An Efficient Architecture of Intra Prediction and TQ/ IQIT Module of Video Encoder

Kibum Suh*

Department of Rail Electrical System, Woosong University, Deajeon, South Korea; kbsuh@wsu.ac.kr

Abstract

Objectives: In this paper, an efficient architecture of intra prediction module of H.264 high profile encoder is proposed. This module can be operated in 308 cycles for one macroblock. **Methods/Statistical Analysis**: The plane mode removal and SAD (Sum of Absolute Difference) distortion calculation are adopted to reduce the hardware cost and cycle. The sharing method of the Q (Quantization) and IQ (Inverse Quantization) modules for I4MB and I8MB prediction, calculation method of the DC value of I16MB and chroma predictions in prediction cycles to speed up the macroblock processing cycle are proposed. **Findings**: The proposed hardware was verified with the vector generated by reference C using JM13.2. The designed circuit has 250 K gate counts by using TSMC 0.18 um process including SRAM memory and can operate in 160 MHz clock. **Improvements/Applications**: The cycle for one macroblock is reduced compared with other architectures.

Keywords: High Profile, Integer Transform, Intra-Prediction, Inverse Integer Transforms, Inverse Quantization, Quantization

1. Introduction

Video coding standard H.264/AVC, Advanced Video Coding (AVC) was proposed by the Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG. Intra prediction is one of the techniques employed to improve the coding efficiency of H.264/AVC. In the baseline and main profiles, intra prediction is performed with two block sizes for 13 prediction modes and the best intra mode is selected among the 13 modes for luma pixels. But in the high profile, intra prediction is performed with three block sizes ($4 \ge 4$, $16 \ge 16$ and $8 \ge 8$) for 22 prediction modes (9 modes for $4 \ge 4$ blocks, 9 modes for $8 \ge 8$ blocks and 4 modes for $16 \ge 16$

As a result, intra prediction requires large computation time, which is comparable to the computation amount of JPEG2000¹. Accordingly, there have been researches for the faster computation to reduce the time complexity of intra prediction.

A number of efficient algorithms have been proposed to speed up the computation of intra prediction¹⁻¹². One of the common approaches reduces the computational complexity of intra prediction by predicting subset of modes and evaluating the RD cost only for a subset of them. The evaluated prediction modes are selected based on local edge information^{1,13,14} or the prediction modes from neighboring blocks.

In these techniques, decided prediction modes may not be the optimal mode. In another technique for fast intra prediction, one of the 4 x 4 or 16 x 16 intra predictions is skipped¹¹. The smoothness of a macroblock is estimated and 16 x 16 intra prediction is chosen in the case of smooth macroblock. Otherwise, 4 x 4 intra predictions are performed. An early decision to skip intra prediction has been proposed for fast intra prediction³. Here, the results of inter prediction is used for the decision of skipped macroblock. For the skipped macroblock, 4 x 4 and 16 x 16 intra predictions can be eliminated and good compression efficiency can be obtained.

Intra prediction is often implemented by hardware. A number of studies have been published recently on the hardware implementation of intra predictions^{2.4,15-18}. The DC components of 4×4 DCT coefficients are pre-calculated when 16 x 16 intra prediction is performed⁴. This technique

^{*}Author for correspondence

enables pipelining of Transform and Quantization (TQ) and Inverse Quantization and Inverse Transform (IQIT) of the results from 16 x 16 intra prediction.

Redundant operation in predictor generation is minimized¹⁵ at the cost of increased execution time. In all these researches, the hardware resources for 4 x 4 intra prediction and reconstruction are often idle and wasted. Because intra prediction of a 4 x 4 block depends on the reconstructed pixels in its neighboring blocks, intra prediction cannot proceed before the completion of the reconstruction of the neighboring blocks. Since the reconstruction of neighboring pel for the I4MB and I8MB prediction is need for intra-prediction stage, DCT and IDCT and Q and IQ operation are parallel processed in the proposed architecture. In this architecture, Q and IQ module are shared by I4MB and I8MB by appropriate cycle scheduling. For the prediction cycle, only 208 cycles are needed. In previously published architecture⁴, the DC components of 4 x 4 DCT coefficients are precalculated when 16 x 16 intra prediction is performed. But, in this paper, we also propose the DC value calculation logic which can be shared for the SAD distortion and I16MB DC value calculation and chroma 8 x 8 DC value calculation. It can be used for the pipelining of TQ and IQIT of the results from 16 x 16 intra predictions and 8 x 8 chroma predictions. Therefore, hardware reduction of prediction for calculation logic of DC value is achieved.

