

INTERNATIONAL  
STANDARD

ISO/IEC  
23000-19

First edition  
2018-01

---

---

**Information technology — Multimedia  
application format (MPEG-A) —**

**Part 19:  
Common media application format  
(CMAF) for segmented media**

*Technologies de l'information — Format pour application multimédia  
(MPEG-A) —*

*Partie 19: Format CMAF (Common Media Application Format) pour  
médias segmentés*

STANDARDISO.COM : Click to view the full PDF of ISO/IEC 23000-19:2018



Reference number  
ISO/IEC 23000-19:2018(E)

© ISO/IEC 2018



**COPYRIGHT PROTECTED DOCUMENT**

© ISO/IEC 2018

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office  
CP 401 • Ch. de Blandonnet 8  
CH-1214 Vernier, Geneva, Switzerland  
Tel. +41 22 749 01 11  
Fax +41 22 749 09 47  
copyright@iso.org  
www.iso.org

Published in Switzerland

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 23000-19:2018

# Contents

	Page
<b>Foreword</b>	<b>vi</b>
<b>Introduction</b>	<b>vii</b>
<b>1 Scope</b>	<b>1</b>
<b>2 Normative references</b>	<b>1</b>
<b>3 Terms and definitions</b>	<b>2</b>
<b>4 Abbreviated terms</b>	<b>6</b>
<b>5 Document organization</b>	<b>6</b>
<b>6 CMAF hypothetical application model, media object model, and profiles</b>	<b>7</b>
6.1 Overview of the hypothetical application model and media object model	7
6.2 CMAF content processing model	9
6.3 Late binding CMAF track synchronization	10
6.4 Adaptive switching of CMAF tracks in CMAF switching sets	11
6.5 CMAF specified objects and profiles	12
6.5.1 Object derivation and interoperability code points	12
6.5.2 Encoded media objects	12
6.5.3 Logical media object sets	12
6.5.4 Addressable media objects	12
6.5.5 CMAF profiles, brand, and identifiers	13
6.6 CMAF media object model	14
6.6.1 CMAF fragments	14
6.6.2 CMAF tracks	15
6.6.3 CMAF track files	15
6.6.4 CMAF segments	16
6.6.5 CMAF chunks	16
6.6.6 CMAF switching sets and adaptive switching	17
6.6.7 CMAF selection sets and late binding	20
6.6.8 CMAF presentation timing model	21
6.6.9 Manifest information	23
6.6.10 CMAF addressable media objects, resources, and resource identifiers	24
<b>7 CMAF track format</b>	<b>24</b>
7.1 Overview	24
7.2 CMAF brands	24
7.3 CMAF media objects	25
7.3.1 CMAF boxes	25
7.3.2 CMAF track media objects	28
7.3.3 CMAF addressable media objects	32
7.3.4 CMAF switching sets	34
7.3.5 CMAF selection sets	37
7.3.6 CMAF presentations	38
7.4 Additional boxes, not defined in the ISO Base Media File Format	38
7.4.1 Track Encryption Box ('tenc')	38
7.4.2 Sample Encryption Box ('senc')	39
7.4.3 Protection System Specific Header Box ('pssh')	39
7.4.4 Media profile specific boxes	39
7.4.5 Event Message Box ('emsg')	39
7.5 Constraints on ISO Base Media File Format boxes	40
7.5.1 Movie Header Box ('mvhd')	40
7.5.2 Metadata Boxes	40
7.5.3 Kind Box ('kind')	40
7.5.4 Track Header Box ('tkhd')	40
7.5.5 Media Header Box ('mdhd')	41
7.5.6 Video Media Header Box ('vmhd')	41

7.5.7	Sound Media Header Box ('smhd')	41
7.5.8	Subtitle Media Header Box ('sthd')	41
7.5.9	Data Reference Box ('dref')	42
7.5.10	Sample Description Box ('stsd')	42
7.5.11	Protection Scheme Information Box ('sinf')	42
7.5.12	Track contained media sample information boxes	42
7.5.13	Edit List Box ('elst')	43
7.5.14	Track Extends Box ('trex')	43
7.5.15	Movie Fragment Header Box ('mfhd')	44
7.5.16	Track Fragment Header Box ('tfhd')	44
7.5.17	Track Run Box ('trun')	44
7.5.18	Sample Group Description Box ('sgpd')	45
7.5.19	Media Data Box ('mdat')	45
7.5.20	Sub-sample Information Box ('subs')	45
<b>8</b>	<b>Common Encryption of CMAF tracks</b>	<b>45</b>
8.1	Multiple DRM system support	45
8.2	Track encryption	46
8.2.1	General requirements	46
8.2.2	CMAF track constraints	47
8.2.3	Encryption constraints	48
8.2.4	CMAF presentation encryption	49
<b>9</b>	<b>Video CMAF tracks</b>	<b>49</b>
9.1	Overview	49
9.2	General video CMAF track format	50
9.2.1	General video CMAF track structure and constraints	50
9.2.2	Video Media Header ('vmhd')	50
9.2.3	Track Header Box ('tkhd')	51
9.2.4	Sample Description Box ('stsd')	51
9.2.5	Video CMAF fragment presentation time	52
9.2.6	Video media sample dependencies	52
9.2.7	Video edit lists	52
9.2.8	General video CMAF fragment random access constraints	52
9.2.9	Additional random access pictures within CMAF video fragments	53
9.2.10	Image framing and encoding constraints	53
9.2.11	General video CMAF switching set constraints	53
9.3	NAL structured video CMAF tracks	55
9.3.1	Overview	55
9.3.2	CMAF track format constraints for NAL structured video	55
9.3.3	NAL structured video Access Units contained in media samples	56
9.3.4	NAL structured video coding sequences corresponding to CMAF fragments	56
9.3.5	Elementary stream constraints	57
9.3.6	General CMAF switching set constraints for NAL structured video	57
9.3.7	Single initialization CMAF switching set constraints for NAL structured video tracks and media profiles	57
9.4	AVC video CMAF tracks	58
9.4.1	Storage of AVC elementary streams	58
9.4.2	Constraints on AVC elementary streams	59
9.5	AVC video Internet Media Type parameters	61
9.5.1	AVC signalling of "codecs" parameters	61
<b>10</b>	<b>Audio CMAF tracks</b>	<b>61</b>
10.1	Overview	61
10.2	General audio CMAF track format	61
10.2.1	Derivation	61
10.2.2	Track Header Box ('tkhd')	61
10.2.3	Sound Media Header Box ('smhd')	62
10.2.4	Sample Description Box ('stsd')	62
10.2.5	AudioSampleEntry	62

10.2.6	Audio offset edit list .....	62
10.3	AAC audio CMAF tracks .....	62
10.3.1	Overview .....	62
10.3.2	"codecs" parameter signalling .....	62
10.3.3	Considerations for AAC audio encoding .....	63
10.3.4	AAC track constraints .....	64
10.3.5	AAC elementary stream constraints .....	65
10.4	AAC core audio CMAF media profile .....	66
10.5	AAC adaptive switching audio CMAF media profile .....	67
10.5.1	General constraints .....	67
10.5.2	CMAF fragment encoding constraints .....	67
10.5.3	General considerations and requirements .....	67
10.5.4	Constraints for AAC-LC .....	68
10.5.5	Constraints for HE-AAC .....	68
10.5.6	Constraints for HE-AACv2 .....	69
<b>11</b>	<b>Subtitles and captions .....</b>	<b>70</b>
11.1	Overview .....	70
11.2	WebVTT .....	70
11.3	IMSC text and image tracks .....	71
11.3.1	General .....	71
11.3.2	Common constraints .....	71
11.3.3	IMSC1 text track constraints .....	71
11.3.4	IMSC1 image track constraints .....	72
11.4	CTA-608 and CTA-708 .....	72
11.5	Metadata for subtitles .....	72
<b>12</b>	<b>CMAF media profiles and CMAF presentation profiles .....</b>	<b>73</b>
12.1	CMAF media profiles .....	73
12.1.1	General guidelines for specifying CMAF media profiles .....	73
12.1.2	Guidelines for audio CMAF media profiles .....	74
12.1.3	Guidelines for video CMAF media profiles .....	74
12.2	CMAF presentation profiles .....	75
12.2.1	General .....	75
12.2.2	CMAF profile conformance .....	75
<b>Annex A</b> (normative) <b>CMAF presentation profiles and media profiles .....</b>	<b>78</b>	
<b>Annex B</b> (normative) <b>HEVC video CMAF track format and CMAF media profiles .....</b>	<b>82</b>	
<b>Annex C</b> (informative) <b>Subsampling of NAL structured video tracks in CMAF switching sets .....</b>	<b>88</b>	
<b>Annex D</b> (informative) <b>Hypothetical player model .....</b>	<b>98</b>	
<b>Annex E</b> (informative) <b>Event messages .....</b>	<b>101</b>	
<b>Annex F</b> (informative) <b>Error handling for missing media .....</b>	<b>102</b>	
<b>Annex G</b> (informative) <b>Recommendations for AAC CMAF switching set encoding .....</b>	<b>103</b>	
<b>Bibliography .....</b>	<b>106</b>	

## Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see [www.iso.org/directives](http://www.iso.org/directives)).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see [www.iso.org/patents](http://www.iso.org/patents)).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: [www.iso.org/iso/foreword.html](http://www.iso.org/iso/foreword.html).

This document was prepared by Technical Committee ISO/IEC JTC 1, *Information technology, SC 29, Coding of audio, picture, multimedia and hypermedia information*.

A list of all parts in the ISO/IEC 23000 series can be found on the ISO website.

## Introduction

Common Media Application Format (CMAF) combines and constrains several MPEG specifications to define a multimedia format that is optimized for delivery of a single adaptive multimedia presentation to a variety of devices, using a variety of adaptive streaming, broadcast, download, and storage methods.

Several MPEG specifications have been adopted for much of the video delivered over the Internet and other IP networks (cellular, cable, broadcast, etc.). Various organizations have taken MPEG's core coding, file format and system standards and combined them into their own specifications for their specific application. While these specifications are similar, their differences result in unnecessary duplication of engineering effort and duplication of identical content in slightly different formats, which results in increased storage and delivery costs.

CMAF provides a common media specification that application specifications, such as MPEG Dynamic Adaptive Streaming over HTTP (DASH), can reference and a common media format that allows a single encoded multimedia presentation to be used by many applications.

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 23000-19:2018

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 23000-19:2018

# Information technology — Multimedia application format (MPEG-A) —

## Part 19: Common media application format (CMAF) for segmented media

### 1 Scope

This document specifies the CMAF multimedia format, which contains segmented media objects optimized for streaming delivery and decoding on end user devices in adaptive multimedia presentations.

CMAF specifies a track format derived from the ISO Base Media File Format, then derives addressable media objects from CMAF tracks that can be used for storage and delivery.

CMAF specifies sets of tracks that share encoding and packaging constraints that enable the selection of multiple tracks to form a multimedia presentation and allow seamless switching of alternative encodings of the same content at different bit rates, frame rates, resolution, etc.

CMAF specifies a hypothetical application model that determines how tracks in a CMAF presentation are intended to be combined and synchronized to form a multimedia presentation. The model abstracts delivery to allow any delivery method. The hypothetical application model assumes a manifest and player, but CMAF does not specify a manifest, player, or delivery protocol, with the intent that any that support the hypothetical application model can be used.

CMAF specifies media profiles and brands that constrain media encoding and packaging of CMAF tracks to enable seamless adaptive switching of tracks and allow devices to identify compatible content by its brand.

CMAF specifies presentation profiles that conditionally require sets of CMAF tracks conforming to specified media profiles and allow content creators and devices to identify compatible multimedia presentations.

CMAF enables extensibility by specifying how new media profiles and presentation profiles can be specified and identified and includes guidelines for those specifications.

### 2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 14496-1, *Information technology — Coding of audio-visual objects — Part 1: Systems*

ISO/IEC 14496-3, *Information technology — Coding of audio-visual objects — Part 3: Audio*

ISO/IEC 14496-10, *Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding*

ISO/IEC 14496-12, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

ISO/IEC 14496-14, *Information technology — Coding of audio-visual objects — Part 14: MP4 file format*

ISO/IEC 14496-15, *Information technology — Coding of audio-visual objects — Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format*

ISO/IEC 14496-30, *Information technology — Coding of audio-visual objects — Part 30: Timed text and other visual overlays in ISO base media file format*

ISO/IEC 23001-7, *Information technology — MPEG systems technologies — Part 7: Common encryption in ISO base media file format files*

ISO/IEC 23008-2, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding*

ISO/IEC 23009-1, *Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats*

IETF RFC 5234<sup>1)</sup>, *Augmented BNF for Syntax Specifications: ABNF*

IETF RFC 6381, *The 'Codecs' and 'Profiles' Parameters for "Bucket" Media Types*

ITU-R Recommendation BT.709, *Parameter values for the HDTV standards for production and international programme exchange*

ITU-R Recommendation BT.1886, *Reference electro-optical transfer function for flat panel displays used in HDTV studio production*

ITU-R Recommendation BT.2020<sup>2)</sup>, *Parameter values for ultra-high definition television systems for production and international programme exchange*

ITU-R Recommendation BT.2035, *A reference viewing environment for evaluation of HDTV program material or completed programmes*

ITU-R Recommendation BT.2100-0:2016<sup>3)</sup>, *Image parameter values for high dynamic range television for use in production and international programme exchange*

ITU-T Recommendation X.667:2014<sup>4)</sup>, *Information technology — Open Systems Interconnection — Procedures for the operation of OSI Registration Authorities: Generation and registration of Universally Unique Identifiers (UUIDs) and their use as ASN.1 object identifier components*

ANSI/CTA-608-E R-2014<sup>5)</sup>, *Line 21 Data Services*

ANSII/CTA-708-E<sup>6)</sup>, *Digital Television (DTV) Closed Captioning*

W3C<sup>7)</sup>, *TTML Profiles for Internet Media Subtitles and Captions 1.0 (W3C IMSC1)*

W3C<sup>8)</sup>, *TTML Media Type Definition and Profile Registry, W3C Working Group Note (W3C TTML Registry)*

### 3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

- 1) Available at <https://tools.ietf.org/html/rfc5234>
- 2) Available at <http://www.itu.int/rec/R-REC-BT.2020/en>
- 3) Available at [https://www.itu.int/dms\\_pubrec/itu-r/rec/bt/R-REC-BT.2100-0-201607-I!!PDF-E.pdf](https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.2100-0-201607-I!!PDF-E.pdf)
- 4) Available at <https://www.itu.int/rec/T-REC-X.667>
- 5) Available at [http://www.techstreet.com/standards/cta-608-e-r2014?product\\_id=1815447](http://www.techstreet.com/standards/cta-608-e-r2014?product_id=1815447)
- 6) Available at [http://www.techstreet.com/standards/cta-708-e?product\\_id=1860354](http://www.techstreet.com/standards/cta-708-e?product_id=1860354)
- 7) Available at <http://www.w3.org/TR/ttml-imsc1>
- 8) Available at <https://www.w3.org/TR/ttml-profile-registry>

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

- IEC Electropedia: available at <http://www.electropedia.org/>
- ISO Online browsing platform: available at <http://www.iso.org/obp>

### 3.1 Media objects

#### 3.1.1

##### **CMAF fragment**

encoded ISOBMFF media segment conforming to CMAF constraints

#### 3.1.2

##### **CMAF header**

sequence of CMAF constrained ISOBMFF boxes that do not reference any *media samples* (3.3.15), but are associated with a *CMAF track* (3.2.1) and necessary for the decoding of its *CMAF fragments* (3.1.1)

#### 3.1.3

##### **CMAF addressable media object**

CMAF media object packaged for storage or delivery

Note 1 to entry: Examples include a *CMAF track file* (3.1.6) containing a *CMAF header* (3.1.2) and *CMAF fragments* (3.1.1), or a *CMAF segment* (3.1.5) containing one or more CMAF fragments, or a *CMAF chunk* (3.1.4) containing a partial sequence of the *media samples* (3.3.15) of a CMAF fragment.

#### 3.1.4

##### **CMAF chunk**

CMAF media object that contains a consecutive subset of the media samples (3.3.15) of a *CMAF fragment* (3.1.1), where only the first CMAF chunk of a CMAF fragment is constrained to be an adaptive switching (3.3.9) point

#### 3.1.5

##### **CMAF segment**

*CMAF addressable media object* (3.1.3) consisting of one or more consecutive *CMAF fragments* (3.1.1) from the same *CMAF track* (3.2.1)

Note 1 to entry: A “CMAF segment” is conformant to an “ISOBMFF segment” and a “DASH segment”.

#### 3.1.6

##### **CMAF track file**

one *CMAF track* (3.2.1) stored consecutively in a single ISOBMFF file with the earliest *CMAF fragment* (3.1.1) constrained to start at decode time zero

### 3.2 Logical structure

#### 3.2.1

##### **CMAF track**

sequence of *CMAF fragments* (3.1.1) that are consecutive in presentation time, contain one media stream, conform to the 'cmfc' brand, including an associated *CMAF header* (3.1.2) that can initialize playback

#### 3.2.2

##### **CMAF switching set**

set of one or more *CMAF tracks* (3.2.1), where each track is an alternative encoding of the same source content, and are constrained to enable seamless track switching (3.3.9)

#### 3.2.3

##### **aligned CMAF switching set**

set of *CMAF switching sets* (3.2.2), the *CMAF tracks* (3.2.1) of which all contain alternative encodings of the same source content in time-aligned *CMAF fragments* (3.1.1), but all CMAF tracks do not conform to a single CMAF switching set

### 3.2.4

#### **CMAF selection set**

set of one or more *CMAF switching sets* (3.2.2), where each CMAF switching set encodes an alternative aspect of the same presentation over the same time period, only one of which is intended to be played at a time, e.g. an alternative language or codec

### 3.2.5

#### **CMAF presentation**

set of one or more *CMAF selection sets* (3.2.4) that can be simultaneously decoded to produce a multimedia user experience, potentially including synchronized audio, video, and subtitles

### 3.2.6

#### **CMAF media profile**

encoding constraint on a *CMAF track* (3.2.1) and its contained *media samples* (3.3.15) associated with a CMAF compatibility brand

### 3.2.7

#### **CMAF presentation profile**

requirement on the *CMAF media profiles* (3.2.6) contained in a *CMAF presentation* (3.2.5)

### 3.2.8

#### **required media profile**

*CMAF media profile* (3.2.6) conditionally required by a *CMAF presentation profile* (3.2.7)

### 3.2.9

#### **manifest**

document describing one or more *CMAF presentations* (3.2.5)

Note 1 to entry: Manifest formats are not specified in this document.

## 3.3 Application model

### 3.3.1

#### **CMAF hypothetical application model**

*CMAF presentation* (3.2.5) application model based on *late binding* (3.3.3) and synchronization of *CMAF tracks* (3.2.1) that partly determines the CMAF track encoding constraints necessary for an intended CMAF presentation

### 3.3.2

#### **player**

component of the *CMAF hypothetical application model* (3.3.1) responsible for interpreting a *manifest* (3.2.9), requesting resources, and rendering a *CMAF presentation* (3.2.5)

### 3.3.3

#### **late binding**

*selection* (3.3.8) and synchronization of separately stored *CMAF tracks* (3.2.1) by a *player* (3.3.2) resulting in a synchronized multimedia presentation

### 3.3.4

#### **CMAF presentation timeline**

timeline shared by all *CMAF tracks* (3.2.1) in a *CMAF presentation* (3.2.5), starting at CMAF presentation time zero, which is coincident with the earliest *media samples* (3.3.15) intended for presentation

### 3.3.5

#### **presentation time offset**

earliest presentation time of each *CMAF track* (3.2.1) at the start of a *CMAF presentation* (3.2.5)

Note 1 to entry: Presentation time offset is an encoded property of tracks in a presentation, but it can also refer to that value stored in a *manifest* (3.2.9).

**3.3.6****CMAF fragment duration**

sum of the *media sample* (3.3.15) durations documented in the *TrackFragmentRunBox* of all *MovieFragmentHeaderBoxes* in the *CMAF fragment* (3.1.1)

**3.3.7****CMAF presentation duration**

sum of the *CMAF fragment durations* (3.3.6) of the longest *CMAF track* (3.2.1) in a *CMAF presentation* (3.2.5), starting from its earliest presentation time on the *CMAF presentation timeline* (3.3.4)

**3.3.8****selection**

choice of a *CMAF track* (3.2.1) from alternatives in a selection set (e.g. selecting an audio track by language), possibly by user action or stored user preference

**3.3.9****switching**

changing to a different *CMAF track* (3.2.1) during presentation, including adaptively switching between *CMAF fragments* (3.1.1) in a *CMAF switching set* (3.2.2)

**3.3.10****seamless switching**

*switching* (3.3.9) between *CMAF tracks* (3.2.1) without interrupting presentation of the media content, i.e. decoding *media samples* (3.3.15), at the same time and quality as though their containing *CMAF track* was decoded without switching

**3.3.11****CMAF switching set constraints**

*CMAF media profile* (3.2.6) constraints that enable seamless switching (3.3.9) between *CMAF tracks* (3.2.1) in a *CMAF switching set* (3.2.2) conforming to that media profile

**3.3.12****single initialization CMAF switching set constraints**

additional *CMAF switching set constraints* (3.3.11) so *CMAF fragments* (3.1.1) do not depend on a different *CMAF header* (3.1.2) when switching (3.3.9)

**3.3.13****resource identifier**

externally specified identifier that identifies a *CMAF addressable media object* (3.1.3)

Note 1 to entry: An example is a URI or other object identifier specified by a delivery protocol and *manifest* (3.2.9).

**3.3.14****stream access point**

*media sample* (3.3.15) random access property, numbered as in ISO/IEC 14496-12:2015, Annex I

**3.3.15****media sample**

*media data* in a *CMAF fragment* (3.1.1) associated with a single decode start time and duration

Note 1 to entry: The term “sample” is often used in the context of video to refer to the spatial samples of an image and in the context of audio to refer to PCM waveform samples. In this document, each type of sample is identified by a defined term. A media sample defined by ISOBMFF is always identified by the term “media sample”. The word “sample” is frequently used in ISOBMFF to refer to objects and parameters such as a “sample entry”, “sample size”, etc., and those terms are used without modification in this document.

**3.3.16****audio PCM sample**

digital sample quantizing the amplitude of an audio waveform at regular and frequent intervals, e.g. 48 kHz

## 3.3.17

**video spatial sample**

quantized values representing the colour and brightness of an area of an image corresponding to a two-dimensional spatial tessellation of the image

## 3.3.18

**subsampling**

video encoding using a smaller number of *video spatial samples* (3.3.17) than the source video, that number being an integer submultiple that can be scaled to the source video size based on video stream parameters without position shift or picture aspect ratio distortion

## 4 Abbreviated terms

The following abbreviated terms are used in this document.

AU	Access Unit
CDN	Content Delivery Network
CMAF	Common Media Application Format
CVS	Coded Video Sequence [A sequence of media samples (coded video frames), starting with a SAP type 1 or 2, and including all media samples prior to the next SAP type 1 or 2 in decoding order.]
DASH	Dynamic Adaptive Streaming over HTTP
ISOBMFF	ISO Base Media File Format, defined in ISO/IEC 14496-12
KID	Key Identifier as defined in ISO/IEC 23001-7
NAL	Network Adaptation Layer
PCM	Pulse Code Modulation
PPS	Picture Parameter Set
SAP	Stream Access Point as defined in ISO/IEC 14496-12
SEI	Supplemental Enhancement Information
SPS	Sequence Parameter Set
VCL	Video Coding Layer
VPS	Video Parameter Set
VUI	Video Usability Information

## 5 Document organization

First-time readers of this document are advised to start with [Clause 6](#) for a description of the objects and terminology specified, the CMAF object model, and the hypothetical application model, which defines how these objects can be combined to form adaptive multimedia presentations.

The normative specifications in [Clause 7](#) through [Clause 12](#) are terse to facilitate development and testing and assume an understanding of [Clause 6](#). [Clause 7](#) specifies ISO Base Media File Format boxes and structures such as movie fragments and tracks that are used to construct all CMAF media objects. [Clauses 8](#) through [11](#) contain details specific to encryption, audio, video, and subtitle tracks. [Clause 12](#) specifies the combination of CMAF tracks and media profiles into CMAF presentations. It also

recommends how to specify additional CMAF media profiles and presentation profiles, which can be specified by other documents and organizations.

CMAF presentation profiles and CMAF media profiles are specified in annexes to allow the addition of new profiles without changing the core document. Additional informative annexes have been added to provide explanations and recommendations on specific topics.

The following is a list of the main clauses of this document, with a brief description of each.

[Clause 6](#) describes the segmented media encoding and playback model using the media objects defined by the CMAF.

[Clause 7](#) describes the use of ISO Base Media File Format for the Common Media Application Format brand.

[Clause 8](#) describes how digital rights management information and encryption is applied to the Common Media Application Format.

[Clause 9](#) describes the general video track format, constraints for NAL structured video tracks, and the AVC video track format.

[Clause 10](#) describes the general audio track format and specifies two AAC audio CMAF media profiles.

[Clause 11](#) describes the subtitle track format, CMAF media profiles for WebVTT and IMSC1 TTML subtitles, and signalling of CTA 608/708 captions embedded in video streams.

[Clause 12](#) describes the general requirements for CMAF media profiles and CMAF presentation profiles.

[Annex A](#) describes several CMAF media profiles, their compatibility brands, and a CMAF presentation profile that conditionally requires some of those media profiles.

[Annex B](#) describes packaging and codec constraints for some CMAF media profiles using the HEVC video codec.

[Annex C](#) describes framing and encoding CMAF switching sets using subsampling and scaling of video to provide seamless playback with adaptive bit rate and scaling.

[Annex D](#) describes examples of player track selection, synchronization, and adaptive switching of a CMAF presentation.

[Annex E](#) describes the use of event messages attached to media objects to deliver metadata.

[Annex F](#) describes maintaining presentation timing and delivery in the event of missing media samples and resources.

[Annex G](#) describes encoding recommendations for AAC audio CMAF tracks conforming to adaptive CMAF switching sets.

## 6 CMAF hypothetical application model, media object model, and profiles

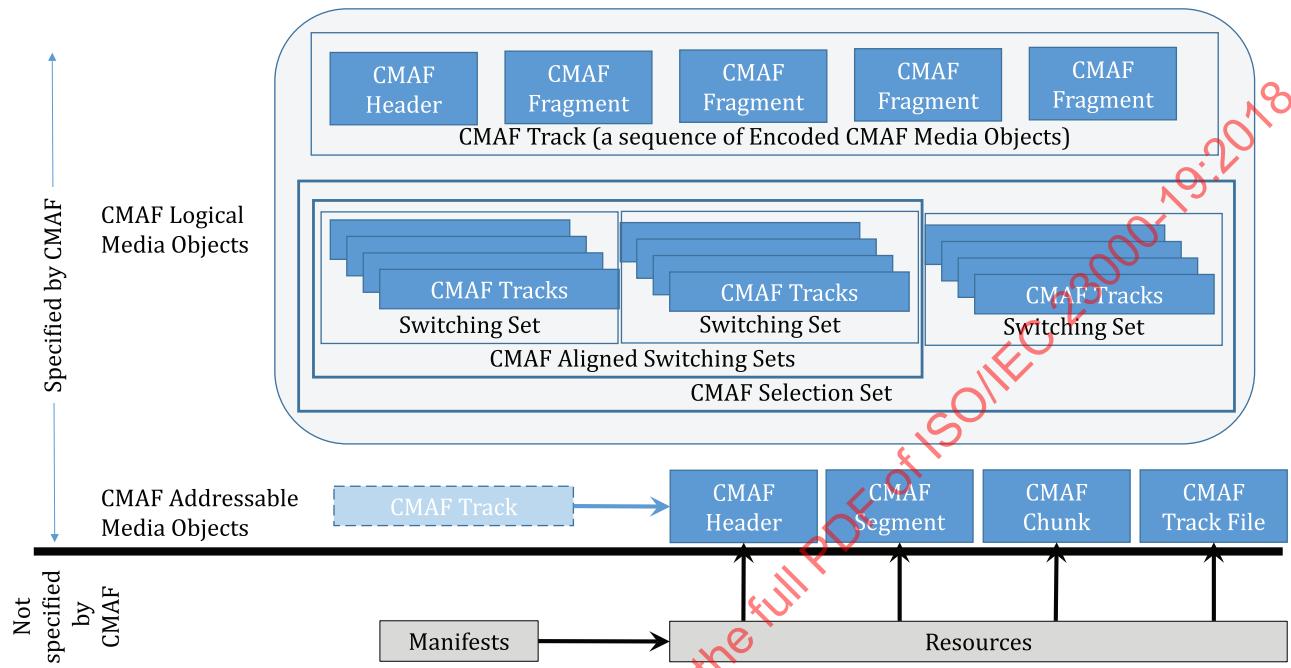
### 6.1 Overview of the hypothetical application model and media object model

CMAF defines a hypothetical application model so that encoding to that model results in consistent CMAF track encoding, representation in manifests, track selection, late binding, synchronization, decoding, and rendering of CMAF presentations.

Decoding requirements can be inferred from encoding constraints and the hypothetical application model, but are not directly specified by CMAF. CMAF does not specify manifest formats or associated resource identification and transport. However, CMAF does specify CMAF addressable media objects derived from encoded CMAF fragments, which can be referenced as resources by a manifest. External specifications can define how a manifest describes a CMAF presentation, including identifying CMAF

addressable media objects as resources and representing their logical relationships determined by the CMAF tracks, CMAF switching sets, CMAF selection sets, and CMAF presentations they are derived from.

**Figure 1** illustrates the media objects that are specified by CMAF, starting with the encoded CMAF fragments that form CMAF tracks, then logical CMAF track sets determined by CMAF track encoding constraints, then derived CMAF addressable media objects that can package encoded CMAF fragments or their media samples for storage and delivery.



