INTERNATIONAL STANDARD

ISO/IEC 15444-3

Second edition
2007-05-01

Information technology — JPEG 2000 image coding system: Motion JPEG 2000

Technologies de l'information — Système de codage d'image JPEG 2000: Motion JPEG 2000

Reference number
ISO/IEC 15444-3:2007(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

CONTENTS

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may involve the use of a patent, as indicated in Annex H.

ISO/IEC 15444-3 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information,* in collaboration with ITU-T. The identical text is published as ITU-T Rec. T.802.

This second edition cancels and replaces the first edition (ISO/IEC 15444-3:2002), which has been technically revised. It also incorporates the amendment ISO/IEC 15444-3:2002/Amd.2:2003.

ISO/IEC 15444 consists of the following parts, under the general title *Information technology — JPEG 2000 image coding system*:

— *Part 1: Core coding system*

— *Part 2: Extensions*

— *Part 3: Motion JPEG 2000*

— *Part 4: Conformance testing*

— *Part 5: Reference software*

— *Part 6: Compound image file format*

— *Part 8: Secure JPEG 2000*

— *Part 9: Interactivity tools, APIs and protocols*

— *Part 10: Extensions for three-dimensional data*

— *Part 11: Wireless*

— *Part 12: ISO base media file format*

The following part is under preparation:

— *Part 13: An entry level JPEG 2000 encoder*

Introduction

This Recommendation | International Standard is the consequent revision of the Motion JPEG 2000 specification, based on the common text of the MP4 and MJ2 formats, which is called the ISO Base Media File Format.

This Recommendation | International Standard specifies the use of the wavelet-based JPEG 2000 codec for the coding and display of timed sequences of images. It has been defined by ISO/IEC JTC 1/SC 29/WG 1 as Part 3 of the JPEG 2000 International Standard. In this Recommendation | International Standard, a file format is defined, and guidelines for the use of the JPEG 2000 codec for timed sequences are supplied. The Motion JPEG 2000 file format MJ2 is designed to contain one or more motion sequences of JPEG 2000 images, with their timing, and also optional audio annotations, all composed into an overall presentation.

To promote interoperability between MJ2 encoders and decoders and to test these systems for compliance to this Recommendation | International Standard, a framework of compliance testing is provided. Compliance testing is the testing of a candidate product for the existence of specific characteristics required by a standard. It involves testing the capabilities of an implementation against both the compliance requirements in the relevant standard and the statement of the implementation's capability.

Motion JPEG 2000 is expected to be used in a variety of applications, particularly where the codec is already available for other reasons, or where the high-quality frame-based approach, with no inter-frame coding, is appropriate. These application areas include:

- digital still cameras;
- error-prone environments such as wireless and the Internet;
- PC-based video capturing;
- high-quality digital video recording for professional broadcasting and motion picture production from film-based to digital systems; and
- high-resolution medical and satellite imaging.

Motion JPEG 2000 is a flexible format, permitting a wide variety of usages, such as editing, display, interchange, and streaming.

INTERNATIONAL STANDARD
ITU-T RECOMMENDATION

Information technology – JPEG 2000 image coding system: Motion JPEG 2000

1 Scope

This Recommendation | International Standard specifies the use of the wavelet-based JPEG 2000 codec for the coding and display of timed sequences of images (motion sequences), possibly combined with audio, and composed into an overall presentation. In this Recommendation | International Standard, a file format is defined, and guidelines for the use of the JPEG 2000 codec for motion sequences are supplied. This Recommendation | International Standard also specifies profiles and the framework, concepts, methodology for testing and the criteria to be achieved to claim compliance to this Recommendation | International Standard.

2 Normative references

The following Recommendations and International Standards contain provisions which, through reference in this text, constitute provisions of this Recommendation | International Standard. At the time of publication, the editions indicated were valid. All Recommendations and Standards are subject to revision, and parties to agreements based on this Recommendation | International Standard are encouraged to investigate the possibility of applying the most recent edition of the Recommendations and Standards listed below. Members of IEC and ISO maintain registers of currently valid International Standards. The Telecommunication Standardization Bureau of the ITU maintains a list of currently valid ITU-T Recommendations.

– ITU-T Recommendation T.800 (2002) | ISO/IEC 15444-1:2004, *Information technology – JPEG 2000 image coding system: Core coding system.*

– ITU-T Recommendation T.803 (2002) | ISO/IEC 15444-4:2004, *Information technology – JPEG 2000 image coding system: Conformance testing.*

– ISO 639-2:1998, *Codes for the representation of names of languages – Part 2: Alpha-3 code.*

– ISO/IEC 14496-1:2004, *Information technology – Coding of audio-visual objects – Part 1: Systems* [particularly the syntax description language (SDL), clause 14].

– ISO/IEC 15444-12:2005, *Information technology – JPEG 2000 image coding system – Part 12: ISO base media file format* (technically identical to ISO/IEC 14496-12).

3 Definitions

For the purposes of this Recommendation | International Standard, the following definitions apply.

3.1 Motion sequence: A timed sequence of JPEG 2000 images.

4 Compatibility and technology derivation

4.1 Family members

This is a stand-alone Recommendation | International Standard; it defines the file format for MJ2. However, it stands as a member of a family of Recommendations | International Standards with common formatting.

The other family members include:

- the JPEG 2000 single image format, JP2;
- the ISO Base Media File Format, on which the MP4 format is based;
- the QuickTime file format, on which the ISO Base Media format is based.

These specifications share a common definition for the structure of a file (a sequence of objects, called boxes here and atoms in QuickTime), and a common definition of the general structure of an object (the size and type).

All these specifications require that readers ignore objects that are unrecognizable to them.

This Recommendation | International Standard takes precedence over those from which it inherits, in any case where there are differences or conflicts; however, no such conflicts are known to exist.

4.2 ISO Media file inheritance and compatibility

The Motion JPEG 2000 file format is defined as derived from the ISO Base Media file format. Notwithstanding anything in that base specification, hint tracks are not a normative part of this Recommendation | International Standard. Clauses 7 (Streaming Support) and 10 (RTP Hint Track Format), though compatible with this Recommendation | International Standard, do not form normative parts of this Recommendation | International Standard. They may be used as a compatible, optional, extension, but are not required for compatibility with this Recommendation | International Standard. There may be license implications in the use of this or other compatible extensions to this format.

4.3 JP2 inheritance and compatibility

The still image format, JP2, defines a number of boxes. The signature box from that specification shall be present. If the JP2 specification requires a particular position (e.g., first in the file), that positioning shall be followed here.

It is permissible under this Recommendation | International Standard to make a file that adheres to both this Recommendation | International Standard and the JP2 specification. In that case:

1) The compatibility list shall include all the compatible brands.
2) The objects (boxes or atoms) required by the JP2 specification shall also be present.
3) The objects (boxes or atoms) optional in the JP2 specification may also be present.

A still image reader, reading a file which contains both a presentation (conformant to this Recommendation | International Standard) and a still image, would 'see' only the still image. Likewise, a motion reader would 'see' only the presentation. A more powerful reader may display both, or offer the user a choice.

The JP2 specification includes an optional IPR (Intellectual Property Rights) box which is therefore also optional in this Recommendation | International Standard. Among other issues this addresses unique identification and protection of content.

4.4 Conformance

Implementations of Motion JPEG 2000 decoders shall support JPEG 2000 image sequences, as well as raw and twos-complement audio if audio output is available. They may also support compressed audio, using MP4 formats, or other track types from MPEG-4. The support of such MPEG-4 tracks is not required; however, readers shall not fail if they are present. If MPEG-4 composition (BIFS) is used, then the simple composition used in this Recommendation | International Standard should also be set up in such a way that a reader not implementing BIFS will display a suitable result.

Files conformant with this Recommendation | International Standard shall contain at least one Motion JPEG 2000 video track. They may contain more video tracks, uncompressed audio, or compressed MP4 audio.

4.5 Profiles and levels

There are two tools for profiling Motion JPEG 2000 files.

The first consists of the optional specification of tools and levels of the JPEG 2000 coding system (codestream features). These are indicated in the optional sample description extension JP2 Profile Box (see clause 5).

The second tool allows a file overall to be identified as belonging to a definition which forms a proper subset of the general specification. Such definitions might restrict such features as:

- the use of data references, and multiple files;
- the layout order of the boxes, and the data within the boxes (e.g., that data is in time order and interleaved);

- the use of profiles of the JPEG 2000 codestream;
- the existence of other tracks, and their format (e.g., audio, MPEG-7, etc.).

The conformance to these restricted profiles is indicated in the file type box by the addition of the compatible profiles as brands within the compatibility list. Annex A defines the available profiles in this Recommendation | International Standard.

4.6 Visual composition

Composition of multiple image sequences in a 2D environment can be achieved by using multiple video tracks which overlap in time. Their composition is defined by the following structures:

- The matrix in the track header specifies their positioning and scaling.
- The layer field in the track header specifies the front-to-back ordering of the tracks.
- The graphics mode and opcolor fields in the video media header are used to specify the ways in which each track is composited onto the existing image (this compositing is performed from back to front).

