



(86) Date de dépôt PCT/PCT Filing Date: 2013/01/25  
(87) Date publication PCT/PCT Publication Date: 2013/08/01  
(45) Date de délivrance/Issue Date: 2021/04/20  
(85) Entrée phase nationale/National Entry: 2014/06/25  
(86) N° demande PCT/PCT Application No.: JP 2013/000397  
(87) N° publication PCT/PCT Publication No.: 2013/111605  
(30) Priorité/Priority: 2012/01/25 (US13/358,414)

(51) Cl.Int./Int.Cl. *H04N 19/573* (2014.01),  
*H04N 19/159* (2014.01), *H04N 19/46* (2014.01),  
*H04N 19/58* (2014.01), *H04N 19/70* (2014.01)  
(72) Inventeurs/Inventors:  
DESHPANDE, SACHIN G., US;  
SEGALL, CHRISTOPHER A., US  
(73) Propriétaire/Owner:  
VELOS MEDIA INTERNATIONAL LIMITED, IE  
(74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : PROCEDES DE DECODAGE VIDEO ET PROCEDES DE CODAGE VIDEO  
(54) Title: VIDEO DECODING METHODS AND VIDEO ENCODING METHODS

(57) **Abrégé/Abstract:**

A method for decoding video bitstream is provided. the method includes: receiving a reference picture set parameter from said video bitstream; decoding a current picture using inter prediction based on a reference picture set; and storing said decoded picture to be referred for future inter prediction, wherein said reference picture set is decoded by using at least: (a) one or more reference picture identifiers each of which being based on a selected number of least significant bits (LSB) of the picture order count (POC) for a reference picture; and (b) a signal to specify whether or not subsequent data to determine the MSB of said POC exists.

## (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization  
International Bureau



(10) International Publication Number  
**WO 2013/111605 A1**

(43) International Publication Date  
1 August 2013 (01.08.2013)

(51) International Patent Classification:  
*H04N 7/32* (2006.01)

(21) International Application Number:  
PCT/JP2013/000397

(22) International Filing Date:  
25 January 2013 (25.01.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
13/358,414 25 January 2012 (25.01.2012) US

(71) Applicant: SHARP KABUSHIKI KAISHA [JP/JP]; 22-22, Nagaike-cho, Abeno-ku, Osaka-shi, Osaka, 5458522 (JP).

(72) Inventors: DESHPANDE, Sachin G.. SEGALL, Christopher A..

(74) Agent: HARAKENZO WORLD PATENT & TRADE-MARK; Daiwa Minamimorimachi Building, 2-6, Tenjin-bashi 2-chome Kita, Kita-ku, Osaka-shi, Osaka, 5300041 (JP).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— with international search report (Art. 21(3))

(54) Title: VIDEO DECODING METHODS AND VIDEO ENCODING METHODS

(57) Abstract: A method for decoding video bitstream is provided. the method includes: receiving a reference picture set parameter from said video bitstream; decoding a current picture using inter prediction based on a reference picture set; and storing said decoded picture to be referred for future inter prediction, wherein said reference picture set is decoded by using at least: (a) one or more reference picture identifiers each of which being based on a selected number of least significant bits (LSB) of the picture order count (POC) for a reference picture; and (b) a signal to specify whether or not subsequent data to determine the MSB of said POC exists.



WO 2013/111605 A1

## VIDEO DECODING METHODS AND VIDEO ENCODING METHODS

### FIELD OF THE INVENTION

The present invention relates to video encoding and/or decoding.

### BACKGROUND OF THE INVENTION

Digital video is typically represented as a series of images or frames, each of which contains an array of pixels. Each pixel includes information, such as intensity and/or color information. In many cases, each pixel is represented as a set of three colors, each of which may be defined by eight bit color values.

Video coding techniques, for example H.264/MPEG-4 AVC (H.264/AVC), typically provide higher coding efficiency at the expense of increasing complexity. Increasing image quality requirements and increasing image resolution requirements for video coding techniques also increase the coding complexity. Video decoders that are suitable for parallel decoding may improve the speed of the decoding process and reduce memory requirements; video encoders that are suitable for parallel encoding may improve the speed of the encoding process and reduce memory requirements.

H.264/MPEG-4 AVC [Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, "H.264: Advanced video coding for generic audiovisual services," ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG4-Part 10), November 2007], and similarly the JCT-VC, ["Draft Test Model Under Consideration", JCTVC-A205, JCT-VC Meeting, Dresden, April 2010 (JCTVC)],

are video codec (encoder/decoder) specifications that decode pictures based upon reference pictures in a video sequence for compression efficiency.

### SUMMARY OF THE INVENTION

It is desired to provide more efficient techniques for decoding pictures based upon reference pictures.

According to the present invention, there is provided a method for decoding video bitstream comprising decoding a current picture using inter prediction based on a reference picture set; and storing said decoded picture to be referred for inter prediction, wherein said reference picture set is decoded by using at least (a) one or more reference picture identifiers each of which being based on a selected number of least significant bits (LSB) of the picture order count (POC) for a reference picture; and (b) a signal to specify whether or not data exists wherein the data is used to determine a value of picture order count of a long term reference picture in a long term reference picture set of the current picture.

- 2 -

According to the present invention, there is provided a method for encoding video bitstream comprising encoding a current picture using inter prediction based on a reference picture set, wherein said reference picture set is encoded by using at least (a) one or more reference picture identifiers each of which being based on a selected number of least significant bits (LSB) of the picture order count (POC) for a reference picture; and (b) a signal to specify whether or not data exists wherein the data is used to determine a value of picture order count of a long term reference picture in a long term reference picture set of the current picture.

The foregoing and other features, and advantages of the invention will be more readily understood upon consideration of the following detailed description of the invention, taken in conjunction with the accompanying drawings.

#### BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates a H.264/AVC video encoder.

FIG. 2 illustrates a H.264/AVC video decoder.

FIG. 3 illustrates an exemplary slice structure.

FIG. 4 illustrates another exemplary slice structure.

FIG. 5 illustrates reconstruction of an entropy slice.

FIG. 6 illustrates reconstruction of an portion of the entropy slice of FIG. 5.

FIG. 7 illustrates reconstruction of an entropy slice with an omitted LSB count value.

FIG. 8 illustrates reconstruction of an entropy slice with a long term picture value.

FIG. 9 illustrates reconstruction of an entropy slice by selecting a first preceding frame with a long term picture value.

FIG. 10 illustrates reconstruction of an entropy slice by using duplicate long term picture frame having the same least significant bit count value.

FIG. 11A illustrates a technique for selecting a reference frame.

FIG. 11B illustrates a technique for selecting a reference frame.

FIG. 12 illustrates another technique for selecting a reference frame.

FIG. 13A illustrates another technique for selecting a reference frame.

FIG. 13B illustrates another technique for selecting a reference frame.

FIG. 14 illustrates another technique for selecting a reference frame.

#### DESCRIPTION OF THE PREFERRED EMBODIMENTS

While any video coder/decoder (codec) that uses encoding/decoding may be

- 3 -

accommodated by embodiments described herein, exemplary embodiments are described in relation to an H.264/AVC encoder and an H.264/AVC decoder merely for purposes of illustration. Many video coding techniques are based on a block-based hybrid video-coding approach, wherein the source-coding technique is a hybrid of inter-picture, also considered inter-frame, prediction, intra-picture, also considered intra-frame, prediction and transform coding of a prediction residual. Inter-frame prediction may exploit temporal redundancies, and intra-frame and transform coding of the prediction residual may exploit spatial redundancies.

FIG. 1 is a block diagram illustrating an exemplary encoder 104 for an electronic device 102. It should be noted that one or more of the elements illustrated as included within the electronic device 102 may be implemented in hardware, and/or software. For example, the electronic device 102 includes an encoder 104, which may be implemented in hardware and/or software.

The electronic device 102 may include a supplier 134. The supplier 134 may provide picture or image data (e.g., video) as a source 106 to the encoder 104. Non limiting examples of the supplier 134 include image sensors, memory, communication interfaces, network interfaces, wireless receivers, ports, video frame content, previously encoded video content, non-encoded video content, etc.

The source 106 may be provided to an intra-frame prediction module and reconstruction buffer 140. The source 106 may also be provided to a motion estimation and motion compensation module 166 and to a subtraction module 146.

The intra-frame prediction module and reconstruction buffer 140 may generate intra mode information 158 and an intra signal 142 based on the source 106 and reconstructed data 180. The motion estimation and motion compensation module 166 may generate inter mode information 168 and an inter signal 144 based on the source 106 and a reference picture buffer 196 signal 198.

The reference picture buffer 196 signal 198 may include data from one or more reference pictures stored in the reference picture buffer 196. The reference picture buffer 196 may also include an RPS index initializer module 108. The initializer module 108 may process reference pictures corresponding to the buffering and list construction of an RPS.

The encoder 104 may select between the intra signal 142 and the inter signal 144 in accordance with a mode. The intra signal 142 may be used in order to exploit spatial characteristics within a picture in an intra coding mode. The inter signal 144 may be used in order to exploit temporal characteristics between pictures in an inter coding mode. While in the intra coding mode, the intra signal 142 may be provided to

the subtraction module 146 and the intra mode information 158 may be provided to an entropy coding module 160. While in the inter coding mode, the inter signal 144 may be provided to the subtraction module 146 and the inter mode information 168 may be provided to the entropy coding module 160.

Either the intra signal 142 or the inter signal 144 (depending on the mode) is subtracted from the source 106 at the subtraction module 146 in order to produce a prediction residual 148. The prediction residual 148 is provided to a transformation module 150. The transformation module 150 may compress the prediction residual 148 to produce a transformed signal 152 that is provided to a quantization module 154. The quantization module 154 quantizes the transformed signal 152 to produce transformed and quantized coefficients (TQCs) 156.

The TQCs 156 are provided to an entropy coding module 160 and an inverse quantization module 170. The inverse quantization module 170 performs inverse quantization on the TQCs 156 to produce an inverse quantized signal 172 that is provided to an inverse transformation module 174. The inverse transformation module 174 decompresses the inverse quantized signal 172 to produce a decompressed signal 176 that is provided to a reconstruction module 178.

The reconstruction module 178 may produce reconstructed data 180 based on the decompressed signal 176. For example, the reconstruction module 178 may reconstruct (modified) pictures. The reconstructed data 180 may be provided to a deblocking filter 182 and to the intra-frame prediction module and reconstruction buffer 140. The deblocking filter 182 may produce a filtered signal 184 based on the reconstructed data 180.

The filtered signal 184 may be provided to a sample adaptive offset (SAO) module 186. The SAO module 186 may produce SAO information 188 that is provided to the entropy coding module 160 and an SAO signal 190 that is provided to an adaptive loop filter (ALF) 192. The ALF 192 produces an ALF signal 194 that is provided to the reference picture buffer 196. The ALF signal 194 may include data from one or more pictures that may be used as reference pictures.