2. The Proposed Hardware Architecture

The proposed hardware architecture of intra prediction, mode decision and TQ/IQIT block is shown in Figure 1. As the intra prediction utilizes neighboring pixels to generate the value of the predicted pixel, the neighboring pixel information has to be stored before prediction. These pixels are stored in intra prediction SRAM or intra prediction



Figure 1. The proposed intra prediction architecture.

flip-flops as shown in Figure 1, where the SRAM contains the horizontal pixel of encoded image and intra prediction flip-flops contains the vertical 32 pixels. This module performs the intra prediction of luma 8×8 , 4×4 , 16×16 and chroma 8×8 . As shown in the Figure, 4×4 , 8×8 and 16×16 prediction (chroma prediction) mode decision is done simultaneously. Each prediction module consists of prediction pel calculation logic, Sum of Absolute Error (SAE) calculation logic and mode selection logic. It computes SAE value and selects mode having the smallest SAE value among available modes (9 modes for the 4×4 and 8×8 predictions, 4 modes for the 16×16 and chroma 8×8 predictions). In the prediction cycle, the predicted values of each best mode are stored in 8×8 pred_sram, 4×4 pred_sram and 16×16 pred_sram.

After determining the selected mode among the I4MB, I16MB and I8MB mode, DCT/Q/IQ/IDCT operations are performed and the coefficient data to be coded is entered into VLC SRAM. The reconstructed pixels are stored into SRAM in De-blocking Filter (DB) module after Reconstruction (REC) process.

2.1 Luma 4 x 4 and 8 x 8 Intra Prediction Architecture

Figure 2 shows the architecture of luma 4 x 4 and 8 x 8 intra predictions. The prediction pel generation module generate predicted values for 9 prediction mode, using the pixels read from the intra prediction flip/flops, SAE calculation module computes Sum Of Absolute Difference value (SAD).

In the SAE calculation module, we adopt the SAD distortion measure rather than the hadamard distortion. In general, most of motion estimators are implemented in hardware based on the SAD measure. In our case, inter RD_cost could be only calculated with SAD since our motion estimator is based on SAD distortion measure. Since high profile encoder is more concentrated for B and



Figure 2. Intra 4 x 4 and 8 x 8 prediction module.

P slice, the deterioration of image quality using SAD distortion measure is smaller than that of baseline profile. 8 pixel parallelisms are used for processing to reduce cycle. The SAE calculation time takes two cycles for intra 4 x 4 and eight cycles for intra 8 x 8. The mode selection logic selects the best mode which has the smallest RD-cost. Timing diagrams for intra 4 x 4 and 8 x 8 prediction are given in Figures 3 and 4 respectively, Here, the intra 4 x 4 prediction take two cycles for SAE calculation, one cycle for best mode selection, seven cycles for TQ and IQIT calculation and two cycles for update. Since 13 cycles are needed for a 4 x 4 block, the total prediction takes 208 (13 x 16) cycles for one macroblock.

The intra 8 x 8 prediction take eight cycles for SAE calculation, one cycle for best mode selection and 28 cycles for TQ, IQIT with update. So, since 38 cycles are needed for one 8 x 8 block and as one macroblock has four 8 x 8 blocks, the total prediction takes 152 (38 x 4) cycles.

In the proposed intra prediction module, 4 x 4 predictions and 8 x 8 predictions are carried out simultaneously. For the 4 x 4 and 8 x 8 predictions, two different DCT modules are used but only one Q and IQ modules is used by proper scheduling of operation cycle of intra 8 x 8 and 4 x 4.



Figure 3. Intra 4 x 4 prediction and TQ/IQIT cycle.



Figure 4. Intra 8 x 8 prediction and TQ/IQIT cycle.

In the Table 1, the thickly shaded section indicate Q and IQ cycle of 4 x 4 and the lightly shaded section indicate Q and IQ cycle of 8 x 8.

2.2 Luma 16 x 16 and Chroma 8 x 8 Intra Prediction Architecture

In our previously published paper⁴, we proposed the method of solving the cycle overhead problem in the case of 16MB. For the post processing without overhead, the DC value is calculated in the I16MB prediction cycle. But in this paper, we propose very simple architecture to calculate DC value and apply for the chroma 8 x 8 DC value in addition. For the 16 x 16 prediction, SAE calculation can be expressed like Equation (1).

$$SAE_value = \sum_{y=0}^{15} \sum_{x=0}^{15} |org[x, y] - pred[x, y]|$$
(1)

In H.264, a 4×4 forward DCT transform is shown as:

$$Y = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \begin{bmatrix} x_{00} & x_{01} & x_{02} & x_{03} \\ x_{10} & x_{11} & x_{12} & x_{13} \\ x_{20} & x_{21} & x_{22} & x_{23} \\ x_{30} & x_{31} & x_{32} & x_{33} \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 1 \\ 1 & 1 & -1 & -2 \\ 1 & -1 & -1 & 2 \\ 1 & -2 & 1 & -1 \end{bmatrix} \right).$$
(2)

Where $x_{ii} = org[i, j] - pred[i, j]$.

