High Throughput Hardware Design for AV1 Paeth and Smooth Intra Modes

Marcel Corrêa[1,2], Bianca Waskow[1], Bruno Zatt[1],
Daniel Palomino[1], Guilherme Corrêa[1], Luciano Agostini[1]

{mmcorrea, bhwaskow, zatt, dpalomino, gcorrea, agostini}@inf.ufpel.edu.br

[1]Video Technology Research Group (ViTech), Federal University of Pelotas (UFPel), Brazil
[2]Sul-rio-grandense Federal Institute of Education, Science and Technology (IFSul), Brazil

***Abstract*—Developed by AOMedia industry consortium and released in June 2018, AV1 is an open-source and royalty-free video coding format. The main goal of AV1 is to deliver substantial compression gains over state-of-the-art codecs such as VP9 and HEVC, while keeping a practical decoding complexity, hardware feasibility and its open and free status. This paper presents a high throughput hardware architecture for four important AV1 intra prediction coding modes: Paeth, Smooth, Smooth Vertical and Smooth Horizontal. The proposed architecture was designed to support all 19 block sizes specified by AV1 and to process every single combination of these blocks according to the 10-way partition tree, with a throughput of UHD 4K (3840x2160 pixels) videos at up to 30 frames per second. When synthesized to the TSMC 40nm cell library targeting a frequency of 648MHz, the proposed design used 109.57K gates and showed a power dissipation and an energy efficiency of 16.1mW and 1.23pJ/sample respectively. No other works were found in the literature describing hardware designs for AV1 intra prediction.**

Keywords—AV1, Hardware Design, Intra Prediction, Video Coding.

I. INTRODUCTION

The continuous growth in consumption of digital videos over the internet, including services such as video-on-demand, video sharing on social networks, as well as high definition video conferences, are reaching the limits of available bandwidth of the telecommunication infrastructures. To address this situation, standardization bodies have been developing video coding standards during the last years. The most recent SIO/IEC and ITU-T standard is the High Efficiency Video Coding (HEVC) [1-2], which has many of its techniques protected by patents.

A key factor in the success of the internet is that its core technologies are open and freely implementable, and digital videos are a central part of the internet experience nowadays, consuming the majority of all the internet traffic [3], making open-source and royalty-free video formats highly desirable. This motivated the Alliance for Open Media (AOM) creation intending to generate a new highly efficient and royalty-free video encoder. The AOM is an industry consortium with more than 30 leading high-tech companies [4]. The developed encoder was called AOMedia Video 1 (AV1) [4-5] and it was released in June 2018, achieving substantial compression gains over other high-performance codecs such as its predecessor VP9 [6-7] and HEVC.

The AV1 encoder should become highly adopted in the near future since the most relevant content streaming, software and hardware companies are founding members of AOM. However, to achieve a satisfactory rate-distortion performance, a new state-of-the-art codec always brings a wide set of techniques, thus adding a considerable computational effort to the encoder and making video coding an infeasible task for software solutions when real-time processing and very high resolutions are desired, even when running on the most advanced available general-purpose processors.

The objective of this work is the hardware design of a high-throughput architecture for the AV1 Paeth, Smooth, Smooth Vertical and Smooth Horizontal intra predictors supporting all block sizes specified by the AV1 format (without any video quality degradation), targeting the processing UHD 4K (3840x2160 pixels) videos in real time.

Since the AV1 specification was very recently released, at the best of the authors' knowledge, no other published work focusing on hardware designs for the AV1 intra prediction is available in the literature. Several published works related to the intra prediction of recent codecs are available in the literature, such as [8-16], but these solutions are not compliant with AV1 and it is impossible to do a fair comparison of the current work with these previous works, because: (a) the Paeth predictor is a novel mode which was not present in any previous codec, (b) although the Smooth predictor can be considered a variation of the HEVC Planar mode, the Smooth Vertical and Horizontal predictors are not, and (c) the AV1 block partitioning structure is very different even from the most advanced previous codecs.

II. AV1 INTRA PREDICTION BACKGROUND

In AV1, a video frame is partitioned in superblocks (SBs) of size 128x128 or 64x64 pixels, according to a bitstream flag. From there, SBs are predicted in raster order within the frame. To deliver an optimal prediction, an encoder can further divide each SB using a 10-way partition tree structure [4], as illustrated in Fig. 1. In the figure, blue filled blocks are final partition modes, but all four partitions of the unfilled block can be recursively divided based on the same 10-way tree structure, down to 4x4, which is the smallest supported block size. The numbers inside each partition indicate the prediction order within the block.

