Video Prediction Block Structure and the Emerging High Efficiency Video Coding Standard

Shan Liu and Shawmin Lei
MediaTek USA Inc.
2860 Junction Avenue, San Joes, CA 95134, U.S.A.
E-mail: {shan.liu, Shawmin.lei}@mediatek.com Tel: +1 408-526-1899

Abstract— **In the ISO/IEC 14496-10 |ITU-T H.264 advanced video coding (AVC) standard, the prediction block sizes can be 16x16, 8x8 and 4x4 for Intra prediction; 16x16, 16x8, 8x16, 8x8, 8x4, 4x8 and 4x4 for Inter prediction. In the first HEVC test model (HM1.0), each 2Nx2N Intra CU may consist of either one 2Nx2N prediction unit (PU) or four NxN prediction units; while each 2Nx2N Inter CU may consist of one 2Nx2N PU, two 2NxN or Nx2N PUs or four NxN PUs. Since then, some investigations have been made to the prediction block structure based on HM1.0, including the removal of NxN prediction partition mode for all coding units (CU) except the smallest CU, and the removal of 4x4 Inter prediction. Experimental results show that with both these simplifications, the encoder complexity can be greatly reduced at the minimum cost of coding efficiency. Therefore both of these two suggestions were adopted by the HEVC standard.**

I. Introduction

Video content has continued being an increase presence in our lives. It is reported that by year 2011, over half of the Internet traffic is video. In fact, it is predicted that within the next three years online video will account for 90 percent of all Internet traffic. Modern video coding technologies allow the possibilities to store and transmit a vast amount of video data in a more efficient and effective format. Thus, the importance of digital video coding technologies has been well recognized and promoted.

Since launching the High Profile of H.264/MPEG-4 AVC, the two video coding standardization organizations namely the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG), have been actively seeking new developments in order to achieve the next major step in video compression. In early 2010, a Joint Collaborative Team on Video Coding (JCT-VC) was formed from the two groups and a joint Call for Proposals (CfP) [1] was issued. The first JCT-VC meeting was held in April 2010, Dresden, Germany, in which, 27 respondents submitted complete documentations of their proposed algorithms and systems. After reviewing the status of efforts, JCT-VC decided to move forward to the development of the new High Efficiency Video Coding (HEVC) standard. In January 2011, the first HEVC Test Model (HM1.0) [2] was released based on a set of adopted coding tools.

The new video coding standard HEVC aims at reducing half of the bit rate at the same picture quality, compared with the preceding H.264/ MPEG-4 AVC standard. In order to achieve this goal, some new coding technologies were proposed and adopted, such as using large block sizes, quart-tree coding and transform structures, etc. This paper describes several contributions since HM1.0 was formed, with a focus on prediction block structures. Some of them have been included in the Committee Draft (CD) [3] of HEVC.

This paper is organized as follows. Section II overviews the prediction block structures in prior arts. Section III describes the proposed prediction block structures for HEVC since HM1.0. Section IV reports and analyzes the experimental results on coding efficiency and complexity.

II. Prediction Block Structures in Prior Arts

A. *Prediction Block Structure in AVC*

The basic coding unit in AVC is macroblock [4]. Each macroblock is 16 pixels wide by 16 pixels high. Within each 16x16 macroblock, Intra prediction can be performed on 16x16, 8x8 and 4x4 blocks. For 4x4 and 8x8 luma Intra predictions, eight directional prediction modes and a DC prediction mode can be used. The directional modes are illustrated in Fig. 1. For 16x16 luma Intra prediction and 8x8 chroma Intra predictions, three of the above eight modes are allowed, i.e. Vertical, Horizontal and DC mode, plus another prediction mode called Plane mode.

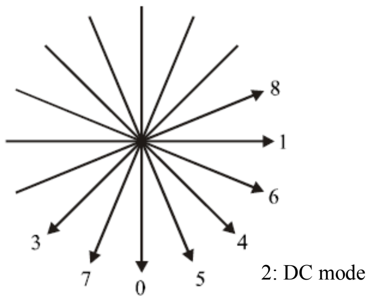

Fig. 1. AVC Intra prediction mode directions.

AVC supports seven block sizes for Inter prediction from 16x16 to 4x4 luminance samples as shown in Fig. 2. Each 16x16 luma macroblock may be split up in four ways, i.e. 16x16, 16x8, 8x16 or 8x8, as shown in Fig. 2 (a). After that, each 8x8 luma sub-macroblock may also be split up in four ways, i.e. 8x8, 8x4, 4x8 and 4x4, as shown in Fig. 2 (b). A separate motion vector or a separate Inter prediction mode is selected and transmitted for each partition. This tree-like partition structure enables the possiblity of representing areas with different characteristics differently. In general, large partitions are more suitable for more homogeneous areas while small partitions are more effective for detailed areas.

