FINAL
DRAFT

AMENDMENT

ISO/IEC 14496-10:2004 FDAM 1

ISO/IEC JTC **1**

Secretariat: **ANSI**

Voting begins on:
2004-12-07

Voting terminates on:
2005-02-07

Information technology — Coding of audio-visual objects —

Part 10: Advanced Video Coding

AMENDMENT 1: AVC professional extensions

Technologies de l'information — Codage des objets audiovisuels —
Partie 10: Codage visuel avancé
AMENDEMENT 1: Extensions AVC professionnelles

Please see the administrative notes on page iii

Reference number
ISO/IEC 14496-10:2004/FDAM 1:2004(E)

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 be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Amendment 1 to ISO/IEC 14496-10:2004 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information.*

Introduction

This document is the Fidelity Range Extensions Amendment to ITU-T Rec. H.264 | ISO/IEC 14496-10.

The Recommendation | International Standard has been modified in areas where extended sample bit depth or alternative chroma format support requires new or substitute text in order to implement an extended-capability encoder or decoder. The original Recommendation | International Standard specified operation with 8 bits per sample and 4:2:0 chroma format (a format in which the two chroma arrays each have half the horizontal and vertical resolution of the corresponding luma array). This amendment extends this capability to support more than 8 bits per sample and additional chroma formats as follows.

– Monochrome (a format in which only a luma array is present)

– 4:2:2 (a format in which the two chroma arrays each have half the horizontal resolution and the same vertical resolution as in the corresponding luma array)

– 4:4:4 (a format in which the two chroma arrays each have the same horizontal and vertical resolution as in the corresponding luma array)

Additionally, auxiliary pictures can be included in the bitstream for such purposes as alpha blending composition.

The variety of colour spaces that can be explicitly indicated as the source of the video content has also been extended.

Four features have been added to the specified decoding process for improved coding efficiency.

– Support for adaptive selection between 4x4 and 8x8 block sizes for the luma spatial transform, enabling adaptivity to the correlation characteristics of the source video data.

– Support for encoder-specified frequency-dependent scaling matrices for transform coefficients, enabling optimization of quantization operation to take advantage of frequency-dependent human visual system sensitivity.

– Support for a residual colour transform, enabling efficient representation of video in alternative colour spaces (such as RGB) without requiring increased bit depth for decoded pictures and without imposing colour-space conversion rounding error in a manner cascaded with the encoding and decoding processes.

– Support for a transform bypass representation, enabling relatively-efficient lossless video representation.

Three new supplemental enhancement information (SEI) messages have been added to enable enhanced uses of the decoded video.

– Description of the characteristics of film grain for film and video, for use when it is advantageous to synthesize simulated film grain with specified characteristics as a post-process after decoding.

– Indication of a preference between the display of pictures that precede or follow the application of the deblocking filter process that is part of the decoding process for enhancement of the perceptual quality of the decoded video.

– Indication of the use of the video bitstream for stereo video display.

In the following paragraphs, the phrase ITU-T Rec. H.264 | ISO/IEC 14496-10 version 1 refers to the first (2003) approved version of this Recommendation | International Standard. The phrase ITU-T Rec. H.264 | ISO/IEC 14496-10 version 2 refers to the integrated text containing the corrections specified in technical corrigendum 1 (2004). The phrase ITU-T Rec. H.264 | ISO/IEC 14496-10 version 3 refers to the integrated text containing both technical corrigendum 1 (2004) and this fidelity range extensions amendment.