From (2), we can obtain the DC value Y_{00} as shown in (3):

$$Y_{00} = (x_{00} + x_{01} + x_{02} + x_{03} + x_{10} + x_{11} + x_{12} + x_{13} + x_{20} + x_{21} + x_{22} + x_{23} + x_{30} + x_{31} + x_{32} + x_{33}).$$
 (3)

From the Equation (1) and (3), we can see that SAE value is sum of absolute difference and DC value is sum of difference. So we concluded that DC value and SAD value can be obtained by using the same hardware.

As shown in the Figure 5, the SAE calculation and DC_value calculation is calculated by using the mux, which select the output of the ABS (absolute value) logic and subtraction logic. The DC_value calculated for the 4 x 4 block is stored in the DC_register (16 coefficients) in the Figure 5 and is sent to the had-amard transform and quantized after 16 coefficients calculation is done.

For the 8 x 8 chroma prediction, the 8 DC values of Cb, Cr are stored in DC value register and is sent to 2 x 2 hadamard transform and quantized. Since, we remove the plane prediction mode of intra 16 x 16 and chroma 8 x 8 for its high complexity in hardware cost, the number of

	SAE calculation & DCT			Best	0		IQIDCT Operation			
count	Org input	transpose register	DCT output	selected Mode	Q input	Q output	IQ output	DCT input	IDCT output	Update
0	0,1,2,3 4,5,6,7				8 x 8 Q in(3)	8 x 8 Q out(1)	8 x 8 IQ(0)			
1	8,9,10,11 12,13,14,15			cal_SAE	8 x 8 Q in(4)	8 x 8 Q out(2)	8 x 8 IQ(1)			
2				cal_SAE	8 x 8 Q in(5)	8 x 8 Q out(3)	8 x 8 IQ(2)			
3				mode_sel	8 x 8 Q in(6)	8 x 8 Q out(4)	8 x 8 IQ(3)			
4	0,1,2,3 4,5,6,7				8 x 8 Q in(7)	8 x 8 Q out(5)	8 x 8 IQ(4)			
5	8,9,10,11 12,13,14,15	0,1,2,3 4,5,6,7	0,4,8,12 1,5,9,13		4 x 4 Q in(0)	8 x 8 Q out(6)	8 x 8 IQ(5)			
6		8,9,10,11 12,13,14,15	2,6,10,14 3,7,11,15		4 x 4 Q in(1)	8 x 8 Q out(7)	8 x 8 IQ(6)			
7						4 x 4 Q out(0)	8 x 8 IQ(7)			
8						4 x 4 Q out(1)	0,4,8,12 1,5,9,13	0,4,8,12 1,5,9,13		
9							2,6,10,14 3,7,11,15	2,6,10,14 3,7,11,15	0,1,2,3 4,5,6,7	
10					8 x 8 Q in(0)				2,6,10,14 3,7,11,15	
11					8 x 8 Q in(1)					update_v
12					8 x 8 Q in(2)	8 x 8 Q_ out(0)				update_v

 Table 1. Intra prediction 4 x 4 cycle and TQ/IQIT hardware sharing cycle

processing core for the 16 x 16 and 8 x 8 prediction is 3 as shown in the Figure 6.

The two different processing elements PE0 and PE1 are presented because DC value is calculated after determining the best mode decision and calculated by using the PE0. The degradation of image quality by the plane mode removal receives the more small effect than baseline. So, in the proposed architecture, we perform hadamard transform after the DC value calculation during the intra 16 x 16 prediction cycle. The chroma DC value is hadamard transformed after the DC value calculation during the chroma prediction. So, we can eliminate the cycle to get DC value in post processing.

As shown in the Figure 7, the intra 16 x 16 prediction take 32 cycles for SAE calculation, one cycle for best mode selection, 33 cycles for 16 x 16 DC value, 17 cycles for chroma prediction and 17 cycles for DC value of Cb, Cr each. So, the total prediction takes 118 cycles.





3. The Experimental Results

Figure 8 is the total cycle of proposed intra prediction. As shown in the Figure, it can divide to PRE processing



Figure 6. Intra prediction decision logic for 16 x 16 and chroma 8 x 8.



