



(51) International Patent Classification:  
H04N 19/00 (2014.01)

(21) International Application Number:  
PCT/CN2019/102284

(22) International Filing Date:  
23 August 2019 (23.08.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
62/722,237 24 August 2018 (24.08.2018) US  
62/734,277 21 September 2018 (21.09.2018) US  
62/735,100 22 September 2018 (22.09.2018) US  
62/735,927 25 September 2018 (25.09.2018) US  
16/548,340 22 August 2019 (22.08.2019) US

(71) Applicant: **MEDIATEK INC.** [CN/CN]; No. 1, Dusing 1st Rd., Hsinchu Science Park, Hsinchu City, Taiwan 30078 (CN).

(72) Inventors: **HSIANG, Shih-Ta**; No. 1, Dusing 1st Rd., Hsinchu Science Park, Hsinchu City, Taiwan 30078 (CN). **CHUANG, Tzu-Der**; No. 1, Dusing 1st Rd., Hsinchu Science Park, Hsinchu City, Taiwan 30078 (CN). **CHEN,**

**Ching-Yeh**; No. 1, Dusing 1st Rd., Hsinchu Science Park, Hsinchu City, Taiwan 30078 (CN).

(74) Agent: **BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.**; 16th Fl., Block A, Corporate Square, No.35 Jinrong Street, Beijing 100033 (CN).

(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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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, SA, 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, ST, 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,

(54) Title: CODING TRANSFORM COEFFICIENTS WITH THROUGHPUT CONSTRAINTS

(57) Abstract: A video coder that constrains the total number of regular bins used for entropy coding syntax elements of a current block is provided. The video coder entropy encodes or decodes the syntax elements selectively as either regular bins using context modeling or as bypass bins without context modeling. A constraint is specified to limit a total number of regular bins used for entropy coding the syntax elements of the current block. There may be no constraint limiting a number of regular bins specific to an individual syntax element of the current block.

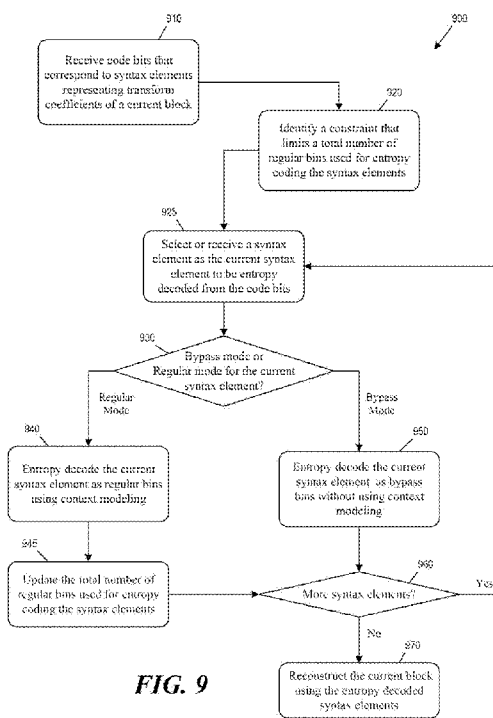


FIG. 9

WO 2020/038465 A1

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,  
KM, ML, MR, NE, SN, TD, TG).

**Published:**

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

## CODING TRANSFORM COEFFICIENTS WITH THROUGHPUT CONSTRAINTS

CROSS REFERENCE TO RELATED PATENT APPLICATION(S)

5 [0001] The present disclosure is part of a non-provisional application that claims the priority benefit of U.S. Provisional Patent Application No. 62/722,237 filed on 24 August, 2018, U.S. Provisional Patent Application No. 62/734,277 filed on 21 September, 2018, U.S. Provisional Patent Application No. 62/735,100 filed on 22 September, 2018, and U.S. Provisional Patent Application No. 62/735,927 filed on 25 September, 2018, and U.S. Patent Application No. 16/548,340 filed on 22 August, 2019. Contents of above-listed applications are herein incorporated by reference.

10

TECHNICAL FIELD

[0002] The present disclosure relates generally to video processing. In particular, the present disclosure relates to methods of coding transform coefficients.

15

BACKGROUND

[0003] Unless otherwise indicated herein, approaches described in this section are not prior art to the claims listed below and are not admitted as prior art by inclusion in this section.

20 [0004] High-efficiency video coding (HEVC) is the latest international video coding standard developed by the Joint Collaborative Team on Video Coding (JCT-VC). The input video signal is predicted from the reconstructed signal, which is derived from the coded picture regions. The prediction residual signal is processed by a linear transform. The transform coefficients are quantized, and entropy coded together with other side information in the bitstream. The reconstructed signal is generated from the prediction signal and the reconstructed residual signal after inverse transform on the de-quantized transform coefficients. The reconstructed signal is further processed by in-loop filtering for removing coding artifacts. The decoded pictures are stored in the frame  
25 buffer for predicting the future pictures in the input video signal.

[0005] In HEVC, a coded picture is partitioned into non-overlapped square block regions represented by the associated coding tree units (CTUs). A coded picture can be represented by a collection of slices, each comprising an integer number of CTUs. The individual CTUs in a slice are processed in a raster scanning order. A bi-predictive (B) slice may be decoded using intra prediction or inter prediction using at most two motion vectors and reference indices to predict the sample values of each block. A predictive (P) slice is decoded using intra prediction or inter prediction using at most one motion vector and reference index to predict the sample values of each block. An intra (I) slice is decoded using intra prediction only.

30

[0006] A CTU can be partitioned into multiple non-overlapped coding units (CUs) using the recursive quadtree (QT) structure to adapt to various local motion and texture characteristics. A CTU may also be  
35 partitioned into one or multiple smaller size CUs by a quadtree with nested multi-type tree using binary and ternary split. The resulting CU partitions can be in square or rectangular shapes.

[0007] One or more prediction units (PU) are specified for each CU. The prediction unit, together with the associated CU syntax, works as a basic unit for signaling the predictor information. The specified prediction process is employed to predict the values of the associated pixel samples inside the PU. A CU can be further partitioned using the residual quadtree (RQT) structure for representing the associated prediction residual signal. The leaf nodes of the RQT correspond to the transform units (TUs). A transform unit is comprised of a transform block (TB) of luma samples of size 8x8, 16x16, or 32x32 or four transform blocks of luma samples of size 4x4, and two corresponding transform blocks of chroma samples of a picture in 4:2:0 color format. An integer transform is applied to a transform block and the level values of quantized coefficients together with other side information are entropy coded in the bitstream.

[0008] The terms coding tree block (CTB), coding block (CB), prediction block (PB), and transform block (TB) are defined to specify the 2-D sample array of one color component associated with CTU, CU, PU, and TU, respectively. Thus, a CTU consists of one luma CTB, two chroma CTBs, and associated syntax elements. A similar relationship is valid for CU, PU, and TU. The tree partitioning is generally applied simultaneously to both luma and chroma, although exceptions apply when certain minimum sizes are reached for chroma.

[0009] A coded block flag (CBF) is utilized to signal if there is any non-zero transform coefficient in a transform block. When the CBF is equal to 0, the associated transform block is not further coded and all the coefficients in the current transform block are inferred to be equal to 0. Otherwise, the associated transform block contains at least one nonzero transform coefficient. A nonzero transform block is further divided into non-overlapped subblocks. A syntax element coded\_sub\_block\_flag may be signaled to indicate whether a current subblock contains any nonzero coefficients. When coded\_sub\_block\_flag is equal to 0, the associated transform subblock is not further coded and all the coefficients in the current transform subblock are inferred to be equal to 0. Otherwise, the associated transform block contains at least one non-zero transform coefficient. The values of the transform coefficient levels in the associated subblock are entropy coded using multiple subblock coding passes. In each coding pass, the individual transform coefficients are visited once according to a pre-defined scanning order.

[0010] In HEVC, a syntax element sig\_coeff\_flag is signaled in the 1<sup>st</sup> subblock coding pass to indicate whether the absolute value of a current transform coefficient level is greater than 0. A syntax element coeff\_abs\_level\_greater1\_flag is further signaled in the 2<sup>nd</sup> coding pass for a current coefficient with sig\_coeff\_flag equal to 1 to indicate whether the absolute value of the associated transform coefficient level is greater than 1. A syntax element coeff\_abs\_level\_greater2\_flag is further signaled in the 3<sup>rd</sup> coding pass for a current coefficient with coeff\_abs\_level\_greater1\_flag equal to 1 to indicate whether the absolute value of the associated transform coefficient level is greater than 2. The sign information and the remaining level values are further signaled by syntax elements coeff\_sign\_flag and coeff\_abs\_level\_remaining in the 4<sup>th</sup> coding and 5<sup>th</sup> subblock coding passes, respectively.

## SUMMARY

[0011] The following summary is illustrative only and is not intended to be limiting in any way. That is, the following summary is provided to introduce concepts, highlights, benefits and advantages of the novel and non-

obvious techniques described herein. Select and not all implementations are further described below in the detailed description. Thus, the following summary is not intended to identify essential features of the claimed subject matter, nor is it intended for use in determining the scope of the claimed subject matter.

5 [0012] Some embodiments of the disclosure provide a video coder (e.g., a video encoder or a video decoder) that constrains the total number of regular bins used for entropy coding syntax elements of a current block is provided. The video coder entropy encodes or decodes the syntax elements selectively as either regular bins using context modeling or as bypass bins without context modeling. A constraint is specified to limit a total number of regular bins used for entropy coding the syntax elements. There may be no constraint limiting a number of regular bins specific to an individual syntax element.

10 [0013] In some embodiments, the plurality of syntax elements are coded in a plurality of coding passes. A total number of regular bins used for entropy coding the syntax elements in a first coding pass is limited by the constraint. The plurality of syntax elements specify a plurality of transform coefficients, wherein an absolute value of a transform coefficient is indicated by at least a first flag, a second flag, and a third flag that are entropy coded in the first coding pass.

15 [0014] In some embodiments, the constraint limiting the number of regular bins is larger for a first, larger block than for a second, smaller block. In some embodiments, the constraint limiting the number of regular bins is different for a first color component than for a second color component. In some embodiments, the constraint limiting the number of regular bins is computed by multiplying a default constraint with a predefined factor that is derived from a ratio of a number of coded subblocks in the current block and a number of all subblocks in the  
20 current block.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0015] The accompanying drawings are included to provide a further understanding of the present disclosure and are incorporated in and constitute a part of the present disclosure. The drawings illustrate implementations of  
25 the present disclosure and, together with the description, serve to explain the principles of the present disclosure. It is appreciable that the drawings are not necessarily in scale as some components may be shown to be out of proportion than the size in actual implementation in order to clearly illustrate the concept of the present disclosure.

[0016] FIG. 1 conceptually illustrates a video coder selectively entropy coding syntax elements of a subblock of a block in either regular mode or bypass mode based on a constraint on numbers of regular bins used for the  
30 subblock.

[0017] FIG. 2 illustrates coding passes that includes flags that indicate the absolute value of a transform coefficient level when entropy coding a subblock of a transform block.

[0018] FIG. 3 illustrates sharing the allowed number of regular bins across the subblocks of a current block.

[0019] FIG. 4 illustrates an example video encoder.

35 [0020] FIG. 5 illustrates portions of the video encoder that selectively entropy encodes syntax elements of a subblock of a block in either regular mode or bypass mode based on a constraint that limits the number of regular bins used for the subblock.

[0021] FIG. 6 conceptually illustrates a process that selectively entropy encodes syntax elements of a subblock of a block in either regular mode or bypass mode based on a constraint that limits the number of regular bins used for the subblock.

[0022] FIG. 7 illustrates an example video decoder.

5 [0023] FIG. 8 illustrates portions of the video decoder that selectively entropy decodes syntax elements of a subblock of a block in either regular mode or bypass mode based on a constraint that limits the number of regular bins used for the subblock.

[0024] FIG. 9 conceptually illustrates a process that selectively entropy decodes syntax elements of a subblock of a block in either regular mode or bypass mode based on a constraint that limits the number of regular  
10 bins used for the subblock.

[0025] FIG. 10 conceptually illustrates an electronic system with which some embodiments of the present disclosure are implemented.

#### DETAILED DESCRIPTION

15 [0026] In the following detailed description, numerous specific details are set forth by way of examples in order to provide a thorough understanding of the relevant teachings. Any variations, derivatives and/or extensions based on teachings described herein are within the protective scope of the present disclosure. In some instances, well-known methods, procedures, components, and/or circuitry pertaining to one or more example implementations disclosed herein may be described at a relatively high level without detail, in order to avoid  
20 unnecessarily obscuring aspects of teachings of the present disclosure.

[0027] Joint Video Expert Team (JVET) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 are currently in the process of establishing a next-generation video coding standard. Some of the work is specified in Versatile Video Coding (VVC) Working Draft (WD) 2 (B. Brossey et al., "Versatile Video Coding (Draft 2)," Joint Video Expert Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), Doc. JVE061001,  
25 11th Meeting: Ljubljana, SI, 10–18 July 2018).

[0028] VVC WD2 specifies that the transform coefficients of a coded block may be quantized by dependent scalar quantization. The selection of one of the two quantizers is specified by a state machine with four states. The state for a current transform coefficient is determined by the state and the parity of the absolute level value for the preceding transform coefficient in scanning order. The syntax elements sig\_coeff\_flag, par\_level\_flag and rem\_abs\_gt1\_flag are signaled in the 1st subblock coding pass. The partially reconstructed absolute value of a transform coefficient level from the 1st is given by  
30

[0029] 
$$\text{AbsLevelPass1} = \text{sig\_coeff\_flag} + \text{par\_level\_flag} + 2 * \text{rem\_abs\_gt1\_flag}$$

[0030] Context selection for entropy coding sig\_coeff\_flag is dependent on the state for the current coefficient. par\_level\_flag is thus signaled in the 1st coding pass for deriving the state for the next coefficient. The syntax elements rem\_abs\_gt2\_flag, abs\_remainder, and coeff\_sign\_flag are further signaled in the 2nd, 3rd, and  
35 4th coding passes, respectively. The fully reconstructed absolute value of a transform coefficient level is given by

[0031]  $AbsLevel = AbsLevelPass1 + 2 * (rem\_abs\_gt2\_flag + abs\_remainder)$  The transform coefficient level is given by

[0032]  $TransCoeffLevel = (2 * AbsLevel - (QState > 1 ? 1 : 0)) * (1 - 2 * coeff\_sign\_flag)$ ,

[0033] where QState indicates the state for the current transform coefficient.

### 5 I. Coding Transform Coefficients

[0034] Some embodiments of the disclosure provide a video coder that signals or parses the absolute value of a transform coefficient level using multiple coding passes. In some embodiments, the fully reconstructed absolute value of a transform coefficient level can be represented by

[0035]  $AbsLevel = (AbsLevelRS1 \ll 1) + ParityBit$ ,

10 [0036] where ParityBit indicates the parity of the absolute value of a transform coefficient level and AbsLevelRS1 indicates the absolute value of a transform coefficient level downward shifted by 1. The syntax elements sig\_coeff\_flag, coeff\_abs\_level\_greater1\_flag and par\_level\_flag may be signaled in the 1st coding pass. When sig\_coeff\_flag is equal to 0, the associated transform coefficient level is equal to 0. Otherwise, coeff\_abs\_level\_greater1\_flag is further signaled to indicate whether the absolute value of the associated transform  
15 coefficient level is greater than 1. When coeff\_abs\_level\_greater1\_flag is equal to 0, AbsLevelRS1 is set equal to 0 and ParityBit is set equal to sig\_coeff\_flag for the associated transform coefficient. Otherwise, par\_level\_flag is further signaled to indicate the value of ParityBit for the associated transform coefficient level. The further information about AbsLevelRS1 and the sign for the associated transform coefficient level can be further signaled using one or multiple coding passes.

