INTERNATIONAL STANDARD

ISO/IEC 23001-11

Second edition
2019-03

Information technology — MPEG systems technologies —

Part 11: Energy-efficient media consumption (green metadata)

Technologies de l'information — Technologies des systèmes MPEG —

Partie 11: Consommation des supports éconergétiques (métadonnées vertes)

Reference number
ISO/IEC 23001-11:2019(E)

Contents

Page

Foreword .... **v**

Introduction .... **vi**

1 Scope **1**

2 Normative references **1**

3 Terms, definitions, symbols, abbreviated terms and conventions **2**
- 3.1 Terms and definitions 2
- 3.2 Symbols and abbreviated terms 3
- 3.3 Conventions 4
  - 3.3.1 Arithmetic operators 4
  - 3.3.2 Mathematical functions 5

4 Functional architecture **5**
- 4.1 Description of the functional architecture 5
- 4.2 Definition of components in the functional architecture 6

5 Decoder power reduction **7**
- 5.1 General 7
- 5.2 Complexity metrics for decoder-power reduction 7
  - 5.2.1 General 7
  - 5.2.2 Syntax 7
  - 5.2.3 Signalling 10
  - 5.2.4 Semantics 10
- 5.3 Interactive signalling for remote decoder-power reduction 26
  - 5.3.1 General 26
  - 5.3.2 Syntax 26
  - 5.3.3 Signalling 26
  - 5.3.4 Semantics 26

6 Display power reduction using display adaptation **26**
- 6.1 General 26
- 6.2 Syntax 26
  - 6.2.1 Systems without a signalling mechanism from the receiver to the transmitter 26
  - 6.2.2 Systems with a signalling mechanism from the receiver to the transmitter 27
- 6.3 Signalling 27
  - 6.3.1 General 27
  - 6.3.2 Systems without a signalling mechanism from the receiver to the transmitter 28
  - 6.3.3 Systems with a signalling mechanism from the receiver to the transmitter 28
- 6.4 Semantics 28

7 Energy-efficient media selection **29**
- 7.1 General 29
- 7.2 Syntax 30
- 7.3 Signalling 30
- 7.4 Semantics 30
  - 7.4.1 Decoder-power indication metadata semantics 30
  - 7.4.2 Display-power indication metadata semantics 31

8 Metrics for quality recovery after low-power encoding **31**
- 8.1 General 31
- 8.2 Syntax 31
- 8.3 Signalling 32
- 8.4 Semantics 32

9 Conformance and reference software **32**

Annex A (normative) **Supplemental enhancement information (SEI) syntax** **33**

Annex B (informative) **Implementation guidelines for the usage of green metadata** **37**

Annex C (normative) **Conformance and reference software** **62**
Bibliography **66**

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.

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 document 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).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC 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) or the IEC list of patent declarations received (see http://patents.iec.ch).

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

For an explanation of 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 www.iso.org/iso/foreword.html.

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

This second edition cancels and replaces the first edition (ISO/IEC 23001-11:2015), which has been technically revised. It also incorporates the Amendments ISO/IEC 23001-11:2015/Amd 1:2016 and ISO/IEC 23001-11:2015/Amd 2:2018. The main changes compared to the previous edition are as follows:

— specification of an HEVC SEI message carrying green metadata and modification of text specifying the carriage of green Metadata in an AVC SEI message so that the AVC and HEVC SEI messages are consistent;

— inclusion of Annex C which specifies conformance-verification procedures for the power-reduction technologies specified in this document, precises the role of the reference software for each technology and gives the links to reference softwares and test vectors.

— specification of HEVC Complexity metrics and improvement of the existing AVC Complexity metrics.

A list of all parts in the ISO/IEC 23001 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

This document specifies the metadata (green metadata) that facilitates reduction of energy usage during media consumption as follows:

— the format of the metadata that enables reduced decoder power consumption;

— the format of the metadata that enables reduced display power consumption;

— the format of the metadata that enables media selection for joint decoder and display power reduction;

— the format of the metadata that enables quality recovery after low-power encoding.