The entropy coding module 160 may code the TQCs 156 to produce a bitstream 114. Also, the entropy coding module 160 may code the TQCs 156 using Context-Adaptive Variable Length Coding (CAVLC) or Context-Adaptive Binary Arithmetic Coding (CABAC). In particular, the entropy coding module 160 may code the TQCs 156 based on one or more of intra mode information 158, inter mode information 168 and SAO information 188. The bitstream 114 may include coded picture data. The encoder often encodes a frame as a sequence of blocks, generally

referred to as macroblocks.

Quantization, involved in video compression such as HEVC, is a lossy compression technique achieved by compressing a range of values to a single value. The quantization parameter (QP) is a predefined scaling parameter used to perform the quantization based on both the quality of reconstructed video and compression ratio. The block type is defined in HEVC to represent the characteristics of a given block based on the block size and its color information. QP, resolution information and block type may be determined before entropy coding. For example, the electronic device 102 (e.g., the encoder 104) may determine the QP, resolution information and block type, which may be provided to the entropy coding module 160.

The entropy coding module 160 may determine the block size based on a block of TQCs 156. For example, block size may be the number of TQCs 156 along one dimension of the block of TQCs. In other words, the number of TQCs 156 in the block of TQCs may be equal to block size squared. For instance, block size may be determined as the square root of the number of TQCs 156 in the block of TQCs. Resolution may be defined as a pixel width by a pixel height. Resolution information may include a number of pixels for the width of a picture, for the height of a picture or both. Block size may be defined as the number of TQCs 156 along one dimension of a 2D block of TQCs.

In some configurations, the bitstream 114 may be transmitted to another electronic device. For example, the bitstream 114 may be provided to a communication interface, network interface, wireless transmitter, port, etc. For instance, the bitstream 114 may be transmitted to another electronic device via LAN, the Internet, a cellular phone base station, etc. The bitstream 114 may additionally or alternatively be stored in memory on the electronic device 102 or other electronic device.

FIG. 2 is a block diagram illustrating an exemplary decoder 212 on an electronic device 202. The decoder 212 may be included for an electronic device 202. For example, the decoder 212 may be a HEVC decoder. The decoder 212 and/or one or more of the elements illustrated as included in the decoder 212 may be implemented in hardware and/or software. The decoder 212 may receive a bitstream 214 (e.g., one or more encoded pictures included in the bitstream 214) for decoding. In some configurations, the received bitstream 214 may include received overhead information, such as a received slice header, received PPS (or picture parameter set), received buffer description information, etc. The encoded pictures included in the bitstream 214 may include one or more encoded reference pictures and/or one or more other encoded pictures.

Received symbols (in the one or more encoded pictures included in the bitstream 214) may be entropy decoded by an entropy decoding module 268, thereby producing a motion information signal 270 and quantized, scaled and/or transformed coefficients 272.

The motion information signal 270 may be combined with a portion of a reference frame signal 298 from a frame memory 278 at a motion compensation module 274, which may produce an inter-frame prediction signal 282. The quantized, scaled and/or transformed coefficients 272 may be inverse quantized, inverse scaled and inverse transformed by an inverse module 262, thereby producing a decoded residual signal 284. The decoded residual signal 284 may be added to a prediction signal 292 to produce a combined signal 286. The prediction signal 292 may be a signal selected from either the inter-frame prediction signal 282 or an intra-frame prediction signal 290 produced by an intra-frame prediction module 288. In some configurations, this signal selection may be based on (e.g., controlled by) the bitstream 214.

The intra-frame prediction signal 290 may be predicted from previously decoded information from the combined signal 292 (in the current frame, for example). The combined signal 286 may also be filtered by a de-blocking filter 294. The resulting filtered signal 296 may be written to frame memory 278. The resulting filtered signal 296 may include a decoded picture.

The frame memory 278 may include a DPB (or display picture buffer) as described herein. The DPB may include one or more decoded pictures that may be maintained as short or long term reference frames. The frame memory 278 may also include overhead information corresponding to the decoded pictures. For example, the frame memory 278 may include slice headers, PPS information, , buffer description information, etc. One or more of these pieces of information may be signaled from an encoder (e.g., encoder 104). The frame memory 278 may provide a decoded picture 218.

An input picture comprising a plurality of macroblocks may be partitioned into one or several slices. The values of the samples in the area of the picture that a slice represents may be properly decoded without the use of data from other slices provided that the reference pictures used at the encoder and the decoder are the same and that de-blocking filtering does not use information across slice boundaries. Therefore, entropy decoding and macroblock reconstruction for a slice does not depend on other slices. In particular, the entropy coding state may be reset at the start of each slice. The data in other slices may be marked as unavailable when defining neighborhood availability for both entropy decoding and reconstruction. The slices may be entropy decoded and reconstructed in parallel. No intra prediction and motion-vector



- 7 -

prediction is preferably allowed across the boundary of a slice. In contrast, de-blocking filtering may use information across slice boundaries.

FIG. 3 illustrates an exemplary video picture 90 comprising eleven macroblocks in the horizontal direction and nine macroblocks in the vertical direction (nine exemplary macroblocks labeled 91-99). FIG. 3 illustrates three exemplary slices: a first slice denoted "SLICE #0" 89, a second slice denoted "SLICE #1" 88 and a third slice denoted "SLICE #2" 87. An H.264/AVC decoder may decode and reconstruct the three slices 87, 88, 89 in parallel. Each of the slices may be transmitted in scan line order in a sequential manner. At the beginning of the decoding/reconstruction process for each slice, entropy decoding 268 is initialized or reset and macroblocks in other slices are marked as unavailable for both entropy decoding and macroblock reconstruction. Thus, for a macroblock, for example, the macroblock labeled 93, in "SLICE #1", macroblocks (for example, macroblocks labeled 91 and 92) in "SLICE #0" may not be used for entropy decoding or reconstruction. Whereas, for a macroblock, for example, the macroblock labeled 95, in "SLICE #1", other macroblocks (for example, macroblocks labeled 93 and 94) in "SLICE #1" may be used for entropy decoding or reconstruction. Therefore, entropy decoding and macroblock reconstruction proceeds serially within a slice. Unless slices are defined using a flexible macroblock ordering (FMO), macroblocks within a slice are processed in the order of a raster scan.

Flexible macroblock ordering defines a slice group to modify how a picture is partitioned into slices. The macroblocks in a slice group are defined by a macroblock-to-slice-group map, which is signaled by the content of the picture parameter set and additional information in the slice headers. The macroblock-to-slice-group map consists of a slice-group identification number for each macroblock in the picture. The slice-group identification number specifies to which slice group the associated macroblock belongs. Each slice group may be partitioned into one or more slices, wherein a slice is a sequence of macroblocks within the same slice group that is processed in the order of a raster scan within the set of macroblocks of a particular slice group. Entropy decoding and macroblock reconstruction proceeds serially within a slice group.

FIG. 4 depicts an exemplary macroblock allocation into three slice groups: a first slice group denoted "SLICE GROUP #0" 86, a second slice group denoted "SLICE GROUP #1" 85 and a third slice group denoted "SLICE GROUP #2" 84. These slice groups 84, 85, 86 may be associated with two foreground regions and a background region, respectively, in the picture 90.

A picture may be partitioned into one or more slices, wherein a slice may be

- 8 -

self-contained in the respect that values of the samples in the area of the picture that the slice represents may be correctly reconstructed without use of data from other slices, provided that the reference pictures used are identical at the encoder and the decoder. All reconstructed macroblocks within a slice may be available in the neighborhood definition for reconstruction.

A slice may be partitioned into more than one entropy slice, wherein an entropy slice may be self-contained in the respect that the area of the picture that the entropy slice represents may be correctly entropy decoded without the use of data from other entropy slices. The entropy decoding 268 may be reset at the decoding start of each entropy slice. The data in other entropy slices may be marked as unavailable when defining neighborhood availability for entropy decoding.

A device configured for decoding pictures obtains or otherwise receives a bitstream that includes a series of pictures, including a current picture. The device further obtains a reference picture set (RPS) parameter that may be used for the identification of other frames that may be used for the decoding of the current picture or for the decoding of pictures subsequent to the current picture in the order that pictures are signaled in the bitstream.

A RPS provides an identification of a set of reference pictures associated with the current frame. A RPS may identify reference pictures that are prior to the current picture in display order that may be used for inter prediction of the current picture and/or identify reference pictures that are after the current picture in display order that may be used for inter prediction of the current picture. For example, if the system receives frame 1, 3, 5 and 5 uses 3 for reference, and, an encoder uses frame 1 for the prediction of frame 7. Then, the RPS for 5 may signal to keep both frame 3 and 1 in the frame memory 278 even though frame 1 is not used for reference of frame 5. In one embodiment, the RPS for 5 may be [-2 -4]. Additionally, the frame memory 278 may be referred to the display picture buffer, or equivalently DPB. For this example, the frame number corresponds to the display order, or output order, of the frames.

A RPS describes one or more reference pictures that should be maintained, at least for a limited time duration, in the decoded picture buffer (DPB) for subsequent use. This identification of the RPS may be included in the slice header of each picture, together with a picture, and/or together with a group of pictures. In one embodiment, a list of RPS may be sent in a picture parameter set (PPS). Then, the slice header may identify one of the RPS sent in the PPS to be used for the slice. For example, a RPS for a group of pictures may be signaled in a picture parameter set (PPS). Any pictures in the DPB that are not a part of the RPS for the current frame may be marked as "unused

for reference".

A DPB may be used to store reconstructed (e.g., decoded) pictures at the decoder. These stored pictures may then be used, for example, in an inter-prediction technique. Also, a picture in the DPB may be associated with a picture order count (POC). The POC may be a variable that is associated with each encoded picture and that has a value that increases with increasing picture position in an output order. In other words, the POC may be used by the decoder to deliver the pictures in the correct order for display. The POC may also be used for identification of reference pictures during construction of a reference picture list and identification of decoded reference pictures. Furthermore, the POC may be used for identification of pictures that are lost during transmission from an encoder to a decoder.

Referring to FIG. 5, one example of a set of frames 300 provided from an encoder to a decoder is illustrated. Each of the frames may have an associated POC 310. As illustrated, the POC may increment from a minus number through a large positive number. In some embodiments, the POC may only increment from zero through a larger positive number. The POC is typically incremented by one for each frame, but in some cases one or more POC are skipped or otherwise omitted. For example, the POC for a set of frames in the encoder may be, 0, 1, 2, 3, 4, 5, etc. For example, the POC for the same or another set of frames in the encoder may be, 0, 1, 2, 4, 5, etc., with POC 3 being skipped or otherwise omitted.

As the POC becomes sufficiently large, a significant number of bits would be necessary to identify each frame using the POC. The encoder may reduce the number of bits used to identify a particular POC by using a selected number of least significant bits (LSB) of the POC to identify each frame, such as 4 bits. Since the reference frames used for decoding the current frame are often temporally located proximate to the current frame, this identification technique is suitable and results in a reduction in the computational complexity of the system and an overall reduction in the bit rate of the video. The number of LSB to use to identify the pictures may be signaled in the bit stream to the decoder.