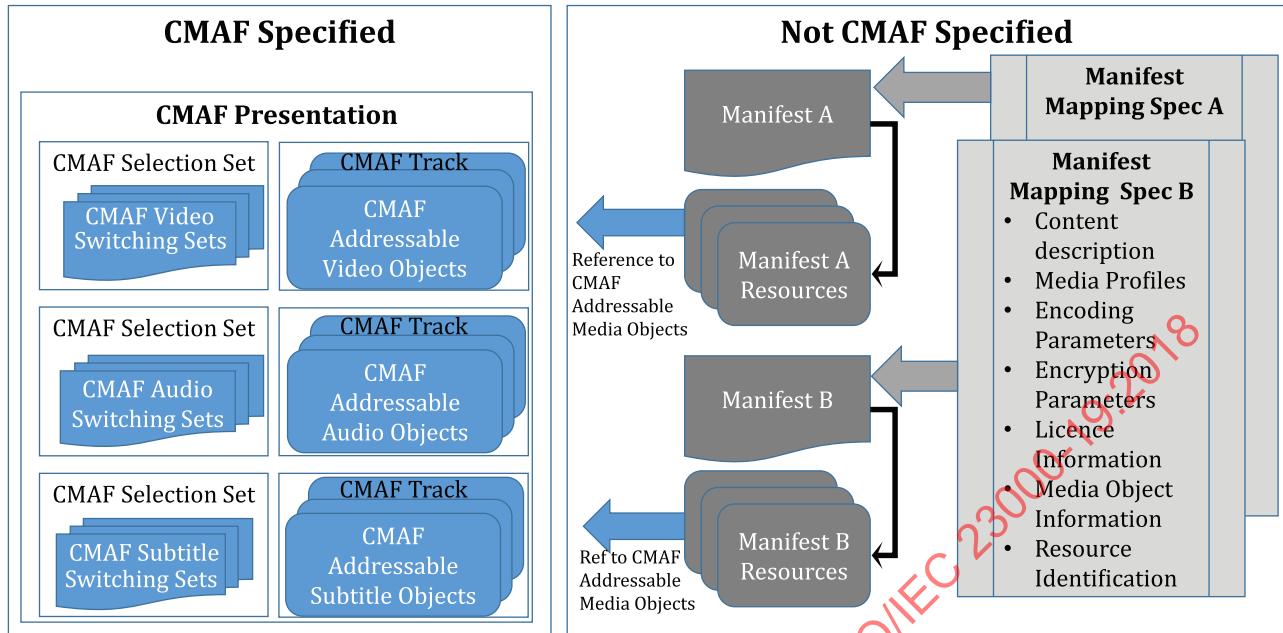
**Figure 1 — Media objects specified in CMAF and presented by externally specified applications, such as adaptive streaming**

**Figure 1** illustrates the mapping between CMAF specified CMAF presentations, and externally specified manifests and resources. Multiple manifests may reference the same CMAF presentation and CMAF addressable media objects. Specification of manifests and resource delivery is outside the scope of this document.

To accurately represent a CMAF presentation, a manifest will describe CMAF track relationships determined by each track's source content and CMAF track encoding constraints, e.g. that CMAF tracks belong to the same CMAF switching set, which belongs to a CMAF selection set. CMAF groups CMAF tracks based on their encoding constraints in logical media objects called CMAF selection sets and CMAF switching sets that also determine intended use in late binding, track selection, seamless switching, and synchronization. Additional CMAF track metadata such as CMAF media profile brands, "codecs" parameters, language fields, etc. can be included in manifests to enable adaptive track selection and playback, optimized for each user and device.

Manifests can reference CMAF addressable media objects by resource identifiers used by manifests and servers to select the identified CMAF addressable media objects for delivery and playback. Multiple CMAF addressable media object types are specified for different delivery use cases. Use cases include prerecorded content that is downloaded or streamed as files, and live and on demand adaptive streaming over the Internet. The size of the CMAF addressable media objects can be optimized for efficient download and CDN caching, or fast bit rate switching and low latency, depending on the application.

**Figure 2** illustrates the relationship between CMAF and streaming specifications that can define a mapping between their manifest and resource formats, and CMAF presentations and the CMAF addressable media objects they include.

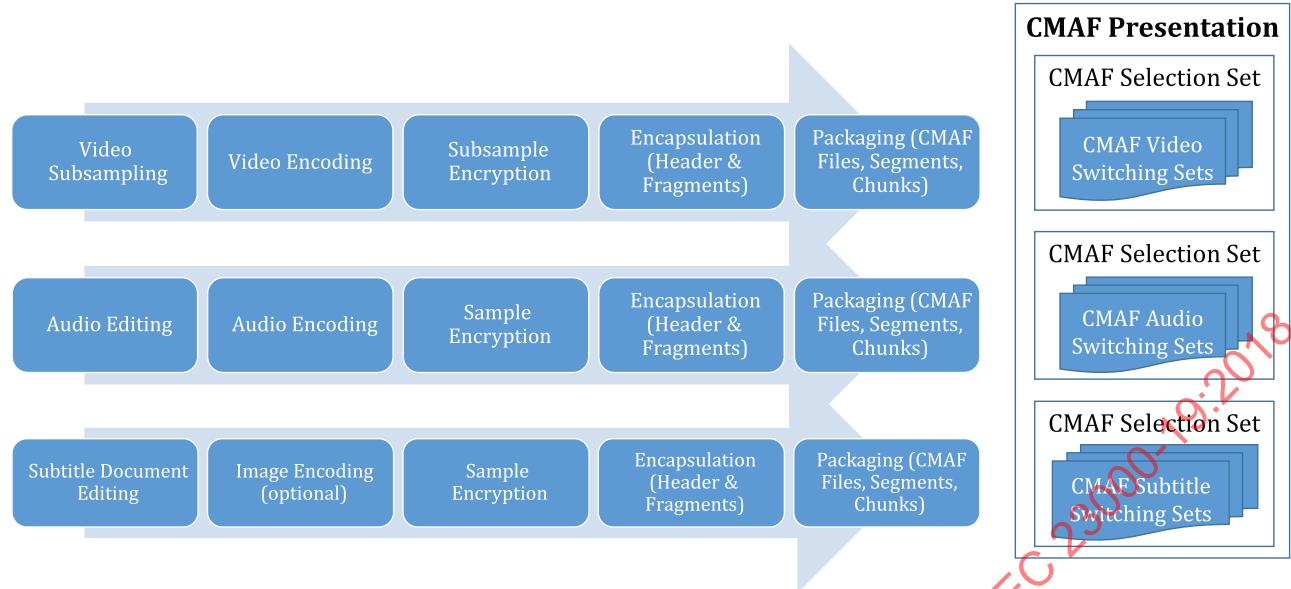


**Figure 2 — CMAF hypothetical application model using externally defined manifests that describe the CMAF presentations and media objects**

## 6.2 CMAF content processing model

The CMAF content processing model is shown in [Figure 3](#). Each CMAF presentation is composed of one or more media content components from different source material, for example, audio components in different languages, video components with different views of the same subject, subtitles with different languages or functions, etc. A typical presentation includes audio, video, and subtitles. Presentations are also possible that only include audio components, or include multiple audio components, or include multiple video sources, e.g. side-by-side video, picture in picture, sign language overlay, etc.

Synchronized CMAF tracks can be created and encoded at different times and synchronized on playback if they share a common presentation timeline. Content packaged with the appropriate CMAF fragment decode and presentation times can form a multimedia presentation conforming to the hypothetical application model. Each CMAF track can be encoded, encrypted, and packaged by an independent encoder if each encoder constrains timing, encoding and encryption parameters to the constraints of the intended CMAF presentation, CMAF selection set, and CMAF switching set. CMAF tracks in a CMAF switching set can be independently encoded by multiple encoders with the necessary CMAF fragment time alignment by deriving CMAF fragment timing, spatial and temporal subsampling, etc. from the parameters of the shared media source they encode from.



**Figure 3 — CMAF presentation content generation**

As illustrated in [Figure 3](#), video media components are often preprocessed (e.g. subsampled) and encoded in multiple media streams, typically different in bit rate and encoded resolution. Media streams can be encrypted and encapsulated into CMAF fragments, which can then be packaged as CMAF addressable media objects appropriate for one or more delivery methods. CMAF addressable media objects can be made available as resources and identified as defined by other media application specifications.

CMAF addressable media objects are derived from CMAF fragments and CMAF chunks as specified in [Clause 7](#), and inherit CMAF fragment constraints such as alignment of media sample sequences within a CMAF fragment, CMAF fragment time alignment within a CMAF switching set, CMAF track and CMAF media profile constraints, etc. This makes CMAF fragment encoding and decoding independent of the CMAF addressable media objects used to store and transport the CMAF fragments. But CMAF addressable media objects need not be constructed by packaging complete CMAF fragments as long as the resulting CMAF addressable media objects conform to CMAF fragment and CMAF track constraints.

For live streaming, a CMAF fragment in each CMAF track can be encoded simultaneously from each media content source component, and each CMAF fragment packaged and made available as soon as possible as a CMAF addressable media object, such as a CMAF segment or a CMAF chunk.

### 6.3 Late binding CMAF track synchronization

The feature of synchronizing separately stored CMAF tracks during playback is referred to as *late binding*. The late binding timing model and accurate recording of CMAF fragment presentation times during encoding are necessary to synchronize late bound CMAF tracks during delivery and presentation.

In the CMAF hypothetical application model, all CMAF tracks in a CMAF presentation share the same CMAF presentation timeline, which has a value of zero at the earliest media sample intended for presentation. All CMAF tracks in a CMAF presentation also share a common track decode timeline origin. A media sample of a CMAF track T1 that has a CMAF track presentation time of X is presented at the same time as the media sample of any other CMAF track T2 that also has the CMAF track presentation time X (the duration X having been divided by each track's timescale).

As defined in ISOBMFF, the presentation time of a media sample in a file is obtained from the decode time (derived from the `baseMediaDecodeTime` in the `TrackFragmentBaseMediaDecodeTime` `Box` and the duration of previous samples) and possibly from composition time offsets and edit lists that offset each track's presentation time on the file's movie presentation timeline, which starts at zero.

Equivalent decode and presentation times in different tracks can have different integer values if the tracks have different timescales.

CMAF tracks are not contained in a single multitrack ISOBMFF file with a shared movie timeline, but the CMAF presentation timeline provides an equivalent shared timeline and start point for all CMAF tracks in a CMAF presentation. If a CMAF presentation contains only CMAF track files, the presentation time of the first media sample of each CMAF track file is 0. However, in CMAF presentations using CMAF tracks, the presentation time of the earliest presented media sample can be a non-zero value, but is limited to an equivalent presentation time across all CMAF tracks. All CMAF tracks in a CMAF presentation are required to share a common timeline origin, but they may use different timescales, such as 90 kHz for video and 48 kHz for audio, so an equivalent decode or presentation time can have a different integer value stored in `baseMediaDecodeTime`.

A shared CMAF presentation timeline permits synchronization of different CMAF tracks at the CMAF player by, for example, starting presentation with the earliest media sample of the earliest video CMAF fragment and simultaneously presenting selected audio and subtitle CMAF tracks at their presentation time equivalent to the presentation time of the earliest video media sample. If an audio or subtitle CMAF fragment contains media samples with earlier presentation times than the earliest video media sample, the leading media samples are not expected to be presented. If a CMAF presentation contains no video, then the starting CMAF presentation time can be the earliest audio media sample.

A CMAF track's presentation time at the start of the CMAF presentation can be specified in a manifest presentation time offset so long as the equivalent offset is specified for all CMAF tracks in a CMAF presentation. For instance, each period in a DASH manifest is considered a CMAF presentation. The requirement that CMAF presentations share a common presentation timeline origin means that all presentation time offsets in a DASH Period will be equivalent values.

#### 6.4 Adaptive switching of CMAF tracks in CMAF switching sets

CMAF tracks are contained in CMAF switching sets. Only CMAF tracks encoded from the same media content component can belong to the same switching set, but CMAF tracks encoded from the same media source may also be grouped into different switching sets that each have different encoding constraints necessary for seamless adaptive switching.

CMAF tracks in a CMAF switching set have the following characteristics.

- The CMAF tracks are alternative encodings of a single media content component and media type, e.g. the same audio or video source.
- The CMAF tracks are perceptually equivalent, e.g. the same aspect ratio, colour space, duration, etc., typically resulting from encoding the same media source, i.e. master file or input stream.
- The CMAF tracks conform to a common CMAF media profile.
- The CMAF tracks are seamlessly switchable based on the general constraints on CMAF tracks and encoding constraints defined by each CMAF media profile in the CMAF switching set.

Track switching refers to the presentation of decoded media samples of one CMAF track up to a presentation time  $t$ , and presentation of decoded media samples of another CMAF track from time  $t$  onwards. For instance, this could be accomplished using two decoders, downloading overlapping portions of any two streams, decoding both streams, and switching on any decoded media sample. Track switching could be user initiated or programmatic.

Track switching within a CMAF switching set conforming to a CMAF media profile allows a player to download a sequence of non-overlapping CMAF fragments from different CMAF tracks and feed them to a single decoder for seamless playback. When a player performs CMAF track switching automatically during playback in response to available bandwidth, video quality, decoding capacity, etc., that is called “adaptive switching”.

## 6.5 CMAF specified objects and profiles

### 6.5.1 Object derivation and interoperability code points

Common Media Application Format specifies track formats and media formats for encoding multimedia and metadata derived from the fragmented ISO Base Media File Format and various audio, video, and subtitle media formats.

- CMAF specifies encoded media objects in a CMAF track format that applies to all media types and determines the objects that are encoded and decoded.
- CMAF specifies logical sets of encoded media objects and their general encoding constraints.
- CMAF specifies profiles that determine the specific encoding constraints on media objects that conform to identified interoperability points called profiles.
- CMAF addressable media objects are derived by packaging CMAF headers, CMAF chunks, and CMAF fragments as ISOBMFF data structures available for storage and delivery.

CMAF addressable media objects inherit the properties of encoded media objects, their profiles, and logical media object and set constraints, and make the process of encoding and decoding of CMAF fragments independent of the delivery method and packaging. Different delivery specifications can map the same CMAF addressable media objects to their different manifests to enable interoperability. CMAF addressable media objects, such as CMAF chunks and CMAF segments, are logically derived from CMAF fragments, but it is up to implementations whether to physically derive them from CMAF fragments by post processing or encode and package them directly.

### 6.5.2 Encoded media objects

- CMAF specifies **CMAF headers** and **CMAF fragments** that form **CMAF tracks** that conform to **CMAF media profiles**. These are the encoded objects from which all other objects are derived.

### 6.5.3 Logical media object sets

- CMAF specifies **CMAF switching sets** that logically group CMAF tracks that are alternative encodings of the same content constrained to enable seamless switching. A CMAF switching set is constrained to simplify seamless switching between tracks by requiring time-aligned CMAF fragments, constrained encoding parameters, and shared encryption keys.
- CMAF specifies **aligned CMAF switching sets** that logically group multiple CMAF switching sets that are alternative encodings of the same content with time-aligned CMAF fragments. Some players can seamlessly switch between time-aligned CMAF fragments in different CMAF switching sets that are encoded with different characteristics, such as different codecs or encryption keys.
- CMAF specifies **CMAF selection sets** that logically group alternative CMAF switching sets based on their content and encoding constraints, e.g. alternative languages or codecs. At most, one CMAF track at a time is intended to be selected from each CMAF selection set in a CMAF presentation.
- CMAF specifies **CMAF presentations** that logically group CMAF tracks containing related synchronized content that can be selected, streamed, adaptively switched, decoded, decrypted, and synchronized on playback to render a multimedia presentation.

### 6.5.4 Addressable media objects

- CMAF specifies **CMAF addressable media objects** that are derived from CMAF tracks and fragments, and can be used for storage and delivery of a CMAF presentation. CMAF addressable media objects include **CMAF headers**, **CMAF track files**, **CMAF segments**, and **CMAF chunks**. CMAF fragment encoding and decoding are independent of the addressable media objects used to transport those CMAF fragments.

## 6.5.5 CMAF profiles, brand, and identifiers

### 6.5.5.1 Overview

CMAF specifies the following brands and identifiers to support the creation and identification of interoperable content. They provide standardized conformance points for consistent content creation and validation. Players and devices can use these brands and identifiers to signal or recognize playback compatibility.

- CMAF specifies CMAF media profiles and ISOBMFF brands for widely used audio, video, and subtitle formats. Each CMAF media profile specifies constraints on codecs, media samples, CMAF fragments, and CMAF tracks to identify encoder/decoder interoperability and optional functionality including random access and seamless adaptive switching specified as CMAF switching set constraints. CMAF also specifies guidelines for the specification of CMAF media profiles and matching brands to enable other specifications to define CMAF media profiles that conform to general CMAF requirements. See [Table A.1](#), [Table B.1](#), [Table A.2](#), and [Table A.3](#).
- CMAF specifies CMAF presentation profiles and identifiers that conditionally require inclusion of CMAF tracks conforming to specific CMAF media profiles. CMAF presentations conforming to the CMAF presentation profiles in [A.1](#) are intended to be compatible with most media playback devices.

### 6.5.5.2 CMAF presentation profile identifiers

CMAF presentation profile identifiers identify CMAF presentations containing CMAF tracks conforming to CMAF media profiles that are conditionally required by the presentation profile. Players can rely on the availability of required CMAF tracks based on the presentation profile identifier. Other CMAF switching sets and media profiles can also be included in the CMAF presentation to enable additional features that might not be supported by all players conforming to the CMAF presentation profile. A CMAF presentation may conform to and signal multiple presentation profiles if all the required CMAF tracks for each presentation profile are available. Presentation profile conformance includes start alignment and synchronization of all CMAF tracks.

### 6.5.5.3 CMAF media profiles and ISOBMFF brands

CMAF file and media profile compatibility brands identify a CMAF track's file and media conformance to CMAF track and media profile constraints. Each media profile defines codec-specific properties in addition to those required by all CMAF tracks, such as the CMAF header sample entry, media sample format, CMAF fragment constraints, media sample presentation synchronization, scaling, mixing, layering, random access, and encoding constraints such as codec profiles and levels. CMAF media profiles can specify constraints between multiple CMAF tracks in a CMAF switching set, and thereby imply player requirements necessary to seamlessly switch between CMAF fragments in those CMAF switching sets.

### 6.5.5.4 Single initialization CMAF switching set constraints identifier

Single initialization constraints specify that a CMAF switching set conforms to general CMAF switching set constraints and to additional CMAF media profile defined constraints that enable seamless switching and decoding without dependence on the CMAF header for each CMAF track. CMAF switching set single initialization constraints can be signalled in a manifest, such as the `bitstreamSwitching` attribute in a DASH manifest. Players can take advantage of these constraints by processing a single CMAF header only once prior to sequencing CMAF fragments from the CMAF switching set.

### 6.5.5.5 ISOBMFF segment brands for CMAF media objects

Segment type compatibility brands are specified in [7.2](#) that can be stored in a `SegmentTypeBox` that start-delimits and identifies the type of CMAF media object that follows.

### 6.5.5.6 Specification of CMAF presentation and media profiles

The general requirements for CMAF media profiles and CMAF presentation profiles are specified in [Clause 12](#).

This document contains the following CMAF profiles.

- CMAF presentation profiles are specified in [A.1.2](#).
- CMAF media profiles are specified in [A.2](#), [A.3](#), [A.4](#), and [B.5](#).

Additional CMAF media profiles can be specified if they conform to the general CMAF track requirements, specify and register ISOBMFF brands, and define codec-specific CMAF track formats, codec constraints, and optional CMAF switching set constraints for seamless switching following the guidelines in [12.1](#).

Additional CMAF presentation profiles can be specified if they conform to the general CMAF requirements and specify a URL that identifies the conditionally required media profiles, as described in [12.2](#).

## 6.6 CMAF media object model

### 6.6.1 CMAF fragments

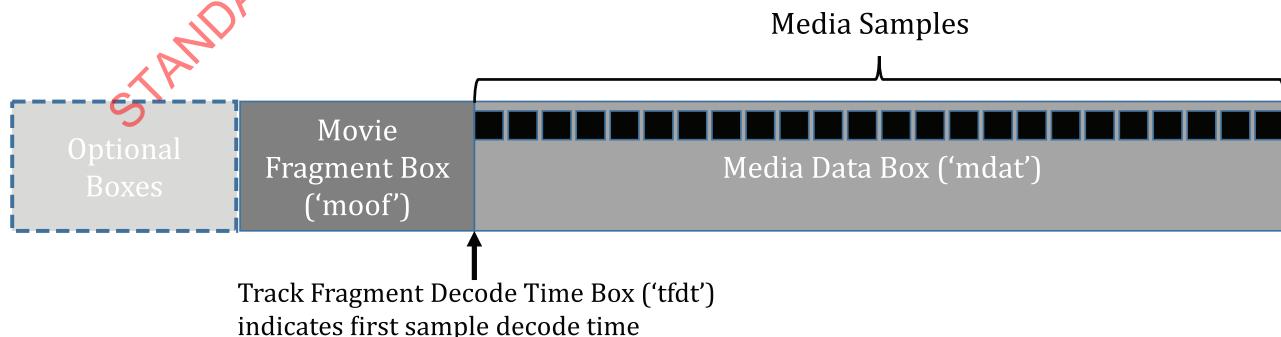
In the CMAF hypothetical application model, CMAF fragments are the media objects that are encoded and decoded.

Each CMAF media profile further constrains CMAF fragments and media sample data.

The CMAF media profiles in [Annex A](#) and their referenced track formats constrain CMAF fragments to be decodable independently of each other; for example, NAL structured video CMAF fragments contain one or more complete coded video sequences to make them randomly accessible and independently decodable. Audio CMAF fragments contain a sequence of audio access units, and subtitle CMAF fragments contain a single subtitle document in a media sample, constrained so that each CMAF fragment is randomly accessible and decodable. Common Encryption can also signal the necessary decryption parameters in each CMAF fragment to make it independently decryptable, in combination with an associated CMAF header and keys.

A CMAF fragment typically consists of one MovieFragmentBox and MediaDataBox pair, but can contain more than one of these pairs. When a fragment contains multiple pairs like this, each pair is called a CMAF chunk, and each CMAF chunk contains a consecutive subset of the CMAF fragment's media samples.

This is illustrated in [Figure 4](#).

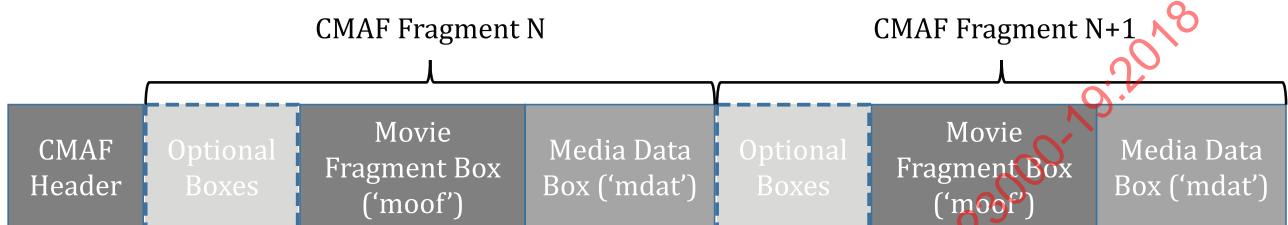


**Figure 4 — Example CMAF fragment data structure containing a single CMAF chunk**

## 6.6.2 CMAF tracks

A CMAF track is a continuous sequence of one or more CMAF fragments in presentation order conforming to a CMAF media profile and an associated CMAF header. The CMAF header contains a MovieBox sufficient to process and present all CMAF fragments in the CMAF track. A CMAF track can be produced by an encoder and ISOBMFF file packager, but it is made accessible in the form of CMAF addressable media objects that can be referenced as resources defined by an external media application specification.

This is illustrated in [Figure 5](#).



**Figure 5 — CMAF track data structure**

See [Clauses 7](#) through [11](#) for additional details on the construction of CMAF fragments and CMAF tracks of different media types.

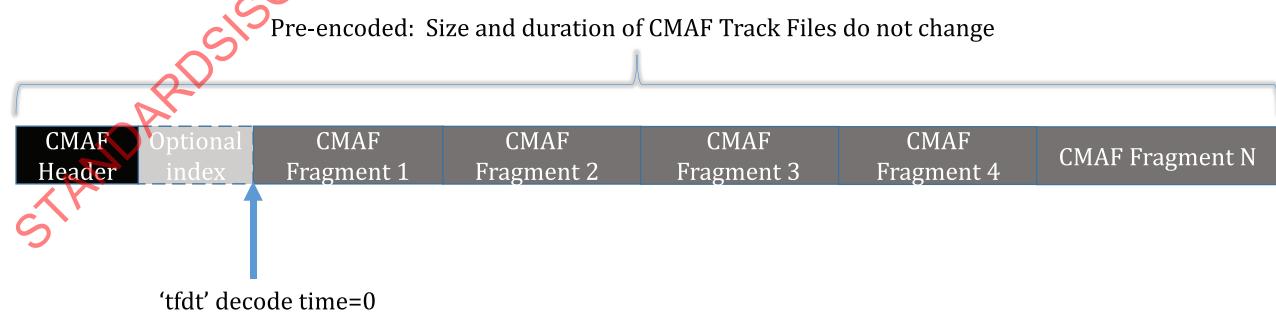
## 6.6.3 CMAF track files

A CMAF track file is a CMAF addressable media object consisting of a CMAF track stored in a single ISOBMFF file. A CMAF track file is a stored CMAF track with the following constraints.

- It starts with a CMAF header.
- The CMAF header is followed by a continuous sequence of one or more CMAF fragments stored in presentation order.
- The first CMAF fragment has a `baseMediaDecodeTime` of zero.

Additional boxes, such as `SegmentIndexBoxes`, can be present between the CMAF header and the first CMAF fragment.

This is illustrated in [Figure 6](#).



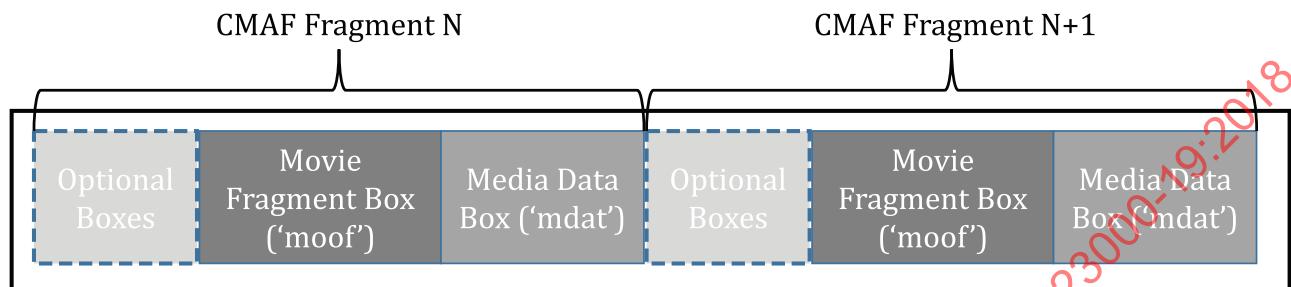
**Figure 6 — CMAF track file, addressable media object**

A sequence of one or more CMAF fragments can be requested from a CMAF track file resource, for example, using a resource identifier based on HTTP 1.1 consisting of the file URL and a byte range. CMAF resource identifiers can be listed in a manifest or determined by some other delivery format defined method, such as downloading a `SegmentIndexBox` to determine the byte ranges of CMAF fragments. The manifest and request method are out of scope of CMAF. A CMAF track file can also be downloaded or progressively downloaded.

### 6.6.4 CMAF segments

A CMAF segment is a CMAF addressable media object containing one or more consecutive CMAF fragments from a CMAF track. External application specifications can define how to reference a CMAF segment with a resource identifier that can be used by servers and manifests to reference and deliver each CMAF segment in a CMAF presentation.

This is illustrated in [Figure 7](#).



A CMAF Segment contains a consecutive sequence of one or more CMAF Fragments from a CMAF Track

**Figure 7 — Example CMAF segment, addressable media object**

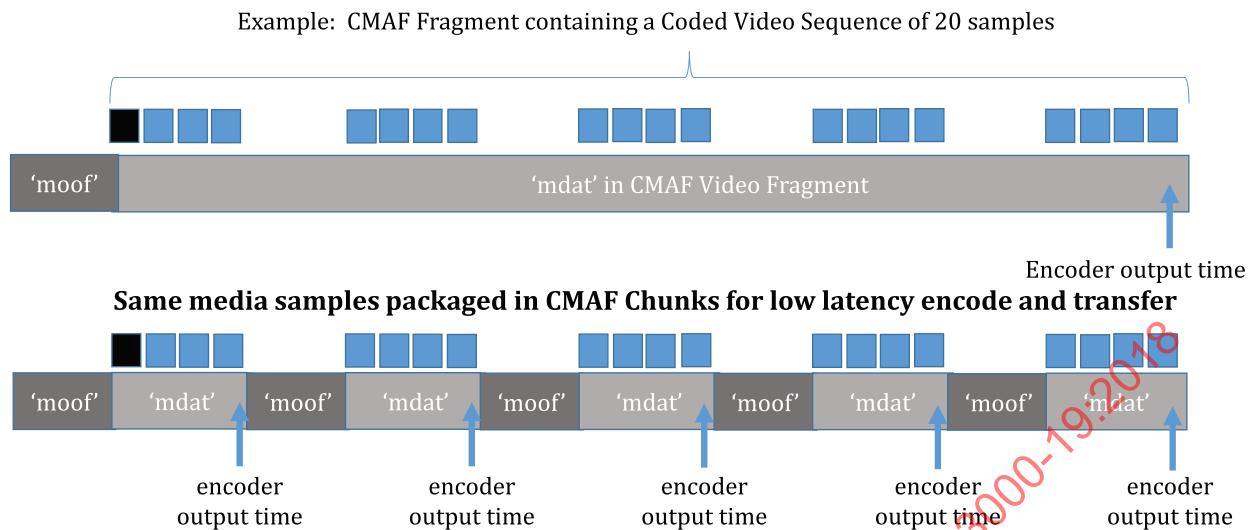
CMAF video fragment durations are typically 2 seconds to 6 seconds for coding efficiency, and CMAF segment durations are typically not greater than 10 seconds to 12 seconds to limit delivery latency in live streaming and limit bitrate rate adaptation response time.

Subtitle segment durations are usually close to video segment durations in live presentations to avoid increasing presentation delay. In a prerecorded CMAF presentation, a single subtitle CMAF segment can have a duration up to the duration of the CMAF track that contains it.

### 6.6.5 CMAF chunks

CMAF chunks are CMAF addressable media objects that contain a consecutive subset of the media samples in a CMAF fragment. CMAF chunks can be used by a delivery protocol to deliver media samples as soon as possible during live encoding and streaming, i.e. typically less than a second. The same media samples delivered in a CMAF segment containing one or more CMAF fragments would be delayed for the duration of the CMAF segment, i.e. typically several seconds. CMAF chunks enable the progressive encoding, delivery, and decoding of each CMAF fragment. A MovieFragmentBox at the start of each CMAF chunk provides access to the contained media samples. Broadcast, unicast, and multicast protocols can deliver and identify CMAF chunks through various methods, including resource IDs that include both a CMAF segment number corresponding to a CMAF fragment and the number of the CMAF chunk within the fragment to allow both fragment and chunk identification.