The numbering in this fidelity range extensions amendment text is relative to the text of ITU-T Rec. H.264 | ISO/IEC 14496-10 version 2. When a numbered item (i.e., a clause, subclause, figure, table, or equation) or associated content is being replaced or modified, the same number is used for the modified numbered item. When a numbered item is inserted between prior numbered items, the number of the corresponding numbered item immediately preceding it is used and the letter 'a' is appended to this number. When, after this one such inserted numbered item, another numbered item is inserted, the letter 'a' is replaced by the letter 'b' to indicate their relative order, and so on, following ordinary English alphabetical order. In the text of ITU-T Rec. H.264 | ISO/IEC 14496-10 version 3 (which contains the integrated fidelity range extensions amendment text), the inserted numbered items are to be assigned the corresponding number in their numerical order without any such letters, and any subsequent numbered items are to be assigned later numbers to avoid conflicts. The purpose of the numbering convention in this amendment text is to avoid renumbering of numbered items in ITU-T Rec. H.264 | ISO/IEC 14496-10 version 2 while drafting this amendment. Therefore, if the addition of a

numbered item does not require renumbering of numbered items in ITU-T Rec. H.264 | ISO/IEC 14496-10 version 2, the final number is assigned to the numbered item herein.

For example, the equation inserted after Equation 5-3 receives the number 5-3a in this text and will receive the number 5-4 in ITU-T Rec. H.264 | ISO/IEC 14496-10 version 3. That causes all equation numbers starting with 5-4 and larger in ITU-T Rec. H.264 | ISO/IEC 14496-10 version 2 to increment by one after the integration of this amendment.

Within clause 3, numbers have been added using subordinate clauses as in ITU-T Rec. H.264 | ISO/IEC 14496-10 version 2. After integration to form version 3, all subclauses of clause 3 should be made into a single consecutive list of subclauses all at the same level of subclause hierarchy.

INTERNATIONAL STANDARD

ITU-T RECOMMENDATION

Information technology – Coding of audio-visual objects – Advanced Video Coding

AMENDMENT 1
AVC fidelity range extensions

1) New Subclause 0.3a "Publication and versions of this specification"

Insert a new subclause 0.3a as follows.

0.3a Publication and versions of this specification

This subclause does not form an integral part of this Recommendation | International Standard.

This specification has been jointly developed by ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group. It is published as technically-aligned twin text in both organizations ITU-T and ISO/IEC.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 1 refers to the first (2003) approved version of this Recommendation | International Standard.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 2 refers to the integrated text containing the corrections specified in technical corrigendum 1 (2004).

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 3 (the current specification) refers to the integrated text containing both technical corrigendum 1 (2004) and Amendment 1 "Fidelity range extensions".

2) Subclause 0.5 "Overview of the design characteristics"

*In subclause 0.5, replace the sentence "*The algorithm is not lossless, as the exact source sample values are typically not preserved through the encoding and decoding processes.*" with "*With the exception of the transform bypass mode of operation for lossless coding in the High 4:4:4 profile and the I_PCM mode of operation in all profiles, the algorithm is typically not lossless, as the exact source sample values are typically not preserved through the encoding and decoding processes.*".*

3) Subclause 0.6 "How to read this specification"

*In subclause 0.6, replace the sentence "*Annex A defines three profiles (Baseline, Main, and Extended), each being tailored to certain application domains, and defines the so-called levels of the profiles.*" with "*Annex A specifies seven profiles (Baseline, Main, Extended, High, High 10, High 4:2:2 and High 4:4:4), each being tailored to certain application domains, and defines the so-called levels of the profiles.*".*

4) New subclause 3.4.1 "alpha blending"

Insert a new subclause 3.4.1 as follows.

3.4.1 alpha blending: A process not specified by this Recommendation | International Standard, in which an *auxiliary coded picture* is used in combination with a *primary coded picture* and with other data not specified by this Recommendation | International Standard in the *display process.* In an alpha blending process, the

samples of an *auxiliary coded picture* are interpreted as indications of the degree of opacity (or, equivalently, the degrees of transparency) associated with the corresponding *luma* samples of the *primary coded picture.*

5) New subclause 3.5.1 "auxiliary coded picture"

Insert a new subclause 3.5.1 as follows.

