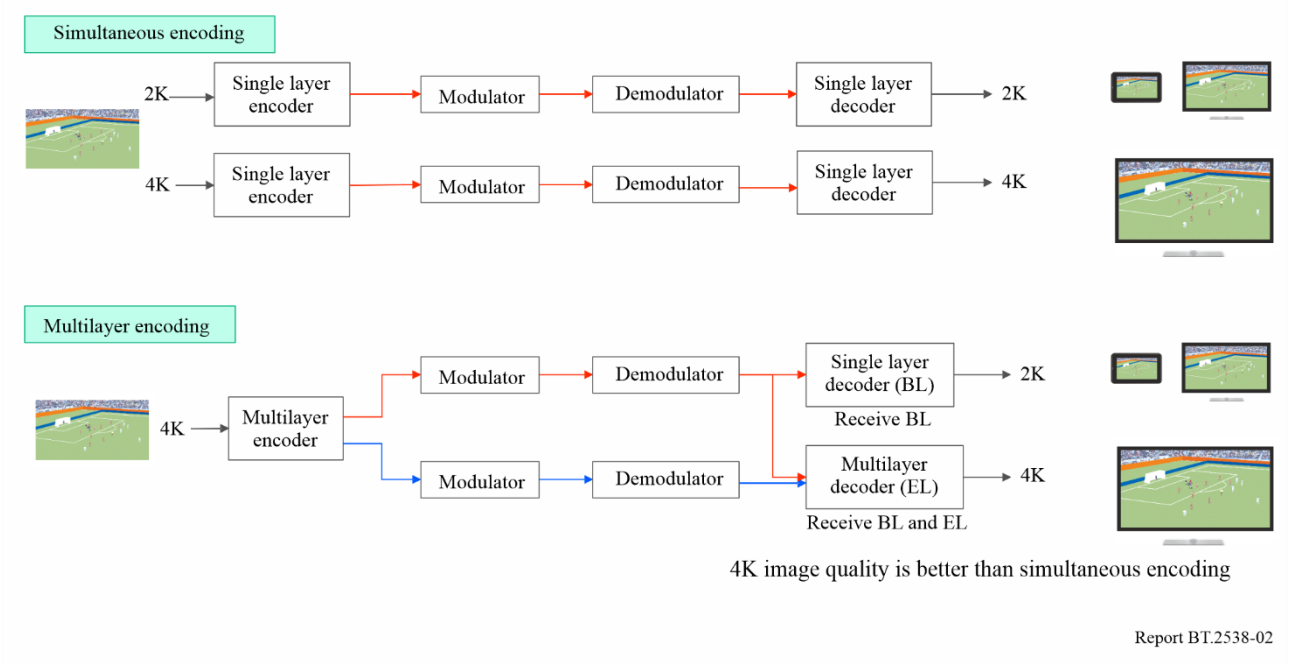
In case of terrestrial or satellite broadcasting, the overall bit rate that can be transmitted is fixed. For this reason, when broadcasting for both mobile receivers (low-resolution video) and TV sets (high-resolution video), the total bit rate needs to be shared. Therefore, when the bit rate of one service is increased, that of the other service is decreased, resulting in a problem of degraded visual quality.

3.1.2 Solution with multilayer profiles

Multilayer (scalable) coding can be applied where low-resolution video for mobile receivers is coded as BL and high-resolution video for TV sets is coded as EL. The home TV sets can decode and display the high-resolution video by receiving both bitstreams for mobile receivers and home TV sets.

FIGURE 2

Comparison between simultaneous encoding and multilayer encoding



3.1.3 Features

Coding efficiency can be improved by utilizing correlation between BL and EL, resulting in the improvement of visual quality of high-resolution video for home TV sets. On the other hand, since the decoding processing of EL is conducted after BL decoding, decoder complexity increases approximately 1.25 times when the low-resolution video is 2K and the high-resolution video is 4K compared to the case of 4K single-layer decoding.

3.1.4 Alternative solutions

3.1.4.1 Implementation using simultaneous encoding

Low-resolution and high-resolution video are encoded separately and broadcast to mobile and fixed receivers, respectively. A low-cost single-layer decoder compatible with the Main 10 profile is likely to be the first to reach the market. Therefore, separated broadcasting services for mobile and fixed receivers can be more feasible at early stage of deployment. On the other hand, the compression efficiency of separate encoding could be degraded compared to multilayer coding, resulting in visual quality degradation.

3.2 Use Case 2: Live content overlaid on offline-encoded video

3.2.1 Issues

In conventional broadcasting, urgent and important information reporting incidents, accidents, and disasters (hereafter referred to as live content) is often overlaid as text information on the TV programme in Japan. To immediately convey those kinds of important information, real time encoding system is adopted, where TV programmes are always encoded in real time whether important information is overlaid or not.