[Figure 8](#) illustrates that CMAF chunks inherit the media sample constraints of the CMAF fragment they are derived from. In this example of a video CMAF fragment containing a coded video sequence, the first media sample of the first CMAF chunk is SAP type 1 or 2, e.g. an IDR picture in an AVC CMAF fragment. The switching and splicing constraints are determined by the CMAF fragment. Boxes that lead the MovieFragmentBox in a CMAF fragment can also lead the MovieFragmentBox of the first CMAF chunk, although they are not shown in this example.

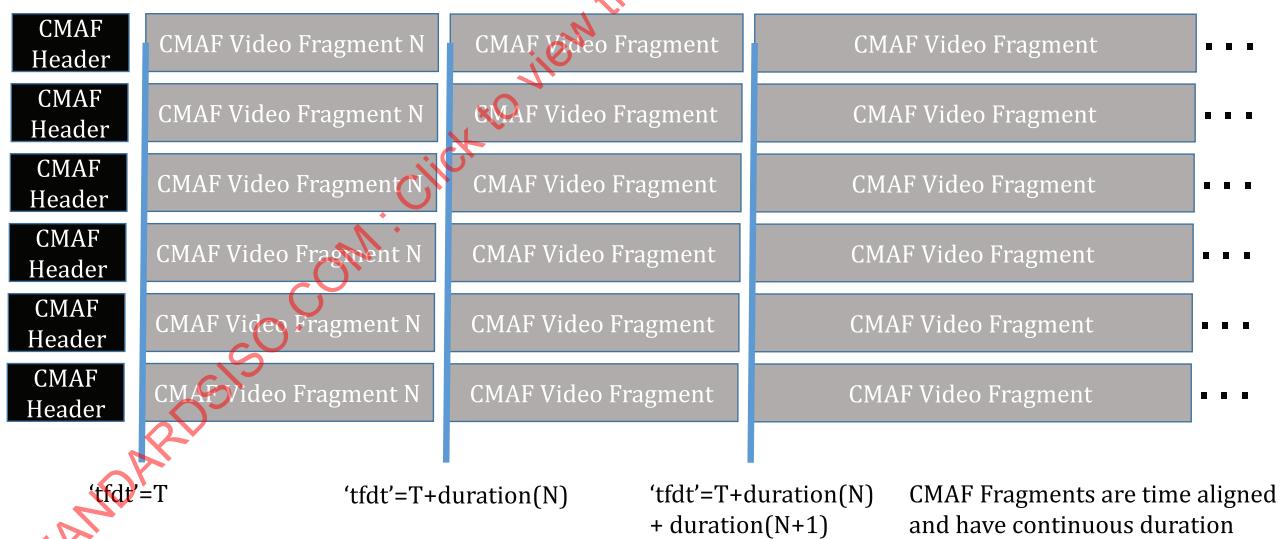


**Figure 8 — CMAF chunk, addressable media objects**

### 6.6.6 CMAF switching sets and adaptive switching

### 6.6.6.1 Overview

As illustrated in [Figure 9](#), CMAF switching sets contain time-aligned CMAF fragments that start with stream access points (SAP type 1 or 2) and timestamps to simplify switching between tracks by sequencing CMAF fragments from different CMAF tracks during playback.



**Figure 9 — A CMAF switching set**

The manifest description of CMAF switching sets and the CMAF tracks they contain is out of scope of the CMAF specification, but manifests are expected to contain enough CMAF track and addressable media object information to enable automatic CMAF track selection and adaptive switching by players.

### 6.6.6.2 Decoding adaptively switched CMAF tracks

Each CMAF media profile can optionally specify one or more CMAF switching set constraints that enable players to seamlessly switch between alternative CMAF fragments in a CMAF switching set. CMAF switching set constraints enable player functions including decoder and decryptor initialization,

video scaling, audio mixing, media sample presentation timing, etc. that are intended to make adaptive switching as perceptually seamless as possible.

The media profiles in the CMAF specification reference or define CMAF switching set constraints that enable a stream of CMAF headers and CMAF fragments to decode on decoders without downloading overlapping CMAF fragments or requiring the use of multiple decoders.

An adaptively delivered CMAF switching set results in a timed sequence of CMAF headers and fragments a player selects from multiple CMAF tracks contained in a CMAF switching set. The encoding constraints of the CMAF switching set and CMAF media profile determine how the CMAF switching set can be encoded and decoded. They are independent from CMAF addressable media object packaging applied during storage or transport.

The CMAF track format specifies two types of CMAF switching set processing, determined by specific constraints on NAL structured video track formats referenced by some CMAF media profiles. If a CMAF switching set conforms to general constraints, a CMAF track's header is assumed to be processed before every switch to that CMAF track. Additional constraints are specified for a CMAF switching set conforming to single initialization constraints that do not depend on processing a new CMAF header to switch to other CMAF tracks within a CMAF switching set.

[Figure 10](#) illustrates the player generated bitstreams resulting from the two types of CMAF switching set constraints and the decoding process each enables. Those two processes and bitstreams are described as “single initialization track switching” and “multiple initialization track switching”.

One method of single initialization track switching relies on a “common” CMAF header containing all referenced sample entries and decoding parameters in the CMAF switching set. It is assumed that players download CMAF headers only once per CMAF track. Consequently, new video parameter sets and indexes referenced by portions of a CMAF track encoded after the CMAF header has been downloaded would not be available to a player. Another method relies on parameter sets delivered in-band when those parameter sets are not included in the initially downloaded CMAF header or differ between CMAF tracks.

Audio CMAF switching sets typically only require single initialization because parameters that are allowed to change are contained in the audio media samples and CMAF fragments.

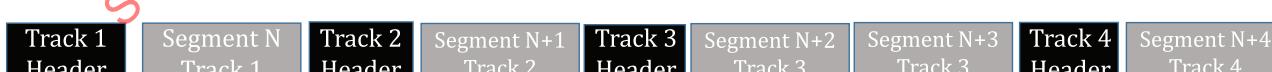
Subtitle CMAF tracks are typically not adaptively switched, and therefore have a single CMAF track per CMAF switching set.

See [7.3.4](#) for details on CMAF switching sets.

### Single Initialization Switching Process



### Multiple Initialization Switching Process



**Figure 10 — Adaptively switched resource streams for different CMAF switching set constraints**

#### 6.6.6.3 Displaying adaptively switched video CMAF tracks

The normative behaviour of a video decoder processing a single track is well-specified in each video codec specification, such as AVC, but display processing is considered out of scope of those codec specifications. For instance, AVC does not specify the conversion by a display processor of YCbCr 4:2:0

video spatial samples and colour subsamples output by a decoder to 32-bit or 48-bit RGB pixels in a size, colour space, resolution, and refresh rate appropriate for a specific display.

Adaptive switching, enabled by CMAF media profile constraints, allows a player's decoder and display system to deterministically scale different video spatial sampling used to encode different CMAF fragments to the same display size, position, and refresh rate. The encoding of CMAF tracks in a CMAF switching set is constrained so that a sequence of CMAF fragments can be accurately rescaled to the same presentation height and width with precise registration and visual continuity of the video content, including its size, shape, location, colour, brightness, etc. The variations allowed between CMAF tracks conforming to CMAF switching set constraints determine the scaling behaviour needed in display processors, but display processing and other player requirements are not directly specified by CMAF.

#### 6.6.6.4 Video parameter set storage and processing

The CMAF header of a CMAF video track needs to be processed on each CMAF track switch if decoding parameters and parameter indexes are not identical between CMAF tracks, and parameters that can change are not signalled in each CMAF fragment. This is often the case when each video CMAF track stores the video parameter set(s) for that CMAF track in its CMAF header sample table, e.g. the 'avc1' sample and track format. In that case, a decoder and display processor need to index the correct encoded video spatial sample counts and cropping parameters from a sample entry containing the correct SPS, PPS, and VUI information matching the parameter index encoded in each video NAL slice header. If CMAF tracks in a CMAF switching set use different spatial subsampling, they will have different parameter sets and sample entries, and different CMAF tracks might use the same parameter set index value for a different parameter set. That makes it necessary to process the CMAF header and sample table of each CMAF track when switching to that CMAF track so that the correct decoding parameters in a sample entry are in effect.

Multiple initialization switching can be used to process all CMAF switching sets, but may not result in seamless presentation when a system processes the entire CMAF header rather than just the video parameters allowed to change in a CMAF switching set, such as height, width, and cropping parameters. Multiple initialization adaptive switching is necessary for a NAL structured video CMAF switching set unless it conforms to single initialization constraints specified by its CMAF media profile.

A CMAF header of a CMAF video track only needs to be processed once for a CMAF switching set conforming to single initialization constraints. There could be several ways to achieve this functionality that can be specified by each CMAF media profile.

Two methods of specifying single initialization constraints are the following.

- A common CMAF header and sample entries for a CMAF switching set

In this case, each CMAF header and its sample table contain all parameter sets and sample entries used to encode CMAF fragments in the CMAF switching set. All the parameter indexes are unique and correspond to the index values encoded in the video NAL slice headers. There would be a different sample entry for each different subsampling used in the CMAF switching set.

- Generic CMAF headers and parameter sets stored in each CMAF fragment

In this case, each CMAF header contains initialization information in the video configuration record and/or additional boxes as specified in [Clause 9](#) and/or a manifest, but the SPS, PPS, and VUI information used for decoding is stored and indexed within each coded video sequence and CMAF fragment. The initialization information is sufficient for the CMAF media profile identified by brand in each CMAF header, and re-initialization is only necessary if switching to a higher CMAF media profile. For example, when a CMAF switching set contains some CMAF tracks that conform to both SD and HD profile, but others that only conform to HD profile, a decoder initialized to HD could switch all CMAF tracks without reinitializing, but a decoder initialized to SD would need to limit playback to SD media profile CMAF tracks, or reinitialize to HD, if possible.

For NAL structured video, single initialization encoding can be accomplished by encoding the Sequence Parameter Set and Picture Parameter Set NALs in the IDR access unit that begins that coded video

sequence (e.g. using the 'avc3' sample description and track format). Then the decoder and display processor can reference the parameter set necessary to decode and scale each CMAF fragment from the parameter set stored in the first media sample of each CMAF fragment.

Continuity between a sequence of CMAF presentations is not specified by CMAF (typically a manifest defines sequences of CMAF presentations). However, seamless playback of sequenced CMAF presentations is possible with CMAF presentations that conform to the same single initialization constraints and signal in-band parameters in each CMAF fragment.

### 6.6.7 CMAF selection sets and late binding

A CMAF selection set is a set of CMAF switching sets, where each CMAF switching set encodes an alternative aspect of the same presentation over the same presentation time, for example, different audio languages, video camera angles, video formats, or codecs.

CMAF players are expected to select one CMAF switching set from each CMAF selection set at the start of playback based on player compatibility and user preferences. Users or playback applications may switch between CMAF switching sets in a CMAF selection set during playback, but seamless presentation is not expected, either because the content differs (e.g. a different language or camera view) or because CMAF fragment time alignment and decoding are not constrained to decode seamlessly.

Aligned CMAF switching sets are encodings of the same source content with time-aligned CMAF fragments to enable compatible players to seamlessly switch between tracks in different but aligned CMAF switching sets that contain CMAF tracks that do not conform to all the constraints of a single CMAF switching set.

For example, in [Figure 11](#), the SD, HD, and UHD10 CMAF media profile CMAF switching sets could be aligned CMAF switching sets, and some players could seamlessly switch between them. In one case, only the encryption key might be different between the SD, HD and UHD CMAF switching sets. In another case, codecs could differ. Each CMAF switching set individually conforms to normal constraints for its media profile, and each player can determine if it can process the differences between aligned CMAF switching sets seamlessly.

Late binding allows CMAF tracks to be encoded once and used in different combinations. See [6.3](#) for more information.

Audio Selection Set	English AAC stereo CMAF Switching Set (single Track)
	French AAC stereo CMAF Switching Set (single Track)
	English multichannel CMAF Switching Set (single Track)
	French multichannel CMAF Switching Set (single Track)
Subtitle Selection Set	English WebVTT description CMAF Switching Set (single Track)
	English TTML description CMAF Switching Set (single Track)
	French WebVTT dub CMAF Switching Set (single Track)
	French TTML dub CMAF Switching Set (single Track)
Video Selection Set	SD Media Profile CMAF Switching Set (multiple Tracks)
	HD Media Profile CMAF Switching Set (multiple Tracks)
	UHD10 Media Profile CMAF Switching Set (multiple Tracks)

**Figure 11 — Example of CMAF selection sets**

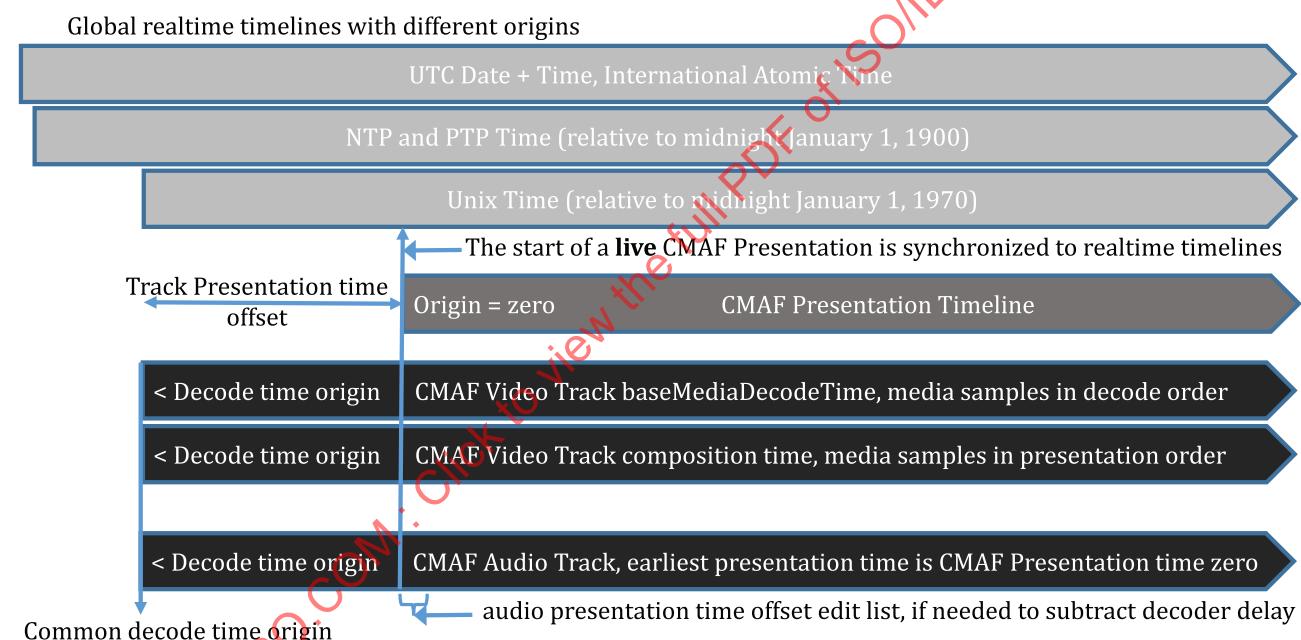
The description of selection sets in a manifest is not defined by the CMAF specification; however, each manifest format is expected to provide sufficient information to enable players to select optimal CMAF switching sets from CMAF selection sets.

When there are multiple tracks in a selection set that differ by language, the language of each track needs to be identified to at least the precision needed to differentiate the tracks, using the language field and optionally the `ExtendedLanguageBox` as defined in ISO/IEC 14496-12. See also [7.5.5](#).

When there is more than one track that is visually presented (e.g. video and subtitles), the visual layering of the tracks is indicated by the track header's layer field. CMAF recommends that subtitle CMAF tracks be positioned on layer -1 (ISOBMFF tracks with lower layer numbers overlay higher layer numbered tracks) and video CMAF tracks may be packaged with layer zero, or values relative to subtitles per the content creator's intended effect.

### 6.6.8 CMAF presentation timing model

There are multiple timelines involved in synchronizing a CMAF presentation. Each timeline has a timescale in units per second, increases at a continuous rate over time, and has an origin where the measure equals zero.



**Figure 12 — CMAF timelines and synchronization model**

All CMAF tracks in a CMAF presentation are specified in [Clause 7](#) to measure `baseMediaDecodeTime` in the `TrackFragmentBaseMediaDecodeTimeBox` from the same timeline origin and adjust each media sample's presentation time relative to its decode time using composition offsets and offset edit lists, when necessary. The resulting presentation timelines and decode timelines all reference the same (coincident) timeline origin. The timelines and their relationships are illustrated in [Figure 12](#).

The timelines are:

- Track decode time** — Is determined by the storage sequence and duration of each media sample, and for any media sample, equals the sum of all prior media sample durations added to the `baseMediaDecodeTime` of the first CMAF fragment.
- Track composition time** — Is determined by each sample's composition offset in the `TrackRunBox` relative to its decode time and is used to reorder video media samples to their presentation order.

- c) **Track presentation time** — Is determined by applying any offset edit list in the CMAF header to each media sample's composition time, as shown for a CMAF audio track in [Figure 12](#).
- d) **CMAF presentation timeline** — Is defined in CMAF to start with a CMAF presentation time equal to zero, at the earliest video media sample presentation time, or the earliest audio media sample presentation time if there are no video CMAF tracks. The CMAF track presentation time at CMAF presentation time zero is its "presentation time offset", which can also be represented in some manifest formats, such as DASH.
- e) **Wall clock time** — This timeline is important for live presentations, e.g. UTC time at the time of CMAF chunk availability. A wall clock timestamp may be stored in a `ProducerReferenceTimeBox` in a CMAF chunk or fragment, and that timestamp linked to a media sample decode time. Manifests can indicate the wall clock time that coincides with the start of a CMAF presentation so a player can determine when each CMAF fragment or chunk will be available for download.

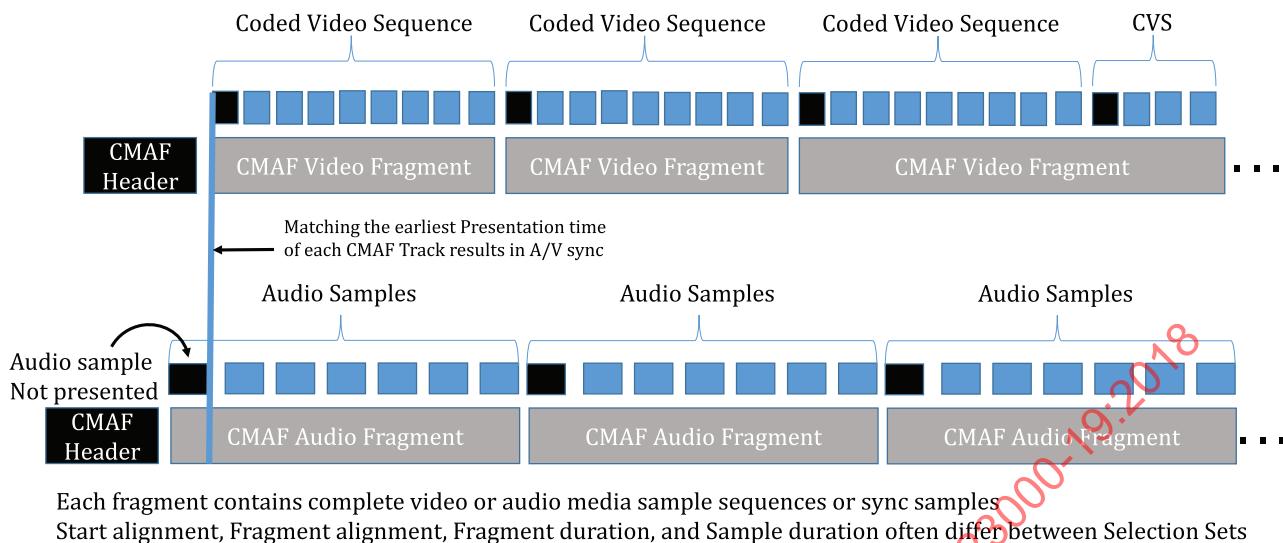
CMAF tracks in a CMAF presentation are required to use the same decode timeline origin, similar to the shared movie timeline of multiple tracks in one ISOBMFF file. But CMAF tracks can start at a non-zero decode time (`baseMediaDecodeTime`) stored in the earliest video CMAF fragment. The CMAF presentation timeline starts at zero, coincident with the earliest media sample presentation time, and is equivalent to the ISOBMFF movie timeline.

As illustrated in [Figure 13](#), CMAF tracks with audio CMAF fragments that overlap the earliest video media sample are intended to start presenting simultaneously with the earliest video media sample. Audio CMAF track presentation time is determined by the value of `baseMediaDecodeTime` recorded in each CMAF chunk or fragment, and any presentation time offset in an edit list, if an edit list is present in the CMAF track's CMAF header.

ISOBMFF defines files and requires that the decode time of each media sample is the sum of all prior media sample durations in that track in stored order. The first media sample in each track in a file therefore has a decode time of zero. Movie fragmentation is optional. In contrast, all media samples in CMAF tracks are stored in movie fragments, and the first CMAF fragment can have a non-zero `baseMediaDecodeTime` in the `TrackFragmentBaseMediaDecodeTimeBox`. The decode time of each media sample equals the sum of prior media sample durations in the CMAF track added to the `baseMediaDecodeTime` of the first CMAF fragment in the CMAF track. The decode time of each media sample also equals the sum of prior media sample durations in the CMAF fragment that contains it added to the CMAF fragment's `baseMediaDecodeTime`.

Manifests can specify a presentation time offset for each CMAF switching set to determine the CMAF track presentation time at the start of each CMAF presentation. The start of CMAF presentations is assumed to be the earliest video media sample when video is included in a CMAF presentation, otherwise the earliest audio media sample. To maintain audio, video, and subtitle synchronization encoded in the CMAF tracks, the presentation time offset of every CMAF switching set in a manifest will be equivalent, which means equal presentation time offsets, but possibly different integer values and timescales per CMAF switching set. Matching presentation time offsets plus CMAF presentation time in each CMAF track being presented maintains the encoded synchronization between audio, video, and text content on playback.

A manifest can select different start and end times with presentation time offsets and durations to present different timespans of the CMAF presentation timeline to play a portion of a CMAF presentation.



**Figure 13 — Audio/video synchronization and start alignment**

CMAF tracks containing video use negative composition offsets where necessary to reorder media samples without adding composition delay, so that the earliest composition time and presentation time of each CMAF fragment equal its earliest decode time (with the optional exception of CMAF track files, mentioned above). The `baseMediaDecodeTime` in the `TrackFragmentBaseMediaDecodeTimeBox` is the earliest media sample presentation time of each CMAF fragment, and the `baseMediaDecodeTime` of the earliest video CMAF fragment is the presentation time offset of the CMAF track at CMAF presentation time zero.

CMAF video track files, illustrated in [Figure 6](#), can be encoded using positive composition offsets and packaged with an edit list in the CMAF header to remove composition delay. All other video CMAF tracks use negative composition offsets, not edit lists, to remove composition delay.

During random access and “trick play” (fast forward, reverse, slow motion, jump, etc.), the `TrackFragmentBaseMediaDecodeTimeBox` and the `TrackRunBox` can be used to determine each media sample decode time relative to the CMAF fragment `baseMediaDecodeTime`, then any edit list present in the associated CMAF header can be used to calculate the presentation time offset relative to each media sample’s composition time.

Decode time discontinuities between CMAF fragments can result from starting a new CMAF presentation in a playlist, for example, in sequenced programs or programs interspersed with prerecorded ads. A manifest can locate the new CMAF presentation on its manifest presentation timeline and adjust CMAF switching set presentation time offsets accordingly.

However, in cases where a decode time discontinuity results from one or more missing or damaged CMAF fragments in a continuous CMAF presentation, players can use the `baseMediaDecodeTime` of the next available CMAF fragment to resume presentation synchronized to the current presentation timeline. Error concealment could include repeating video frames and skipping over missing content in the case of on demand or buffered media, but all CMAF tracks have to skip the same duration to maintain A/V synchronization.

## 6.6.9 Manifest information

Although CMAF does not define the form or the content of the manifest, it does define its role. A manifest is a document that describes one or more CMAF presentations, e.g. an MPEG DASH Media Presentation Description (MPD). A manifest is responsible for describing the combination and synchronization of independently packaged CMAF tracks grouped in CMAF switching sets and selection sets to form a synchronized multimedia presentation.

A manifest provides the player with information to select, initialize, start align, and synchronize the CMAF track(s) to be played and identify CMAF media objects as resources to access and possibly download them. CMAF tracks and CMAF fragments contain sufficient information to enable decryption, decoding, and presentation scheduling. A manifest can also provide information on delivery protocol, network management, authorization, license acquisition, etc., in addition to resource identification and presentation description. The manifest can also signal that tracks conform to a CMAF media profile.

In cases where there are multiple CMAF tracks in a CMAF switching set or multiple CMAF presentations in a sequence, a manifest can indicate to a player that the CMAF switching set conforms to a CMAF media profile and CMAF switching set constraints, so a player can select and initialize compatible CMAF tracks and can switch seamlessly between those CMAF tracks, reinitializing only when necessary.

### 6.6.10 CMAF addressable media objects, resources, and resource identifiers

CMAF headers, CMAF chunks, and CMAF fragments can be packaged and referenced as CMAF addressable media objects for storage and delivery. Each CMAF addressable media object can be referenced as a resource as specified by an external specification, e.g. MPEG DASH (ISO/IEC 23009-1).

External specifications can define resource identifiers and the CMAF addressable media objects they identify so that servers, CDNs, and manifests can access CMAF addressable media objects and sequence them for decoding and presentation. A common example is a CMAF addressable media object packaged as the body of an HTTP response and identified as a resource by a unique segment URL. A CMAF track file can be segmented during delivery by adding a byte range to each HTTP request for the file's URL, using byte ranges matching one or more CMAF fragments or the CMAF header of the CMAF track file.

Use of the same URI per CMAF addressable media object is recommended to improve efficiency in content distribution networks and caches, even when different manifests or delivery protocols are used. The use of different URIs for the same CMAF addressable media object causes duplicate storage and delivery of the same CMAF media objects. For example, in cases where the same CMAF segments can be delivered over broadcast, multicast, and unicast, a CMAF player can request a CMAF segment with a single resource identifier and retrieve it from the player's cache regardless of which network path and encapsulation delivered it.

## 7 CMAF track format

### 7.1 Overview

The CMAF track format is derived from the ISO Base Media File Format in this clause and a compatibility brand is specified. Several CMAF media objects are derived from the CMAF track format.

### 7.2 CMAF brands

[Clause 7](#) defines the requirements and constraints that apply to all CMAF tracks.

A CMAF track should include the '`cmfc`' file brand, indicating conformance with [Clause 7](#) and [Clause 8](#) (specifying file format and optional encryption, but not media encoding). This file brand also declares that boxes specified in Common Encryption (ISO/IEC 23001-7) and DASH (ISO/IEC 23009-1) may be present.

The primary constraints of the '`cmfc`' file type brand are described in [7.3.2.2](#).

Media profiles and their compatibility brands may specify additional box requirements and CMAF track constraints, such as the CMAF media profiles that reference the NAL structured video CMAF track format defined in [Clause 9](#), incorporating boxes specified in ISO/IEC 14496-15, and AAC audio CMAF track format specified in [Clause 10](#), incorporating boxes specified in ISO/IEC 14496-3.

CMAF also specifies addressable media objects in [7.3.3](#) that are derived from CMAF tracks and have segment type brands that may be included in a `SegmentTypeBox` prepended to an addressable media object to identify its type (see [Table 1](#)).

**Table 1 — CMAF brands**

Brand	Location	Conformance requirements
'cmfc'	FileTypeBox and SegmentTypeBox	<a href="#">7.3</a>
'cmfs'	SegmentTypeBox	<a href="#">7.3.3.1</a>
'cmfl'	SegmentTypeBox	<a href="#">7.3.3.2</a>
'cmff'	SegmentTypeBox	<a href="#">7.3.2.3</a>

An ISO brand, such as 'iso9', that indicates the ISOBMFF boxes actually present in each CMAF track should be listed in `compatible_brands` to improve interoperability with file readers. The 'isoX' brand (where "X" represents a number or letter) specified in each new edition of ISO/IEC 14496-12 determines what box versions are allowed in CMAF tracks, unless specifically constrained by the 'cmfc' brand specified in [Clause 7](#).

If 'cmfc' is the major\_brand, the minor\_version shall be set to 0, and file names should use the file extensions in [Table 2](#). Otherwise, file names should use the file extension and Internet Media Type specified to match the major brand, e.g. \*.mp4, \*.3gp, \*.uvv, \*.uva, etc.

NOTE It is expected that file readers read possible future versions of CMAF that increment the `minor_version` number.

**Table 2 — Common Media Application Format file extensions**

Track type	File extension	Internet Media Type (MIME type)
Video	.cmfv	video/mp4
Audio	.cmfa	audio/mp4
Text (subtitle)	.cmft	application/mp4

Each CMAF track has a CMAF header associated with it, although the CMAF header and CMAF fragments might not be stored as an ISOBMFF file, and might not be stored at all, i.e. only existing temporarily as streams.

## 7.3 CMAF media objects

### 7.3.1 CMAF boxes

CMAF tracks shall include the following ISOBMFF boxes with nesting, optionality, and cardinality specified in the following tables, with box definitions incorporated by reference to the clauses listed in the "Specification" column that reference ISO/IEC 14496-12, ISO/IEC 23001-7, or ISO/IEC 23009-1.

Some boxes are additionally constrained by CMAF, as specified by clauses in this document referenced in the "Constraints" column of [Tables 3](#) through [5](#). The normative references are dated, e.g. the 2015 edition of ISO/IEC 14496-12, so the box versions and features defined in that edition are valid in a CMAF track unless additionally constrained by CMAF or by the 'isoX' brand promised in the `FileTypeBox`.

CMAF addressable media objects are derived from the CMAF track, CMAF header, CMAF chunk, and CMAF fragment recorded media objects and specify additional box constraints in [7.3.3](#).

Additional boxes and constraints are specified by CMAF track formats and CMAF media profiles for specific audio, video, and subtitle formats (see [7.4.4](#)).

Legend for <a href="#">Table 3</a> , <a href="#">Table 4</a> , and <a href="#">Table 5</a>									
The “Format Req.” column indicates the number of boxes that are required to be present in a CMAF track, where:									
“*” means “zero or more” may be present;									
“+” means “one or more” shall be present;									
“0/1” indicates only one box may be present, and it is optional;									
“1” indicates one box shall be present;									
“CR” indicates that a box is conditionally required as specified in the CMAF track format or a specific media profile, and shall be present under those conditions (see “Constraints” column).									