As illustrated, with LSB being 4 bits, the LSB index repeats every 16 values ( $2^4$ ) when the selected number of LSB of the POC is 4. Thus, frame 0 has a LSB having a value of 0, frame 1 has a LSB having a value of 1, ..., frame 14 has a LSB having a value of 14, frame 15 has a LSB having a value of 15. However, frame 16 again has a LSB having a value of 0, frame 17 again has a LSB having a value of 1, and frame 20 has a LSB having a value of 4. The LSB identifier (generally also referred to as the LSB of the POC or, equivalently, POC LSB) may have the characteristic of  $LSB =$

POC % 16, where % is the remainder after dividing by 16 ( $2^{\text{number of least significant bits}}$  which in this case is 4). Similarly, if the selected number of LSBs to identify a POC is N bits, the LSB identifier may have the characteristic of  $\text{LSB} = \text{POC} \% (2^N)$  where  $2^N$  denotes 2 raised to the power of N. Rather than including the POC the bitstream to identify frames, the encoder preferably provides the LSB index (generally also referred to as the LSB of the POC or, equivalently, POC LSB), in the bitstream to the decoder.

The reference frames used for inter prediction of a current frame, or frames subsequent to the current frame, may be identified with an RPS using either relative (e.g., delta) referencing (using the difference between POC values, or alternatively a deltaPOC and a currentPOC, for example) or absolute referencing (using the POC, for example). In some embodiments, frames identified with relative referencing may be called a short term reference frame, and frames identified with an absolute referencing may be called a long term reference frame. For example, the frame identified by POC 5 310 and signaled to the decoder as LSB 5 320 in the bitstream may have an associated RPS 330 of [-5, -2, -1]. The meaning of the RPS values is described later.

Referring to FIG. 6, illustrating a portion of FIG. 5, the RPS of [-5, -2, -1] refers to frames that include the fifth previous frame 320, second previous frame 321, and first previous frame 322 relative to the current frame. This in turn refers to the POC values of 0, 3, and 4, respectively as illustrated in FIG. 6 for the current frame with POC value of 5. Typically, the RPS refers to the difference in between the POC value of the current frame and the POC value of the previous frame. For example, the RPS of [-5, -2, -1] for a current frame having a POC value of 5, refers to frames having POC values of  $5 \text{ minus } 5 = 0$ ;  $5 \text{ minus } 2 = 3$ ; and  $5 \text{ minus } 1 = 4$ . The RPS can also include frames in the future. These may be indicated with positive values in the RPS (positive deltaPOC values).

In the case that the POC values are not sequential, such as in the case that one or more POC values are skipped or otherwise omitted in parts of the bitstream, the difference between the POC value of the current frame and POC value of the previous frame may be different than the number of frames output between the previous frame and current frame such as illustrated in FIG. 7. As shown in FIG. 7, the RPS of [-5, -2, -1] refers to frames that include the fifth previous frame 320, second previous frame 321, and first previous frame 322 relative to the POC of the frame identified with POC value equal to 5. The RPS may be signaled in the bitstream in any suitable manner, such as provided together with the frame or provided together with a set of frames.

Referring to FIG. 8, another technique for signaling the reference frames is to

use an absolute reference, generally referred to as a long term picture, in the RPS associated with a frame. The decoding process, such as the motion vector prediction technique, may be different depending if the reference frame is signaled using an absolute reference or a relative reference. The absolute reference (referred to as LT for convenience) refers to a particular LSB count value associated with a reference frame, such as a previous or subsequent frame. For example, the absolute reference of  $LT = 3$  (LT3) would refer to a reference frame having a POC LSB value of 3. Accordingly, a RPS of [LT3, -5] would refer to a reference frame having POC LSB value of 3 and a reference frame with a POC equal to the POC of the current frame minus 5. In FIG. 8, this corresponds to the reference frame with POC equal to 3 444 and the reference frame with POC equal to 0 320. Typically, the LT3 refers to the first previous frame relative to the current frame having a POC LSB value of 3. In one embodiment, LT3 refers to the first previous frame relative to the current frame in output order having a POC LSB value of 3. In a second embodiment, LT3 refers to the first previous frame relative to the current frame in transmission order having a POC LSB value of 3. While such a system is suitable for many bit streams, it is not sufficiently robust to select a frame with a LSB count value of 3 that is different than the immediately previous frame having a LSB count value of 3.

Referring to FIG. 9, for example, if the encoder was encoding frame 31 (POC=31) and the system signals the use of the long term picture with POC LSB=0 (LT0), then this would refer A to frame 16 (POC=16) since it is the first previous frame with LSB=0. However, the encoder may desire to signal the long term picture frame 0, which likewise has a POC LSB count value of 0, but this may not be accomplished with such a first previous referencing scheme. To overcome this limitation, one technique is to increase the number of least significant bits used to signal the long term frame POC LSB. While such an increase in the number of least significant bits is possible, it results in substantial additional bits being added to the bitstream.

A more preferred technique that results in fewer additional bits being added to the bitstream is to signal a different long term picture than the first immediately preceding frame with a corresponding POC LSB value. For example, the system could indicate the RPS of the current frame having an absolute reference as [LT0|2] where the 0 refers to the POC LSB value and 2 refers to which of the previous frames with POC LSB value equal to 0 to use, which in this case would be the second previous POC LSB value of 0 (e.g., frame 0 in FIG. 9). If no second reference is included then the system may default to the immediately preceding frame with a POC LSB=0 [LT0] (e.g., frame 16 in FIG. 9).

In many cases, the frequency of occurrence of the desire to signal a frame that is not the first immediately preceding frame with the corresponding POC LSB value using absolute referencing will be relatively infrequent. To further reduce the overall bit rate indicating which frame to use, while permitting the capability of signaling a different frame than the first immediately preceding frame with the corresponding POC LSB value using absolute referencing, the system may use a duplication technique. For example, the RPS may be structured as follows, [LT0, LT0|3]. The duplication of the LT0 within the same RPS signals the decoder to use a different frame having a POC LSB value of 0, which in this case would be the third previous occurrence of the POC LSB value of 0. In general, aside from the potential that a particular POC LSB value would not be included in a particular cycle of POC LSB values, the desired POC LSB value will correspond to a frame of the indicated previous occurrence. Here, a cycle of POC LSB values denotes a set of frames that when ordered in output order do not contain the same POC LSB value and are not separated in output order by frames not in the set.

Referring to FIG. 10, the duplication technique may be indicated as follows. The RPS includes a signal of a long term picture having a POC LSB value 400 (e.g., [LT3]). The same RPS includes another signal of a long term picture having the same POC LSB value 410 (e.g., [LT3, LT3]). The same RPS includes another signal of the second long term picture having the same LSB count value 410 indicating the location of the desired frame 420 [LT3, LT3|2].

The signaling of the location of the desired frame may be performed in any suitable manner. Referring to FIGS. 11A-11B for example, the location may be one or more previous cycles of the POC LSB values for the desired frame relative to the current frame, such as the third previous cycle. Referring to FIG. 12 for example, the location may be based upon an absolute number of frames offset from the current frame. Referring to FIGS. 13A-13B for example, the location may be one or more previous cycles of the POC LSB values relative to the first immediately preceding frame with the desired POC LSB value. Referring to FIG. 14 for example, the location may be based upon an absolute number of frames offset relative to the first immediately preceding frame with the desired POC LSB value.

One exemplary implementation of such a technique may use the following syntax.

[Table 1]

slice_header() {	Descriptor
<b>lightweight_slice_flag</b>	u(1)
if( !lightweight_slice_flag ) {	
<b>slice_type</b>	ue(v)
<b>pic_parameter_set_id</b>	ue(v)
if( !idrPicFlag ) {	
<b>idr_pic_id</b>	ue(v)
<b>no_output_of_prior_pics_flag</b>	u(1)
}	
else {	
<b>pic_order_cnt_lsb</b>	u(v)
<b>short_term_ref_pic_set_pps_flag</b>	u(1)
if( !short_term_ref_pic_set_pps_flag )	
short_term_ref_pic_set( num_short_term_ref_pic_sets )	
else	
<b>short_term_ref_pic_set_idx</b>	u(v)
if( long_term_ref_pics_present_flag ) {	
<b>num_long_term_pics</b>	ue(v)
for( i = 0; i < num_long_term_pics; i++ ) {	
<b>delta_poc_lsb_lt_minus1[ i ]</b>	ue(v)
if( deltaPOCLSBCheck(i)==1 ) {	
<b>delta_poc_msb_lt_minus1[ i ]</b>	ue(v)
}	
<b>used_by_curr_pic_lt_flag[ i ]</b>	u(1)
}	
}	
}	
if( slice_type == P    slice_type == B ) {	
<b>num_ref_idx_active_override_flag</b>	u(1)
if( num_ref_idx_active_override_flag ) {	
<b>num_ref_idx_l0_active_minus1</b>	ue(v)
if( slice_type == B )	
<b>num_ref_idx_l1_active_minus1</b>	ue(v)
}	
}	
}	
...	
}	

When the 'lightweight\_slice\_flag' is equal to 1 specifies that the value of slice header syntax elements not present shall be inferred to be equal to the value of slice header syntax elements in a proceeding slice, where a proceeding slice is defined as the slice containing treeblock with location (LCUAddress - 1). The lightweight\_slice\_flag shall be equal to 0 when LCUAddress equal to 0. Here, a treeblock may be a macroblock and LCUAddress denotes the spatial location of the treeblock within a picture.

The 'slice\_type' specifies the coding type of the slice as follows:

[Table 2]

slice_type	Name of slice_type
0	P (P slice)
1	B (B slice)
2	I (I slice)

When nal\_unit\_type is equal to 5 (IDR picture), slice\_type shall be equal to 2. When max\_num\_ref\_frames is equal to 0, slice\_type shall be equal to 2.

'pic\_parameter\_set\_id' specifies the picture parameter set in use. The value of pic\_parameter\_set\_id shall be in the range of 0 to 255, inclusive.

'idr\_pic\_id' identifies an IDR picture, which denotes a picture that does not use previously transmitted pictures for reference. The values of idr\_pic\_id in all the slices of an IDR picture shall remain unchanged. When two consecutive access units in decoding order are both IDR access units, the value of idr\_pic\_id in the slices of the first such IDR access unit shall differ from the idr\_pic\_id in the second such IDR access unit. The value of idr\_pic\_id shall be in the range of 0 to 65535, inclusive.

'no\_output\_of\_prior\_pics\_flag' specifies how the previously decoded pictures in the decoded picture buffer are treated after decoding of an IDR picture. When the IDR picture is the first IDR picture in the bitstream, the value of no\_output\_of\_prior\_pics\_flag has no effect on the decoding process. When the IDR picture is not the first IDR picture in the bitstream and the value of pic\_width\_in\_luma\_samples or pic\_height\_in\_luma\_samples, which denote the dimensions of the pictures, or max\_dec\_frame\_buffering, which denotes the maximum amount of reordering required at a decoder to convert a sequence of frames in transmission order to a sequence of frames in display order, derived from the active sequence parameter set is different from the value of pic\_width\_in\_luma\_samples or



pic\_height\_in\_luma\_samples or max\_dec\_frame\_buffering derived from the sequence parameter set active for the preceding picture, no\_output\_of\_prior\_pics\_flag equal to 1 may (but should not) be inferred by the decoder, regardless of the actual value of no\_output\_of\_prior\_pics\_flag.