On the other hand, in over-the-top (OTT) media distribution, an offline encoding and transmission system (video on demand (VoD) service) that generates an encoded bitstream in advance is common. Since encoding complexity is less restrictive in OTT than live broadcasting, the degree of optimization of encoding control is higher than that of the real-time encoding system, which results

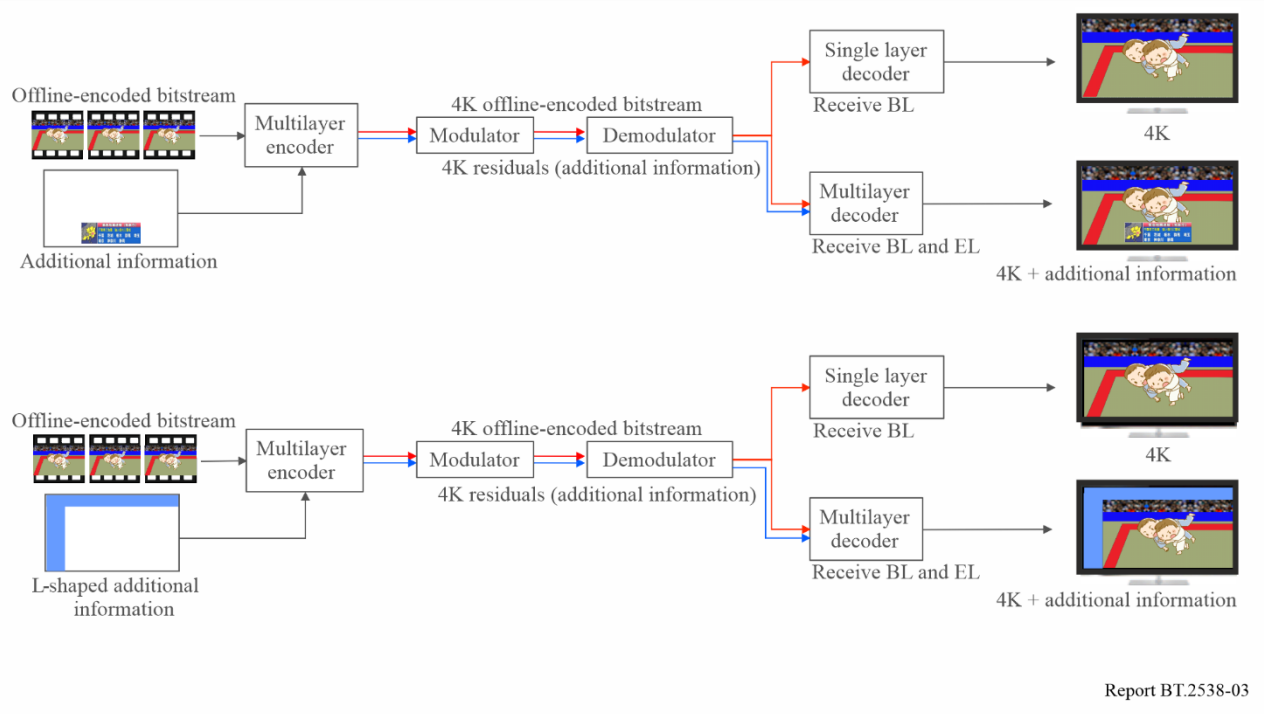
in higher visual quality. In broadcasting, it is also expected that higher image quality will be achieved while maintaining a lower bit rate by introducing offline encoding. The problem is how to superimpose the live content in offline-encoded video.

3.2.2 Implementation with multilayer profiles

The main content is offline-encoded and reused as BL. The live content is overlaid on top of decoded video of BL and encoded in real time as EL. The receiver can display the main content by decoding only the BL bitstream, while the video superimposed with the live content can be displayed by decoding both BL and EL bitstreams. When highly urgent live content is transmitted, an emergency flag of the system layer can be used as a trigger to make the receiver forcibly decode both BL and EL so that no one misses those types of the important information.

FIGURE 3

Live content overlay using multilayer coding



3.2.3 Features

In the transmission of offline-encoded video, it is possible to convey live content on top of the main content, which is not feasible in conventional broadcasting. The types of live content are not limited to text information, but also include L-shape information which is often conveyed in Japan for weather warning. Since EL decoding involves BL decoding, the number of pixels to be decoded is doubled compared to conventional single-layer coding. On the other hand, introducing offline encoding with a higher degree of optimization of encoding control is expected to improve visual quality.

3.2.4 Alternative solutions

3.2.4.1 Implementation by video synthesis on the receiver side

Live content is separately encoded in real time as a live content bitstream different from the offline-coded main content bitstream. Additionally, metadata including arrangement information and

presentation time information is transmitted. The receiver decodes both bitstreams and synthesizes those decoded videos based on the transmitted metadata.

This solution is feasible with low-cost Main 10 profile decoders that will be firstly available on the market. In addition, by adopting the display on a Web browser, it is expected that it will be possible to easily cope with the complicated overlaying layout. On the other hand, since it is necessary to decode the main content and the live content independently, the number of pixels to be decoded is increased by the number of pixels of the live content. In addition, high-performance video processing may be required for content synthesis control.

3.2.4.2 Implementation by content composition on the sending side

As with the conventional broadcast systems, all content is synthesized on the sending side. A real-time encoder is used to deliver the synthesized content.

Offline encoding cannot be achieved since synthesized content is encoded in real time, so that coding efficiency using this approach is lower than that of implementation with multilayer profiles.

3.2.4.3 Implementation by sub-picture coding

Sub-picture coding of the VVC standard divides the whole content into multiple tiles and encodes each tile independently to achieve parallel and partial decoding. By sub-picture, it is possible to replace a part of the main content by replacing a part of the bitstream within the bitstream of the main content. At the sending side, the additional bitstream which corresponds to the live content is transmitted in addition to the bitstream of the main content. Since the tile layout is arranged at the time of encoding, content synthesis at the receiver side is unnecessary.

Assuming that the live content will be superimposed on various areas for each content, it is necessary to configure the complicated layout of sub-pictures even though the live content is not always activated. In addition, there are constraints on tile division in VVC, so that the above-mentioned complicated layout might be impossible. Furthermore, the complex tile layout may complicate verification of decoding operations during receiver development.