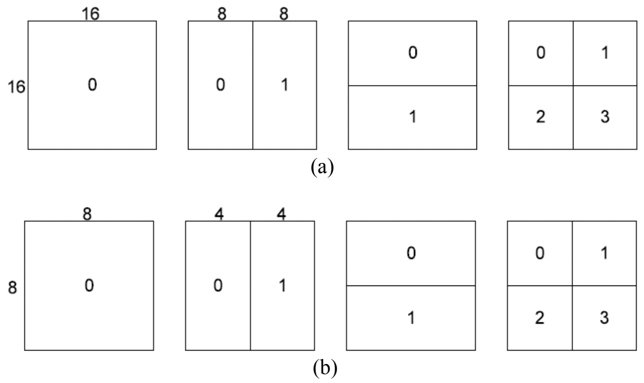

Fig. 2. (a) AVC Inter prediction block structure by 16x16 macroblock, (b) AVC Inter prediction block structure by 8x8 sub-macroblock.

B. Prediction Block Structure in HM1.0

HEVC introduced the concepts of coding unit (CU), prediction unit (PU) and transform unit (TU). By definition [5], CU is the basic video coding, processing and splitting unit. PU is the basic unit for Intra and Inter predictions. A PU can be various size or shape, but not exceeding a CU boundary. A CU can consist of one or multiple PUs. TU is the basic unit of transform. A TU size is independent from PU but cannot be bigger than the CU it belongs to. In HM1.0, a 2Nx2N Intra CU may consist of either one 2Nx2N PU or four NxN PUs; a 2Nx2N Inter CU may consist of one 2Nx2N PU, two 2NxN or Nx2N (rectangular) PUs or four NxN PUs. The prediction block structure in HM1.0 is illustrated in Fig. 3.

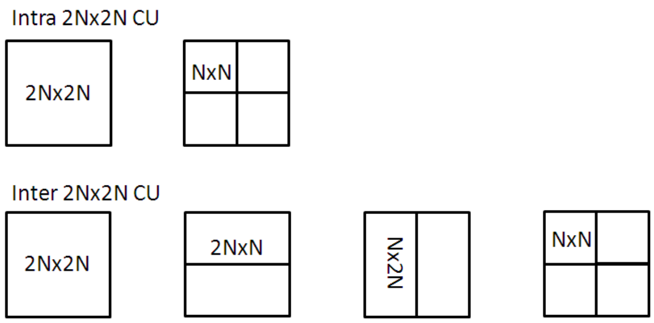

Fig. 3. Prediction block structure and PU types in HM1.0.

In HM1.0, there are 17 prediction modes (mode 0-16 in Fig. 4.) allowed for 4x4 Intra PUs, 34 modes (mode 0-33 in Fig. 4.) allowed for 8x8, 16x16 and 32x32 PUs, three modes (mode 0-2 in Fig. 4.) allowed for 64x64 and bigger PUs. Similar to AVC, a separate motion vector or a separate Inter prediction mode is selected and transmitted for each Inter PU.

III. PROPOSED PREDICTION BLOCK STRUCTURES

A. Removal of NxN Prediction Partition

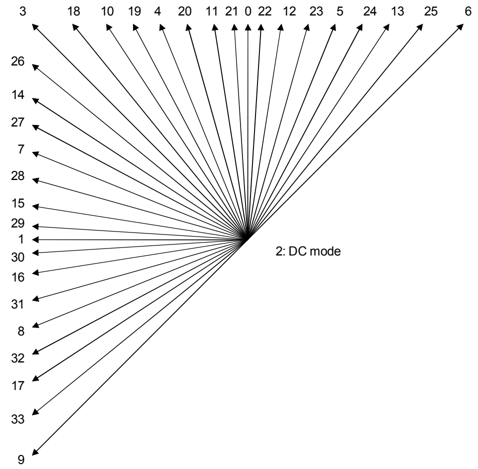

Fig. 4. Intra prediction modes in HEVC HM1.0.

HEVC defined quad-tree based CU structure, that is, a larger CU can be split to four smaller CUs, each of which has half size width and height as the larger CU. As illustrated in Fig. 5., the $2N_Kx2N_K$ CU in k[th] depth can be split into four $2N_{K+1}x2N_{K+1}$ CU in depth (k+1), where N_K equals to two times N_{K+1}. For example, under the current HM common test conditions, the largest CU (LCU) is set as 64 by 64 pixels, i.e. N=32; the smallest CU (SCU) is set as eight by eight pixels, i.e. N=4. Each CU split is signal by a split_flag.