Applications requiring more complex compositing may use the BIFS system from MPEG-4, optionally. The matrix, graphics mode, and layers should be set up so that a reader not implementing BIFS displays the desired result. Matrix values which occur in the headers specify a transformation of video images for presentation. The point (p,q) is transformed into (p',q') using the matrix as follows:

```
(p q 1) *   | a   b     u       |       = (m n z)
            | c   d     v       |
            | x   y     w       |

m = ap + cq + x; n = bp + dq + y; z = up + vq + w;

p' = m/z; q' = n/z
```

The coordinates {p,q} are on the decompressed frame, and {p',q'} are at the rendering output. Therefore, for example, the matrix {2,0,0, 0,2,0, 0,0,1} exactly doubles the pixel dimension of an image. The coordinates transformed by the matrix are not normalized in any way, and represent actual sample locations. Therefore {x,y} can, for example, be considered a translation vector for the image.

The coordinate origin is located at the upper left corner, and X values increase to the right, and Y values increase downwards. {p,q} and {p',q'} are to be taken as absolute pixel locations relative to the upper left hand corner of the original image (after scaling to the size determined by the track header's width and height) and the transformed (rendering) surface, respectively.

Each track is composed using its matrix as specified into an overall image; this is then transformed and composed according to the matrix at the movie level in the MovieHeaderBox. It is application-dependent whether the resulting image is 'clipped' to eliminate pixels, which have no display, to a vertical rectangular region within a window, for example. So for example, if only one video track is displayed and it has a translation to {20,30}, and a unity matrix is in the MovieHeaderBox, an application may choose not to display the empty "L" shaped region between the image and the origin.

All the values in a matrix are stored as 16.16 fixed-point values, except for u, v and w, which are stored as 2.30 fixed-point values. For upwards compatibility into the MPEG-4 BIFS (scene composition) system, matrices used here restrict (u,v,w) to be (0,0,1), for which the hex values are (0,0,0x40000000). This permits the simple composition used here to be mapped into BIFS if a scene later requires full scene management.

The values in the matrix are stored in the order {a,b,u, c,d,v, x,y,w}.

Tracks are composed to the presentation surface from back (highest layer number) to front (lowest layer number), against an indeterminate initial colour. There are various composition modes available; the backmost (first-rendered) track would normally use 'copy' as the initial image is indeterminate. Subsequent layers can then be composed on top in a variety of ways. Table 1 details the composition modes available. Note that (currently) only the 'transparent' mode uses the opcolor field.

Table 1 – Graphics composition modes

Mode	Code	Description
Copy	0x0	Copy the source image over the destination.
Transparent	0x24	Replace the destination pixel with the source pixel if the source pixel is not equal to the opcolor. (Also known as 'blue-screen').
Alpha	0x100	Replace the destination pixel with a blend of the source and destination pixels, with the proportion controlled by the alpha channel. The alpha channel is applied to all channels.
Pre-multiplied black alpha	0x102	Pre-multiplied with black means that the colour components of each pixel have already been blended with a black pixel, based on their alpha channel value. Effectively, this means that the image has already been combined with a black background, which must be removed before composition.
Component alpha	0x110	One or more alpha channels are present, which are applied to individual colour channels, and the image must be composed channel-by-channel.

Images are only alpha-composed if both the graphics composition mode requests alpha composition, and the images contain alpha channels, as declared by the Channel Definition Box inside the JP2 Header Box. Therefore the graphics mode can be used to prevent alpha composition of an image with alpha channels, if that is desired.

If there is a single alpha channel applied to the entire image, then the value of the graphics must be 'Alpha' if that channel is a straight 'Opacity' channel, and must be 'Pre-multiplied black alpha' if that channel is a 'Pre-multiplied' opacity channel. If there are one or more alpha channels in the image which are applied to individual channels and not to the whole image, and alpha composition is desired, then the 'Component alpha' value must be used for the graphics mode. Support of 'Component alpha' composition is optional in this Recommendation | International Standard.

The alpha blending formulas are defined in ITU-T Rec. T.800 | ISO/IEC 15444-1.

NOTE – Use of the "transparent" opcode may be yield unexpected results when the image codestreams are compressed in a non-reversible fashion, or are subject to scaling in quality or resolution, either during or after content production. Such operations are not guaranteed to preserve individual sample values precisely.

4.7 Box order

All JPEG 2000 files start with a signature box, therefore this rule is added to the Box Order rules:

1) The JP2 Signature Box and File Type Box **shall** occur first and second in the file.

5 File identification

In the file type compatibility box, the brand shall be 'mjp2' for files conforming to this Recommendation | International Standard, and 'mjp2' shall be a member of the compatibility list.

See Annex A for a complete list of all profile brand names.

The preferred file extension is '.mj2'. The MIME type video/mjp2 is used, as defined in the appropriate RFC.

6 Required additions

6.1 Sample description box

6.1.1 Definition

Box Types: 'mjp2', 'raw '?, 'twos'
Container: Sample Table Box ('stbl')
Mandatory: Yes
Quantity: Exactly one

A Motion JPEG 2000 visual sample entry shall contain a JP2 Header Box from ITU-T Rec. T.800 | ISO/IEC 15444-1; however, the sRGB YCC enumerated colour space (codepoint 18) from ITU-T Rec. T.801 | ISO/IEC 15444-2 may also be used to identify the colour space used, in addition to ITU-T Rec. T.800 | ISO/IEC 15444-1 numerated colour spaces (such as sRGB and greyscale). If the JP2 Header Box indicates the presence of alpha channels, then the 'depth' field in the VisualSampleEntry must also indicate their presence, with the value 0x20. Similarly, if the JP2 Header Box defines a monochrome image with no alpha, the 'depth' field must contain the value indicating grayscale (0x28). Otherwise, the 'depth' field must declare colour images (0x18).

If two fields are present in the samples, the JP2 Header Box applies to the complete image, not to each field individually. Therefore the height as declared in the JP2 Header Box and the VisualSampleEntry applies to the entire de-interlaced image.

NOTE – This means that you may not be able to construct a legitimate JP2 file by composing a single field's codestream together with the JP2 header box found in the sample description.

The sample format for Motion JPEG 2000 data is a set of boxes. Currently this Recommendation | International Standard permits only JP2 Codestream Boxes ('jp2c') as defined in the JP2 specification. If there is no Field Coding Box present, or the field count is 1, the sample shall contain precisely one codestream box. If the field count is 2, then there shall be two codestream boxes. Other boxes, if present in the sample, shall be ignored. The last (or only) box in the sample may have a value of 0 for its length field, indicating it extends to the end of the sample, as indicated by the sample size given in the sample size table. The actual codestreams presented to the decoder are formed by concatenating the contents of the JP2 Prefix Box, if any, in the sample description before each codestream presented in the jp2c box(es) in the samples. If field coding is used, the same prefix is concatenated before both fields. Typically, the prefix will contain a JPEG 2000 main header; however, this is not required in the general case, though specific profiles may limit the use of the prefix box.

If the codestreams used in a sequence conform to a specific profile of the JPEG 2000 coder, a JP2 Profile Box may be used to indicate such conformance.

The visual sample entry may optionally contain a field-ordering box (see below). If `fieldcount` is 2, each field will be half the height of the overall image, as declared in the 'height' field of the sample description. To be precise, if the height field contains the value H, then the field with the topmost scanline has ((H+1) div 2) lines, and the other field has (H div 2) lines. The utility of the prefix box may be diminished when field coding is used, especially when H is odd.

The original format of the material (interlace or progressive) may be documented by the Original Format Box. The `original_fieldcount` must take the value 1 (progressive) or 2 (interlaced), When the value is 2, the `original_fieldorder` documents whether the topmost line came from the earlier or later field. The documentation of the original format is independent of whether the material has been coded as progressive (frame-based) or interlaced, which is documented by the field coding box. This box is informative only and may assist readers in display or transcoding. For example, material that was originally interlaced but is encoded frame-based and must be displayed on an interlaced display can be positioned so that the display interlace matches the original material's interlace.

The values present in the VisualSampleEntry, its constituent boxes including the JP2 Header Box, and the codestreams that these boxes describe, must agree, to the extent that the format and precision of fields allow. This agreement includes, but is not limited to, width and height information, and the resolution declaration (within the accuracy permitted by the different representations). Files with conflicts are non-conforming and readers may attempt to decide which values are correct, or reject the file.

The fields `horizresolution` and `vertresolution` document the highest resolution component of the image (which is typically, but not required to be, the luminance, in a sub-sampled image).

For audio tracks, in the formats defined here (with code points `'raw '` and `'twos'`), the data is stored as uncompressed samples. If stereo is stored, the data consists of interleaved left/right samples. The `raw` format uses offset-binary; for 8-bit samples, values range from 0 to 255, with 128 indicating silence. For `'twos'`, 8-bit values range from –128 to 127, with 0 being silence. Values of 16 bits are similarly derived (with the bytes in network byte order, or big-endian format).