This metadata facilitates reduced energy usage during media consumption without any degradation in the quality of experience (QoE). However, it is also possible to use this metadata to get larger energy savings, but at the expense of some QoE degradation.

Information technology — MPEG systems technologies —

Part 11:
Energy-efficient media consumption (green metadata)

1 Scope

This document specifies metadata for energy-efficient decoding, encoding, presentation and selection of media.

The metadata for energy-efficient decoding specifies two sets of information: complexity metrics (CM) metadata and decoding operation reduction request (DOR-Req) metadata. A decoder uses CM metadata to vary operating frequency and thus reduce decoder power consumption. In a point-to-point video conferencing application, the remote encoder uses the DOR-Req metadata to modify the decoding complexity of the bitstream and thus reduce local decoder power consumption.

The metadata for energy-efficient encoding specifies a quality metric that is used by a decoder to reduce the quality loss from low-power encoding.

The metadata for energy-efficient presentation specifies RGB-component statistics and quality levels. A presentation subsystem uses this metadata to reduce power by adjusting display parameters, based on the statistics, to provide a desired quality level from those provided in the metadata.

The metadata for energy-efficient media selection specifies decoder operation reduction ratios (DOR-Ratios), RGB-component statistics and quality levels. The client in an adaptive streaming session uses this metadata to determine decoder and display power-saving characteristics of available video representations and to select the representation with the optimal quality for a given power-saving.

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 13818-1, *Information technology — Generic coding of moving pictures and associated audio information — Part 1: Systems*

ISO/IEC 14496-10:—[1]), *Information technology — Coding of audio-visual objects — Part 10: Advanced video coding*

ISO/IEC 23001-10, *Information technology — MPEG systems technologies — Part 10: Carriage of timed metadata metrics of media in ISO base media file format*

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:—[2]), *Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats*

ISO/IEC/TR 23009-3, *Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 3: Implementation guidelines*

1) Under preparation. Stage at the time of publication: ISO/IEC DIS 14496-10:2018.

2) Under preparation. Stage at the time of publication: ISO/IEC FDIS 23009-1:2019.

3 Terms, definitions, symbols, abbreviated terms and conventions

For the purposes of this document, the terms and definitions given in ISO/IEC 14496-10, ISO/IEC 23008-2 and ISO/IEC 23009-1 and the following apply.

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

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1 Terms and definitions

3.1.1
alpha-point deblocking instance
APDI
single filtering operation that produces either a single, filtered output p'_0 or a single, filtered output q'_0, where p'_0 and q'_0 are filtered samples across a 4x4 block edge

3.1.2
chroma_format_idc
chroma sampling relative to the luma sampling

3.1.3
deblocking filtering instance
single filtering operation that produces either a single, filtered output p' or a single, filtered output q', where p' and q' are filtered samples across a 8x8 block edge

3.1.4
decoding process
process that reads a bitstream and derives decoded pictures from it

Note 1 to entry: This process is specified in ISO/IEC 14496-10 or ISO/IEC 23008-2.

3.1.5
display process
process that takes, as its input, the cropped decoded pictures that are the output of the *decoding process* (3.1.4)

3.1.6
encoder
embodiment of an *encoding process* (3.1.7)

3.1.7
encoding process
process that produces a bitstream

Note 1 to entry: The bitstream produced is conforming to ISO/IEC 14496-10 or ISO/IEC 23008-2.

3.1.8
no-quality-loss operating point
NQLOP
metadata-enabled operating point associated with the largest display-power reduction that can be achieved without any quality loss (infinite PSNR)

3.1.9
non-zero block
block containing at least one non-zero transform coefficient

3.1.10
peak signal
maximum permissible *RGB component* (3.1.16) in a *reconstructed frame* (3.1.14)

Note 1 to entry: For 8-bit video, the peak signal is 255.