In order to increase the flexibility of layout, it is necessary to finely divide the input image into the tiles, but this reduces the coding efficiency for the main content. In addition, since the replacement is limited to sub-picture units, the L-shape live content which requires resizing of the main content shown in Fig. 3 is not feasible.

3.3 Use Case 3: Premium content distribution linked to the Internet

3.3.1 Issues

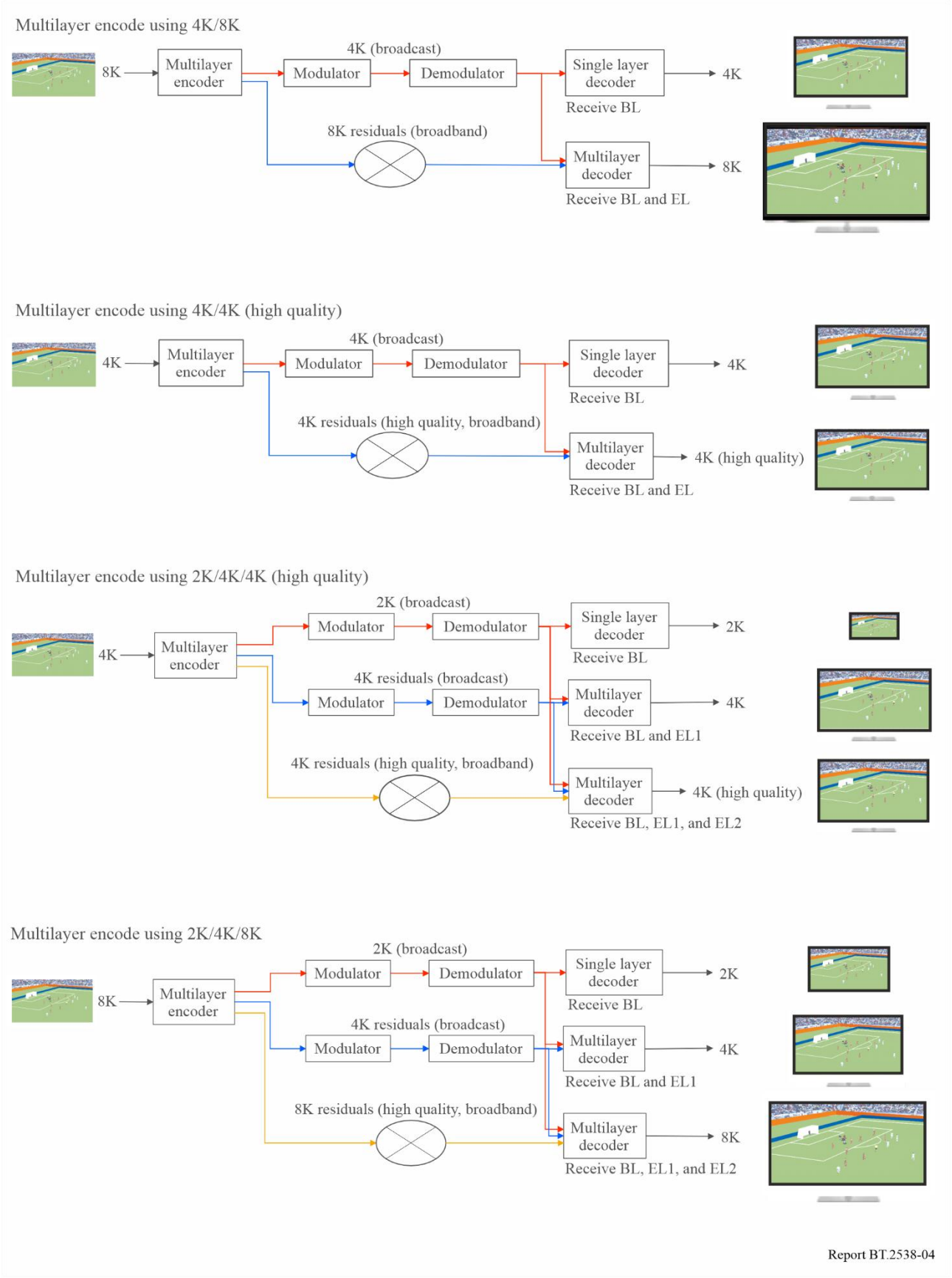
By linking broadcasting and broadband, it is possible to provide premium services that bring added value (high image quality, high resolution, etc.) that could not be achieved with conventional broadcasting services.

3.3.2 Implementation with multilayer profiles

High-quality video services are realized as a premium service by providing the EL bitstream via broadband, while the BL bitstream is provided via radio wave broadcasting as a basic service. Receivers that support both broadcast and broadband can receive, decode, and display high-quality video.

FIGURE 4

Premium content via broadband in combination with broadcast



3.3.3 Features

When several contents are transmitted through different routes, signal synchronization is required at the receiver side due to the difference in transmission delay for each route. However, by distributing the high-quality video as EL through the broadcasting path, it is possible to reduce the bit rate to be transmitted through radio wave broadcasting as compared with the case of simulcast. However, the number of pixels to be decoded increases 1.25 times (2K/4K case) to 1.31 times (2K/4K/8K case) compared to that of the conventional single-layer case since EL decoding requires BL decoding.

With this method, as an extended service, it becomes possible to provide additional distribution services that are not feasible to distribute due to transmission capacity issues in radio wave broadcasting. In addition, when the transmission band is expanded by using broadband or channel bonding, the extended service leaves existing services unchanged.

3.3.4 Alternative solutions

3.3.4.1 Implementation by changing the transmission method

High-quality video services will be provided only by broadcast waves by increasing the bit rate of broadcasting through channel bonding and the use of frequencies that have become available after the transition to the new system. However, when transitioning from the existing system to the new one, a certain period of transition time may be required to achieve backward compatibility, so that the feasibility of this alternative might be low.

