FINAL
DRAFT

INTERNATIONAL STANDARD

ISO/IEC FDIS 23001-11

ISO/IEC JTC **1**/SC **29**

Secretariat: **JISC**

Voting begins on:
2022-11-23

Voting terminates on:
2023-01-18

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)

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.

Reference number
ISO/IEC FDIS 23001-11:2022(E)

Contents

Page

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 or www.iec.ch/members_experts/refdocs).

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 https://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. In the IEC, see www.iec.ch/understanding-standards.

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 third edition cancels and replaces the second edition (ISO/IEC 23001-11:2019), which has been technically revised.

The main changes are as follows:

— 6.2 related to complexity metrics for decoder-power reduction is amended by the specification of a new VVC SEI message carrying complexity metrics for decoder-power reduction.

— Clause 9 related to metrics for quality recovery after low-power encoding is amended by the specification of additional metrics for quality recovery after low-power encoding in the newly added VVC SEI message.

— 6.3 related to interactive signalling for remote decoder-power reduction is amended by adding new syntax elements allowing a finer control by decoder of the encoding operations.

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

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 and www.iec.ch/national-committees.

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.

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 quality metrics that are 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 DOR-Req parameters, 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.

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.

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

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*

ISO/IEC 23090-3, *Information technology — Coded representation of immersive media — Part 3: Versatile video coding*

3 Terms, definitions, symbols, abbreviated terms

3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 14496-10, ISO/IEC 23008-2, ISO/IEC 23009-1, ISO/IEC 23090-3 and the following apply.

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

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

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

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
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 an 8x8 and 4x4 block edge for HEVC and VVC, respectively

3.1.3
decoding process
process that reads a bitstream and derives decoded pictures from it

Note 1 to entry: The decoding process is specified in ISO/IEC 14496-10, ISO/IEC 23008-2 or ISO/IEC 23090-3.

3.1.4
encoding process
process that produces a bitstream

Note 1 to entry: The bitstream shall conform to ISO/IEC 14496-10, ISO/IEC 23008-2 or ISO/IEC 23090-3.

3.1.5
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)

3.1.6
non-zero block
block containing at least one non-zero transform coefficient

3.1.7
peak signal
maximum permissible RGB component in a reconstructed frame

Note 1 to entry: For N-bit video, peak signal is $(2^N – 1)$.

3.1.8
period
interval over which complexity-metrics metadata are applicable

3.1.9
pixel
smallest addressable element in an all-points addressable display device

3.1.10
reconstructed frames
frames obtained after applying RGB colour-space conversion and cropping to the specific decoded picture or pictures for which display power-reduction metadata are applicable

3.1.11
RGB component
single sample representing one of the three primary colours of the RGB colour space

3.1.13
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

3.2.1 Symbols

+	addition
-	subtraction (as a two-argument operator) or negation (as a unary prefix operator)
*	multiplication

/ integer division with truncation of the result toward zero. For example, 7 / 4 and −7 / −4 are truncated to 1 and −7 / 4 and 7 / −4 are truncated to −1

÷ division in mathematical equations where no truncation or rounding is intended

3.2.2 Abbreviations

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
CMP	cubemap projection format
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
ERP	equi-rectangular projection format
Fps	frames per second
FS	fresh start
GP	good picture
HEVC	high efficiency video coding – ISO/IEC 23008-2
HCMP	hemisphere cubemap projection format
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
SSIM	structural similarity index measure
VVC	versatile video coding – ISO/IEC 23090-3
XSD	cross-segment decoding
wPSNR	weighted peak signal-to-noise ratio
WS-PSNR	weighted to spherically uniform peak signal-to-noise ratio

4 Conventions

4.1 Arithmetic operators

x^y	exponentiation
x/y	division where no truncation or rounding is intended
	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
$\sum_{p \text{ in } B}^{y} f(p)$	summation of $f(p)$ with p taking all integer location values in a block B in a picture
$x \% y$	Modulus. Remainder of x divided by y, defined only for integers x and y with $x >= 0$ and $y > 0$

4.2 Relational operators

>	greater than
>=	greater than or equal to
<	less than
<=	less than or equal to
= =	equal to
!=	not equal to

When a relational operator is applied to a syntax element or variable that has been assigned the value "na" (not applicable), the value "na" is treated as a distinct value for the syntax element or variable. The value "na" is considered not to be equal to any other value.

4.3 Bit-wise operators

x >> y	arithmetic right shift of a two's complement integer representation of x by y binary digits

This function is defined only for non-negative integer values of y. Bits shifted into the most significant bits (MSBs) as a result of the right shift have a value equal to the MSB of x prior to the shift operation.

x << y arithmetic left shift of a two's complement integer representation of x by y binary digits

This function is defined only for non-negative integer values of y. Bits shifted into the least significant bits (LSBs) as a result of the left shift have a value equal to 0.