**Table 3 — CMAF header boxes**

NL 0	NL 1	NL 2	NL 3	NL 4	NL 5	Format Req.	Specification	Constraints	Description
ftyp						1	ISO/IEC 14496-12	CMAF <a href="#">7.2</a>	File Type and Compatibility
moov						1	ISO/IEC 14496-12		Container for functional metadata
	mvhd					1	ISO/IEC 14496-12	CMAF <a href="#">7.5.1</a>	Movie header
	trak					1	ISO/IEC 14496-12		Container for each track
	tkhd					1	ISO/IEC 14496-12	CMAF <a href="#">7.5.4</a>	Track header
	edts					CR	ISO/IEC 14496-12	CMAF <a href="#">7.5.13</a>	Edit Box
		elst				1	ISO/IEC 14496-12	CMAF <a href="#">7.5.13</a>	Edit List Box
		mdia				1	ISO/IEC 14496-12		Track Media Information
		mdhd				1	ISO/IEC 14496-12	CMAF <a href="#">7.5.5</a>	Media Header
		hdlr				1	ISO/IEC 14496-12		Declares the media handler type
		elng				0/1	ISO/IEC 14496-12		Extended Language Tag
		minf				1	ISO/IEC 14496-12		Media Information container
		vmhd				CR	ISO/IEC 14496-12	CMAF <a href="#">7.5.6</a>	Video Media Header
		smhd				CR	ISO/IEC 14496-12	CMAF <a href="#">7.5.7</a>	Sound Media Header
		sthd				CR	ISO/IEC 14496-12	CMAF <a href="#">7.5.8</a>	Subtitle Media Header
		dinf				1	ISO/IEC 14496-12		Data Information Box
		dref				1	ISO/IEC 14496-12	CMAF <a href="#">7.5.9</a>	Data Reference Box, declares source of media data in track

Table 3 (continued)

NL 0	NL 1	NL 2	NL 3	NL 4	NL 5	Format Req.	Specification	Constraints	Description
				stbl		1	ISO/IEC 14496-12	CMAF <a href="#">7.5.12</a>	Sample Table Box, container for the time/space map
				stsd		1	ISO/IEC 14496-12	CMAF <a href="#">7.5.10</a>	Sample Descriptions (see <a href="#">Table 4</a> for additional details)
				stts		1	ISO/IEC 14496-12	CMAF <a href="#">7.5.12</a>	Decoding Time to Sample
				stsc		1	ISO/IEC 14496-12	CMAF <a href="#">7.5.12</a>	Sample-to-Chunk
				stsz/stz2		1	ISO/IEC 14496-12	CMAF <a href="#">7.5.12</a>	Sample Size Box
				stco		1	ISO/IEC 14496-12	CMAF <a href="#">7.5.12</a>	Chunk Offset
				sgpd	CR	ISO/IEC 14496-12	CMAF <a href="#">7.5.18</a>	Sample Group Description Box	
				stss	CR	ISO/IEC 14496-12	CMAF <a href="#">7.5.12</a>	Sync Sample Box	
	udta					0/1	ISO/IEC 14496-12		User Data Box
		cppt				*	ISO/IEC 14496-12		Copyright Box
		kind				*	ISO/IEC 14496-12	CMAF <a href="#">7.5.3</a>	Track Kind Box
	mvex					1	ISO/IEC 14496-12		Movie Extends Box
		mehd				0/1	ISO/IEC 14496-12	CMAF <a href="#">7.5.1</a>	Movie Extends Header
		trex				1	ISO/IEC 14496-12	CMAF <a href="#">7.5.14</a>	Track Extends Box
	pssh					*	ISO/IEC 23001-7	CMAF <a href="#">7.4.3</a>	Protection System Specific Header Box

Table 4 — CMAF header protected sample entry boxes

NL 5	NL 6	NL 7	NL 8	Format Req.	Specification	Constraints	Description
stsd				1	ISO/IEC 14496-12		Sample description box
	sinf			CR	ISO/IEC 14496-12	CMAF <a href="#">7.5.11</a>	Protection Scheme Information Box
		frma		1	ISO/IEC 14496-12		Original Format Box
		schm		1	ISO/IEC 14496-12		Scheme Type Box
		schi		1	ISO/IEC 14496-12		Scheme Information Box
			tenc	1	ISO/IEC 23001-7	CMAF <a href="#">7.4.1</a>	Track Encryption Box

NOTE [Table 4](#) is a continuation of [Table 3](#) showing nesting levels 5 to 8 separately to reduce table width.

**Table 5 — CMAF chunk, CMAF fragment, CMAF segment, and CMAF track file boxes**

NL 0	NL 1	NL 2	NL 3	NL 4	NL 5	Format Req.	Specifica-tion	Constraints	Description
styp						0/1	ISO/IEC 14496-12		Segment Type
prft						0/1	ISO/IEC 14496-12		Producer Reference Time
emsg						*	ISO/IEC 23009-1	CMAF <a href="#">7.4.5</a>	Event Message
moof						+	ISO/IEC 14496-12		Movie Fragment
	mfhd					1	ISO/IEC 14496-12		Movie Fragment Header
	traf					1	ISO/IEC 14496-12		Track Fragment
	tfhd					1	ISO/IEC 14496-12	CMAF <a href="#">7.5.16</a>	Track Fragment Header
	tfdt					1	ISO/IEC 14496-12		Track Fragment Base Media Decode Time
	trun					1	ISO/IEC 14496-12	CMAF <a href="#">7.5.17</a>	Track Fragment Run Box
	senc					0/1	ISO/IEC 23001-7	CMAF <a href="#">8.2.2.1</a>	Sample Encryption Box
	saio					CR	ISO/IEC 14496-12	CMAF <a href="#">8.2.2.1</a>	Sample Auxiliary Information Offsets Box
	saiz					CR	ISO/IEC 14496-12	CMAF <a href="#">8.2.2.1</a>	Sample Auxiliary Information Sizes Box
	sbgp					*	ISO/IEC 14496-12		Sample to Group Box
	sgpd					*	ISO/IEC 14496-12	CMAF <a href="#">7.5.17</a>	Sample Group Description Box
	subs					CR	ISO/IEC 14496-12	CMAF <a href="#">7.5.19</a>	Sub-sample Information Box
mdat						+	ISO/IEC 14496-12	CMAF <a href="#">7.5.18</a>	Media Data container for media samples

### 7.3.2 CMAF track media objects

#### 7.3.2.1 CMAF header

A CMAF header is defined in [3.1.2](#) to be a “sequence of CMAF constrained ISOBMFF boxes that do not reference any media samples, but are associated with a CMAF track and necessary for the decoding of its CMAF fragments”.

A CMAF header conforms to the following constraints.

- A CMAF header shall contain the set of boxes in [Table 3](#) and [Table 4](#) with the conditions and optionality listed.
- Each CMAF header shall form a valid CMAF track, as specified in [7.3.2.2](#), when followed by a continuous sequence of associated CMAF fragments in decode order.

c) A CMAF header shall be conformant with ISO/IEC 14496-12 and the following additional constraints and requirements.

- 1) The CMAF header shall start with a `FileTypeBox`.
- 2) The CMAF header shall include exactly one `MovieBox`.
- 3) The `MovieBox` shall start with a `MovieHeaderBox`, as constrained in [7.5.1](#).
- 4) The `MovieBox` shall contain exactly one track containing media data as specified in [7.3.2.2](#).

NOTE Timed metadata tracks can be provided as separate CMAF tracks in a separate selection set.

- 5) The `MovieBox` shall contain a `MovieExtendsBox`, as defined in ISO/IEC 14496-12, to indicate that the file contains `MovieFragmentBoxes`.
- 6) The `MovieExtendsBox` may contain a `MovieExtendsHeaderBox`, as defined in ISO/IEC 14496-12, and if so shall provide the overall duration of the CMAF track. If the duration is unknown, this box shall be omitted.

d) If the CMAF header requires sample entries with a decoder configuration record, as specified by a CMAF media profile, with decoding and rendering parameter limits, such as codec profile, level, image height, width, etc., then the first sample entry shall contain parameters equal or greater than the maximum values contained in all CMAF fragments in the track to enable a single initialization of media decoding and rendering systems to render all CMAF fragments in the CMAF track.

### 7.3.2.2 CMAF track

A CMAF track is defined in [3.2.1](#) to be a “sequence of CMAF fragments that are consecutive in presentation time, contain one media stream, conform to the 'cmfc' brand, including an associated CMAF header that can initialize playback”.

**Table 6 — CMAF track structure**

NL 0	Format Req.	Specification	CMAF constraints	Description
CMAF Header	1		CMAF <a href="#">7.3.2.1</a>	CMAF header
CMAF Fragment	1+		CMAF <a href="#">7.3.2.4</a>	CMAF fragment

A CMAF track has the structure shown in [Table 6](#) and conforms to the following constraints.

- a) A CMAF track shall conform to the 'cmfc' brand and contain the set of boxes in [Table 3](#), [Table 4](#), and [Table 5](#), with the conditions and optionality listed.
- b) The concatenation of a CMAF header and all CMAF fragments in the CMAF track in consecutive decode order shall be a valid fragmented ISOBMFF file, with the exception that the first CMAF fragment in a CMAF track may have a non-zero `baseMediaDecodeTime`.
- c) Each CMAF fragment in a CMAF track shall have `baseMediaDecodeTime` equal to the sum of all prior CMAF fragment durations added to the first fragment's `baseMediaDecodeTime`. A CMAF fragment duration is the sum of the media sample durations, documented in the `TrackFragmentRunBox` in the `MovieFragmentHeaderBox`.

NOTE Valid CMAF tracks do not have media time discontinuities resulting from missing media samples or fragments. Gaps in decode time would result in audio-video synchronization errors. For recommendations on handling missing media samples and missing CMAF fragments, see [Annex F](#).

- d) Each CMAF track contains a single ISOBMFF track and `TrackBox`, as determined by CMAF header constraints specified in [7.3.2.1](#).

Additional constraints specific for encryption, video, audio, and subtitle CMAF tracks are specified in [Clauses 8 to 11](#) and derived from the general CMAF track format specified in [Clause 7](#).

### 7.3.2.3 CMAF chunk

A CMAF chunk is defined in [3.1.4](#) to be a “CMAF media object that contains a consecutive subset of the media samples of a CMAF fragment, where only the first CMAF chunk of a CMAF fragment is constrained to be an adaptive switching point”.

CMAF chunks are the smallest CMAF media object that can be encoded, and they can be referenced as addressable media objects.

**Table 7 — CMAF chunk structure**

NL 0	Format Req.	Specification	CMAF constraints	Description
styp	0/1	ISO/IEC 14496-12	CMAF <a href="#">7.2</a>	Segment Type Signalling compatibility to CMAF Chunk
prft	0/1	ISO/IEC 14496-12		Producer Reference Time
emsg	*	ISO/IEC 23009-1	CMAF <a href="#">7.4.5</a>	Event Message
moof	1	ISO/IEC 14496-12	CMAF <a href="#">Table 5</a>	Movie Fragment box and the boxes it contains
mdat	1	ISO/IEC 14496-12	CMAF <a href="#">7.5.19</a>	Media Data container for media samples

Each CMAF chunk shall conform to the following constraints.

- A CMAF chunk shall contain the boxes in [Table 7](#) with the conditions and optionality indicated.
- The MovieFragmentBox shall conform to the constraints of the 'cmfc' brand specified in [Table 5](#), such as containing only one TrackFragmentBox that contains only one TrackRunBox.

NOTE 1 Since there is only one TrackRunBox per MovieFragmentBox, all media samples of a CMAF chunk are located in a single track run in one MediaDataBox.

- The MediaDataBox shall contain all media samples referenced by the TrackRunBox and should immediately follow the TrackRunBox in byte order.
- All media samples in a CMAF chunk shall be addressed by byte offsets in the TrackRunBox that are relative to the first byte of the MovieFragmentBox. See ISO/IEC 14496-12.
- A CMAF chunk shall contain a consecutive subset of the media samples of a CMAF fragment.
- A CMAF chunk shall contain a TrackFragmentDecodeTimeBox containing the baseMediaDecodeTime of the first media sample.
- A sequence of CMAF chunks that spans the decode time and duration of a CMAF fragment containing those media samples shall contain all the media samples in the CMAF fragment, without duplicates, i.e. CMAF chunks in a CMAF track shall not overlap or have gaps in decode time.
- The decode time and CMAF track presentation times of a media sample in a CMAF track shall not change whether the sample is contained in a CMAF chunk or a CMAF fragment.

NOTE 2 Because the CMAF fragment and the first CMAF chunk both start with the same media sample, they both have the same baseMediaDecodeTime.

- i) A CMAF chunk may include a SegmentTypeBox preceding the MovieFragmentBox and may be referenced as a CMAF addressable media object. The SegmentTypeBox should include the CMAF chunk brand '`cmaf`' and may include any valid `compatible_brands` listed in the FileTypeBox of the CMAF track's CMAF header.
- j) CMAF chunks may be referenced as CMAF addressable media objects, and optional file level boxes prepended when they are formatted as CMAF addressable media objects.

NOTE 3 Preceding boxes indicated in [Table 7](#) can be prepended to CMAF chunks at time of encoding or prepended to CMAF chunks and CMAF segments at time of delivery. It is expected that players will process the boxes prepended to the start of each CMAF addressable media object, but they will ignore these boxes when located within a CMAF addressable media object, i.e. not prepended to the first CMAF chunk in a CMAF segment. The SegmentTypeBox is specified by ISO/IEC 14496-12 to be the first box of an ISOBMFF segment and can be ignored in other locations.

NOTE 4 A SegmentTypeBox prepended to CMAF chunks or fragments at the time of encoding can be useful as a media object delimiter and identifier. It allows indexing and packaging operations to rapidly locate and identify CMAF chunks and fragments within a CMAF track without deep inspection of elementary streams or other clues to identify CMAF fragment boundaries and preceding boxes intended to be part of each media object.

#### 7.3.2.4 CMAF fragment

A CMAF fragment is defined in [3.1.1](#) to be an “encoded ISOBMFF media segment conforming to CMAF constraints”.

A CMAF fragment contains one or more CMAF chunks, therefore CMAF chunks may be present in CMAF addressable media objects, including CMAF segments and CMAF track files, and a CMAF chunk may be referenced as a CMAF addressable media object (see [7.3.3.2](#)).

**Table 8 — CMAF fragment structure**

NL 0	Format Req.	Specification	CMAF constraints	Description
styp	0/1	ISO/IEC 14496-12	CMAF <a href="#">7.2</a>	Segment Type Signalling compatibility to CMAF Fragment
prft	0/1	ISO/IEC 14496-12		Producer Reference Time
emsg	*	ISO/IEC 23009-1	CMAF <a href="#">7.4.5</a>	Event Message
CMAF Chunk	1+	CMAF <a href="#">7.3.2.3</a>		CMAF Chunk

Each CMAF fragment conforms to the following constraints.

- a) A CMAF fragment shall contain the boxes in [Table 8](#), with the conditions and optionality indicated.
- b) Each CMAF fragment, in combination with its associated CMAF header and optional decryption key(s), shall contain sufficient metadata to enable the CMAF fragment to be decoded, decrypted, and displayed when it is independently accessed.

NOTE 1 For example, if sample groups and sample group descriptions are used to signal encryption key changes, then a SampleGroupDescriptionBox and SampleToGroupBox probably need to be present in the TrackFragmentBox to make the CMAF fragment randomly accessible and decryptable.

- c) The first CMAF chunk in a CMAF fragment may be preceded by other boxes as long as the CMAF fragment remains conformant, including one SegmentTypeBox and/or one or more ProducerReferenceTimeBox(es) and/or DASHEventMessageBox(es). (See [7.4.5](#) and [Annex E](#) for more information on Event Messages).

- d) A CMAF fragment with a SegmentTypeBox preceding the first MovieFragmentBox should include the CMAF fragment brand 'cmff' and may include any compatible\_brands listed in the FileTypeBox of the CMAF track's CMAF header.
- e) A CMAF fragment can be referenced as a CMAF segment and may have a SegmentTypeBox prepended that includes the CMAF segment brand 'cmfs' to identify the start of the CMAF segment.

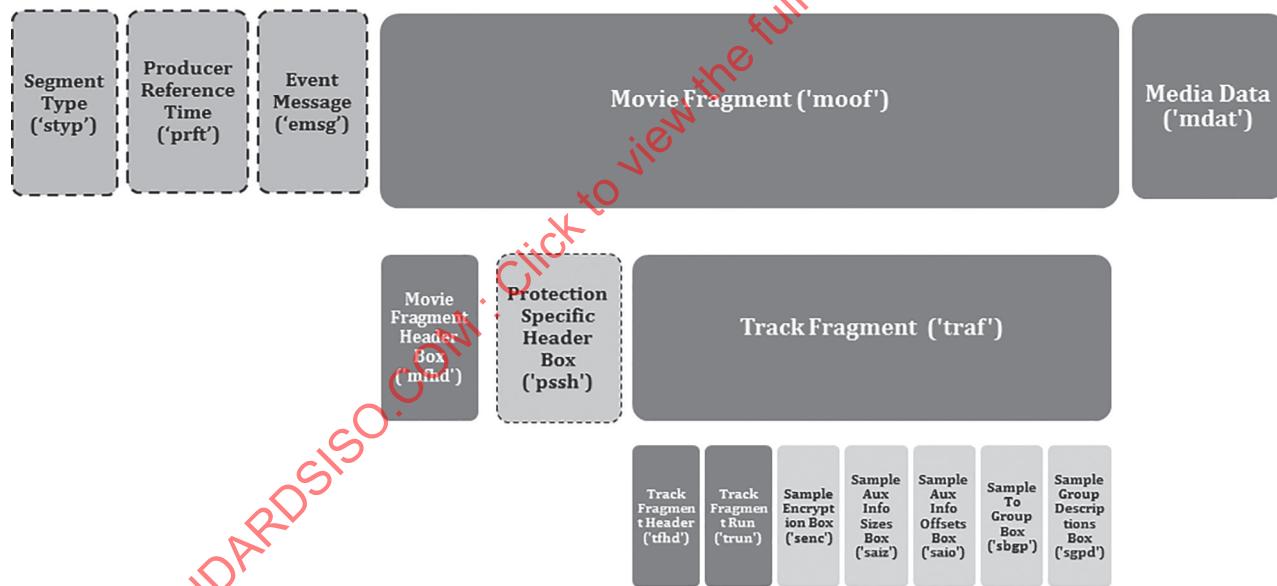
NOTE 2 It is possible that the first CMAF chunk in the first CMAF fragment in a CMAF segment would be preceded by a SegmentTypeBox that includes the compatible\_brands 'cmfl', 'cmff', and 'cmfs' to indicate that it is the start of a CMAF segment, fragment, and chunk. The start of a CMAF fragment that contains multiple CMAF chunks can be identified by including both the 'cmff' and 'cmfl' brands, whereas subsequent CMAF chunks in the CMAF fragment would only include the 'cmfl' brand.

NOTE 3 As noted in [7.3.2.3](#), only a SegmentTypeBox at the start of a CMAF addressable media object is expected to be processed by players, whereas a SegmentTypeBox located within a CMAF addressable media object is expected to be ignored by players.

- f) Each CMAF fragment in a CMAF track should have a duration of at least one second, with the possible exception of the first and last CMAF fragments of the CMAF track.

NOTE 4 Additional constraints specific to video, audio, and subtitle media types are specified in [Clauses 9 to 11](#), and CMAF media profile and codec related CMAF fragment constraints are specified in [Clause 12](#), [Annex A](#) and [Annex B](#).

[Figure 14](#) illustrates the box sequence (left to right) and containment (top to bottom) of a CMAF fragment.



NOTE Lower boxes indicate containment in the box above. The sequence of boxes contained in the TrackFragmentBox ('traf') is only recommended. Light shaded boxes are always optional. Light boxes in the bottom row are required or conditionally required when encryption is used. This example contains one CMAF chunk ('moof', 'mdat') in the CMAF fragment.

**Figure 14 — CMAF fragment box sequence and containment**

### 7.3.3 CMAF addressable media objects

#### 7.3.3.1 CMAF segment

A CMAF segment is defined in [3.1.5](#) to be a “CMAF addressable media object consisting of one or more consecutive CMAF fragments from the same CMAF track”.

**Table 9 — CMAF segment structure**

NL 0	Format Req.	Specification	CMAF constraints	Description
styp	0/1	ISO/IEC 14496-12		Segment Type Signalling compatibility to CMAF Segment
prft	0/1	ISO/IEC 14496-12		Producer Reference Time
emsg	*	ISO/IEC 23009-1	CMAF <a href="#">7.4.5</a>	Event Message
CMAF Fragment	1+	CMAF <a href="#">7.3.2.4</a>		CMAF Fragment

Each CMAF segment conforms to the following constraints.

- A CMAF segment shall include the boxes in [Table 9](#), with the conditions and optionality listed.
- A CMAF segment shall contain one or more complete and consecutive CMAF fragments in decode order.
- A CMAF segment may include a ProducerReferenceTimeBox and/or DASHEventMessageBox(es) preceding the first MovieFragmentBox.
- A CMAF segment may include a SegmentTypeBox preceding all other boxes in the CMAF segment. The SegmentTypeBox should include the CMAF segment brand 'cmfs' and may include any compatible\_brands listed in the FileTypeBox of the CMAF track's CMAF header.

**NOTE** Preceding boxes can be prepended to CMAF chunks or CMAF fragments at time of encoding or prepended to CMAF segments at time of delivery. It is expected that players will process the boxes prepended to a CMAF addressable media object, but they will ignore these boxes when located within a CMAF addressable media object, i.e. not prepended to the first CMAF chunk in a CMAF segment. The SegmentTypeBox is specified by ISO/IEC 14496-12 to be the first box of an ISOBMFF segment and can be ignored in other locations.

### 7.3.3.2 CMAF chunk

A CMAF chunk is specified as a data structure within a CMAF fragment in [7.3.2.3](#), but a CMAF chunk may also be referenced as a CMAF addressable media object, e.g. for low latency live streaming.

There is no direct restriction on the SAP type of the first media sample in a CMAF chunk, except by inheritance of CMAF fragment media sample constraints specified by a CMAF media profile. Because video CMAF media profiles typically constrain the initial media sample type to be SAP 1 or 2, the first CMAF chunk in a CMAF fragment inherits that media sample constraint.

Adaptive switching in players is normally limited to the earliest CMAF chunk in a CMAF fragment. Prepending a SegmentTypeBox to each CMAF chunk with appropriate brands can indicate the start of a CMAF chunk that is also the start of a CMAF fragment.

A SegmentTypeBox, ProducerReferenceTimeBox, and/or DASHEventMessageBox(es) may be prepended to a CMAF chunk at time of encoding or time of delivery. In a CMAF addressable media object, only when these boxes precede the first MovieFragmentBox are they expected to be processed by a player.

### 7.3.3.3 CMAF track file

A CMAF track file is defined in [3.1.6](#) to be “one CMAF track stored consecutively in a single ISOBMFF file with the earliest CMAF fragment constrained to start at decode time zero”.

A CMAF track is a logical collection of CMAF fragments and a CMAF header that might never be stored or stored in that form. In contrast, a CMAF track file is a single file that contains one CMAF track, stored in sequence, and is therefore addressable as a file. Since CMAF chunks and CMAF fragments can be preceded by a `SegmentTypeBox`, CMAF media objects may be identified by `SegmentTypeBox` within a CMAF track. CMAF media objects may be indexed in a `SegmentIndexBox` to enable addressing CMAF media objects in a CMAF track file by byte range.

**Table 10 — CMAF track file structure**

NL 0	Format Req.	Specification	CMAF constraints	Description
CMAF Header	1	CMAF <a href="#">7.3.2.1</a>	CMAF <a href="#">7.2</a>	CMAF Header
sidx	0/1	ISO/IEC 14496-12		Segment Index
CMAF Fragment	1+	CMAF <a href="#">7.3.2.4</a>	CMAF <a href="#">7.3.2.4</a>	CMAF Fragments

Each CMAF track file conforms to the following constraints.

- A CMAF track file shall include the boxes in [Table 10](#), with the conditions and optionality listed.
- Additional boxes, such as `SegmentIndexBox` (es), may be present between the CMAF header and the first CMAF fragment.
- If `SegmentIndexBoxes` exist, each subsegment referenced in the `SegmentIndexBox` shall be a single CMAF fragment contained in the CMAF track file.
- A video CMAF track file may use an offset edit list to adjust the earliest presentation time of the first presented video media sample by subtracting any composition delay added by the use of a v0 `TrackRunBox` that uses positive composition offset values to reorder video frames from decode order to presentation order. See [7.5.13](#).
- A video CMAF track file may use v1 `TrackRunBoxes` using negative composition offsets to adjust the composition time of the earliest presented video media sample in each CMAF fragment to its `baseMediaDecodeTime` and the earliest video media sample in the CMAF track file to zero, without using an offset edit list.

### 7.3.4 CMAF switching sets

#### 7.3.4.1 General constraints on a CMAF switching set

A CMAF switching set is defined in [3.2.2](#) to be a “set of one or more CMAF tracks, where each track is an alternative encoding of the same source content, and are constrained to enable seamless track switching”.

These general CMAF switching set constraints, plus specific constraints defined by a CMAF media profile that the CMAF switching set conforms to, are intended to enable seamless adaptive switching by typical players.

Each CMAF switching set conforms to the following constraints.

- A CMAF switching set shall contain one or more CMAF tracks, each of which is a synchronized encoding of the same source content.
- A CMAF switching set shall contain CMAF tracks of only one media type, i.e. audio or video or subtitles.
- All CMAF tracks in a CMAF switching set shall have the same duration (computed from the media sample durations, optionally stored in a `MovieExtendsHeaderBox`).

- d) All CMAF tracks in a CMAF switching set shall contain the same number of CMAF fragments.
- e) For any CMAF fragment in one CMAF track in a CMAF switching set, there shall be a CMAF fragment with the same decode time in all other CMAF tracks.
- f) All CMAF tracks in a CMAF switching set shall have the same value of `baseMediaDecodeTime` in the first CMAF fragment's `TrackFragmentBaseMediaDecodeTimeBox`, measured from the same timeline origin.
- g) The presentation time of the earliest media sample of the earliest CMAF fragment in each CMAF track shall be equal.
- h) Alternative CMAF fragments in a CMAF switching set with the same `baseMediaDecodeTime` shall contain alternative encodings of the same source media samples.

NOTE 1 These constraints do not prohibit different frame rates in different CMAF tracks in a CMAF switching set. However, these constraints restrict the combinable frame rates to those that result in identical CMAF fragment durations for alternative CMAF fragments.

- i) All CMAF tracks in a CMAF switching set shall conform to one CMAF media profile. However, CMAF tracks included in a CMAF switching set may contain different media profile brands that all conform to one CMAF media profile. See [12.1](#) for additional CMAF switching set constraints specified by CMAF media profiles.

NOTE 2 CMAF media profiles specify the maximum limits for encoding parameters, so subset or lower media profiles that can be decoded by a higher media profile decoder conform to the higher media profile.

- j) CMAF header parameters shall not differ between CMAF tracks, except as allowed in [Table 11](#). CMAF media profiles may specify additional constraints, as noted in the table.
- k) Additional CMAF header boxes specified in [Table 3](#) and [Table 4](#) that are not listed in [Table 11](#) may contain different boxes and information.

**Table 11 — General constraints on CMAF header boxes in CMAF switching sets**

Box	General CMAF header constraints in a CMAF switching set
<code>ftyp</code>	<i>Shall be identical except for media profile brands [see i) in <a href="#">7.3.4.1</a>]</i>
<code>mvhd</code>	<i>Shall be identical except for creation_time and modification_time</i>
<code>tkhd</code>	<i>Shall be identical except for width, height, creation_time, and modification_time. See NOTE 1.</i>
<code>trex</code>	<i>Identical</i>
<code>elst</code>	<i>Shall be identical except for video CMAF track files with a different composition offset</i>
<code>mdhd</code>	<i>Shall be identical except for creation_time and modification_time</i>
<code>mehd</code>	<i>Identical</i>
<code>meta</code>	<i>May contain different boxes and data</i>
<code>udta</code>	<i>May contain different boxes and data</i>
<code>cprt</code>	<i>Identical</i>
<code>kind</code>	<i>Identical</i>
<code>elng</code>	<i>Identical</i>
<code>hdlr</code>	<i>Identical</i>
<code>vmhd</code>	<i>Identical</i>
<code>smhd</code>	<i>Identical</i>

NOTE 1 Track width and height can differ, but picture aspect ratio is the same for all CMAF tracks.

NOTE 2 Sample entry constraints for CMAF switching sets are defined by each CMAF media profile, its specified or referenced ISOBMFF track format, and its CMAF switching set constraints. A sample description can contain multiple sample entries.

**Table 11** (continued)

Box	General CMAF header constraints in a CMAF switching set
sthd	<i>Identical</i>
dref	<i>Identical</i>
stsd	<i>Sample entries shall have the same codingname (four-character code) and conform to other CMAF track format and media profile specified constraints. See NOTE 2.</i>
sgpd	<i>May differ</i>
pssh	<i>Identical</i>
sinf	<i>Identical</i>
schi	<i>Identical</i>
schm	<i>Identical</i>
frma	<i>Identical</i>
tenc	<i>Shall be identical except for initialization vector values</i>

NOTE 1 Track width and height can differ, but picture aspect ratio is the same for all CMAF tracks.

NOTE 2 Sample entry constraints for CMAF switching sets are defined by each CMAF media profile, its specified or referenced ISOBMFF track format, and its CMAF switching set constraints. A sample description can contain multiple sample entries.

There are two functional types of CMAF switching set initialization constraints:

- 1) Multiple initialization — CMAF switching sets conforming to the general constraints above may rely on decoding parameters stored in each CMAF track's header to decode a CMAF fragment from that CMAF track. This implies that a player needs to access and process the CMAF header for each CMAF track it switches to.
- 2) Single initialization — If CMAF headers are constrained to be equivalent within a CMAF switching set, or CMAF tracks only change parameters that are stored in each CMAF fragment, then a player only needs to access one CMAF header and each CMAF fragment to find all the necessary decoding and rendering parameters. A player only needs to initialize the CMAF switching set once, then it can continue decoding CMAF fragments from any CMAF track in the CMAF switching set.

The two types of CMAF switching set constraints and processing are illustrated in [Figure 10](#).

#### 7.3.4.2 CMAF switching set single initialization constraints

Encoding of CMAF tracks in a CMAF switching set can be further constrained to allow one-time initialization of the CMAF switching set using a single CMAF header (see [Figure 10](#)). Each CMAF media profile or the CMAF track format it references can specify single initialization CMAF switching set constraints.

In general, single initialization constraints can be specified by:

- a “common CMAF header” in all CMAF tracks containing all decoding, decryption, and rendering parameters required by any CMAF fragment in the CMAF switching set;
- “inband parameter sets” stored and referenced in each CMAF fragment, where CMAF headers are only used to initialize decoding and display with the appropriate codec, profile, level, resolution, etc. This is typically the case for audio and can be specified for inband video parameter media sample formats such as 'avc3' sample entry;
- a combination of both CMAF header and CMAF fragment stored parameters, but with the constraint that the combination of any CMAF header in the CMAF switching set and CMAF fragment will contain the necessary decoding parameters.