6.1.2 Syntax

```
      // Visual Sequences

class MJ2SampleEntry() extends VisualSampleEntry ('mjp2'){
      JP2HeaderBox();
      FieldCodingBox(); // optional
      MJP2ProfileBox(); // optional
      MJP2PrefixBox();  // optional
      MJP2SubSamplingBox();         // optional
      MJP2OriginalFormatBox();      // optional
}

      // Field-Based Coding
```

```
class FieldCodingBox() extends Box('fiel'){
      int(8)      fieldcount;
      int(8)      fieldorder; // both storage and temporal order
}

class MJP2OriginalFormatBox() extends Box('orfo'){
      int(8)      original_fieldcount;
      int(8)      original_fieldorder;
}

class MJP2ProfileBox() extends FullBox('jp2p', 0, 0){
      unsigned int(32)[] compatible_brands;
}


class MJP2PrefixBox() extends Box('jp2x'){
      int(8)[]    data; // the data is the initial codestream part
}

class MJP2SubSamplingBox () extends Box('jsub'){
      unsigned    int(8) horizontal_sub;
      unsigned    int(8) vertical_sub;
      unsigned    int(8) horizontal_offset;
      unsigned    int(8) vertical_offset;
}

      // Audio Sequences

class MJ2AudioSampleEntry() extends AudioSampleEntry (AudioFormat){
}
```

6.1.3 Semantics

`AudioFormat` is either 'raw ' or 'twos'.

`Compressorname` the value "\017Motion JPEG 2000" is recommended (\017 is 15, the length of the string as a byte)

`depth` takes one of the following values:

0x18 – images are in colour with no alpha;

0x28 – images are in grayscale with no alpha;

0x20 – images have alpha (gray or colour).

`compatible_brands` is a list, filled to the end of the containing box, of JPEG 2000 profiles, to which the associated codestreams conform; see the JPEG 2000 specification for the defined values.

`Horizontal_sub` and `vertical_sub` indicate whether the chroma components of a YCbCr encoding were downsampled in the codestream; the value indicates the number of luminance samples to a single chroma sample in the given direction. This can assist decoders in memory allocation, or in using optimized sub-sampled display interfaces.

`Horizontal_offset` and `vertical_offset` specify the offset of the first chroma sample from the first luminance sample, as measured on the sample grid. If a CRG marker is present in the codestream, these values take precedence over those in the codestream. See Annex C for example values.

`fieldcount` specifies the number of fields in the samples, and shall be 1 or 2.

`fieldorder` describes the order of the two fields, and is only relevant if `fieldcount` equals 2:

0 Field coding unknown;

1 Field with the topmost line is stored first in the sample; fields are in temporal order;

6 Field with the topmost line is stored second in the sample; fields are in temporal order.

`original_fieldcount` specifies the number of fields in original material before encoding, and shall be 1 or 2.

`original_fieldorder` describes the order of the two fields, and is only relevant if `original_fieldcount` equals 2:

0 Field coding unknown;

11 Topmost line came from the earlier field;

16 Topmost line came from the later field.

7 Template fields used

Motion JPEG 2000 uses simple composition, as defined above. Therefore the following fields are required to be correctly set in conforming files:

1) For Visual composition: `matrix` in the Movie Header Box and Track Header Box; layer; graphicsmode and `opcolor` in the Video Media Header Box; and `depth` in a VisualSampleEntry;

2) For Audio Composition: `volume` in the Movie Header Box and Track Header Box; `balance` in the Sound Media Header Box;

3) For variable-rate playback, `rate` in the Movie Header Box;

4) To describe the audio and visual sample correctly, `horizresolution`, and `vertresolution` in a VisualSampleEntry; and `channelcount`, `samplesize` and the `samplerate` in an AudioSampleEntry.

8 Definition of compliance points

8.1 General

This clause describes a number of compliance points (Cpoints) for this Recommendation | International Standard. The points and parameters are described to provide assistance in designing a compliant decoder. Actual compliance is determined by the test methods in clause 10 and the codestreams, reference images, and tolerances in clause 11. The definitions of compliance points in this clause are useful for the design of an encoder. The parameters may correspond to particular parts of an implementation.

Because of resource limitations, implementations of Motion JPEG 2000 sometimes will not be able to decode a codestream in its entirety. This clause defines various parameters for which a specific implementation might be limited. A set of values for every parameter defines a compliance point. Thus, an implementation of a particular Cpoint must guarantee resources as defined in all the parameters.

8.2 H, W, C: Image size guarantees

Decoders may be limited in the size of the output image that they are capable of producing, due to physical display characteristics or memory limitations. H, W, and C are respectively the largest height, width, and number of components that are required to be decoded for a decoder in the compliance point. Codestreams containing more samples than the H, W, and C for a Cpoint shall still be decoded, provided they contain a resolution equal to or less than H × W. Compliance for these codestreams is based on the ability to decode at the largest size smaller than or equal to that specified by the decoder's Cpoint, while preserving aspect ratio. The requested image size is defined by the height and width fields in the applicable 'VisualSampleEntry' from the MJ2 file.

Equations 8-1 and 8-2 express these restrictions. The maximum $r \le T_L$ that satisfies both conditions is the number of levels that must be decoded. T_L is defined in 8.9. The variables w_r and h_r denote requested width and height from the VisualSampleEntry segment defined in 6.1. The term $\min_{\forall i}(N_L(i))$ denotes the minimum number of decomposition levels declared in any COD or COC marker segments whether used in main header or in tile-part headers as defined in Annex A of ITU-T Rec. T.800 | ISO/IEC 15444-1.

If a non-negative r does not exist to satisfy both conditions for any tile or for the whole image, then no decoder obligation exists. A decoder claiming compliance at some Cpoint with image dimensions $H \times W$ and number of components C, must also be capable of decoding any sequences with width less than or equal to W, height less than or equal to H, and number of components less than or equal to C. For each Cpoint, the minimum values for H, W, and C are specified in Table 2.

$$\left\lceil \frac{w_r}{2^{\min_{\forall i}(N_L(i))-r}} \right\rceil \le W \tag{8-1}$$

$$\left\lceil \frac{h_r}{2^{\min_{\forall i}(N_L(i))-r}} \right\rceil \le H \tag{8-2}$$

8.3 N_{cb}: Code-block parsing guarantee

Decoders need not decode compressed bits that cannot be recovered from the codestream due to excessive parser memory being required. An upper bound for the parser state memory required to reach a point x in the codestream may be determined from the total number of code-blocks for which state information must be kept, the total number of precincts for which a packet has been encountered, and the total number of components of the codestream.

At position x in the codestream, $N_{cb}(x)$ is defined as the total number of code-blocks in every precinct where the first header byte of at least one received packet for the precinct lies outside the range 0x80 to 0x8F.

Decoders are permitted to stop parsing the codestream at the point, x, once $N_{cb}(x) > N_{cb}$, where N_{cb} is defined for each compliance point. Decoders are permitted to stop parsing the codestream once packet headers with more than N_{cb} code-blocks have been encountered. Code-blocks in packets prior to the packet with the N_{cb}th code-block shall be decoded up to the limits of other parameters in the compliance point.

NOTE – Packet headers with the first bit set to 0 are defined as empty. The above definition adds all the code-blocks associated with such precincts to N_{cb} for these empty packets because a decoder requires more memory for these packets than for packets starting in the listed range.

8.4 N_{comp}: Component parsing guarantee

Decoders could be required to buffer information about each component for many thousands of components just to parse a codestream. To limit the required memory, decoders are permitted to stop parsing the codestream at a point, x, once the following condition is reached:

$$C_{max}(x) > N_{comp}$$

where $C_{max}(x)$ is defined as the largest component index for which a packet has been encountered up to point x regardless of the emptiness or the relevance of the packet.

Code-blocks in packets prior to the above stop condition shall be decoded up to the limits of other parameters in the compliance point.

8.5 L_{body}: Coded data buffering guarantee

The parser state memory described in 8.3 is required to parse packets regardless of whether their code-blocks are relevant to the dimensions and number of components for which compliance is being claimed. For those code-blocks that are relevant, the implementation is required to store the recovered packet bytes. These are the code bytes that are processed by the block decoder (Annexes C and D of ITU-T Rec. T.800 | ISO/IEC 15444-1).

After a given number of decoded codestream bytes, x, the quantity $L_{body}(x)$ is defined as the total number of packet bytes that have been encountered so far in packets whose precincts are relevant to the dimensions and components for which compliance is being claimed. Although some implementations may be able to decode some of these packet bytes incrementally, L_{body} represents an upper bound on the number of packet bytes that must be stored by the decoder prior to decoding. If the number of relevant packet bytes exceeds L_{body}, then the Implementation Under Test (IUT) is allowed to stop reading the codestream and to decode the code-blocks obtained up to the limits of other parameters in the compliance point.