3.3.4.2 Implementation by adding a transmission route

The premium services are provided only using broadband. Since the premium service is independent from broadcast waves, the required bit rate is higher than that with multilayer profiles.

3.4 Use Case 4: Sub-content delivery based on viewers' demands

3.4.1 Issues

Conventional broadcasting services assume that an unspecified number of viewers watch the same video. Therefore, it has not been possible to provide different videos according to the needs of each viewer. In recent years, there is a growing need for improved accessibility to broadcasting services such as sign language services. Also in sports programmes, there is another need for a multi-view service that allows viewers to select their favourite view among videos taken from various viewpoints. In this way, in addition to the main content, it is necessary for a content provider to deliver sub-contents according to a viewer's wishes.

3.4.2 Implementation with multilayer profiles

Main content is delivered as a basic service, and sub-content that meets viewer needs is provided as an additional service by encoding the main content as BL and the sub-content as EL. By doing so, it becomes possible for the viewer to select the reception layer and selectively display the content that the viewer wants to view.

3.4.3 Features

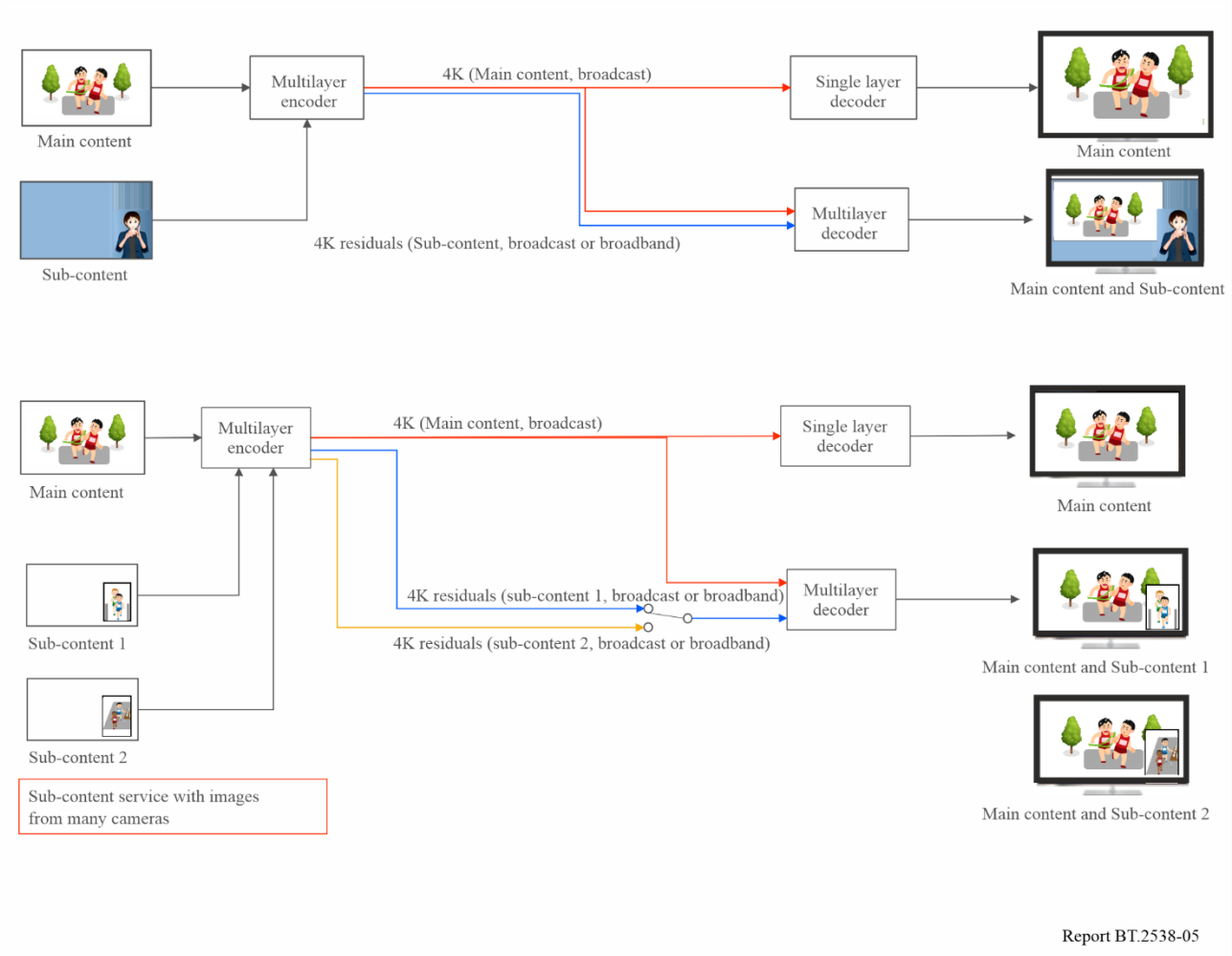
The layout of sub-contents is managed at the time of production, so that no technical restrictions on the layout are imposed. The receiver needs to select only the layer to be decoded by the user instruction. Video synthesis processing is not required, which results in no difference in appearance between models or manufacturer of the receivers. In addition, since the viewer cannot freely change the layout of the sub-contents, the content is always displayed as intended by the production side. Moreover, extension of the service without affecting existing receivers can be feasible since the

existing receivers without decoding capability of bitstreams with multilayer profiles can also decode BL which contains the main content.