Although AV1 supports many different block sizes ranging from 128x128 to 4x4 due to its flexible partition structure, only partitions of size 64x64 or smaller are supported by the intra predictors, according to the following 19 supported transform sizes: 4x4, 8x8, 16x16, 32x32, 64x64, 4x8, 8x4, 8x16, 16x8, 16x32, 32x16, 32x64, 64x32, 4x16, 16x4, 8x32, 32x8, 16x64 and 64x16 [5].

AV1 supports 56 different directional predictors, corresponding to angles between 36 and 212 degrees, aiming

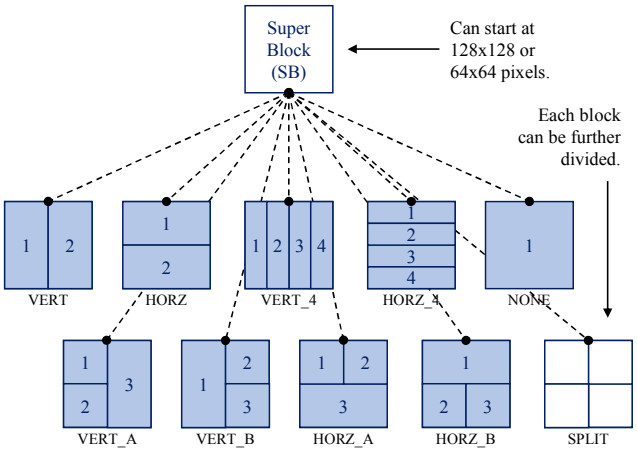

Fig. 1. AV1 10-way block partition tree structure.

at several varieties of spatial redundancies in directional textures. It also supports a variety of non-directional predictors that consider gradients, spatial correlation of samples and coherence of luminance and chrominance planes. These modes are: (i) DC, also present in other formats; (ii) Paeth, which will be better described in the next sections; (iii) Smooth, with three modes that will also be detailed next; and (iv) Chroma-from-luma, intending to use luma characteristics to encode the chroma samples. There are also two optional predictors: (i) Recursive-based-filtering (with five modes), which provide substantial gains when encoding luminance blocks at cost of an increase of data dependency and bitstream overhead, and the (ii) Pallete predictor, which is particularly efficient for screen content, also at cost of bitstream overhead.

Generally, the intra prediction of a single block, referred as a 2D array called *Pred*, requires reference samples from previously reconstructed blocks located on the left and above of the current block. The number of reference samples needed for each reference array, referred as *AboveRow* (row above the current block) and *LeftCol* (column in the left side of current block), is equal to *height*+*width*+1 (according to the block dimensions).

This work focuses on four of the non-directional predictors: Paeth, Smooth, Smooth Vertical and Smooth Horizontal. The Paeth compares the vertically aligned reference sample, the horizontally aligned reference sample and the top-left reference sample to select as the predicted sample with the lowest gradient [4]. The Smooth Vertical use linear interpolation to generate filtered samples from *LeftCol*[*height*-1] and *AboveRow*[0 to *width*-1]. The Smooth Horizontal is an analogue process using samples from *LeftCol*[0 to *height*-1] and *AboveRow*[*width*-1]. The Smooth predictor is an arithmetic average between samples from the Smooth Vertical and Smooth Horizontal predicted blocks.

The algorithms for the Smooth Vertical, Smooth Horizontal and Paeth predictors are described in Figs. 2-4. The arrays *smWeightsY* and *smWeightsX* mentioned in the first two algorithms refer to the constant coefficients ranging from 4 to 255 used in the interpolation. For example, if the size is 8, then the array is set to {255, 197, 146, 105, 73, 50, 37, 32}. For all arrays of coefficients, refer to [5]. Fig. 5. shows a hypothetical application of the mentioned predictors for a given set of reference samples.

```
FOR i in 0 to height-1:
 FOR j in 0 to width-1:
  x =  smWeightsY[i] * AboveRow[j]
  y = (256 - smWeightsY[i]) * LeftCol[h – 1]
  Pred[i][j] = (x + y + 128) / 256
 END FOR
END FOR
```

Fig. 2. AV1 Smooth Vertical algorithm for a block of any size.

```
FOR i in 0 to height-1:
 FOR j in 0 to width-1:
  x =  smWeightsX[j] * LeftCol[i]
  y = (256 - smWeightsX[j]) * AboveRow[w – 1]
  Pred[i][j] = (x + y + 128) / 256
 END FOR
END FOR
```