3.1.11
period
interval over which complexity-metrics metadata are applicable

3.1.12
PicSizeInMbs
product of the picture width and the picture height in units of macroblocks

3.1.13
pixel
smallest addressable element in an all-points addressable display device

3.1.14
reconstructed frames
frames obtained after applying *RGB colour-space* (3.1.15) conversion and cropping to the specific decoded picture or pictures for which display power-reduction metadata are applicable

3.1.15
RGB colour space
colour space based on the red, green and blue primaries

3.1.16
RGB component
single sample representing one of the three primary colours of the *RGB colour space* (3.1.15)

3.1.17
separate_colour_plane_flag
flag that, when set, specifies that the three colour components of the 4:4:4 chroma format are coded separately

3.1.18
six-tap filtering
STF
single application of the 6-tap filter to generate a single filtered sample for fractional positions using the samples at integer-sample positions

3.2 Symbols and abbreviated terms

For the purposes of this document, the symbols and abbreviated terms given in the following apply:

APDI	alpha-point deblocking instance
ASIC	application specific integrated circuit
AVC	advanced video coding — ISO/IEC 14496-10
BMFF	base media file format
CM	complexity metric
CMOS	complementary metal oxide semiconductor
CPU	central processing unit

DASH	dynamic adaptive streaming over HTTP
DOR-Ratio	decoding operation reduction ratio
DOR-Req	decoding operation reduction request
DVFS	dynamic voltage frequency scaling
Fps	frames per second
FS	fresh start
GP	good picture
HEVC	high efficiency video coding — ISO/IEC 23008-2
Mbps	mega bits per second
MPD	media presentation description
MSD	mean square difference
MV	motion vector
NQLOP	no-quality-loss operating point
PSNR	peak signal to noise ratio
QoE	quality of experience
RBLL	remaining battery life level
RGB	red, green, blue
SEI	supplemental enhancement information
SP	start picture
STF	six-tap filtering
XSD	cross-segment decoding

3.3 Conventions

3.3.1 Arithmetic operators

+	Addition
−	Subtraction (as a two-argument operator) or negation (as a unary prefix operator)
*	Multiplication
x^y	Exponentiation

x/y	Division where no truncation or rounding is intended
$\frac{x}{y}$	Division where no truncation or rounding is intended
$\sum_{i=x}^{y} f(i)$	Summation of *f*(*i*) with *i* taking all integer values from *x* up to and including *y*

3.3.2 Mathematical functions

Mathematical functions in this document are defined as follows:

$$\text{Abs}(x)=\begin{cases}-x, & x<0\\ x, & x\geq 0\end{cases} \quad (1)$$

$$\text{Clip}(x)=\begin{cases}x, & x<256\\ 255, & \text{otherwise}\end{cases} \quad (2)$$

Floor(*x*) is the greatest integer less than or equal to *x* (3)

Log10(*x*) returns the base-10 logarithm of *x* (4)

Round(*x*) = Sign(*x*) * Floor(Abs(*x*) + 0.5) (5)

$$\text{Sign}(x)=\begin{cases}-1, & x<0\\ 1, & x\geq 0\end{cases} \quad (6)$$

4 Functional architecture

This clause is informative and placed here to provide context.

4.1 Description of the functional architecture

Figure 1 shows the functional architecture utilizing green metadata in this document. The media pre-processor is applied to analyse and to filter the content source and a video encoder is used to encode the content to a bitstream for delivery. The bitstream is delivered to the receiver and decoded by a video decoder with the output rendered on a presentation subsystem that implements a display process.