20 [0037] When coeff\_abs\_level\_greater1\_flag is greater than 1 for a current coefficient, the video coder may further signal a syntax element coeff\_abs\_level\_rs1\_gt1\_flag to indicate whether AbsLevelRS1 is greater than 1. When coeff\_abs\_level\_rs1\_gt1\_flag is equal to 0, AbsLevelRS1 is set equal to 1 for the associated transform coefficient level. Otherwise, AbsLevelRS1 is inferred to be greater than 1 and the remaining value of AbsLevelRS1 is further signaled. In some embodiments, the syntax element coeff\_abs\_level\_rs1\_gt1\_flag is  
25 signaled in the same coding pass as coeff\_abs\_level\_greater1\_flag. In some embodiments, coeff\_abs\_level\_rs1\_gt1\_flag is signaled in a separate coding pass after the coding pass for signaling coeff\_abs\_level\_greater1\_flag.

[0038] When coeff\_abs\_level\_rs1\_gt1\_flag is greater than 1 for a current coefficient, the video coder may further signal a syntax element coeff\_abs\_level\_rs1\_gt2\_flag to indicate whether AbsLevelRS1 is greater than 2.  
30 When coeff\_abs\_level\_rs1\_gt2\_flag is equal to 0, AbsLevelRS1 is set equal to 2 for the associated transform coefficient level. Otherwise, AbsLevelRS1 is inferred to be greater than 2 and the remaining value of AbsLevelRS1 is further signaled. coeff\_abs\_level\_rs1\_gt2\_flag may be signaled in the same coding pass as coeff\_abs\_level\_rs1\_gt1\_flag or signaled in a separate coding pass after the coding pass for signaling coeff\_abs\_level\_rs1\_gt1\_flag. In some embodiments, syntax elements coeff\_abs\_level\_greater1\_flag,  
35 coeff\_abs\_level\_rs1\_gt1\_flag and coeff\_abs\_level\_rs1\_gt2\_flag are all signaled in one coding pass.

[0039] In some embodiments, the video coder performs entropy coding of the residual block quantized by using dependent scalar quantization. The video coder utilizes the syntax elements sig\_coeff\_flag,

coeff\_abs\_level\_greater1\_flag, par\_level\_flag, coeff\_abs\_level\_rs1\_gt1\_flag, coeff\_abs\_level\_rs1\_gt2\_flag, and coeff\_abs\_level\_rs1\_remainder for coding the absolute value of a transform coefficient level. **Table 1** below provides a syntax table for decoding a residual transform block. The syntax element coeff\_abs\_level\_rs1\_remainder is employed to signal the remaining absolute value of an associated coefficient level when coeff\_abs\_level\_rs1\_gt2\_flag is equal to 1. The syntax elements sig\_coeff\_flag, coeff\_abs\_level\_greater1\_flag, coeff\_abs\_level\_rs1\_gt1\_flag, coeff\_abs\_level\_rs1\_gt2\_flag, and coeff\_abs\_level\_rs1\_remainder are inferred to be equal to 0 when not coded.

**Table 1: Residual Coding Syntax**

residual_coding( x0, y0, log2TbWidth, log2TbHeight, cIdx ) {	Descriptor
if( transform_skip_enabled_flag && ( cIdx != 0    cu_mts_flag[ x0 ][ y0 ] == 0 ) && ( log2TbWidth <= 2 ) && ( log2TbHeight <= 2 ) )	
<b>transform_skip_flag</b> [ x0 ][ y0 ][ cIdx ]	ae(v)
<b>last_sig_coeff_x_prefix</b>	ae(v)
<b>last_sig_coeff_y_prefix</b>	ae(v)
if( last_sig_coeff_x_prefix > 3 )	
<b>last_sig_coeff_x_suffix</b>	ae(v)
if( last_sig_coeff_y_prefix > 3 )	
<b>last_sig_coeff_y_suffix</b>	ae(v)
log2SbSize = ( Min( log2TbWidth, log2TbHeight ) < 2 ? 1 : 2 )	
numSbCoeff = 1 << ( log2SbSize << 1 )	
lastScanPos = numSbCoeff	
lastSubBlock = ( 1 << ( log2TbWidth + log2TbHeight - 2 * log2SbSize ) ) - 1	
do {	
if( lastScanPos == 0 ) {	
lastScanPos = numSbCoeff	
lastSubBlock--	
}	
lastScanPos--	
xS = DiagScanOrder[ log2TbWidth - log2SbSize ][ log2TbHeight -	

log2SbSize ] [ lastSubBlock ][ 0 ]	
yS = DiagScanOrder[ log2TbWidth - log2SbSize ][ log2TbHeight - log2SbSize ] [ lastSubBlock ][ 1 ]	
xC = ( xS << log2SbSize ) + DiagScanOrder[ log2SbSize ][ log2SbSize ][ lastScanPos ][ 0 ]	
yC = ( yS << log2SbSize ) + DiagScanOrder[ log2SbSize ][ log2SbSize ][ lastScanPos ][ 1 ]	
} while( ( xC != LastSignificantCoeffX )    ( yC != LastSignificantCoeffY ) )	
QState = 0	
for( i = lastSubBlock; i >= 0; i-- ) {	
startQStateSb = QState	
xS = DiagScanOrder[ log2TbWidth - log2SbSize ][ log2TbHeight - log2SbSize ] [ lastSubBlock ][ 0 ]	
yS = DiagScanOrder[ log2TbWidth - log2SbSize ][ log2TbHeight - log2SbSize ] [ lastSubBlock ][ 1 ]	
inferSbDcSigCoeffFlag = 0	
if( ( i < lastSubBlock ) && ( i > 0 ) ) {	
coded_sub_block_flag[ xS ][ yS ]	ae(v)
inferSbDcSigCoeffFlag = 1	
}	
firstSigScanPosSb = numSbCoeff	
lastSigScanPosSb = -1	
for( n = ( i == lastSubBlock ) ? lastScanPos - 1 : numSbCoeff - 1; n >= 0; n-- )	
{	
xC = ( xS << log2SbSize ) + DiagScanOrder[ log2SbSize ][ log2SbSize ][ n ][ 0 ]	
yC = ( yS << log2SbSize ) + DiagScanOrder[ log2SbSize ][ log2SbSize ][ n ][ 1 ]	
if( coded_sub_block_flag[ xS ][ yS ] && ( n > 0	

!inferSbDcSigCoeffFlag ) ) {	
sig_coeff_flag[ xC ][ yC ]	ae(v)
}	
ParityBit[n] = sig_coeff_flag[ xC ][ yC ]	
if( sig_coeff_flag[ xC ][ yC ] ) {	
if( lastSigScanPosSb == -1 )	
lastSigScanPosSb = n	
firstSigScanPosSb = n	
coeff_abs_level_greater1_flag [ n ]	ae(v)
if( coeff_abs_level_greater1_flag [ n ] ){	
par_level_flag [ n ]	ae(v)
coeff_abs_level_rs1_gt1_flag [ n ]	ae(v)
ParityBit[n] = par_level_flag [ n ]	
}	
}	
AbsLevelPass1[ xC ][ yC ] = 2 * ( coeff_abs_level_greater1_flag [ n ] + coeff_abs_level_rs1_gt1_flag [ n ] ) + ParityBit [ n ]	
if( dep_quant_enabled_flag )	
QState = QStateTransTable[ QState ][ ParityBit [ n ] ]	
}	
for( n = numSbCoeff - 1; n >= 0; n-- ) {	
if(coeff_abs_level_rs1_gt1_flag [ n ] )	
coeff_abs_level_rs1_gt2_flag [ n ]	ae(v)
}	
for( n = numSbCoeff - 1; n >= 0; n-- ) {	
xC = ( xS << log2SbSize ) + DiagScanOrder[ log2SbSize ][ log2SbSize ][ n ][ 0 ]	
yC = ( yS << log2SbSize ) + DiagScanOrder[ log2SbSize ][ log2SbSize ][ n ][ 1 ]	

if(coeff_abs_level_rs1_gt2_flag[ n ] )	
coeff_abs_level_rs1_remainder[ n ]	
AbsLevel[ xC ][ yC ] = AbsLevelPass1[ xC ][ yC ] + 2 * (coeff_abs_level_rs1_gt2_flag [ n ] + coeff_abs_level_rs1_remainder [ n ] )	
}	
if( dep_quant_enabled_flag    !sign_data_hiding_enabled_flag )	
signHidden = 0	
else	
signHidden = ( lastSigScanPosSb - firstSigScanPosSb > 3 ? 1 : 0 )	
for( n = numSbCoeff - 1; n >= 0; n-- ) {	
xC = ( xS << log2SbSize ) + DiagScanOrder[ log2SbSize ][ log2SbSize ][ n ][ 0 ]	
yC = ( yS << log2SbSize ) + DiagScanOrder[ log2SbSize ][ log2SbSize ][ n ][ 1 ]	
if( sig_coeff_flag[ xC ][ yC ] && ( !signHidden    ( n != firstSigScanPosSb ) ) )	
coeff_sign_flag[ n ]	ae(v)
}	
if( dep_quant_enabled_flag ) {	
QState = startQStateSb	
for( n = numSbCoeff - 1; n >= 0; n-- ) {	
xC = ( xS << log2SbSize ) + DiagScanOrder[ log2SbSize ][ log2SbSize ][ n ][ 0 ]	
yC = ( yS << log2SbSize ) + DiagScanOrder[ log2SbSize ][ log2SbSize ][ n ][ 1 ]	
if( sig_coeff_flag[ xC ][ yC ] )	
TransCoeffLevel[ x0 ][ y0 ][ cIdx ][ xC ][ yC ] = ( 2 * AbsLevel[ xC ][ yC ] - ( QState > 1 ? 1 : 0 ) ) * ( 1 - 2 *	

coeff_sign_flag[ n ] )	
QState = QStateTransTable[ QState ][ ParityBit [ n ] ]	
} else {	
sumAbsLevel = 0	
for( n = numSbCoeff - 1; n >= 0; n-- ) {	
xC = ( xS << log2SbSize ) + DiagScanOrder[ log2SbSize ][ log2SbSize ][ n ][ 0 ]	
yC = ( yS << log2SbSize ) + DiagScanOrder[ log2SbSize ][ log2SbSize ][ n ][ 1 ]	
if( sig_coeff_flag[ xC ][ yC ] ) {	
TransCoeffLevel[ x0 ][ y0 ][ cIdx ][ xC ][ yC ] = AbsLevel[ xC ][ yC ] * ( 1 - 2 * coeff_sign_flag[ n ] )	
if( signHidden ) {	
sumAbsLevel += AbsLevel[ xC ][ yC ]	
if( ( n == firstSigScanPosSb ) && ( sumAbsLevel % 2 ) == 1 )	
TransCoeffLevel[ x0 ][ y0 ][ cIdx ][ xC ][ yC ] = -TransCoeffLevel[ x0 ][ y0 ][ cIdx ][ xC ][ yC ]	
}	
}	
}	
}	
}	
if( cu_mts_flag[ x0 ][ y0 ] && ( cIdx == 0 ) && !transform_skip_flag[ x0 ][ y0 ][ cIdx ] && ( ( CuPredMode[ x0 ][ y0 ] == MODE_INTRA && numSigCoeff > 2 )    ( CuPredMode[ x0 ][ y0 ] == MODE_INTER ) ) ) {	
mts_idx[ x0 ][ y0 ]	ae(v)
}	

[0040] Some embodiments of the disclosure relate to complexity reduction by reducing the number of coding passes for entropy coding a transform block or a sub-block. For example, in some embodiments, the syntax element `rem_abs_gt2_flag` may be coded in the same coding pass as the syntax element `rem_abs_gt1_flag`. In some embodiments, the information on the signs and remaining values of transform coefficients are entropy coded using CABAC in a bypass mode and may be signaled in one coding pass. In one embodiment, all syntax elements entropy coded using CABAC in a regular mode are signaled in one coding pass. All syntax elements entropy coded using CABAC in a bypass mode are signaled in another coding pass.

## *II. Constraint on Regular Bins for a Subblock or a Coding Pass*

[0041] For achieving high compression efficiency, the context-based adaptive binary arithmetic coding (CABAC) mode, or known as regular mode, is employed for entropy coding the values of the syntax elements in HEVC. As the arithmetic coder in the CABAC engine only encode the binary symbol values, the CABAC operation first needs to convert the value of a syntax element into a binary string, the process commonly referred to as binarization. During the coding process, the probability models are gradually built up from the coded symbols for the different contexts. The selection of the modeling context for coding the next binary symbol can be determined by the coded information. Values of syntax elements processed in the regular mode are referred to as “regular bins” and entropy coded by using context modeling. Symbols may also be coded without the context modeling stage and assume an equal probability distribution, commonly referred to as the bypass mode, for improving bitstream parsing throughput rate. Values of syntax elements processed in the bypassed mode are referred to as “bypass bins” and entropy coded by without context modeling. In HEVC, the values of the syntax elements such as `coded_sub_block_flag`, `sig_coeff_flag`, `coeff_abs_level_greater1_flag`, and `coeff_abs_level_greater2_flag` in a transform subblock are coded in the regular mode as regular bins. The values of the syntax elements `coeff_sign_flag` and `coeff_abs_level_remaining` in a transform subblock are coded in the bypass mode as bypass bins. In order to limit the total number of the regular bins for entropy coding transform coefficient levels in a sub-block under the worst-case scenario, each subblock is allowed to code only up to eight `coeff_abs_level_greater1_flag` values and one `coeff_abs_level_greater2_flag` value. In this way, the maximum number of the regular bins in each subblock is limited to 26.

[0042] Some embodiments of the disclosure relate to the constraints on the maximum allowed numbers of CABAC regular bins for coding a transform subblock. Having the constraint for regular bins of a subblock facilitates controlling the bitstream parsing throughput rate per subblock (e.g., under the worst-case condition) because entropy coding using CABAC in the regular mode involves higher complexity than in the bypass mode. In some embodiments, a video coder may apply the constraint on the maximum allowed number of the regular bins in one subblock or one subblock coding pass but have no specific constraints on the maximum allowed numbers of the regular bins for individual syntax elements present in the said subblock or the said subblock coding pass. The video coder may track the accumulated number of consumed regular bins in a current subblock or a current coding pass. When the specified maximum allowed number of regular bins is reached, the video coder may have the CABAC engine switched to the bypass mode for the remaining coding passes in the current subblock. Or the video

coder may terminate the coding passes in the regular CABAC mode. The remaining absolute values of the transform coefficient levels are coded by the subblock coding passes in the bypass mode.

[0043] FIG. 1 conceptually illustrates a video coder selectively entropy coding syntax elements of a subblock of a block in either regular mode or bypass mode based on a constraint on numbers of regular bins used for the subblock. The figure illustrates a current block 100 that is being coded (encoded or decoded). The current block 100 may be a transform block or TU of a block of pixels such as a coding unit (CU). The current block 100 includes  $4 \times 4 = 16$  subblocks, including a subblock 110.

[0044] The subblock 110 includes multiple syntax elements. These syntax elements (SEs) may represent several transform coefficients (e.g., Coef1-Coef5) of the current block. Each transform coefficient may be represented by several syntax elements (e.g., SE1-SE6). The syntax elements of a transform coefficient may include flags that indicate the absolute value and the parity of the transform coefficient.

[0045] These syntax elements are entropy coded using CABAC either as regular bins (to be coded in regular mode by context modeling) or bypass bins (to be coded in bypass mode without context modeling). The video coder imposes a constraint 115. In some embodiments, the constraint limits a maximum allowed number of CABAC regular bins for the transform block 100, e.g., the maximum allowed numbers of CABAC regular bins for coding the subblock 110. The constraint provides a budget of regular bins that is shared by the syntax elements of the block or of the subblock. Entropy coding each syntax element as regular bins consumes the budget until there is no more budget left.

[0046] In the example of FIG. 1, the constraint 115 for the subblock 110 is 32. Some of the syntax elements are coded by using regular mode (syntax elements illustrated as shaded). These syntax elements include Coef1:SE1, Coef1:SE2, Coef1:SE3, Coef1:SE4, Coef2:SE1, Coef2:SE2, Coef2:SE3, Coef2:SE4, and Coef3:SE1. Entropy encoding these syntax elements consumes 30 regular bins. To encode any more syntax elements as regular bins would exceed the subblock constraint of 32. Consequently, all other syntax elements are coded as bypass bins (illustrated as unshaded).