Figure 7. DC extraction timing diagram.



Figure 8. The total cycle of proposed intra prediction.

and POST processing. The PRE processing is a section for the intra prediction and post processing is a section for DCT, Q, IQ and IDCT after mb_type decision. In the PRE processing, it takes 210 cycles in the 4 x 4 and the POST processing takes 98 cycles for the TQ/IQIT and reconstruction performance of luma and chroma. The total prediction of one macroblock takes 308 cycles. In the post processing, the chormacoeff_cost and coeff_cost of inter macroblock is considered in the design.

Since we use the sad distortion measure and plane removal for lowering the hardware complexity, the PSNR

curves are obtained for performance comparison of the proposed architecture. We used the coding IBBPBBP (IDR15) sequence for the foreman CIF 300 frame image, and used 22, 24, 26, 28, 30, 32 and 34 for the QP value. From these experiments we get the PSNR values for the luminance and chrominance.

Table 2 shows the PSNR values for the 4 experiments. First column shows the results of JM 13.2 using hadamard distortion (ME: SATD, IP: SATD) for both motion estimator and intra prediction. Second column shows the results of JM 13.2 using the SAD for ME and SATD for intra prediction and third column shows the results in the case of using the SAD distortion for both ME and intra prediction. Last column shows the results of plane removal with SAD distortion for both ME and intra prediction.

Figure 9 shows the rate-distortion curves plotted by using the Table 2. In the Figure, the results of this experiments shows that SAD distortion measure has small PSNR degradation compared with SATD distortion measure. It is known that hadamard based rd cost calculation can provide about 0.3 dB coding quality improvement as compared with SAD based one. But in the case of high profile sequence which has many P and B picture, there is small degradation (less than 0.1 dB) of SAD distortion compare with SATD.

Second experiment shows the case of adaptation of the different distortion measure for ME and intra prediction and that the deterioration is the greatest of the all experiments. In the Figure 9, the lowest curve is the curve of 2nd experiment. For the experiment for other sequence for FULL HD and D1, we obtained the similar results.

Table 3 shows synthesized gate size. The proposed architecture has the 250,000 gate size at 160 MHz clock. Table 4 shows the comparison with other architectures.



Figure 9. The rate distortion curve (CABAC: rd-off).

Forman CIF(298fr)	ME : SATD IP : SATD		ME : SAD IP : SATD		ME : SAD IP : SAD		ME : SAD IP : SAD without plane	
QP	PSNR	bit rate	PSNR	Bit rate	PSNR	bit rate	PSNR	bit rate
22	43.07952	340961	43.0602	346572	43.0637	345447	43.0699	346116
24	41.92593	240524	41.9127	245109	41.9208	243092	41.9182	243240
26	40.89436	176870	40.8824	178102	40.8729	177876	40.8761	177949
28	39.91394	131512	39.8970	132677	39.9168	132917	39.9052	132814
30	39.12155	98157	39.1258	98257	39.1354	99028	39.1382	98504
32	38.21222	73761	38.2000	74547	38.2038	74316	38.2175	74477
34	37.37096	57025	37.3691	57330	37.3879	57198	37.3898	57088
Better PSNR	1		4		2		3(almost same as 2)	

Table 2.The PSNR values of experiments

Table 3.The synthesized gate size

	DAC 2008[17]	ICASSP 2007[18]	Proposed
CMOS Tech.	UMC 0.13µm	UMC 0.18µm	TSMC 0.18µm
Profile	H.264 High@Level4	H.264 Baseline	H.264 High@Level4
Resolution	1920x1080 @30fps	1280x720 @30fps	1920x1080 @60fps
Cycle/MB	N/A	560 Cycle/MB	308 Cycle/MB
Gate Count	164K	72 K (only I frame)	250K
Operating Frequency (for 30fps)	145MHz@1080p (30fps)	61Mhz@720p(30fps)	152MHz@1080p(60fps)

Table 4.The comparison with other architectures

Block	Gate		
intra16 x16, chroma 8 x 8 prediction	19,485		
luma 4 x 4, luma 8 x 8 prediction_ TOP	70,597		
TOP_DCTQ_IQIDCT	102,916		
Others	57,316		
Total	250,314		

The comparison shows that our proposed architecture has larger gate size compared with De-Wei^{17,18}, but has a better performance for the MB processing cycles.

4. Conclusion

Efficient hardware architecture of high profile intra prediction for H.264/AVC based on AMBA AHB bus is visualized. The existing JM13.2 software tool is not suitable to verify our encoder architecture. Hence a reference C for our architecture was developed and tested successfully.