When sub-contents are transmitted using a different route such as broadband, an additional synchronization process is required in receivers due to the difference in transmission delay. In addition, since decoding of EL always requires decoding of BL, the number of pixels to be decoded is doubled compared to the conventional method in which a single video is transmitted at the time of transmission.

FIGURE 5

Sub-content delivery using multilayer coding



3.4.4 Alternative solutions

3.4.4.1 Implementation by sub-picture coding

Similar to use case 2, sub-picture division can be applied (see § 3.2.4.3). Sub-content is encoded individually and delivered as a bitstream independent from the main content. The receiver replaces a part of bitstream of the main content with the sub-content bitstream selected by the viewer instruction.

It has the same layout disadvantages as use case 2 (see § 3.2.4.3) and restricts the position of multi-view video in production. In addition, since the replacement is limited to a unit of sub-picture, it is impossible to realize L-shape sub-content such as sign language illustrated in Fig. 5.

3.4.4.2 Implementation by screen synthesis in the receiver

A main content and sub-content are encoded individually and transmitted with auxiliary information indicating layout control. The receiver decodes each bitstream and performs video synthesis based on the layout information. Broadcasting services can be realized without technical restrictions on production. Although the implementation complexity increases, viewers can freely change the layout of sub-contents. Since it is necessary to decode the main content and the sub-contents, multiple decoders are required, and the number of pixels to be decoded increases by the number of pixels of the sub-contents compared to the conventional method.

Although multiple decoders and video synthesis processing are required, this can be feasible with low-cost Main 10 profile decoders that will be firstly available in the market. In addition, it is expected that the introduction of Web browser decoding will make it possible to easily cope with the complicated content layout. However, video synthesis control may require high-performance video processing.

3.4.4.3 Realization by screen synthesis on the sending side

As with the conventional broadcast systems, all content is synthesized on the sending side; however, it is not possible to meet the needs of the viewers to freely select the video based on the viewer's need.

4 Signalling of video layer information in VVC multilayer bitstreams

This section describes the signalling of layer structure in a VVC multilayer bitstream.

The VVC standard introduces the concept of an "output layer set" (OLS) to signal the layer structure. A VVC multilayer bitstream may contain multiple OLSs, each assigned an index (OLS idx). The receiver can access the intended service by specifying the OLS idx corresponding to the desired service. An OLS consists of one or more video layers and one or more output video layers. All output video layers within an OLS are decoded simultaneously. Each video layer, except for the base layer (BL), may have decoding dependencies on other layers. A video layer may be decoded either for internal processing or for display output. OLS information is signalled in the Video Parameter Set (VPS) NAL unit.

A multilayer bitstream includes an OLS that contains only the BL, allowing decoders compatible with single layer profiles (e.g. the Main 10 profile) to decode the BL bitstream.

Decoders compatible with multilayer profiles (e.g. the Multilayer Main 10 profile) can switch between layers for display output. There are two methods for output switching: OLS-based output switching and layer-based output switching.

– In OLS-based output switching, the decoder selects and decodes a single OLS, which contains only one output layer. Switching can occur at each coded video sequence start AU (CVSS AU). Decoder initialization occurs during OLS switching, similar to conventional channel zapping, which results in temporary video interruption.
– In layer-based output switching, the decoder selects one output layer from multiple output layers within the selected OLS. Since all output layers within the OLS are decoded simultaneously, instantaneous switching at every picture is possible, although this requires higher processing capability on the receiver side.

Table 2 shows OLS information for the use cases described in § 3. Figures 6 (a) through 6 (d) illustrate the OLS-based output switching and layer-based output switching.

In the use Cases 1 and 3, where the layer to be viewed is assumed to be fixed, only OLS-based switching is described in Table 2 and Figs. 6 (a) and 6 (c). This is because the displayed layer is

uniquely determined by the receiving device or environment, and the viewer is not expected to actively switch between layers.

In contrast, in the use cases 2 and 4, where the viewer may choose to access additional content (e.g. premium content or sub-content), which may involve actively switching the displayed layer, both OLS-based and layer-based output switching are available options, as shown in Table 2 and Figs. 6 (b) and 6 (d).

TABLE 2

OLS information in use cases described in Section 3

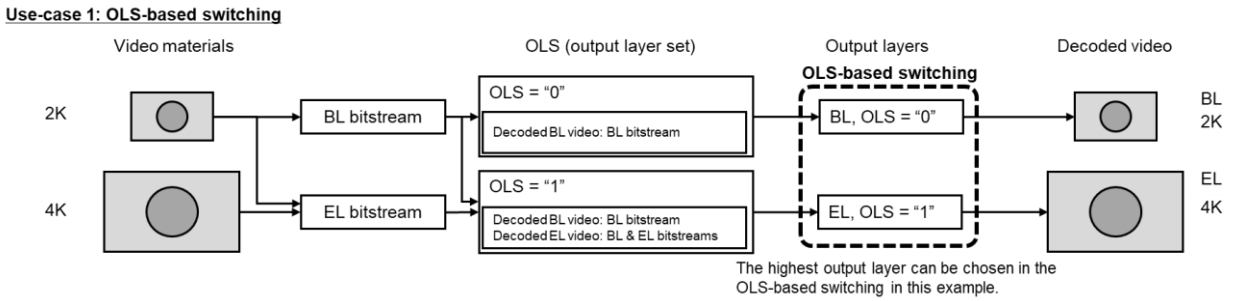
Use case		Decoding dependency	Output switching	OLS information		
				OLS idx	Video layers in OLS	Output video layers (1)
1	Optimization of frequency resource utilization (2 layers, BL: 2K, EL: 4K)	EL refers to BL	OLS-based	0	BL	BL
				1	BL, EL	EL (2)
2	Overlaying live content on top of offline-encoded video (2 layers, BL: 4K, EL: additional information)	EL refers to BL	OLS-based	0	BL	BL
				1	BL, EL	EL (2)
			Layer-based	0	BL	BL
				1	BL, EL	BL or EL
3	Premium content distribution via the Internet (3 layers, BL: 2K, EL1: 4K, EL2: 8K via Internet)	EL1 refers to BL EL2 refers to BL and EL1	OLS-based	0	BL	BL
				1	BL, EL1	EL1 (2)
				2	BL, EL1, EL2	EL2 (2)
4	Sub-content delivery based on viewer needs (3 layers, BL: main, EL1: main and sub-1, EL2: main and sub-2)	EL1 refers to BL EL2 refers to BL	OLS-based	0	BL	BL
				1	BL, EL1	EL1 (2)
				2	BL, EL2	EL2 (2)
			Layer-based	0	BL	BL
				1	BL, EL1, EL2	BL, EL1 or EL2