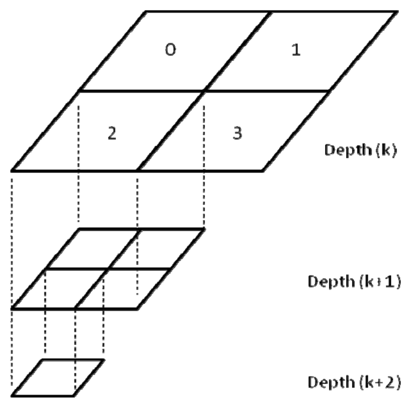

Fig. 5. Quad-tree CU structure in HEVC.

Recall that in HM1.0 a 2Nx2N CU can be split into four NxN PUs, then the prediction block size of an NxN PU in a depth(k) CU is actually the same as the prediction block size of a 2Nx2N PU in a depth(k+1) CU. Therefore, it was proposed in the 4[th] JCT-VC meeting to normatively remove the NxN PU from CUs in all depths except SCU [6]. This applies to both Intra and Inter CUs [6][7] and thus, a 2Nx2N Intra CU has only one PU size which is 2Nx2N except that an Intra SCU may also consist of four NxN PUs. Similarly, an Inter 2Nx2N CU may consist of one 2Nx2N PU, two 2NxN or Nx2N PUs in all CU depths, and possibly four NxN PUs if the CU is an SCU.

The removal of NxN PU consequently modified the codeword binarization for signaling the partition modes. As illustrated in Table I, in HM1.0, one bin is needed to signal whether 2Nx2N or NxN is selected as the partition mode for an Intra CU for all CU depths; while in the proposed scheme, this is only needed for SCU. In all other CU depths, there is no codeword required to signal an Intra PU type; the Intra prediction partition mode is inferred as 2Nx2N.

TABLE I

(a) Partition mode and binarization in HM1.0

Slice Type	Value of part_mode	Pred_Part_Mode	Bin string (HM1.0)
INTRA	0	INTRA_PART_2Nx2N	1
	1	INTRA_PART_NxN	0
INTER	0	INTER_PART_2Nx2N	1
	1	INTER_PART_2NxN	01
	2	INTER_PART_Nx2N	001
	3	INTER_PART_NxN	0001
	4	INTRA_PART_2Nx2N	00001
	5	INTRA_PART_NxN	00000

(b) Partition mode and binarization in the proposed scheme

Slice Type	Value of part_mode	PartMode	Bin string (Proposed)	
			cLog2CUSize > Log2MinCUSize	cLog2CUSize== Log2MinCUSize
INTRA	0	INTRA_PART_2Nx2N	Inferred	1
	1	INTRA_PART_NxN	n/a	0
INTER	0	INTER_PART_2Nx2N	1	1
	1	INTER_PART_2NxN	01	01
	2	INTER_PART_Nx2N	001	001
	3	INTER_PART_NxN	n/a	0001
	4	INTRA_PART_2Nx2N	000	00001
	5	INTRA_PART_NxN	n/a	00000

B. *Removal of 4x4 Inter Prediction*

During the investigation of various prediction block structures and their impacts on coding efficiency and complexity, it was observed that the 4x4 Inter prediction does not contribute much to coding efficiency improvement under HM test conditions [6]. Besides the fact that there are more powerful coding tools in HEVC compared with the preceding standards such as H.264/MPEG-4 AVC, one main reason is that the resolutions of HEVC test sequences (range from 416x240 up to 2560x1600) are much higher than some of the popular formats (e.g. CIF and QCIF) that have been used in the past. Therefore, the small size (i.e. 4x4) Inter prediction which used to bring the coding gain and belong to the core part of the preceding standard e.g. AVC, now becomes less helpful. Based on the fact that the resolutions of nowadays video contents are mostly WQVGA and above, and the belief that future video contents are moving towards higher resolutions, 4x4 inter prediction was removed from the proposed HEVC main profile. Results reported very minor impact (<0.1% BD-rate increase on average, detailed results in Section IV) on objective quality; no visible subjective quality degradation was observed under HEVC common test conditions.

Incorporated with the proposed removal of NxN partition, optional 4x4 Inter prediction, the unification of P/B prediction and partition mode binarization [10], the prediction and partition modes and binarization are finalized in HEVC Committee Draft (CD) [3] as shown in Table II.