The module is designed using the Verilog HDL, can be operated in 308 cycles for one macroblock. We reduce the hardware cost and cycles for macroblock through the plane prediction removal, SAD calculation and 8 pixel parallelisms. We also propose the DC value calculation logic which can be shared for the SAD distortion and I16MB DC value calculation and chroma 8 x 8 DC value calculation. It can be used for the pipelining of TQ and IQIT of the results from 16 x 16 intra predictions and 8 x 8 chroma predictions. Therefore, hardware reduction of prediction for calculation logic of DC value is achieved.

The proposed architecture has the 250,000 gate size and we confirmed that it can operate Full HD 1080@ 60fps at 152 MHz clock.

6. References

- 1. Huang YW, Hsieh BY, Chen TC, Chen LG. Analysis, fast algorithm and VLSI architecture design for H.264/AVC intra frame coder. IEEE Transaction on Circuits System Video Technology. 2005 Mar; 15(3):378–400.
- 2. Pan F, Lin X, Rahardja S, Lim KP, Li ZG, Wu D, Wu S. Fast mode decision algorithm for intra prediction in H.264/

AVC video coding. IEEE Transaction on Circuits System Video Technology. 2005 Jul; 15(7):813–22.

- Meng B, Au OC, Wong CW, Lam HK. Efficient intra prediction mode selection for 4 x 4 blocks in H.264. Proceeding of IEEE International Conference on Multimedia and Expo; 2003. p. 521–4.
- Suh K, Park S, Cho H. An efficient hardware architecture of intra prediction and TQ/IQIT module for H.264 encoder. ETRI Journal. 2005 Oct; 27(5):511–24.
- Meng B, Au OC, Wong CW, Lam HK. Efficient intra prediction algorithm in H.264. Proceeding of International Conference on Image Processing; 2003. p. 837–40.
- Lin YK, Chang TS. Fast mode decision algorithm for intra prediction in H.264/AVC. Proceeding of IEEE International Conference on Image Processing. 2005. p. 585–8.
- Wang TC, Huang YW, Fang HC, Chen LG. Performance analysis of hardware oriented algorithm modifications in H.264. Proceeding International Conference on Acoustic, Speech and Signal Processing; 2003. p. 493–6.
- Fu F, Lin X, Xu L. Fast intra prediction algorithm in H.264/ AVC. Proc Int Conf Signal Processing. 2004. p. 1191–4.
- Tsai AC, Wang JF, Lin WG, Yang JF. A simple and robust direction detection algorithm for fast H.264 intra prediction. Proceeding International Conference on Multimedia and Expo; 2007. p. 1587–90.
- Jafari M, Kasaei S. Adaptive search range decision for fast intraand inter-prediction mode decision in H.264/AVC. Indian Journal of Science and Technology. 2011 Sept; 4(9):1137–46.
- 11. Kim JH, Kim BG, Kim ST, Cho CS. A fast intra-mode decision algorithm for P-Slices in H.264/AVC video coding.

Proceeding International Conference on Consumer Electronics; 2007. p. 1–2.

- Kun Z, Chun Y, Qiang L, Yuzhou Z. A fast block type decision method for H.264/AVC intra prediction. Proceeding International Conference on Advanced Communication Technology; 2007. p. 673–6.
- Jung J, Kwon DN. DCT based fast 4 x 4 intra-prediction mode selection. Proceeding of IEEE Conference on Consumer Communication and Networking Conference; 2007.
- Rajeev PMS, Beulet A. An efficient hardware realization of distributed arithmetic based discrete cosine transform. Indian Journal of Science and Technology. 2015 Oct; 8(25):1–4.
- Kao YC, Chih Kuo H, Lin YT, Hou CW, Li YH, Huang HT, Lin YL. A high-performance VLSI architecture for intra prediction and mode decision in H.264/AVC video encoding. Proceeding IEEE Asia Pacific Conference Circuits and Systems; 2006. p. 562–5.
- Sahin E, Hamzaoglu I. An efficient hardware architecture for H.264 intra prediction algorithm. Proceeding of Conference on Design Automation and Test in Europe; 2007. p. 1–6.
- Lin YK, Li DW, Lin CC, Kuo TY, Wu SJ, Tai WC, Chang WC, Chang TS. A 240 mW, 10 mm² 1080 p H.264/AVC high profile encoder chip. DAC. 2008 Jun; 2008:78–83.
- Li DW, Ku CW, Cheng CC, Lin YK, Chang TS. A 61 MHz 72 K Gates 1280 x 720 30 fps H.264 intra encoder. Proceeding International Conference on Acoustic, Speech and Signal Processing; 2007 Apr. p. 801–4.