INTERNATIONAL STANDARD

ISO/IEC 21122-2

Second edition
2022-03

Information technology — JPEG XS low-latency lightweight image coding system —

Part 2: Profiles and buffer models

Technologies de l'information — Système de codage d'images léger à faible latence JPEG XS —

Partie 2: Profils et modèles tampons

Reference number
ISO/IEC 21122-2:2022(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 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 second edition cancels and replaces the first edition (ISO/IEC 21122-2:2019), which has been technically revised.

The main changes are as follows:

— addition of new profiles to compress colour filter array images (CFA images), to allow mathematically lossless image compression, and to compress 4:2:0 colour sampled images.

A list of all parts in the ISO/IEC 21122 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 is part of a series of standards for a low-latency lightweight image coding system, denoted as JPEG XS.

While ISO/IEC 21122-1 specifies a full set of compression coding tools needed to satisfy all of the requirements of JPEG XS, a targeted application can often work with a simpler and reduced set of coding tools, and with or without tighter constraints, to meet its targeted goals. For this reason, profiles, levels, and sublevels are defined in this document. These three concepts facilitate partial and reduced complexity implementations of ISO/IEC 21122-1 for such specific application use cases, while also safeguarding interoperability.

This document specifies a limited number of profiles to represent interoperability subsets of the codestream syntax specified in ISO/IEC 21122-1 with each profile serving specific application use cases. In other word, profiles select a subset of the available coding tools. In addition, levels and sublevels provide limits to the maximum throughput in respectively the encoded (codestream) and the decoded (spatial/pixel) domains. In this way, profiles, levels and sublevels allow designing cost-efficient implementations that serve the needs of the desired applications.

In addition to being light-weight, another major requirement of JPEG XS is to allow low end-to-end latency, limited to a fraction of the frame size. To ensure this low-latency property, this document also specifies a buffer model, consisting of a decoder model and a transmission channel model. The models show the interaction of a hypothetical reference decoder, including its smoothing buffer with a constant bitrate channel feeding this buffer. The size of the decoder smoothing buffer is computed from the profile, level, and sublevel. Codestreams are formed such that the buffer of a decoder, operating according to this buffer model, never overflows or underflows. In effect, the buffer model provides encoders with the necessary information to generate codestreams that can be decoded by an arbitrary decoder implementation, ensuring system interoperability.

In addition to the size of the decoder smoothing buffer, end-to-end latency also depends on the latency inherent to each processing step of the encoding-decoding chain whose methods are described in ISO/IEC 21122-1. To help implementers estimate the latency of their device, this document gives extra information on the minimum latency that can be achieved by the different methods described in ISO/IEC 21122-1.

Information technology — JPEG XS low-latency lightweight image coding system —

Part 2: Profiles and buffer models

1 Scope

This document defines a number of subsets of the syntax specified in ISO/IEC 21122-1 as profiles. It also defines lower bounds on the throughput in the decoded domain via levels and the encoded domain via sublevels that a conforming decoder implementation shall support. Furthermore, it defines a buffer model to ensure interoperability between implementations in the presence of a latency constraint.

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 21122-1, *JPEG XS low-latency lightweight image coding system — Part 1: Core coding system*

3 Terms, definitions, symbols and abbreviated terms

3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 21122-1 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
blanking codestream fragment
placeholder *codestream fragment* representing blanking periods

3.1.2
buffer model
combination of a *decoder model* and a *channel model* whose behaviour can be defined by a set of parameters

3.1.3
buffer model instance
specific configuration of a *buffer model* specified by the assignment of well-defined values to the buffer model parameters

3.1.4
channel model
model describing the temporal behaviour of the *transmission channel* connecting an encoder and a decoder

3.1.5
coded codestream fragment
continuous sequence of bits in the codestream containing exactly one packet body and a well-defined number of packet headers, markers and marker segments

3.1.6
codestream fragment
either *coded codestream fragment*, or *blanking codestream fragment*

3.1.7
cycle
single clock period of an encoder or decoder clocked implementation

3.1.8
decoder model
combination of a *decoder unit* and a *decoder smoothing buffer*

3.1.9
decoder smoothing buffer
memory buffer that is used to level out changes in the number of bits read by a *decoder unit* per time unit

3.1.10
decoder unit
module reading a variable number of bits per time unit to generate decoded output *pixels* with a fixed rate

3.1.11
decomposition level
set of wavelet coefficients resulting from a particular *level* of recursive application of a wavelet transform

3.1.12
encoder model
combination of an *encoder unit* and an *encoder smoothing buffer*

3.1.13
encoder smoothing buffer
memory buffer that is used to level out changes in the number of bits generated by an *encoder unit* per time unit

3.1.14
encoder unit
module transforming a sequence of input *pixels* with constant rate into a conforming codestream, producing a bit sequence with variable number of bits generated per time unit

3.1.15
fill level
number of bits stored in the encoder or *decoder smoothing buffer*

3.1.16
horizontal blanking period
timespan expressed in units of the grid point sampling rate between the last *pixel* of an image line — not being the last line of an image — and the first pixel of the next image line

3.1.17
level
defined set of constraints on the amount of decoded *samples* to be processed by an encoder or decoder, both in the spatial and time dimensions

Note 1 to entry: The same set of levels is defined for all profiles. Individual implementations may, within the specified constraints, support a different level for each supported profile.