TABLE II

Partition mode and binarization with proposed methods as in HEVC CD

PredMode	Value of part_mode	PartMode	Bin string		
			cLog2CUSize > Log2MinCUSize	cLog2CUSize = = Log2MinCUSize	
				cLog2CUSize = =3 && !inter_4x 4_enabled_flag	cLog2CUSize > 3 || inter_4x 4_enabled_flag
MODE_INTRA	0	PART_2Nx2N	-	1	1
	1	PART_NxN	-	0	0
MODE_INTER	0	PART_2Nx2N	1	1	1
	1	PART_2NxN	011	01	01
	2	PART_Nx2N	001	00	001
	3	PART_NxN	-	-	000
	4	PART_2NxnU	0100	-	-
	5	PART_2NxnD	0101	-	-
	6	PART_nLx2N	0000	-	-
	7	PART_nRx2N	0001	-	-

IV. Performance Analyses

A. *Removal of NxN Prediction Partition*

The implementation was built on top of TMuC0.9, a previous implementation that is close to HM1.0 [2]. The test conditions were as defined in JCTVC-C500 [9]. In this experiment, total 18 video sequences are tested, which are grouped into five classes by various resolutions and content characteristics. Class A contains two 4K (2560x1600) sequences, class B contains five full HD (1080p) sequences, class C contains four WVGA sequences, class D contains four WQVGA sequences and class E contains three 720p sequences with slow motions. Three configurations are tested, i.e., All Intra, Random access and Low delay. Anchor data was generated by TMuC0.9 software under the common test conditions [9], with adopted high efficiency coding tools enabled. The results for the proposed scheme, i.e. removing NxN prediction partitions for all CU depths except SCU, are compared with the anchor data and presented in average B-D rate [8] per class.

The experimental results are reported in Table III. It is shown that with the proposed method for eliminating the unnecessary redundancy during the encoder search loop, the encoding run-time is significantly reduces for all three configurations. At the same time, the impact on coding efficiency is very minor, i.e. up to 0.2% BD-rate increase on average.

TABLE III

Results for the proposed method of removing NxN partition from all CU depths except SCU

	All Intra		
	Y BD-rate	U BD-rate	V BD-rate
Class A (4K)	0.3	0.5	0.7
Class B (1080p)	0.2	0.3	0.4
Class C (WVGA)	0.2	0.2	0.3
Class D (WQVGA)	0.1	0.1	0.1
Class E (720p)	0.3	0.7	0.7
All	0.2	0.3	0.4
Enc Time[%]	72%		
Dec Time[%]	98%		

	Random access		
	Y BD-rate	U BD-rate	V BD-rate
Class A (4K)	0.2	0.2	0.2
Class B (1080p)	0.2	0.2	0.2
Class C (WVGA)	0.2	0.1	0.2
Class D (WQVGA)	0.1	-0.1	0.3

Class E (720p)			
All	0.2	0.1	0.2
Enc Time[%]	88%		
Dec Time[%]	99%		

	Low delay		
	Y BD-rate	U BD-rate	V BD-rate
Class A (4K)			
Class B (1080p)	0.1	0.1	0.0
Class C (WVGA)	0.1	0.2	0.2
Class D (WQVGA)	0.1	-0.1	0.4
Class E (720p)	0.0	0.2	0.5
All	0.1	0.1	0.3
Enc Time[%]	86%		
Dec Time[%]	98%		

B. *Removal of 4x4 Inter Prediction*

The implementation was built on top of TMuC0.9, same test conditions as described in section IV-A were used in this experiment. Anchor data was generated by TMuC0.9. The test results shown in Table IV include both removal of NxN prediction partition (except SCU) and removal of 4x4 Inter prediction. Under the common test conditions [9], the maximum CU size is 64x64 and the minimum CU size is 8x8.

Removal of 4x4 Inter prediction only affects configurations that involve Inter slices. Hence the “All Intra” results are identical to that have been shown in Table VI. The results for Random access and Low delay configurations including both removal of NxN prediction partition (except SCU) and disabling 4x4 Inter prediction are reported in Table IV. It is shown that by removing 4x4 Inter prediction, encoder search complexity can be further reduced, as reflected in reduced encoding run-time. At the same time, the impact on coding efficiency is negligible, i.e. less than 0.1% BD-rate increase on average on top of the removal of NxN.

In overall, the results show that by including both tools (i.e. removal of NxN prediction partition and disabling 4x4 Inter prediction) the encoding complexity can be significantly reduced, as in around 20-30% encoding run-time reduction, with very minor impact on coding efficiency, as in 0.2% BD-rate increase on average. Hence, both methods were adopted into the proposed HEVC main profile.