Fig. 3. AV1 Smooth Horizontal algorithm for a block of any size.

```
FOR i in 0 to height-1:
 FOR j in 0 to width-1:
  base = AboveRow[j] + LeftCol[i] – AboveRow[-1]
  pLeft    = ABS(base - LeftCol[i])
  pTop     = ABS(base - AboveRow[j])
  pTopLeft = ABS(base - AboveRow[-1])

  IF pLeft <= pTop AND pLeft <= pTopLeft:
   PredBlock[i][j] = LeftCol[i]
  ELSE IF pTop <= pTopLeft:
   PredBlock[i][j] = AboveRow[j]
  ELSE
   PredBlock[i][j] = AboveRow[-1]
  END IF
 END FOR
END FOR
```

Fig. 4. AV1 Paeth algorithm for a block of any size.

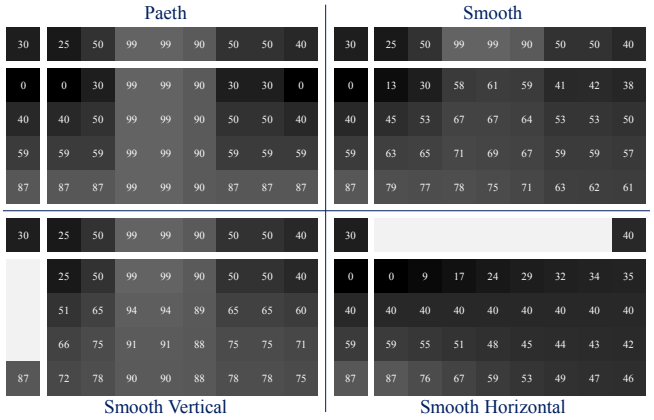

Fig. 5. Application of the predictors for an 8x4 luminance block. The colors used represent the real 8-bit luminance values.

It is important to mention the computational effort of the AV1 intra prediction when compared to previous codecs. For each 128x128 SB in AV1, there are 7,476 blocks of different sizes that must be evaluated to get the optimal combination of sub-blocks in terms of rate-distortion. For the same area, the HEVC standard allows an evaluation of 1,360 blocks of sizes 32x32, 16x16, 8x8 and 4x4. Therefore, a single AV1 intra predictor must process ~5.5 times more blocks than an HEVC predictor to achieve the same frame rate.

III. AV1 Intra Predictors Evaluation

The Paeth, Smooth, Smooth Vertical and Smooth Horizontal are among 11 non-directional predictors that can be selected for a given predicted luminance block

(chrominance blocks only use seven non-directional modes). If the optional predictors Pallete and Recursive-based-filtering are disabled, then the predictors discussed in this work represent four among five modes available for luminance prediction and four among six for modes available for chrominance prediction.

To evaluate the real relevance of the Paeth, Smooth, Smooth Vertical and Smooth Horizontal predictors in terms of compression efficiency, several software experiments were conducted using the AOMedia Codec v1.0.0 reference software. A total of 22 test sequences from the Common Test Conditions [17] were encoded with different quantization parameters (22, 32, 42 and 52). The six first frames of each video were considered in these experiments.

Table I shows the mode distribution for the encoded video sequences. For luminance samples, one can note that the Paeth, Smooth, Smooth Vertical and Smooth Horizontal predictors occur in 17.34% of the cases among 67 different modes available. For the chrominance plane, among the 62 modes, it occurs in 20.15% of the cases. This implies that this subset of predictors is very important among all available predictors for luminance and chrominance.

Fig. 6 illustrates the distribution of block sizes encoded using Paeth, Smooth, Smooth Vertical and Smooth Horizontal predictors for luminance samples. It shows that 61.4% of the blocks encoded using these modes are of the five symmetrical sizes and 38.6% are of the remaining 14 asymmetrical sizes. By far, the most signaled block size is 16x16, which is used almost 25% of the times.

The proposed hardware was designed to support all available block sizes, thus this solution does not cause any degradation in the encoding efficiency.

TABLE I. MODE DISTRIBUTION FOR ENCODED VIDEO SEQUENCES.