3.1.18
nominal bits per pixel value
mean number of bits allocated per encoded *pixel* which is used to derive the *sublevel* constraints by assuming an image with well-defined dimensions and frame rate derived from the *level*

3.1.19
profile
specified subset of the codestream syntax together with admissible parameter values

3.1.20
sampling grid point
position on the *sample grid*, specified by integer horizontal and vertical offset relative to the origin of the sample grid

3.1.21
smoothing buffer unit
level and *sublevel* dependent number of bits by which the smoothing buffer size of the *decoder model* is specified

3.1.22
start of transmission
SoT
time at which the *transmission channel* starts transmission relative to the start of encoding of the first *codestream fragment* of a codestream

3.1.23
sublevel
defined set of constraints on the amount of codestream bits to be processed by an encoder or decoder, per unit of time, per column, and per image

Note 1 to entry: The same set of sublevels is defined for all profiles. Individual implementations may, within the specified constraints, support a different sublevel for each supported profile.

3.1.24
transmission channel
facility transferring bits from a source entity to a target entity

3.1.25
transmission channel capacity
maximum number of bits per time unit that a *transmission channel* can transfer from a source entity to a target entity

3.1.26
vertical blanking period
timespan in units of the grid point sampling rate between the last line of an image — including the *horizontal blanking periods* — and the first line of the next image

3.2 Abbreviated terms

bpp	bits per pixel
CFA	colour filter array
DWT	discrete wavelet transform
IDWT	inverse discrete wavelet transform
RCT	reversible colour transform
IRCT	inverse reversible colour transform

3.3 Symbols

$C(i)$	codestream i
D_{c2d}	number of clock cycles between the first bit written into the decoding smoothing buffer and the decoding start of the first codestream fragment of a stream of codestream fragments
$F_{\mathrm{first}}(C(i))$	first codestream fragment of codestream $C(i)$
$F_{\mathrm{last}}(C(i))$	last codestream fragment of codestream $C(i)$
H_{f}	height of the image in sampling grid points
$H_{\max}$	maximum picture height in sampling grid points
$L_{\max}$	maximum number of sampling grid points per image
$l_{\mathrm{enc}}(t)$	fill level of the encoding smoothing buffer in bits at the end of cycle t
$l_{\mathrm{dec}}(t)$	fill level of the decoding smoothing buffer in bits at the end of cycle t
$l_{\mathrm{enc,max}}$	capacity in bits of the encoding smoothing buffer
$l_{\mathrm{dec,max}}$	capacity in bits of the decoding smoothing buffer
$\tilde{l}_{\mathrm{dec}}(t)$	number of bits that can be read from the decoding smoothing buffer in cycle t
$l_{\mathrm{sum}}(t)$	sum of encoder and decoder smoothing buffer fill level in bits at cycle t
$\mathbb{N}$	all integer numbers being strictly larger than zero
$\mathbb{N}_0$	all integer numbers being greater than or equal to zero
$N_{\mathrm{b,x}}$	size of the horizontal blanking line in sampling grid point clock periods
$N_{\mathrm{b,y}}$	size of the vertical blanking period in sampling grid lines
N_{bpp}	nominal number of bits allocated per pixel for compression
N_{c}	number of components in an image
$N_{\mathrm{cg}}(f)$	number of coefficient groups within codestream fragment f
$N_{\mathrm{cg,hz}}$	number of coefficient groups associated to a codestream fragment representing a horizontal blanking period

$N_{cg,vt}$ number of coefficient groups associated to a codestream fragment representing a vertical blanking period

$N_f(i)$ number of codestream fragments within a codestream i

N_g number of coefficients in a code group

$N_{L,x}$ number of horizontal decomposition levels

$N_{L,y}$ number of vertical decomposition levels

$N_{p,x}$ number of precincts per sampling grid line

$N_{p,y}$ number of precincts per sampling grid column

N_{sbu} number of decoder smoothing buffer units for a given profile

$\mathbb{Q}$ set of rational numbers

$r_{dec}(t)$ number of bits read and removed from the decoder smoothing buffer in clock cycle t