NOTE Enabling changes in inband parameters over time is useful for live streaming, where content from different encodings can be sequenced in a CMAF presentation, such as programs and ads in a live “channel”, or when a live video contribution feed is received containing inband parameters that change and are not known in advance when a CMAF header is packaged and delivered to a player.

### 7.3.4.3 CMAF switching set single initialization constraints identifier

The following identifier shall signal CMAF switching set single initialization constraints.

urn:mpeg:cmaf:siss

### 7.3.4.4 Aligned CMAF switching set constraints

Aligned CMAF switching sets are defined in 3.2.3 to be a “set of CMAF switching sets, the CMAF tracks of which all contain alternative encodings of the same source content in time-aligned CMAF fragments, but all CMAF tracks do not conform to a single CMAF switching set”.

Aligned CMAF switching sets conform to the following constraints.

- a) An aligned CMAF switching set shall contain two or more CMAF switching sets.
- b) Aligned CMAF switching sets shall contain CMAF switching sets that are encoded from the same media stream.
- c) Aligned CMAF switching sets shall contain CMAF switching sets of equal duration.
- d) Aligned CMAF switching sets shall contain the same number of CMAF fragments in every CMAF track.
- e) Aligned CMAF switching sets shall contain CMAF fragments in every CMAF track with matching baseMediaDecodeTime and duration, for each CMAF fragment in a CMAF track.

NOTE Some CMAF players can seamlessly switch between CMAF tracks in different CMAF switching sets that are aligned, such as CMAF switching sets with different encryption keys and their associated DRM license(s).

### 7.3.5 CMAF selection sets

A CMAF selection set is defined in 3.2.4 to be a “set of one or more CMAF switching sets, where each CMAF switching set encodes an alternative aspect of the same presentation over the same time period, only one of which is intended to be played at a time, e.g. an alternative language or codec”.

CMAF selection sets conform to the following constraints.

- a) A CMAF selection set shall contain one or more CMAF switching sets.
- b) All CMAF switching sets within a CMAF selection set shall be of the same media type, e.g. audio, video, or subtitles.
- c) All switching sets within a CMAF selection set shall be of the same duration, within a tolerance of the longest CMAF fragment duration of any CMAF track in the selection set.
- d) Different CMAF switching sets that encode the same content with different codecs or video formats may be contained within a CMAF selection set to enable a player to select the most compatible codec or video format and the CMAF switching set that contains it.
- e) A CMAF selection set may contain multiple aligned CMAF switching sets, which encode the same content in different CMAF switching sets that have time-aligned CMAF fragments, but with different keys, codecs, etc.
- f) One CMAF track should be presented from each selection set in a presentation, e.g. one audio, one video, and one subtitle selection set.

NOTE A subtitle track can be selected, but not displayed by user control, while IMSC1 titles marked as “forced” are automatically displayed.

### 7.3.6 CMAF presentations

A CMAF presentation is defined in [3.2.5](#) to be a “set of one or more CMAF selection sets that can be simultaneously decoded to produce a multimedia user experience, potentially including synchronized audio, video, and subtitles”.

A CMAF presentation conforms to the following constraints.

- a) All CMAF tracks in a CMAF presentation shall have the same timeline origin.
- b) Different CMAF switching sets may use different CMAF track timescale values. In that case, `baseMediaDecodeTime` integer values will be different for the same presentation time. But they are “equivalent” presentation times when dividing by each track’s timescale yields the same presentation time value.
- c) CMAF fragment presentation time delays relative to decode time shall be removed using an offset edit list (see [7.5.13](#)), negative composition offsets (see [7.5.16](#)) or encoder compensation (see [10.3.3.2](#)).
- d) All CMAF tracks in a CMAF presentation containing a video switching set shall be start aligned with CMAF presentation time zero equal to the earliest video media sample presentation start time in the earliest CMAF fragment (see [6.6.8](#)).
- e) All CMAF tracks in a CMAF presentation that does not contain video shall be start aligned with the CMAF presentation time zero equal to the earliest audio media sample presentation start time in the earliest CMAF fragment.
- f) Media samples in audio and subtitle CMAF tracks with earlier presentation times than CMAF presentation time zero should not be presented.
- g) Media samples in audio and subtitle CMAF tracks whose durations overlap the earliest video media sample presentation time should be partially presented, starting at the earliest video media sample presentation time.
- h) The duration of a CMAF presentation shall be the duration of its longest CMAF track.
- i) CMAF tracks in a CMAF presentation shall equal the CMAF presentation duration, within a tolerance of the longest video CMAF fragment duration.

NOTE The CMAF hypothetical application model synchronizes media samples by presentation times relative to a common timeline origin. CMAF track presentation time is determined by each CMAF fragment’s stored `baseMediaDecodeTime` and any composition or edit list offsets present. Synchronization during late binding does not rely on externally stored presentation time offsets. This means that audio, video and subtitle media samples that are intended to be played simultaneously in a CMAF presentation need to have overlapping presentation time ranges, where the presentation time range starts at the media sample presentation time ([6.2](#)) and ends at the media sample presentation time plus the media sample duration.

## 7.4 Additional boxes, not defined in the ISO Base Media File Format

### 7.4.1 Track Encryption Box ('tenc')

The `TrackEncryptionBox` ('tenc') specified in ISO/IEC 23001-7 indicates that media samples in the track might be encrypted using a specific scheme of Common Encryption, identified by a four-character code, such as 'cenc' or 'cbc'. The `TrackEncryptionBox` contains parameters and default values that apply to an entire track. See [Clause 8](#) for additional information.

#### 7.4.2 Sample Encryption Box ('senc')

When Common Encryption sample auxiliary information is used:

- it shall be present in each CMAF fragment;
- the `SampleEncryptionBox` should be used to store sample auxiliary information.

See [Clause 8](#) for more information.

#### 7.4.3 Protection System Specific Header Box ('pssh')

The `ProtectionSystemSpecificHeaderBox` is specified in the Common Encryption specification (ISO/IEC 23001-7) for the storage of content protection system information such as license acquisition information and DRM licenses. This box was primarily designed to store license information in download files. Either version 1 or version zero may be used, as constrained in [Clause 8](#).

CMAF streaming applications should signal license acquisition information in the manifest and should not duplicate the information in this box in CMAF headers.

CMAF fragments may contain DRM information in this box intended for delivery to all players during playback.

See [Clause 8](#) for more information.

#### 7.4.4 Media profile specific boxes

Audio and video track formats typically derive sample entries and define decoder configuration boxes that can be required by one or more CMAF media profiles. For example, the `AVCConfigurationBox` is defined in ISO/IEC 14496-15 and required by the NAL structured video CMAF track format specified in [Clause 9](#) and referenced by the CMAF video media profiles specified in [Annex A](#). Other CMAF media profiles may specify boxes that are required by a '`cmfc`' conformant CMAF track that also conforms to the CMAF media profile brand.

#### 7.4.5 Event Message Box ('emsg')

The enhanced `DASHEventMessageBox` ('`emsg`') described in this clause is version 1 of the `DASHEventMessageBox`, specified in ISO/IEC 23009-1.

Version 1 of this box adds the field `presentation_time`, which makes event message timing independent of box location in the CMAF track. Version 1 should be used for event messages in CMAF fragments and addressable media objects.

`presentation_time` provides the presentation time of the event measured on the CMAF track's presentation timeline, in the timescale declared in its `MovieHeaderBox`.

`message_data` is the body of the event message. The syntax and semantics of this field are defined by the owner of the scheme identified in the `scheme_id_uri` field. Message schemes may be defined for specific applications and users, or standardized for global use, such as SCTE-35 advertisement and program segmentation markers.

NOTE 1 `DASHEventMessageBoxes` can be included in CMAF fragments to indicate ad insertion points, etc. in the media stream, then other `DASHEventMessageBoxes` added to CMAF segments or CMAF chunks at time of delivery, e.g. to trigger manifest updates.

A `DASHEventMessageBox` in a CMAF track shall contain the value in its `timescale` field equal to the value of the `timescale` field in the `MediaHeaderBox` of the CMAF track that contains it.

If version 0 is used, then DASH defines the timing of an Event Message related to the earliest media sample presentation time of a DASH segment using the field `presentation_time_delta`, which "provides the media presentation time delta of the media presentation time of the event and the earliest

presentation time in this segment." For CMAF fragments, the presentation\_time\_delta shall equal the media presentation time of the event minus the earliest presentation time of the following CMAF fragment.

NOTE 2 The earliest decode time in a CMAF track file is zero (defragmented or not), and if an edit list is present, the earliest presentation time is the earliest media sample composition time adjusted by the edit list offset.

See [Annex E](#) for more information on the use of event messages.

## 7.5 Constraints on ISO Base Media File Format boxes

### 7.5.1 Movie Header Box ('mvhd')

In the MovieHeaderBox, the value of the duration field should be set to zero to indicate that the MovieBox contains no media samples and therefore has no duration.

NOTE The duration field in the MediaHeaderBox ('mdhd') applies to the TrackBox ('trak'), which contains no media samples in a CMAF track. The duration of a CMAF track can optionally be stored in the fragment\_duration field of the MovieExtendsHeaderBox ('mehd'), which is equal to the sum of all CMAF fragment durations in the CMAF track. If the duration is unknown, this box is omitted.

The fields rate, volume, and matrix shall be set to their default values.

### 7.5.2 Metadata Boxes

Metadata, carried in either UserDataBox or MetaBoxes, may be present. When present, they shall not occur at file level, i.e. they can only be contained in another box, as permitted by ISO/IEC 14496-12.

### 7.5.3 Kind Box ('kind')

The KindBox ('kind') may be used to store the role of a CMAF track. The KindBox is stored in the UserDataBox of the TrackBox, as documented in the ISO Base Media File Format (ISO/IEC 14496-12).

Any track can be labelled with role information describing the intended purpose of the track. This information can be captured at the time of encoding and later copied to a manifest describing the CMAF tracks in a selection set so that a user or an automatic algorithm can make an appropriate selection.

The KindBox can contain one or more tags from a variety of places, including:

- the DASH specification ISO/IEC 23009-1, as identified by the schemeURI "urn:mpeg:dash:role:2011" (without the quotation marks);
- the W3C HTML5 specification of track 'kind', as identified by the schemeURI "about:html-kind".

Where multiple schemes define the same concepts, the DASH Role scheme should be used.

### 7.5.4 Track Header Box ('tkhd')

CMAF TrackHeaderBoxes shall conform to ISO/IEC 14496-12 with the following additional constraints.

- The field duration shall be set to a value of zero ('0'), indicating no media samples are referenced from the TrackBox ('trak').
- The field matrix shall be set to their default values as defined in ISO/IEC 14496-12, except to indicate video orientation (i.e. portrait or landscape orientation relative to the captured scene). See [9.2.3](#).

- The following fields shall be set to default values as defined in ISO/IEC 14496-12, unless specified otherwise in this document.
  - The layer field should equal 0 or greater for normally presented video tracks.
  - The layer field should equal -1 for subtitle tracks so they are normally presented over the video.
  - The width and height fields for a non-visual track (e.g. audio) shall be 0.
  - As defined in ISO/IEC 14496-12, the width and height fields for a CMAF video track shall specify the track's normalized presentation size as fixed-point 16.16 values expressed in square pixels after decoder cropping, and in the case of video encoded with a non-square video spatial sample shape, after horizontal scaling has been applied. See [9.2.3](#) for normalized width and height calculation.
  - Subtitle tracks may set width and height to an intended layout size, in which case the text layout engine or graphics engine can scale the width and height to match the video display aperture (player implementation dependent).
  - As defined in ISO/IEC 14496-30, subtitle tracks encoded as text may use relative position coordinates and font sizes so that the text layout engine can adjust glyph and layout size to match the final video display aperture without relying on image scaling. For such tracks, the value of zero width and height should be used to indicate that the data can be rendered at any size, and the layout size may be determined by matching the size of the video display aperture.
  - For scalable text and subtitle tracks, the flag `track_size_is_aspect_ratio` may also be used.

### 7.5.5 Media Header Box ('mdhd')

The CMAF MediaHeaderBoxes shall conform to ISO/IEC 14496-12 with the following additional constraints.

- The value of the `duration` field should be set to a value of zero ('0') (see [7.5.1](#)).
- Where possible, the value of the `timescale` field should be chosen such that when the frame rate is constant, the value of the media sample duration may also be constant.
- All tracks that are language-specific should identify the language as precisely as possible (e.g. a text track whose language can be written in different scripts should identify which script is used). When the language is not relevant or not known, the 'und' (undetermined) language tag should be used.

### 7.5.6 Video Media Header Box ('vmhd')

The VideoMediaHeaderBox shall conform to ISO/IEC 14496-12 with the constraints specified in [9.2.2](#).

### 7.5.7 Sound Media Header Box ('smhd')

The SoundMediaHeaderBox shall conform to ISO/IEC 14496-12 and the following constraint.

The field `balance` shall equal 0 (centre).

### 7.5.8 Subtitle Media Header Box ('sthd')

The SubtitleMediaHeaderBox shall conform to ISO/IEC 14496-12 with constraints specified in ISO/IEC 14496-30.

NOTE Subtitle media uses the 'subt' handler\_type in the HandlerBox in the MediaHeaderBox.

### 7.5.9 Data Reference Box ('dref')

DataReferenceBoxes in a CMAF track shall conform to ISO/IEC 14496-12 with the following additional constraints.

- The DataReferenceBox shall contain a single entry with the entry\_flags field set to 0x000001 (which means that the media data is in the same file as the MovieBox containing this data reference).

### 7.5.10 Sample Description Box ('stsd')

The SampleDescriptionBox in a CMAF track shall conform to version 0 as defined in ISO/IEC 14496-12 with the following additional constraints.

- Sample entries for encrypted tracks (those containing any encrypted media sample data) shall encapsulate the existing sample entry with the appropriate four-character-code listed in ISO/IEC 14496-12 and include a ProtectionSchemeInfoBox ('sinf') that conforms to ISO/IEC 14496-12 and [Clause 8](#).
- Constraints on visual sample entries are specified in [9.2.4](#).
- Constraints on audio sample entries are specified in [10.2.5](#).
- Constraints on subtitle sample entries are specified in ISO/IEC 14496-30.

NOTE A CMAF SampleDescriptionBox can contain multiple sample entries.

### 7.5.11 Protection Scheme Information Box ('sinf')

CMAF shall use Common Encryption (ISO/IEC 23001-7) for CMAF tracks containing one or more encrypted CMAF fragments and use Scheme Signalling as defined in ISO/IEC 23001-7. An encrypted CMAF track shall include at least one ProtectionSchemeInfoBox ('sinf') containing a TrackEncryptionBox ('tenc') identifying a scheme specified in ISO/IEC 23001-7.

### 7.5.12 Track contained media sample information boxes

All boxes in the SampleTableBox have a sample count of 0 because CMAF does not reference media samples from the TrackBox. The mandatory boxes of ISO/IEC 14496-12 are mandatory, even if they document no samples.

The following boxes therefore shall have an entry\_count of zero:

- TimeToSampleBox ('stts');
- SampleToChunkBox ('stsc');
- ChunkOffsetBox ('stco');
- SampleSizeBox or CompactSampleSizeBox ('stsz' or 'stz2');
- SyncSampleBox ('stss'), if present.

NOTE 1 The presence of an empty SyncSampleBox in a CMAF header indicates that not all media samples in the CMAF track are sync samples.

NOTE 2 Media sample size, duration, and dependency information can be found in the TrackRunBox (es) in each CMAF fragment or CMAF chunk.

### 7.5.13 Edit List Box ('elst')

In two cases, an `EditBox` containing a single `EditListBox` with the following constraints may be present in the CMAF header of a CMAF track to adjust the presentation time of all media samples in the CMAF track.

- a) The first case is a video CMAF track file using `v0 TrackRunBoxes` with positive composition offsets to reorder video media samples, in which case an `EditListBox` shall be used to adjust the earliest video sample to ISO/IEC 14496-12 "movie" presentation time zero.
- b) The second case is an audio CMAF track where each media sample's presentation time does not equal its composition time. In that case, the CMAF track shall include an `EditListBox` with a value of `media_time` equal to the CMAF track presentation time delay.

See [7.3.6](#) and [Figure 12](#) for information on CMAF timelines and start alignment.

- The `EditBox` shall contain a single `EditListBox`.
- The value of `entry_count` field in the `EditListBox` shall be set to 1.
- The value of the `media_rate_integer` field shall be set to 1 and the value of the `media_rate_fraction` field shall be set to 0.
- The value of the `segment_duration` field shall be set to 0.
- The value of `media_time` = sample presentation time - sample composition time (sample composition time = sample decode time + sample composition offset).

NOTE Since there is no media in the `MovieBox`, the duration in the edit is zero. However, as noted in ISO/IEC 14496-12, movie fragments implicitly extend any edit represented by an edit list in the `MovieBox`. This means that any offset edit list in the CMAF header is applied to the composition times of all samples to determine each sample's presentation time.

According to ISOBMFF, samples with a sum of movie presentation time and duration less than zero are not meant to be presented.

Using offset edits to set the presentation time of the earliest presented sample of each CMAF track to an equivalent CMAF track presentation time enables synchronization of all CMAF tracks to a common CMAF presentation timeline.

Additional constraints specific to certain media types are in the following subclauses:

- audio edit lists in [10.2.6](#);
- video edit lists in [9.2.7](#).

### 7.5.14 Track Extends Box ('trex')

A `TrackExtendsBox` shall be present in a CMAF track since it is a fragmented file as defined in ISO/IEC 14496-12.

CMAF track `default_sample_flags` shall not be necessary to determine sample dependencies in CMAF fragments or CMAF chunks. `default_sample_flags`, `sample_flags` and `first_sample_flags` shall be set in the `TrackFragmentHeaderBox` and/or `TrackRunBox` to provide sample dependency information within each CMAF chunk and CMAF fragment.

CMAF track default sample parameters shall not be necessary to determine media sample duration, size, or sample description index. Default values or per sample values shall be stored in each CMAF chunk's `TrackFragmentHeaderBox` and/or `TrackRunBox`.

NOTE Default flags and sample parameters can be set and ignored in the `TrackExtendsBox`, as long as those values are also set in all CMAF chunks and CMAF fragments so each CMAF fragment is decodable without access to that track's CMAF header.

### 7.5.15 Movie Fragment Header Box ('mfhd')

A MovieFragmentHeaderBox in a CMAF track shall conform to ISO/IEC 14496-12.

NOTE The sequence\_number integer value is not required to be unique within a CMAF track nor to increase with decode time.

### 7.5.16 Track Fragment Header Box ('tfhd')

A TrackFragmentHeaderBox in a CMAF track shall conform to ISO/IEC 14496-12 with the following additional constraints.

- The track\_ID field shall contain the same value as the track\_ID in the associated CMAF header.
- The base-data-offset-present flag (in the tf\_flags field) shall be set to zero.
- The default-base-is-moof flag (in the tf\_flags field) shall be set to one.
- Every TrackFragmentBox shall contain a TrackFragmentBaseMediaDecodeTimeBox, as defined in ISO/IEC 14496-12, to provide the decode time of the first media sample in the track fragment.

NOTE The baseMediaDecodeTime of the first available CMAF fragment in a CMAF track can be non-zero.

### 7.5.17 Track Run Box ('trun')

A TrackRunBox in a CMAF track shall conform to ISO/IEC 14496-12 with the following additional constraints.

- Version 1 shall be used for video CMAF tracks, except in the case of a video CMAF track file, in which case either version 1 or version 0 may be used. Other CMAF track types should use version 0.
- When the version field is set to "1", the sample\_composition\_time\_offset is a signed 32-bit integer measured by the timescale of the track, which reorders video media samples from decode order to presentation order, and in that case, the composition time of the first presented media sample in a CMAF fragment shall be the same as the first media sample decode time (baseMediaDecodeTime).
- The data-offset-present flag (in the tf\_flags field) shall be set to true in order to indicate that the data\_offset field is present and contains the byte offset from the start of this fragment's MovieFragmentBox to the start of the first media sample in the following MediaDataBox.

NOTE This is called movie fragment relative addressing in ISO/IEC 14496-12.

- default\_sample\_flags, sample\_flags and first\_sample\_flags shall be set in the TrackFragmentHeaderBox and/or TrackRunBox to provide sample dependency information within each CMAF chunk and CMAF fragment.
- Default values or per sample values of sample duration and sample size shall be stored in each CMAF chunk's TrackRunBox and/or TrackFragmentHeaderBox.
- Within a CMAF track, any TrackRunBox that describes any non-sync media samples shall identify sample dependency with the CMAF chunk and CMAF fragment using a combination of the sample\_flags and first\_sample\_flags fields and default values in the TrackFragmentHeaderBox:
  - sample\_is\_non\_sync\_sample shall be 0 for SAP type 1 or 2, and 1 otherwise;
  - an empty SyncSampleBox shall be present in the track.

NOTE ISO/IEC 14496-12 specifies that absence of the SyncSampleBox indicates that all media samples are sync samples in the track, which allows a reader to know that all subsequent CMAF fragments will also consist of sync samples. If a SyncSampleBox is present, then dependency flags in each CMAF fragment indicate which media samples are sync samples, since the header contains no media samples and the SyncSampleBox therefore lists no media samples.

### 7.5.18 Sample Group Description Box ('sgpd')

As required by ISO/IEC 14496-12, when sample group descriptions can change within a CMAF track, a SampleGroupDescriptionBox shall be stored in each CMAF fragment that references that sample group description. If sample group information is the same for all CMAF fragments in a CMAF switching set, it may be stored in a SampleGroupDescriptionBox in the CMAF header SampleTableBox ('stbl'). Either version 1 or 2 may be used (version 0 is deprecated by ISOBMFF).

EXAMPLE When Common Encryption is used and KID values can change per CMAF fragment, a SampleToGroupBox stored in each TrackFragmentBox will reference a SampleGroupDescriptionBox containing the KID for that sample group, and both are stored in the TrackFragmentBox in order to support random access.

Pre-roll sample groups can be used for some audio as described in [10.3.3.1](#).

### 7.5.19 Media Data Box ('mdat')

Each CMAF fragment shall contain one or more MediaDataBox(es) ('mdat') containing media samples. The MediaDataBox conforms to the definition in ISO/IEC 14496-12.

Each MediaDataBox in a CMAF chunk shall be immediately preceded by the MovieFragmentBox that references the media samples it contains.

### 7.5.20 Sub-sample Information Box ('subs')

Each CMAF fragment in a TTML image subtitle track of CMAF media profile 'im1i', as specified in [11.3.4](#), shall contain a SubSampleInformationBox in the TrackFragmentBox that indexes any images contained in the subtitle media sample. The field entry\_count shall equal 1.

When one or more images are present in the subtitle media sample, the value of subsample\_count shall equal 1 for the first image sub-sample, and the subsample\_count of the TTML document shall equal 0.

When no images are present in the subtitle media sample, the value of subsample\_count shall equal 0.

Version 0 of this box should be used, unless subsample\_size exceeds a 16-bit byte address size (65 kibibytes), in which case version 1 should be used.

The fields subsample\_priority, discardable, and codec\_specific\_parameters are undefined.

## 8 Common Encryption of CMAF tracks

### 8.1 Multiple DRM system support

Common Encryption (ISO/IEC 23001-7) enables multiple DRM systems to provide a license for an encrypted CMAF track. The default\_KID identifies the key and license required, and a registered SystemID identifies a Common Encryption capable DRM system. License acquisition information can be provided to identify which DRM systems can provide licenses. License acquisition information usually includes the URL of an authorization and license server, DRM SystemID, DRM client identification, type of license requested, media key identifier, etc. and may be stored in a ProtectionSystemSpecificHeaderBox, a manifest, or an application in order to assist players in requesting a license to decrypt a CMAF track.

A single key and license may be sufficient to access all tracks in a presentation, or high value content may use different keys for audio and video tracks, since the audio path may be less secure. A content provider may also use different keys and possibly licenses for different qualities, such as SD, HD, and UHD.

When streaming, it is recommended that any license acquisition information used to acquire CMAF decryption key(s) be signalled in a manifest or application, and not in the CMAF header. This enables a CMAF player application to parse license acquisition information one time, rather than each time the CMAF header is processed for adaptive switching purposes.

Manifest signalling makes it easier to:

- add or change license information without editing media files;
- offer different types of licenses for the same CMAF presentation, e.g. subscription, rental, ownership, SD, HD, UHD, etc.;
- support different distribution channels and license servers with the same media;
- request a license before the live media becomes available to authorize or purchase playback rights and download a license in advance.

## 8.2 Track encryption

### 8.2.1 General requirements

Encrypted media sample data in a CMAF track shall use an encryption scheme defined in ISO/IEC 23001-7. CMAF presentation profiles can constrain the allowable schemes in that CMAF presentation profile. See [A.1](#).

Encrypted NAL structured video tracks shall use a media subsample encryption scheme specified in ISO/IEC 23001-7, which defines partial encryption of NAL units in a video media sample to allow access to unencrypted video NAL headers and slice headers in an encrypted NAL structured video elementary stream.

Encrypted non-video tracks shall use the schemes specified in ISO/IEC 23001-7, which define a full media sample encryption method for each scheme.

Conditional requirements are defined in ISO/IEC 23001-7 and [8.2.2.1](#) in this document to include a `SampleAuxiliaryInformationOffsetsBox` and a `SampleAuxiliaryInformationSizesBox` in the `TrackFragmentBox`. As specified in ISO/IEC 14496-12, both boxes have a default `aux_info_type` of 'cenc'.

The `SampleEncryptionBox` ('senc') for storage of sample auxiliary information in encrypted CMAF tracks is documented in [7.4.2](#) and its use specified in [8.2.2.1](#).

The `TrackEncryptionBox` version 1 or zero for signalling CMAF track encryption parameters and their defaults is documented in [7.4.1](#) and its use specified in [8.2.2.2](#).

The following additional constraints apply to all encrypted CMAF tracks.

- All key identifier values shall uniquely identify one and only one key within their scope of use. To ensure this level of uniqueness, it is strongly recommended that the key identifier values be UUIDs generated according to X.667. A UUID shall be stored in the KID field as 16 bytes, sequenced as specified in X.667:2014, 6.2.
- A KID value may be represented in text form as a hyphenated hexadecimal string, as specified in X.667:2014, 6.4.

## 8.2.2 CMAF track constraints

### 8.2.2.1 Sample Encryption Box ('senc') and sample auxiliary information

Sample auxiliary information, such as per media sample initialization vectors and subsample byte ranges, should be stored in a `SampleEncryptionBox ('senc')` as described in 7.4.2. Sample auxiliary information may be stored in any valid box within a CMAF chunk's or CMAF fragment's `MovieFragmentBox`, such as a '`uuid`' box.

For encrypted CMAF fragments that contain sample auxiliary information, each `TrackFragmentBox` shall contain a `SampleAuxiliaryInformationOffsetsBox` with an `aux_info_type` value or default of '`cenc`' as defined in ISO/IEC 23001-7 to provide access to sample auxiliary information data.

The `SampleAuxiliaryInformationOffsetsBox` shall conform to the following constraints.

- a) The `entry_count` field of the `SampleAuxiliaryInformationOffsetsBox` shall equal 1. Therefore, data in the `SampleEncryptionBox` or other sample auxiliary information storage box shall be contiguous for all of the media samples in the movie fragment.
- b) CMAF movie fragments use movie fragment relative addressing where no base data offset is provided in the track fragment header. Therefore, the `offset` field is calculated as the difference between the first byte of the containing `MovieFragmentBox` and the first byte of the first `InitializationVector` in the sample auxiliary information.

The `SampleAuxiliaryInformationSizesBox` may specify the size of the sample auxiliary data for each media sample with a type of '`cenc`', as defined in ISO/IEC 23001-7, with the following constraints.

- 1) If subsample encryption is not used, the `SampleAuxiliaryInformationSizesBox` may be omitted.

NOTE 1 For full media sample encryption with an IV per media sample, the size of the sample auxiliary information equals `default_Per_Sample_IV_Size` in the `TrackEncryptionBox` (see ISO/IEC 23001-7).

- 2) If subsample encryption is used and all the media samples have the same number of subsamples, then the size of the sample auxiliary information will be the same for all of the media samples. In that case, the `default_sample_info_size` of the `SampleAuxiliaryInformationSizesBox ('saiz')` may be used instead of storing a size per media sample.
- 3) If `Per_Sample_IV_Size` is also zero (because the scheme uses constant IVs and no subsamples), then there would be no sample auxiliary information, and the `SampleEncryptionBox`, `SampleAuxiliaryInformationSizesBox`, and `SampleAuxiliaryInformationOffsetsBox` should be omitted.

NOTE 2 Sample auxiliary information is located by movie fragment byte offsets stored in the `SampleAuxiliaryInformationOffsetsBox`, and in some cases, by size information stored in the `SampleAuxiliaryInformationSizesBox`. Sample auxiliary information, such as per-sample initialization vectors and subsample byte ranges, is not intended to be read directly from the `SampleEncryptionBox` or other storage box. The `SampleAuxiliaryInformationOffsetsBox` should always be used to locate sample auxiliary information.

### 8.2.2.2 Track Encryption Box ('tenc')

A `TrackEncryptionBox` specified in 7.4.1 shall be present in a CMAF header if any media samples in the track are encrypted.

### 8.2.2.3 Protection System Specific Header Box ('pssh')

Common Encryption specifies version zero and version one `ProtectionSystemSpecificHeaderBox`. `ProtectionSystemSpecificHeaderBoxes` can be used to store licenses in downloaded files, signal in CMAF headers whether license downloads may be needed, and deliver licenses, keys,

and usage information in CMAF fragments. `ProtectionSystemSpecificHeaderBoxes` contain a `SystemID` that uniquely identifies the protection system intended to use the information. Information contained in the `data[]` array is considered opaque to players, file parsers, and other DRM systems and might be encrypted by the protection system.

Common Encryption (ISO/IEC 23001-7) also specifies XML elements to contain license acquisition information for use in manifests. For CMAF presentations, manifest signalling is recommended.

**NOTE** A player can download all licenses that will be needed for playback as soon as it parses a manifest and before downloading any CMAF headers. It is particularly useful to acquire licenses in advance of a large live streaming event that would result in a large number of synchronized license requests when triggered by the simultaneous arrival of the first CMAF headers or CMAF fragments. Purchases, etc. can be completed prior to media availability.

The following constraints apply to `ProtectionSystemSpecificHeaderBox`.