'pic\_order\_cnt\_lsb' specifies the picture order count modulo MaxPicOrderCntLsb for the current picture. The length of the pic\_order\_cnt\_lsb syntax element is  $\log_2(\text{max\_pic\_order\_cnt\_lsb\_minus4} + 4)$  bits. The value of the pic\_order\_cnt\_lsb shall be in the range of 0 to MaxPicOrderCntLsb - 1, inclusive. When pic\_order\_cnt\_lsb is not present, pic\_order\_cnt\_lsb shall be inferred to be equal to 0. Here, pic\_order\_cnt\_lsb indicates the number of LSBs in POC LSB.

'short\_term\_ref\_pic\_set\_pps\_flag' equal to 1 specifies that the short-term reference picture set of the current picture shall be created using syntax elements in the active picture parameter set, which contains syntax elements that may be shared between multiple pictures. short\_term\_ref\_pic\_set\_pps\_flag equal to 0 specifies that the short-term reference picture set of the current picture shall be created using syntax elements in the short\_term\_ref\_pic\_set() syntax structure in the slice header. In some embodiments, a short-term reference picture set denotes a pictures set that only uses delta referencing.

'short\_term\_ref\_pic\_set\_idx' specifies the index to the list of the short-term reference picture sets specified in the active picture parameter set that shall be used for creation of the reference picture set of the current picture. The syntax element short\_term\_ref\_pic\_set\_idx shall be represented by  $\text{ceil}(\log_2(\text{num\_short\_term\_ref\_pic\_sets}))$  bits. The value of short\_term\_ref\_pic\_set\_idx shall be in the range of 0 to num\_short\_term\_ref\_pic\_sets - 1, inclusive, where num\_short\_term\_ref\_pic\_sets is the syntax element from the active picture parameter set.

The variable StRpsIdx is derived as follows.

[Table 3]

<pre> If( short_term_ref_pic_set_pps_flag )     StRpsIdx = short_term_ref_pic_set_idx ELSE     StRpsIdx = num_short_term_ref_pic_sets </pre>
--

- 16 -

'num\_long\_term\_pics' specifies the number of the long-term reference pictures that are to be included in the long-term reference picture set of the current picture. The value of num\_long\_term\_pics shall be in the range of 0 to max\_num\_ref\_frames - NumNegativePics[ StRpsIdx ] - NumPositivePics[ StRpsIdx ], inclusive. When not present, the value of num\_long\_term\_pics shall be inferred to be equal to 0. In some embodiments, the long-term reference pictures denote reference pictures that are transmitted with absolute referencing.

'delta\_poc\_lsb\_lt\_minus1[ i ]' is used to determine the value of the least significant bits of the picture order count value of the i-th long-term reference picture that is included in the long-term reference picture set of the current picture. delta\_poc\_lsb\_lt\_minus1[ i ] shall be in the range of 0 to MaxPicOrderCntLsb - 1, inclusive. In some embodiments, delta\_poc\_lsb\_lt\_minus1[i] denotes POC LSB of the i-th long-term reference picture.

The variable DeltaPocLt[ i ] is derived as follows.

[Table 4]

```

If (i == 0)
    DeltaPocLt[ i ] = delta_poc_lsb_lt_minus1[ i ] + 1
Else
    DeltaPocLt[ i ] = delta_poc_lsb_lt_minus1[ i ] + 1 + DeltaPocLt[ i - 1 ]

```

The value of DeltaPocLt[ i ] shall be in the range of 0 to MaxPicOrderCntLsb, inclusive.

deltaPOCLSBCheck(i) is a function as follows:

[Table 5]

```

deltaPOCLSBCheck(int i)
{
    for(m=0; m<i; m++)
    {
        if(delta_poc_lsb_lt_minus1[i]==delta_poc_lsb_lt_minus1[m])
        {
            return 1;
        }
    }
    return 0;
}

```

- 17 -

'delta\_poc\_msb\_lt\_minus1[i]' is together with delta\_poc\_lsb\_lt\_minus1[i] used to determine the value of picture order count of the i-th long term reference picture that is included in the long-term reference picture set of the current reference picture.

The variable delta\_poc\_msb\_lt\_minus1[i] is derived as follows:

[Table 6]

```

for(n=0;n<i;n++)
{
    deltaNumSameLSBs=0;
    if(delta_poc_lsb_lt_minus1[i]==delta_poc_lsb_lt_minus1[n])
    {
        if(deltaNumSameLSBs==0)
        {
            delta_poc_msb_lt_minus1[i]=PicOrderCntMsb[i]-1;
            deltaNumSameLSBs++;
        }
        else
        {
            delta_poc_msb_lt_minus1[i]=PicOrderCntMsb[i]-
delta_poc_msb_lt_minus1 [n-1];
        }
    }
}

```

In an alternative embodiment instead of sending element delta\_poc\_msb\_lt\_minus1 when the delta\_poc\_lsb\_lt\_minus1 values are same, a poc\_msb\_lt\_minus1 or poc\_msb\_lt element may be sent. Here poc\_msb\_lt\_minus1 indicates POC value of the reference picture-1. This may be absolute POC value. Similarly poc\_msb\_lt indicates POC value of reference picture. Again this may be absolute POC value.

'used\_by\_curr\_pic\_lt\_flag[ i ]' equal to 0 specifies that the i-th long-term reference picture included in the long-term reference picture set of the current picture is not used for reference, or inter-frame prediction, by the current picture.

'num\_ref\_idx\_active\_override\_flag' equal to 1 specifies that the syntax element num\_ref\_idx\_l0\_active\_minus1 is present for P and B slices and that the syntax element num\_ref\_idx\_l1\_active\_minus1 is present for B slices. num\_ref\_idx\_active\_override\_flag equal to 0 specifies that the syntax elements num\_ref\_idx\_l0\_active\_minus1 and num\_ref\_idx\_l1\_active\_minus1 are not present.

When the current slice is a P or B slice and `field_pic_flag` is equal to 0 and the value of `num_ref_idx_l0_default_active_minus1` in the picture parameter set exceeds 15, `num_ref_idx_active_override_flag` shall be equal to 1.

When the current slice is a B slice and `field_pic_flag` is equal to 0 and the value of `num_ref_idx_l1_default_active_minus1` in the picture parameter set exceeds 15, `num_ref_idx_active_override_flag` shall be equal to 1.

'`num_ref_idx_l0_active_minus1`' specifies the maximum reference index for reference picture list 0 that shall be used to decode the slice.

When the current slice is a P or B slice and `num_ref_idx_l0_active_minus1` is not present, `num_ref_idx_l0_active_minus1` shall be inferred to be equal to `num_ref_idx_l0_default_active_minus1`.

The range of `num_ref_idx_l0_active_minus1` is specified as follows:

If `field_pic_flag` is equal to 0, `num_ref_idx_l0_active_minus1` shall be in the range of 0 to 15, inclusive. When `MbaffFrameFlag` is equal to 1, `num_ref_idx_l0_active_minus1` is the maximum index value for the decoding of frame macroblocks and  $2 * \text{num\_ref\_idx\_l0\_active\_minus1} + 1$  is the maximum index value for the decoding of field macroblocks.

Otherwise (`field_pic_flag` is equal to 1), `num_ref_idx_l0_active_minus1` shall be in the range of 0 to 31, inclusive.

'`num_ref_idx_l1_active_minus1`' specifies the maximum reference index for reference picture list 1 that shall be used to decode the slice.

When the current slice is a B slice and `num_ref_idx_l1_active_minus1` is not present, `num_ref_idx_l1_active_minus1` shall be inferred to be equal to `num_ref_idx_l1_default_active_minus1`.

The range of `num_ref_idx_l1_active_minus1` is constrained as specified in the semantics for `num_ref_idx_l0_active_minus1` with l0 and list 0 replaced by l1 and list 1, respectively.

The operation `deltaPOCLSBCheck(int i)` determines the same POC LSB is transmitted from the encoder to the decoder using absolute referencing for the current frame. In an alternative embodiment, determining if the same POC LSB is transmitted can be accomplished by checking if the value '`delta_poc_lsb_lt_minus1`' is equal to a value known to both the encoder and decoder. For example, '`delta_poc_lsb_lt_minus1`' equal to 0 could denote the POC LSB is the same as the previously transmitted POC LSB. Alternatively, '`delta_poc_lsb_lt_minus1`' equal to  $2^N - 1$ , where  $N$  denotes the number of bits used to transmit POC LSB and known to the both the encoder and decoder, 0 could denote the POC LSB is the same as the

- 19 -

previously transmitted POC LSB. In alternative embodiments, the value 'delta\_poc\_lsb\_lt\_minus1' is replaced with the syntax element 'delta\_poc\_lsb\_lt', which is generally equal to 'delta\_poc\_lsb\_lt\_minus1' plus 1. In these embodiments, the 'delta\_poc\_lsb\_lt' equal to a value known to both the encoder and decoder can indicate the picture transmitted using absolute referencing has the same POC LSB as the previous picture transmitted using absolute referencing in the same RPS. For example, 'delta\_poc\_lsb\_lt' equal to 0 could denote the POC LSB is the same as the previously transmitted POC LSB. Alternatively, 'delta\_poc\_lsb\_lt' equal to  $2^N$ , where N denotes the number of bits used to transmit POC LSB and known to the both the encoder and decoder, 0 could denote the POC LSB is the same as the previously transmitted POC LSB.

For long term reference picture set the decoding process may be done as follows:

[Table 7]

```

for( i = 0, j = 0, k = 0; i < num_long_term_pics; i++ ) {
    PocMSB=0;
    if(deltaPOCLSBCheck(i)==1)
    {
        for(n=0;n<i;n++)
        {
            PocMSB=0; deltaNumSameLSBs=0;
            if(delta_poc_lsb_lt_minus1[i]==delta_poc_lsb_lt_minus1[n])
            {
                if(deltaNumSameLSB==0)
                {
                    PocMSB =delta_poc_msb_lt_minus1[i];
                    deltaNumSameLSBs++;
                }
                else
                {
                    PocMSB +=delta_poc_msb_lt_minus1 [n];
                }
            }
        }
    }
    if( used_by_curr_pic_lt_flag[ i ] )
        PocLtCurr[ j++ ] = PocMSB+(( PicOrderCntVal - DeltaPocLt[ i ]
+ MaxPicOrderCntLsb ) %
        MaxPicOrderCntLsb)
    else
        PocLtFoll[ k++ ] = PocMSB+(( PicOrderCntVal - DeltaPocLt[ i ]
+ MaxPicOrderCntLsb ) %
        MaxPicOrderCntLsb)
}

```

- 20 -

The terms and expressions which have been employed in the foregoing specification are used therein as terms of description and not of limitation, and there is no intention, in the use of such terms and expressions, of excluding equivalents of the features shown and described or portions thereof, it being recognized that the scope of the invention is defined and limited only by the claims which follow.