4.4 Assignment operators

= assignment operator

++ increment, i.e., x++ is equivalent to $x = x + 1$; when used in an array index, evaluates to the value of the variable prior to the increment operation

− − decrement, i.e., x− − is equivalent to $x = x - 1$; when used in an array index, evaluates to the value of the variable prior to the decrement operation

+= increment by amount specified, i.e., x += 3 is equivalent to $x = x + 3$, and x += (−3) is equivalent to $x = x + (-3)$

−= decrement by amount specified, i.e., x −= 3 is equivalent to $x = x - 3$, and x −= (−3) is equivalent to $x = x - (-3)$

4.5 Range notation

$x = y..z$ x takes on integer values starting from y to z, inclusive, with x, y, and z being integer numbers and z being greater than or equal to y

4.6 Mathematical functions

Mathematical functions are defined as follows:

$$\text{Abs}(x) = \begin{cases} -x, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (4\text{-}1)$$

$$\text{Clip}(x) = \begin{cases} x, & x < 256 \\ 255, & \text{otherwise} \end{cases} \quad (4\text{-}2)$$

$$\text{Clip3}(x, y, z) = \begin{cases} x & ; \quad z < x \\ y & ; \quad z > y \\ z & ; \quad \text{otherwise} \end{cases} \quad (4\text{-}3)$$

Floor(x) is the greatest integer less than or equal to x (4-4)

Log10(x) returns the base-10 logarithm of x (4-5)

Round(x) = Sign(x) * Floor(Abs(x) + 0.5) (4-6)

$$\text{Sign}(x) = \begin{cases} -1, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (4\text{-}7)$$

x^y specifies x to the power of y (4-8) power(x, y) specifies x to the power of y (4-9)

4.7 Specification of syntax functions and descriptors

The following function is used in the specification of the syntax:

read_bits(n) reads the next n bits from the bitstream and advances the bitstream pointer by n bit positions. When n is equal to 0, read_bits(n) is specified to return a value equal to 0 and to not advance the bitstream pointer.

The following descriptors specify the parsing process of each syntax element:

— u(n): unsigned integer using n bits. The parsing process for this descriptor is specified by the return value of the function read_bits(n) interpreted as a binary representation of an unsigned integer with most significant bit written first.

— s(n): signed integer using n bits. The parsing process for this descriptor is specified by the return value of the function read_bits(n) interpreted as a two's complement integer representation with most significant bit written first.

5 Functional architecture

5.1 Description of the functional architecture

Figure 1 shows the functional architecture utilizing green metadata. 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.

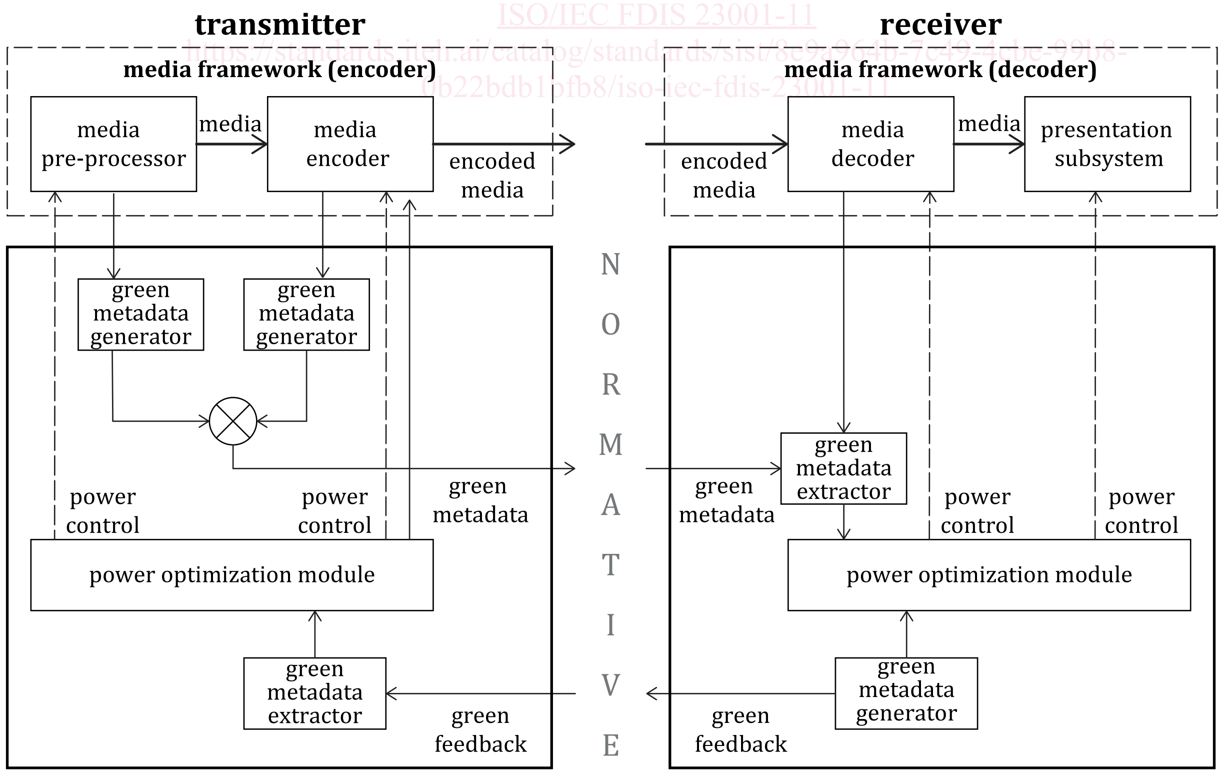

Figure 1 — Functional architecture

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 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.