[0047] The example of FIG. 1 also illustrates several coding passes 121, 122, and 123. In the first coding pass 121, the video coder entropy encodes SE1-SE4 of each coefficient, specifically in the order of Coef1:SE1, Coef1:SE2, Coef1:SE3, Coef1:SE4, Coef2:SE1, Coef2:SE2, Coef2:SE3,...etc. In the second coding pass 122, the video coder entropy encodes SE5 of each coefficient, specifically in the order of Coef1:SE5, Coef2:SE5, Coef3:SE5, Coef4:SE5,...etc. In the third coding pass 123, the video coder entropy encodes SE6 of each coefficient, specifically in the order of Coef1:SE6, Coef2:SE6, Coef3:SE6, Coef4:SE6,...etc.

[0048] In some embodiments, the constraint 115 limits the total number of regular bins that are used by all syntax elements by the subblock, regardless of which coding passes they are coded in. In some embodiments, the constraint 115 limits the total number of regular bins that are used by syntax elements coded during the first coding pass 121 (or the leading coding pass) but does not limit the number of regular bins used during other coding passes.

[0049] In some embodiments, regardless of whether the constraint 115 is applicable to all syntax elements of the subblock 110 or only to syntax elements coded in the first coding pass 121, the video coder has no specific constraints on the maximum allowed numbers of the regular bins for individual syntax elements present in the

subblock 110 or in the first coding pass 121. In some embodiments, a video coder may apply the constraint on the maximum allowed number of the regular bins in one subblock or one subblock coding pass (e.g., the first coding pass 121) but also apply specific constraints on the maximum allowed numbers of the regular bins for individual syntax elements present in the subblock or the subblock coding pass.

5       **[0050]** In some embodiments, HEVC's constraints on the specified maximum allowed numbers of regular bins for syntax elements `coeff_abs_level_greater1_flag` and `coeff_abs_level_greater2_flag` are removed. Specifically, the syntax elements `coeff_abs_level_greater1_flag` and `coeff_abs_level_greater2_flag` are coded in a leading subblock coding pass. The leading subblock coding pass in the regular CABAC mode may be terminated once the accumulated number of the regular bins in the current subblock reaches the specified maximum allowed  
10 number 25 or 26. The remaining absolute values of the transform coefficient levels are coded by the syntax element `coeff_abs_level_remaining` in a bypass mode. In this way, the video coder may still meet the worst-case constraint on the number of regular bins per subblock while being able to use more regular bins than the maximum numbers originally specified for syntax elements `coeff_abs_level_greater1_flag` and `coeff_abs_level_greater2_flag`.

15       **[0051]** In some embodiments, a video coder may apply a constraint on the maximum allowed number of regular bins in the leading subblock coding pass for entropy coding the transform coefficient levels generated by dependent scalar quantization. The video coder may track the accumulated number of consumed regular bins in a current leading coding pass for signaling `sig_coeff_flag`, `par_level_flag` and `rem_abs_gt1_flag`, or for signaling `sig_coeff_flag`, `coeff_abs_level_greater1_flag`, `par_level_flag` and `coeff_abs_level_rs1_gt1_flag`.

20       **[0052]** When the accumulated number of regular bins is greater than a specified threshold value, the video coder may switch its CABAC engine to bypass mode for coding the remaining pass in bypass bins. Or the video coder may terminate the current leading coding pass. The remaining absolute values of the transform coefficient levels are coded by the subblock coding pass in the bypass mode for signaling `abs_remainder` or for signaling `coeff_abs_level_rs1_remainder`.

25       **[0053]** In some embodiments, a video coder may apply a constraint on the maximum allowed number of regular bins for entropy coding a transform subblock resulted from dependent scalar quantization. The video coder may track the accumulated number of consumed regular bins for signaling `sig_coeff_flag`, `par_level_flag`, `rem_abs_gt1_flag` and `rem_abs_gt2_flag` or for signaling `sig_coeff_flag`, `coeff_abs_level_greater1_flag`, `par_level_flag`, `coeff_abs_level_rs1_gt1_flag` and `coeff_abs_level_rs1_gt2_flag`. When the accumulated number of regular bins is greater than a specified threshold value, the remaining absolute values of the transform coefficient  
30 levels are all coded by the subblock coding pass in the bypass mode for signaling `abs_remainder` or for signaling `coeff_abs_level_rs1_remainder`.

35       **[0054]** In some embodiments, the `coeff_abs_level_rs1_gt1_flag` may be integrated into the first coding pass. Specifically, `gt1_flag` (`coeff_abs_level_greater1_flag`, or `rem_abs_gt1_flag`) is signaled before the `par_flag` (the parity bit, `par_level_flag`), and the signaling of `gt2_flag` (`coeff_abs_level_rs1_gt1_flag`, or `rem_abs_gt2_flag`) is moved to the first coding pass. Some example of the codewords for different levels are listed in Table 2.

**Table 2: Examples of the Codewords for Different Levels**

Abs	1 <sup>st</sup> Pass	2 <sup>nd</sup> Pass
-----	----------------------	----------------------

Level	Sig_flag	Gt1_flag	Par_flag	Gt2_flag	Rem_abs
0	0	-	-	-	-
1	1	0	-	-	-
2	1	1	0	0	-
3	1	1	1	0	-
4	1	1	0	1	0
5	1	1	1	1	0
6	1	1	0	1	1
7	1	1	1	1	1
8	1	1	0	1	2
9	1	1	1	1	2

[0055] Since the context modeling of sig\_flag (sig\_coeff\_flag) is related to the state of dependent quantization determined by the parity bit of the previous coefficient, and the state transition is dependent on the parity bit of the level, the modified parsing process is described as below. During the parsing stage of the sig\_flag, the context models for even level will be prepared and stored in the local buffer. If the sig\_flag is 0, the stored context models for even level is applied to the next sig\_flag. If the sig\_flag is 1, gt1\_flag is parsed. During the parsing stage of gt1\_flag, the context models for odd level is loaded in and then stored in the local buffer. If the gt1\_flag is 0, the stored context models for odd level can be applied to the next sig\_flag. If the gt1\_flag is 1, then both the par\_flag and gt2\_flag are parsed. These two bins are parsed in order, that is, the bins for par\_flag is parsed the bins for gt2\_flag. If the par\_flag is 0, the context models for even level will be loaded in during the parsing stage of gt2\_flag. If the par\_flag is 1, the stored context models for odd level can be used for the next sig\_flag. In one embodiment, one bin is always parsed after the parity bit in the first coding pass. After moving the gt2\_flag to the first coding pass, the context modeling can be modified accordingly.

[0056] In some embodiments, the value of the variable sumAbs1 becomes the sum of partially reconstructed absolute levels after the first coding pass, which is equal to sum of the coeff\_partial (= sig\_flag + gt1\_flag + par\_flag + (gt2\_flag << 1)) or min(4+(x&1), x) in the five neighboring positions, where x is the absolutely level of the coefficient. In some embodiments, the value of sumAbs1 only takes sig\_flag, par\_flag, and gt1\_flag into account. Specifically, the sumAbs1 is equal to the sum of min(coeff\_partial, 3) or (sig\_flag + gt1\_flag + par\_flag) or min(2+(x&1), x) in the five neighboring positions. In some embodiment, the video coder parses at least one bin after par\_flag in the same pass. Therefore, the time segment spent on parsing other bins can be used to prepare the context models for the next sig\_flag.

[0057] FIG. 2 illustrates coding passes that includes flags that indicate the absolute value of a transform coefficient level when entropy coding a subblock of a transform block. The figure illustrates coding passes for entropy coding transform coefficients of a subblock of a transform block. Each transform coefficient includes syntax elements that indicates the absolute value (sig\_flag, gt1\_flag, gt2\_flag), the parity (par\_flag), the remainder (absr), and the sign of the transform coefficient.

[0058] The sig\_flag may refer to the sig\_coeff\_flag, or any other flag that indicates whether the absolute value of a corresponding transform coefficient level (e.g., value of AbsLevelRS1) is greater than 0. The gt1\_flag may refer to the coeff\_abs\_level\_greater1\_flag or any other flag that is used to indicate whether the absolute value of a corresponding transform coefficient is greater than 1. The gt2\_flag may refer to the coeff\_abs\_level\_rs1\_gt1\_flag, or any other flag that is used to indicate whether the absolute value of a corresponding transform coefficient is greater than 3. The par\_flag may refer to the par\_level\_flag, or any other flag that indicates the parity of the absolute value of a transform coefficient level (e.g., the value of ParityBit). The remainder may be indicated by a flag such as abs\_remainder or coeff\_abs\_level\_remaining. The sign may be indicated by a flag such as coeff\_sign\_flag.

[0059] FIG. 2 illustrates two schemes 210 and 220 of coding passes. In the first scheme 210 of coding passes, the sig\_flag, the gt1\_flag, and the par\_flag of each transform coefficient are coded in a first coding pass, while the gt2\_flag, is coded in a second coding pass, the remainder is coded in a third coding pass, and the sign is coded in a fourth coding pass. In other words, not all indicators of absolute values of transform coefficients are in the same coding pass. In the second scheme 220 of coding passes, the sig\_flag, the gt1\_flag, the par\_flag and the gt2\_flag of each transform coefficient are coded in a same first coding pass. The remainder and the sign are coded in second and third coding passes, respectively. In other words, at least three, or all, indicators of absolute value of a transform coefficient are in the same coding pass. Regardless of whether gt2\_flag is coded in the first coding pass or a subsequent coding pass, in some embodiments, the total number of regular bins used to entropy code the syntax elements of the first coding pass or the subblock of the transform block is limited by a constraint or budget as described reference to FIG. 1 above.

[0060] In some embodiments, the maximum number of regular bins of some syntax elements or the maximum number of regular bins of some coding passes may be further specified. For example, the maximum number of regular bins for coeff\_abs\_level\_rs1\_gt2\_flag or the maximum number of regular bins for coding coeff\_abs\_level\_rs1\_gt2\_flag in a coding pass can be specified by the video coder. When coding these syntax elements in a coding pass, when the accumulated number of regular bins is greater than a specified threshold value, the video coder may switch its CABAC engine to bypass mode for coding the remaining bins of the coding pass as bypass bins or terminate the current (leading) coding pass. The remaining absolute values of the transform coefficient levels are coded by the subblock coding pass in the bypass mode. In some embodiments, some of the syntax elements are not counted for the regular bin number constraint or can always be coded in regular bins even when there is no allowed regular bin quota or there is no budget left for regular bins. For example, in some embodiments, the sig\_coeff\_flag is coded in regular bin regardless of any regular bin constraint for its coding pass. For another example, in some embodiments, the sig\_coeff\_flag and the par\_level\_flag are coded in regular bin (regardless of any regular bin constraint for its coding pass). In some embodiments, if there is no quota or no budget left for using regular bins to code par\_level\_flag and rem\_abs\_gt1\_flag in the coding pass, then only the sig\_coeff\_flag is coded in the coding pass as regular bins. The state for context modeling of sig\_coeff\_flag may remain the same or transition with the parsed result of sig\_coeff\_flag.

[0061] In some embodiments, in the coding pass, if the remaining available regular bin number is not larger than the required regular bin number of a coefficient in this coding pass, i.e., if the remaining budget for regular bins for this coding pass is insufficient to meet the required number of regular bins for coding the coefficient, the video coder may switch its CABAC engine to the bypass mode for the remainder of the coding pass, or the video coder may terminate the current leading coding pass. The remaining absolute values of the transform coefficient levels are all coded by the subblock coding pass in the bypass mode. This required remaining budget for coding a coefficient as regular bins in the coding pass is also referred as the termination threshold for the coefficient. For example, if `sig_coeff_flag`, `par_level_flag`, and `rem_abs_gt1_flag` are coded in a coding pass, the termination threshold (for using regular bins) is 3. If `sig_coeff_flag`, `par_level_flag`, `rem_abs_gt1_flag`, and `rem_abs_gt2_flag` are coded in a coding pass, the termination threshold is 4.

[0062] In some embodiments, the specified constraints on the maximum allowed numbers of regular bins for coding transform subblocks may be pre-defined or explicitly signaled in the bitstream, for example, using one or multiple high-level syntax sets such as sequence parameter set (SPS), picture parameter set (PPS) or slice header. The specified constraints may be dependent on the profile and the level indicated in the coded bitstream. Different constraints may be employed for different color components. Different constraints may be employed for different transform subblock sizes. In some embodiments, a video coder may limit the maximum allowed number of regular bins for a 2x2 subblock to be equal to one quarter of the maximum allowed number of regular bins for a 4x4 subblock. In some embodiments, a video coder may limit the maximum allowed number of regular bins for a 4x4 chroma subblock to be equal to one half of the maximum allowed number of regular bins for a 4x4 luma subblock. The specified constraint for a 4x4 transforms subblock may be signaled in the SPS. In one example, the specified constraint for a 4x4 luma subblock, 4x4 chroma subblock, 2x2 chroma subblock may be signaled in the SPS. In one embodiment, the maximum allowed number of regular bins for a 4x4 chroma subblock is equal to or smaller than the maximum allowed number of regular bins for a 4x4 luma subblock. In another embodiment, the maximum allowed number of regular bins for a 2x2 chroma subblock is equal to or smaller than the maximum allowed number of regular bins for a 2x2 chroma subblock.

[0063] In some embodiments, the maximum regular bins of a 4x4 luma subblock can be 25, 30, or 32, the maximum regular bins of a 4x4 chroma subblock can be 25, 16, or 15, the maximum regular bins of a 2x2 chroma subblock can be 2, 3, 4, 5, 6, 7, or 8. For example, the maximum regular bins of {a 4x4 luma subblock, a 4x4 chroma subblock, a 2x2 chroma subblock} can be {25, 25, 8}, {25, 25, 6}, {25, 25, 7}, {32, 16, 4}, {32, 32, 8}, {30, 16, 4}, {30, 15, 4}, or {30, 15, 3}.

[0064] In some embodiments, the maximum regular bins of coding `coeff_abs_level_rs1_gt1_flag` can also be specified. For example, the maximum regular bins of the `coeff_abs_level_rs1_gt1_flag` in a 4x4 luma subblock can be 2, 3, 4, or 5, the maximum regular bins of the `coeff_abs_level_rs1_gt1_flag` in a 4x4 chroma subblock can be 0, 1, 2, 3, or 4, the maximum regular bins of the `coeff_abs_level_rs1_gt1_flag` in a 2x2 chroma subblock can be 0, 1, or 2. For example, the maximum regular bins of {a 4x4 luma subblock, the `coeff_abs_level_rs1_gt1_flag` in a 4x4 luma subblock, a 4x4 chroma subblock, the `coeff_abs_level_rs1_gt1_flag` in a 4x4 chroma subblock, a 2x2 chroma subblock, the `coeff_abs_level_rs1_gt1_flag` in a 2x2 chroma subblock } can be {25, 4, 25, 4, 8, 2}, {25, 4, 25, 4, 6,

2}, {25, 4, 25, 4, 6, 1}, {25, 3, 25, 3, 6, 2}, {25, 3, 25, 3, 6, 1}, {25, 2, 25, 2, 6, 1}, {25, 2, 25, 2, 6, 0}, {25, 1, 25, 1, 6, 1}, {25, 1, 25, 1, 6, 0}, {25, 3, 25, 3, 7, 2}, {32, 4, 16, 2, 4, 1}, {32, 4, 16, 2, 4, 0}, {32, 4, 16, 4, 4, 1}, {32, 4, 16, 4, 4, 0}, {32, 3, 16, 2, 4, 1}, {32, 3, 16, 2, 4, 0}, {30, 4, 16, 2, 4, 1}, {30, 4, 16, 2, 4, 0}, {30, 3, 16, 2, 4, 1}, {30, 3, 16, 2, 4, 0}, {30, 4, 15, 2, 4, 1}, {30, 4, 15, 2, 4, 0}, {30, 3, 15, 2, 4, 1}, {30, 3, 15, 2, 4, 0}, {30, 4, 15, 2, 3, 1}, {30, 4, 15, 2, 3, 0}, {30, 3, 15, 2, 3, 1}, {30, 3, 15, 2, 3, 0}, {32, 4, 32, 4, 8, 1}.