The embodiments of the present invention in which an exclusive property or privilege is claimed are defined as follows:

1. A decoding device for decoding a current picture from an encoded video bitstream using a reference picture set, the decoding device comprising:

a processor; and

a memory comprising instructions, which when executed by the processor, cause the decoding device to:

receive the encoded video bitstream;

decode first information in a slice header of the encoded video bitstream, wherein the first information is a value of the least significant bits (LSB) of a picture order count (POC) of a reference picture in the reference picture set;

after decoding the first information, determine third information which is used to determine whether second information is included in the encoded video bitstream, wherein the second information is a value of the most significant bits (MSB) of the POC of the reference picture in the reference picture set;

if the determined second information is included in the encoded video bitstream, decoding the second information included in the slice header of the encoded video bitstream after the first information;

decode the reference picture set using at least the first information and the third information;

decode the current picture to generate a current decoded picture by using inter prediction based on the reference picture set; and

store the current decoded picture to be referred for future inter prediction.

2. A method for decoding a current picture from an encoded video bitstream using a reference picture set, comprising:

receiving the encoded video bitstream;

decoding first information in a slice header of the encoded video bitstream, wherein the first information is a value of the least significant bits (LSB) of a picture order count (POC) of a reference picture in the reference picture set;

after decoding the first information, determining third information which is used to determine whether second information is included in the encoded video bitstream, wherein the second information is a value of the most significant bits (MSB) of the POC of the reference picture in the reference picture set;

if the determined second information is included in the video bitstream, decoding the second information included in the slice header of the encoded video bitstream after the first information;

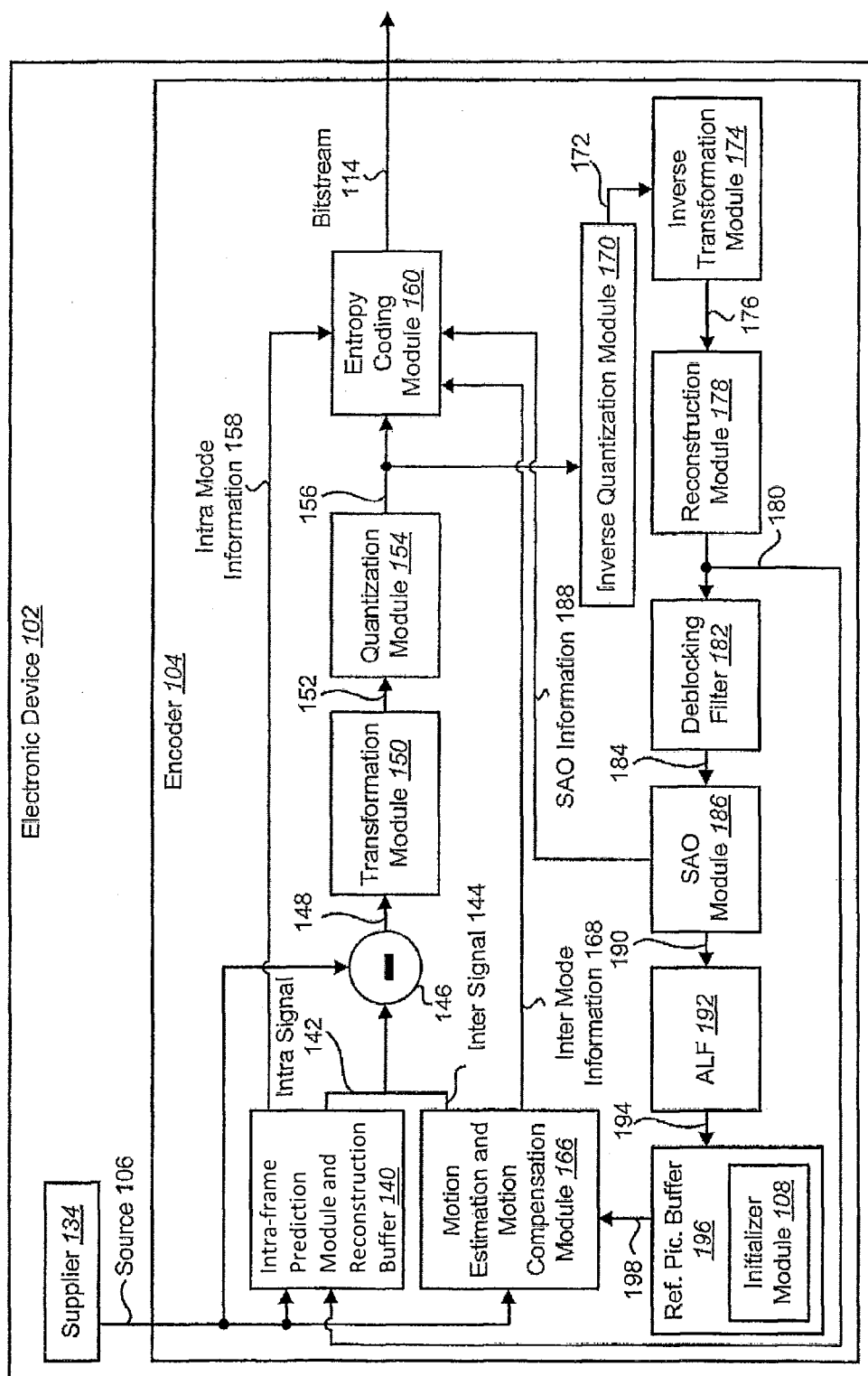
decoding the reference picture set using at least the first information and the third information;

decoding the current picture to generate a current decoded picture by using inter prediction based on the reference picture set; and

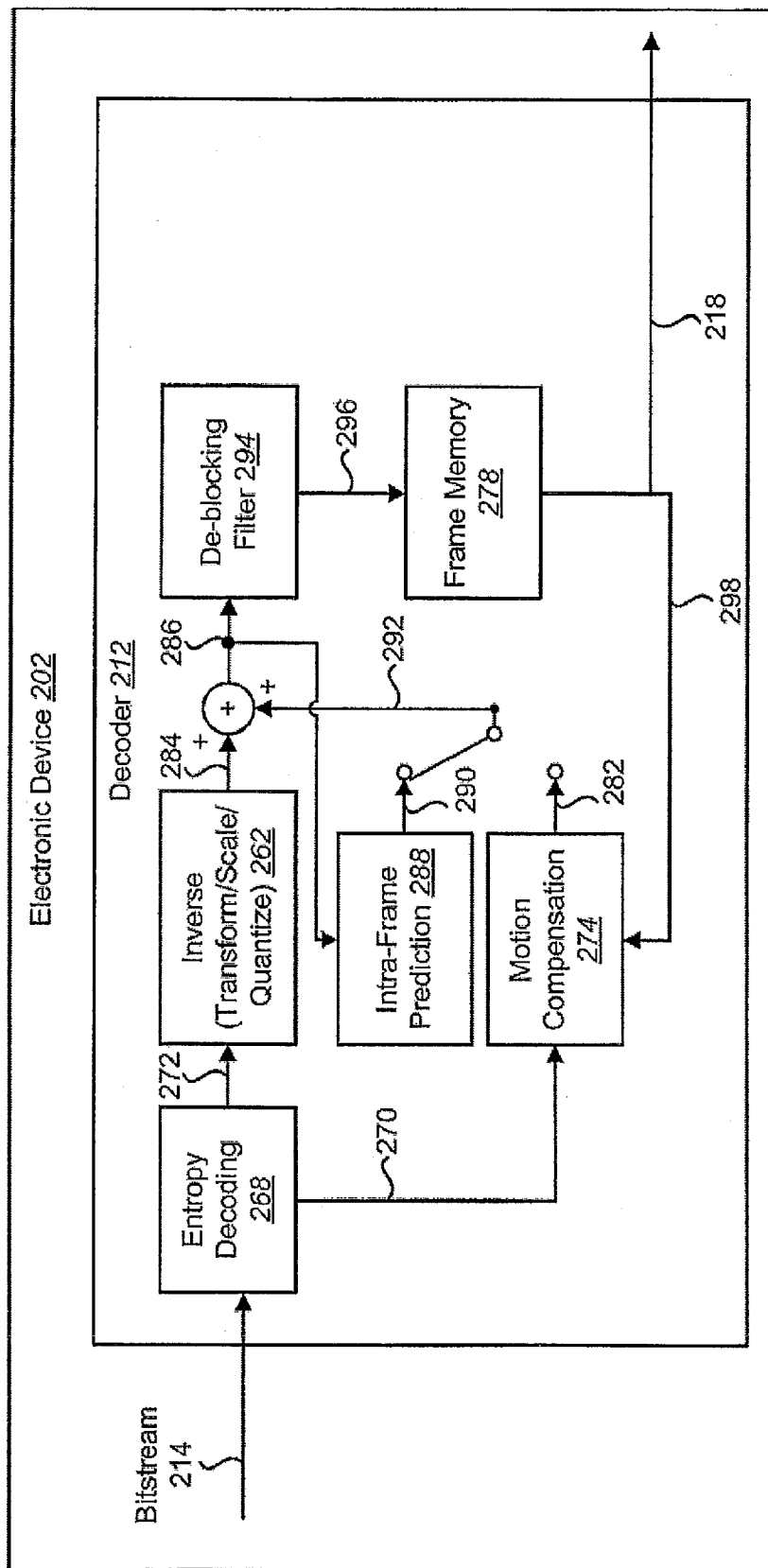
storing the current decoded picture to be referred for future inter prediction.