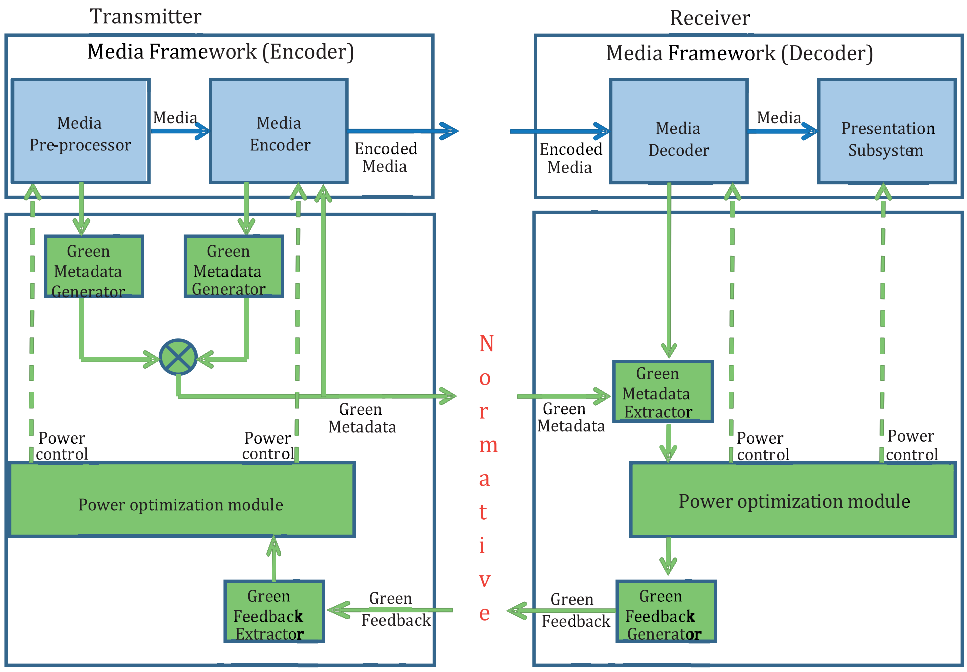

Figure 1 — Functional architecture

The green metadata is extracted from either the media encoder or the media pre-processor. In both cases, the green metadata is multiplexed or encapsulated in the conformant bitstream. Such green metadata is used at the receiver to reduce the power consumption for video decoding and presentation. The bitstream is packetized and delivered to the receiver for decoding and presentation. At the receiver, the metadata extractor processes the packets and sends the green metadata to a power optimization module for efficient power control. For instance, the power optimization module interprets the green metadata and then applies appropriate operations to reduce the video decoder's power consumption when decoding the video and also to reduce the presentation subsystem's power consumption when rendering the video. In addition, the power-optimization module can collect receiver information, such as remaining battery capacity, and send it to the transmitter as green feedback to adapt the encoder operations for power-consumption reduction.

The normative aspect of this document is limited to the green metadata and green feedback in Figure 1.

4.2 Definition of components in the functional architecture

Green metadata generator

— Generates metadata from either the video encoder or the content pre-processor.

Green metadata extractor

— Interprets the bitstream syntax information and sends it to the power optimization module in the receiver.

Green feedback generator

— Generates feedback information for the transmitter.

— Communicates with the transmitter through a feedback channel, if available, for energy-efficient processing.

Green feedback extractor

— Receives the feedback from the receiver and sends it to the power optimization module in the transmitter.

Power optimization module in the transmitter

— Collects platform statistics such as the remaining battery capacity of the device in which the transmitter resides.

— Controls the operation of the green metadata generator, video encoder and content pre-processor.

— Processes green feedback.

Power optimization module in the receiver

— Processes the green-metadata information and applies appropriate operations for power-consumption control.

— Collects platform statistics such as remaining battery capacity of the device in which the receiver resides.

— Sends requests to Green feedback generator.

5 Decoder power reduction

5.1 General

Energy-efficient decoding is achieved with two types of metadata: complexity metrics (CMs) metadata and decoding operation reduction request (DOR-Req) metadata. A decoder may use CMs metadata to vary operating frequency and thus reduce decoder power consumption. In a point-to-point video conferencing application, the remote encoder may use the DOR-Req metadata to modify the decoding complexity of the bitstream and thus reduce local decoder power consumption.

5.2 Complexity metrics for decoder-power reduction

5.2.1 General

With respect to the functional architecture in Figure 1, the green-metadata generator provides CMs that indicate the picture-decoding complexity of an AVC or HEVC bitstream to the decoder.

5.2.2 Syntax

The syntax for the AVC CMs is given in Table 1.