8.6 M: Decoded bit-plane guarantee

The decoder shall decode all of the packet bytes recovered by the parser in accordance with the requirements described above. This obligation is limited to the most significant M bit-planes of each code-block. Specifically, the block decoder must correctly decode the first $3(M - P_B)–2$ coding passes, if available, of any relevant code-block, b, where P_b is the number of zero-valued most significant bit-planes signalled in the relevant packet header as described in Annex B of ITU-T Rec. T.800 | ISO/IEC 15444-1. The decoder is free to decode any number of additional coding passes for any code-block. Codestreams with large values for the number of guard bits will have a larger number of zero-valued most significant bit-planes, and thus a decoder of any given Cpoint will decode fewer useful bit-planes. Likewise, codestreams with large values for the shift in the RGN marker segment may have fewer bit-planes decoded.

8.7 P: 9-7I precision guarantee

Codestreams that make use of the irreversible 9-7 discrete wavelet transform will require dequantization, the 9-7 inverse discrete wavelet transform, and potentially the inverse irreversible component transform (ICT). The precision values for the wavelet transform are chosen to allow high quality imagery at various bit-depths, e.g., 8, 12, or 16 bits per sample. However, for Cpoint-0, the accuracy of the 9-7I filter required is set such that it is possible to be compliant by decoding and inverse quantizing and performing a 5-3I (irreversible 5-3) inverse wavelet transform. This allows lower cost decoders to be used for the lowest compliance point only. For higher compliance points, using the 5-3 filter in place of the 9-7 filter will not be sufficient to pass the compliance tests.

Using the 5-3 inverse wavelet transform to decode imagery compressed with the 9-7 wavelet introduces signal dependent noise. For example, errors are highest around edges in the imagery. Because induced errors are signal dependent, there is no "precision" specified for the implementation of the wavelet transform for Cpoint-0. Instead, the bounds on accuracy of the 9-7 transform have been set for each Cpoint-0 reference image to allow an implementation to use the 5-3I inverse wavelet filter. Using the 5-3I inverse wavelet transform instead of a 9-7I filter does not relieve a decoder of the requirement to perform inverse quantization.

For compliance points other than Cpoint-0, the precision guarantee in Table 2 refers to the implementation's minimum word size that will achieve the target MSE values for the test streams.

To facilitate end-to-end testing for compliance, dequantization may be performed using mid-point rounding. That is, the value of r in Equation G-6 of ITU-T Rec. T.800 | ISO/IEC 15444-1 can be $r = 1/2$. Implementations under test may provide the option of using different values for the reconstruction parameter, r; however, if the value $r = 1/2$ is supported and employed for compliance testing this will typically increase the ease of passing.

8.8 B: 5-3R precision guarantee

A decoder is expected to implement the reversible 5-3R IDWT exactly, for component bit-depths of B bits/sample or less, as specified in the SIZ marker segment (see Annex A of ITU-T Rec. T.800 | ISO/IEC 15444-1). If a codestream employs the reversible component transform (RCT) and the IUT claims compliance at 3 or more components, it must be able to perform both the 5-3R IDWT and the inverse RCT exactly for bit-depths of B bits/sample or less.

8.9 T_L: Transform level guarantee

For each Cpoint, a decoder is expected to be able to synthesize a minimum number of levels of the IDWT, T_L. For codestreams that contains more than T_L decomposition levels, the decoded image from a compliant decoder in a given Cpoint may include only the top resolution levels.

8.10 L: Layer guarantee

For each Cpoint, a decoder is expected to decode a minimum number of layers, L, in a codestream. For codestreams that contain more than L layers, the decoded image from a compliant decode in a given Cpoint may include only the top L layers. This relieves compliant decoders from the burden of decoding inefficient codestreams with an excessive number of layers.

8.11 Progressions

For all Cpoints, a decoder is expected to decode all possible progressions as specified in the COD marker segment. If a POC marker segment is used in a codestream, Cpoint-0 to Cpoint-3 decoders shall decode packets associated with the first progression order specified in the POC marker segment for that tile. Additional packets in the tile may be skipped.

8.12 Tiles

If an image is divided into tiles, the following restrictions apply to tile dimensions:

$$128 \le XTsiz / \min(XRsiz^i, YRsiz^i) \le 1024 \tag{8-3}$$

$$XTsiz = YTsiz \tag{8-4}$$

8.13 Tile-parts

Codestreams may contain multiple tile-parts for each tile. Profile-0 codestreams require all initial tile-parts to appear in spatial order in the codestream before other tile parts. Cpoint-0 to Cpoint-3 decoders may ignore tile-parts beyond the first even if N_{cb} or L_{body} has not been reached.

8.14 Precincts

Tiles may contain several precincts. Cpoint-0 decoders need only decode the first precinct in each subband of each tile.

8.15 Frame-rate and bit-rate

Frame-rate: A compliant real-time decoder must report the lowest frame rate that it can sustain when decoding all frames, as well as the number of skipped frames (fields) when it achieves real-time.

Bit-rate: A compliant real-time decoder must also report the highest bit-rate which can always be fully decoded in real-time (bit-rate guarantee).

8.16 Profile: Codestream guarantee

Profiles provide limits on the codestream syntax parameters. Two profiles are defined in this Recommendation | International Standard in Annex A, labelled 'unrestricted' and 'simple'. Conformance testing of the rich feature set of unrestricted codestreams is not targeted in this Recommendation | International Standard. Thus, for all Cpoints compliant decoders need only to handle motion representations in MJ2 simple profile, indicated by the brand 'mj2s' in top-level file-type box.

NOTE – Conforming to simple profile means that restrictions of Profile-0 defined in A.10 of ITU-T Rec. T.800 | ISO/IEC 15444-1 apply to embedded codestreams, with the exception of tile dimensions where *YTsiz* = *XTsiz* = 128 is replaced by definitions of 8.12.

9 Compliance point definitions

Table 2 defines four compliance points in terms of the parameters.

Table 2 – Definitions of compliance points (Cpoint) for Part-3

Parameter	Cpoint-0	Cpoint-1	Cpoint-2	Cpoint-3
W × H(Size)	360 × 288	720 × 576	1920 × 1080	4096 × 3112
C(Components)	3	3	4	4
N_{cb}	399	1371	8428	50656
N_{comp}	4	4	4	4
L_{body}	2^{17} bytes	2^{20} bytes	2^{23} bytes	2^{26} bytes
M	11	13	15	19
P	Low enough to allow 5 × 3 I decoding of 9 × 7 I data	16-bit fixed point implementation	16-bit fixed point implementation	20-bit fixed point implementation
B	8	10	12	16
T_L	3	4	5	5
L	15	15	15	15
Progressions	For all Cpoints, a decoder is expected to decode all possible progressions as specified in the COD marker segment. If a POC marker segment is used in a codestream, a Cpoint-0 decoder shall decode packets associated with the first progression order specified in the POC marker segment for that tile. Additional packets in the tile may be skipped. For all other Cpoints, packets may be skipped only due to other limitations (e.g., N_{cb} and L_{body}) and there is no explicit limitation on the number of progression order changes that may occur.			
Tiles	Single tile image or square tiles with dimensions ranging from 128 to 1024	Single tile image or square tiles with dimensions ranging from 128 to 1024	Single tile image or square tiles with dimensions ranging from 128 to 1024	Single tile image or square tiles with dimensions ranging from 128 to 1024
Tile-parts	Decode only first tile-part per tile	Decode only first tile-part per tile	Decode only first tile-part per tile	Decode only first tile-part per tile
Precincts	Decode first precinct per subband	Decode all precincts	Decode all precincts	Decode all precincts
File format	MJ2 simple profile	MJ2 simple profile	MJ2 simple profile	MJ2 simple profile

10 Definition of test methods

Compliance testing procedures apply as defined in Annex B of ITU-T Rec. T.803 | ISO/IEC 15444-4, with the following extensions:

1) A particular executable test suite (ETS) defines the test codestreams (TCS), output images and error tolerances. This is done in clause 11 for the four defined compliance points by taking specified frames from MJ2 sequences. Implementations under test (IUT) must therefore be able to output decoded visual samples in a format (see Annex G) that they can be compared as defined in ITU-T Rec. T.803 | ISO/IEC 15444-4. Any visual composition transformations do not apply to this test method.

2) In addition to the testing of single visual samples IUT will be evaluated by using MJ2 reference sequences defined in clause 11. The test procedure is:

   - to decode all frames (fields) of a sequence as defined in 8.15. Therefore it must report, that no frames (fields) were skipped;
   - to decode in correct frame order and correct field order. Therefore it must report, that the correct frame (field) order has been respected;
   - to report the lowest frame rate it can sustain as defined in 8.15;
   - to report the highest bit-rate which can always be fully decoded as defined in 8.15.

Table 3 lists the obligations for an IUT evaluated to be compliant to a certain Cpoint.

Table 3 – Obligations for IUTs

	Cpoint-0 TCS	Cpoint-1 TCS	Cpoint-2 TCS	Cpoint-3 TCS
Cpoint-0 IUT	Must decode everything	Must decode up to compliance point parameters defined in Cpoint-0	Must decode up to compliance point parameters defined in Cpoint-0	Must decode up to compliance point parameters defined in Cpoint-0
Cpoint-1 IUT	Must decode everything	Must decode everything	Must decode up to compliance point parameters defined in Cpoint-1	Must decode up to compliance point parameters defined in Cpoint-1
Cpoint-2 IUT	Must decode everything	Must decode everything	Must decode everything	Must decode up to compliance point parameters defined in Cpoint-2
Cpoint-3 IUT	Must decode everything	Must decode everything	Must decode everything	Must decode everything

11 Executable test suite (ETS)

This clause defines four ETSs for the compliance points (Cpoints) defined in clause 9. Additional ETSs may be made available after the publication of this Recommendation | International Standard; see http://www.jpeg.org/software for the latest set of ETSs.