3.5.1 **auxiliary coded picture**: A *picture* that supplements the *primary coded picture* that may be used in combination with other data not specified by this Recommendation | International Standard in the *display process.* An auxiliary coded picture has the same syntactic and semantic restrictions as a monochrome *redundant coded picture.* An auxiliary coded picture must contain the same number of *macroblocks* as the *primary coded picture*. Auxiliary coded pictures have no normative effect on the *decoding process*. See also *primary coded picture* and *redundant coded picture.*

6) New subclause 3.39.1 "display process"

Insert a new subclause 3.39.1 as follows.

3.39.1 **display process**: A process not specified in this Recommendation | International Standard having, as its input, the cropped decoded *pictures* that are the output of the *decoding process.*

7) New subclause 3.59.1 "interpretation sample value"

Insert a new subclause 3.59.1 as follows:

3.59.1 **interpretation sample value**: A possibly-altered value corresponding to a decoded sample value of an *auxiliary coded picture* that may be generated for use in the *display process*. Interpretation sample values are not used in the *decoding process* and have no normative effect on the *decoding process*.

8) Subclause 5.7 "Mathematical functions"

Replace Equation 5-3 of subclause 5.7 with the following two equations.

$$Clip1_Y(x) = Clip3(0, (1 << BitDepth_Y) - 1 , x) \quad (5\text{-}3)$$

$$Clip1_C(x) = Clip3(0, (1 << BitDepth_C) - 1 , x) \quad (5\text{-}3a)$$

9) Subclause 6.2 "Source, decoded, and output picture formats"

Replace the content of subclause 6.2 with the following

6.2 Source, decoded, and output picture formats

This subclause specifies the relationship between source and decoded frames and fields that is given via the bitstream.

The video source that is represented by the bitstream is a sequence of either or both frames or fields (called collectively pictures) in decoding order.

The source and decoded pictures (frames or fields) are each comprised of one or more sample arrays:

- Luma (Y) only (monochrome), with or without an auxiliary array
- Luma and two Chroma (YCbCr or YCgCo), with or without an auxiliary array
- Green, Blue and Red (GBR, also known as RGB), with or without an auxiliary array

– Arrays representing other unspecified monochrome or tri-stimulus colour samplings (for example, YZX, also known as XYZ), with or without an auxiliary array

For convenience of notation and terminology in this specification, the variables and terms associated with these arrays are referred to as luma (or L or Y) and chroma, where the two chroma arrays are referred to as Cb and Cr; regardless of the actual colour representation method in use. The actual colour representation method in use can be indicated in syntax that is specified in Annex E. The (monochrome) auxiliary arrays, which may or may not be present as auxiliary pictures in a coded video sequence, are optional for decoding and can be used for such purposes as alpha blending.

The variables SubWidthC, and SubHeightC are specified in Table 6-1, depending on the chroma format sampling structure, which is specified through chroma_format_idc. An entry marked as "-" in Table 6-1 denotes an undefined value for SubWidthC or SubHeightC. Other values of chroma_format_idc, SubWidthC, and SubHeightC may be specified in the future by ITU-T | ISO/IEC.

Table 6-1 –SubWidthC, and SubHeightC values derived from chroma_format_idc

chroma_format_idc	Chroma Format	SubWidthC	SubHeightC
0	monochrome	-	-
1	4:2:0	2	2
2	4:2:2	2	1
3	4:4:4	1	1

In monochrome sampling there is only one sample array, which shall nominally be considered the luma array.

In 4:2:0 sampling, each of the two chroma arrays has half the height and half the width of the luma array.

In 4:2:2 sampling, each of the two chroma arrays has the same height and half the width of the luma array.

In 4:4:4 sampling, each of the two chroma arrays has the same height and width as the luma array.