Table 1 — Syntax for the AVC CMs

	Size (bits)	Descriptor
period_type	8	unsigned integer
if (period_type = = 2) || (period_type == 7) {		
num_seconds	16	unsigned integer
}		
else if (period_type = = 3) || (period_type == 8) {		
num_pictures	16	unsigned integer
}		
if (period_type == 8) {		

Table 1 *(continued)*

temporal_map		
for (t=0; t<8; t++) {		
if ((temporal_map>>t)%2 == 1)		
num_pictures_in_temporal_layers[t]		
}		
}		
if (period_type <= 3) {		
portion_non_zero_8x8_blocks	8	unsigned integer
portion_intra_predicted_macroblocks	8	unsigned integer
portion_six_tap_filterings	8	unsigned integer
portion_alpha_point_deblocking_instances	8	unsigned integer
}		
else if (period_type == 4) {		
for (i=0; i<= num_slice_groups_minus1; i++) {		
num_slices_minus1[i]	16	unsigned integer
}		
for (i=0; i<= num_slice_groups_minus1; i++) {		
for (j=0; j<=num_slices_minus1[i]; j++) {		
first_mb_in_slice[i][j]	16	unsigned integer
portion_non_zero_8x8_blocks[i][j]	8	unsigned integer
portion_intra_predicted_macroblocks[i][j]	8	unsigned integer
portion_six_tap_filterings[i][j]	8	unsigned integer
portion_alpha_point_deblocking_instances[i][j]	8	unsigned integer
}		
}		
}		
else if (period_type >= 5) && (period_type <= 8) {		
num_layers_minus1	16	unsigned integer
for (l=0; l<= num_layers_minus1; l++) {		
picture_parameter_set_id[l]	8	unsigned integer
priority_id[l]	6	unsigned integer
dependency_id[l]	3	unsigned integer
quality_id[l]	4	unsigned integer
temporal_id[l]	3	unsigned integer
portion_non_zero_8x8_blocks[l]	8	unsigned integer
portion_intra_predicted_macroblocks[l]	8	unsigned integer
portion_six_tap_filterings[l]	8	unsigned integer
portion_alpha_point_deblocking_instances[l]	8	unsigned integer
}		
}		

The syntax for the HEVC CMs is given in Table 2.

Table 2 — Syntax for the HEVC CMs

	Size (bits)	Descriptor
period_type	8	unsigned integer
if (period_type == 2) {		
num_seconds	16	unsigned integer
}		
else if (period_type == 3) {		
num_pictures	16	unsigned integer
}		
if (period_type <= 3) {		
portion_non_zero_blocks_area	8	unsigned integer
if (portion_non_zero_blocks_area != 0) {		
portion_8x8_blocks_in_non_zero_area	8	unsigned integer
portion_16x16_blocks_in_non_zero_area	8	unsigned integer
portion_32x32_blocks_in_non_zero_area	8	unsigned integer
}		
portion_intra_predicted_blocks_area	8	unsigned integer
if (portion_intra_predicted_blocks_area == 255) {		
portion_planar_blocks_in_intra_area	8	unsigned integer
portion_dc_blocks_in_intra_area	8	unsigned integer
portion_angular_hv_blocks_in_intra_area	8	unsigned integer
}		
else {		
portion_blocks_a_c_d_n_filterings	8	unsigned integer
portion_blocks_h_b_filterings	8	unsigned integer
portion_blocks_f_i_k_q_filterings	8	unsigned integer
portion_blocks_j_filterings	8	unsigned integer
portion_blocks_e_g_p_r_filterings	8	unsigned integer
}		
portion_deblocking_instances	8	unsigned integer
}		
else if (period_type == 4) {		
max_num_slices_tiles_minus1	16	unsigned integer
for (t=0; t<=max_num_slices_tiles_minus1; t++) {		
first_ctb_in_slice_or_tile[t]	16	unsigned integer
portion_non_zero_blocks_area[t]	8	unsigned integer
if (portion_non_zero_blocks_area[t] != 0) {		
portion_8x8_blocks_in_non_zero_area[t]	8	unsigned integer
portion_16x16_blocks_in_non_zero_area[t]	8	unsigned integer
portion_32x32_blocks_in_non_zero_area[t]	8	unsigned integer
}		
portion_intra_predicted_blocks_area[t]	8	unsigned integer
if (portion_intra_predicted_blocks_area[t] == 255) {		
portion_planar_blocks_in_intra_area[t]	8	unsigned integer
portion_dc_blocks_in_intra_area[t]	8	unsigned integer
portion_angular_hv_blocks_in_intra_area[t]	8	unsigned integer