[0065] In some embodiments, a video coder may have the constraints on the maximum allowed numbers of regular bins specified for entropy coding some sizes of transform units or transform blocks. The constraints on the maximum allowed numbers of regular bins for a current subblock can be derived from the constraints specified for the related transform unit or transform block. For example, the constraints on the maximum allowed numbers of regular bins for 4x4 transform subblocks and 2x2 transform subblocks can be derived from the constraints on the maximum allowed numbers of regular bins for 4x4 transform blocks and 2x2 transform blocks, respectively, in a video coder.

[0066] In some embodiments, when the `rem_abs_gt2_flag` is included in the leading coding pass (e.g., the first coding pass) and the regular bin number constraint value of `rem_abs_gt2_flag` is specified separately (from the regular bin number constraint value of `sig_coeff_flag`, `par_level_flag`, `rem_abs_gt1_flag`) and the `rem_abs_gt2_flag` is signaled, the remaining level of the coefficient is coded when the partial sum of the coefficient ( $= \text{sig\_flag} + \text{gt1\_flag} + \text{par\_flag} + (\text{gt2\_flag} \ll 1)$  or  $= \text{sig\_flag} + \text{par\_flag} + (\text{gt1\_flag} \ll 1) + (\text{gt2\_flag} \ll 1)$ ) is equal to or larger than 5 if the parity bit is signaled before `rem_abs_gt1_flag` and is equal to or larger than 4 if the parity bit is signaled after `rem_abs_gt1_flag`. Otherwise, the remaining level of the coefficient is not coded. If the `rem_abs_gt2_flag` is not signaled because that the available number of regular bin of `rem_abs_gt2_flag` is zero, the remaining level of the coefficient is coded when the partial sum of the coefficient is equal to or larger than 4 if the parity bit is signaled before `rem_abs_gt1_flag` and is equal to or larger than 2 if the parity bit is signaled after `rem_abs_gt1_flag`. Otherwise, the remaining level of the coefficient is not coded. However, if available number of regular bins is smaller than a threshold, in one embodiment, the leading coding pass is terminated, the level is coded with Golomb-Rice code directly if the coefficient is not coded by the leading pass. When the leading coding pass is terminated, the remaining level coding pass is performed. The remaining level coding pass starts from the last position of the current subblock to the first position of the subblock.

[0067] In some embodiments, the maximum allowed numbers of regular bins is specified for a TU. The threshold for terminating regular bins coding (or the budget of regular bins) of a TU can be derived from the threshold for terminating regular bins coding of a 4x4 sub-block. For example, if the TU is a luma 8x8 block and the threshold of a luma 4x4 block is 32, the threshold of this 8x8 luma TU is  $32 \times 4 = 128$ . The threshold value can be different for different color component or different TU size.

[0068] In some embodiments, the constraints on the maximum allowed numbers of regular bins is specified for a transform subblock, however, the constraint number for the allowed regular bins may depend on the current TU size, TU type, TU width, TU height, total number of subblocks in current TU, subblock size, subblock width, subblock height, color component, last significant coefficient position, last significant subblock position, or any

combination of above. For example, the constraint number may depend on total number of subblocks in current TU, color component, subblock size, and last significant subblock position.

[0069] For coding the coefficients of subblocks of the TU, the allowed number of regular bins value can be shared across the subblocks. For example, in the first sub-block of the TU, the allowed number of regular bins is 128. When coding the first subblock, the allowed number of regular bins is decreased when using regular bin for coding coefficients. The remaining allowed value is used for the next subblock of the TU. In some embodiments, the number of required coded subblock can be derived after (or based on) the encoded/decoded the last significant coefficient position or the last significant subblock position. For example, a 16x16 TU has 16 4x4 subblock. The total number of allowed regular bins for the TU can be  $32 \times 16 = 512$ . After decoding the last significant coefficient position, if the number of required coded subblocks is 8, then each subblock can use 64 regular bins ( $64 \times 8 = 512$ ).

[0070] In some embodiments, when the last significant subblock position is determined, the number of subblocks to be coded can be determined. If the number of subblock to be coded is smaller than the total number of subblocks in current TU, i.e., some subblock is skipped in entropy coding, the constraint on the maximum allowed numbers of regular bins per subblock can be increased.

[0071] FIG. 3 illustrates sharing the allowed number of regular bins across the subblocks of a current block. The figure illustrates a first block 310 and a second block 320 that are transform blocks or TUs. Each of the first block 310 and the second block 320 has  $4 \times 4 = 16$  subblocks. Each of the first block 310 and the second block 320 has a total constraint or budget of regular bins of 512 for the block.

[0072] All of the subblocks of the first block 310 are coded (illustrated as shaded) such that the position of the last coded subblock is subblock 319, which is the 16th subblock in the first block 310. Each subblock of the first block 310 is given a constraint or budget for regular bins of  $512 \div 16 = 32$ .

[0073] On the other hand, only 8 of the subblocks of the second block 320 are coded such that the position of the last coded subblock is subblock 325, which is the 8th subblock in second block 320. Each coded subblock of the second block 320 is given a constraint or budge for regular bins of  $512 \div 8 = 64$ .

[0074] In some embodiments, the constraint on the regular bins for a subblock is computed by multiplying a default constraint value by a factor. (The default constraint in the example of FIG. 3 is 32.) The factor is related to a ratio computed as (the total number of subblocks in current TU)/(the number of subblock to be coded), or  $\text{floor}((\text{the total number of subblocks in current TU})/(\text{the number of subblock to be coded}))$ , where  $\text{floor}(x)$  means to find the maximum integer value that smaller than or equal to  $x$ . In some embodiments, some predefined factor are specified. The predefined factor can be chosen from  $\{4, 2, 1.5, 1.25, 1\}$ , or other values based on power of two denominators that can be computed by left/right shift and add operations. Let the total number of subblocks in current TU be defined as  $A$ , and the number of subblock to be coded be defined as  $B$ . In one example, if  $B \times 4 \leq A$ , then the factor can be 4. Otherwise, if  $B \times 2 \leq A$ , then the factor can be 2. Otherwise, if  $B \times 3 \leq A \times 2$ , then the factor can be 1.5. Otherwise, if  $B \times 5 \leq A \times 4$ , then the factor can be 1.25. Otherwise, the factor can be 1. In another example, if  $B \times 4 \leq A$ , then the factor can be 4. Otherwise, if  $B \times 3 \leq A$ , then the factor can be 3. Otherwise, if  $B \times 2 \leq A$ , then the factor can be 2. Otherwise, if  $B \times 3 \leq A \times 2$ , then the factor can be 1.5. Otherwise, if  $B \times 5 \leq A \times 4$ ,

then the factor can be 1.25. Otherwise, the factor can be 1. For different syntax type or different subblock size or different color component, the factor can be different.

### *III.Example Video Encoder*

[0075] FIG. 4 illustrates an example video encoder 400. As illustrated, the video encoder 400 receives input video signal from a video source 405 and encodes the signal into bitstream 495. The video encoder 400 has several components or modules for encoding the signal from the video source 405, at least including some components selected from a transform module 410, a quantization module 411, an inverse quantization module 414, an inverse transform module 415, an intra-picture estimation module 420, an intra-prediction module 425, a motion compensation module 430, a motion estimation module 435, an in-loop filter 445, a reconstructed picture buffer 450, a MV buffer 465, and a MV prediction module 475, and an entropy encoder 490. The motion compensation module 430 and the motion estimation module 435 are part of an inter-prediction module 440.

[0076] In some embodiments, the modules 410 – 490 are modules of software instructions being executed by one or more processing units (e.g., a processor) of a computing device or electronic apparatus. In some embodiments, the modules 410 – 490 are modules of hardware circuits implemented by one or more integrated circuits (ICs) of an electronic apparatus. Though the modules 410 – 490 are illustrated as being separate modules, some of the modules can be combined into a single module.

[0077] The video source 405 provides a raw video signal that presents pixel data of each video frame without compression. A subtractor 408 computes the difference between the raw video pixel data of the video source 405 and the predicted pixel data 413 from the motion compensation module 430 or intra-prediction module 425. The transform module 410 converts the difference (or the residual pixel data or residual signal 409) into transform coefficients (e.g., by performing Discrete Cosine Transform, or DCT). The quantization module 411 quantizes the transform coefficients into quantized data (or quantized coefficients) 412, which is encoded into the bitstream 495 by the entropy encoder 490.

[0078] The inverse quantization module 414 de-quantizes the quantized data (or quantized coefficients) 412 to obtain transform coefficients, and the inverse transform module 415 performs inverse transform on the transform coefficients to produce reconstructed residual 419. The reconstructed residual 419 is added with the predicted pixel data 413 to produce reconstructed pixel data 417. In some embodiments, the reconstructed pixel data 417 is temporarily stored in a line buffer (not illustrated) for intra-picture prediction and spatial MV prediction. The reconstructed pixels are filtered by the in-loop filter 445 and stored in the reconstructed picture buffer 450. In some embodiments, the reconstructed picture buffer 450 is a storage external to the video encoder 400. In some embodiments, the reconstructed picture buffer 450 is a storage internal to the video encoder 400.

[0079] The intra-picture estimation module 420 performs intra-prediction based on the reconstructed pixel data 417 to produce intra prediction data. The intra-prediction data is provided to the entropy encoder 490 to be encoded into bitstream 495. The intra-prediction data is also used by the intra-prediction module 425 to produce the predicted pixel data 413.

[0080] The motion estimation module 435 performs inter-prediction by producing MVs to reference pixel data of previously decoded frames stored in the reconstructed picture buffer 450. These MVs are provided to the

motion compensation module 430 to produce predicted pixel data.

[0081] Instead of encoding the complete actual MVs in the bitstream, the video encoder 400 uses MV prediction to generate predicted MVs, and the difference between the MVs used for motion compensation and the predicted MVs is encoded as residual motion data and stored in the bitstream 495.

5 [0082] The MV prediction module 475 generates the predicted MVs based on reference MVs that were generated for encoding previously video frames, i.e., the motion compensation MVs that were used to perform motion compensation. The MV prediction module 475 retrieves reference MVs from previous video frames from the MV buffer 465. The video encoder 400 stores the MVs generated for the current video frame in the MV buffer 465 as reference MVs for generating predicted MVs.

10 [0083] The MV prediction module 475 uses the reference MVs to create the predicted MVs. The predicted MVs can be computed by spatial MV prediction or temporal MV prediction. The difference between the predicted MVs and the motion compensation MVs (MC MVs) of the current frame (residual motion data) are encoded into the bitstream 495 by the entropy encoder 490.

15 [0084] The entropy encoder 490 encodes various parameters and data into the bitstream 495 by using entropy-coding techniques such as context-adaptive binary arithmetic coding (CABAC) or Huffman encoding. The entropy encoder 490 encodes various header elements, flags, along with the quantized transform coefficients 412, and the residual motion data as syntax elements into the bitstream 495. The bitstream 495 is in turn stored in a storage device or transmitted to a decoder over a communications medium such as a network.

20 [0085] The in-loop filter 445 performs filtering or smoothing operations on the reconstructed pixel data 417 to reduce the artifacts of coding, particularly at boundaries of pixel blocks. In some embodiments, the filtering operation performed includes sample adaptive offset (SAO). In some embodiment, the filtering operations include adaptive loop filter (ALF).

[0086] FIG. 5 illustrates portions of the video encoder 400 that selectively entropy encodes syntax elements of a subblock of a block in either regular mode or bypass mode based on a constraint that limits the number of regular bins used for the subblock. The figure illustrates the components of the entropy encoder module 490.

25 [0087] A syntax element generator 502 generates syntax elements to be entropy coded, including syntax elements that represent transform coefficients of the current subblock that are based on the quantized coefficients 412. The syntax elements are generated according to predetermined coding passes (e.g., sig\_flag, gt1\_flag, gt2\_flag, and par\_flag in a first coding pass).

30 [0088] The selected syntax elements are converted into bin strings 510 by the binarizer 506 (for non-binary valued syntax elements) or directly used as bin strings (for binary valued syntax elements). A regular/bypass mode switch 508 determines whether the syntax elements bin strings 510 are entropy coded as regular bins or bypass bins. Regular bins are processed by context modeler 512 and regular encoding engine 514 into code bits 518 using context modeling. Bypass bins are processed by bypass encoding engine 516 into code bits 518 without using context modeling. The code bits 518 are stored as part of the bitstream 495.

35 [0089] The regular/bypass mode switch 508 is controlled by a regular/bypass mode selection module 520. The regular/bypass mode selection module makes its decision as to whether to select regular mode or bypass mode

by using a regular bins counter 522. The regular bins counter 522 counts the number of regular bins that are encoded for the current subblock, and the regular/bypass mode switch 508 allows bin string to be encoded as regular bins as long as the regular bins counter is not larger or equal to a constraint 524. Thus, the constraint 524 limits the number of regular bins for the current subblock or the current coding pass of the current subblock. The constraint 524 may be determined based on the size of the subblock, number of coded subblock in the current block, total number of subblock in the current block, color component of the current subblock, or other properties of the current block or subblock.

**[0090] FIG. 6** conceptually illustrates a process 600 that selectively entropy encodes syntax elements of a current block in either regular mode or bypass mode based on a constraint that limits the number of regular bins used for a transform block. In some embodiments, one or more processing units (e.g., a processor) of a computing device implementing the video encoder 400 performs the process 600 by executing instructions stored in a computer readable medium. In some embodiments, an electronic apparatus implementing the video encoder 400 performs the process 600.

**[0091]** The video encoder receives (at step 610) input data associated with syntax elements representing transform coefficients of a current block. The transform coefficients may be that of a transform block of the current block, or that of a subblock of the transform block. In some embodiments, the syntax elements are to be entropy encoded in multiple coding passes. For each transform coefficient, at least a first flag, a second flag, and a third flag indicating an absolute value of the transform coefficient are to be entropy encoded in a first coding pass. The first flag may indicate whether an absolute value of the transform coefficient level is greater than 0 (e.g., sig\_flag). The second flag may indicate whether the absolute value of the transform coefficient level is greater than 1 (e.g., gt1\_flag). The third flag may indicate whether the absolute value of the transform coefficient level is greater than 2 or 3 (e.g., gt2\_flag). In some embodiments, a parity of the absolute value of the transform coefficient (e.g., par\_flag) is also entropy coded in the first coding pass, while a remainder of the transform coefficient is entropy coded in a second coding pass, and a sign of the transform coefficient is entropy coded in a third coding pass. The regular bins of the transform coefficient are entropy coded in the first coding pass by using a context model that is determined based on the first flag, the second flag, the third flag, and the parity of the absolute value of the transform coefficient.

**[0092]** The video encoder identifies (at step 620) a constraint that limits a total number of regular bins used for entropy encoding the syntax elements. In some embodiments, the constraint is different for a first color component than for a second color component. In some embodiments, the constraint is determined based on the size of the transform block. In some embodiments, the constraint is computed by multiplying a default constraint with a predefined factor that is derived from a ratio of a number of coded subblocks in the current block and a number of all subblocks in the current block. In some embodiments, a total number of regular bins used for entropy coding the syntax elements in the first coding pass is limited by the constraint. In some embodiments, there is no constraint limiting a number of regular bins that is specific to an individual syntax element.

**[0093]** The video encoder selects or receives (at step 625) a syntax element as the current syntax element to be entropy encoded. The selection of the syntax element may be determined by the current coding pass. The video

encoder then determines (at step 630) whether to use bypass mode or regular mode for encoding a current syntax element. In some embodiments, the video decoder keeps track of the number or the count of regular bins that have already been used for encoding the syntax elements. The number of regular bins is compared with the constraint identified in the step 620, which limits a maximum allowed number of CABAC regular bins for the transform block, e.g., the maximum allowed numbers of CABAC regular bins for coding a subblock. If encoding the current syntax element as regular bins will exceed the constraint, the process proceeds to 650. If encoding the current syntax element as regular bins will not exceed the constraint, the process proceeds 640. In some embodiments, the video coder may code certain syntax elements as regular bins regardless of the constraint.