Each ETS consists of motion sequence and single image codestreams, reference decoded images, and tolerance values for MSE and peak error.

11.1 Test sequences

The following compressed test sequences listed in Tables 4 to 7 are attached to this Recommendation | International Standard.

Table 4 – List of test sequences (Cpoint-3)

Sequence name	Resolution	Video type	Number of frames	Format	Bit depth
cp31.mj2	4096 × 3112	24P	80	RGB 4:4:4	16

Table 5 – List of test sequences (Cpoint-2)

Sequence name	Resolution	Video type	Number of frames	Format	Bit depth
cp21.mj2	1920 × 1080	30P	240	YCC 4:2:2	8
cp22.mj2	1920 × 1080	24P	90	RGB 4:4:4	12
cp23.mj2	1280 × 1024	15P	90	YCC 4:4:4	10/11

Table 6 – List of test sequences (Cpoint-1)

Sequence name	Resolution	Video type	Number of frames	Format	Bit depth
cp11.mj2	720 × 480	30P	150	YCC 4:2:0	8
cp12.mj2	640 × 480	24P	90	RGB 4:4:4	8
cp13.mj2	720 × 576	50I	200 (400 fields)	YCCM 4:2:0:4	8

Table 7 – List of test sequences (Cpoint-0)

Sequence name	Resolution	Video type	Number of frames	Format	Bit depth
cp01.mj2	360 × 240	30P	150	YCC 4:2:0	8
cp02.mj2	320 × 240	24P	150	RGB 4:4:4	8
cp03.mj2	352 × 288	25P	200	Mono	8

NOTE – Resolutions in the above tables are in the form width × height, referred to the component(s) with the least sub-sampling. The column Video type in the above tables are in the form ##[P|I] with '##' indicating the original frame/field rate per second, 'P' indicating progressive scan (i.e., single frame images) and 'I' indicating interlaced scan (i.e., images consisting of two fields). Format information explains colour representation and sub-sampling style of colour components. Sequence 'cp13.mj2' has a 4th component which represents a two-level mask used during encoding.

11.2 Cpoint-3

11.2.1 Codestreams

The test codestreams for this ETS as listed in Tables 8 and 9 are in the directory 'Cpoint3'. There are motion sequences, with names of the form 'xxx.mj2' and single image codestreams, with names of the form 'xxx_###.j2k', where '###' is the frame number extracted from the corresponding motion sequence.

Table 8 – Reference motion sequences (Cpoint-3)

TCS	Number of frames	Bit depth	Width	Height	Video type	Format
cp31.mj2	80	16	4096	3112	24P	RGB 4:4:4

Table 9 – Reference image and allowable error (Cpoint-3)

TCS	Reference image file name	Frame number	Bit depth	Width	Height	Peak	MSE
cp31_045.j2k	cp31_045_1.pgx cp31_045_2.pgx cp31_045_3.pgx	45	16 16 16	4096 4096 4096	3112 3112 3112	0 0 0	0 0 0

11.2.2 Reference decoded images

The reference decoded single images for this ETS are in the directory 'reference_Cpoint3'. The filenames are of the form 'xxx_###_?.pgx' where '###' is the frame number (starting from 1) extracted from the corresponding motion sequence and '?' denotes the component number (starting from 1).

11.2.3 Tolerances

The maximum allowable MSE and peak errors are listed in Table 9, along with information on image properties.

11.3 Cpoint-2

11.3.1 Codestreams

The test codestreams for this ETS as listed in Tables 10 and 11 are in the directory 'Cpoint2'. There are motion sequences, with names of the form 'xxx.mj2' and single image codestreams, with names of the form 'xxx_###.j2k', where '###' is the frame number extracted from the corresponding motion sequence.

Table 10 – Reference motion sequences (Cpoint-2)

TCS	Number of frames	Bit depth	Width	Height	Video type	Format
cp21.mj2	240	8	1920	1080	24P	YCC 4:2:2
cp22.mj2	90	12	1920	1080	24P	RGB 4:4:4
cp23.mj2	90	10/11	1280	1024	15P	YCC 4:4:4

Table 11 – Reference image and allowable error (Cpoint-2)

TCS	Reference image file name	Frame number	Bit depth	Width	Height	Peak	MSE
cp21_045.j2k	cp21_045_1.pgx cp21_045_2.pgx cp21_045_3.pgx	45	8 8 8	1920 960 960	1080 1080 1080	0 0 0	0 0 0
cp22_045.j2k	cp22_045_1.pgx cp22_045_2.pgx cp22_045_3.pgx	45	12 12 12	1920 1920 1920	1080 1080 1080	14 11 18	6.46 4.28 8.18
cp23_045.j2k	cp23_045_1.pgx cp23_045_2.pgx cp23_045_3.pgx	45	10 11 11	1280 1280 1280	1024 1024 1024	2 5 5	0.42 0.64 0.52

11.3.2 Reference decoded images

The reference decoded single images for this ETS are in the directory 'reference_Cpoint2'. The filenames are of the form 'xxx_###_?.pgx' where '###' is the frame number (starting from 1) extracted from the corresponding motion sequence and '?' denotes the component number (starting from 1).

11.3.3 Tolerances

The maximum allowable MSE and peak errors are listed in Table 11, along with information on image properties.

11.4 Cpoint-1

11.4.1 Codestreams

The test codestreams for this ETS as listed in Tables 12 and 13 are in the directory 'Cpoint1'. There are motion sequences, with names of the form 'xxx.mj2' and single image codestreams, with names of the form 'xxx_###.j2k', where '###' is the frame number extracted from the corresponding motion sequence.

Table 12 – Reference motion sequences (Cpoint-1)

TCS	Number of frames	Bit depth	Width	Height	Video type	Format
cp11.mj2	150	8	720	480	30P	YCC 4:2:0
cp12.mj2	90	8	640	480	24P	RGB 4:4:4
cp13.mj2	200 (400 fields)	8	720	576	50I	YCCM 4:2:0:4

Table 13 – Reference image and allowable error (Cpoint-1)

TCS	Reference image file name	Frame number	Bit depth	Width	Height	Peak	MSE
cp11_045.j2k	cp11_045_1.j2k cp11_045_2.j2k cp11_045_3.j2k	45	8 8 8	720 360 360	480 240 240	4 3 4	0.09 0.12 0.06
cp12_045.j2k	cp12_045_1.j2k cp12_045_2.j2k cp12_045_3.j2k	45	8 8 8	640 640 640	480 480 480	0 0 0	0 0 0
cp13_045_top.j2k	cp13_045_top_1.j2k cp13_045_top_2.j2k cp13_045_top_3.j2k	45	8 8 8	720 360 360	576 288 288	3 1 1	0.19 0.09 0.08
cp13_045_bot.j2k	cp13_045_bot_1.j2k cp13_045_bot_2.j2k cp13_045_bot_3.j2k	45	8 8 8	720 360 360	576 288 288	3 2 2	0.20 0.08 0.08

11.4.2 Reference decoded images

The reference decoded single images for this ETS are in the directory 'reference_Cpoint1'. The filenames are of the form 'xxx_###_?.pgx' where '###' is the frame number (starting from 1) extracted from the corresponding motion sequence and '?' denotes the component number (starting from 1). For interlaced material, the top and bottom fields are denoted with 'top' and 'bot' together with the frame number.

11.4.3 Tolerances

The maximum allowable MSE and peak errors are listed in Table 13, along with information on image properties.

11.5 Cpoint-0

11.5.1 Codestreams

The test codestreams for this ETS as listed in Tables 14 and 15 are in the directory 'Cpoint0'. There are motion sequences, with names of the form 'xxx.mj2' and single image codestreams, with names of the form 'xxx_###.j2k', where '###' is the frame number extracted from the corresponding motion sequence.

Table 14 – Reference motion sequences (Cpoint-0)

TCS	Number of frames	Bit depth	Width	Height	Video type	Format
cp01.mj2	150	8	360	240	30P	YCC 4:2:0
cp02.mj2	90	8	320	240	30P	RGB 4:4:4
cp03.mj2	200	8	352	288	25P	Mono

Table 15 – Reference image and allowable error (Cpoint-0)

TCS	Reference image file name	Frame number	Bit depth	Width	Height	Peak	MSE
cp01_045.j2k	cp01_045_1.j2k cp01_045_2.j2k cp01_045_3.j2k	45	8 8 8	360 180 180	240 120 120	3 1 1	0.16 0.18 0.17
cp02_045.j2k	cp02_045_1.pgx cp02_045_2.pgx cp02_045_3.pgx	45	8 8 8	320 320 320	240 240 240	0 0 0	0 0 0
cp03_045.j2k	cp03_045.pgx	45	8	352	288	3	0.21

11.5.2 Reference decoded images

The reference decoded single images for this ETS are in the directory 'reference_Cpoint0'. The filenames are of the form 'xxx_###_?.pgx' or 'xxx_###.pgx' where '###' is the frame number (starting from 1) extracted from the corresponding motion sequence and '?' denotes the component number (starting from 1).