5.2 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.

6 Decoder power reduction

6.1 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.

6.2 Complexity metrics for decoder-power reduction

6.2.1 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, HEVC or VVC bitstream to the decoder.

6.2.2 Syntax

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

Table 1 — Syntax for the AVC CMs

<table>
<tr><th></th><th>Descriptor</th></tr>
<tr><td>period_type</td><td>u(8)</td></tr>
<tr><td>if ((period_type = = 2) || (period_type = = 7)) {</td><td></td></tr>
<tr><td> num_seconds</td><td>u(16)</td></tr>
<tr><td>}</td><td></td></tr>
<tr><td>else if ((period_type = = 3) || (period_type = = 8)) {</td><td></td></tr>
<tr><td> num_pictures</td><td>u(16)</td></tr>
<tr><td>}</td><td></td></tr>
<tr><td>if (period_type = = 8) {</td><td></td></tr>
<tr><td> temporal_map</td><td>u(8)</td></tr>
<tr><td> for (t=0; t<8; t++) {</td><td></td></tr>
<tr><td>   if ((temporal_map>>t)%2 = = 1)</td><td></td></tr>
<tr><td>   num_pictures_in_temporal_layers[t]</td><td>u(16)</td></tr>
<tr><td> }</td><td></td></tr>
<tr><td>}</td><td></td></tr>
<tr><td>if (period_type <= 3) {</td><td></td></tr>
<tr><td> portion_non_zero_8x8_blocks</td><td>u(8)</td></tr>
<tr><td> portion_intra_predicted_macroblocks</td><td>u(8)</td></tr>
<tr><td> portion_six_tap_filterings</td><td>u(8)</td></tr>
<tr><td> portion_alpha_point_deblocking_instances</td><td>u(8)</td></tr>
<tr><td>}</td><td></td></tr>
<tr><td>else if (period_type = = 4) {</td><td></td></tr>
<tr><td> for (i=0; i<= num_slice_groups_minus1; i++) {</td><td></td></tr>
<tr><td>   num_slices_minus1[i]</td><td>u(16)</td></tr>
<tr><td> }</td><td></td></tr>
<tr><td> for (i=0; i<= num_slice_groups_minus1; i++) {</td><td></td></tr>
<tr><td>   for (j=0; j<=num_slices_minus1[i]; j++) {</td><td></td></tr>
<tr><td>   first_mb_in_slice[i][j]</td><td>u(16)</td></tr>
<tr><td>   portion_non_zero_8x8_blocks[i][j]</td><td>u(8)</td></tr>
</table>

Table 1 *(continued)*

portion_intra_predicted_macroblocks[i][j]	u(8)
portion_six_tap_filterings[i][j]	u(8)
portion_alpha_point_deblocking_instances[i][j]	u(8)
}	
}	
}	
else if (period_type >= 5) && (period_type <= 8) {	
num_layers_minus1	u(16)
for (l=0; l<= num_layers_minus1; l++) {	
picture_parameter_set_id[l]	u(8)
priority_id[l]	u(6)
dependency_id[l]	u(3)
quality_id[l]	u(4)
temporal_id[l]	u(3)
portion_non_zero_8x8_blocks[l]	u(8)
portion_intra_predicted_macroblocks[l]	u(8)
portion_six_tap_filterings[l]	u(8)
portion_alpha_point_deblocking_instances[l]	u(8)
}	
}	

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

Table 2 — Syntax for the HEVC CMs

	Descriptor
period_type	u(8)
if (period_type = = 2) {	
num_seconds	u(16)
}	
else if (period_type = = 3) {	
num_pictures	u(16)
}	
if (period_type <= 3) {	
portion_non_zero_blocks_area	u(8)
if (portion_non_zero_blocks_area != 0) {	
portion_8x8_blocks_in_non_zero_area	u(8)
portion_16x16_blocks_in_non_zero_area	u(8)
portion_32x32_blocks_in_non_zero_area	u(8)
}	
portion_intra_predicted_blocks_area	u(8)
if (portion_intra_predicted_blocks_area = = 255) {	
portion_planar_blocks_in_intra_area	u(8)
portion_dc_blocks_in_intra_area	u(8)
portion_angular_hv_blocks_in_intra_area	u(8)
}	
else {	
portion_blocks_a_c_d_n_filterings	u(8)