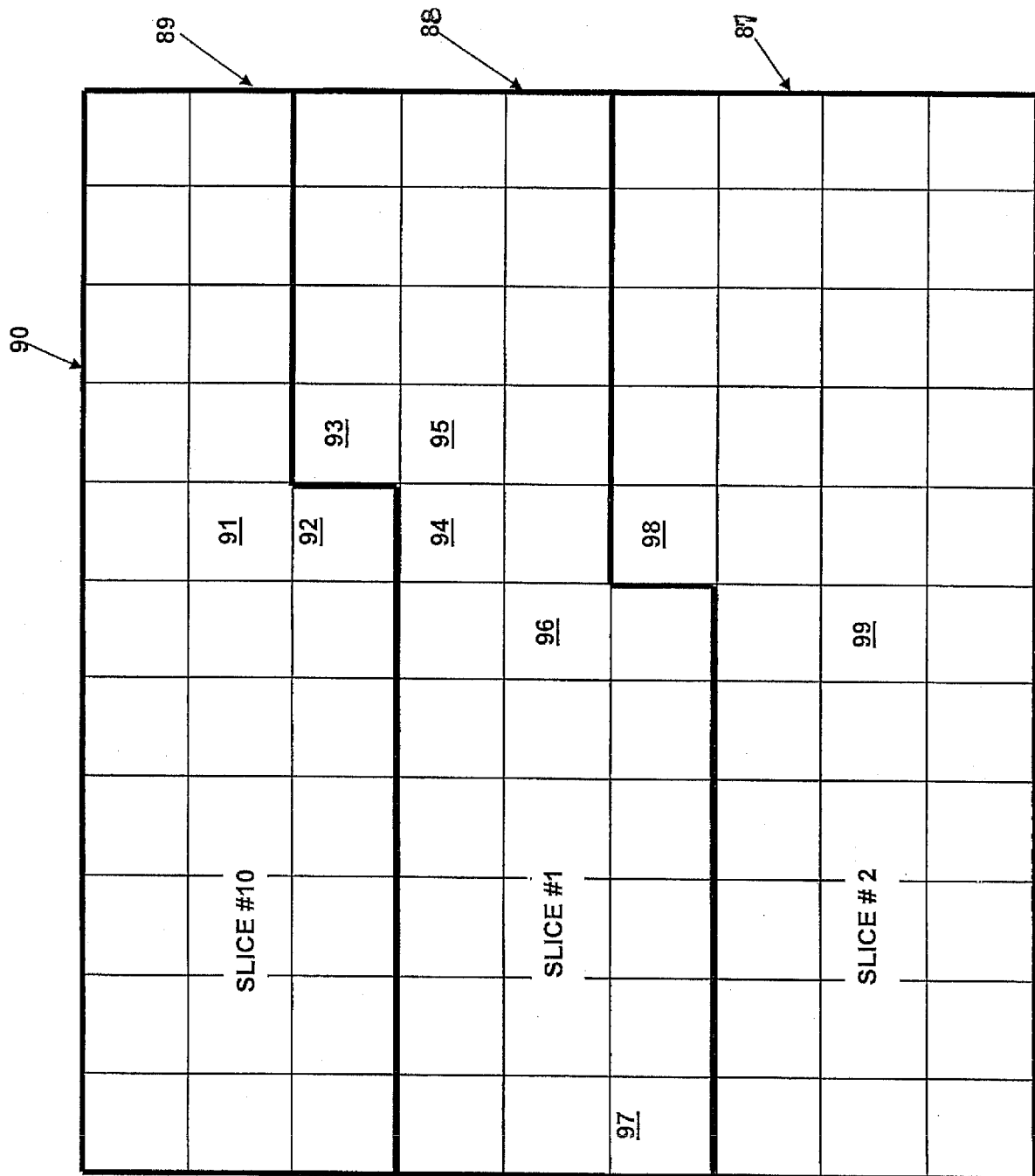
[Fig. 1]



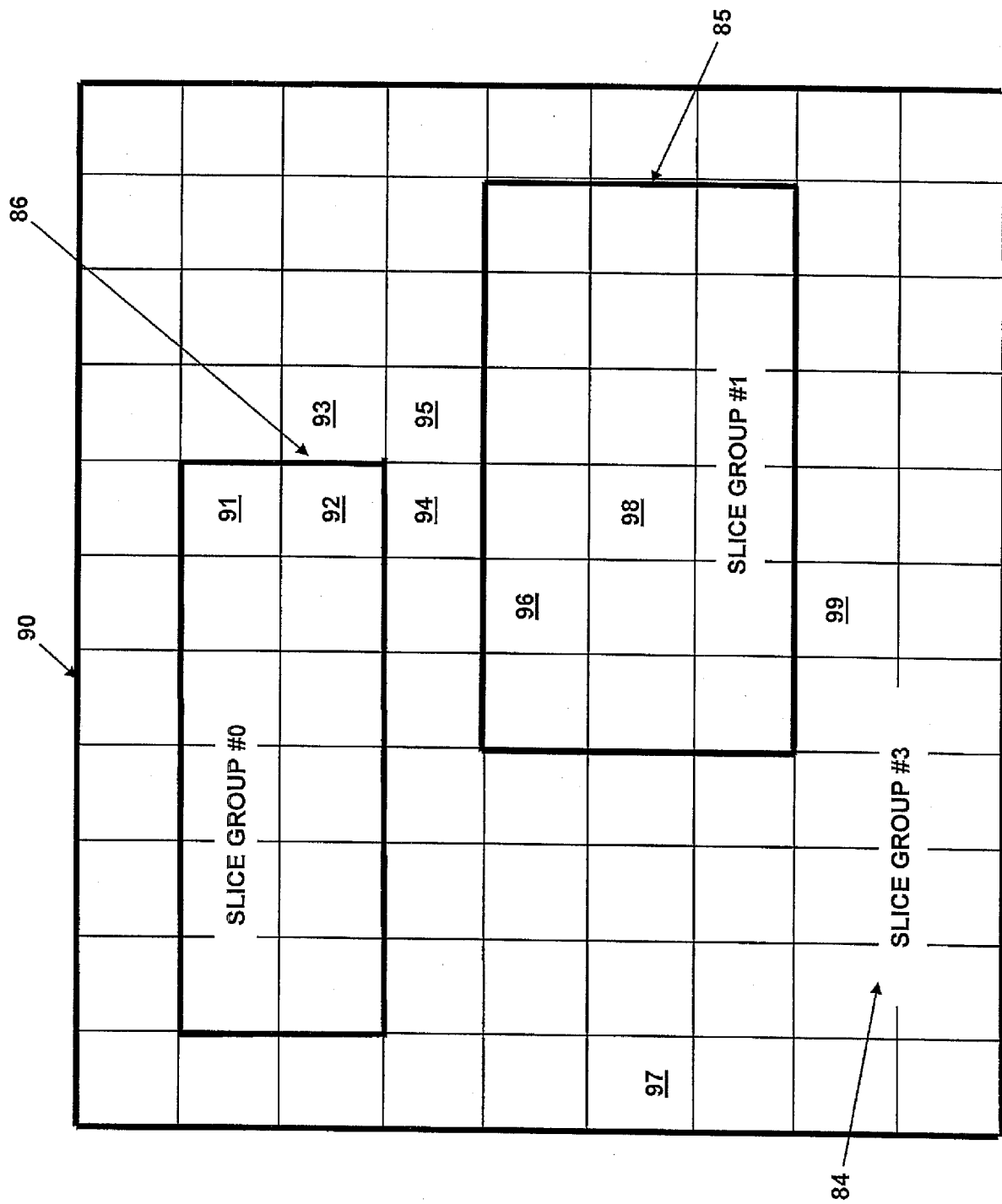
[Fig. 2]



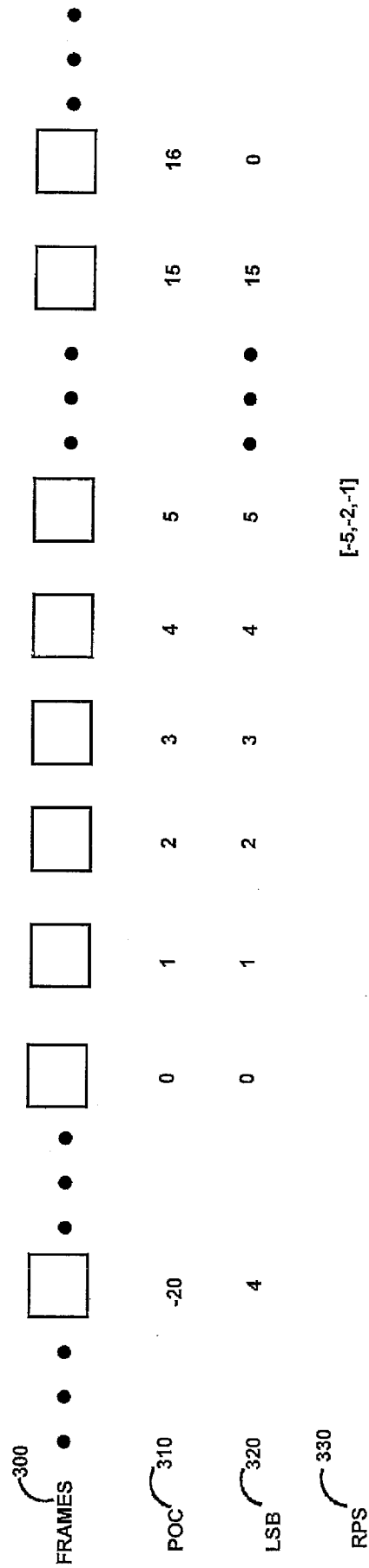
[Fig. 3]



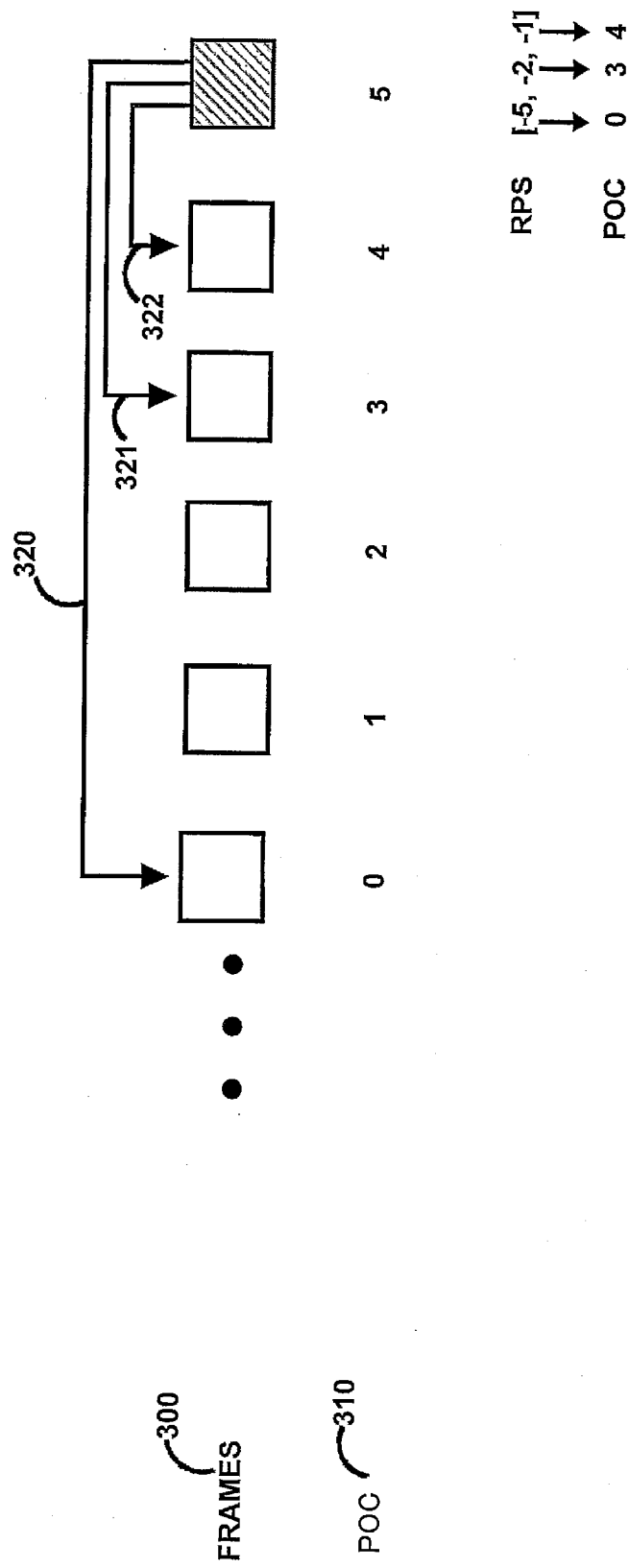
[Fig. 4]



