A High Throughput Hardware Architecture Targeting the AV1 Paeth Intra Predictor

Marcel Corrêa, Bianca Waskow, Jones Goebel, Daniel Palomino, *Member, IEEE*
Guilherme Corrêa, *Member, IEEE*, Luciano Agostini, *Senior Member, IEEE*

***Abstract*— AV1 is an open-source and royalty-free video coding format, which was developed by the AOMedia industry consortium and released in June 2018 as the state-of-the-art in video coding. The main goal of AV1 development was to achieve substantial compression gain over high-performance codecs such as VP9 and HEVC, while keeping a practical decoding complexity, hardware feasibility and its open and free status. This paper presents a highly parallelized hardware architecture for the AV1 Paeth intra predictor supporting all 19 block sizes allowed, capable of processing UHD 4K videos at 120 frames per second. When synthesized to the TSMC 40nm cell library targeting a frequency of 315MHz, the proposed design used 247.28K gates and showed a power dissipation and an energy efficiency of 268.36mW and 179.74pJ/sample respectively.**

***Index Terms*—AV1, Hardware Design, Intra Prediction, Paeth Mode, 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 for many years, the most recent being 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]. Hence, open-source and royalty-free video formats are highly desirable. This motivated the Alliance for Open Media (AOMedia) creation intending to generate a new highly efficient and royalty-free video encoder. The AOMedia is an industry consortium of more than 30 leading high-tech companies [4], such as Google, Apple, Amazon, Cisco, Facebook, Netflix, Intel, and others. 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 [1-2].

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 AOMedia. 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 complexity 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 main objective of this work is the hardware design of a highly parallel architecture for the AV1 Paeth intra predictor supporting all block sizes specified by the AV1 format (without any video quality degradation), targeting the processing of UHD 8K (7680x4320 pixels) and 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. Therefore, this is the first work to propose a hardware design for the AV1 Paeth intra predictor. Even with several published works about HEVC and VP9 intra predictors, a fair comparison is not possible, because the Paeth predictor is a novel mode which was not present in any previous codecs.

This paper is organized as follows: Section II presents the AV1 intra prediction basics and the Paeth predictor algorithm, section III presents software experiments to evaluate the relevance of the Paeth predictor, section IV proposes the hardware architecture for the Paeth predictor, section V presents synthesis results, and finally, section VI concludes this

Paper submitted for review on September 28th, 2018.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and also by CNPq and FAPERGS Brazilian research support agencies. The authors would like to thank the support received for the developed work.

M. Corrêa is a professor at the Federal Institute of Education, Science and Technology Sul-rio-grandense, Bagé, Brazil (IFSul) (e-mail: marcelcorrea@ifsul.edu.br).

D. Palomino, G. Corrêa and L. Agostini are professors at the Federal University of Pelotas (UFPel), Brazil (e-mail: {dpalomino, gcorrea, agostini}@inf.ufpel.edu.br).

M. Corrêa, J. Goebel and B. Waskow are, respectively, doctorate, master and undergraduate students at the Federal University of Pelotas (UFPel), Brazil (e-mail: {mmcorrea, bhwaskow, jwgoebel}@inf.ufpel.edu.br).

The authors are members of the Video Technology Research Group (ViTech) of the Postgraduate Program in Computation (PPGC) at UFPel (https://wp.ufpel.edu.br/vitech/en/introduction/).

978-1-7281-0453-9/19/$31.00 ©2019 IEEE

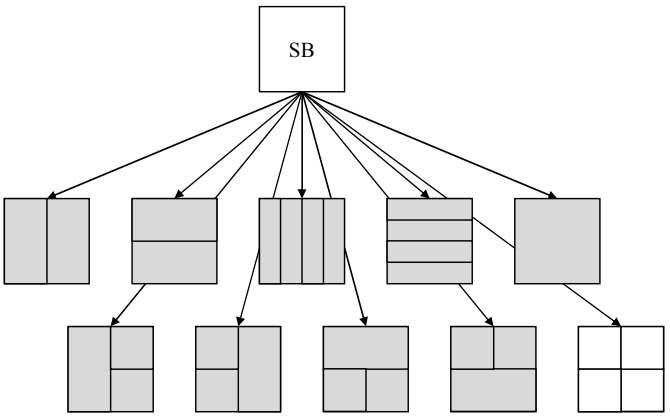

Fig. 1. AV1 10-way block partition tree structure.

paper.