The width and height of the luma sample arrays are each an integer multiple of 16. In bitstreams using 4:2:0 chroma sampling, the width and height of chroma sample arrays are each an integer multiple of 8. In bitstreams using 4:2:2 sampling, the width of the chroma sample arrays is an integer multiple of 8 and the height is an integer multiple of 16. The height of a luma array that is coded as two separate fields or in macroblock-adaptive frame-field coding (see below) is an integer multiple of 32. In bitstreams using 4:2:0 chroma sampling, the height of each chroma array that is coded as two separate fields or in macroblock-adaptive frame-field coding (see below) is an integer multiple of 16. The width or height of pictures output from the decoding process need not be an integer multiple of 16 and can be specified using a cropping rectangle.

The syntax for the luma and (when present) chroma arrays are ordered such when data for all three colour components is present, the data for the luma array is first, followed by any data for the Cb array, followed by any data for the Cr array, unless otherwise specified.

The width of fields coded referring to a specific sequence parameter set is the same as that of frames coded referring to the same sequence parameter set (see below). The height of fields coded referring to a specific sequence parameter set is half that of frames coded referring to the same sequence parameter set (see below).

The number of bits necessary for the representation of each of the samples in the luma and chroma arrays in a video sequence is in the range of 8 to 12, and the number of bits used in the luma array may differ from the number of bits used in the chroma arrays.

When the value of chroma_format_idc is equal to 1, the nominal vertical and horizontal relative locations of luma and chroma samples in frames are shown in Figure 6-1. Alternative chroma sample relative locations may be indicated in video usability information (see Annex E).

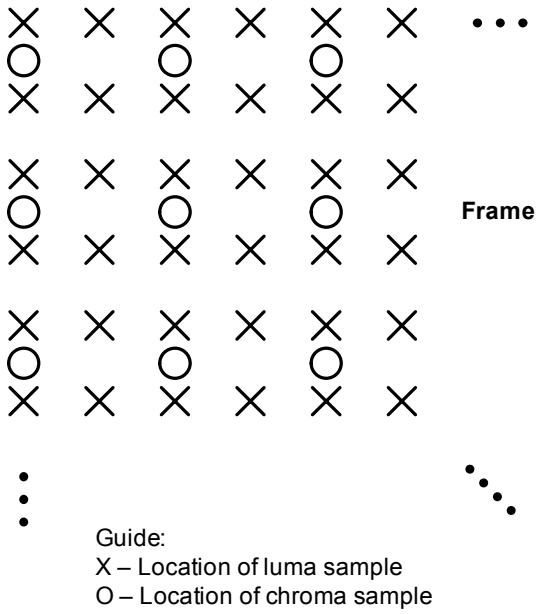

Figure 6-1 – Nominal vertical and horizontal locations of 4:2:0 luma and chroma samples in a frame

A frame consists of two fields as described below. A coded picture may represent a coded frame or an individual coded field. A coded video sequence conforming to this Recommendation | International Standard may contain arbitrary combinations of coded frames and coded fields. The decoding process is also specified in a manner that allows smaller regions of a coded frame to be coded either as a frame or field region, by use of macroblock-adaptive frame-field coding.

Source and decoded fields are one of two types: top field or bottom field. When two fields are output at the same time, or are combined to be used as a reference frame (see below), the two fields (which shall be of opposite parity) are interleaved. The first (i.e., top), third, fifth, etc. rows of a decoded frame are the top field rows. The second, fourth, sixth, etc. rows of a decoded frame are the bottom field rows. A top field consists of only the top field rows of a decoded frame. When the top field or bottom field of a decoded frame is used as a reference field (see below) only the even rows (for a top field) or the odd rows (for a bottom field) of the decoded frame are used.

When the value of chroma_format_idc is equal to 1, the nominal vertical and horizontal relative locations of luma and chroma samples in top and bottom fields are shown in Figure 6-2. The nominal vertical sampling relative locations of the chroma samples in a top field are specified as shifted up by one-quarter luma sample height relative to the field-sampling grid. The vertical sampling locations of the chroma samples in a bottom field are specified as shifted down by one-quarter luma sample height relative to the field-sampling grid. Alternative chroma sample relative locations may be indicated in the video usability information (see Annex E).