[0094] At step 640, the video encoder entropy encodes the current syntax element as regular bins using context modeling. The video encoder also updates (at step 645) the total number or count of regular bins used for entropy coding the syntax elements. The process 600 then proceeds to 660.

[0095] At step 650, the video encoder entropy encodes the current syntax element as bypass bins without using context modeling. The process then proceeds to 660.

[0096] At step 660, the video encoder determines whether there are more syntax elements to be entropy encoded for the current transform block, the current subblock, or for the current coding pass. If so, the process returns to step 625 to entropy encode more syntax elements. Otherwise, the process proceeds to 670 to store the entropy encoded syntax elements as code bits in a bitstream.

#### *IV. Example Video Decoder*

[0097] FIG. 7 illustrates an example video decoder 700. As illustrated, the video decoder 700 is an image-decoding or video-decoding circuit that receives a bitstream 795 and decodes the content of the bitstream into pixel data of video frames for display. The video decoder 700 has several components or modules for decoding the bitstream 795, including some components selected from an inverse quantization module 705, an inverse transform module 710, an intra-prediction module 725, a motion compensation module 730, an in-loop filter 745, a decoded picture buffer 750, a MV buffer 765, a MV prediction module 775, and a parser 790. The motion compensation module 730 is part of an inter-prediction module 740.

[0098] In some embodiments, the modules 710 – 790 are modules of software instructions being executed by one or more processing units (e.g., a processor) of a computing device. In some embodiments, the modules 710 – 790 are modules of hardware circuits implemented by one or more ICs of an electronic apparatus. Though the modules 710 – 790 are illustrated as being separate modules, some of the modules can be combined into a single module.

[0099] The parser 790 (or entropy decoder) receives the bitstream 795 and performs initial parsing according to the syntax defined by a video-coding or image-coding standard. The parsed syntax element includes various header elements, flags, as well as quantized data (or quantized coefficients) 712. The parser 790 parses out the various syntax elements by using entropy-coding techniques such as context-adaptive binary arithmetic coding (CABAC) or Huffman encoding.

[00100] The inverse quantization module 705 de-quantizes the quantized data (or quantized coefficients) 712 to obtain transform coefficients, and the inverse transform module 710 performs inverse transform on the

transform coefficients 716 to produce reconstructed residual signal 719. The reconstructed residual signal 719 is added with predicted pixel data 713 from the intra-prediction module 725 or the motion compensation module 730 to produce decoded pixel data 717. The decoded pixels data are filtered by the in-loop filter 745 and stored in the decoded picture buffer 750. In some embodiments, the decoded picture buffer 750 is a storage external to the video decoder 700. In some embodiments, the decoded picture buffer 750 is a storage internal to the video decoder 700.

**[00101]** The intra-prediction module 725 receives intra-prediction data from bitstream 795 and according to which, produces the predicted pixel data 713 from the decoded pixel data 717 stored in the decoded picture buffer 750. In some embodiments, the decoded pixel data 717 is also stored in a line buffer (not illustrated) for intra-picture prediction and spatial MV prediction.

**[00102]** In some embodiments, the content of the decoded picture buffer 750 is used for display. A display device 755 either retrieves the content of the decoded picture buffer 750 for display directly or retrieves the content of the decoded picture buffer to a display buffer. In some embodiments, the display device receives pixel values from the decoded picture buffer 750 through a pixel transport.

**[00103]** The motion compensation module 730 produces predicted pixel data 713 from the decoded pixel data 717 stored in the decoded picture buffer 750 according to motion compensation MVs (MC MVs). These motion compensation MVs are decoded by adding the residual motion data received from the bitstream 795 with predicted MVs received from the MV prediction module 775.

**[00104]** The MV prediction module 775 generates the predicted MVs based on reference MVs that were generated for decoding previous video frames, e.g., the motion compensation MVs that were used to perform motion compensation. The MV prediction module 775 retrieves the reference MVs of previous video frames from the MV buffer 765. The video decoder 700 stores the motion compensation MVs generated for decoding the current video frame in the MV buffer 765 as reference MVs for producing predicted MVs.

**[00105]** The in-loop filter 745 performs filtering or smoothing operations on the decoded pixel data 717 to reduce the artifacts of coding, particularly at boundaries of pixel blocks. In some embodiments, the filtering operation performed includes sample adaptive offset (SAO). In some embodiment, the filtering operations include adaptive loop filter (ALF).

**[00106] FIG. 8** illustrates portions of the video decoder 700 that selectively entropy decodes syntax elements of a subblock of a block in either regular mode or bypass mode based on a constraint that limits the number of regular bins used for the subblock. The figure illustrates the components of the entropy decoder module 790.

**[00107]** The entropy decoder 790 receives code bits 818 from the bitstream 795. A regular/bypass mode switch 808 determines whether the code bits are to be decoded as regular bins or bypass bins. Code bits that are to be decoded as regular bins are processed by context modeler 812 and a regular decoding engine 814 into bin strings 810 using context modeling. Code bits that are to be decoded as bypass bins are processed by a bypass coding engine 816 into bin strings 810 without using context modeling.

**[00108]** The bin strings 810 are converted into syntax elements by the de-binarizer 806 (for non-binary valued syntax elements) or directly used as syntax elements (for binary valued syntax elements). A syntax element converter 802 converts the syntax elements into signals or values to be used by various parts of the video decoder

700, including converting syntax elements of transform coefficient (e.g., sig\_flag, gt1\_flag, gt2\_flag, and par\_flag) into quantized coefficients 712 for the inverse quantizer 705. Syntax elements are converted according to their orderings specified by predetermined coding passes.

[00109] The regular/bypass mode switch 808 is controlled by a regular/bypass mode selection module 820. The regular/bypass mode selection module makes its decision as to whether to select regular mode or bypass mode by using a regular bins counter 822. The regular bins counter 822 counts the number of regular bins that are decoded for the current subblock, and the regular/bypass mode switch 808 allows code bits to be decoded as regular bins as long as the regular bins counter is not larger or equal to a constraint 824. Thus, the constraint 824 limits the number of regular bins for the current subblock or the current coding pass of the current subblock. The constraint 824 may be determined based on the size of the subblock, number of coded subblock in the current block, total number of subblock in the current block, color component of the current subblock, or other properties of the current block and subblock.

[00110] FIG. 9 conceptually illustrates a process 900 that selectively entropy decodes syntax elements of a subblock of a block in either regular mode or bypass mode based on a constraint that limits the number of regular bins used for a transform block. In some embodiments, one or more processing units (e.g., a processor) of a computing device implementing the video decoder 700 performs the process 900 by executing instructions stored in a computer readable medium. In some embodiments, an electronic apparatus implementing the video decoder 700 performs the process 900.

[00111] The video decoder receives (at step 910) code bits from a bitstream that correspond to a plurality of syntax elements representing transform coefficients of a current block. The transform coefficients may be that of a transform block of the current block, or that of a subblock of the transform block. In some embodiments, the syntax elements are to be entropy decoded in multiple coding passes. For each transform coefficient, at least a first flag, a second flag, and a third flag indicating an absolute value of the transform coefficient are to be entropy decoded in a first coding pass. The first flag may indicate whether an absolute value of the transform coefficient level is greater than 0 (e.g., sig\_flag). The second flag may indicate whether the absolute value of the transform coefficient level is greater than 1 (e.g., gt1\_flag). The third flag may indicate whether the absolute value of the transform coefficient level is greater than 2 or 3 (e.g., gt2\_flag). In some embodiments, a parity of the absolute value of the transform coefficient (e.g., par\_flag) is also entropy coded in the first coding pass, while a remainder of the transform coefficient is entropy coded in a second coding pass, and a sign of the transform coefficient is entropy coded in a third coding pass. The regular bins of the transform coefficient are entropy coded in the first coding pass by using a context model that is determined based on the first flag, the second flag, the third flag, and the parity of the absolute value of the transform coefficient.

[00112] The video decoder identifies (at step 920) a constraint that limits a total number of regular bins used for entropy decoding the syntax elements. In some embodiments, the constraint is different for a first color component than for a second color component. In some embodiments, the constraint is determined based on a size of the transform block. In some embodiments, the constraint is computed by multiplying a default constraint with a predefined factor that is derived from a ratio of a number of coded subblocks in the current block and a number of

all subblocks in the current block. In some embodiments, a total number of regular bins used for entropy decoding the syntax elements in the first coding pass is limited by the constraint. In some embodiments, there is no constraint limiting a number of regular bins that is specific to an individual syntax element.

5 [00113] The video decoder selects or receives (at step 925) a syntax element as the current syntax element to be entropy decoded from the code bits of the bitstream. The selection of the syntax element may be determined by the current coding pass. The video decoder then determines (at step 930) whether to use bypass mode or regular mode for decoding a current syntax element. In some embodiments, the video decoder keeps track of the number or the count of regular bins that have already been used for decoding the syntax elements. The number of regular bins is compared with the constraint identified in step 920, which limits a maximum allowed number of CABAC  
10 regular bins for the transform block, e.g., the maximum allowed numbers of CABAC regular bins for coding a subblock.

[00114] If decoding the current syntax element as regular bins will exceed the constraint, the process proceeds to 950. If decoding the current syntax element as regular bins will not exceed the constraint, the process proceeds 940. In some embodiments, the video coder may decode certain syntax elements as regular bins  
15 regardless of the constraint.

[00115] At step 940, the video decoder entropy decodes the current syntax element as regular bins using context modeling. The video decoder also updates (at step 945) the total number or count of regular bins used for entropy decoding the syntax elements. The process 900 then proceeds to 960.

20 [00116] At step 950, the video decoder entropy decodes the current syntax element as bypass bins without using context modeling. The process then proceeds to 960.

[00117] At step 960, the video decoder determines whether there are more syntax elements for the current subblock or the current coding pass of the current subblock to be entropy decoded. If so, the process returns to step 925 to entropy decode more syntax elements. Otherwise, the process proceeds to 970 to reconstruct the current block using entropy decoded syntax elements such as transform coefficients.

#### 25 *V. Example Electronic System*

[00118] Many of the above-described features and applications are implemented as software processes that are specified as a set of instructions recorded on a computer readable storage medium (also referred to as computer readable medium). When these instructions are executed by one or more computational or processing unit(s) (e.g., one or more processors, cores of processors, or other processing units), they cause the processing unit(s) to  
30 perform the actions indicated in the instructions. Examples of computer readable media include, but are not limited to, CD-ROMs, flash drives, random-access memory (RAM) chips, hard drives, erasable programmable read only memories (EPROMs), electrically erasable programmable read-only memories (EEPROMs), etc. The computer readable media does not include carrier waves and electronic signals passing wirelessly or over wired connections.

[00119] In this specification, the term “software” is meant to include firmware residing in read-only memory  
35 or applications stored in magnetic storage which can be read into memory for processing by a processor. Also, in some embodiments, multiple software inventions can be implemented as sub-parts of a larger program while remaining distinct software inventions. In some embodiments, multiple software inventions can also be

implemented as separate programs. Finally, any combination of separate programs that together implement a software invention described here is within the scope of the present disclosure. In some embodiments, the software programs, when installed to operate on one or more electronic systems, define one or more specific machine implementations that execute and perform the operations of the software programs. In some embodiments, a coding operation means an encoding operation or a decoding operation.

**[00120] FIG. 10** conceptually illustrates an electronic system 1000 with which some embodiments of the present disclosure are implemented. The electronic system 1000 may be a computer (e.g., a desktop computer, personal computer, tablet computer, etc.), phone, PDA, or any other sort of electronic device. Such an electronic system includes various types of computer readable media and interfaces for various other types of computer readable media. Electronic system 1000 includes a bus 1005, processing unit(s) 1010, a graphics-processing unit (GPU) 1015, a system memory 1020, a network 1025, a read-only memory 1030, a permanent storage device 1035, input devices 1040, and output devices 1045.

**[00121]** The bus 1005 collectively represents all system, peripheral, and chipset buses that communicatively connect the numerous internal devices of the electronic system 1000. For instance, the bus 1005 communicatively connects the processing unit(s) 1010 with the GPU 1015, the read-only memory 1030, the system memory 1020, and the permanent storage device 1035.

**[00122]** From these various memory units, the processing unit(s) 1010 retrieves instructions to execute and data to process in order to execute the processes of the present disclosure. The processing unit(s) may be a single processor or a multi-core processor in different embodiments. Some instructions are passed to and executed by the GPU 1015. The GPU 1015 can offload various computations or complement the image processing provided by the processing unit(s) 1010.

**[00123]** The read-only-memory (ROM) 1030 stores static data and instructions that are used by the processing unit(s) 1010 and other modules of the electronic system. The permanent storage device 1035, on the other hand, is a read-and-write memory device. This device is a non-volatile memory unit that stores instructions and data even when the electronic system 1000 is off. Some embodiments of the present disclosure use a mass-storage device (such as a magnetic or optical disk and its corresponding disk drive) as the permanent storage device 1035.

**[00124]** Other embodiments use a removable storage device (such as a floppy disk, flash memory device, etc., and its corresponding disk drive) as the permanent storage device. Like the permanent storage device 1035, the system memory 1020 is a read-and-write memory device. However, unlike storage device 1035, the system memory 1020 is a volatile read-and-write memory, such a random access memory. The system memory 1020 stores some of the instructions and data that the processor uses at runtime. In some embodiments, processes in accordance with the present disclosure are stored in the system memory 1020, the permanent storage device 1035, and/or the read-only memory 1030. For example, the various memory units include instructions for processing multimedia clips in accordance with some embodiments. From these various memory units, the processing unit(s) 1010 retrieves instructions to execute and data to process in order to execute the processes of some embodiments.

[00125] The bus 1005 also connects to the input and output devices 1040 and 1045. The input devices 1040 enable the user to communicate information and select commands to the electronic system. The input devices 1040 include alphanumeric keyboards and pointing devices (also called “cursor control devices”), cameras (e.g., webcams), microphones or similar devices for receiving voice commands, etc. The output devices 1045 display images generated by the electronic system or otherwise output data. The output devices 1045 include printers and display devices, such as cathode ray tubes (CRT) or liquid crystal displays (LCD), as well as speakers or similar audio output devices. Some embodiments include devices such as a touchscreen that function as both input and output devices.

[00126] Finally, as shown in FIG. 10, bus 1005 also couples electronic system 1000 to a network 1025 through a network adapter (not shown). In this manner, the computer can be a part of a network of computers (such as a local area network (“LAN”), a wide area network (“WAN”), or an Intranet, or a network of networks, such as the Internet. Any or all components of electronic system 1000 may be used in conjunction with the present disclosure.

[00127] Some embodiments include electronic components, such as microprocessors, storage and memory that store computer program instructions in a machine-readable or computer-readable medium (alternatively referred to as computer-readable storage media, machine-readable media, or machine-readable storage media). Some examples of such computer-readable media include RAM, ROM, read-only compact discs (CD-ROM), recordable compact discs (CD-R), rewritable compact discs (CD-RW), read-only digital versatile discs (e.g., DVD-ROM, dual-layer DVD-ROM), a variety of recordable/rewritable DVDs (e.g., DVD-RAM, DVD-RW, DVD+RW, etc.), flash memory (e.g., SD cards, mini-SD cards, micro-SD cards, etc.), magnetic and/or solid state hard drives, read-only and recordable Blu-Ray® discs, ultra-density optical discs, any other optical or magnetic media, and floppy disks. The computer-readable media may store a computer program that is executable by at least one processing unit and includes sets of instructions for performing various operations. Examples of computer programs or computer code include machine code, such as is produced by a compiler, and files including higher-level code that are executed by a computer, an electronic component, or a microprocessor using an interpreter.

[00128] While the above discussion primarily refers to microprocessor or multi-core processors that execute software, many of the above-described features and applications are performed by one or more integrated circuits, such as application specific integrated circuits (ASICs) or field programmable gate arrays (FPGAs). In some embodiments, such integrated circuits execute instructions that are stored on the circuit itself. In addition, some embodiments execute software stored in programmable logic devices (PLDs), ROM, or RAM devices.