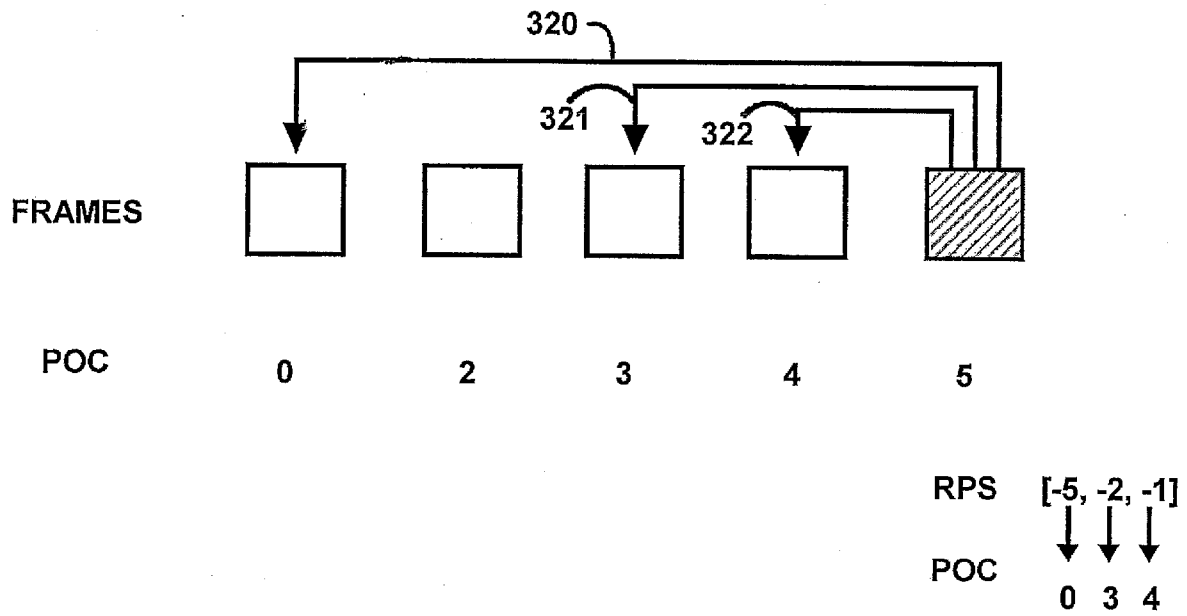
[Fig. 5]



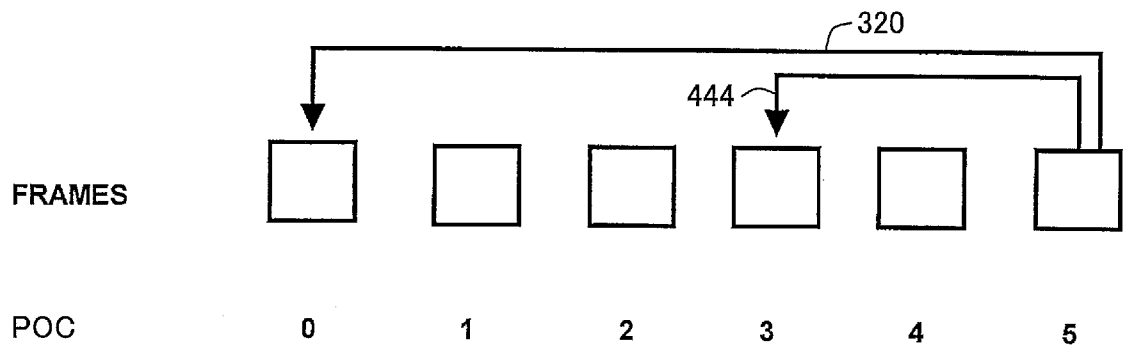
[Fig. 6]



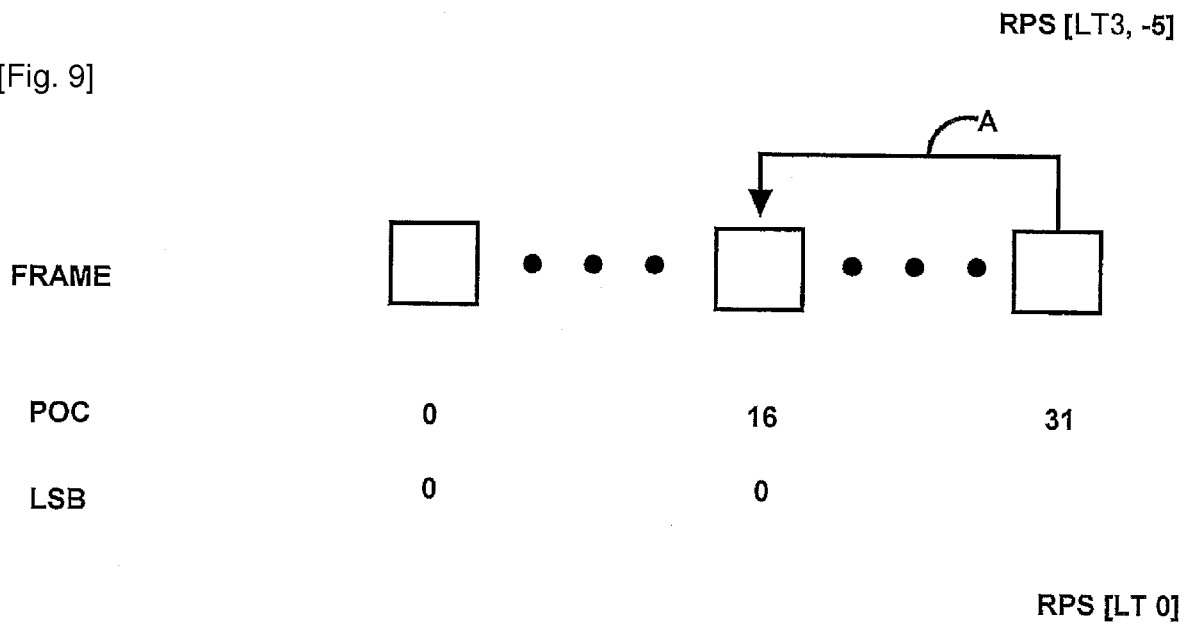
[Fig. 7]



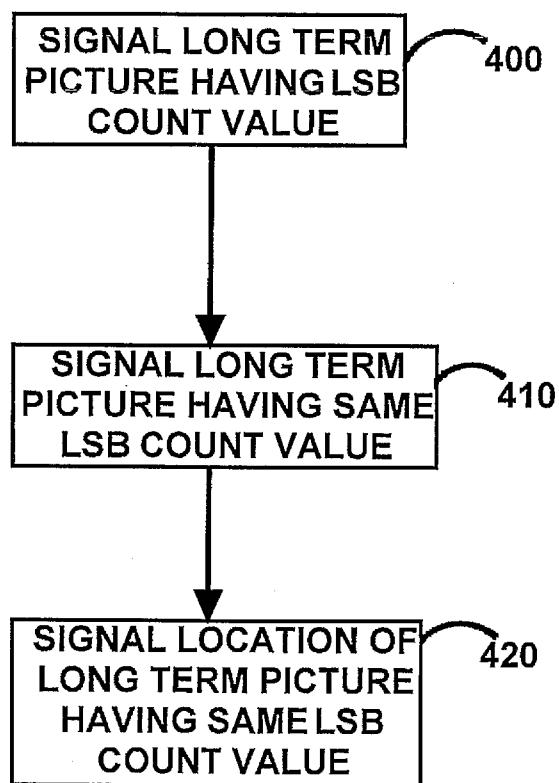
[Fig. 8]



[Fig. 9]

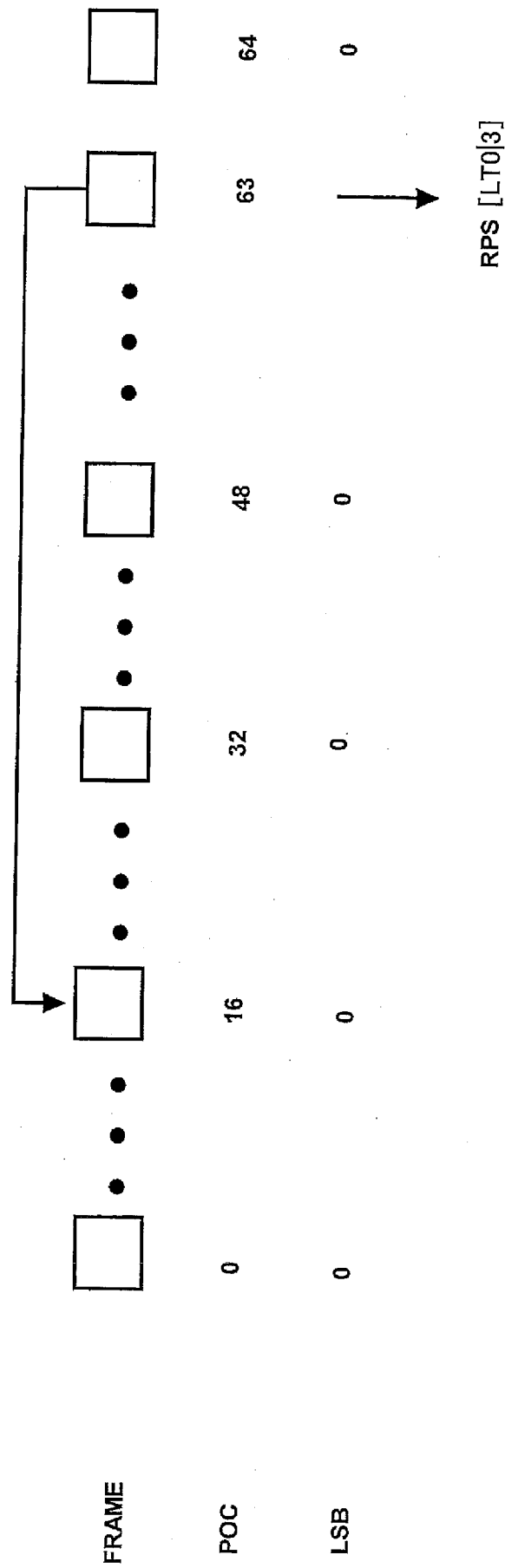


[Fig. 10]