(1) All output video layers within the OLS are decoded simultaneously.

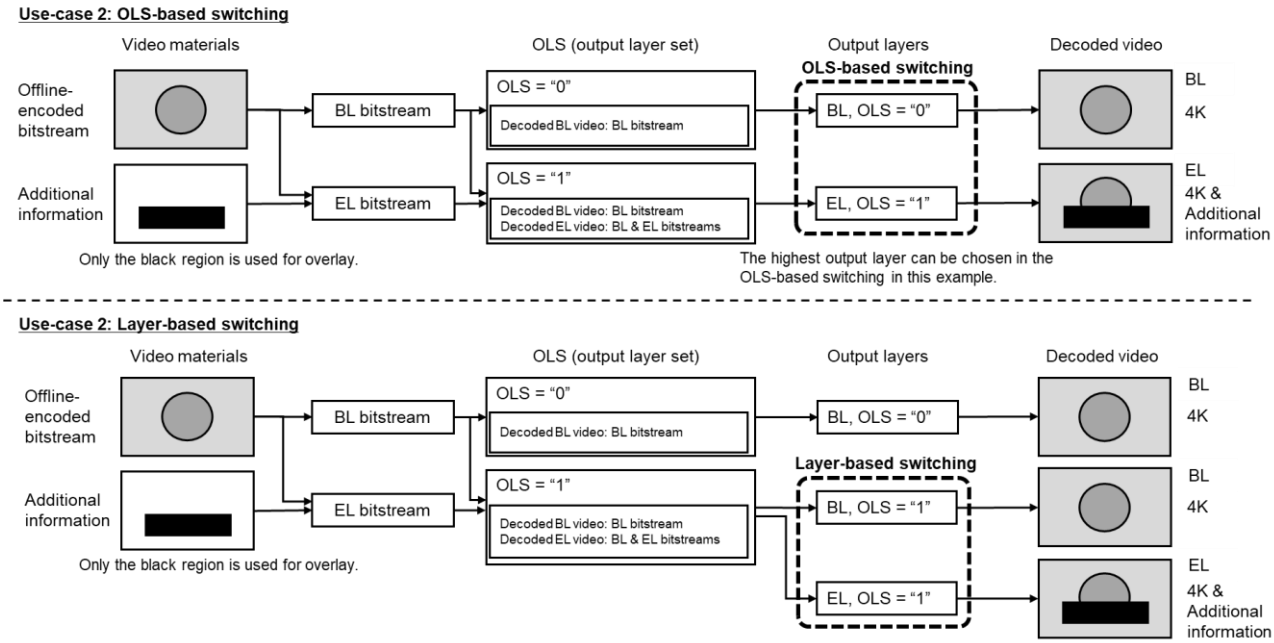
(2) The highest output layer can be chosen in the OLS-based switching in this example.

FIGURE 6

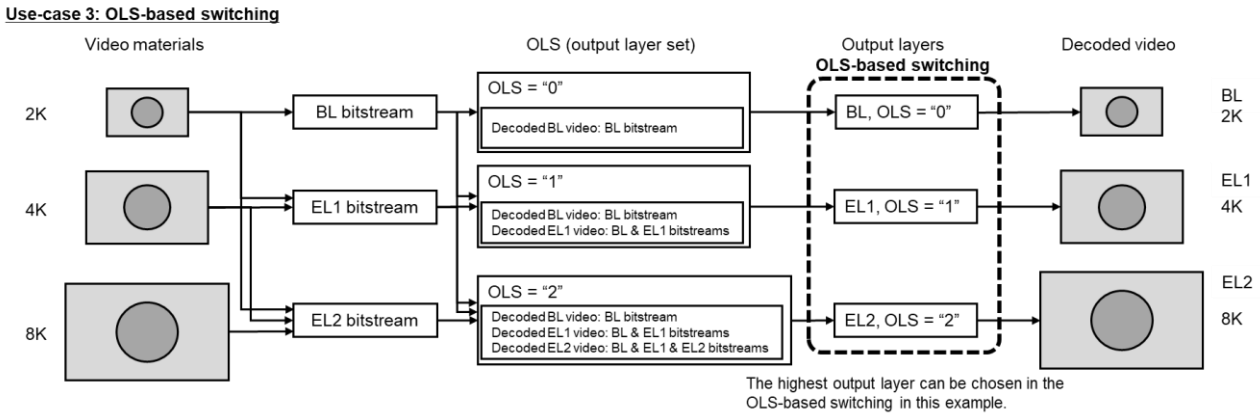
(a) OLS-based output switching in use Case 1