[00129] As used in this specification and any claims of this application, the terms “computer”, “server”, “processor”, and “memory” all refer to electronic or other technological devices. These terms exclude people or groups of people. For the purposes of the specification, the terms display or displaying means displaying on an electronic device. As used in this specification and any claims of this application, the terms “computer readable medium,” “computer readable media,” and “machine readable medium” are entirely restricted to tangible, physical objects that store information in a form that is readable by a computer. These terms exclude any wireless signals, wired download signals, and any other ephemeral signals.

[00130] While the present disclosure has been described with reference to numerous specific details, one of ordinary skill in the art will recognize that the present disclosure can be embodied in other specific forms without departing from the spirit of the present disclosure. In addition, a number of the figures (including FIG. 6 and FIG. 9) conceptually illustrate processes. The specific operations of these processes may not be performed in the exact order shown and described. The specific operations may not be performed in one continuous series of operations, and different specific operations may be performed in different embodiments. Furthermore, the process could be implemented using several sub-processes, or as part of a larger macro process. Thus, one of ordinary skill in the art would understand that the present disclosure is not to be limited by the foregoing illustrative details, but rather is to be defined by the appended claims.

10 *Additional Notes*

[00131] The herein-described subject matter sometimes illustrates different components contained within, or connected with, different other components. It is to be understood that such depicted architectures are merely examples, and that in fact many other architectures can be implemented which achieve the same functionality. In a conceptual sense, any arrangement of components to achieve the same functionality is effectively "associated" such that the desired functionality is achieved. Hence, any two components herein combined to achieve a particular functionality can be seen as "associated with" each other such that the desired functionality is achieved, irrespective of architectures or intermediate components. Likewise, any two components so associated can also be viewed as being "operably connected", or "operably coupled", to each other to achieve the desired functionality, and any two components capable of being so associated can also be viewed as being "operably couplable", to each other to achieve the desired functionality. Specific examples of operably couplable include but are not limited to physically mateable and/or physically interacting components and/or wirelessly interactable and/or wirelessly interacting components and/or logically interacting and/or logically interactable components.

[00132] Further, with respect to the use of substantially any plural and/or singular terms herein, those having skill in the art can translate from the plural to the singular and/or from the singular to the plural as is appropriate to the context and/or application. The various singular/plural permutations may be expressly set forth herein for sake of clarity.

[00133] Moreover, it will be understood by those skilled in the art that, in general, terms used herein, and especially in the appended claims, e.g., bodies of the appended claims, are generally intended as "open" terms, e.g., the term "including" should be interpreted as "including but not limited to," the term "having" should be interpreted as "having at least," the term "includes" should be interpreted as "includes but is not limited to," etc. It will be further understood by those within the art that if a specific number of an introduced claim recitation is intended, such an intent will be explicitly recited in the claim, and in the absence of such recitation no such intent is present. For example, as an aid to understanding, the following appended claims may contain usage of the introductory phrases "at least one" and "one or more" to introduce claim recitations. However, the use of such phrases should not be construed to imply that the introduction of a claim recitation by the indefinite articles "a" or "an" limits any particular claim containing such introduced claim recitation to implementations containing only one such recitation, even when the same claim includes the introductory phrases "one or more" or "at least one"

and indefinite articles such as "a" or "an," e.g., "a" and/or "an" should be interpreted to mean "at least one" or "one or more;" the same holds true for the use of definite articles used to introduce claim recitations. In addition, even if a specific number of an introduced claim recitation is explicitly recited, those skilled in the art will recognize that such recitation should be interpreted to mean at least the recited number, e.g., the bare recitation of "two recitations," without other modifiers, means at least two recitations, or two or more recitations. Furthermore, in those instances where a convention analogous to "at least one of A, B, and C, etc." is used, in general such a construction is intended in the sense one having skill in the art would understand the convention, e.g., "a system having at least one of A, B, and C" would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc. In those instances where a convention analogous to "at least one of A, B, or C, etc." is used, in general such a construction is intended in the sense one having skill in the art would understand the convention, e.g., "a system having at least one of A, B, or C" would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc. It will be further understood by those within the art that virtually any disjunctive word and/or phrase presenting two or more alternative terms, whether in the description, claims, or drawings, should be understood to contemplate the possibilities of including one of the terms, either of the terms, or both terms. For example, the phrase "A or B" will be understood to include the possibilities of "A" or "B" or "A and B."

**[00134]** From the foregoing, it will be appreciated that various implementations of the present disclosure have been described herein for purposes of illustration, and that various modifications may be made without departing from the scope and spirit of the present disclosure. Accordingly, the various implementations disclosed herein are not intended to be limiting, with the true scope and spirit being indicated by the following claims.

## CLAIMS

1. A method comprising:  
receiving input data associated with a plurality of syntax elements representing transform coefficients of a  
5 current block; and  
entropy coding the plurality of syntax elements selectively as either regular bins using context modeling or as  
bypass bins without context modeling, wherein a number of the regular bins used for entropy coding the plurality  
of syntax elements is limited by a constraint.
- 10 2. The method of claim 1, wherein the plurality of syntax elements are coded in a plurality of coding passes,  
and wherein a total number of the regular bins used for entropy coding the syntax elements in a first coding pass is  
limited by the constraint.
- 15 3. The method of claim 2, wherein an absolute value for a transform coefficient is indicated by one or more  
of a first flag, a second flag, and a third flag that are entropy coded in the first coding pass, and wherein:  
the first flag indicates whether an absolute value of the transform coefficient is greater than 0,  
the second flag indicates whether the absolute value of the transform coefficient is greater than 1, and  
the third flag indicates whether the absolute value of the transform coefficient is greater than 3.
- 20 4. The method of claim 3, wherein:  
a parity of the absolute value of the transform coefficient is entropy coded in the first coding pass,  
a remainder of the transform coefficient is entropy coded in a second coding pass, and  
a sign of the transform coefficient is entropy coded in a third coding pass.
- 25 5. The method of claim 4, wherein the regular bins of the first coding pass are entropy coded by using one or  
more context models that are dependent on one or more indicators of absolute values of one or more neighboring  
transform coefficients.
- 30 6. The method of claim 5, wherein at least one context model is selected conditioned on a value of the third  
flag.
7. The method of claim 1, wherein the plurality of syntax elements representing transform coefficients of a  
transform block.
- 35 8. The method of claim 7, wherein the constraint limiting the number of the regular bins is determined based  
on a size of the transform block.

9. The method of claim 1, wherein the constraint limiting the number of the regular bins is different for a first color component than for a second color component.

5 10. The method of claim 1, wherein the constraint limiting the number of the regular bins is computed by multiplying a default constraint with a predefined factor that is derived from a ratio of a number of coded subblocks in the current block and a number of all subblocks in the current block.

10 11. The method of claim 1, wherein there is no constraint limiting a number of the regular bins that is specific to an individual syntax element.

12. The method of claim 1, wherein the constraint limiting the number of the regular bins limits a number of regular bins used for a subblock of the current block.

15 13. An electronic apparatus comprising:  
a decoder circuit capable of:

receiving from a bitstream code bits that correspond to a plurality of syntax elements representing transform coefficients of a current block;

20 entropy decoding the code bits into the plurality of syntax elements selectively as either regular bins using context modeling or as bypass bins without context modeling, wherein a number of the regular bins used for entropy decoding the plurality of syntax elements is limited by a constraint; and

reconstructing the current block using the entropy decoded syntax elements.

14. An electronic apparatus comprising:

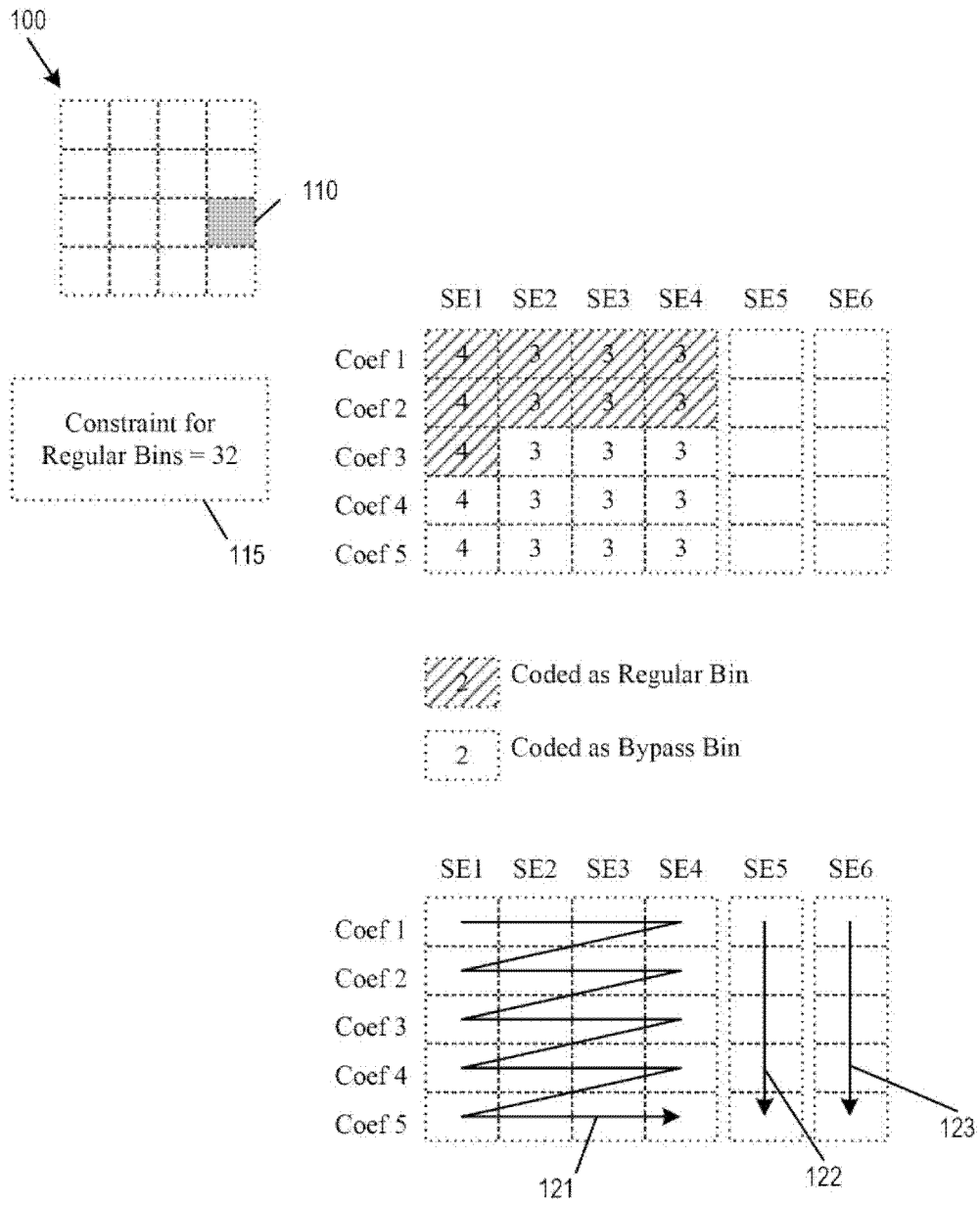
an encoder circuit capable of:

25 receiving a plurality of syntax elements representing transform coefficients of a current block;

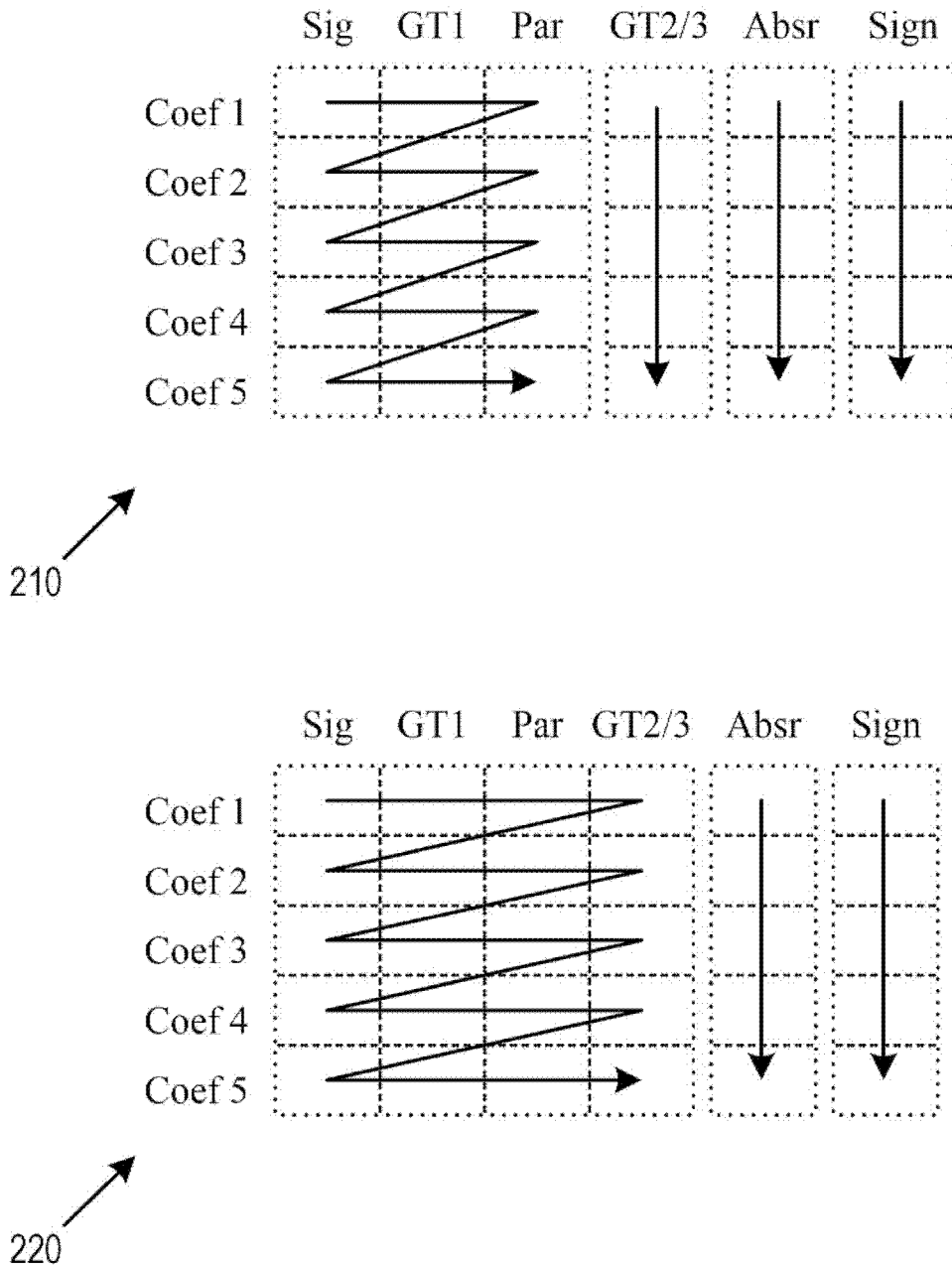
entropy encoding the plurality of syntax elements selectively as either regular bins using context modeling or as bypass bins without context modeling, wherein a number of the regular bins used for entropy encoding the plurality of syntax elements is limited by a constraint; and

storing the entropy encoded syntax elements in a bitstream.