NOTE – The shifting of the chroma samples is in order for these samples to align vertically to the usual location relative to the full-frame sampling grid as shown in Figure 6-1.

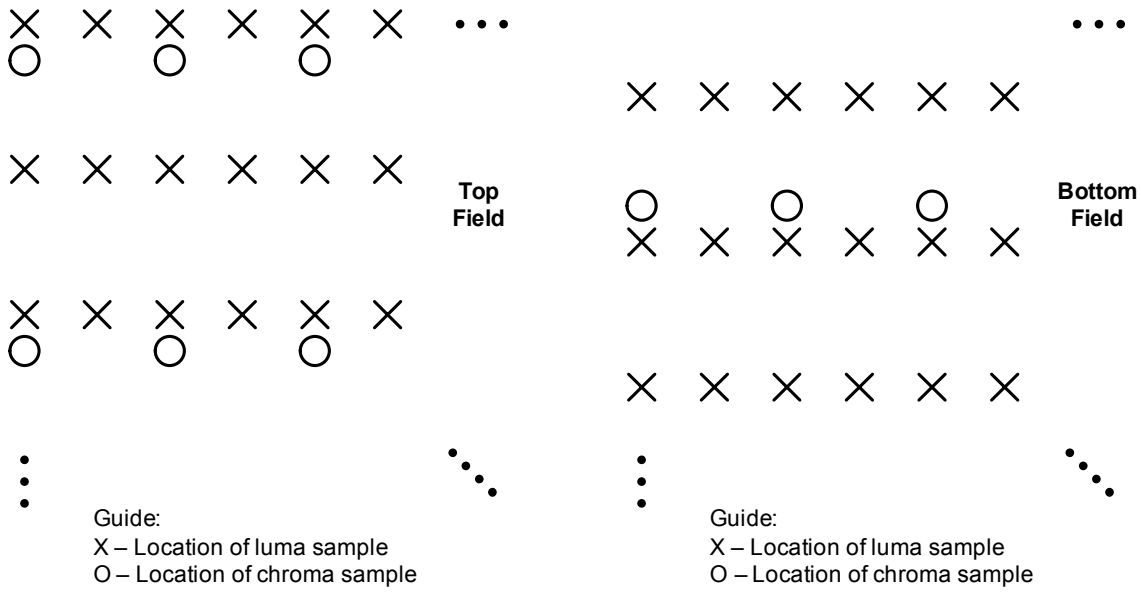

Figure 6-2 – Nominal vertical and horizontal sampling locations of 4:2:0 samples in top and bottom fields

When the value of chroma_format_idc is equal to 2, the chroma samples are co-sited with the corresponding luma samples and the nominal locations in a frame and in fields are as shown in Figure 6-2a and Figure 6-2b, respectively.

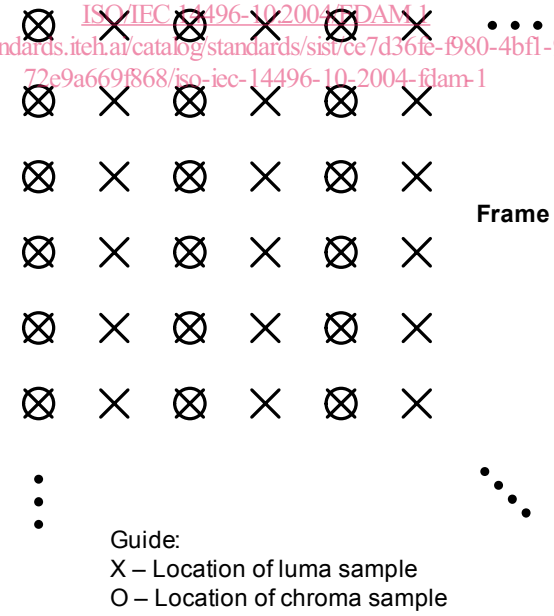

Figure 6-2a – Nominal vertical and horizontal locations of 4:2:2 luma and chroma samples in a frame