II. AV1 Intra Prediction

In AV1, a frame is partitioned in superblocks (SBs) of size 128x128 or 64x64 pixels. From there, SBs can be further partitioned using a 10-way partition tree structure [4], as illustrated in Fig. 1. In the figure, gray partition modes are final, but all four blocks of the white partition mode can be further divided based on the same 10-way tree structure, down to 4x4, which is the smallest size allowed.

The AV1 intra prediction process is invoked for intra blocks to predict a part of the block corresponding to a transform block. When the transform size is smaller than the block size, this process is invoked multiple times in raster order within a single block, allowing the next transform block to use the reconstruction as a better reference. 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, 64x16} [5].

Generally, the prediction of a single block, referred as a 2D array called *PredBlock*, 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* and *LeftCol*, is equal to *height*+*width*+1 (the same diagonally aligned sample appears in both arrays). Fig. 2 illustrates a 4x8 block (white squares) to be predicted using eight reference samples from left neighboring transform block(s), eight reference samples from above neighboring transform block(s) and a single reference sample from an above left neighboring transform block.

AV1 supports 56 different directional modes for intra prediction, corresponding to angles between 36 and 212 degrees, aiming at several varieties of spatial redundancies in directional textures. It also supports a variety of non-directional modes that consider gradients, spatial correlation of samples and coherence of luminance and chrominance planes, including two modes developed particularly for screen content [4].

This work focuses on one of the non-directional modes, called Paeth predictor, which compares the vertically aligned

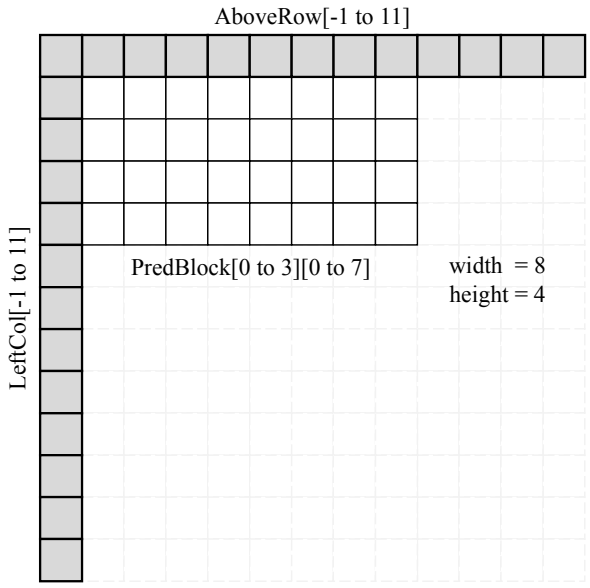

Fig. 2. A 4x8 block to be predicted by using 25 reference samples from left neighboring transform blocks.

reference sample, the horizontally aligned reference sample and the top-left reference sample, and selects as the predicted sample the reference from the direction with the lowest gradient [4]. From the set of reference samples shown in Fig. 2, the Paeth predictor requires only the subset composed of *AboveRow*[-1 to *width*-1] and *LeftCol*[-1 to *height*-1].

The algorithm (Fig. 3) for a given predicted block, according to [5], is as follows: for each predicted sample, firstly, a *base* value measures how far the intensities of the top and left samples together are from the top-left sample. Secondly, values *pLeft*, *pTop* and *pTopLeft* measure the differences of each reference between the *base* value. Finally, the reference with the smallest absolute difference is selected as predicted sample.