Class of Predictor	Number of Modes	Luma Blocks Coded (%)	Chroma Blocks Coded (%)
Directional	56	47.64	27.37
DC	1	7.58	38.82
Smooth	**3**	**13.68**	**15.39**
Paeth	**1**	**3.66**	**4.76**
Recursive-based-filtering	5	27.22	n/a
Chroma-from-luma	1	n/a	13.65
Pallete	1	0.22	0.01

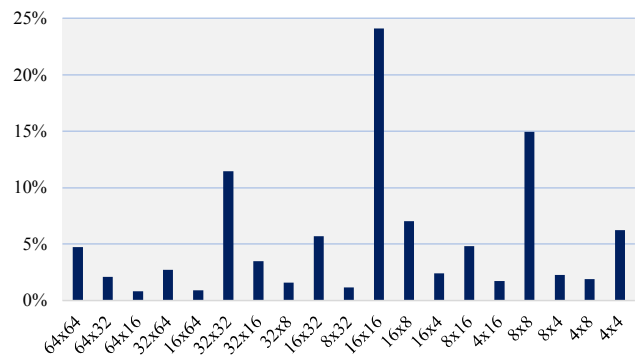

Fig. 6. Percentage of each block size among Paeth, Smooth, Smooth Vertical and Smooth Horizontal predicted blocks for luminance samples.

IV. PREDICTOR HARDWARE DESIGN

The designed architecture, presented in Fig. 7, works at the 128x128 SB level. The hardware receives as input up to 130 8-bit reference samples (*AboveRow*[-1 to 63] and *LeftCol*[-1 to 63]) associated to a full 64x64 block inside a SB, which is the biggest size supported by the AV1 intra predictors. The reference samples are stored in buffers until the entire block is predicted.

The architecture is composed of four main unities: (i) Smooth VH (Vertical), (ii) Smooth VH (Horizontal), (iii) Paeth, and (iv) Smooth. These unities are fully combinational and used a multiplierless strategy together with common subexpressions sharing and operand isolation to decrease the power dissipation, to increase the processing rates and to reduce the hardware costs. These four unities are connected by three buffers, one to store the reference samples, one to store the Smooth results and one to store the Paeth results. The Control Unit synchronizes all operations to generate the correct results at the architecture outputs.

Since the Smooth Horizontal algorithm is a transposed version of the Smooth Vertical algorithm, a single hardware was designed for both modes, which was instanced twice. The Paeth Unit and Smooth VH Unit (when processing the Vertical mode) are able to predict an entire column (regardless of the *height*) in one clock cycle. When processing the Horizontal mode, the Smooth VH Unit is able to predict an entire row in one cycle. Thus, the number of clock cycles required per predicted block is always determined by the largest of *width* and *height*.

A. Smooth VH Unit

As explained in Figs. 2-3, each predicted sample from Smooth Vertical and Smooth Horizontal modes are result of a linear interpolation between two reference samples. Each predicted sample from a column (for Vertical mode) or from a row (for Horizontal) mode has a constant coefficient associated to it ranging from 4 to 255.

To process an entire column/row in a single clock cycle, the Smooth VH Unit has ten parallel multiple constant multiplication (PMCM) units, two for each size (4, 8, 16, 32 and 64). Five of the PMCM units apply the multiplications to the reference sample that varies for every column/row (x), and five apply the multiplication to the reference sample that stays constant for the entire block (y). Each PMCM unit was built with adders, subtractors and shifts only.

Fig. 8 shows how the PMCM unit for x (size: 8) is implemented. The prediction of an entire row/column of the same size also requires PMCM for y (not shown). The prediction of the next column/row requires a new x, resulting in switching activity in the PMCM unit for x, but does not require a new y, which means that the PMCM unit for y switches only once per predicted block.

In order to prevent unnecessary switching activity, the Smooth VH Unit receives from the Control Unit the size of the current block allowing the application of operand isolation. This way, only the two PMCM units of the current size receive reference samples, while the other eight have their inputs set to zero, reducing the power dissipation.

B. Smooth Unit

As Fig. 7 shows, the intermediate results of both Smooth VH Units are organized in buffers and sent to the Smooth