- `ProtectionSystemSpecificHeaderBoxes` should not be present in CMAF headers, except as allowed below, and may be ignored if present.
- A common `ProtectionSystemSpecificHeaderBox` and additional `ProtectionSystemSpecificHeaderBoxes` in CMAF header may be used with presentation-specific playback applications that can read the KID from the common '`pssh`' and implement application-specific license management. A common `ProtectionSystemSpecificHeaderBox` is defined by W3C Encrypted Media Extension specification as a version one `ProtectionSystemSpecificHeaderBox` with a W3C designated `SystemID`, and the track's `default_KID` listed in the version one `ProtectionSystemSpecificHeaderBox` `KID` array, but it contains no data in the `data[]` array.
- Version one `ProtectionSystemSpecificHeaderBoxes` may be present in CMAF fragments for delivering protection system information, such as encrypted keys in licenses, for use by the DRM system with a `SystemID` matching the `ProtectionSystemSpecificHeaderBox` `SystemID`.
- Multiple `ProtectionSystemSpecificHeaderBoxes` with different `SystemIDs` may be present to enable different protection systems on different devices.

For example, version one `ProtectionSystemSpecificHeaderBoxes` in CMAF fragments can be used to deliver the same encrypted licenses to all players by a streaming or broadcast service, but only users who have purchased and downloaded an entitlement license bound to their particular player can decrypt the licenses in the CMAF fragments and decrypt media samples with the keys in the inband licenses. Since one license decrypts the other, they are said to be "chained", and chained licenses in combination with key changes can periodically verify that each player has a valid entitlement license to decrypt the inband licenses and continue the presentation (this process is often called "key rotation").

### 8.2.3 Encryption constraints

#### 8.2.3.1 General

- For a given KID, initialization vectors and counter values shall follow the guidelines outlined in ISO/IEC 23001-7, including the requirement that the combination of values only be used on one cipher block.
- Initialization vectors for use with the '`cenc`' scheme shall conform to ISO/IEC 23001-7 and shall be limited to 8-bytes to avoid block counter value overlap.
- Each KID and `default_KID` shall never reference more than one key value for all CMAF tracks in a CMAF presentation.

**NOTE** It is possible for two different KIDs to reference the same key value.

- All CMAF tracks in a CMAF switching set shall use the same `default_KID` and key value. Any additional keys described by sample groups may be stored in version 1 `ProtectionSystemSpec`

ificHeaderBoxes in CMAF fragments, identifying the contained KID(s), protected by DRM specific methods.

- It is recommended that any textual representation of KID and SystemID (e.g. in manifests) uses the hexadecimal string representation specified in X.667:2014 6.4, derived from the 16 byte binary representation specified in X.667:2014 6.2 and equivalent byte arrays specified in ISO/IEC 23001-7 boxes.

The following additional constraints shall be applied to the encryption of NAL structured video tracks.

- Slice headers, NAL type headers, NAL size headers, and all non-video NALs shall be unencrypted using subsample encryption, as required by ISO/IEC 23001-7.

This constraint is recommended in CENC, but required in CMAF.

- Video VCL data shall be protected by subsample encryption, with bytesOfProtectedData spanning all complete 16-byte blocks in the VCL slice data.
- 'cenc' scheme bytesOfProtectedData shall be a multiple of 16 bytes.

This constraint is recommended in CENC, but required in CMAF.

- 'cbc5' scheme bytesOfProtectedData shall start on the first complete byte of video data following the slice header and end on the end of the last complete 16-byte block of slice data in the video NAL unit.

These are constraints that are recommended in CENC, but required in CMAF.

### 8.2.3.2 Clear samples within an encrypted CMAF track

In an encrypted track, the isProtected flag in the TrackEncryptionBox shall be set to 1, indicating that all media samples are protected by default. Sample groups may indicate unprotected media samples, as specified in ISO/IEC 23001-7.

Each CMAF fragment shall be constrained to media samples that are all protected or all unprotected, not a mix.

### 8.2.4 CMAF presentation encryption

It is recommended that encryption keys not be shared between audio and video CMAF switching sets in a protected CMAF presentation. Audio decoding and decryption systems may have lower key security than video decoding systems so should not use and expose a key also used for video.

Premium UHD/HDR content may require hardware security for keys, a secure video path, secure output interfaces, etc. only available on some devices and DRM systems. Lower resolution video content with different keys could enable playback on devices that are less secure.

Standard definition, high definition, and/or ultra-high definition content in one CMAF presentation can be sold separately or accessed on different devices with different DRM licenses, keys, and DRM playback requirements. Separate CMAF switching sets for SD, HD and UHD may be functionally combined and possibly seamlessly switched using aligned CMAF switching sets, as specified in [7.3.4.4](#).

## 9 Video CMAF tracks

### 9.1 Overview

CMAF video tracks are derived from the general CMAF track format specified in [Clauses 7](#) and [8](#).

This clause specifies general constraints on all video CMAF tracks and video CMAF switching sets, then derives more specific constraints for NAL structured video from that. Then, AVC CMAF track format is derived from NAL structured video CMAF tracks.

Video media profiles can be derived from the general video track format, NAL structured video track format, or AVC video track format, as appropriate.

This clause specifies:

- general constraints on a video CMAF track, in addition to conforming to the '`cmfc`' brand;
- general constraints on multiple CMAF tracks conforming to a CMAF switching set intended to enable seamless adaptive switching of video;
- NAL structured video constraints for CMAF switching sets;
- single initialization constraints for NAL structured video CMAF switching sets;
- constraints between aligned CMAF switching sets containing NAL structured video;
- AVC video CMAF track constraints, used to derive the AVC video CMAF media profiles defined in [A.2](#).

The CMAF track format for NAL structured video is derived from ISO/IEC 14496-15 and requires boxes defined in that specification.

Some CMAF video media profiles using the AVC codec are specified in [Annex A](#). They define widely used video formats, such as SD, HD, and UHD, by specifying codec parameters, such as the maximum codec profile and level that can be encoded, and video parameters such as maximum resolution, frame rate, bit depth, transfer function, colour space, colour subsampling, etc.

Video CMAF tracks that carry a compatibility brand indicating parameter limits that also conform to a higher CMAF media profile (i.e. a lower codec profile and level) are considered to also conform to that higher CMAF media profile. For example, a CMAF track with an SD (standard definition) AVC compatibility brand also conforms to a CMAF HD (high definition) AVC media profile. In general, a media profile is considered contained in another media profile if it can always be decoded and rendered by all decoders conforming to the containing media profile.

Other CMAF media profiles for video codecs may be defined in other specifications, as long as they conform to CMAF track constraints specified in [Clauses 7](#) and [8](#) and recommendations in [Clause 12](#). Video media profiles may optionally specify CMAF track constraints for CMAF switching sets to enable seamless adaptive switching and may optionally specify single initialization CMAF switching set constraints.

For additional information, see [Clause 12](#) and [Annex A](#).

## 9.2 General video CMAF track format

### 9.2.1 General video CMAF track structure and constraints

Video CMAF tracks shall conform to the '`cmfc`' compatibility brand as specified in [Clauses 7](#) and [8](#) and general CMAF video track constraints specified in [9.2](#).

### 9.2.2 Video Media Header ('`vmhd`')

`VideoMediaHeaderBoxes` in a video CMAF track shall conform to ISO/IEC 14496-12 with the following additional constraints.

- The following fields shall be set to their default values as defined in ISO/IEC 14496-12:
  - `version=0`;

- graphicsmode=0;
- opcolor={0, 0, 0}.

### 9.2.3 Track Header Box ('tkhd')

For video tracks, the fields of the `TrackHeaderBox` shall be set to the values constrained below and specified in ISO/IEC 14496-12.

- flags = 0x0000007
- The values of `width` and `height` (specified in ISO/IEC 14496-12 as the decoded and cropped image size in video spatial samples measured on a uniformly sampled square grid) in a CMAF `TrackHeaderBox` shall be normalized to width and height of the encoded video, as defined below:
  - The normalized presentation height shall be the number of vertical video spatial samples after codec cropping parameters are applied.

NOTE 1 The `height` field of the visual sample entry is the number of encoded vertical video spatial samples after cropping.

- The normalized presentation `width` shall be the number of horizontal video spatial samples after codec cropping parameters are applied, then multiplied by the video spatial sample aspect ratio. The video spatial sample aspect ratio is defined by the `PixelAspectRatioBox`, if present in the sample entry, or by a codec specific method.

NOTE 2 The `width` field in the visual sample entry is the number of encoded horizontal video spatial samples after cropping, not the value of the track header `width` field.

- The `CleanApertureBox` should not be present (since the cropped aperture is defined to be "clean").
- The value of the `matrix` field signals the video orientation. Non-identity matrices shall be rotations in multiples of 90 degrees.
  - When video is not rotated, `matrix` shall be {0x00010000, 0, 0, 0x00010000, 0, 0, 0x40000000}.
  - When video should be rotated 90 degrees clockwise for display, `matrix` should be {0, 0x00010000, 0, 0xFFFF0000, 0, 0, height<<16, 0, 0x40000000}.
  - When video should be rotated 180 degrees for display, `matrix` should be {0xFFFF0000, 0, 0, 0, 0xFFFF0000, 0, width<<16, height<<16, 0x40000000}.
  - When video should be rotated 90 degrees counter-clockwise for display, `matrix` should be {0, 0xFFFF0000, 0, 0x00010000, 0, 0, 0, width<<16, 0x40000000}.

NOTE 3 A player is expected to adapt the picture aspect ratio (track header `width/height`) of the decoded and cropped video image to the size and shape of the current display. The player can frame the video aperture with letterbox bars, pillarbox bars, crop to fit, size to a window, etc. A player needs to scale, and possibly rotate, all CMAF fragments from a CMAF switching set to the selected video aperture to maintain seamless playback without size, shape, or location errors.

### 9.2.4 Sample Description Box ('sts<sub>d</sub>')

The `SampleDescriptionBox` in a video track shall contain at least one visual sample entry which shall include:

- `width` and `height` field values of the first sample entry equal to the largest cropped horizontal and vertical video spatial sample counts of any image in the track;

- a decoder configuration record that:
  - should signal the lowest profile, level, height and width values that are equal or greater than all CMAF fragments in the CMAF track.

**NOTE** Although it is valid to signal a higher profile and level than is necessary to decode the CMAF fragments in the CMAF track, that could unnecessarily exclude decoders capable of decoding the CMAF fragments, but not able to initialize the unnecessarily high profile and level signalled in the CMAF header.

### 9.2.5 Video CMAF fragment presentation time

CMAF video tracks shall either

- a) contain version 1 TrackRunBoxes in CMAF video fragments with composition offsets (negative composition offsets where necessary) to adjust the earliest media sample presentation time in each CMAF fragment to equal the earliest media sample decode time, which is equal to the baseMediaDecodeTime field of the TrackFragmentBaseMediaDecodeTimeBox, or
- b) contain an offset edit list in the associated CMAF header to subtract the composition delay added by positive composition offsets, if and only if the CMAF track is contained in a CMAF track file containing version 0 TrackRunBoxes.

A video track shall not use both signed composition offsets and an EditListBox.

### 9.2.6 Video media sample dependencies

- Within a video CMAF track, any TrackRunBox that describes any non-sync pictures shall identify picture dependencies using a combination of the sample\_flags and first\_sample\_flags fields, or default flags in the corresponding TrackFragmentHeaderBox:
  - sample\_is\_non\_sync\_sample shall be 0 for SAP type 1 or 2, and 1 otherwise;
  - sample\_depends\_on should be 1 or 2 (the value 2 identifies I pictures);
  - sample\_is depended\_on should be 2 for disposable pictures.

### 9.2.7 Video edit lists

When unsigned composition offsets are used in version zero TrackRunBoxes in a CMAF track file, the earliest video media sample will normally have a composition time that is greater than zero (whenever video media samples are reordered).

The earliest video media sample in a CMAF track file is defined to have presentation time zero, so an offset edit list, as specified in 7.5.13, shall be used when the earliest media sample's composition time is not zero. In this case, the value of the media\_time field shall be set to the composition time of the earliest presented video media sample in the CMAF track file.

For avoidance of doubt, all other video CMAF tracks (not in CMAF track files) shall use signed composition offsets in v1 TrackRunBoxes to adjust the earliest video media sample presentation time in the CMAF track to the baseMediaDecodeTime in the first CMAF fragment and shall not include an edit list.

### 9.2.8 General video CMAF fragment random access constraints

- a) The first media sample shall be Stream Access Point (SAP) type 1 or 2, as defined in ISO/IEC 14496-12 and a video CMAF media profile that specifies these SAP types for a specific video codec and track format.

**NOTE** SAP type 1 or 2 means that video media samples in a coded video sequence can be decoded in stored order from the start of the CMAF fragment and presented with the correct presentation timing and ordering.

- b) Each video CMAF fragment shall contain sufficient metadata to be decrypted and correctly displayed, possibly in combination with its CMAF header and separately delivered decryption keys.

### 9.2.9 Additional random access pictures within CMAF video fragments

When coded video sequences have durations longer than 2 seconds, pictures of SAP type 3 should be encoded every 2 seconds or less to provide additional random access video media samples.

For longer coded video sequences and resulting CMAF fragment durations, additional type 3 SAPs ("open GOP" independently decodable pictures) enable independent picture decoding for fast forward, fast reverse, and resumption of normal playback, while improving visual uniformity and lowering bit rate relative to groups of pictures with type 1 or 2 SAPs limited to the equivalent random access duration.

### 9.2.10 Image framing and encoding constraints

- a) CMAF video tracks shall only encode video spatial samples intended for presentation. Image padding used to adjust the picture aspect ratio, such as letterbox and pillarbox bars, should not be encoded.
- b) The video spatial samples intended for presentation should be upper left justified and only one row and/or one column of partially filled coding blocks encoded if the image height or width is not a multiple of the coding block size.
- c) Decoder cropping parameters shall be encoded in the video stream to remove video spatial samples in a partially filled coding block that are not intended for presentation.
- d) Each video CMAF fragment in a CMAF track or CMAF switching set may be encoded with different vertical and horizontal video spatial sample counts (and different corresponding sample entries or inband parameters), so devices are expected to scale each CMAF fragment to the same display aperture and maintain the correct picture aspect ratio, size, and position of each CMAF fragment during playback of CMAF switching sets.

NOTE Each player and display system is expected to frame the decoded and cropped video spatial samples to a player determined video display aperture using methods such as scaling, stretching, cropping, padding with letterbox or pillarbox bars, framing in a window, etc.

### 9.2.11 General video CMAF switching set constraints

#### 9.2.11.1 Video CMAF switching set constraints

- a) CMAF tracks in a CMAF switching set shall be encoded with the same display aspect ratio, although sample aspect ratio and size may differ. All video CMAF fragments within a CMAF switching set are intended to be displayed with the same height, width, and position.
- b) Video tracks in a CMAF switching set shall conform to the general CMAF switching set constraints in 7.3.4 that require each CMAF switching set to be encoded from the same source content and with the same visual characteristics such as transfer function, colour volume, colour primaries, colour subsampling, dynamic range, brightness, bit depth, and presentation timing, which would typically be the result of encoding from a single video master.
- c) Video tracks in a CMAF switching set shall conform to initialization constraints specified by the CMAF media profile they conform to, so a player and decoder can determine when CMAF headers need to be processed during adaptive switching.
- d) Switching between CMAF tracks at the start of CMAF fragments in a video CMAF switching set should result in continuous presentation and appearance when the CMAF fragments are scaled to the same device determined display aperture.

### 9.2.11.2 Aligned video CMAF switching set constraints

Aligned CMAF switching sets can enable players to switch between CMAF tracks in different CMAF switching sets. In aligned CMAF switching sets, CMAF fragments are time aligned so that track switching can be seamless at CMAF fragment boundaries, depending on the capabilities of the player and the differences between the aligned video CMAF switching sets.

Aligned video CMAF switching sets shall conform to [7.3.4.4](#).

Video media samples with the same presentation time in aligned video CMAF switching sets shall contain perceptually equivalent images that only differ by their encoding or encryption, i.e. codec, resolution, bit rate, frame rate, default key ID, etc.

A common example of aligned video CMAF switching sets is one CMAF switching set containing CMAF conforming to the CMAF HHD8 media profile, and another CMAF switching set containing CMAF tracks conforming to the CMAF UHD8 media profile, both encoded from the same source, using the same codec, with time aligned CMAF fragments. Because of different content protection requirements for high definition and ultra-high definition content, in this example, the high definition and ultra-high definition CMAF tracks are encrypted with different keys and digital rights management system playback rules.

An HHD8 capable player could select and play the HHD8 CMAF switching set normally. A UHD8 capable player could select and seamlessly switch both the UHD8 and HHD8 CMAF tracks in aligned CMAF switching sets, applying the different keys and licenses as needed.

### 9.2.11.3 Multiple initialization switching of video tracks

A CMAF header from each track in the switching set need only be downloaded once (it is assumed they do not change unless a new CMAF presentation is started, e.g. a DASH Period).

CMAF switching sets conforming to general constraints or single initialization constraints can be adaptively switched using multiple initialization, meaning that the CMAF header of a CMAF track is processed before decoding the first CMAF fragment from that track on every track switch. The dynamic reinitialization behaviour of parsers, decoders, and display systems is rarely specified and may not be seamless.

It is sometimes possible for players that have full control over both switching and decoding to evaluate changes in CMAF header parameters when switching and leave most parameters unchanged to minimize presentation disruption. In playback environments, such as Web browsers using an HTML5 media source extension (MSE) buffer, appending a new CMAF header can cause playback interruption when the media pipeline is fully reinitialized when each CMAF header is appended.

### 9.2.11.4 Single initialization switching of video tracks

CMAF switching sets containing video CMAF tracks that conform to single initialization constraints can be decoded and displayed by sequencing CMAF fragments from the same CMAF switching set after initializing once with a CMAF header.

Parameters that are allowed to change within a CMAF track or between CMAF tracks in a CMAF switching set can be contained in each CMAF fragment so that full reinitialization with a CMAF header is avoided on track switches. Parameters allowed to change are typically vertical and horizontal video spatial sample count and cropping, so changes in decoder configuration, display buffer sizes, etc. that might interrupt presentation can be avoided.

In a single initialization switching set with hierarchically nested CMAF media profiles, a CMAF header is sufficient to decode any CMAF tracks with equal or lower CMAF media profiles and resolutions than the initialized CMAF media profile and CMAF track.

Single initialization constraints for NAL structured video are specified in [9.3.7](#).

## 9.3 NAL structured video CMAF tracks

### 9.3.1 Overview

The NAL structured video CMAF track format shall conform to the general video CMAF track format specified in [9.2](#), the NAL structured video track format specified in ISO/IEC 14496-15, with the constraints specified in [9.3](#).

### 9.3.2 CMAF track format constraints for NAL structured video

#### 9.3.2.1 Track Header Box ('t\_khd')

ISO/IEC 14496-12 specifies that the values of width and height are set to the decoded image size on a uniformly sampled square grid.

For NAL structured video CMAF tracks, the values are additionally constrained as follows.

The values of width and height shall be normalized to the width and height of the NAL structured video, as defined below.

- The normalized presentation height shall be the number of vertical video spatial samples in the Sequence Parameter Set NAL VUI after cropping parameters are applied.

NOTE 1 The height field of the visual sample entry is also the number of encoded vertical video spatial samples after cropping.

- The normalized presentation width shall be the number of horizontal video spatial samples after Sequence Parameter Set NAL VUI cropping parameters are applied, then multiplied by the video spatial sample aspect ratio. The sample aspect ratio is specified by aspect\_ratio\_idc (and if applicable by the sar\_width/sar\_height ratio) in SPS VUI parameters and the PixelAspectRatioBox if present in the sample entry.

NOTE 2 The width field of the visual sample entry is the number of encoded horizontal video spatial samples after cropping.

- The CleanApertureBox should not be present, since the cropped aperture is defined to be active image ("clean") in a video CMAF track.

#### 9.3.2.2 Sample Description Box ('sts\_d')

The SampleDescriptionBox in a video track shall contain one or more visual sample entries conforming to ISO/IEC 14496-12 and ISO/IEC 14496-15.

The first VisualSampleEntry in the SampleDescriptionBox:

- shall include width and height field values that equal or exceed the largest cropped horizontal and vertical video spatial sample counts in any Sequence Parameter Set referenced by a video slice in the NAL structured video track;
- should contain no more than one NAL video parameter set, e.g. for AVC video, one set of SPS and PPS NALs with VUI in the AVCDecoderConfigurationRecord;
- and contain a decoder configuration record that:
  - should signal the lowest codec profile, level, height and width values that are equal to or greater than the values required for all the CMAF fragments in the CMAF track. See general constraints in [9.2.4](#);
  - should set LengthSizeMinusOne field to the value "3" (to indicate 4 bytes) to address large VCL NALs and simplify conversion of elementary streams between MPEG-2 TS bytestreams with startcodes and ISO/IEC 14496-15 with NAL length headers;

NOTE 1 The size of the NAL header length field defined in video tracks conforming to ISO/IEC 14496-15 is stored in the field `LengthSizeMinusOne` in the corresponding decoder configuration record, e.g. for AVC video in the `AVCDecoderConfigurationRecord`.

- shall contain one or more `ColorInformationBoxes` with sub-type '`nclx`' and a `PixelAspectRatioBox` '`pasp`', as documented in ISO/IEC 14496-12, if the first sample entry contains no SPS NAL with VUI in the decoder configuration record.

NOTE 2 A decoder configuration record without a video parameter set is valid for a sample description such as '`avc3`' or '`hev1`' that can store the video parameter set NALs necessary for decoding and display in each CMAF fragment.

Any subsequent sample entries in the `SampleDescriptionBox`:

- should contain one NAL video parameter set including VUI in each visual sample entry's decoder configuration box, e.g. for AVC, the '`avcc`' box;
- may be unreferenced by media samples in this CMAF track or other CMAF tracks in a CMAF switching set;
- may be constrained by dependency on other CMAF tracks and CMAF headers in a CMAF switching set. See [9.3.6](#).

NOTE 3 CMAF headers can contain `VisualSampleEntries` intended for use by other CMAF tracks in a CMAF switching set or for the purpose of initializing decoding and display, but not referenced by any slice header in a media sample.

### 9.3.3 NAL structured video Access Units contained in media samples

Each media sample shall contain one NAL structured video Access Unit for one presentation time and duration, as defined in ISO/IEC 14496-15 and ISO/IEC 14496-12.

NOTE As specified in ISO/IEC 14496-15, timing information provided within a video elementary stream is ignored. Instead, media sample timing in the `TrackRunBox` determines picture presentation time and duration.

Access Units shall conform to a sample description specified in ISO/IEC 14496-15 and signalled in the sample entry `codingname` field.

Each Access Unit may start with an Access Unit Delimiter NAL.

Each Access Unit shall be stored as a media sample in a `MediaDataBox` in a CMAF chunk and/or CMAF fragment.

Access Units conforming to sample descriptions with inband parameters, such as '`avc3`' and '`hev1`', may retain filler data (NAL units or SEI messages) and SEI messages that might change hypothetical reference decoder bitstream conformance if removed.

The filler data and SEI messages may be retained if HRD conformance is desired.

### 9.3.4 NAL structured video coding sequences corresponding to CMAF fragments

Each CMAF fragment shall contain one or more complete coded video sequences, as specified by the video codec. Consequently, the first media sample in all NAL video CMAF fragments is SAP type 1 or 2, as specified in ISO/IEC 14496-12.

NAL structured video sample descriptions that allow inline parameter sets (e.g. '`avc3`' and '`hev1`') shall contain all SPS and PPS NALs referenced by a Coded Video Sequence in the first Access Unit of that coded video sequence, immediately following the Access Unit Delimiter NAL, if present; followed by SEI NALs, if present; followed by the VCL NAL(s) of the first Access Unit.

If sample entries also exist in the CMAF tracks' CMAF header using the same sample entry and parameter set indexes as an inband parameter set, then they shall contain the same SPS and PPS NALs in their decoder configuration record as the inband parameter set.

### 9.3.5 Elementary stream constraints

#### 9.3.5.1 Video colour and dynamic range mastering

Video streams conforming to CMAF shall be conformant to a CMAF media profile and that profile brand should be included in the CMAF header. CMAF media profiles can constrain video transfer characteristics, colour parameters, and grading. See [A.2](#).

For all video media profiles, unless signalled otherwise by a mechanism defined in that video media profile, default transfer characteristics and grading shall be assumed. Default grading is defined as a viewing environment that complies with BT.2035 for presentation on a display which uses the electro-optic transfer function specified in BT.1886 with a peak luminance of 100 cd/m<sup>2</sup>.

Video may be graded for presentation on a display which uses an electro-optic transfer function not specified in BT.1886 and/or with a peak luminance greater than 100 cd/m<sup>2</sup>, in which case the grading profile should be signalled by VUI and/or SEI messages contained in the sample entry decoder configuration box, or other mechanism defined in the video media profile.

#### 9.3.5.2 Caption data in SEI messages

CTA 608/708 caption data (CTA-608-E, CTA-708-E) may be stored in video SEI messages defined by the associated video codec specification (see [11.4](#)). Video embedded captions are not considered a CMAF subtitle track, and it is expected that many CMAF compatible players will ignore these messages in video CMAF tracks. The presence of caption data in such SEI messages should be indicated by the addition of the 'ccea' compatibility brand in the CMAF header, as specified in [A.4](#).

For NAL structured video, CTA 608/708 caption data (CTA-608-E, CTA-708-E) may be stored in SEI messages in coded video sequences in CMAF fragments described as user data registered by Rec. ITU-T T.35, with SEI payloadType = 4 and the registered identifier in the field user\_data\_registered\_itu\_t\_t35. See ISO/IEC 23008-2:2015, D.2.6.

Captions intended for CMAF applications should be carried in subtitle CMAF tracks to make delivery optional and selectable by each player, to allow several alternative languages, roles, etc. without duplicating the same video tracks with different SEI messages. Subtitle tracks also enable parsing and presentation by player applications when a device does not contain a built-in subtitle decoder.

### 9.3.6 General CMAF switching set constraints for NAL structured video

Each CMAF track in a CMAF switching set containing NAL structured video conforming to ISO/IEC 14496-15 (NAL structured video track format) shall conform to general CMAF switching set constraints in [7.3.4](#) and general video CMAF switching set requirements in [9.2.11](#).

### 9.3.7 Single initialization CMAF switching set constraints for NAL structured video tracks and media profiles

NAL structured video CMAF switching sets conforming to the general CMAF switching set constraints for NAL structured video in [9.3.6](#) may also conform to single initialization constraints specified below.

Single initialization NAL structured video CMAF switching sets:

- a) Shall conform to the general single initialization constraints in [9.2.11.4](#).

- b) Shall index and store video parameter sets (SPS and PPS NAL units) that are referenced by each video coding layer slice NAL for decoding and display, using parameter sets stored in:
  - 1) one or more sample entries in a CMAF header, in which case one CMAF header shall contain sample entries sufficient to decode and display every CMAF track in the CMAF switching set, or shall contain sample entries sufficient to decode and display every CMAF track with a CMAF header containing the same CMAF media profile brand in its `FileTypeBox`. ISO/IEC 14496-15 requires '`avc1`' and '`hvc1`' sample descriptions to use sample entry storage;
  - 2) the first (IDR) access unit of every coded video sequence that references the parameter set from a slice header, as specified in [9.3.4](#). Parameter set indexes shall be valid within each coded video sequence, but need not be valid within other CMAF fragments or CMAF tracks. ISO/IEC 14496-15 allows '`avc3`' and '`hev1`' sample descriptions to use inband parameter set NAL storage, as additionally constrained in this clause;
  - 3) a combination of sample entry parameter storage (1) and inband parameter storage (2) may be used for '`avc3`' and '`hev1`' sample description CMAF tracks. In that case, each sample entry and slice NAL parameter set index shall index the same parameter set if stored in both locations. In addition, any CMAF header shall provide all the sample entries referenced by any CMAF fragment in the CMAF switching set.
- c) The first visual sample entry in each CMAF header shall be sufficient to initialize decoding, decryption, and display of all CMAF tracks in the CMAF switching set, or only CMAF tracks with matching CMAF media profile brands, if all CMAF headers contain CMAF media profile brands. Initialization information can be:
  - 1) the first sample entry in the CMAF header shall contain a decoder configuration record containing a parameter set (e.g. SPS and PPS in '`avcC`') conforming to c), or
  - 2) may contain a sample entry without NALs that shall include one or more `ColorInformationBoxes` with sub-type '`nclx`' and a `PixelAspectRatioBox`, as specified in ISO/IEC 14496-12.
- d) Any CMAF header shall be sufficient to initialize the CMAF switching set once and decode all CMAF fragments in the CMAF switching set, if the CMAF tracks in the CMAF switching set do not contain CMAF media profile brands.
- e) Any CMAF header with a specific CMAF media profile brand shall be sufficient to initialize once and decode the other CMAF tracks with the same CMAF media profile brand, or CMAF media profile brands that conform to the brand initialized, if all CMAF tracks in the CMAF switching set contain CMAF media profile brands.

**NOTE** A manifest can reference a common CMAF header for every CMAF track in the CMAF switching set, or it can reference a different CMAF header for each CMAF media profile in the CMAF switching set when each CMAF header contains a CMAF media profile brand. In the second case, a player can initialize once and play a subset of the CMAF tracks that conform to the initialized media profile. If it initializes the highest signalled CMAF media profile, it can decode all CMAF fragments in the CMAF switching set.

## 9.4 AVC video CMAF tracks

### 9.4.1 Storage of AVC elementary streams

#### 9.4.1.1 Conformance

AVC video tracks shall conform to the NAL structured video format specified in [9.3](#), the AVC specification ISO/IEC 14496-10 and constraints specified below.

### 9.4.1.2 AVC Visual Sample Entry

The syntax and values for visual sample entry shall conform to AVCSampleEntry ('avc1') or AVCSampleEntry ('avc3') as defined in ISO/IEC 14496-15 and the general video sample entry requirements of [9.3.2.2](#).

## 9.4.2 Constraints on AVC elementary streams

### 9.4.2.1 Picture type

All pictures shall be encoded as frames and shall not be encoded as fields.

### 9.4.2.2 Sequence Parameter Sets (SPS)

#### 9.4.2.2.1 SPS field constraints

Sequence Parameter Set NAL units that occur in an AVC video CMAF track shall conform to ISO/IEC 14496-10 with the following additional constraints.