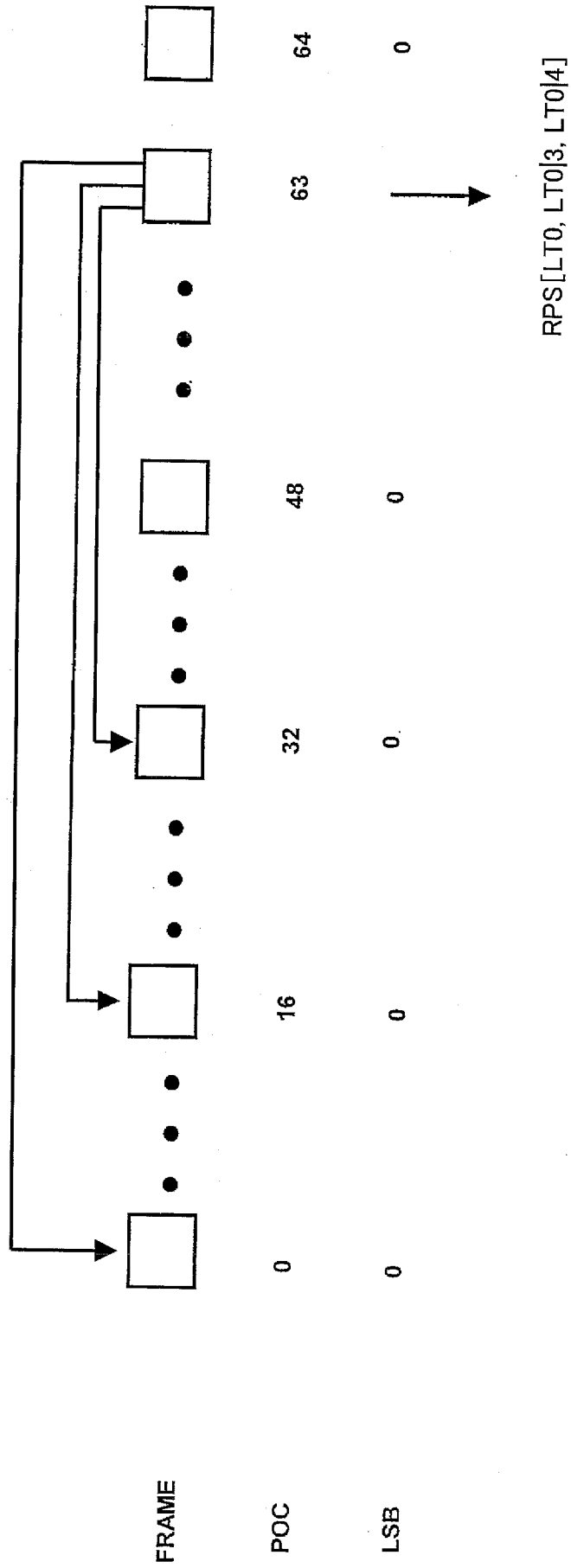




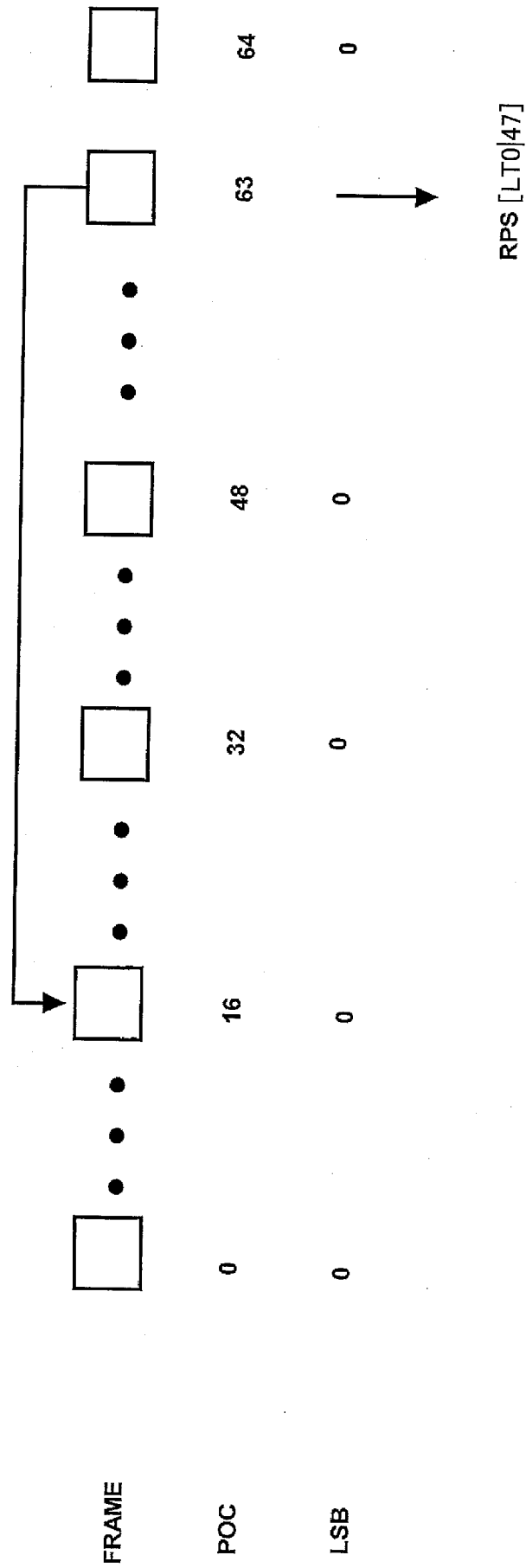
[Fig. 11A]



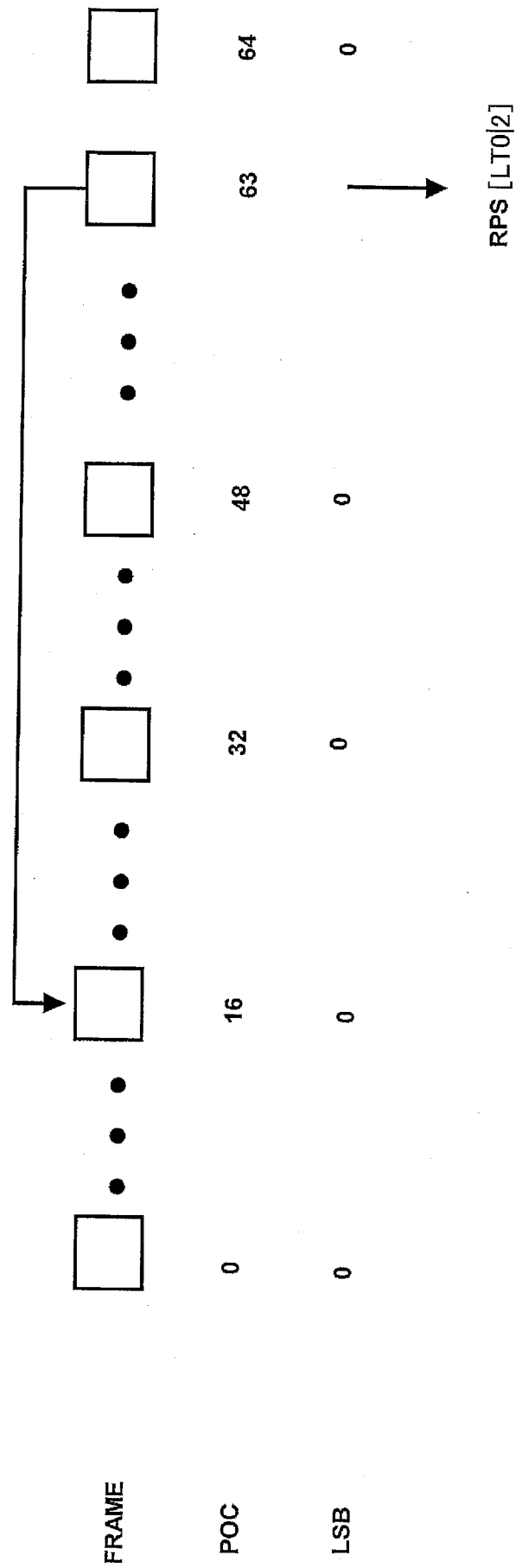
[Fig. 11B]



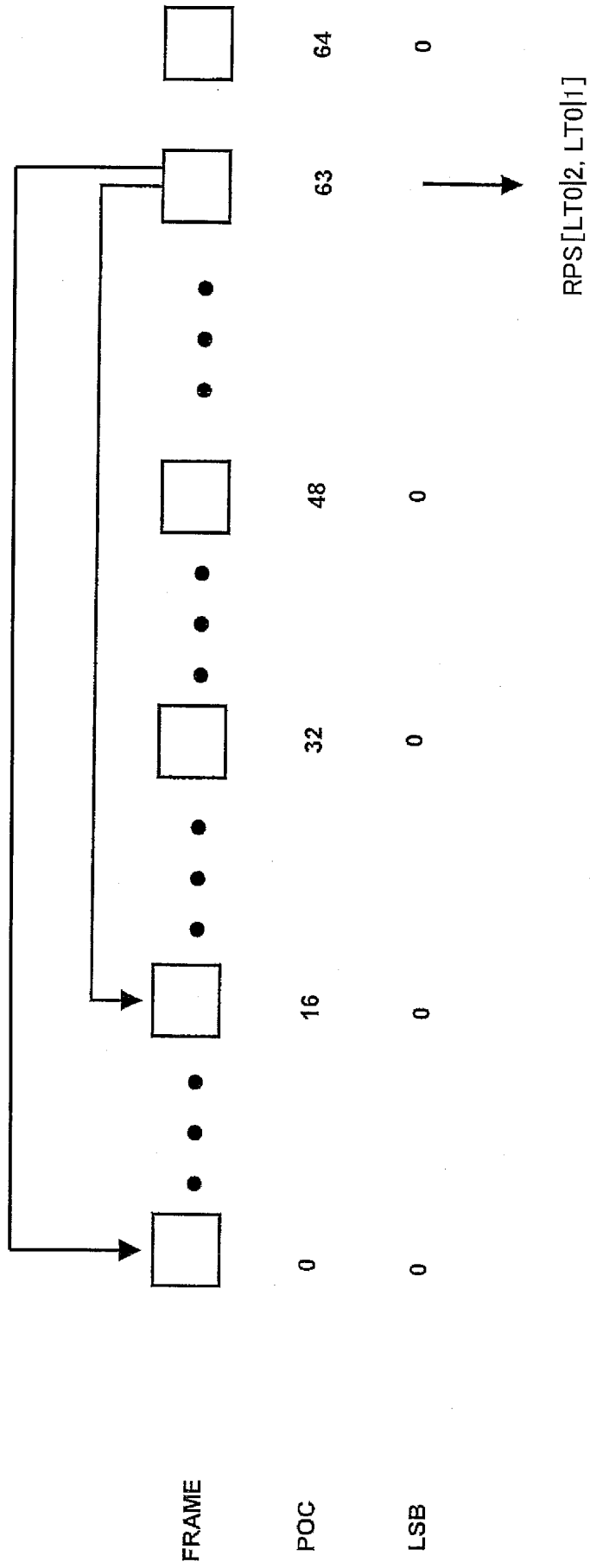
[Fig. 12]



[Fig. 13A]



[Fig. 13B]



[Fig. 14]