R_{trans} transmission channel capacity, expressed in bits per cycle (having a duration of T)

$R_{t,max}(l_m, l_s)$ maximum admissible encoded throughput in bits per second for a given level

$R_{s,max}$ maximum grid point sample rate (in samples per second) at decoder output

$S_{bits}(f)$ number of bits forming the codestream fragment f

$S_{c,max}$ targeted maximum number of bytes of an encoded codestream

$S_{sbu}(l_m, l_s)$ size of the smoothing buffer unit in bytes for level l_m and sublevel l_s

$S_{sbo}(p)$ smoothing buffer offset in bits for a profile p

$S_{sl,max}(l_m, l_s)$ maximum size of an encoded codestream in bytes of level l_m and sublevel l_s

$s_x[i]$ sampling factor of component i in horizontal direction

$s_y[i]$ sampling factor of component i in vertical direction

T_{enc} clock period defining the frequency by which code groups are processed by an encoder

T_{dec} clock period defining the frequency by which code groups are processed by a decoder

$t_{enc,write}(f)$ timestamp in cycles at which the codestream fragment f is written to the encoder smoothing buffer

$t_{dec,start}(f)$ timestamp in cycles at which decoder starts decoding codestream fragment f

$t_{dec,read}(f)$ timestamp in cycles at which codestream fragment f is removed from the decoder smoothing buffer

Tbmd buffer model type

$W_c[i]$ width of component i in samples

$W_{c,max}$ maximum column width in sampling grid points for a given profile

$w_{dec}(t)$ number of bits written into the decoder smoothing buffer in clock cycle t

W_f width of the image in sampling grid points

W_{max} maximum picture width in sampling grid points

$\mathbb{Z}$ set of all integer numbers

4 Conventions

4.1 Conformance language

The keyword "reserved" indicates a provision that is not specified at this time, shall not be used, and may be specified in the future. The keyword "forbidden" indicates "reserved" and in addition indicates that the provision will never be specified in the future.

4.2 Operators

NOTE Many of the operators used in document are similar to those used in the C programming language.

4.2.1 Arithmetic operators

+	addition
−	subtraction (as a binary operator) or negation (as a unary prefix operator)
×	multiplication
/	division without truncation or rounding

4.2.2 Logical operators

||	logical OR
&&	logical AND
!	logical NOT

4.2.3 Relational operators

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

4.2.4 Precedence order of operators

Operators are listed in descending order of precedence. If several operators appear in the same line, they have equal precedence. When several operators of equal precedence appear at the same level in an expression, evaluation proceeds according to the associativity of the operator either from right to left or from left to right.

Operators	Type of operation	Associativity
()	expression	left to right
[]	indexing of arrays	left to right
–	unary negation	
×, /	multiplication, division	left to right
+, –	addition and subtraction	left to right
< , >, ≤, ≥	relational	left to right
&	bitwise AND	left to right
|	bitwise OR	left to right

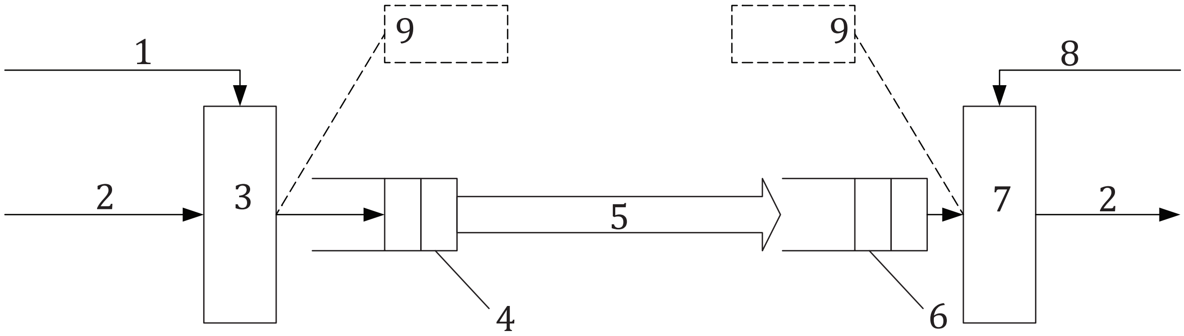
4.2.5 Mathematical functions

$\lceil x \rceil$	ceil of x: returns the smallest integer that is greater than or equal to x
$\lfloor x \rfloor$	floor of x: returns the largest integer that is less than or equal to x
$\|x\|$	absolute value of x, $\|x\|$ equals $-x$ for $x < 0$, otherwise x
sign(x)	sign of x, 0 if x is 0, +1 if x is positive, −1 if x is negative
$\xi(t)$	step function $\xi(t)=\begin{cases}1 & t \geq 0 \\ 0 & \text{otherwise}\end{cases}$
$\max_i(x_i)$	maximum of a sequence of numbers $[x_i]$ enumerated by the index i

5 Buffer model

5.1 General system block diagram

The JPEG XS coding system addresses applications where coded images are transferred from a source to a target, as shown in Figure 1. To this end, the encoder is compressing a continuous stream of input pixels into a sequence of bits. These bits are forwarded by means of a transmission channel to the decoder that decompresses the bits to produce a continuous stream of output pixels.