11.5.3 Tolerances

The maximum allowable MSE and peak errors are listed in Table 15, along with information on image properties.

Annex A

File and codestream profiles

(This annex forms an integral part of this Recommendation | International Standard)

A.1 Profile introduction

This annex normatively specifies the ways in which Motion JPEG 2000 files can be profiled.

There are three areas for which profiles and levels can be declared:

a) The optional still image which may occur at the top-level of a Motion JPEG 2000 file.

b) The characteristics of the individual frames (codestreams) which form a motion sequence.

c) The overall characteristics of the Motion Presentation itself.

The profile for the optional still image is declared by including the correct 'compatible brands' in the top-level File Type Box. These profiles (brands) are specified for JP2 files, and that specification shall be followed here.

Profile code-points should be used to profile the images that form the frames of motion sequences. These code-points are placed within the JP2 Profile Box within the sample description.

The characteristics of the motion presentation are declared by placing the brands of the compatible profiles in the top-level file-type box. This Recommendation | International Standard defines two profiles:

1) Unrestricted. The brand 'mjp2' indicates unrestricted conformance to this Recommendation | International Standard and shall be placed in the top-level File Type Box in all files.

2) Simple. The brand 'mj2s' indicates the simple Motion JPEG 2000 profile, as defined below.

A.2 Motion JPEG 2000 simple profile

File conforming to the simple profile have the following characteristics:

1) Exactly one video track is present.

2) At most a single audio track, using only 8- or 16-bit raw audio, is present.

3) Each track shall have exactly one sample description, used by all samples.

4) The sample rate of the audio, if present, shall not exceed 48 kHz.

5) The frame rate of the video shall not exceed 30 frames per second.

6) The video codestream profile shall be profile 0 for both the motion sequence and the still image, if present.

7) The file is self-contained; no data references are used, and therefore all media data is contained within the single file.

8) The media data in the Media Data Box(es) is placed within the box(es) in temporal order.

9) If more than one track is present, the media data for the tracks is interleaved, with a granularity no greater than the greater of:

   a) the duration of a single 'sample' (in file format terms) or;

   b) one second.

10) The transformation matrices used are restricted to uniform scaling and rotation by multiples of 90º.

Annex B

Guidelines for use of the JPEG 2000 codec

(This annex does not form an integral part of this Recommendation | International Standard)

B.1 Introduction

Certain artifacts have been observed when using ITU-T Rec. T.800 | ISO/IEC 15444-1 codec to create motion sequences. Experiments have been performed to identify the cause of these artifacts and indicate encoding strategies that would be likely to mitigate them. This annex is intended to review the results of the experiments that were performed and recommend possible methods that may be employed. These methods are not necessarily the best or only approach to addressing these observed artifacts. This annex is intended as a point of reference for individuals implementing this Recommendation | International Standard (Motion JPEG 2000).

Subclause B.2 suggests particular frequency weighting for motion sequences. Subclause B.3 addresses sub-sampling issues.

B.2 Frequency weighting for motion sequences

The objective of employing frequency weighting to motion images is to explore methods that might improve subjective quality. Tuning of weighting values for chroma component is an important area to investigate. An experiment showed that there are some flickering artifacts when displaying a moving sequence of images although individual pictures do not show an annoying artifact. The frequency weighting was observed to reduce these artifacts in all test sequences, especially at low to middle bitrate (for example 0.25 to 0.50 bpp for CIF). A finer and better adjustment of visual weights to improve the image sequence's quality may provide additional improvements.

For colour images, the frequency weighting tables of the Y, Cr, and Cb components should differ in order to take advantage of the properties of the human visual system. For example, it is usually desirable to emphasize the luminance component more than the chrominance components.

Table B.1 specifies a set of CSF weights which were designed for the luminance component based on the CSF value at the mid-frequency of each subband. The other two tables (Tables B.2 and B.3) specify sets of CSF weights for Cb and Cr. Note that the tables are intended for a 5-level wavelet decomposition.

The table does not include the weight for the lowest frequency subband, nLL, which is always 1. Levels 1, 2, …, 5 denote the subband levels in low to high frequency order. (HL, LH, HH) denotes the three frequency orientations within each subband.

Table B.1 – CSF weights for luminance

Level	HL	LH	HH
1	1.000000	1.000000	1.000000
2	1.000000	1.000000	1.000000
3	0.999994	0.999994	0.999988
4	0.837755	0.837755	0.701837
5	0.275783	0.275783	0.090078

Table B.2 – CSF weights for Cb

Level	HL	LH	HH
1	0.812612	0.812612	0.737656
2	0.679829	0.679829	0.567414
3	0.488887	0.488887	0.348719
4	0.267216	0.267216	0.141965
5	0.089950	0.089950	0.027441

Table B.3 – CSF weights for Cr

Level	HL	LH	HH
1	0.856065	0.856065	0.796593
2	0.749805	0.749805	0.655884
3	0.587213	0.587213	0.457826
4	0.375176	0.375176	0.236030
5	0.166647	0.166647	0.070185

B.3 Encoder sub-sampling of components

In the most of all cases, the format of motion sequences is either YcbCr (BT-601-5 for standard TV and BT-709-4 for HDTV) or YUV. In these, colour components are already sub-sampled and the sub-sampled signal may be the input of the Motion JPEG 2000 encoder. Although the encoder sub-sampling of colour components is not recommended in ITU-T Rec. T.800 | ISO/IEC 15444-1, the sub-sampling may be commonly used in Motion JPEG 2000. Therefore, CSF weights also should be weighted directly to sub-sampled colour components.

Sub-sampling of chroma components in JPEG 2000 can be achieved in two ways:

a) the highest-frequency subbands may be omitted; and

b) if no ICT is used, then the codestream permits the components to differ in size.

Annex C

Indicating sub-sampling chroma offset

(This annex does not form an integral part of this Recommendation | International Standard)

This annex provides informative material for the Sub-sampling Box 'jsub' (see 6.1), with offset values for common sub-sampling formats.

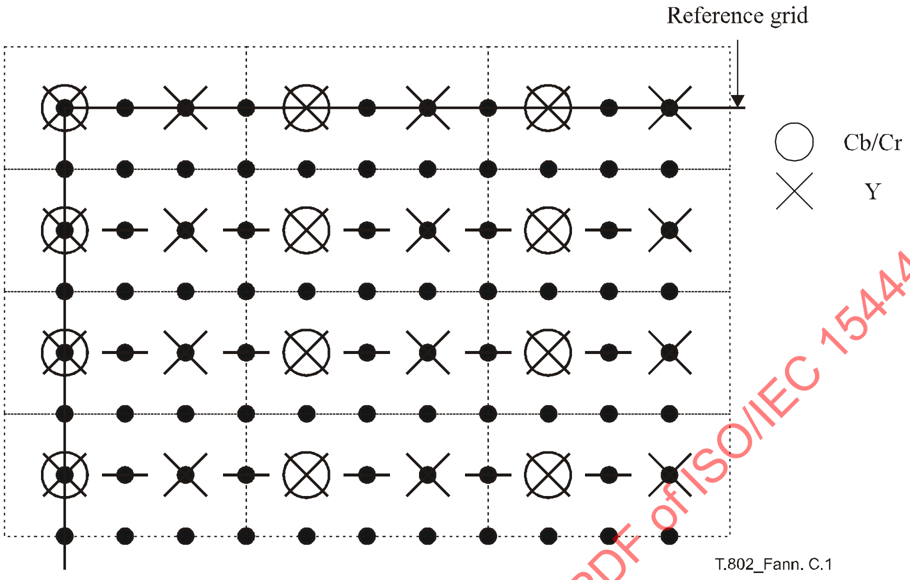

Figure C.1 – 4:2:2 format (co-sited)

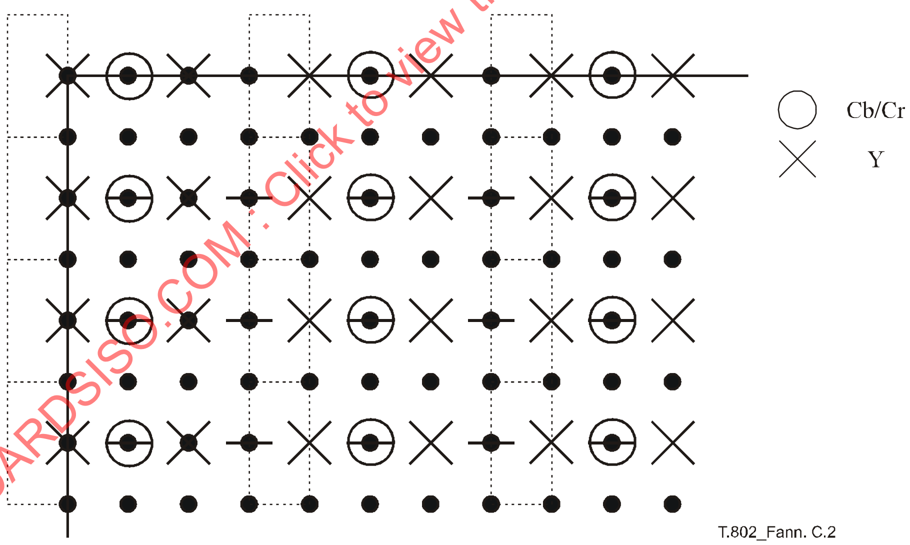

Figure C.2 – 4:2:2 format (centred)