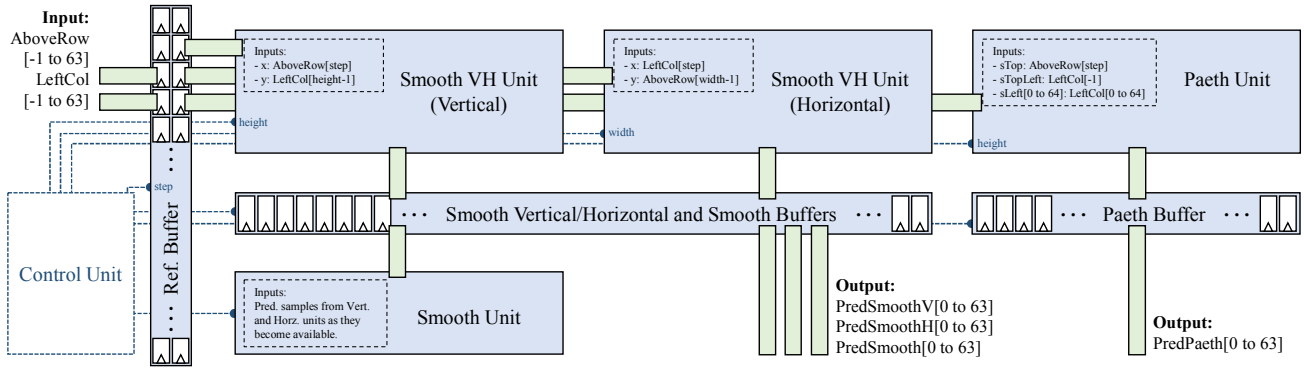

Fig. 7. Global architecture diagram.

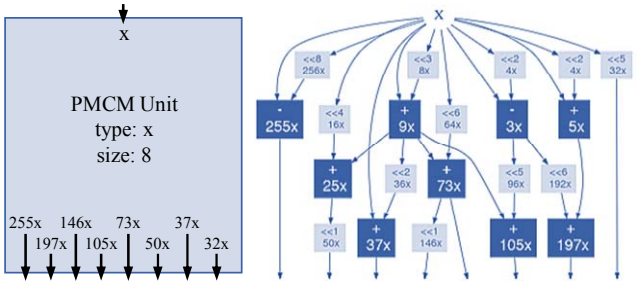

Fig. 8. Implementation of a PMCM for the *smWeightsX* of size 8, showing the efficient reuse of subexpressions.

Unit, which is a simple array of adders responsible for taking the arithmetic average of the Smooth Vertical and Horizontal predictions and to generate the Smooth prediction.

C. Paeth Unit

Basically, the Paeth Unit is able to process an entire column by first calculating (1-2), which are redundant for every predicted sample from the column, where *sTop* is the sample from *AboveRow* vertically aligned with the current column and *sTopLeft* is the sample from *LeftCol*[-1].

$$pLeft = \text{ABS}(sTop - sTopLeft) \quad (1)$$

$$pTopLeft\_temp = sTop - (sTopLeft << 1) \quad (2)$$

Secondly, by calculating 64 values of (3-4) all in parallel, where *sLeft*[*i*] is the sample horizontally aligned with the position *i* from the current column (read from *LeftCol*[0 to 63]).

$$pTop[i] = \text{ABS}(sTop - sLeft[i]) \quad (3)$$

$$pTopLeft[i] = pTopLeft\_temp + sLeft[i] \quad (4)$$

And, finally, by sending the *pLeft*, *pTop*[0 to 63] and *pTopLeft*[0 to 63] to 64 comparation units.

D. Control Unit

The Control Unit is a Finite State Machine (FSM) responsible of making all combinations of block partitions allowed by the AV1, according to the partition tree (Fig. 1). For a given partition, the FSM will send the block *height* to the Smooth VH Unit (Vertical) and to the Paeth Unit, and the block *width* to the Smooth VH Unit (Horizontal).

The global architecture is able to predict all 7,476 possible blocks inside a 128x128 SB in 28,416 clock cycles. The three target throughputs were: UHD 4K at 30 fps, FHD 1080p (1920x1080 pixels) at 60 and 30 fps. Each target requires the processing of 22,781, 11,390 and 5,695 SBs per second, respectively, leading to target frequencies of 648, 324 and 162MHz.

V. Synthesis Results

The presented architecture was synthesized to the TSMC 40nm 1.1v standard-cells library using the Cadence Encounter RTL Compiler 11.10 tool.

When synthesized to the abovementioned library, the proposed design can process UHD 4K videos at 30 fps when running at 648MHz. However, lower target frequencies were also considered. Table II shows the throughput, power dissipation (mW) and energy cost (pJ per predicted sample) for three different operation frequencies of interest. Although the power increases almost twice as the target increases, one can notice that the energy cost gets better as the target increases. For the highest target frequency, the synthesized architecture used 109.57K gates.

At the best of the authors' knowledge, there is no other published work about AV1 intra prediction hardware design, so a fair comparison cannot be done.

TABLE II. Power dissipation and energy cost for three different frequencies of operation.