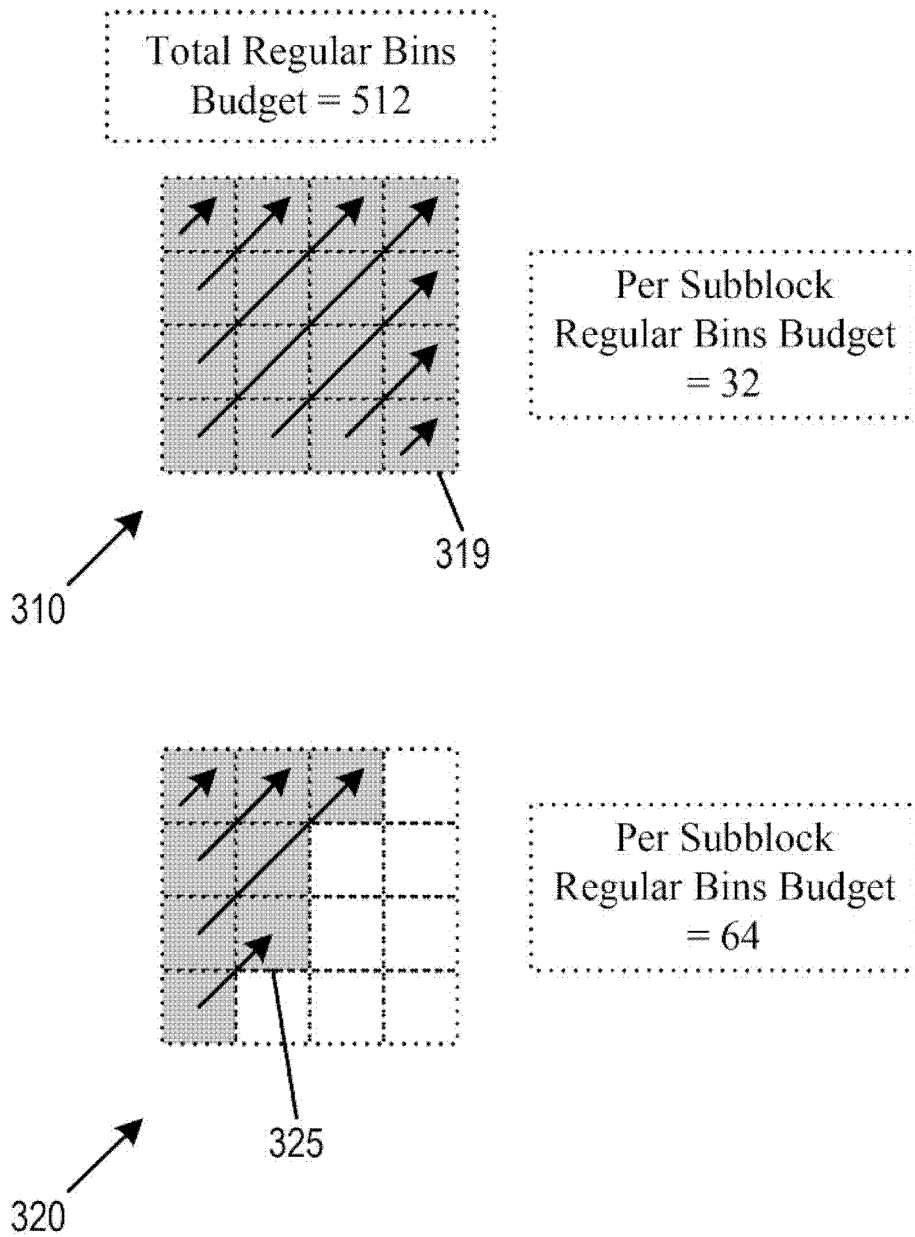
30



**FIG. 1**



**FIG. 2**



**FIG. 3**

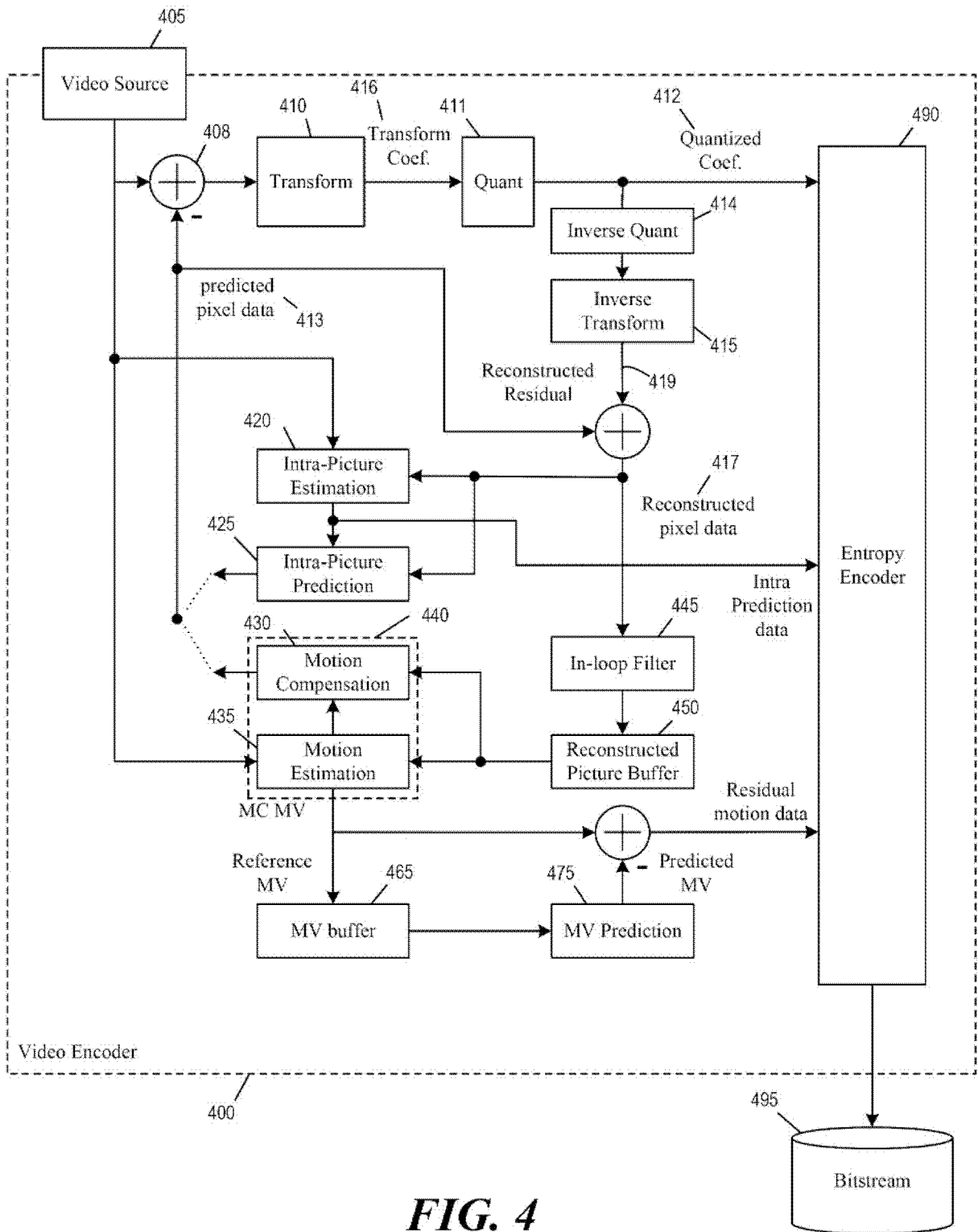


FIG. 4

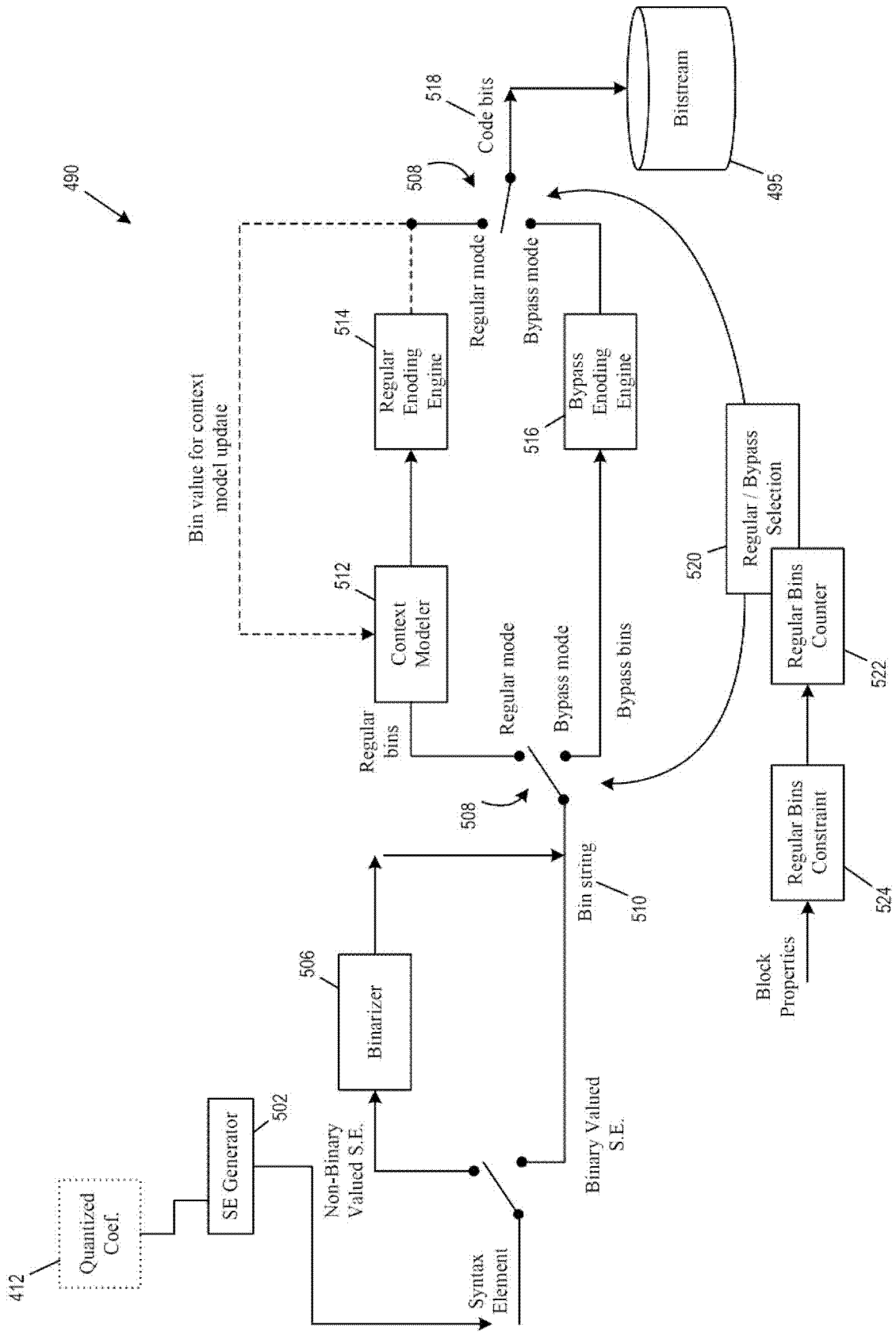
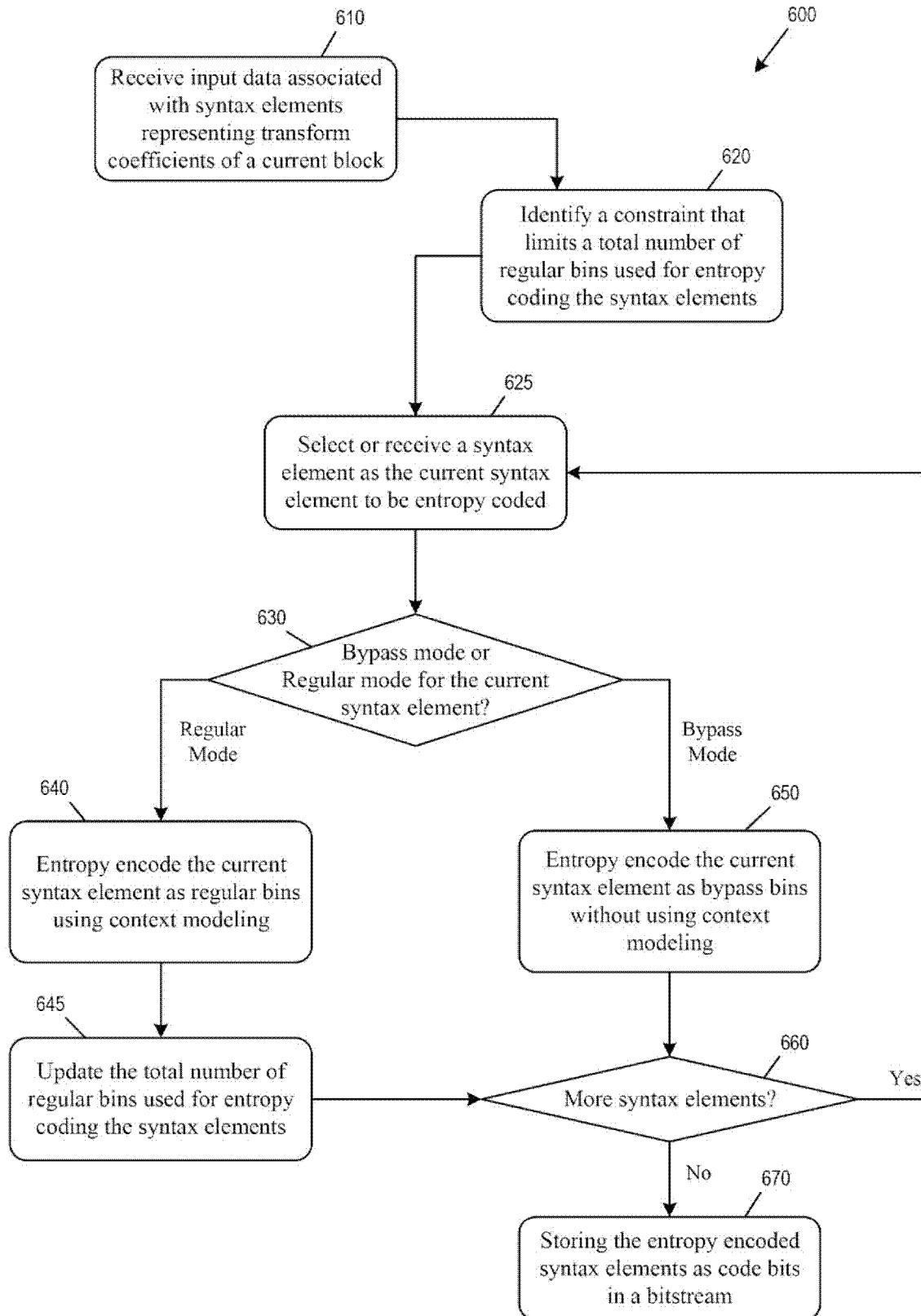