Key

1 encoder clock
2 pixel data
3 encoder unit
4 encoder smoothing buffer
5 transmission channel
6 decoder smoothing buffer
7 decoder unit
8 decoder clock
9 variable bit rate

Figure 1 — General system block diagram

The time instances at which the encoder processes each pixel are determined by an encoding clock. Similarly, the time instances at which the decoder produces each output pixel are determined by a decoding clock. Both clocks are generated by the system.

NOTE In implementations, these clocks can be the same or differ in both frequency and phase. The presented model is independent of whether clocks are synchronized or not.

In accordance with ISO/IEC 21122-1, the pixels of an image are translated into coefficient groups represented as code groups in the codestream. The number of bits necessary to code these code groups may vary from group to group. As a consequence, the encoder writes encoded bits at a variable rate into the encoder smoothing buffer. Similarly, the decoder reads the codestream at a variable rate from the decoder smoothing buffer.

In case the maximum bit rate of the transmission channel is below the peak bit rate generated by the encoder, an encoder smoothing buffer is necessary to decouple generation of bits by the encoder from transmission of bits over the transmission channel. Similarly, a decoder smoothing buffer needs to be provided that decouples the arrival of bits at the rate afforded by the transmission channel and the consumption of bits by the decoder per clock.

Correct operation requires that the decoder buffer never overflows. This is because the decoder is not able to pause the arrival of bits from the transmission channel. Moreover, a buffer underflow in the decoder buffer needs to be avoided. This is because the decoder is required to output pixels in accordance with the timing of its output interface. Hence it needs to be ensured that the bits to be read from the decoding buffer to produce the next pixel in accordance with the decoding clock are available in this decoding buffer.

5.2 Influencing variables on the required buffer sizes

Avoiding any buffer overflow or underflow, as discussed in subclause 5.1, requires sizing the decoder smoothing buffer properly. Moreover, the time at which decoding starts is delayed relative to the starting time of encoding and the start of transmission needs to be carefully set. Those values are influenced by many system parameters, for example:

— The maximum transmission channel bit rate.

— The granularity at which the encoder writes the encoded data and the decoder reads the encoded data.

— The rate control strategy applied by the encoder.

These dependencies cause that encoders and decoders are only interoperable in well-defined conditions. Defining these conditions is the purpose of the buffer model defined in Annex B and Annex C.

5.3 Role of the buffer model

The core coding system defined in ISO/IEC 21122-1 can be implemented on a large variety of platforms using many different implementation strategies. Thus, interoperability cannot be achieved by precisely specifying the temporal behaviour of a conforming decoding implementation. Instead, the buffer model defines a simplified decoder model. Interoperability is then achieved by mandating that a conforming decoder shall decode all bit streams being decodable by the simplified decoder model. Similarly, a conforming encoder shall not create bit streams that cannot be decoded by the simplified decoder model.

To this end, Annex B defines a generic JPEG XS decoder model that precisely defines the temporal behaviour of the decoder model assuming a processing granularity of codestream packets. While such a model already defines some fundamental properties of the decodable codestreams, it is still not sufficient to ensure interoperability. The reason is that otherwise codestreams could be constructed that would only be decodable by the decoder model if the transmission channel could transport bits arbitrarily fast. In practice, this is obviously not the case. Consequently, interoperability also requires defining a channel model over which an encoder sends the codestreams to the decoder.

Annex C defines such a channel model assuming a transmission channel with a fixed upper bit rate that is related to the target compression ratio. Together with the decoder model of Annex B, it defines the packet-based constant bit rate buffer model. It describes the conditions for a low latency interoperability between any conforming encoder and any decoder. These conditions are expressed by buffer model parameters that are specified by the profiles and levels defined in Annex A. The properties of such conforming implementations are exemplified in Annex D. Since these properties are direct consequences of Annex B and Annex C, Annex D is informative only.

6 Interpretation of Bayer data

ISO/IEC 21122-1 defines coding tools and signalling for compression of Bayer-type CFA image data. According to this specification, each sampling grid point represents a super-pixel of four sensor elements containing at least one sample of each component. Thus Bayer data is interpreted as an image having four components, where each sampling grid point describes four spatially disjoint sensor elements (one element per Bayer channel).