- The following fields have pre-determined values as follows:
  - frame\_mbs\_only\_flag shall be set to 1;
  - vui\_parameters\_present\_flag shall be set to 1;
  - gaps\_in\_frame\_num\_value\_allowed\_flag should be set to 0.
- The values of the following fields shall not change throughout a CMAF track.
  - chroma\_format\_idc
  - bit\_depth\_luma\_minus8
  - bit\_depth\_chroma\_minus8
- The maximum values of the following fields are specified by CMAF media profiles in [A.2](#).
  - profile\_idc
  - level\_idc
  - pic\_width\_in\_mbs\_minus1
  - pic\_height\_in\_map\_units\_minus1
- The following fields may change per CMAF fragment within a CMAF track. Changes in these parameters shall require a different sample entry, if the SPS NALs referenced are stored in sample entries.
  - pic\_width\_in\_mbs\_minus1
  - pic\_height\_in\_map\_units\_minus1
  - frame\_crop\_right\_offset
  - frame\_crop\_bottom\_offset
  - max\_num\_ref\_frames

#### 9.4.2.2.2 Visual Usability Information (VUI) parameters

VUI parameters that occur within a CMAF AVC video track shall conform to ISO/IEC 14496-10 with the following additional constraints.

The following fields have pre-determined values as follows:

- `video_signal_type_present_flag` should be set to 1, and the value of `video_full_range_flag` when not present shall be assumed to be 0.

NOTE 1 This indicates normal black “setup”, i.e. black level is 16 for 8-bit video.

- `aspect_ratio_info_present_flag` shall be set to 1. `aspect_ratio_idc` shall not be set to 0 (unknown).

NOTE 2 These parameters refer to the video spatial sample aspect ratio, not picture aspect ratio. Always setting the value distinguishes between content that omitted the value because it intended the default or just failed to set it properly.

- `overscan_info_present_flag`, if present, shall be set to 0.

NOTE 3 Exact scan encoding of the active image is used for reliable image framing by devices and precise adaptive scaling during adaptive switching.

- If `video_signal_type_present_flag` is set to 1, `colour_description_present_flag` should be set to 1.

NOTE 4 As defined in ISO/IEC 14496-10, if the `colour_description_present_flag` is set to 1, the `colour_primaries`, `transfer_characteristics` and `matrix_coefficients` fields are present.

- If `colour_description_present_flag` is set to 1, then `colour_primaries`, `transfer_characteristics` and `matrix_coefficients` shall be set to the values used in the AVC video CMAF track.

- If `colour_description_present_flag` is set to 0, the following values shall be assumed in AVC video CMAF tracks:

- `colour_primaries` = 1;
- `transfer_characteristics` = 1;
- `matrix_coefficients` = 1.

- The presence and values of the following fields shall not change in a CMAF track.

- `low_delay_hrd_flag`
- `colour_primaries`
- `transfer_characteristics`
- `matrix_coefficients`

#### 9.4.2.3 Picture cropping

Any picture cropping needed due to partially filled coding blocks shall be indicated by the SPS cropping parameters `frame_crop_bottom_offset` or `frame_crop_right_offset`. See [Annex C](#) for examples.

NOTE The NAL video CMAF track format fixes top and left cropping values to zero (default when omitted), which means the image is upper left justified within the coded blocks, and there is at most one row and column of partially filled and padded blocks below and/or to the right of the image if the image size is not evenly divisible by the coded block size.

## 9.5 AVC video Internet Media Type parameters

### 9.5.1 AVC signalling of "codecs" parameters

The video codec profile and level of each AVC track and CMAF switching set should be signalled using parameters conforming to RFC 6381 and ISO/IEC 14496-15.

## 10 Audio CMAF tracks

### 10.1 Overview

This clause specifies CMAF audio tracks derived from the CMAF track format, with additional constraints specific to CMAF audio tracks and audio CMAF media profiles.

See [12.1](#) for guidelines on specification and registration of CMAF media profiles.

[Clause 10](#) includes:

- constraints that apply to all audio CMAF tracks;
- constraints that apply to the audio track format and encoding of AAC audio using the system layer specified in ISO/IEC 14496-14;
- an “AAC Core” CMAF media profile that specifies constraints on AAC CMAF tracks to enable interoperability and random access of CMAF switching sets containing a single CMAF track encoded with one of three variants of AAC in stereo or mono;
- an “AAC adaptive” CMAF media profile that specifies additional constraints on CMAF AAC tracks, fragments, media samples, and metadata to allow seamless switching between alternative tracks and bit rates in an AAC adaptive CMAF switching set.

### 10.2 General audio CMAF track format

#### 10.2.1 Derivation

Audio CMAF tracks shall conform to the 'cmfc' compatibility brand and CMAF track format specified in [Clause 7](#) and [Clause 8](#) and the general audio CMAF track constraints specified in [Clause 10](#).

The general CMAF audio track format applies to all audio CMAF tracks, regardless of audio codec or media profile.

#### 10.2.2 Track Header Box ('tkhd')

For audio tracks, the fields of the TrackHeaderBox shall be set to the values specified below. Other fields may be set per ISO/IEC 14496-12.

- flags = 0x000007
- layer = 0
- volume = 0x0100
- matrix = {0x00010000, 0, 0, 0, 0x00010000, 0, 0, 0, 0x40000000} // unity matrix
- width = 0
- height = 0
- duration = 0

### 10.2.3 Sound Media Header Box ('smhd')

The syntax and values for the SoundMediaHeaderBox shall conform to ISO/IEC 14496-12 with constraints specified in [7.5.7](#).

### 10.2.4 Sample Description Box ('stsd')

As specified in ISO/IEC 14496-12, the SampleDescriptionBox contains the AudioSampleEntry box or AudioSampleEntryV1 box.

### 10.2.5 AudioSampleEntry

For all audio media profiles in CMAF, the value of the samplesize parameter in the AudioSampleEntry box defined in ISO/IEC 14496-12 shall be set to 16.

Each AudioSampleEntry or AudioSampleEntryV1 shall contain a "codingnamespecific" Box" containing codec-specific information.

### 10.2.6 Audio offset edit list

An offset edit list in a CMAF audio track shall conform to constraints specified on EditListBoxes in [7.5.13](#).

If audio media sample composition times differ from their intended presentation times in a CMAF presentation, then an offset edit list shall be included in that audio CMAF track to adjust to the intended CMAF track presentation time.

## 10.3 AAC audio CMAF tracks

### 10.3.1 Overview

AAC audio CMAF tracks containing AAC audio as defined in ISO/IEC 14496-3 shall conform to the general audio CMAF track constraints in [10.2](#) and the AAC constraints in [10.3](#). These constraints are used to derive the AAC audio media profiles specified in [10.4](#) and [10.5](#) and listed in [A.3](#).

[Table 12](#) lists the AAC profiles allowed.

**Table 12 — AAC profiles**

AAC profile	codingname	SampleEntry Type
MPEG-4 AAC (AAC-LC)	mp4a	MP4AudioSampleEntry
MPEG-4 high efficiency AAC (HE-AAC)	mp4a	MP4AudioSampleEntry
MPEG-4 high efficiency AAC v2 (HE-AACv2)	mp4a	MP4AudioSampleEntry

The SampleEntry format in the SampleDescriptionBox is the same for each AAC audio profile.

### 10.3.2 "codecs" parameter signalling

The signalling of the MIME "codecs" parameter is per RFC 6381, as shown in [Table 13](#). The third value of the codecs parameter is the audioObjectType, as if explicit hierarchical signalling were used.

**Table 13 — AAC MIME “codecs” parameter according to RFC 6381**

AAC profiles	MIME type	codecs parameter
MPEG-4 AAC (AAC-LC)	audio/mp4	mp4a.40.2
MPEG-4 high efficiency AAC (HE-AAC)	audio/mp4	mp4a.40.5
MPEG-4 high efficiency AAC v2 (HE-AACv2)	audio/mp4	mp4a.40.29

NOTE HE-AAC is a superset of AAC-LC and HE-AACv2 is a superset of HE-AAC. It is assumed that an HE-AACv2 decoder is also capable of decoding HE-AAC or AAC-LC, and an AAC-LC decoder is capable of partially decoding HE-AAC and HE-AACv2 conforming to CMAF constraints (without reproduction of high frequencies coded with SBR).

### 10.3.3 Considerations for AAC audio encoding

#### 10.3.3.1 Overview of AAC presentation timing

The AAC codec uses audio frames of a fixed length and a transform which applies over two frames. To obtain correct audio from a frame, both frames in the transform are needed, and hence the prior encoded frame and the current encoded frame need to be decoded to output the first frame. This is sometimes called “priming” and may be signalled using the ‘roll’ sample group.

A full reconstruction of the first encoded audio frame is sometimes not possible since there is no previous access unit. To still achieve a full reconstruction, a common practice is to add silence to the beginning of the audio signal. A more detailed explanation of this approach can be found in ISO/IEC TR 14496-24.

In practice, an encoder might prepend an arbitrary amount of (invalid) audio waveform samples to the signal. This portion of the audio signal is sometimes called “encoder delay” and varies depending on the implementation.

#### 10.3.3.2 Presentation delay compensation using an edit list

The most common approach to compensate for inserted extra audio is to add an offset edit list to the CMAF header.

In the case where padding has been added to the start of an audio stream, the `media_time` in the edit list is the length (in audio samples, as measured by the timescale of the track) of the inserted audio samples; 2112 is a common example for AAC.

If an edit list is used, a single `EditListBox` shall be recorded in the CMAF header, as specified in [10.2.6](#).

#### 10.3.3.3 Delay compensation before encapsulation

If the content has been generated according to [G.5](#), no `EditListBox` is present.

#### 10.3.3.4 Delay compensation when using additional AAC coding tools

If the SBR and PS coding tools are present, they shall not be considered for the purpose of delay compensation.

#### 10.3.3.5 Loudness and dynamic range control

The audio stream should contain DRC and loudness metadata according to ISO/IEC 14496-3. The audio encoder should set the Program Reference Level to the loudness level of the audio stream.

The audio encoder should generate DRC metadata for light compression encoded in the `dyn_rng_ctl` and `dyn_rng_sgn` fields of `dynamic_range_info()` in the FIL element and DRC metadata for heavy

compression in the `compression_value` field of `MPEG4_ancillary_data()` in the data stream element (DSE).

**NOTE** It is expected that the audio decoder will use the Program Reference Level, if available, to achieve a desired target loudness, if applicable. It is expected that the audio decoder will apply the DRC metadata, if present, according to ISO/IEC 14496-3 including the DRC Presentation Mode value of the `drc_presentation_mode` fields.

### 10.3.4 AAC track constraints

#### 10.3.4.1 Storage of AAC media samples

Storage of AAC elementary stream access units as media samples within a CMAF track shall conform to the CMAF audio track format specified in [10.2](#).

The following additional constraints also apply.

- All audio media samples shall consist of one AAC audio access unit.
- All AAC access units in a CMAF track shall be encoded with one of AAC-LC, HE-AAC or HE-AACv2.
- The values given in `AudioSampleEntry`, `DecoderConfigDescriptor`, and `DecoderSpecificInfo` shall match the corresponding values in the AAC audio bitstream.

#### 10.3.4.2 AAC audio sample entry

##### 10.3.4.2.1 Field values

The syntax and values of the `AudioSampleEntry` shall conform to `MP4AudioSampleEntry` ('mp4a') as defined in ISO/IEC 14496-14.

The sample entry and fields specified in [10.3.4.2](#) shall not change within a CMAF track.

##### 10.3.4.2.2 ESBox

As defined in ISO/IEC 14496-14, the `MP4AudioSampleEntry` in the `SampleDescriptionBox` shall contain an `ESBox`, which contains an `ES_Descriptor`.

##### 10.3.4.2.3 ES\_Descriptor

The syntax and values for `ES_Descriptor` shall conform to ISO/IEC 14496-1, and the fields of the `ES_Descriptor` shall be set to the following values.

- `ES_ID` = 0
- `streamDependenceFlag` = 0
- `URL_flag` = 0
- `OCRstreamFlag` = 0
- `streamPriority` = 0
- `decConfigDescr` = `DecoderConfigDescriptor`
- `slConfigDescr` = `SLConfigDescriptor`, predefined type 2

Descriptors other than those specified in [10.3.4.2.4](#) through [10.3.4.2.6](#) shall not be used.

#### 10.3.4.2.4 DecoderConfigDescriptor

The syntax and values for DecoderConfigDescriptor shall conform to ISO/IEC 14496-1, and the fields of this descriptor shall be constrained to the following values.

- decoderSpecificInfo shall be used and ProfileLevelIndicationIndexDescriptor shall NOT be used.
- objectTypeIndication = 0x40 (audio)
- streamType = 0x05 (audio stream)
- upStream = 0
- decSpecificInfo = AudioSpecificConfig

#### 10.3.4.2.5 AudioSpecificConfig

The syntax and values for AudioSpecificConfig shall conform to ISO/IEC 14496-3.

The following fields of AudioSpecificConfig shall be set to AAC audio CMAF media profile specified values, for example, see [10.4](#).

- audioObjectType
- channelConfiguration
- extensionAudioObjectType
- GASpecificConfig

#### 10.3.4.2.6 GASpecificConfig

The syntax and values for GASpecificConfig shall conform to ISO/IEC 14496-3, and the fields of GASpecificConfig shall be set to the following values.

- frameLengthFlag = 0 (1024 lines IMDCT)
- dependsOnCoreCoder = 0
- extensionFlag = 0

### 10.3.5 AAC elementary stream constraints

#### 10.3.5.1 General encoding constraints

AAC elementary streams shall conform to ISO/IEC 14496-3 and the constraints of an audio CMAF media profile it conforms to, for example, see [A.3](#).

- The elementary stream shall be a raw data stream, i.e. ADTS and ADIF headers shall not be present.
- CMAF fragments containing HE-AAC shall start with a type 1 SAP. Notably, the SBR configuration information shall be in the first packet.
- The transform length of the IMDCT for AAC shall be 1024 audio PCM samples for long blocks and 128 audio PCM samples for short blocks.
- The following parameters shall not change within the elementary stream:
  - audio object type;
  - sampling frequency;

- channel configuration.

### 10.3.5.2 AAC elementary stream syntactic elements

#### 10.3.5.2.1 Syntax and values of syntactic elements

The syntax and values for syntactic elements shall conform to ISO/IEC 14496-3.

The following element shall not be present in an MPEG-4 HE-AAC or HE-AACv2 elementary stream:

- coupling\_channel\_element (CCE).

If the program\_config\_element (PCE) element is present, then it shall only list a set of channels corresponding to one of the fixed channel configurations specific in ISO/IEC 14496-3, and the element shall not change for the duration of the track.

#### 10.3.5.2.2 Arrangement of syntactic elements

The syntax and values for syntactic elements shall conform to ISO/IEC 14496-3.

Syntactic elements shall be arranged in the following order for the channel configurations below.

- <SCE>, <optional additional elements>, <TERM>... for HE-AACv2 and mono HE-AAC or AAC-LC.
- <CPE>, <optional additional elements>, <TERM>... for stereo HE-AAC or AAC-LC.

NOTE Angled brackets (<>) are used above to indicate separate syntactic elements, not stream syntax.

#### 10.3.5.2.3 individual\_channel\_stream

The syntax and values for individual\_channel\_stream shall conform to ISO/IEC 14496-3. The following fields shall be set as defined:

- gain\_control\_data\_present = 0.

#### 10.3.5.2.4 Maximum bit rate

The maximum bit rate of AAC elementary streams shall be calculated in accordance with the AAC buffer requirements as defined in ISO/IEC 14496-3. Only the raw data stream shall be considered in determining the maximum bitrate (system-layer descriptors are excluded).

### 10.4 AAC core audio CMAF media profile

AAC core audio CMAF media profile shall conform to the AAC CMAF track format specified in [10.3.4](#), with the following constraints.

- Each AAC elementary stream shall be encoded using MPEG-4 AAC-LC, HE-AAC Level 2, or HE-AACv2 Level 2. Use of the MPEG-4 HE-AACv2 is recommended for 32 kbps or lower.
- When using HE-AAC and HE-AACv2 bitstreams, explicit backwards compatible signalling shall be used to indicate the use of the SBR and PS coding tools.
- AAC core CMAF tracks shall not exceed two audio channels.
- AAC core elementary streams shall not exceed 48 kHz sampling rate.
- The AAC core FileTypeBox compatibility brand shall be 'caac' and should be used to indicate CMAF tracks that conform to this media profile.

See [A.3](#) for the table of AAC audio CMAF media profiles and their compatibility brands.

## 10.5 AAC adaptive switching audio CMAF media profile

### 10.5.1 General constraints

Producing audio content capable of seamless adaptive switching with codecs available in the AAC core audio CMAF media profile (AAC-LC, HE-AAC, HE-AACv2) requires constrained encoding at CMAF fragment boundaries. [10.5](#) specifies constraints on AAC adaptive CMAF tracks and [Annex G](#) includes recommendations for their use in CMAF switching sets.

- The AAC adaptive audio CMAF media profile `FileTypeBox` compatibility brand shall be '`caaa`' and should be used to indicate CMAF tracks that conform to this CMAF media profile.
- CMAF tracks conforming to AAC adaptive CMAF media profile and brand shall conform to the AAC core CMAF media profile specified in [10.4](#) and the constraints specified in [10.5](#).

See [A.3](#) for the table of AAC audio CMAF media profiles and their compatibility brands.

### 10.5.2 CMAF fragment encoding constraints

To enable seamless switching between AAC CMAF tracks in a CMAF switching set, the AAC CMAF fragments comprising each CMAF track are encoded and packaged for random access and seamless splicing. The process of encoding the AUs (Access Units) requires periodic generation of Stream Access Points (SAP) that are time aligned between CMAF tracks, and encapsulation requires CMAF fragment alignment to those SAPs. Tracks that belong to the same CMAF switching set can be generated from different encoder instances, but shall conform to the same CMAF switching set constraints.

All CMAF tracks in an AAC CMAF switching set shall have AAC SAPs at CMAF fragment boundaries (i.e. the first AU in each CMAF fragment), and corresponding CMAF fragments in alternative CMAF tracks shall be decode time aligned and presentation time aligned.

Video frames usually have well-defined SAP types, such as an IDR frame that always equals SAP type 1 or 2. For audio, frame SAP types are not usually specified. In MPEG-4 AAC, there is no special frame type defined for random access. Nevertheless, audio SAPs of type 1 can be generated by applying the constraints described in the following clauses and are referred to as an "SAP" frame, since there is only one type.

### 10.5.3 General considerations and requirements

The following values shall not change within a CMAF switching set:

- audio object type (AOT);
- channel configuration;
- sampling frequency.

The first point means that all the CMAF tracks within a CMAF switching set share the same AOT, i.e. they either use AAC-LC, HE-AAC or HE-AACv2.

The second and third constraints enable the output PCM to remain constant and not require, for example, a re-configuration of the audio device or sound card. A constant sampling frequency assures that AUs share the same temporal framing and, therefore, the AAC CMAF fragments in alternative CMAF tracks are decode time aligned, so no overlap or gap in presentation time results when switching.

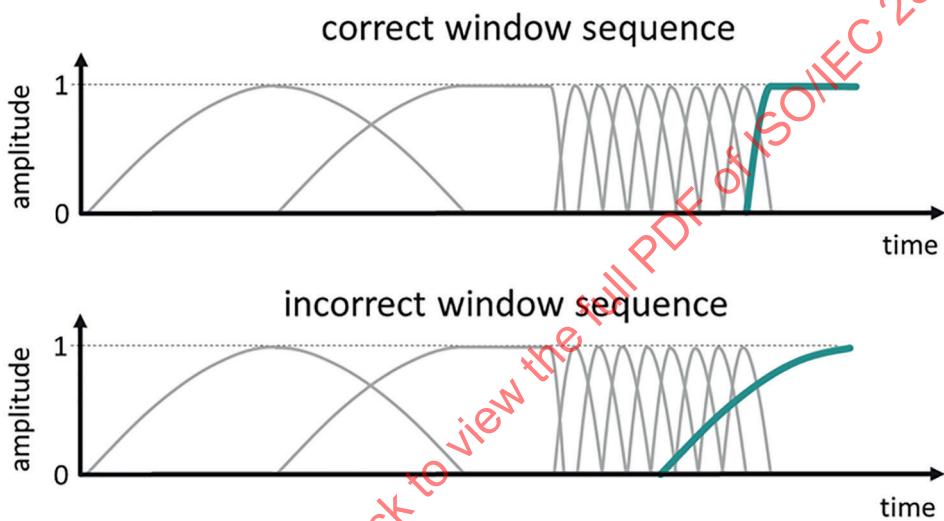
All CMAF tracks in the same CMAF switching set shall use the same method of presentation delay compensation, either an offset edit list specified in [10.3.3.1](#) or decode time compensation specified in [10.3.3.2](#).

#### 10.5.4 Constraints for AAC-LC

To guarantee seamless switching for AAC-LC encoded tracks, the window shape and window sequence shall be constrained as follows.

- To avoid artefacts from not cancelled time alignment components, the window shape and window sequence shall be synchronized across all tracks.
- At each CMAF fragment boundary (i.e. “right window half” of the last frame in a fragment and “left window half” of the first frame in a fragment), all tracks shall use the same window shape and a defined window sequence.

Using a short overlap for the window sequence (i.e. a long start or short window sequence) is recommended in the last frame of a CMAF fragment. This allows for the use of either a short or stop Window Sequence in the SAP frame (i.e. first frame in the next CMAF fragment). This concept is exemplified in [Figure 15](#).



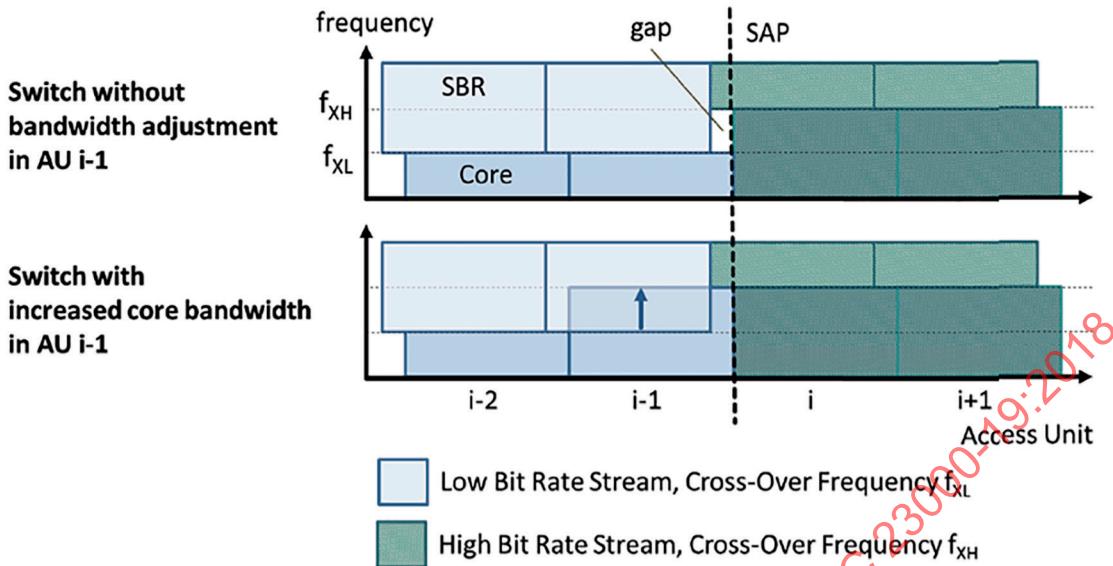
**Figure 15 — Correct and incorrect windowing sequence at fragment boundaries for seamless switching of AAC-LC CMAF tracks**

#### 10.5.5 Constraints for HE-AAC

Since HE-AAC is based on AAC-LC, the restrictions described in [10.5.4](#) apply also here. In addition, the following restrictions apply to the Spectral Band Replication (SBR) tool.

##### 10.5.5.1 Core bandwidth adjustment

The framing of the SBR decoder analysis is delayed by 6 QMF time slots ( $6 \times 64$  audio PCM samples) compared to the framing of the AAC core decoder. Furthermore, the QMF analysis adds another 320 audio PCM samples. This time shift between core and SBR framing might result in an energy gap when switching, for example, from a track encoded with low bit-rate to a track encoded with high bit-rate. This concept is depicted in [Figure 16](#).



**Figure 16 — Illustration of core bandwidth adjustment for HE-AAC switching points**

To avoid this gap in the frequency range, it is recommended to have the AAC core bandwidth of the last frame of a fragment matching the highest AAC/SBR crossover frequency of the supported stream configuration. To properly encode the additional bandwidth, extra bits are necessary and the bit reservoir control should be adapted accordingly.

#### 10.5.5.2 SBR header and time differential coding

In contrast to the AAC configuration that is completely signalled inside the audio specific configuration (ASC), the SBR decoder needs additional configuration parameters. These parameters are transmitted in the SBR header, which may not be contained in every access unit. The MPEG-4 standard recommends a transmission interval of 500 ms or whenever an instantaneous change of header parameters is required. To allow for seamless switching of HE-AAC encoded tracks, an SBR header shall be transmitted in each SAP frame. For frames containing the SBR header, the MPEG-4 audio conformance forbids the use of tools that rely on preceding frames. In other words, time differential coding of parameters contained in the SBR header is forbidden and, hence, this restriction guarantees that SAP frames can be completely decoded and processed.

#### 10.5.5.3 SBR frame class

If the SBR frame class is “VARVAR” or “FIXVAR”, the SBR envelope for this particular frame could reach over the frame borders. To make sure all necessary information is self-contained in the SAP frame, so that it can be fully decoded, the SAP shall always start with a “FIX” border (“FIXVAR” or “FIXFIX”). Consequently, the last frame in a fragment (the frame before the SAP) shall also end with a “FIX” border (“FIXFIX” or “VARFIX”).

#### 10.5.6 Constraints for HE-AACv2

Since HE-AACv2 is based on HE-AAC, the restrictions described in 10.5.5 are valid also in this case. In addition, the following restriction applies to the Parametric Stereo (PS) tool.

Just like for SBR, the PS configuration parameters may not be transmitted with every access unit. To allow for seamless switching of HE-AACv2 encoded tracks, PS header shall be transmitted with each SAP frame. For frames containing a PS header, the MPEG-4 audio conformance again forbids the use of tools that rely on past information, i.e. time differential coding of parameters. These restrictions assure that the SAP frame can be completely decoded and processed. Since HE-AACv2 conformance requires

a PS header delivered together with each SBR header, this requirement is implicitly inherited from the HE-AAC requirements.

See [Annex G](#) for additional guidelines.

## 11 Subtitles and captions

### 11.1 Overview

CMAF defines the following formats for carrying subtitles and captions:

- WebVTT subtitle CMAF tracks;
- IMSC1 subtitle CMAF tracks;
- CTA-608/CTA-708 captions embedded in video CMAF tracks.

The term “subtitles” in this document is used to mean a CMAF track for the visual presentation of text or graphics that is synchronized with video and audio for purposes including “closed captions” for hearing impaired viewing. Subtitles are presented for various purposes such as dialogue language translation, localized titles, commentary, lyrics, and descriptive captions for hearing impaired viewers or presentation situations where audio is unavailable or inappropriate.

CTA-708 (608) “closed captions” embedded in video elementary streams are also described, in part, to distinguish them from CMAF subtitle tracks. Closed captions are commonly used in North America to broadcast descriptive text and dialogue, primarily intended for hearing impaired viewers. See [9.3.5.2](#). Embedded closed captions have limited flexibility because they cannot be independently created, selected, and delivered without duplicating every video track, and they require low level processing of video elementary streams to decode. Closed captions may be present in CMAF video tracks and can be ignored by players that only process subtitle CMAF tracks.

A CMAF presentation can offer the same subtitle content in more than one CMAF track and format to reach players with different decoding capabilities. Because each CMAF track can have different qualities, it is recommended that manifests indicate the preferred CMAF tracks in the selection set.

### 11.2 WebVTT

WebVTT is a subtitle format that is supported by several web browsers. It can also be used to render subtitles in other types of media presentation applications. In CMAF, a WebVTT document is encapsulated in a CMAF fragment contained in a CMAF track. WebVTT CMAF fragments may be sequentially downloaded, synchronized, and presented the same as audio and video CMAF fragments.

The encapsulation and format of WebVTT documents in ISO Base Media movie fragments and tracks shall conform to ISO/IEC 14496-30, to [Clause 7](#) of this document, and to the 'cwvt' media profile listed in [A.4](#) of this document.

As specified in ISO/IEC 14496-30, each WebVTT document is contained in a single ISOBMFF movie fragment. To conform to CMAF, WebVTT movie fragments shall also conform to the CMAF fragment constraints specified in [Clause 7](#).

One CMAF fragment and WebVTT document may span the entire duration of a prerecorded presentation, or multiple CMAF fragments may sequentially span the duration of the presentation with CMAF fragment durations similar to audio and video CMAF fragments, which is necessary to approximate audio and video latency during live encoding and streaming.

WebVTT subtitle tracks are specified in ISO/IEC 14496-30 using a track handler\_type of 'text' with a codingname of 'wvtt'. WebVTT subtitle tracks store a WebVTT document as a media sample containing zero or more WebVTT cues with presentation times contained within the presentation timespan of the media sample and CMAF fragment that contains it. WebVTT subtitle CMAF fragments are synchronized during playback based on each CMAF fragment's presentation time and duration, and

WebVTT cues are decoded and presented within the CMAF fragments presentation time interval based on their WebVTT timestamps.

By design in ISO/IEC 14496-30, no media sample can contain a WebVTT cue with a presentation time that is outside or crosses the media sample and CMAF fragment presentation time interval (i.e. starts before the CMAF fragment starts and/or ends after the CMAF fragment ends, that duration determined by the duration of the media sample contained in the CMAF fragment). ISO/IEC 14496-30 specifies how to split cues into multiple WebVTT documents and media samples to avoid this case.

NOTE Forced titles are supported in TTML/IMSC1, but not WebVTT. A workaround is to create two WebVTT tracks, one with only the forced titles and the other with both forced and regular subtitles. A player would then play the forced-only track by default, and switch to the forced + subtitles track when subtitles are selected.

## 11.3 IMSC text and image tracks

### 11.3.1 General

TTML Profiles for Internet Media Subtitles and Captions 1.0 (IMSC1) is an application of the Timed Text Markup Language (TTML) for subtitle and caption delivery. IMSC1 defines a text-only profile and an image-only profile.

An IMSC1 text track is a CMAF track that conforms to the provisions of [11.3.2](#) and [11.3.3](#).

An IMSC1 image track is a CMAF track that conforms to the provisions of [11.3.2](#) and [11.3.4](#).

### 11.3.2 Common constraints

The CMAF track shall conform to ISO/IEC 14496-30 (W3C IMSC1) and an IMSC1 media profile listed in [Table A.3](#) of [A.4](#).

The namespace field of the XMLSubtitleSampleEntry box shall contain one instance of the string "http://www.w3.org/ns/ttml".

NOTE The schema\_location field of the XMLSubtitleSampleEntry box is not expected to be set for the purpose of profile signalling. The CMAF media profile brand can be used to determine the profile.