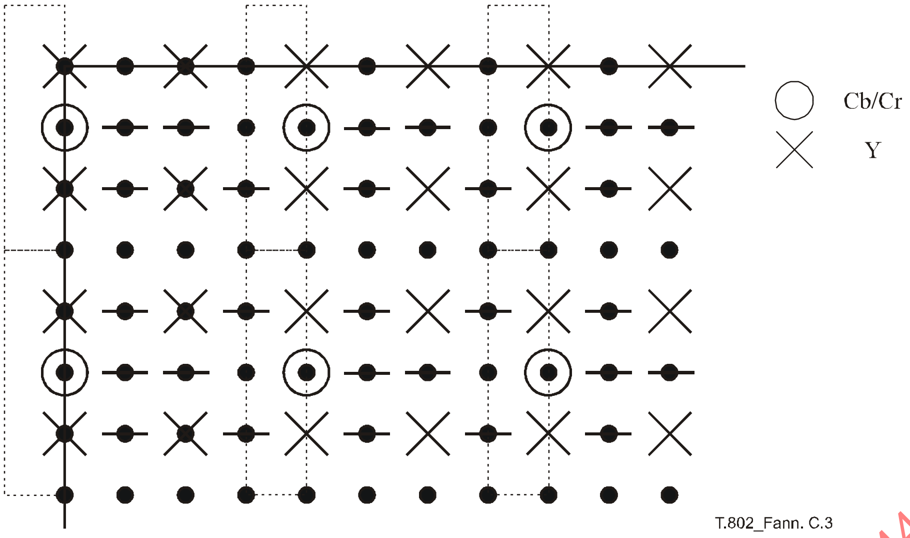

Figure C.3 – 4:2:0 format (co-sited)

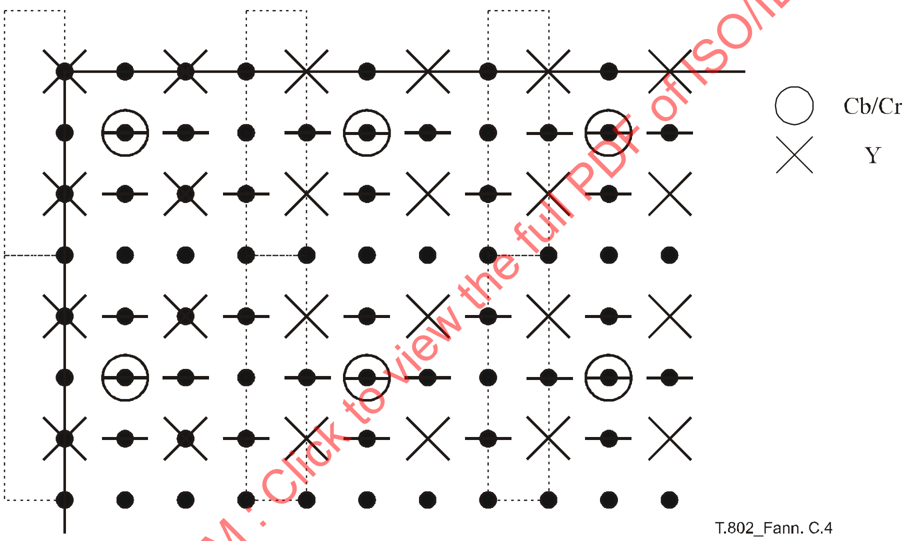

Figure C.4 – 4:2:0 format (centred)

Table C.1 – Chroma phase values

	Figure C.1	Figure C.2	Figure C.3	Figure C.4
Horizontal-offset	0	1	0	1
Vertical-offset	0	0	1	1

Annex D

Field structures for interlace

(This annex does not form an integral part of this Recommendation | International Standard)

In the case of frame-based coding in an interlaced sequence, each sample is composed of lines from two fields. Encoding is illustrated in Figure D.1 and decoding in Figure D.4. In the case of field-based coding, each sample is composed of lines from one of the two fields. There are two types of vertical and temporal positions of samples in an interlaced frame. Figures D.2 and D.5 show encoding and decoding when the topmost line of the frame belongs to the earlier field and the second line of the frame belongs to the later field. Figures D.3 and D.5 show the other case where the topmost line belongs to the later field.

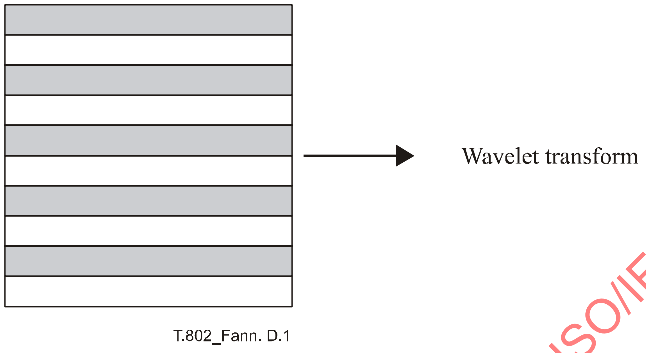

Figure D.1 – Picture structure in frame-based coding

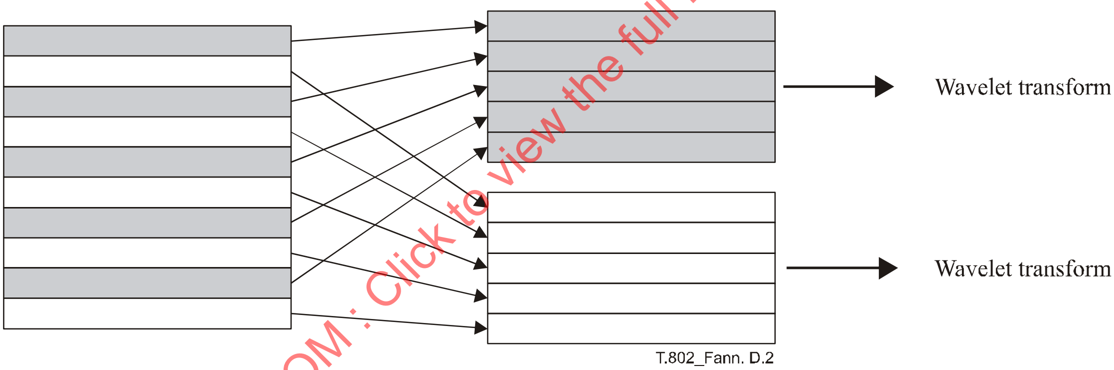

Figure D.2 – Picture structure in field-based coding (topmost line first)

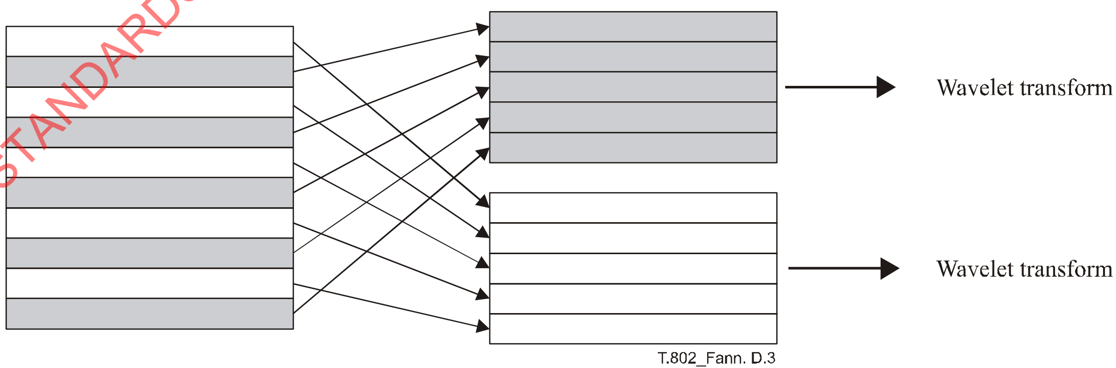

Figure D.3 – Picture structure in field-based coding (topmost line second)