Desired Throughput	Required Frequency (MHz)	Total Power (mW)	Energy Cost (pJ/sample)	Total Area (Kgates)
FHD 1080p at 30 fps	162	5.17	1.58	109.46
FHD 1080p at 60 fps	324	9.12	1.40	109.57
FHD 1080p at 120 fps	648	16.10	1.23	109.57
UHD 4K at 30 fps				

VI. Conclusions

This paper presented a high-throughput hardware architecture for the AV1 Paeth, Smooth, Smooth Vertical and Smooth Horizontal intra predictors. The solution is the first in the literature and is capable of processing UHD 4K resolutions at 30fps. Therefore, the proposed hardware can be employed in high-end devices which must encode and decode high resolution videos in real time.

Acknowledgement

This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* – Brazil (CAPES) – Finance Code 001, and also by the Brazilian research support agencies CNPq and FAPERGS.

REFERENCES

[1] G. Sullivan, J. Ohm, W. Han, T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard", IEEE Trans. Circuits Syst. Video Technol., vol. 22, pp. 1649-1668, September 2012.

[2] Information technology: High efficiency coding and media delivery in heterogeneous environments – Part 2: High efficiency video coding. ISO/IEC 23008-2. 2013.

[3] CISCO, "Cisco Visual Networking Index: Forecast and Methodology, 2016–2021," CISCO, 2017. [Online] https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf

[4] Y. Chen, et al., "An Overview of Core Coding Tools in the AV1 Video Codec," in Picture Coding Symposium, San Francisco CA, United States of America, 2018.

[5] P. Rivaz, J. Haughton, "AV1 Bitstream & Decoding Process Specification," AOMedia, 2018. [Online] https://aomedia.org/av1-bitstream-and-decoding-process-specification/

[6] D. Mukherjee, et al., "A Technical Overview of VP9 - The Latest Open-Source Video Codec," in SMPTE Annual Technical Conference & Exhibition, Hollywood CA, United States of America, 2018.

[7] A. Grange, P. Rivaz, J. Hunt, "VP9 Bitstream & Decoding Process Specification," WebM Project, 2016. [Online] https://www.webmproject.org/vp9/

[8] M. Corrêa, B. Zatt, M. Porto, L. Agostini, "High-throughput HEVC intrapicture prediction hardware design targeting UHD 8K videos," in *IEEE International Symposium on Circuits and Systems* (ISCAS), Baltimore, 2017.

[9] B. Min, Z. Xu, R. Cheung, "A Fully Pipelined Hardware Architecture for Intra Prediction of HEVC," IEEE Trans. Circuits Syst. Video Technol., vol. 27, no. 12, pp. 2702-2713, July 2016.

[10] H. Fang, H. Chen, T. Chang, "Fast intra prediction algorithm and design for high efficiency video coding," in *IEEE International Symposium on Circuits and Systems* (ISCAS), Montreal, 2016.

[11] G. Pastuszak, A. Abramowski, "Algorithm and architecture design of the H.265/HEVC intra encoder," IEEE Trans. Circuits Syst. Video Technol., vol. 26, no. 1, pp. 210-222, May 2015.

[12] J. Zhu, Z. Liu, D. Wang, Q. Han, Y. Song, "HDTV1080p HEVC intra encoder with source texture based CU/PU mode pre-decision," in *19th Asia and South Pacific Design Automation Conferece* (ASP-DAC), Singapore, 2014.

[13] C. Liu, W. Shen, T. Ma, Y. Fan, "A highly pipelined VLSI architecture for all modes and block size intra prediction in HEVC encoder," in *IEEE 10th International Conference on ASIC* (ASICON), Shenzhen, 2013.

[14] M. Zhou, D. Ding, L. Yu, "On hardware architecture and processing order of HEVC intra prediction module," in *Picture Coding Symposium*, San Jose, 2013.

[15] D. Palomino, F. Sampaio, L. Agostini, S. Bampi, A. Susin, "A memory aware and multiplierless VLSI architecture for the complete intra prediction of the HEVC emerging standard," in *IEEE 19th International Conference on Image Processing* (ICIP), Orlando, 2012.

[16] S. Tsai, C. Li, H. Chen, P. Tsung, K. Chen, L. Chen, "A 1062Mpixels/s 8192x4320p high efficiency video coding (H.265) encoder chip," in *Symposium on VLSI Circuits* (VLSIC), Kyoto, 2013.

[17] JCT-VC, "JCTVC-L1100 Common test conditions and software configurations," JCT-VC Meeting, Geneva, January 2013. [Online] http://phenix.int-evry.fr/jct/