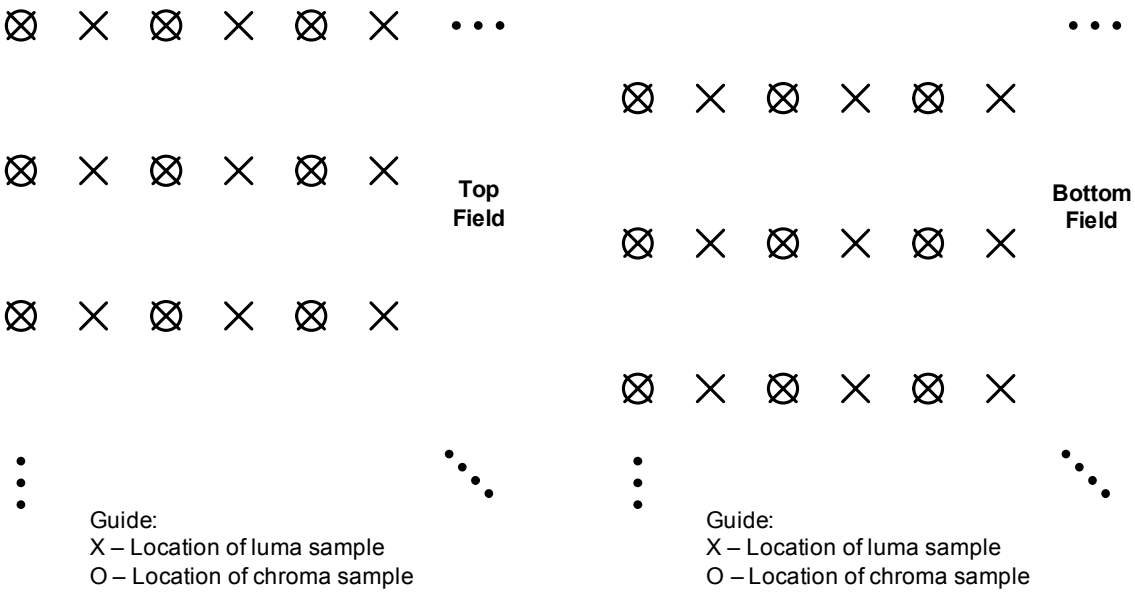

Figure 6-2b – Nominal vertical and horizontal sampling locations of 4:2:2 samples top and bottom fields

When the value of chroma_format_idc is equal to 3, all array samples are co-sited for all cases of frames and fields and the nominal locations in a frame and in fields are as shown in Figure 6-2c and Figure 6-2d, respectively.

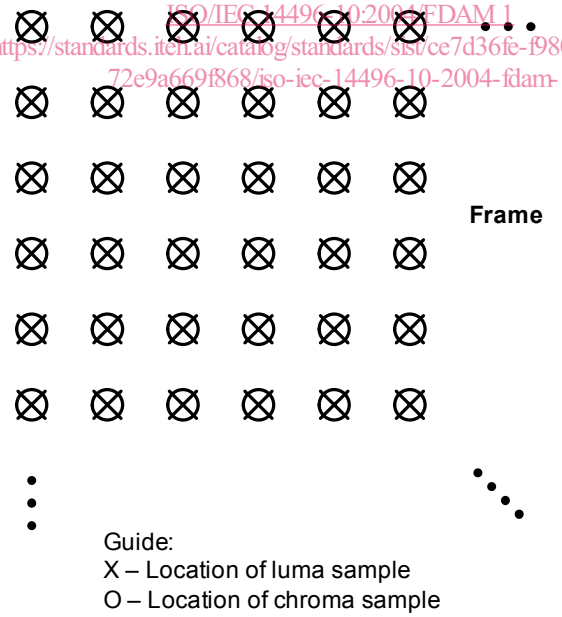

Figure 6-2c – Nominal vertical and horizontal locations of 4:4:4 luma and chroma samples in a frame