Fig. 4 shows an example of the Paeth prediction for a 4x8 block with hypothetical 8-bit reference samples. The *PredBlock*[1][1] is predicted as 50, because *base*=60, *pLeft*=20, *pTop*=0 and *pTopLeft*=20, which leads to the selection of the top reference sample as predicted sample.

```
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. 3. Paeth prediction algorithm for a block of any size supported by the AV1 format.

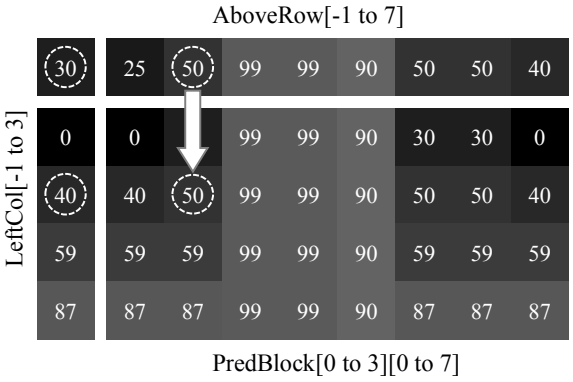

Fig. 4. Example of the Paeth prediction for a 4x8 block with hypothetical 8-bit reference samples.

It is important to note the complexity of the AV1 intra prediction when compared to previous codecs. In AV1, for each 128x128 SB, there are 3,396 blocks of different sizes (total of 311,296 samples) that must be evaluated to get the optimal combination of sub-blocks in terms of rate-distortion. The HEVC standard allows evaluation of 1,360 blocks of size {32x32, 16x16, 8x8, 4x4} (total of 65,536 samples). Therefore, a single AV1 intra predictor must process 4.75 times more samples than an HEVC predictor, in order to achieve the same frame rate.

III. Software Experiments

The Paeth prediction mode is one among 56 directional and 11 non-directional modes that can be selected for a given predicted luminance block (chrominance blocks use seven non-directional modes). To evaluate the relevance of the Paeth mode in terms of efficiency, several software experiments were conducted using the AOMedia Codec v1.0.0 reference software [5]. A total of 22 test sequences from the Common Test Conditions [8] were encoded with different quantization parameters (QP=22, 32, 42 and 52). Six frames for 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 prediction occurs in 3.66% of the cases among 67 different modes available. For the chrominance plane, among the 62 modes, it occurs in 4.76% of the cases. This implies that this mode is one of the most used modes among all available ones for luminance and chrominance

TABLE I. Mode Distribution for Encoded Video Sequences.

Class of Predictor	Number of Modes	Luminance Blocks Coded (%)	Chrominance Blocks Coded (%)
Directional	56	47.64	27.37
Smooth	3	13.68	15.39
DC	1	7.58	38.82
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

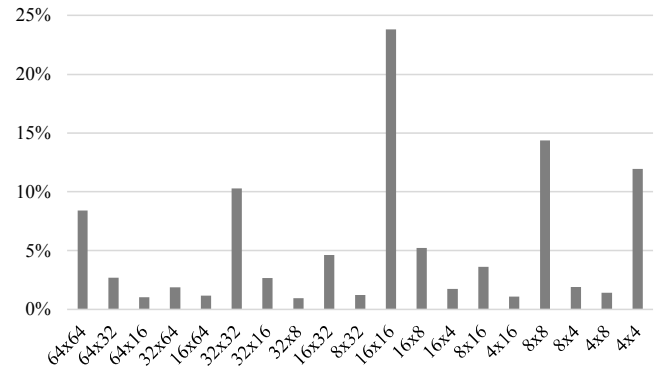

Fig. 5. Percentage of each block size among Paeth predicted blocks for luminance samples.

Fig. 5 illustrates the distribution of blocks encoded using Paeth mode for luminance samples according to their sizes. It shows that 68.81% of the blocks encoded using Paeth are of the five symmetrical sizes and 31.19% are of the remaining 14 asymmetrical sizes. The most signaled block size is 16x16, which is used almost a quarter of the times. Another important consideration is that larger blocks require a larger computational effort to be processed, since they are composed of a larger number of samples. Thus, a hardware architecture must have a throughput high enough to deal in real time with 64x64 blocks.

IV. Proposed Paeth Architecture

The proposed architecture is presented in Fig. 6 and it works at the 128x128 SB level. For each SB in the video stream, a total of 3,396 blocks of different symmetrical and asymmetrical sizes (listed in Fig. 5) must be processed.

The design receives as input 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 in the AV1 intra prediction.

Since the AV1 algorithm has redundant calculations regarding the *base*, *pLeft* and *pTop* values, our design calculates, in parallel, only 32 *pTop* values (one for each line) and 32 *pLeft* values (one for each column). The *pTopLeft* values are exclusive for each position in the predicted block, thus 1,024 *pTopLeft* values are calculated in parallel.