TABLE IV

Results for the proposed method of disabling 4x4 Inter prediction

	Random access		
	Y BD-rate	U BD-rate	V BD-rate
Class A (4K)	0.2	0.2	0.2
Class B (1080p)	0.2	0.2	0.2
Class C (WVGA)	0.3	0.3	0.3
Class D (WQVGA)	0.3	0.3	0.4
Class E (720p)			
All	0.2	0.3	0.3
Enc Time[%]	82%		
Dec Time[%]	99%		

	Low delay		
	Y BD-rate	U BD-rate	V BD-rate
Class A (4K)			
Class B (1080p)	0.1	0.2	0.2
Class C (WVGA)	0.2	0.0	0.1
Class D (WQVGA)	0.3	-0.4	0.3
Class E (720p)	0.1	0.1	0.3
All	0.2	0.0	0.2
Enc Time[%]	79%		
Dec Time[%]	98%		

V. Conclusion

This paper describes two proposals that are related to prediction block structure and the emerging HEVC standard. Firstly, it was proposed to remove the NxN prediction partition mode from 2Nx2N CU except for the smallest CU (SCU), experimental results reported significant encoding complexity reduction, i.e. 30% encoding run-time decrease for All Intra configuration, and 10-15% encoding run-time decrease for Random Access and Low Delay configurations. At the same time, the impact on coding efficiency is very small, i.e. average 0.2% BD-rate increase or less for all configurations. Secondly, it was proposed to remove 4x4 Inter prediction. Experimental results reported negligible impact, i.e. less than 0.1% BD-rate increase on average for Random Access and Low Delay configurations, on coding efficiency with another 8-10% encoding complexity (as in encoding run-time) reduction. These two methods combined together can significantly reduce encoding complexity (as in 20-30% encoding run-time decrease) with very small impact on coding efficiency and thus were both adopted to the proposed HEVC main profile.

References

[1] ITU-T Q6/16 Visual Coding and ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio, *Joint Call for Proposals on Video Compression Technology*, ISO/IEC JTC1/SC29/WG11/N11113, Kyoto, Japan, 22 Jan. 2010.

[2] HEVC Test Model 1.0, http://hevc.kw.bbc.co.uk/trac/browser/tags/HM-1.0, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T VCEG and ISO/IEC MPEG, Jan. 2011.

[3] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan and T. Wiegand, “High efficiency video coding (HEVC) text specification draft 6”, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T VCEG and ISO/IEC MPEG, San Jose, CA, Feb. 2012.

[4] “Advanced video coding for generic audiovisual services,” *Information Technology-Coding of Audio-Visual Objects, Part 10: Advanced Video Coding*, document ITU-T Rec. H.264, ISO/IEC 14496-10:2009, Mar. 2010.

[5] W.-J. Han, J. Min, I.-K. Kim, E. Alshina, A. Alshin, T.Lee, J. Chen, V. Seregin, S. Lee, Y. M. Hong, M.-S. Cheon, N. Shlyakhov, K. McCann, T. Davies and J.-H. Park, “Improved Video Compression Efficiency Through Flexible Unit Representation and Corresponding Extension of Coding Tools”, *IEEE Trans. Circuits and Systems for Video Technology,* Vol. 20, No. 12, pp. 1709-1720, Dec. 2010.

[6] S. Liu, Y.-W. Huang and S. Lei, “Remove Partition Size NxN”, JCTVC-D432, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T VCEG and ISO/IEC MPEG, Daegu, Korea, Jan. 2011.

[7] Frank Bossen, “Three digits to speed up the reference encoder”, JCTVC-D356, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T VCEG and ISO/IEC MPEG, Daegu, Korea, Jan. 2011.

[8] G. Bjontegaard, “Improvements of the BD-PSNR model”, ITU-T Q.6/SG16 VCEG, VCEG-AI11, Berlin, Germany, July 2008.

[9] Frank Bossen, “Common test conditions and software reference configurations”, JCTVC-C500, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T VCEG and ISO/IEC MPEG, Guangzhou, China, Oct. 2010.

[10] S. Oudin, B Bross, Y. Piao, J. Min, J. Park, X. Zhang, S. Liu, S. Lei, W.-J. Chien, J. Sole, M. Karczewicz, “Harmonization of the prediction and partitioning mode binarization of P and B slices”, JCTVC-G1042, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T VCEG and ISO/IEC MPEG, Geneva, Switzerland, Nov. 2011.