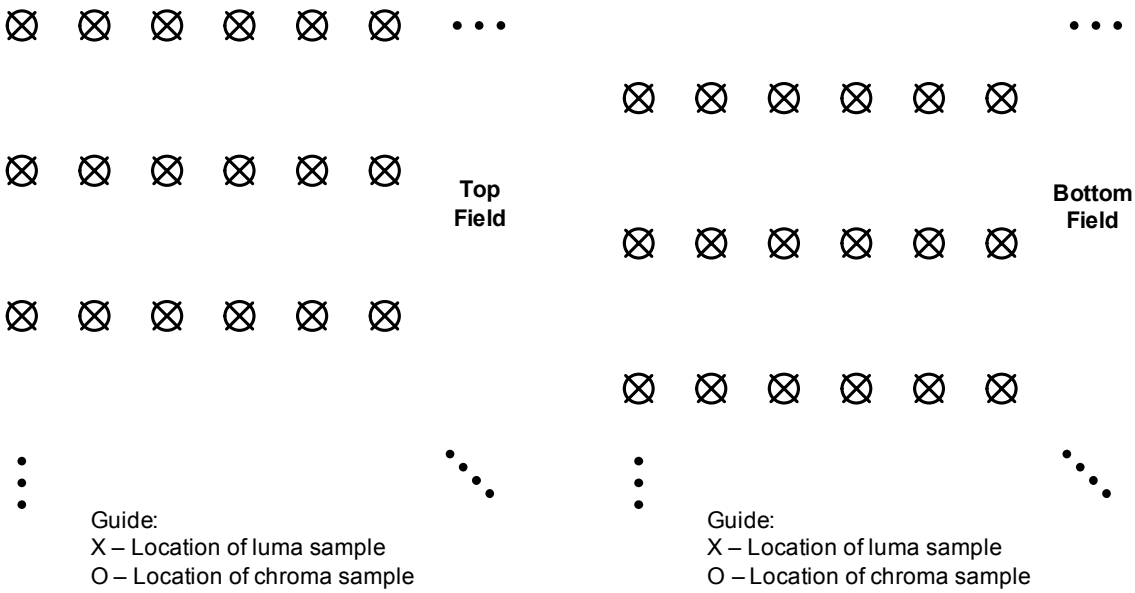

Figure 6-2d – Nominal vertical and horizontal sampling locations of 4:4:4 samples top and bottom fields

The samples are processed in units of macroblocks. The luma array for each macroblock is 16 samples in both width and height. The variables MbWidthC and MbHeightC, which specify the width and height, respectively, of the chroma arrays for each macroblock, are derived as follows.

– If chroma_format_idc is equal to 0 (monochrome), MbWidthC and MbHeightC are both equal to 0 (as no chroma arrays are specified for monochrome video).
– Otherwise, MbWidthC and MbHeightC are derived as

$$MbWidthC = 16 / SubWidthC \quad (6\text{-}0a)$$
$$MbHeightC = 16 / SubHeightC \quad (6\text{-}0b)$$

10) Subclause 6.3 "Spatial subdivision of pictures and slices"

*In subclause 6.3, replace the sentence "*Each macroblock is comprised of one 16x16 luma and two 8x8 chroma sample arrays.*" with "*Each macroblock is comprised of one 16x16 luma array and, when the video format is not monochrome, two corresponding chroma sample arrays*".*

11) New Subclause 6.4.3a "Inverse 8x8 luma block scanning process"

Insert a new subclause 6.4.3a as follows.

6.4.3a Inverse 8x8 luma block scanning process

Input to this process is the index of an 8x8 luma block luma8x8BlkIdx.

Output of this process is the location (x, y) of the upper-left luma sample for the 8x8 luma block with index luma8x8BlkIdx relative to the upper-left luma sample of the macroblock.