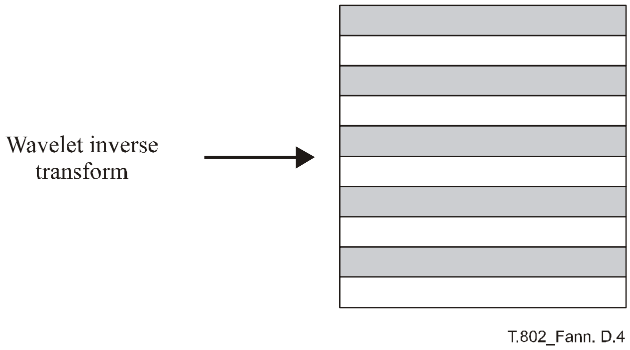

Figure D.4 – Picture structure in frame-based decoding

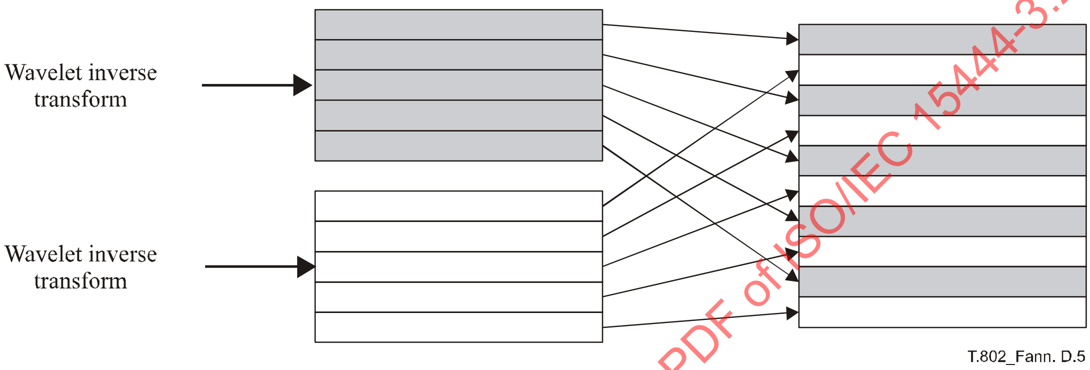

Figure D.5 – Picture structure in field-based decoding (topmost line first)

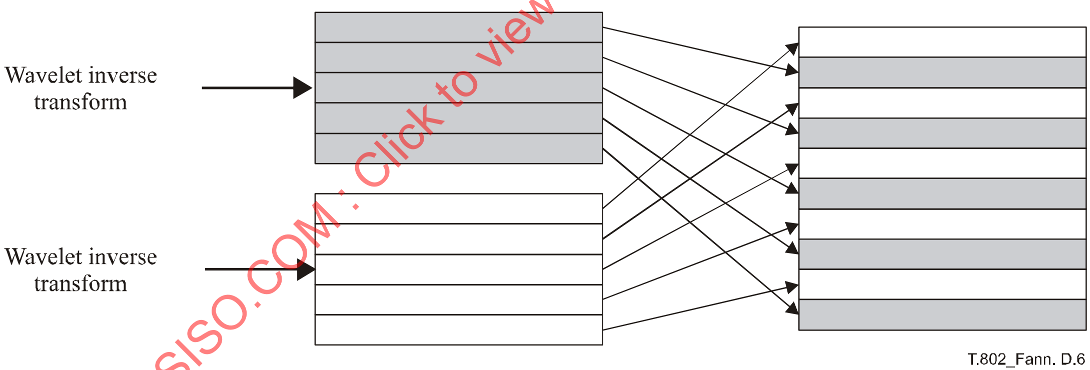

Figure D.6 – Picture structure in field-based decoding (topmost line second)

Annex E

Guidelines for implementing Motion JPEG 2000

(This annex does not form an integral part of this Recommendation | International Standard)

E.1 Introduction

The Motion JPEG 2000 file format is derived from the ISO Base Media File Format. Members from the ISO/IEC JPEG and MPEG committees jointly maintain the ISO Base Media File Format, known as ISO/IEC 15444-12 and ISO/IEC 14496-12 (two standards with identical text). It is expected that the ISO/IEC Base Media File Format will be extended as new features are proposed to support desired new functionality.

When implementing the MJ2 file format, it is important to note that not all features of the ISO Base Media File Format are required. Features that are not required for a compliant Motion JPEG 2000 implementation may be used as an option. However, use of optional features could limit interoperability with other Motion JPEG 2000 compliant implementations. And, use of optional features from the ISO Base Media File Format may require a license. A compliant MJ2 file format is targeted to be IP free.

NOTE – To implement Motion JPEG 2000, both ITU-T Rec. T.802 | ISO/IEC 15444-3 and ISO/IEC 15444-12 documents will be needed and will be supplied by ISO when ordering ITU-T Rec. T.802 | ISO/IEC 15444-3. Since there are features in ISO/IEC 15444-12 that are not part of a compliant Motion JPEG 2000 implementation, please refer to the following guidelines and tables for required (normative) features and optional features. A compliant Motion JPEG 2000 implementation must support required features. A compliant implementation need not support optional features, but it must not fail if optional features are present.

E.2 Guidelines

Motion JPEG 2000 video is stored in video tracks, as documented in the ISO Base Media File Format. This format defines the structure of video tracks in a format-independent fashion. The exact format of the video is declared by the sample description; the sample description for Motion JPEG 2000 is defined in Clause 6 of this Recommendation | International Standard.

The format of the samples is defined by the declarations in the sample description and the corresponding specification. For Motion JPEG 2000, the sample format contains JPEG 2000 code-streams; again, the precise format is defined in ITU-T Rec. T.800 | ISO/IEC 15444-1 and Clause 6 of this Recommendation | International Standard.

Since Motion JPEG 2000 is an I-frame coder (there are no difference frames), some of the structures in the base ISO media file format are not needed: the sync sample table (and shadow sync) are only needed when there are difference frames, and so are unused here. Similarly, the composition time to sample table is only needed when there is decode/composition re-ordering (as happens with B frames), and so is not needed.

Motion JPEG 2000 files may also contain audio. The structure of an audio track is like that of a video track. Support for uncompressed (raw) audio is defined in this Recommendation | International Standard (both sample description and sample format). The definitions for MPEG-4 audio may be found in the MPEG-4 specifications.

Padding bits are only needed for MPEG-4 audio (the padb table) and since support of MPEG-4 audio is optional, so is support of this table.

Hinting and support for hinting (such as hint tracks) are not part of the Motion JPEG 2000 specification, so all the structures relating to hint tracks are not required. Track references are not used by the basic Motion JPEG 2000 specification, but are used for various purposes including hinting; therefore, track reference atoms are optional.

The Motion JPEG 2000 specification uses only audio and video media tracks; the null media header used by other tracks, including MPEG-4 systems tracks, is not used here.

The use of interlaced material is not recommended, but possible if necessary. Refer to clause 6 where organization of media data and meta data is explained. Explicitly, when two fields are present in the samples, the JP2 Header Box applies to the complete image, not to each field individually. Therefore, the height as declared in the JP2 Header Box ('jp2h') and the VisualSampleEntry ('mjp2') applies to the entire de-interlaced image. `Entry-size` in the Sample Size Box ('stsz') is therefore the sum of the two corresponding JP2 Codestream Boxes ('jp2c'). Those codestream boxes shall always appear in temporal order, whilst `fieldorder` entries in the Field Coding Box ('fiel') and `original_fieldorder` entries in MJP2 Original Format Box ('orfo') help properly reconstructing the decoded images.

When encoding video source images provided in a chrominance sub-sampled YCC format, it is recommended to signal this through MJP2 Sub-sampling Box ('jsub'). Decoders should be aware that this box is optional and therefore eventually sub-sampling information shall be retrieved from the JP2 codestream.

Table E.1 – Required features for Motion JPEG 2000

ftyp						*		*file type and compatibility*
moov						*		*container for all the meta-data*
	mvhd					*		*movie header, overall declarations*
	trak					*		*container for an individual track or stream*
		tkhd				*		*track header, overall information about the track*
		tref						*track reference container*
		edts						*edit list container*
			elst					*an edit list*
		mdia				*		*container for the media information in a track*
			mdhd			*		*media header, overall information about the media*
			hdlr			*		*handler, declares the media (handler) type*
			minf			*		*media information container*
				vmhd				*video media header, overall information (video track only)*
				smhd				*sound media header, overall information (sound track only)*
				dinf		*		*data information box, container*
					dref	*		*data reference box, declares source(s) of media data in track*
				stbl		*		*sample table box, container for the time/space map*
					stsd	*		*sample descriptions (codec types, initialization, etc.)*
					stts	*		*(decoding) time-to-sample*
					stsc	*		*sample-to-chunk, partial data-offset information*
					stsz			*sample sizes (framing)*
					stz2			*compact sample sizes (framing)*
					stco	*		*chunk offset, partial data-offset information*
					co64			*64-bit chunk offset*
	mvex							*movie extends box*
		mehd						*movie extends header box*
		trex				*		*track extends defaults*
moof								*movie fragment*
	mfhd					*		*movie fragment header*
	traf							*track fragment*
		tfhd				*		*track fragment header*
		trun						*track fragment run*
mfra								*movie fragment random access*
	tfra							*track fragment random access*
	mfro					*		*movie fragment random access offset*
mdat								*media data container*
free								*free space*
skip								*free space*
	udta							*user-data*
		cprt						*copyright, etc.*