Internally, the architecture has a matrix of 32 by 32 Dedicated Prediction Units (DPUs), as presented in Fig. 7, where each DPU compares the respective *pLeft*, *pTop* and *pTopLeft* values and selects the best reference as the predicted sample accordingly.

This high parallelism allows the processing of 32x32 blocks in a single clock cycle. Blocks of size 64x64, 64x32 and 32x64 must be processed in four, two and two clock cycles respectively. In case of blocks bigger than 32x32, the Sample Selectors switch between the first and second half of the reference arrays. Blocks smaller than 32x32 can also be processed in a single clock cycle by using only a subset of the DPU matrix. In case of sub utilization of the DPU matrix, an operand isolation technique is applied on the inputs.

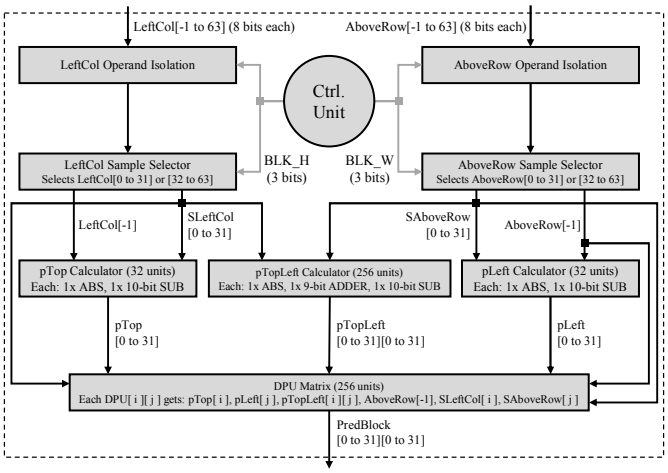

Fig. 6. High level block diagram of the proposed Paeth predictor design.

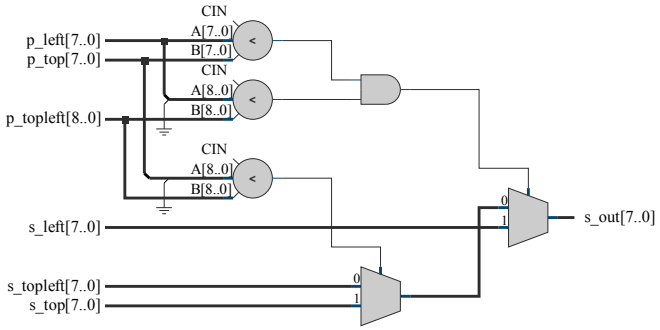

Fig. 7. Paeth Dedicated Prediction Unit (DPU) for a single predicted sample.

The prediction of all 3,396 blocks inside an 128x128 SB takes 3,456 clock cycles. Five target throughputs were considered based on technological tendencies: UHD 4K at 120, 60 and 30 frames per second (fps) and FHD 1080p (1920x1080 pixels) at 60 and 30 fps. Each target requires the processing of 91,125, 45,563, 22,781, 11,390 and 5,695 SBs per second, respectively, leading to frequencies of 315, 158, 79, 40 and 20MHz

V. Synthesis Results

The developed Paeth prediction architecture was fully described in VHDL, synthesized to the TSMC 40nm 1.1v standard-cells library using the Cadence Encounter RTL Compiler 11.10 tool.

When synthesized to the abovementioned cell library, the proposed design can achieve a very high-throughput, processing UHD 8K videos at 30 frames per second (fps) or UHD 4K videos at 120 fps when running at 315MHz. However, lower target frequencies were also considered, as FHD 1080p videos are still very relevant. Table II shows the throughput, power dissipation and energy efficiency for five different operation frequencies of interest. Although the power increases almost twice as the target increases, it can be seen that the energy efficiency, measured in pJ per sample predicted, gets better as the target increases.

For the highest target frequency, the synthesized architecture used 247.28K gates and, for the four lowest frequencies, 241.36K gates were required.

TABLE II. Power dissipation and energy efficiency for five different frequencies of operation.

Desired Throughput	Required Frequency (MHz)	Total Power (mW)	Energy Efficiency (pJ/sample)
FHD 1080p at 30 fps	20	23.71	179.74
FHD 1080p at 60 fps	40	44.29	208.11
FHD 1080p at 120 fps	79	79.58	213.20
UHD 4K at 30 fps			
UHD 4K at 60 fps	158	155.35	237.35
UHD 4K at 120 fps	315	268.36	254.09
UHD 8K at 30 fps			

VI. Conclusions

This paper presented a highly parallelized and energy-aware hardware architecture for the AV1 Paeth intra predictor. The designed solution is the first in the literature and is capable of processing UHD 4K resolutions at 120fps. Therefore, the proposed hardware can be employed in applications targeting UHDTV videos, including video servers which must encode and decode high resolution videos in real time.

Future works include the design of hardware architectures for exclusive features of the AV1 intra prediction, such as: (1) a special non-directional predictor called Recursive-based-filtering, which leads to very high rate-distortion performance but introduces data dependencies among predicted samples, posing a big challenge for real-time processing of high resolutions and frame rates, and (2) the smoothing and upscaling filters of directional predictors, which tend to be a bottleneck when processing directional modes.

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