The XMLSubtitleSampleEntry shall contain a MIMEBox as specified in ISO/IEC 14496-12 and its content\_type field shall be constrained as follows.

- The type shall be "application".
- The subtype shall be "ttml+xml".
- The codecs parameter shall be present and set according to [11.3.3](#) and [11.3.4](#).

The media type of the CMAF track, as specified in RFC 6381, shall conform to the following.

- The type shall be "application".
- The subtype shall be "mp4".
- The codecs parameter shall be present and set according to [11.3.3](#) and [11.3.4](#).

### 11.3.3 IMSC1 text track constraints

All subtitle media samples of the CMAF track shall conform to the Text Profile specified in W3C IMSC1.

The codecs parameter of the content\_type field of the MIME box within the XMLSubtitleSampleEntry box (see [11.3.2](#)) shall contain the value "im1t", which signals that an

IMSC1 text processor is required, as specified in the W3C, TTML Media Type Definition and Profile Registry.

NOTE 1 A document that conforms to EBU-TT-D, as specified in EBU TECH 3380, also generally conforms to IMSC Text Profile.

EXAMPLE 1 "application/ttml+xml;codecs=im1t|etd1" signals that the documents contained in an IMSC1 text track also conform to EBU TECH 3380.

The `codecs` parameter of the media type of the CMAF track shall contain the value "stpp.ttml.im1t". Other TTML profiles the track conforms to may be additionally listed in the `codecs` parameter.

EXAMPLE 2 "application/mp4;codecs=stpp.ttml.im1t" is the media type of an IMSC1 text subtitle CMAF track.

EXAMPLE 3 `@mimeType="application/mp4"` and `@codecs="stpp.ttml.im1t"` signal an IMSC1 text subtitle CMAF track in a DASH manifest.

NOTE 2 [A.4](#) specifies a file compatibility brand that signals that a track is an IMSC1 text subtitle CMAF track.

#### 11.3.4 IMSC1 image track constraints

All subtitle media samples of the CMAF track shall conform to the Image Profile specified in W3C IMSC1.

The `codecs` parameter of the `content_type` field of the MIME box within the `XMLSubtitleSampleEntry` box (see [11.3.2](#)) shall contain the value "im1i", which signals that an IMSC1 image processor is required, as specified in the W3C, TTML Media Type Definition and Profile Registry.

EXAMPLE 1 "application/ttml+xml;codecs=im1i" is a possible value for the `content_type` field of an IMSC1 image track.

EXAMPLE 2 "application/mp4;codecs=stpp.ttml.im1i" is the media type of an IMSC1 image track.

The `codecs` parameter of the media type of the CMAF track shall contain the value "stpp.ttml.im1i".

EXAMPLE 3 `@mimeType="application/mp4"` and `@codecs="stpp.ttml.im1i"` signal an IMSC1 image track in a DASH manifest.

Each `smpte:backgroundImage` attribute shall be a URN as specified in ISO/IEC 14496-30 that may conform to the following `bg-image-urn` syntax expressed using RFC 5234.

```
bg-image-urn = "urn:mpeg:14496-30:subs:" 1*DIGIT
```

The `auxiliary_mime_types` field of the `XMLSubtitleSampleEntry` box includes one instance of the media type "image/png" if any image is used.

NOTE [A.4](#) specifies a file compatibility brand that signals that a track is an IMSC1 image subtitle CMAF track.

#### 11.4 CTA-608 and CTA-708

CTA-608 and CTA-708 are formats developed for delivering closed captions for accessibility purposes for broadcast television in North America. Closed captions can be embedded in the video elementary stream's SEI messages.

NOTE [A.4](#) specifies a file compatibility brand that signals that a video track contains CTA 608/708 captions in SEI messages and can be included in addition to the video CMAF media profile compatibility brand.

#### 11.5 Metadata for subtitles

Text tracks should be labelled with their role. See [7.5.3](#).

Text tracks should be tagged with their language. See [7.5.5](#).

## 12 CMAF media profiles and CMAF presentation profiles

### 12.1 CMAF media profiles

#### 12.1.1 General guidelines for specifying CMAF media profiles

##### 12.1.1.1 CMAF media profile CMAF track format

All CMAF media profiles should specify a CMAF track format, either derived directly from the 'cmfc' brand and format specified in [Clause 7](#) or from a video or audio CMAF track format specified in [Clause 9](#) or [Clause 10](#). The media profile's CMAF track format should specify any requirements and constraints on CMAF headers and CMAF fragments.

##### 12.1.1.2 Codecs, profiles, and levels

All CMAF media profiles should specify codec specific features and operating points, such as profiles and levels, that are necessary to determine encoder/decoder interoperability. The maximum features and limits allowed in content should be specified. When related media profiles are specified, subset/superset relationships between those media profiles should be specified based on the ability of a decoder conforming to superset media profile being able to reliably decode content conforming to a subset media profile.

##### 12.1.1.3 Media access unit mapping to media samples

All CMAF media profiles should specify one or more media access unit formats that are addressed as media samples in a CMAF track. In addition, all CMAF media profiles should specify the sample entry code point for the media samples in a CMAF track.

##### 12.1.1.4 Media access unit sequence mapping to CMAF fragments

All CMAF media profiles should specify any limitation on storing sequences of media samples in a CMAF fragment in order to conform to the general random access and timing constraints of CMAF fragments. If all media samples are sync samples, there are no additional constraints. If sequential media sample dependencies are caused by codec prediction, media sample reordering transforms that overlap media samples, sample group and layer dependencies, etc., then media sample sequences should be specified that allow CMAF fragments to be randomly accessed and optionally, adaptively switched (if the media profile defines adaptive switching functionality and CMAF switching set constraints).

##### 12.1.1.5 CMAF track constraints for CMAF switching sets

CMAF media profiles that support adaptive switching of multiple CMAF tracks in a CMAF switching set should specify the constraints between CMAF tracks in a CMAF switching set. CMAF media profiles should specify subset/superset relationships with compatible CMAF media profiles, e.g. a higher level of the same codec. A CMAF media profile should specify initialization constraints on CMAF switching sets and constraints related to the adaptive switching process enabled.

##### 12.1.1.6 CMAF media profile Internet Media Type

All CMAF media profiles should specify an Internet Media Type and associated parameters, file extensions, etc. for this media profile in a CMAF track or more generally in an ISOBMFF file.

##### 12.1.1.7 CMAF media profile brand

All CMAF media profiles should specify a compatibility brand and register that brand at [www.mp4ra.org](http://www.mp4ra.org). The brand registry should reference a CMAF media profile specification conforming to [12.1.1](#).

## 12.1.2 Guidelines for audio CMAF media profiles

### 12.1.2.1 General

Audio codecs often define features and other details that present options for implementers. In order to increase interoperability, CMAF media profiles usually limit some encoding options and specify an identifier for conforming CMAF tracks. While every codec offers its own set of choices, this subclause presents guidelines for what should be specified by audio CMAF media profiles in the form of testable assertions.

### 12.1.2.2 Audio track format

Audio CMAF media profiles should be derived from the general CMAF audio track format specified in [10.2](#) or the AAC audio CMAF track format specified in [10.2](#).

### 12.1.2.3 Loudness and dynamic range control

Audio CMAF media profiles should specify loudness and dynamic range control information.

### 12.1.2.4 Audio parameters

Audio CMAF media profiles should specify which parameters (such as loudness, dynamic range control information, channel configuration, etc.) are allowed and if they can change within a CMAF track or between CMAF tracks in a CMAF switching set.

### 12.1.2.5 Audio presentation time adjustment

Audio CMAF media profiles should specify any presentation time adjustment relative to decode time necessary on decoding, preferably without using media sample composition offsets in the TrackRunBox.

## 12.1.3 Guidelines for video CMAF media profiles

### 12.1.3.1 General

Video codecs often define features and performance limits, such as profiles, levels, video image characteristics, that present options for implementers. In order to increase interoperability, CMAF media profiles usually limit some encoding options and specify an identifier for conforming CMAF tracks. While every codec offers its own set of choices, this subclause presents guidelines for what should be specified by video CMAF media profiles in the form of testable assertions.

### 12.1.3.2 Video CMAF track format

Video CMAF media profiles should be derived from the general CMAF video track format specified in [9.2](#) or NAL structured video CMAF track format specified in [9.3](#).

### 12.1.3.3 Video image coding characteristics

Video CMAF media profiles should specify allowed options and constraints for image coding, such as colour primaries, transfer functions, colour and luminance matrix, colour spatial subsampling, bit depth, maximum height and width, frame rate limitations, etc.

### 12.1.3.4 Video rendering information

Video CMAF media profiles should specify any metadata in the video stream and CMAF track format that is necessary for accurate video rendering.

### 12.1.3.5 Video presentation time adjustment

Video CMAF media profiles should specify media sample composition offsets and any other means to align the presentation time of each coded video sequence to the earliest media sample decode time in that sequence, preferably using `TrackRunBox` composition offsets to reorder media samples without creating a presentation time delay.

### 12.1.3.6 Seamless adaptive video CMAF switching set constraints

Video CMAF media profiles should specify additional constraints on video CMAF tracks in a CMAF switching set that are necessary to enable seamless adaptive switching of CMAF fragments in decoders that conform to the video CMAF media profile. Only CMAF media profiles that support multiple CMAF tracks in a CMAF switching set should specify these CMAF switching set constraints. Seamless switching constraints should result in perceptually equivalent images when adaptive switching occurs.

### 12.1.3.7 Video decoder and display initialization

Video CMAF media profiles should specify CMAF track constraints and associated initialization process necessary to render a video CMAF track. If a video media profile also supports adaptive switching, it should specify CMAF switching set constraints and associated initialization process necessary to adaptively switch and seamlessly render a video CMAF switching set.

## 12.2 CMAF presentation profiles

### 12.2.1 General

Each CMAF presentation profile defines conditionally required media profiles for CMAF tracks that are required to be available in a CMAF presentation that conforms to the presentation profile.

Required media profiles are a minimum constraint. CMAF presentations containing the required CMAF media profiles may also include additional CMAF media profiles and remain conformant. A CMAF presentation may conform to more than one presentation profile if it contains CMAF tracks in all the conditionally required media profiles.

CMAF presentation profiles may be specified in specifications other than CMAF by entities other than MPEG. Externally specified CMAF presentation profiles should be identified by a URI in the namespace of the specifying entity. If a URL is defined, it should resolve to the specification document that defines the conditionally required media profiles of the presentation profile.

See [A.1](#) for CMAF presentation profiles defined in the CMAF specification.

### 12.2.2 CMAF profile conformance

CMAF defines several conformance points including CMAF fragments, CMAF tracks, CMAF addressable media objects, CMAF media profiles, CMAF switching sets, CMAF selection sets, and CMAF presentations.

CMAF objects and conformance points are hierarchical.

- CMAF tracks are the conformance point for ISOBMFF box, track, and file format constraints associated with the ISOBMFF compatibility brand ('`cmfc`'). CMAF headers, CMAF fragments, and CMAF chunks are objects within the CMAF track structure. CMAF track conformance provides basic compatibility between file packagers, players, parsers, and decryptors, independent of media encoders and decoders.
- General CMAF media track formats for audio, video, and subtitles are conformance points that can be referenced by CMAF media profiles. The general CMAF track formats for audio, video, and subtitles and specific CMAF track formats for NAL structured video and AAC audio are specified in [Clauses 7](#) through [11](#), which are derived from ISO/IEC 14496-2, ISO/IEC 14496-3, ISO/IEC 14496-14, ISO/IEC 14496-15 and ISO/IEC 14496-30. Conformance to one of these media track formats can

be specified by a CMAF media profile and that conformance indicated by the media profile brand. CMAF media track formats support CMAF adaptive multimedia presentations.

- c) CMAF media profiles are conformance points for encoded media data stored in a specified CMAF track format. A media profile compatibility brand indicates encode/decode and adaptive switching conformance of media data, as well as a CMAF track format. CMAF media profile conformance enables encoding and decoding of media samples in CMAF fragments and CMAF tracks, independent of delivery.
- d) CMAF addressable media objects are conformance points for CMAF media objects (headers, chunks, segments, and files) and their brands, which are derived from CMAF tracks, fragments, chunks, and headers, and optimized for external reference, storage, and delivery. They inherit the structure and media constraints of the CMAF track and media profile they contain, and constrain the structure of the CMAF addressable media object independent of the CMAF media. CMAF addressable media objects are a conformance point for streaming protocols and systems that is independent of the manifest format and delivery method used.
- e) CMAF switching sets are a conformance point for a group of CMAF tracks that share content and encoding constraints defined by the CMAF track format and a CMAF media profile to enable seamless adaptive switching. Single initialization CMAF switching set constraints and an associated identifier are defined to indicate CMAF switching set functional conformance to single initialization constraints, the syntax of which is specified by a CMAF media profile. Conformance for content pertains to a set of CMAF tracks, and conformance for decoders pertains to decoding and seamlessly rendering sequences of consecutive CMAF fragments selected from any of the CMAF tracks.
- f) CMAF selection sets are conformance points for a group of synchronized CMAF switching sets that share related content, timing, and media type. For instance, alternative audio CMAF switching sets in a CMAF selection set can cover the same presentation timespan with synchronized alternative content or encoding. Since CMAF tracks contain a single media type (audio, video, or subtitles), multiple CMAF selection sets are necessary to compose a multimedia presentation using the CMAF hypothetical application model and late binding. Conformance indicates that all CMAF tracks are correctly grouped, synchronized, and have approximately the same duration to form a CMAF presentation.
- g) CMAF presentation profiles are conformance points for a group of conditionally required CMAF tracks that conform to required media profiles and form a synchronized multimedia presentation. What content and timing are appropriate for a multimedia presentation is not specified by CMAF and can only be tested based on a presentation author's intended result. However, CMAF presentations can be tested against the CMAF hypothetical application model, which provides a common understanding between content creators and players as to the intended and expected result. A CMAF presentation profile constrains the media profiles, CMAF switching sets, track formats, etc. that it contains and is intended to identify basic content/device compatibility. A player that claims compatibility with a CMAF presentation profile is expected to decode all the required CMAF media profiles and implement the CMAF hypothetical application model to late bind CMAF tracks, synchronize them, decrypt them, render them, and perform seamless adaptive switching, if appropriate. If optional CMAF media profiles are included in a CMAF presentation, a player can select those instead, if that player supports the optional CMAF media profile.

A CMAF media profile that falls within the maximum encoding constraints of a higher CMAF media profile conforms to the higher CMAF media profile and does not need to indicate compatibility with a second CMAF media profile compatibility brand, since it is an intrinsic property of those media profiles.

**EXAMPLE** Standard definition video conforming to the CMAF SD media profile also conforms to the CMAF HD media profile because it will decode on an HD compatible system. It could be included in a CMAF switching set that conforms to the HD media profile. CMAF tracks and media profiles with different source formats, such as different colour spaces or transfer functions, do not conform to a single CMAF switching set or media profile because they are not considered the same source content and would not provide seamless adaptive switching in the same CMAF switching set and CMAF media profile.

A CMAF presentation shall conform to one or more CMAF presentation profiles by containing one or more CMAF tracks for each conditionally required media profile in the CMAF presentation profile. A CMAF presentation can conform to multiple CMAF presentation profiles by containing all the conditionally required CMAF tracks and CMAF media profiles. Additional CMAF media profiles in a CMAF presentation are optional by default when they are not conditionally required in a CMAF presentation profile that the CMAF presentation conforms to.

A presentation profile may specify a single encryption scheme to be used on every encrypted CMAF switching set that is conditionally required in the presentation.

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 23000-19:2018

## Annex A (normative)

### CMAF presentation profiles and media profiles

#### A.1 CMAF presentation profiles

##### A.1.1 Overview

This clause defines some CMAF presentation profiles and their associated CMAF presentation profile identifiers. The CMAF presentation profiles below differ only by the Common Encryption schemes they require (ISO/IEC 23001-7).

**NOTE** In the current market, two encryption schemes, 'cenc' and 'cbc3', are widely deployed on popular platforms. Therefore, this document currently defines one presentation profile for each of these encryption schemes and another for unencrypted CMAF presentations. By defining multiple presentation profiles, a mechanism is provided to select a given encryption scheme by presentation profile within a given deployment. This identifies the potential incompatibilities between content and deployments conforming to similar but different CMAF presentation profiles.

##### A.1.2 CMFHD presentation profile

The CMFHD presentation profile is intended to provide basic interoperability of unprotected content on the widest range of Internet video devices in use today. Other CMAF media profiles that are not required in this presentation profile are considered optional.

Requirements of CMAF presentation profile CMFHD

- Presentation profile ID= “[urn:mpeg:cmaf:presentation\\_profile:cmfhd:2017](urn:mpeg:cmaf:presentation_profile:cmfhd:2017)”
- If containing video, it shall include at least one CMAF switching set constrained to the 'cfhd' media profile in [A.2](#).
- If containing audio, it shall include at least one audio CMAF switching set constrained to the 'caac' media profile defined in [A.3](#).
- If containing subtitle tracks, it shall include at least one CMAF switching set for each language and role in the 'im1t' media profile defined in [A.4](#).
- CMAF tracks containing optional CMAF media profiles in CMAF switching sets should be included in selection sets with CMAF switching sets containing required media profiles.
- All CMAF tracks shall not contain encrypted media samples or a `TrackEncryptionBox`.

##### A.1.3 The CMFHDC presentation profile

The CMFHDC presentation profile is intended to provide interoperability with a large portion of video devices in use today that support content protection using the 'cenc' scheme of Common Encryption (see [8.2.3](#)). Other CMAF media profiles that are not required in this presentation profile are considered optional.

Requirements of CMAF presentation profile CMFHDC

- Presentation profile ID= “[urn:mpeg:cmaf:presentation\\_profile:cmfhdc:2017](urn:mpeg:cmaf:presentation_profile:cmfhdc:2017)”

- If containing video, it shall include at least one CMAF switching set constrained to the 'cfhd' media profile in [A.2](#).
- If containing audio, it shall include at least one audio CMAF switching set constrained to the 'caac' media profile defined in [A.3](#).
- If containing subtitle tracks, it shall include at least one CMAF switching set for each language and role in the 'im1t' media profile defined in [A.4](#).
- CMAF tracks containing optional CMAF media profiles in CMAF switching sets should be included in selection sets with CMAF switching sets containing required media profiles.
- At least one CMAF switching set shall be encrypted. Any CMAF switching set that is encrypted shall be available in 'cenc' Common Encryption scheme specified in [Clause 8](#).

#### A.1.4 The CMFHDs presentation profile

The CMFHDs presentation profile is intended to provide interoperability with a large portion of video devices in use today that support content protection using the 'cbc5' scheme of Common Encryption (see [8.2.3](#)). Other CMAF media profiles that are not required in this presentation profile are considered optional.

Requirements of CMAF presentation profile CMFHDs

- Presentation profile ID= “[urn:mpeg:cmaf:presentation.profile:cmfhds:2017](#)”
- If containing video, it shall include at least one CMAF switching set constrained to the 'cfhd' media profile in [A.2](#).
- If containing audio, it shall include at least one audio CMAF switching set constrained to the 'caac' media profile defined in [A.3](#).
- If containing subtitle tracks, it shall include at least one CMAF switching set for each language and role in the 'im1t' media profile defined in [A.4](#).
- CMAF tracks containing optional CMAF media profiles in CMAF switching sets should be included in selection sets with CMAF switching sets containing required media profiles.
- At least one CMAF switching set shall be encrypted. Any CMAF switching set that is encrypted shall be available in 'cbc5' Common Encryption scheme specified in [Clause 8](#).

#### A.1.5 Protected CMAF presentations

Manifests that include equivalent CMAF presentations in CMFHDC and CMFHDs presentation profiles can declare both presentation profile IDs to indicate that all CMAF switching sets are available with both encryption schemes.

Presentations conforming to the CMFHD presentation profile are not protected by Common Encryption of media sample data, but do not preclude encryption during delivery by methods such as HTTPS or envelope encryption that are removed on arrival, prior to media processing of the unencrypted presentation.

## A.2 AVC video CMAF media profiles and brands

A set of AVC video CMAF media profiles is defined together with an ISOBMFF brand and constraints in [Table A.1](#).

For a CMAF track to comply with one of the media profiles in [Table A.1](#), it

- shall conform to [9.4](#),

- shall not exceed the profile or level listed in the table,
- shall conform to the colour\_primaries, transfer\_characteristics and matrix\_coefficients values from the options listed in the table,
- shall not exceed the width, height or frame rate listed in the table, even if the AVC level would permit higher values, and
- should include the CMAF file brand in its CMAF header.

NOTE CMAF tracks conform to more than one media profile if they meet the requirements for multiple CMAF media profiles. For example, a video CMAF track could conform to both the SD and HD CMAF media profiles if its dimensions are within the constraints for the SD profile (the HD profile has higher dimension constraints and codec level) and the colour\_primaries, transfer\_characteristics and matrix\_coefficients are values which are common to both profiles.

See [Annex C](#) for detailed encoding examples.

**Table A.1 — AVC video CMAF media profiles**

Media profile	Codec	Profile	Level	colour_primaries field in VUI	transfer_characteristics field in VUI	matrix_coefficients field in VUI	Max frame height	Max frame width	Max frame rate	CMAF File Brand
SD	AVC	High	3.1	1 (BT.709) 5 (BT.601-7 PAL/SECAM) 6 (BT.601-7 NTSC)	1 (BT.709 OETF) 6 (BT.601-7 OETF) (see NOTE)	1 (BT.709) 5 (BT.601-7 PAL/SECAM) 6 (BT.601-7 NTSC)	576	864	60	'cfsd'
HD	AVC	High	4.0	1 (BT.709)	1 (BT.709 OETF)	1 (BT.709)	1080	1920	60	'cfhd'
HDHF	AVC	High	4.2	1 (BT.709)	1 (BT.709 OETF)	1 (BT.709)	1080	1920	60	'chdf'
NOTE Values 1 and 6 for transfer_characteristics are functionally equivalent, but the value which refers to the same specification as the colour_primaries is normally used.										

See [Clause 9](#) and [Annex C](#) for general requirements and constraints on CMAF video tracks.

### A.3 Audio CMAF media profiles and brands

The AAC audio CMAF media profiles in [Table A.2](#) are specified in [10.4](#) and [10.5](#). [Table A.2](#) summarizes key parameters and specifies the four-character code that can be used as an ISOBMFF brand to identify the CMAF media profile in CMAF tracks and CMAF media objects.

The AAC adaptive switching audio CMAF media profile ('caaa') is a constrained subset of the AAC core audio CMAF media profile ('caac'), so 'caaa' CMAF tracks always conform to the 'caac' CMAF media profile.

**Table A.2 — AAC audio CMAF media profiles**

Media profile	Codec	Profiles	Levels	Number of channels	File brand
AAC core	AAC	AAC-LC, HE-AAC or HE-AACv2	2	Mono or stereo	'caac'
AAC adaptive	AAC	AAC-LC, HE-AAC or HE-AACv2 (see NOTE)	2	Mono or stereo	'caaa'
NOTE AAC adaptive CMAF media profile is AAC core constrained for adaptive switching.					

See [Clause 10](#) for general constraints on audio CMAF tracks and [12.1.2](#) for the specification of additional audio CMAF media profiles.

#### A.4 Subtitle CMAF media profiles and brands

CMAF specifies three subtitle CMAF track formats as shown in [Table A.3](#) — WebVTT, IMSC1 text, and IMSC1 image — to provide interoperability between CMAF presentations and players. Subtitle media profiles shall conform to the CMAF subtitle track format specified in [Clause 11](#).

**Table A.3 — Subtitle and caption CMAF media profiles**

Media profile	Format	Notes	File brand
WebVTT	Specified in <a href="#">11.2</a>	ISO/IEC 14496-30	'cwvt'
TTML IMSC1 text	Specified in <a href="#">11.3.3</a>	IMSC1 text profile	'im1t'
TTML IMSC1 image	Specified in <a href="#">11.3.4</a>	IMSC1 image profile	'im1i'
CEA	CTA-608 and CTA-708 Specified in <a href="#">11.4</a>	Caption data is embedded in SEI messages in video track; multiple closed caption streams may be present	'ccea'

NOTE The 'ccea' compatibility brand can be included in addition to a video CMAF media profile compatibility brand to indicate the presence of captions embedded in the video elementary stream.

See [Clause 11](#) for more details on subtitle CMAF track formats.

## Annex B (normative)

# HEVC video CMAF track format and CMAF media profiles

### B.1 HEVC video CMAF tracks

HEVC tracks shall conform to [9.3](#), as additionally constrained in this annex.

### B.2 HEVC video track constraints

#### B.2.1 HEVC video CMAF switching set constraints

HEVC video CMAF switching set shall conform to constraints for NAL structured video CMAF switching sets specified in [9.3.6](#) or [9.3.7](#).

#### B.2.2 Sample Description Box ('stsd')

The Sample Description Box ('stsd') shall conform to [9.2.4](#) and contain one or more sample entries.

#### B.2.3 Visual sample entry

The syntax and values of a visual sample entry shall conform to HEVCSampleEntry ('hvc1') or HEVCSampleEntry ('hev1') sample entries as defined in ISO/IEC 14496-15 and constrained as follows.

The HEVCSampleEntry:

- shall contain an HEVCConfigurationBox ('hvcC') box containing an HEVCDecoderConfigurationRecord, as specified in ISO/IEC 14496-15;
- shall set sample entry fields consistent with the sequence parameter set and picture parameter set values in the video track as specified in ISO/IEC 14496-15 and constrained by [B.3](#);
- shall contain only one HEVC parameter set per HEVCSampleEntry;
- should contain SEI messages containing colour and dynamic range mastering and rendering information if default values are not encoded, as required in [9.3.5.1](#);
- shall contain a ColourInformationBox ('colr') with colour\_type 'nclx' and PixelAspectRatioBox ('pasp') when required per [9.3.5.1](#).

NOTE The ColourInformationBox with colour\_type 'nclx' and PixelAspectRatioBox provide equivalent information to SPS VUI and are required when SPS VUI is not present in an 'hev1' sample entry.

#### B.2.4 HEVCDecoderConfigurationRecord colour and dynamic range information

- As specified in [9.3.5.1](#), video captured or colour graded with characteristics other than BT.709 defaults should include one or more SEI NALs with additional transfer characteristics or colour volume information stored in the HEVCDecoderConfigurationRecord to enable colour and dynamic range calibration during decoder and display initialization. This may include the following SEI messages specified in ISO/IEC 23008-2:2015, Annex D:
  - SEI payloadType 137, mastering\_display\_colour\_volume;

- SEI payloadType 144, content\_light\_level\_info;
- SEI payloadType 147, alternative\_transfer\_characteristics.
- CMAF player display processors can use SEI colour grading, colour volume, and dynamic range messages in SEI NALs stored in the HEVCDecoderConfigurationRecord to calibrate rendering for display characteristics and viewing conditions.
- Alternative transfer characteristics in an SEI message can be used to optimize rendering of hybrid log gamma transfer function video on high dynamic range displays while using SPS signalled gamma rendering on standard range displays.
- A CMAF player can pass MaxFALL, MaxCLL, and other SEI message data to a display over a digital video interface, as specified in ANSI/CTA 861-G, to enable colour and transfer function optimization within a compatible UHD/HDR display.
- CMAF switching sets shall be constrained to include identical SEI NALs and SPS VUI colour mastering and dynamic range information in the first sample entry of every CMAF header in the CMAF switching set to provide consistent initialization and calibration.

### B.2.5 Track Header Box ('tkhd')

The requirements of [9.2.3](#) apply.

NOTE Normalized width and height can be derived from the track's visual sample entry for 'hvc1' video media samples (see [9.3.2.2](#)) or from a Sequence Parameter Set NAL in each Coded Video Sequence for 'hev1' video media samples. See [9.3.3](#) and [9.3.4](#) for the storage and semantics of video Sequence Parameter Sets.

## B.3 Media sample and CMAF fragment constraints

### B.3.1 Storage of HEVC elementary streams

#### B.3.1.1 Conformance

HEVC video tracks shall comply with ISO/IEC 14496-15 and [9.3](#).

#### B.3.2 Access Units

Access Units and media samples shall conform to [9.3](#).

Access Units shall conform to the requirements of a media sample of the indicated description ('hvc1' or 'hev1') as specified in ISO/IEC 14496-15.

CMAF fragments containing access units identified by the 'hev1' sample description shall contain all SPS and PPS NALs referenced from a Coded Video Sequence in the first Access Unit of that sequence, immediately following its first Access Unit Delimiter NAL, if an Access Unit Delimiter NAL is present.

Access Units identified by the 'hev1' sample description may retain filler data (in NAL units or SEI messages) and SEI messages that would change hypothetical reference decoder bitstream conformance if removed.

Access Units of type 'hvc1' shall reference a video parameter set in the sample entry of the CMAF header associated with the containing CMAF track.

### B.3.3 Constraints on HEVC elementary streams

#### B.3.3.1 Overview

The following general constraints apply to all CMAF HEVC elementary streams, and their values are additionally constrained in [B.5](#) with constraints on tier, profile, level, resolution, video characteristics, and frame rates specified by HEVC video CMAF media profile in [Table B.1](#).

#### B.3.3.2 Picture type

All pictures shall be encoded as coded frames and shall not be encoded as coded fields.

#### B.3.3.3 Video Parameter Sets (VPS)

Each HEVC video media sample in the CMAF track shall reference the VPS in the CMAF header sample entry. VPS shall not change within CMAF tracks or between CMAF tracks in a CMAF switching set. A CMAF HEVC track shall conform to ISO/IEC 23008-2 with the following additional constraints.

- The following fields shall have pre-determined values as follows.
  - general\_progressive\_source\_flag shall be set to 1.
  - general\_frame\_only\_constraint\_flag shall be set to 1.
  - general\_interlaced\_source\_flag shall be set to 0.
- The condition of the following fields shall not change throughout an HEVC elementary stream:
  - general\_profile\_space
  - general\_profile\_idc
  - general\_tier\_flag
  - general\_level\_idc

#### B.3.3.4 Sequence Parameter Sets (SPS)

##### B.3.3.4.1 SPS fields

Sequence Parameter Set NAL units that occur within a CMAF HEVC track shall conform to ISO/IEC 23008-2 with the following additional constraints.

- The following fields shall have pre-determined values as follows.
  - vui\_parameters\_present\_flag shall be set to 1.

##### B.3.3.4.2 Visual Usability Information (VUI) parameters

VUI parameters that occur within a CMAF HEVC track shall conform to ISO/IEC 23008-2 with the following additional constraints.

- The following fields shall have pre-determined values as follows.
  - aspect\_ratio\_info\_present\_flag shall be set to 1.
  - video\_full\_range\_flag shall be set to 0.
- The following fields have the following values.
  - colour\_description\_present\_flag should be set to 1.