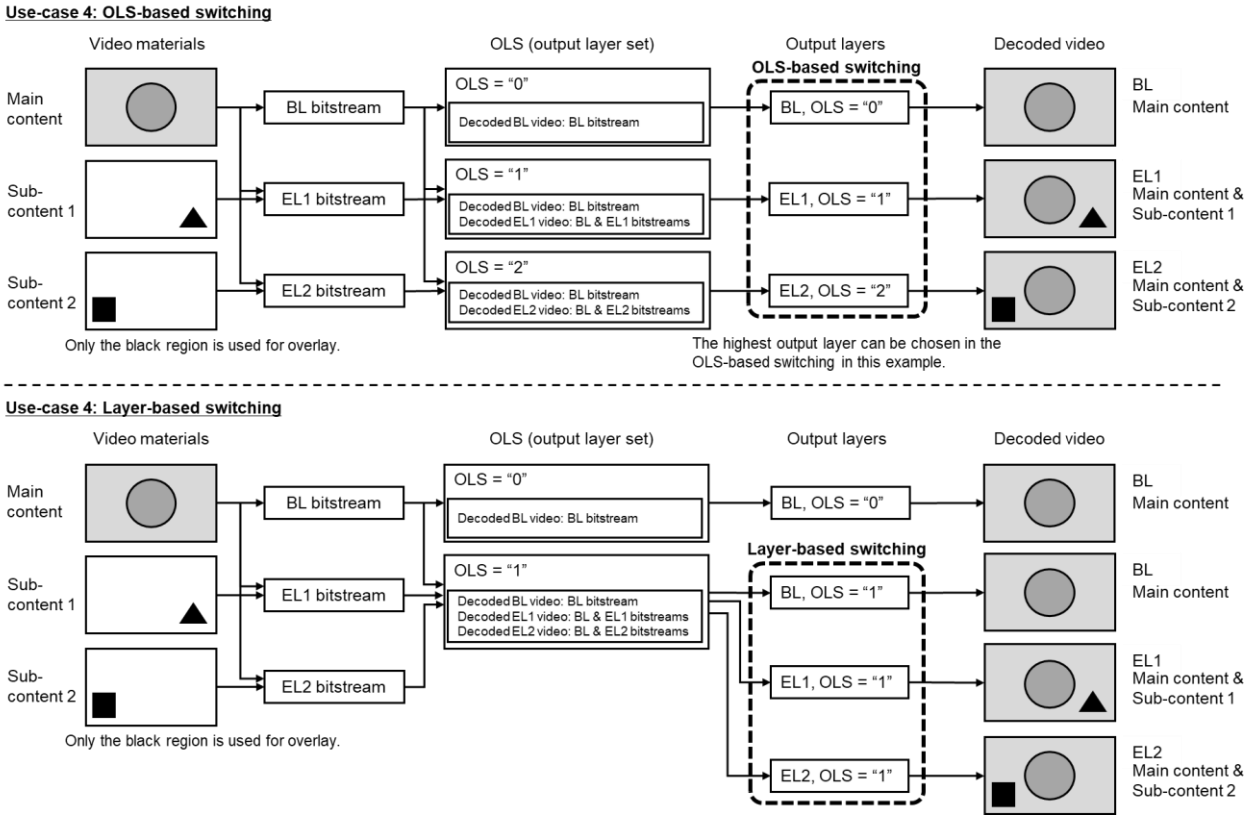
(b) OLS-based output switching and layer-based output switching in use Case 2



(c) OLS-based output switching in use Case 3



(d) OLS-based output switching and layer-based output switching in use Case 4



5 Signalling of video layer information in multiplexing layer

This section describes the signalling of video layer information in multiplexing layer.

In the multiplexing layer, a video preset ID is assigned to each video layer to identify content that can be selected by the viewer. The video preset ID is necessary for the receiver to uniquely determine which video layer to display when a bitstream contains multiple output layers.

Table 3 shows an example of video preset ID assignment for the use cases. As described in § 4, in use cases 1 and 3, where the video content across layers is identical except for resolution, the displayed layer is assumed to be fixed. Therefore, the video preset ID is set to 0 for all layers. However, it is also possible to assign different video preset IDs if the intention is to treat these contents as distinct. In contrast, in the use cases 2 and 4, each layer contains different content, allowing the viewer to actively select which layer to watch. Accordingly, different video preset IDs are assigned to each layer.

Furthermore, by defining audio preset IDs, as detailed in § 6, which assign selectable audio content to each video content, it becomes possible to switch audio content in association with the selected video content.

TABLE 3

Video layer information in multiplexing layer in use cases described in Section 4

Use case		Video layer information	
		Output video layers	Video preset ID
1	Optimization of frequency resource utilization	BL	0
		EL	
2	Overlaying live content on top of offline-encoded video	BL	0
		EL	1
3	Premium content distribution via the Internet (3 layers)	BL	0
		EL1	
		EL2	
4	Sub-content delivery based on viewer needs (3 layers)	BL	0
		EL1	1
		EL2	2

6 Concept of association between video and audio presentations

This section describes the concept of associating video and audio presentations for the use case of "sub-content delivery based on viewers' demands", in which viewers can select their preferred view from videos captured from various viewpoints, as described in § 3.4. In such use cases, it is desirable to provide additional audio content options associated with sub-video content.

MPEG-H 3D audio (3DA) and AC-4 support advanced sound systems (AdvSS) and can deliver multiple audio objects, including dialogues and background sound, within a single audio stream. 3DA uses the concept of "preset" (referred to as "presentation" in AC-4) to identify combinations of audio objects to be played simultaneously, similar to the video preset IDs described in §§ 4 and 5.