FIG. 5



**FIG. 6**

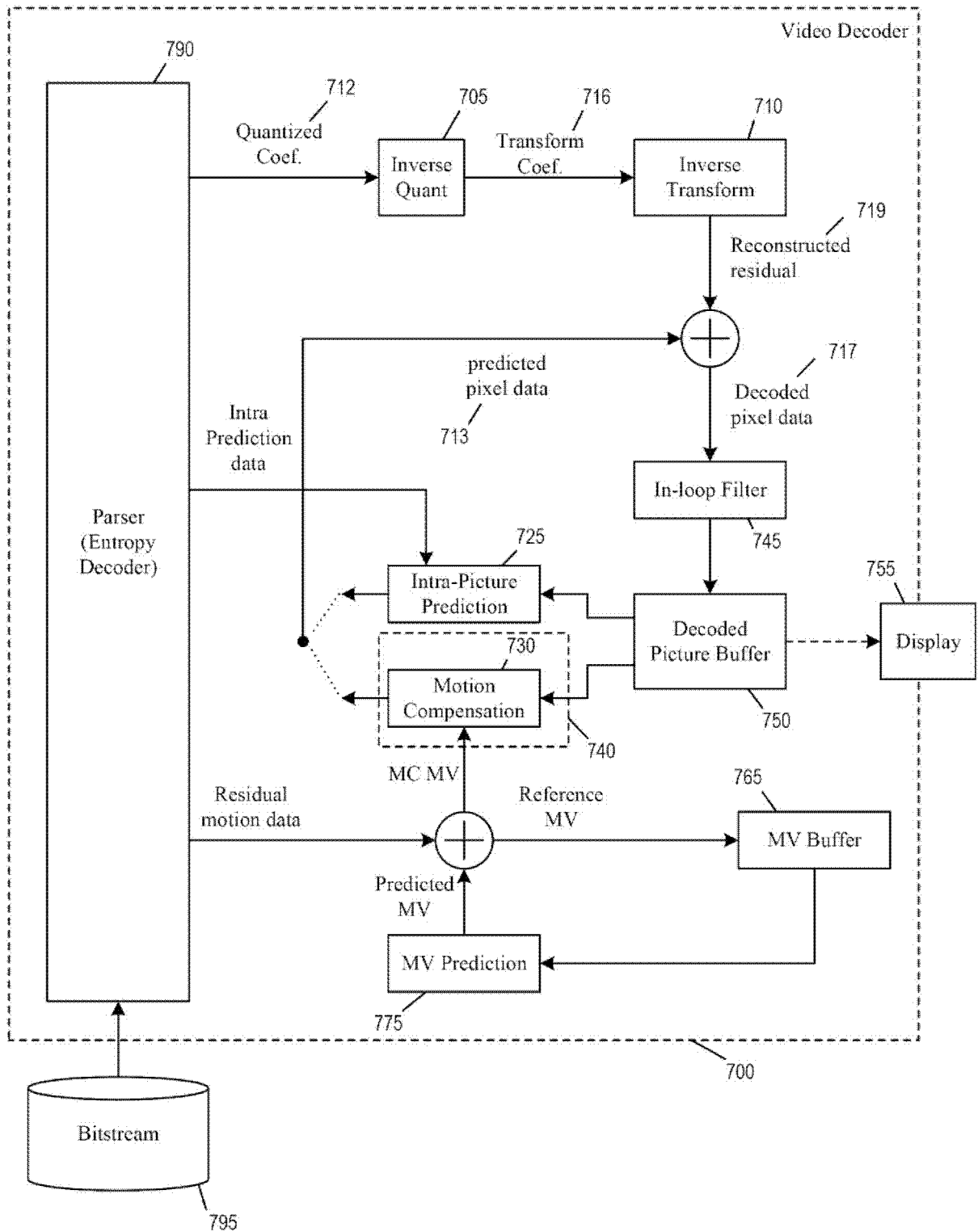


FIG. 7

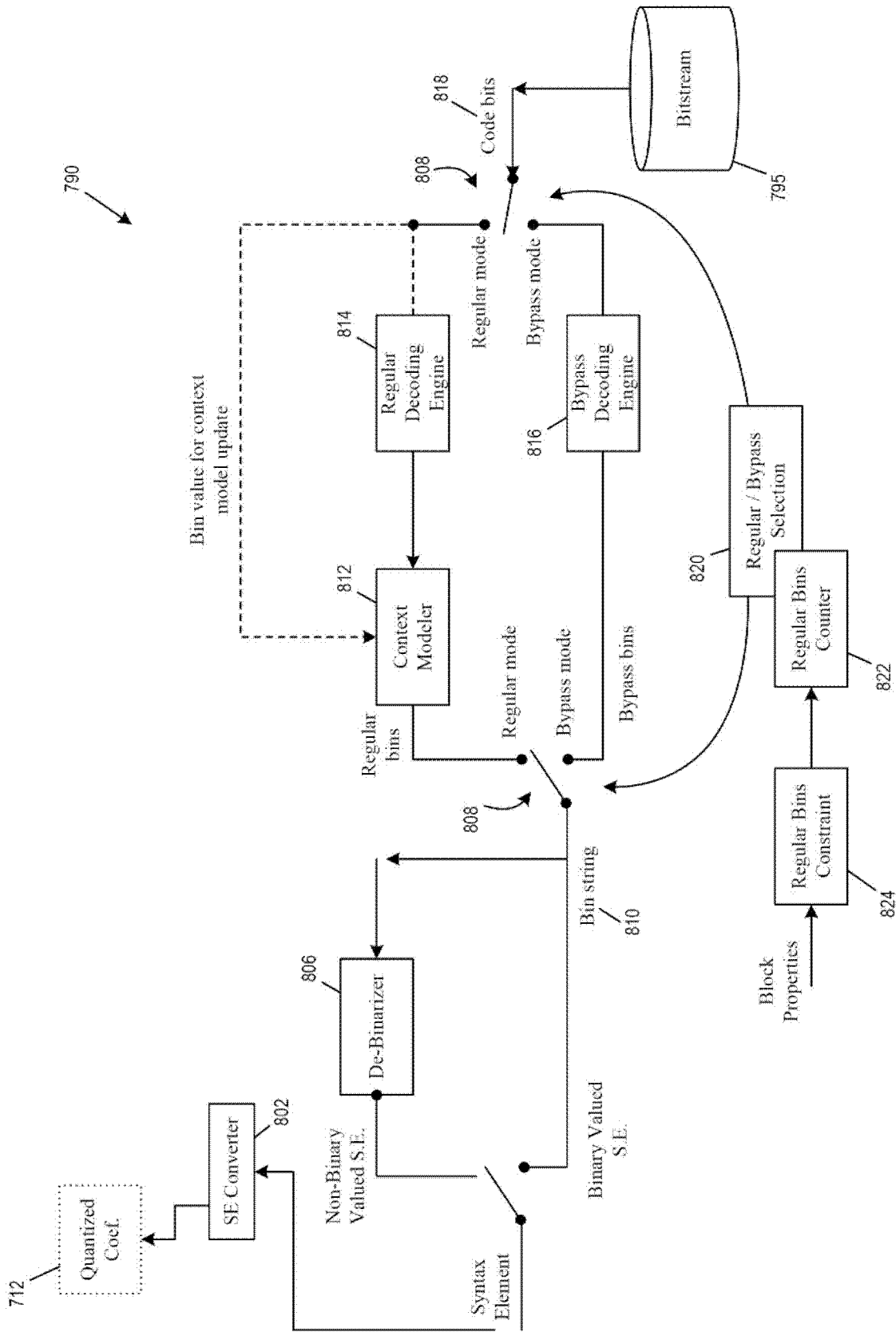
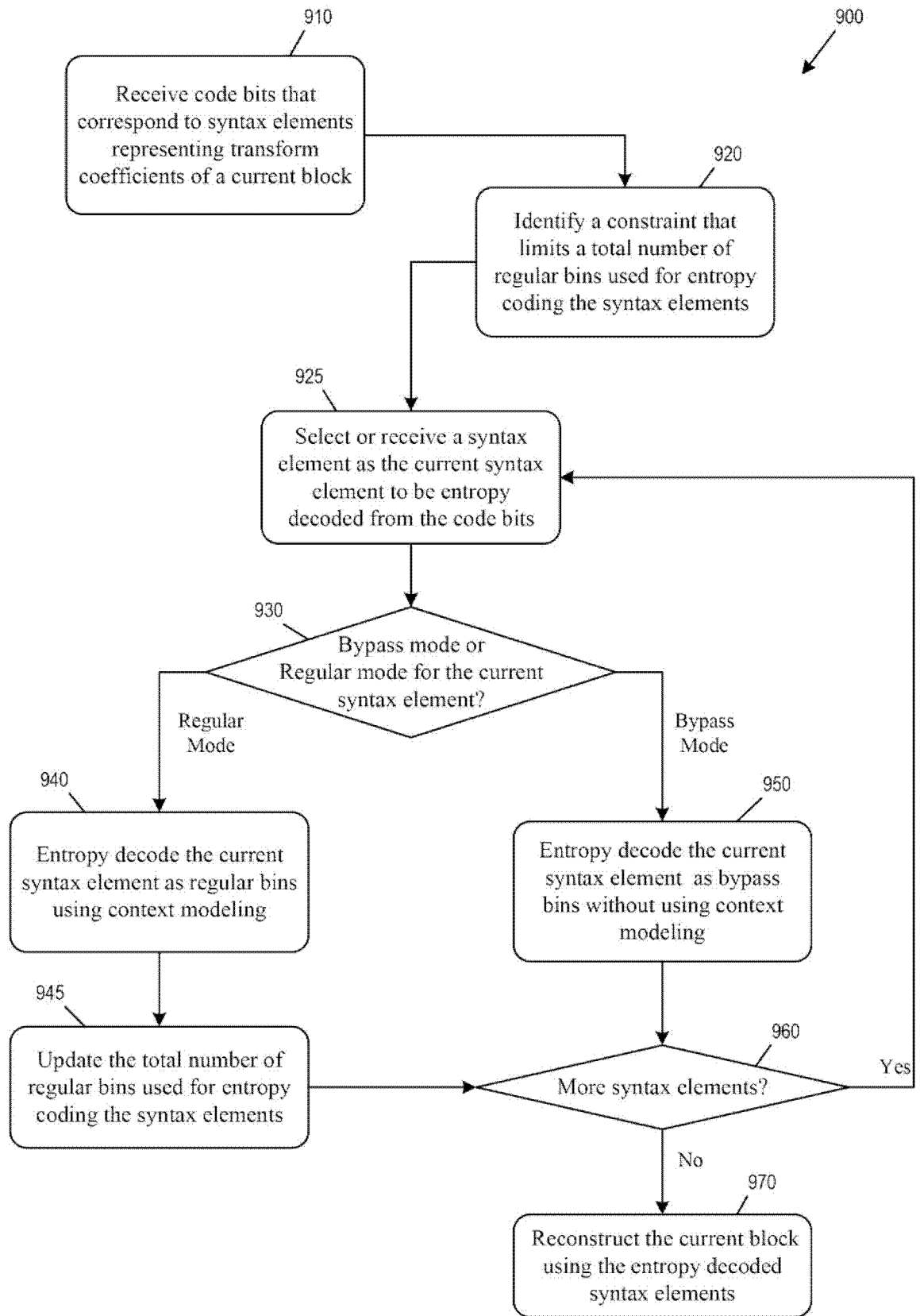
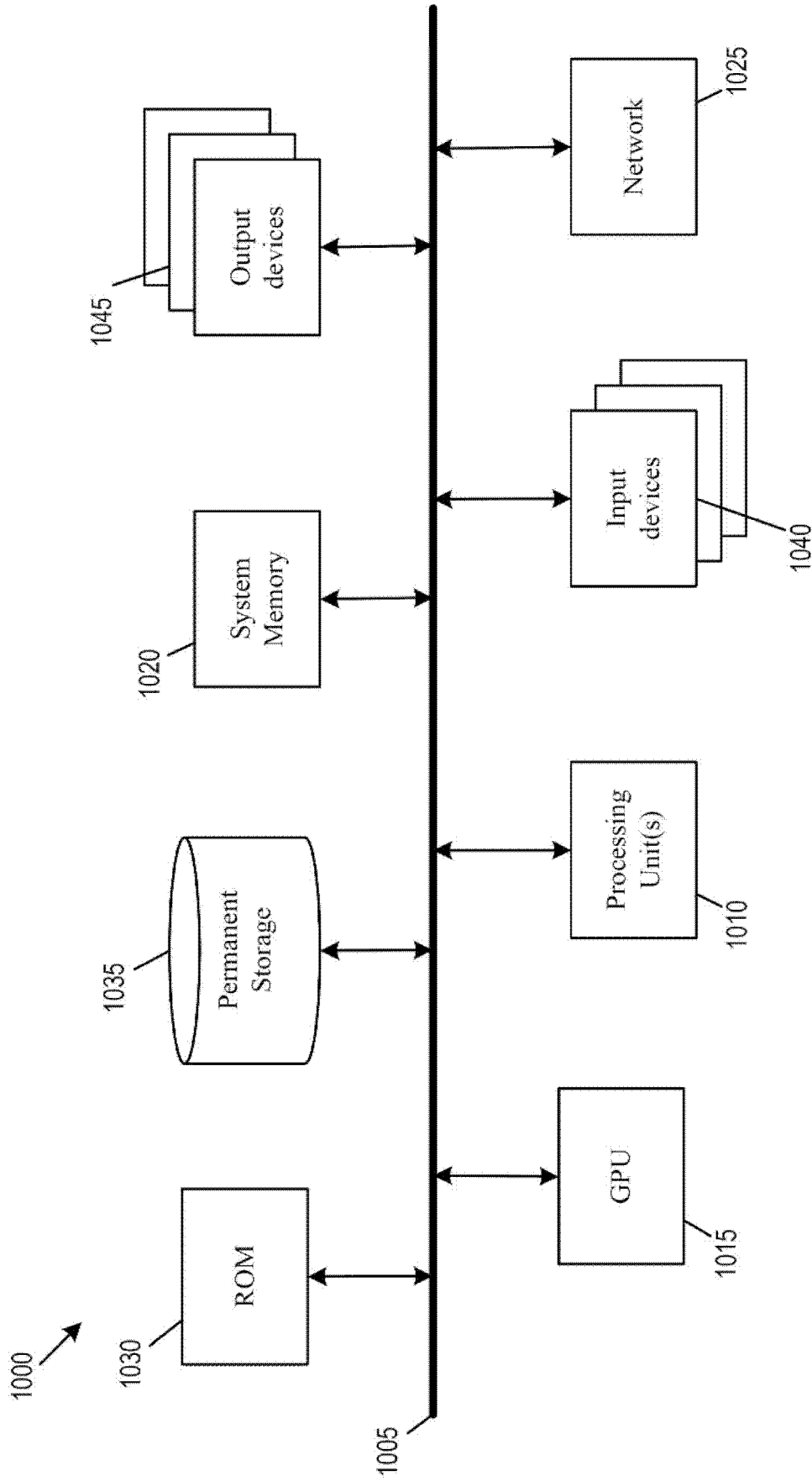


FIG. 8



**FIG. 9**



**FIG. 10**

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/102284

**A. CLASSIFICATION OF SUBJECT MATTER**

H04N 19/00(2014.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC: bypass, regular, bin, syntax element, entropy coding, encoding, transform coefficient, context modeling, constraint, number, threshold

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2017041676 A1 (MEDIATEK SINGAPORE PTE. LTD. ET AL.) 16 March 2017 (2017-03-16) description, paragraphs [0005] - [0022]	1-14
A	CN 103931188 A (MEDIATEK SINGAPORE PTE. LTD.) 16 July 2014 (2014-07-16) the whole document	1-14
A	WO 2013017092 A1 (MEDIATEK INC. ET AL.) 07 February 2013 (2013-02-07) the whole document	1-14
A	US 2016345030 A1 (QUALCOMM INCORPORATED) 24 November 2016 (2016-11-24) the whole document	1-14
A	US 2015092833 A1 (APPLE INC.) 02 April 2015 (2015-04-02) the whole document	1-14

 Further documents are listed in the continuation of Box C. See patent family annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&amp;" document member of the same patent family

Date of the actual completion of the international search

24 October 2019

Date of mailing of the international search report

21 November 2019

Name and mailing address of the ISA/CN

National Intellectual Property Administration, PRC  
6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing  
100088  
China

Authorized officer

ZHANG,Dezhen

Facsimile No. (86-10)62019451

Telephone No. 86-(10)-53961658

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2019/102284**

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
WO	2017041676	A1	16 March 2017	CN	108028919	A	11 May 2018
				WO	2017041271	A1	16 March 2017
				KR	20180044969	A	03 May 2018
				EP	3335421	A1	20 June 2018
				US	2018288416	A1	04 October 2018
-----							
CN	103931188	A	16 July 2014	None			
-----							
WO	2013017092	A1	07 February 2013	CN	106101727	A	09 November 2016
				US	2014192861	A1	10 July 2014
				EP	2988511	A1	24 February 2016
				CN	103597834	A	19 February 2014
				US	2017041022	A1	09 February 2017
				EP	2740272	A1	11 June 2014
				MX	2013015212	A	21 March 2014
-----							
US	2016345030	A1	24 November 2016	EP	3298778	A2	28 March 2018
				JP	2018520562	A	26 July 2018
				WO	2016187567	A2	24 November 2016
				CN	107637080	A	26 January 2018
-----							
US	2015092833	A1	02 April 2015	None			
-----							