Figure 6-6a shows the scan for the 8x8 luma blocks.

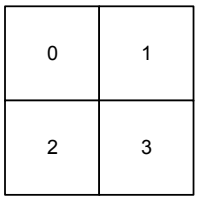

Figure 6-6a – Scan for 8x8 luma blocks

The inverse 8x8 luma block scanning process is specified by

x = InverseRasterScan(luma8x8BlkIdx, 8, 8, 16, 0) (6-16a)

y = InverseRasterScan(luma8x8BlkIdx, 8, 8, 16, 1) (6-16b)

12) Subclause 6.4.8 "Derivation process for neighbouring locations"

In subclause 6.4.8, make the following changes.

Replace the section that states as follows.

Let maxWH be a variable specifying a maximum value of the location components xN, yN, xW, and yW. maxWH is derived as follows.

- If this process is invoked for neighbouring luma locations,

  maxWH = 16 (6-23)

- Otherwise (this process is invoked for neighbouring chroma locations),

  maxWH = 8 (6-24)

Depending on the variable MbaffFrameFlag, the neighbouring luma locations are derived as follows.

- If MbaffFrameFlag is equal to 0, the specification for neighbouring luma locations in fields and non-MBAFF frames as described in subclause 6.4.8.1 is applied.
- Otherwise (MbaffFrameFlag is equal to 1), the specification for neighbouring luma locations in MBAFF frames as described in subclause 6.4.8.2 is applied.

with the following.

Let maxW and maxH be variables specifying maximum values of the location components xN, xW, and yN, yW, respectively. maxW and maxH are derived as follows.

- If this process is invoked for neighbouring luma locations,

  maxW = maxH = 16 (6-23)

- Otherwise (this process is invoked for neighbouring chroma locations),

  maxW = MbWidthC (6-24)

  maxH = MbHeightC (6-24a)

Depending on the variable MbaffFrameFlag, the neighbouring locations are derived as follows.

- If MbaffFrameFlag is equal to 0, the specification for neighbouring locations in fields and non-MBAFF frames as described in subclause 6.4.8.1 is applied.
- Otherwise (MbaffFrameFlag is equal to 1), the specification for neighbouring locations in MBAFF frames as described in subclause 6.4.8.2 is applied.

13) Subclause 6.4.8.1 "Specification for neighbouring luma locations in fields and non-MBAFF frames"

Replace the content of subclause 6.4.8.1 with the following.

6.4.8.1 Specification for neighbouring locations in fields and non-MBAFF frames

The specifications in this subclause are applied when MbaffFrameFlag is equal to 0.

The derivation process for neighbouring macroblock addresses and their availability in subclause 6.4.5 is invoked with mbAddrA, mbAddrB, mbAddrC, and mbAddrD as well as their availability status as the output.

Table 6-3 specifies mbAddrN depending on (xN, yN).

Table 6-3 – Specification of mbAddrN

xN	yN	mbAddrN
< 0	< 0	mbAddrD
< 0	0 .. maxH - 1	mbAddrA
0 .. maxW - 1	< 0	mbAddrB
0 .. maxW - 1	0 .. maxH - 1	CurrMbAddr
> maxW - 1	< 0	mbAddrC
> maxW - 1	0 .. maxH - 1	not available
	> maxH - 1	not available

The neighbouring location (xW, yW) relative to the upper-left corner of the macroblock mbAddrN is derived as

$$xW = (xN + maxW) \% maxW \quad (6\text{-}25)$$

$$yW = (yN + maxH) \% maxH \quad (6\text{-}26)$$

14) Subclause 6.4.8.2 "Specification for neighbouring luma locations in MBAFF frames"

In subclause 6.4.8.2, make the following changes.

Replace the title of the subclause as follows.

6.4.8.2 Specification for neighbouring locations in MBAFF frames

Replace Table 6-4 by the following table.