In the use Case 4, as shown in Table 2 and Fig. 6 (d), three types of video presentations are provided:

– Main video content.
– Main video content and Sub-video content 1.
– Main video content and Sub-video content 2.

These are signalled by video preset IDs 0, 1 and 2, respectively as shown in Table 3.

Similarly, an AdvSS stream can provide multiple types of audio presentations, for example,

– Main audio content.
– Sub-audio content not linked to specific video content.
– Sub-audio content linked to specific video content.

The concept of association between video and audio presentations involves defining available combinations of video and audio presets in the multiplexing layer (e.g. via a descriptor). Table 4 and Fig. 7 show an example of such combinations. An audio preset ID identifies audio content including commentary and its association with video content signalled by video present IDs. For example, a video content may have a default audio content. Additionally, non-default audio content may be selectable for specific video content. Conversely, some audio content may not be selectable for certain video content. Such relationships are explicitly signalled.

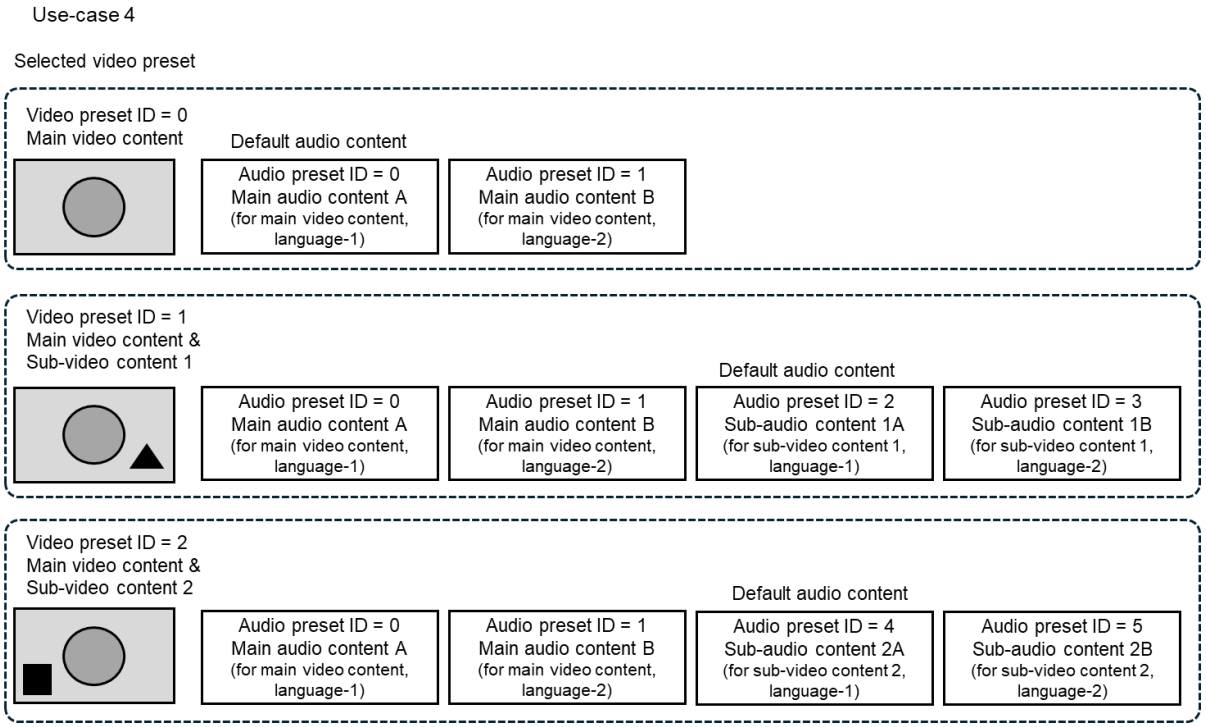
TABLE 4

Example of combinations of video and audio presets in use Case 4

ID	Audio content		Relationship between video and audio presets		
		Commentary	Video preset 0 (Main video content)	Video preset 1 (Main video content + Sub-video content 1)	Video preset 2 (Main video content + Sub-video content 2)
0	Main audio content A	For Main video content in language-1	Default	Selectable	Selectable
1	Main audio content B	For Main video content in language-2	Selectable	Selectable	Selectable
2	Sub-audio content 1A	For Sub-video content 1 in language-1	Non-selectable	Default	Non-selectable
3	Sub-audio content 1B	For Sub-video content 1 in language-2	Non-selectable	Selectable	Non-selectable
4	Sub-audio content 2A	For Sub-video content 2 in language-1	Non-selectable	Non-selectable	Default
5	Sub-audio content 2B	For Sub-video content 2 in language-2	Non-selectable	Non-selectable	Selectable

FIGURE 7

Example of combinations of video and audio presets in use Case 4



In this example:

– When the user selects video preset ID = 0 (Main video content), audio preset ID = 0 (Main audio content A, commentary for Main video content in language-1) is the default, but the user may select audio preset ID = 1 (Main audio content B, commentary for Main video content in language-2).

– When the user selects video preset ID = 1 (Main video content and Sub-video content 1), audio preset ID = 2 (Sub-audio content 1A, commentary for Sub-video content 1 in language-1) becomes the default, with optional selections of audio preset ID = 0, 1 or 3.

– When the user selects video preset ID = 2 (Main video content and Sub-video content 2), audio preset ID = 4 (Sub-audio content 2A, commentary for Sub-video content 2 in language-1) becomes the default, with optional selections of audio preset ID = 0, 1 or 5.